



**HAL**  
open science

**Phytoplancton des plans d'eau d'Outre-Mer:  
développement d'outils de monitoring basés sur l'ADN  
Livrable 3: " Application de l'approche ADNe pour  
recenser le phytoplancton des plans d'eau d'Outre-Mer:  
méthodes et premiers résultats "**

Alexis Canino, Benjamin Alric, Maria-Mercedes Nicolosi-Gelis, Christophe Laplace-Treyture, Agnes Bouchez, Isabelle Domaizon, Frédéric Rimet

► **To cite this version:**

Alexis Canino, Benjamin Alric, Maria-Mercedes Nicolosi-Gelis, Christophe Laplace-Treyture, Agnes Bouchez, et al.. Phytoplancton des plans d'eau d'Outre-Mer: développement d'outils de monitoring basés sur l'ADN Livrable 3: " Application de l'approche ADNe pour recenser le phytoplancton des plans d'eau d'Outre-Mer: méthodes et premiers résultats ". INRAE UMR CARRETEL; INRAE UR EABX. 2023. hal-04116746

**HAL Id: hal-04116746**

**<https://hal.inrae.fr/hal-04116746>**

Submitted on 5 Jun 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Phytoplancton des plans d'eau d'Outre-Mer: développement d'outils de monitoring basés sur l'ADN

**Livrable 3** : « Application de l'approche ADNe pour recenser le phytoplancton  
des plans d'eau d'Outre-Mer: méthodes et premiers résultats »

CANINO Alexis<sup>1,3</sup>, ALRIC Benjamin<sup>1,3</sup>, NICOLOSI-GELIS Maria Mercedes<sup>1,3,4</sup>, LAPLACE-TREYTURE  
Christophe<sup>2,3</sup>, BOUCHEZ Agnès<sup>1,3</sup>, DOMAIZON Isabelle<sup>1,3</sup>, RIMET Frédéric<sup>1,3</sup>

<sup>1</sup>INRAE – UMR CARTELE, F-74500 Thonon-les-Bains, France

<sup>2</sup>INRAE – UR EABX, F-33612 Cestas, France

<sup>3</sup>Pôle R&D ECLA

<sup>4</sup>ILPLA-CONICET-UNLP, Argentina

Février 2023



## TABLE DES MATIÈRES

I.	Contexte.....	1
I.1.	Rappels des travaux réalisés sur la période de 2020 à 2021 .....	1
I.2.	Principaux objectifs pour l'année 2022 .....	2
II.	Campagnes d'échantillonnage.....	3
II.1.	Campagne d'échantillonnage de l'année 2021 .....	3
II.2.	Campagne d'échantillonnage de l'année 2022 .....	4
III.	Traitement des échantillons au laboratoire.....	5
III.1.	Traitement des échantillons environnementaux.....	5
III.2.	Constitution d'échantillons synthétiques .....	7
III.2.1.	Souches phytoplanctoniques utilisées .....	7
III.2.2.	Test de différentes méthodes d'extraction d'ADN pour la constitution des échantillons synthétiques .....	7
III.2.3.	Constitution des communautés synthétiques à partir d'amplicons d'ADN .....	10
III.3.	Amplification PCR et séquençage des régions d'intérêt.....	10
III.4.	Optimisation des protocoles de laboratoire.....	11
III.4.1.	Optimisation de l'approche « Chelex » .....	11
III.4.2.	Optimisation du protocole d'extraction d'ADN à partir de Sterivex utilisant l'automate MagnetaPure32 .....	12
IV.	Développements bioinformatiques .....	13
IV.1.	Mise à jour de l'application Phytool .....	13
IV.2.	Traitement des données de séquençage.....	14
IV.3.	Optimisation de l'analyse des données .....	14
V.	Comparaison des inventaires obtenus en microscopie et en metabarcoding .....	16
V.1.	Échantillons synthétiques .....	16
V.2.	Échantillons environnementaux .....	18
V.2.1.	Ensemble des échantillons : lacs des DROM et lacs péri-alpins .....	18
V.2.2.	Échantillons des Lacs péri-alpins (OLA) .....	24
V.2.3.	Échantillons des Lacs des DROM .....	28
VI.	Tests d'application d'une métrique de bio-indication : MCS de l'IPLAC.....	34
VII.	Références.....	36
VIII.	Annexes.....	38

## LISTES DES FIGURES

- Figure 1.** Photographie de quelques Sterivex en provenance de La Réunion avant l'extraction. .... 5
- Figure 2.** Résumés des protocoles d'extraction d'ADN utilisés pour les échantillons d'eau filtrés sur Sterivex avec 2mL de tampon de lyse (panneau du haut) et sur filtres (panneau du bas). .... 6
- Figure 3.** Espèces de la TCC utilisées dans les échantillons synthétiques en 2021 (« Experiment 1 ») et en 2022 (« Experiment 2 »). L'arbre présente les liens taxonomiques des différentes espèces. .... 7
- Figure 4.** Présentation synthétique des différentes conditions d'extraction testées sur les souches contrôles pour constituer les différentes communautés synthétiques. Différents types de mélanges ont ensuite été constitués (équimolaires, exponentiels et en gradient) et sont détaillés dans la section III.2.3. .... 8
- Figure 5.** Résumé des protocoles d'extraction d'ADN « GEN » et « FIL » (panneau du haut) et par l'approche single-cell « CHE » (panneau du bas). .... 9
- Figure 6.** Photographie de cellules d'*Haematococcus lacustris*, restées intactes après la lyse mécanique du protocole Chelex. .... 12
- Figure 7.** Présentation schématique du fonctionnement de l'algorithme utilisé pour ré-assigner les séquences environnementales en s'appuyant sur les suivis microscopiques. .... 15
- Figure 8.** Corrélation entre la quantité d'ADN introduite et le nombre de reads obtenus pour les 18 espèces dans les 3 échantillons synthétiques (mélanges équimolaire, gradient et exponentiel) pour le marqueur 16S (haut) et 23S (bas) et pour chacune des 4 méthodes d'extraction utilisées (automate - AUT, Chelex - CHE, filtre - FIL, GenElute - GEN). Les résultats sont exprimés en log des abondances relatives de reads et d'ADN d'amplicon par espèce. En bleu : régression sur l'ensemble des données ; en rouge : régression sur un sous-ensemble des données excluant les valeurs extrêmes ; en pointillés : régression 1:1. Les coefficients de corrélations de Pearson sont indiqués ainsi que la significativité de leur p-value (\* <0.01 ; . <0.05). .... 17
- Figure 9.** Relation entre les valeurs de diversité locale (ou alpha-diversité) calculées à partir des données de microscopie et moléculaires (marqueurs 16S et 23S). a) La relation est estimée entre les données de microscopie et les données du marqueur 16S. b) La relation est estimée en considérant les données de microscopies et les données du marqueur 23S. Chaque point noir correspond à un échantillon des DROM ou de métropole. L'alpha-diversité est exprimée selon l'indice de Shannon. Les relations ont été estimées par le calcul du coefficient de corrélation non paramétrique de Spearman. .... 19
- Figure 10.** Analyse en coordonnées principales (PCoA) basée sur une matrice de distance de Bray-Curtis calculée sur les données de microscopie obtenues sur l'ensemble des échantillons provenant des DROM (i.e., Guadeloupe, Guyane et Réunion) et de métropole (i.e., OLA). Chaque point correspond à un échantillon et la couleur à sa provenance : Guadeloupe (rouge), Guyane (jaune), Réunion (bleu), et métropole (vert, lacs OLA). Les échantillons sont projetés sur le plan factoriel formé par les deux premiers axes de l'analyse (% d'inertie expliquée : Axe 1 : 15.58%, Axe 2 : 10.54%). Sur le plan factoriel de l'analyse sont également projetés les 10 taxons les plus abondants dans l'ensemble des échantillons. .... 20
- Figure 11.** Analyse en coordonnées principales (PCoA) basée sur une matrice de distance de Bray-Curtis calculée sur les données moléculaire (marqueur 16S) obtenues sur l'ensemble des échantillons provenant des DROM (i.e., Guadeloupe, Guyane et Réunion) et de métropole (i.e., OLA). Les échantillons sont projetés sur le plan factoriel formé par les deux premiers axes de l'analyse (% d'inertie expliquée : Axe 1 : 26.72%, Axe 2 : 13.48%). Les 10 taxons les plus abondants dans l'ensemble des échantillons sont également projetés sur le plan factoriel de l'analyse. Voir la figure 10 pour plus de détail sur les légendes. .... 21
- Figure 12.** Analyse en coordonnées principales (PCoA) basée sur une matrice de distance de Bray-Curtis calculée sur les données moléculaire (marqueur 23S) obtenues sur l'ensemble des échantillons provenant des DROM (i.e., Guadeloupe, Guyane et Réunion) et de métropole (i.e., OLA). Les échantillons sont projetés sur le plan factoriel formé par les deux premiers axes de l'analyse (% d'inertie expliquée : Axe 1 : 22.39%, Axe 2 : 13.77%). Les 10 taxons les plus abondants dans l'ensemble des échantillons sont également projetés sur le plan factoriel de l'analyse. Voir la figure 10 pour plus de détail sur les légendes. .... 22
- Figure 13.** Analyse de co-inertie décrivant la co-structure entre les assemblages phytoplanctoniques estimés par ACP sur les données de microscopie et moléculaire (16S). Projection des échantillons sur les deux premiers axes de l'analyse. Pour chaque axe, les valeurs entre parenthèses correspondent au pourcentage d'inertie

expliquée. La longueur des flèches indique la différence de composition entre les données morphologiques (les points des flèches) et les données moléculaires (les pointes des flèches). La couleur des points correspond à l'origine des échantillons. ....	23
<b>Figure 14.</b> Analyse de co-inertie décrivant la co-structure entre les assemblages phytoplanctoniques estimés par ACP sur les données de microscopie et moléculaire (23S). Projection des échantillons sur les deux premiers axes de l'analyse. Pour chaque axe, les valeurs entre parenthèses correspondent au pourcentage d'inertie expliquée. La longueur des flèches indique la différence de composition entre les données morphologiques (les points des flèches) et les données moléculaires (les pointes des flèches). La couleur des points correspond à l'origine des échantillons. ....	24
<b>Figure 15.</b> Diagrammes de Venn comparant les taxons identifiés par les trois approches (16S : bleu, 23S : jaune et microscopie : vert). Les différents niveaux taxonomiques sont considérés du phylum à l'espèce. Le nombre (en gras) et le pourcentage (%) de taxons sont indiqués. ....	25
<b>Figure 16.</b> Analyse de co-inertie entre les assemblages phytoplanctoniques estimés sur les données de microscopie et moléculaire avec le marqueur 16S. La longueur des flèches indique la différence de composition entre les données morphologiques (points des flèches) et les données moléculaires (pointes des flèches). Chaque point correspond à un échantillon et sa couleur à sa provenance : Aiguebelette (vert), Annecy (jaune), Bourget (rouge), et Léman (bleu). ....	27
<b>Figure 17.</b> Analyse de co-inertie entre les assemblages phytoplanctoniques estimés par les données de microscopie et moléculaire avec le marqueur 23S. La longueur des flèches indique la différence de composition entre les données morphologiques (points des flèches) et les données moléculaires (pointes des flèches). Chaque point correspond à un échantillon et sa couleur à sa provenance : Aiguebelette (vert), Annecy (jaune), Bourget (rouge), et Léman (bleu). ....	27
<b>Figure 18.</b> Analyse en coordonnées principales (PCoA) basée sur une matrice de distance de Bray-Curtis calculée sur les données de microscopie obtenues à partir des échantillons prélevés dans les 4 sites de la Réunion : Gol (rouge), Saint Andre (jaune), Saint Paul (bleu), Sainte Suzanne (vert). ....	29
<b>Figure 19.</b> Analyse en coordonnées principales (PCoA) basée sur une matrice de distance de Bray-Curtis calculée sur les données moléculaires (marqueur 16S) obtenues à partir des échantillons prélevés dans les 4 sites de la Réunion : Gol (rouge), Saint Andre (jaune), Saint Paul (bleu), Sainte Suzanne (vert). ....	30
<b>Figure 20.</b> Analyse en coordonnées principales (PCoA) basée sur une matrice de distance de Bray-Curtis calculée sur les données moléculaires (marqueur 23S) obtenues à partir des échantillons prélevés dans les 4 sites de la Réunion : Gol (rouge), Saint Andre (jaune), Saint Paul (bleu), Sainte Suzanne (vert). ....	31
<b>Figure 21.</b> Analyse de co-inertie décrivant la co-structure entre les assemblages phytoplanctoniques estimés par des ACP sur les données de microscopie et moléculaires (16S). La longueur des flèches indique la différence de composition entre les données morphologiques (les points des flèches) et les données moléculaires (les pointes des flèches). Chaque point correspond à un échantillon et la couleur à sa provenance : Gol (rouge), Saint Andre (jaune), Saint Paul (bleu), et Sainte Suzanne (vert). ....	32
<b>Figure 22.</b> Analyse de co-inertie décrivant la co-structure entre les assemblages phytoplanctoniques estimés par des ACP sur les données de microscopie et moléculaires (23S). La longueur des flèches indique la différence de composition entre les données morphologiques (les points des flèches) et les données moléculaires (les pointes des flèches). Chaque point correspond à un échantillon et la couleur à sa provenance : Gol (rouge), Saint Andre (jaune), Saint Paul (bleu), et Sainte Suzanne (vert). ....	33
<b>Figure 23.</b> Comparaison des valeurs de MCS obtenues avec la microscopie et l'approche moléculaire 16S. Aucune corrélation significative n'a été obtenue. ....	34
<b>Figure 24.</b> Comparaison des valeurs de MCS obtenues avec la microscopie et l'approche moléculaire 23S. Aucune corrélation significative n'a été obtenue. ....	35

## LISTES DES TABLEAUX

<b>Tableau 1.</b> Liste des masses d'eau des DROM et de métropole échantillonnées lors des campagnes 2021. ....	3
<b>Tableau 2.</b> Synthèse des masses d'eau de métropole suivis à partir de la campagne d'échantillonnage de 2022. .....	4
<b>Tableau 3.</b> Récapitulatif des amorces utilisées pour amplifier la région v3-v4 de l'ARNr 16S et la région UPA de l'ARNr 23S.....	11
<b>Tableau 4.</b> Comparaison des dix taxons les plus abondants détectés par les approches morphologiques et moléculaires. ....	26

# I. Contexte

## I.1. Rappels des travaux réalisés sur la période de 2020 à 2021

Le projet a démarré au second semestre de 2020. Trois objectifs ont été identifiés pour 2020 :

- Le 1<sup>er</sup> objectif a été lié à l'harmonisation de la taxonomie du phytoplancton afin de pouvoir : (1) comparer les inventaires floristiques des DROM entre eux, (2) comparer le contenu des différentes bases de références de barcodes contenant différents marqueurs génétiques (et permettre ainsi une comparabilité des tests *in silico* mais aussi du recouvrement des espèces par chacun des marqueurs génétiques), (3) comparer les inventaires microscopiques avec les inventaires obtenus en metabarcoding.

Un premier algorithme a été écrit, permettant de mettre à jour automatiquement l'ensemble des rangs taxinomiques du phytoplancton d'eau douce référencé en lien avec le logiciel Phytobs (Laplace-Treyture et al. 2017) dont la taxonomie, pour le phytoplancton, est à jour et basée sur la base taxonomique de référence AlgaeBase (Guiry & Guiry 2020). Cet algorithme a été appliqué sur les bases de références publiques curées (SILVA, Quast et al. 2013 ; PR2, Guillou et al. 2013 ; PhytoRef, Decelle et al. 2015 ;  $\mu$ Green, Djemiel et al. 2020 ; Diat.barcode, Rimet et al. 2019) ce qui a permis, dès 2020, de commencer les travaux *in silico* concernant la comparaison des marqueurs génétiques entre eux (gap-analysis, fiches synthétiques décrivant avantages et inconvénients) et la recherche de primers relatifs à ces derniers, pour cibler efficacement le phytoplancton en metabarcoding.

- Le 2<sup>ème</sup> objectif de 2020 a été la recherche de primers. Nous nous sommes concentrés dans un premier temps sur des tests *in silico* de couples d'amorces utilisés dans la littérature mais aussi sur un autre volet où de nouveaux couples de primers ont été conçus depuis les bases de références homogénéisées. Ces nouveaux couples ont été testés à leur tour. Le design de ces derniers a pu être réalisé grâce à une méthode développée en 2020 pour la recherche de couples de primers adaptés au metabarcoding ciblant des organismes appartenant à des clades polyphylétiques et diversifiés taxonomiquement, comme c'est le cas pour le phytoplancton. A l'issue de cette nouvelle méthodologie, certains couples d'amorces ont présenté des résultats concluants (cf. rapport de 2021 et publication en cours).
- Enfin, le 3<sup>ème</sup> objectif a été le développement d'un protocole d'échantillonnage. Ce protocole avait pour but d'être transmis aux bureaux d'études, futurs acteurs des campagnes d'échantillonnage du projet dans les DROM.

Pour l'année 2021 deux objectifs ont été identifiés :

- La mise en place d'une campagne d'échantillonnage pour laquelle les différents bureaux d'études dans les DROM ont été contactés (transmission des instructions, du protocole et du matériel nécessaire). Ces campagnes d'échantillonnages se sont appuyées sur un protocole permettant le prélèvement du phytoplancton dans les plans d'eau en limitant les risques de contamination et en assurant une bonne conservation du matériel biologique jusqu'aux analyses moléculaires. Les campagnes d'échantillonnages ont été réalisées dans les DROM, (1 site en Guadeloupe, 9 sites en Guyane et 5 sites à la Réunion) et en métropole (4 lacs péri-alpins suivis par l'UMR CARTELE).
- La sélection des primers pour laquelle 5 couples de primers ont été sélectionnés sur la base de tests *in silico*. Ces couples de primers amplifient des fragments des gènes 16S et le 23S (i.e., ARNr 16S et 23S, respectivement la petite et grande sous-unité ribosomale des procaryotes et dans ce cas précis, l'unité procaryotique en jeu étant le chloroplaste). Ces 5 couples de primers ont été testés *in vitro* sur des communautés synthétiques (i.e., mélange de cultures de microalgues

de la collection de culture TCC, Rimet et al. 2018) et un échantillon environnemental (Léman, proche du port de l'UMR CARTELE, le 13.04.2021).

- L'algorithme d'homogénéisation taxonomique développé en 2020 a été adapté et mis à disposition dans une application en ligne (interface ShinyApp proposée par Rstudio) en open-access : Phytool ([https://caninuzzo.shinyapps.io/phytool\\_v1/](https://caninuzzo.shinyapps.io/phytool_v1/)). Une nouvelle version a été déployée en 2022 ([https://github-carrtel.shinyapps.io/phytool\\_v2/](https://github-carrtel.shinyapps.io/phytool_v2/)). A l'issue de cette année 2021, les couples de primers à retenir pour la suite du projet ont été définis. Les échantillons environnementaux récupérés en fin d'année 2021 ont été analysés en metabarcoding en 2022 sur la base de ces choix de primers (cf. détails dans le rapport 2021).

## I.2. Principaux objectifs pour l'année 2022

Les principaux objectifs pour l'année 2022 ont été :

- **La récupération début 2022 des échantillons des prélèvements d'eau réalisés durant l'année 2021 dans les DROM et en métropole (grands lacs péri-alpins).** Ces échantillonnages ayant été mutualisés avec les campagnes d'échantillonnages prévus dans le cadre des suivis DCE ou d'autres suivis de routine, nous avons pu bénéficier des comptages phytoplanctoniques correspondants réalisés en microscopie par les bureaux d'études.
- **La réalisation des analyses moléculaires (i.e., extractions d'ADN, amplifications PCR et séquençage) des différents échantillons de la campagne 2021.** Ces analyses ont été réalisées selon la méthodologie d'analyse établie sur la base des résultats des tests menés en 2021 et détaillés dans le rapport 2021 de ce projet.
- **Le lancement d'une nouvelle campagne d'échantillonnage pour l'année 2022.** Les mêmes sites échantillonnés lors de la première campagne de 2021 (DROM, 4 lacs péri-alpins) ont à nouveau été suivis en 2022. Par ailleurs, de nouveaux lacs métropolitains suivis dans le cadre de la DCE ont été rajoutés afin d'avoir un panel plus large de conditions environnementales sur lequel la méthode pouvait être testée.



## II. Campagnes d'échantillonnage

### II.1. Campagne d'échantillonnage de l'année 2021

Disposer pour chaque échantillon traité en metabarcoding, d'un comptage microscopique du phytoplancton est une condition essentielle à l'analyse et à la comparaison entre les données morphologiques et moléculaires. Par conséquent, pour la campagne d'échantillonnage 2021, nous avons pris contact avec les Offices de l'Eau des DROM et les DEAL/DREAL pour demander la permission de contacter les bureaux d'études en charge des suivis DCE dans les DROM. Nous avons pu solliciter ces derniers en vue de mutualiser leurs campagnes d'échantillonnage DCE avec celles que nous souhaitons réaliser. La réponse positive des bureaux d'étude a permis de pouvoir bénéficier d'échantillons filtrés (Sterivex 0.45  $\mu\text{m}$ ) pour les sites référencés dans le Tableau 1. Outre, ces sites dans les DROM, quatre grands lacs péri-alpins ont également été suivis (Tableau 1). Ces lacs péri-alpins sont suivis par l'UMR CARTELE dans le cadre du suivi de l'observatoire OLA (Observatoire des grands Lacs Alpins) et pour lequel des comptages microscopiques sont réalisés. Au total, 19 écosystèmes ont été suivis en 2021 et 175 échantillons ont été obtenus. Bien que le projet PhytoDOM vise à s'intéresser majoritairement à la diversité phytoplanctonique dans les plans d'eau d'Outre-Mer, des échantillons d'autres écosystèmes sont tout aussi importants dans cette phase de développement, pour tester et valider la méthodologie sur une large diversité d'organismes et d'écosystèmes.

**Tableau 1.** Liste des masses d'eau des DROM et de métropole échantillonnées lors des campagnes 2021.

Localisation	Département	Site	Type de masse d'eau	Nombre de station	Nombre d'échantillon		
DROM	Guadeloupe	Gaschet	Retenue	1	5		
		Guyane	Angoulême	Rivière	1	1	
			Athanase	Rivière	1	1	
			Aval Apatou	Rivière	1	1	
			Dégrad Corrèze	Rivière	1	1	
			Fourmi	Rivière	1	1	
			Langatabiki	Rivière	1	1	
			Papaichton	Rivière	1	1	
			Patagai	Rivière	1	1	
			Twenke	Rivière	1	1	
			La Réunion	Bois Rouge	Etang	1	12
				Gol	Etang	3	36
				Grand-Étang	Etang	1	4
	Sainte Suzanne	Rivière		1	12		
Métropole	Haute-Savoie	Saint Paul	Etang	3	36		
		Aiguebelette	Lac	1	8		
		Annecy	Lac	1	15		
		Bourget	Lac	1	18		
		Léman	Lac	1	20		

## II.2. Campagne d'échantillonnage de l'année 2022

Pour la campagne d'échantillonnage de 2022, le partenariat mis en place avec les bureaux d'études en charge des suivis DCE dans les DROM, ainsi que le suivi des quatre grands lacs péri-alpins s'est poursuivi (Tableau 1). La même stratégie a été mise en place en métropole afin d'étendre l'échantillonnage sur un plus grand nombre de sites (Tableau 2) avec les Agences de l'Eau et leurs bureaux d'étude prestataires. Pour l'heure, 91 plans d'eau répartis entre quatre bassins ont été ajoutés à la liste initiale (Tableau 2) représentant 221 échantillons supplémentaires. Il est déjà prévu pour l'année 2023 de poursuivre cette collaboration avec les Agences de l'Eau et leurs bureaux d'études prestataires, notamment par l'échantillonnage de sites dans les bassins non échantillonnés en 2022 (e.g., Rhin-Meuse).

**Tableau 2.** Synthèse des masses d'eau de métropole suivis à partir de la campagne d'échantillonnage de 2022.

Bassin (Agence de l'Eau)	Type de masse d'eau	Nombre de sites	Nombre d'échantillon
Adour-Garonne	Lacs	44	83
Artois-Picardie	Lacs	5	15
Loire-Bretagne	-	-	0
Rhin-Meuse	-	-	0
Rhône-Méditerranée-Corse	Lacs	28	81
Seine-Normandie	Lacs	14	42

### III. Traitement des échantillons au laboratoire

#### III.1. Traitement des échantillons environnementaux

Pour la campagne 2021, un total de 175 échantillons environnementaux (Tableau 1) ont été réceptionnés au laboratoire. Une fois récupérés, les échantillons ont été stockés à  $-20^{\circ}\text{C}$  afin de préserver au mieux l'ADN en vue de son extraction. Le travail en laboratoire a débuté une fois tous les échantillons réceptionnés.

L'ensemble de ces échantillons a été traité dans les premiers mois de l'année 2022 en suivant le même protocole. Cependant, la méthode de filtration utilisée dans les DROM diffère de celle utilisée pour les échantillons des lacs péri-alpins :

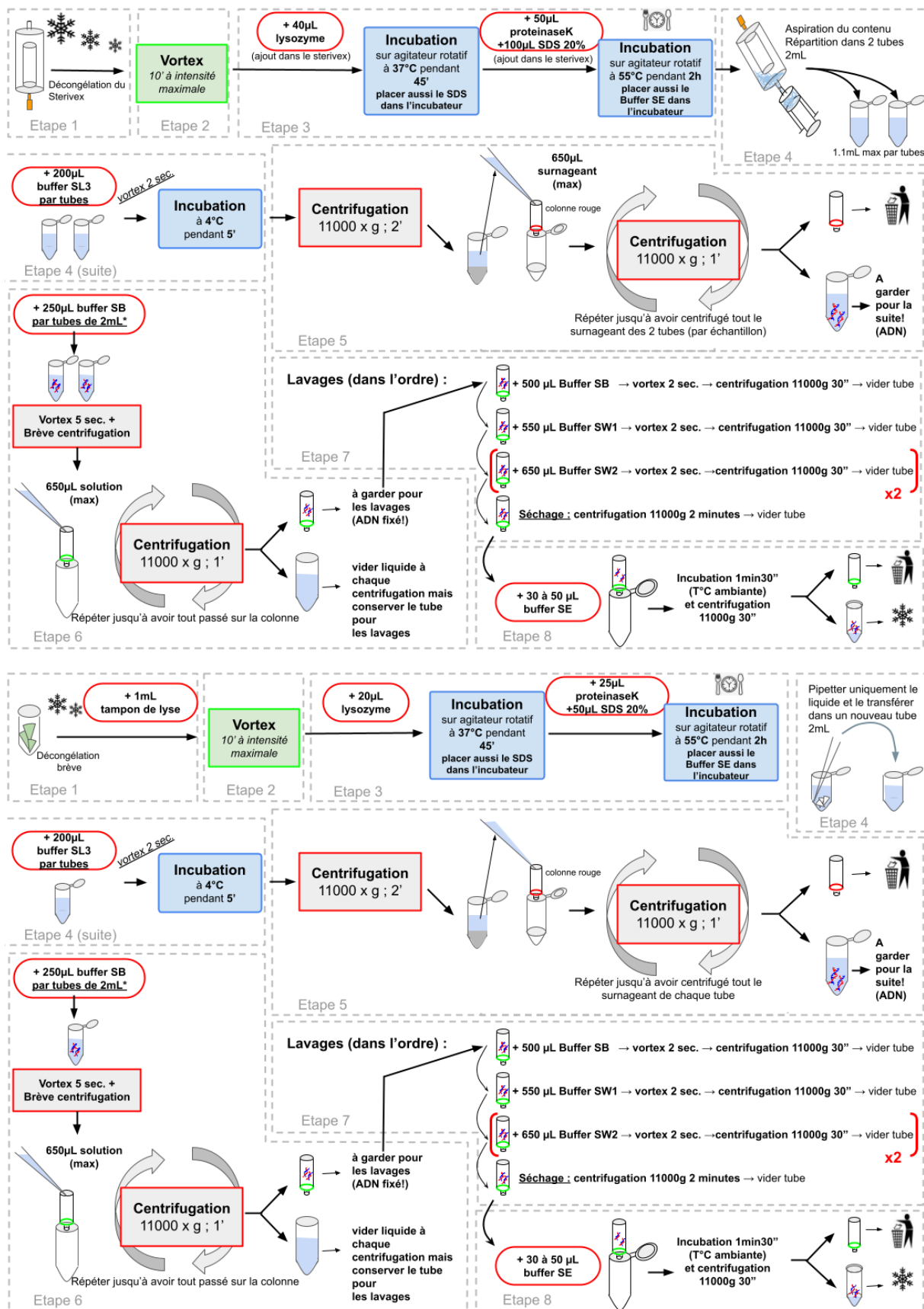
Pour les DROM, des filtres encapsulés Sterivex  $0.45\ \mu\text{m}$  (Figure 1) ont été utilisés par les bureaux d'étude, suivant le protocole développé en 2020 (cf. rapport 2021) qui permet une filtration sur le terrain. Une fixation est réalisée sur le terrain (ajout d'une solution de tampon de lyse) afin que le filtre Sterivex puisse être conservé pendant plusieurs semaines au frais et à l'obscurité, avant d'être envoyé au laboratoire de l'UMR CARTELE pour l'extraction/PCR.

- Pour les lacs péri-alpins, suivis par l'équipe de l'UMR CARTELE, les échantillons d'eau sont directement réceptionnés et filtrés au laboratoire sur des unités de filtration et sur des filtres ouverts présentant la même porosité que les Sterivex :  $0.45\ \mu\text{m}$ . Ce type de filtration est possible car les échantillons d'eau peuvent être filtrés et congelés quelques heures seulement après leur prélèvement. Après filtration, chaque filtre est découpé en morceaux avec des ustensiles stériles pour faciliter l'extraction d'ADN, puis placé dans un tube Eppendorf 2mL et conservé à  $-20^{\circ}\text{C}$  jusqu'à l'extraction/PCR. Ceci permet aussi de s'affranchir des coûts liés aux Sterivex.



**Figure 1.** Photographie de quelques Sterivex en provenance de La Réunion avant l'extraction.

L'extraction d'ADN repose sur l'utilisation du kit NucleoSpin Soil (Macherey-Nagel) avec un protocole adapté pour son utilisation avec des filtres Sterivex (Vautier et al. 2021). Cette méthode a déjà été testée avec succès en 2021. Pour les échantillons filtrés sur filtres ouverts, le protocole d'extraction a été adapté de manière à rester comparable à l'approche Sterivex : les mêmes étapes et réactifs sont utilisés. La Figure 2 résume le protocole utilisé pour un filtre Sterivex ou un filtre ouvert. Les protocoles détaillés sont disponibles en Annexe 1 et 2. Une fois les ADNss extraits des 176 échantillons, leur quantification a été réalisée en utilisant un NanoDrop (NanoDrop ONE ThermoScientific OZYME). En fonction des valeurs obtenues, des dilutions ont été réalisées afin d'avoir des aliquots d'ADN à une concentration de  $5\ \text{ng}\cdot\mu\text{L}^{-1}$  destinés aux amplifications de PCR.



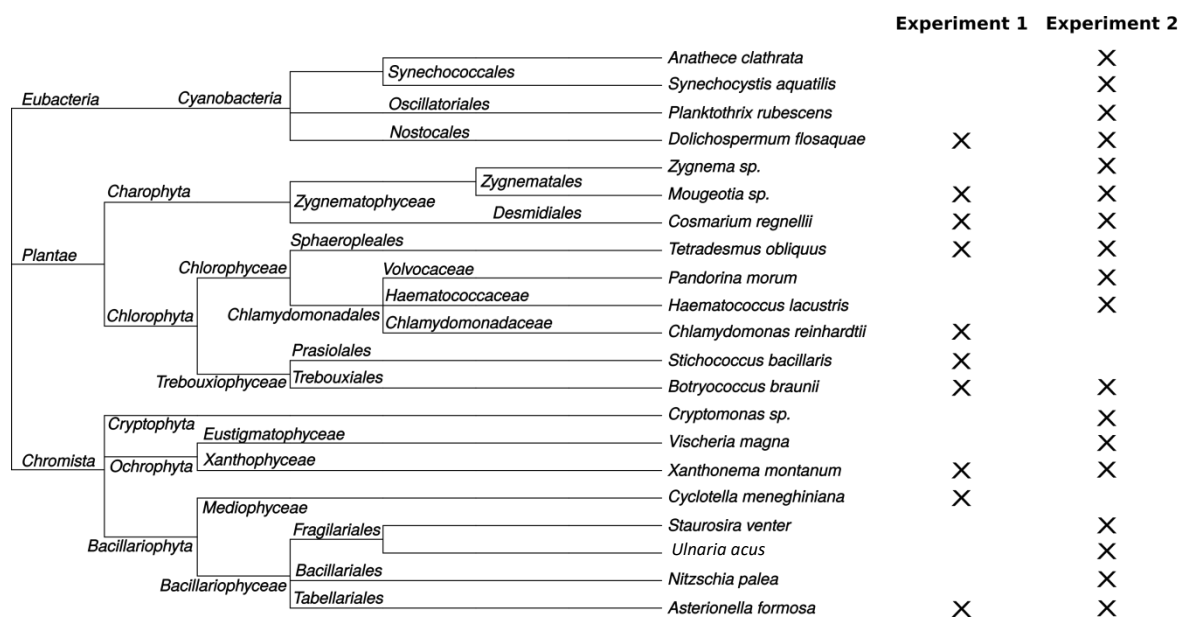
**Figure 2.** Résumés des protocoles d'extraction d'ADN utilisés pour les échantillons d'eau filtrés sur Sterivex avec 2mL de tampon de lyse (panneau du haut) et sur filtres (panneau du bas).

## III.2. Constitution d'échantillons synthétiques

Des échantillons synthétiques, à partir de souches en culture, ont été constitués afin de pouvoir avoir des échantillons de composition contrôlés (échantillons contrôles).

### III.2.1. Souches phytoplanctoniques utilisées

Pour constituer les échantillons synthétiques, 18 espèces différentes, couvrant une large gamme taxonomique de microalgues (Figure 3, experiment 2) et cultivées dans la collection de culture TCC ont été utilisées. Parmi ces dernières, 7 espèces sont communes avec les échantillons synthétiques de l'année passée (Figure 3, experiment 1). Pour chaque souche, l'ADN a été extrait individuellement et amplifié par PCR, afin de produire des mélanges d'amplicons (et non plus des mélanges d'ADN), dans le but d'améliorer la qualité des échantillons contrôles (cf. section III.2.2).



**Figure 3.** Espèces de la TCC utilisées dans les échantillons synthétiques en 2021 (« Experiment 1 ») et en 2022 (« Experiment 2 »). L'arbre présente les liens taxonomiques des différentes espèces.

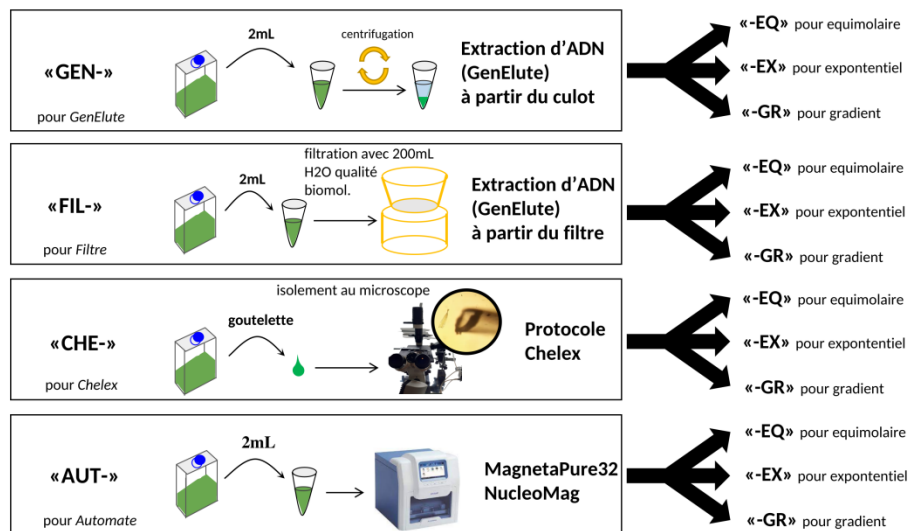
### III.2.2. Test de différentes méthodes d'extraction d'ADN pour la constitution des échantillons synthétiques

Les résultats de séquençage obtenus en 2021 avec les échantillons synthétiques ont mis en évidence certains biais, qui ont été discutés dans le rapport de 2021. En particulier, nous avons montré que les communautés contrôles souffraient d'un manque de précision quant à l'évaluation des concentrations en ADN après extraction. En effet, les ADN appartenant respectivement aux microalgues et aux bactéries présentes dans les cultures ne pouvaient pas être dissociés. Il est alors difficile d'évaluer avec précision la quantité d'ADN d'une souche de microalgue lorsque celle-ci se trouve en culture non axénique avec des bactéries associées.

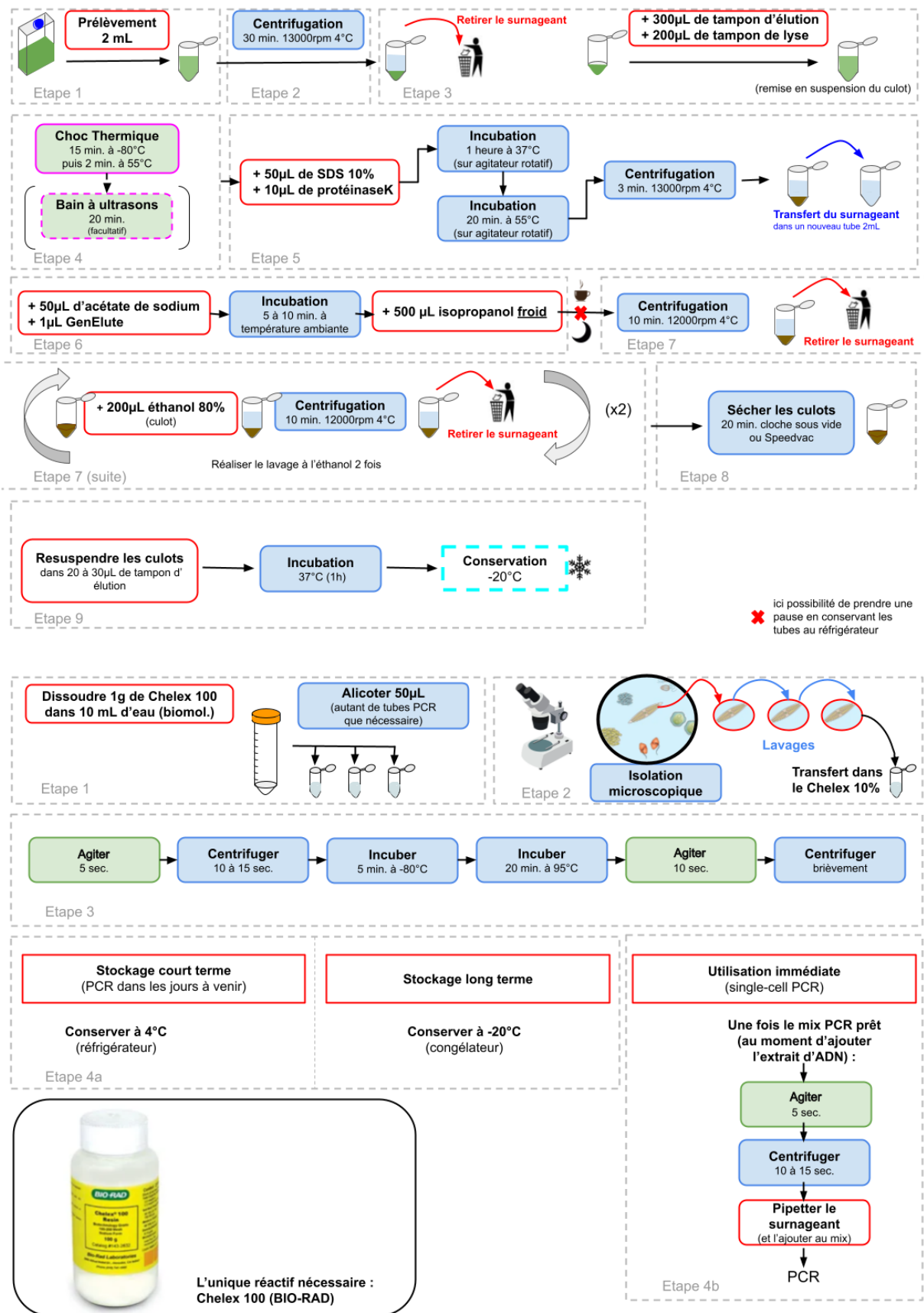
Pour pallier ce biais nous avons testé différentes stratégies d'extraction d'ADN pour constituer des communautés synthétiques. Tout d'abord, chacune des souches a été extraite individuellement selon les 4 méthodes présentées dans la Figure 4 et détaillées ci-dessous :

- Communauté « GEN » : extractions d'ADN « brut » avec la méthode GenElute (le protocole GenElute est en Annexe 3), comme cela avait déjà été fait en 2021 (cf. mocks « MA » livrable 2).

- Communauté « AUT » : extraction d'ADN « brut » avec un protocole utilisant l'automate MagnetaPure32 et le kit d'extraction NucleoMag Microbiome (MACHEREY-NAGEL) en suivant les recommandations du fournisseur. L'intérêt d'utiliser cet automate est le gain de temps pour l'extraction d'un grand nombre d'échantillons.
- Communauté « FIL » : extraction d'ADN « brut » à partir de souches filtrées sur des filtres ouverts dont la porosité est fonction de la taille des souches. Cette filtration a pour objectif de réduire la quantité de bactéries associées aux cultures, censées passer au travers du filtre, même si une importante partie peut rester attachée aux cellules de microalgues. Pour ce faire 2mL de chacune des souches ont été filtrés puis rincés avec un volume de 200 mL d'eau stérile, agissant comme solution de lavage/resuspension des organismes.
- Communauté « CHE » : cette méthode permet de réaliser une PCR sur une ou quelques cellules isolées par microscopie, sans extraction d'ADN préalable. Elle se base sur un protocole utilisant des billes de résines Chelex initialement développé par Walsh et al. (1991) repris par Richlen & Barber (2005). Cette approche est intéressante, notamment pour le séquençage de nouveaux barcodes d'espèces isolées par microscopie. Le protocole de la littérature a été optimisé pour obtenir des résultats concluants ( le protocole est présenté en Figure 5 et l'ensemble du protocole est également disponible en Annexe 4).



**Figure 4.** Présentation synthétique des différentes conditions d'extraction testées sur les souches contrôles pour constituer les différentes communautés synthétiques. Différents types de mélanges ont ensuite été constitués (équimolaires, exponentiels et en gradient) et sont détaillés dans la section III.2.3.



**Figure 5.** Résumé des protocoles d'extraction d'ADN « GEN » et « FIL » (panneau du haut) et par l'approche single-cell « CHE » (panneau du bas).

### III.2.3. Constitution des communautés synthétiques à partir d'amplicons d'ADN

Les extraits de chacune des cultures, réalisés suivant les conditions présentées dans la section III.2.2, ont été utilisés pour constituer 3 différents types de mélange à partir d'amplicons PCR. Les extraits d'ADN ont été amplifiés par PCR (cf. section III.3) afin d'obtenir, comme amplicons, les barcodes relatifs à chacun des couples de primers. Après une étape de purification des amplicons à l'aide du kit Illustra GFX PCR DNA and Gel Band (Cytiva), ceux-ci ont été quantifiés par NanoDrop pour constituer les mélanges d'ADN ci-dessous :

- Mélange équimolaire (« EQ ») : la quantité d'amplicon de chacune des souches est de 20ng.
- Mélange exponentiel (« EX ») : 2 espèces dominent le mélange (50ng d'ADN de chacune) et les autres espèces reproduisent des cibles rares (1ng) ou très rares (0.1ng).
- Mélange par gradient (« GR ») : un gradient de la quantité d'ADN s'étendant de 1ng à 50 ng par souche, avec des paliers de 10ng (3 espèces sont associés à chaque concentration).

Au final, un total de 24 communautés synthétiques a été produit : 2 barcodes investigués (16S et 23S) x 4 méthodes d'extraction (GEN, AUT, FIL, CHE) x 3 assemblages (EQ, EX, GR). Les concentrations d'ADN de chaque taxon dans chacune des communautés sont données en Annexe 8.

### III.3. Amplification PCR et séquençage des régions d'intérêt

Le choix du ou des barcodes constituait une étape importante du rapport de 2021. Il a été choisi de conserver pour cette année 2022 les 2 marqueurs 16S et 23S, afin d'assurer une comparaison entre ces derniers au cas où l'un d'entre eux (en particulier la région UPA de l'ARNr 23S) ne posséderait pas suffisamment de taxons référencés dans la bibliothèque de barcodes Phytool (Canino et al. 2021). Cela permet également d'augmenter les chances de détection des espèces en metabarcoding qui ne seraient pas détectées en microscopie. Les couples de primers sont rappelés dans le Tableau 3 : CYA359F/CYA781R pour l'ARNr 16S et ECLA23S\_F1/ECLA23S\_R1 pour l'ARNr 23S (au vu des tests *in silico*, ce couple d'amorce présentait un nombre de cibles taxonomiques plus important que les autres).

La PCR a été réalisée en deux étapes avec le protocole fourni en 2021 pour la PCR1 et celui de la plateforme de séquençage PGTB (Plateforme de Génomique et Transcriptomique de Bordeaux, <https://pgtb.fr/la-metagenomique-ciblee/>) pour la PCR2.

Pour rappel, pour les PCR1, les mélanges réactionnels d'un volume final de 25µL sont constitués comme suit :

- 2.5µL d'extrait d'ADN à une concentration d'environ 5ng.µL<sup>-1</sup>(<sup>1</sup>)
- 5µL de l'amorce forward à une concentration de 1µM
- 5µL de l'amorce reverse à une concentration de 1µM
- 12.5µL (2X) KAPA HiFi HotStart Ready Mix (mélange prêt de l'enzyme, tampons et nucléotides libres)

Les résultats des amplifications ont été contrôlés sur gel d'agarose (1%) par électrophorèse de 2mL des produits de PCR. Lorsque les résultats s'avéraient de bonne qualité (i.e., amplification réussie, bande nette sans trainées, absence de contamination, etc.), le reste de produit était conservé pour remplir les plaques destinées au séquençage (volume minimal nécessaire par échantillon: 20µL).

<sup>1</sup> Lorsque la concentration initiale de l'échantillon le permettait (i.e., était supérieure à 5ng.µL<sup>-1</sup>)



Deux différents runs de séquençage ont été réalisés sur la plateforme PGTB avec la technologie MiSeq (kit Macro v3, 2×250pb). Ce kit permet d’obtenir 20 à 22.10<sup>6</sup> reads ce qui correspond à un peu plus de 100 000 reads environ par échantillon pour 191 échantillons multiplexés dans chacun des runs. Le premier run concerne les amplicons obtenus avec les amorces ciblant la région v3-v4 de l’ARNr 16S (CYA359F/CYA781R), le second run concerne les amplicons obtenus avec les amorces ciblant la région UPA de l’ARNr 23S (ECLA23S\_F1/ECLA23S\_R1).

**Tableau 3.** Récapitulatif des amorces utilisées pour amplifier la région v3-v4 de l’ARNr 16S et la région UPA de l’ARNr 23S.

Couples d’amorces	Forward	Reverse	Référence
CYA359F/CYA359R [16S v3-V4]	GGGGAATYTTCCGCAATGGG	GACTACWGGGGTATCTAATCCCW TT	Nübel et al. 1997
ECLA23S_F1/ECLA23S_R 1	ACAGWAAGACCCTATGAAGC TT	CCTGTTATCCCTAGAGTAACTT	Cette étude

### III.4. Optimisation des protocoles de laboratoire

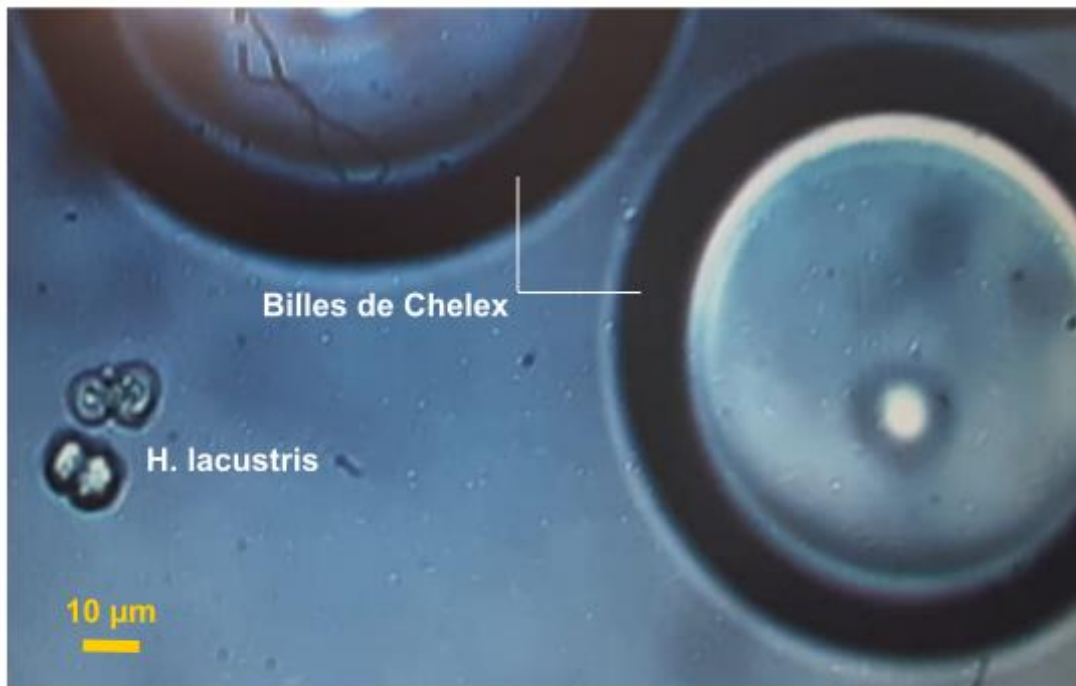
Les différents protocoles pour l’**extraction d’ADN** sont disponibles en Annexe 1 (Nucleospin soil filtre ouvert), 2 (Nucleospin soil Sterivex), 3 (Genelute) et 4 (Chelex).

En annexe sont donnés également les protocoles relatifs aux **filtrations** : sur filtre ouvert (Annexe 5) et sur Sterivex adapté au terrain (Annexe 6), ainsi que le protocole d’**amplification** recommandé par la plateforme PGTB (Annexe 7).

Le protocole actuellement utilisé en routine pour l’extraction d’ADN des échantillons prélevés à l’aide de Sterivex et conservés dans du tampon de lyse présente certains inconvénients. En effet, le protocole est long, ne permet pas de passer beaucoup d’échantillons en une journée (seulement 12 pour 1 manipulateur) et est coûteux en consommables. C’est également en grande partie pour cela que lorsque c’est possible (i.e., lorsque l’échantillon d’eau peut être ramené au laboratoire quelques heures après le prélèvement, tout en le conservant au frais pendant le transport) l’approche utilisant les filtres ouverts, préservés à sec et à -20°C doit être favorisée.

#### III.4.1. Optimisation de l’approche « Chelex »

L’intérêt principal de l’approche « Chelex » est le potentiel qu’elle offre pour séquencer facilement des espèces à partir de seulement quelques cellules (pipetées dans un échantillon naturel), sans avoir à les cultiver, afin de pouvoir compléter facilement les bases de référence de barcodes. Le protocole Chelex (Annexe 4), a dû être adapté par rapport à la littérature (e.g., Walsh et al., 1991) pour permettre une amplification à partir de cellules de microalgues isolées : augmentation du nombre de cellules cibles (nombre dépendant de la taille de ces dernières et de la facilité à les isoler à l’aide de pipettes pasteurs installées sur un micropipeteur) re-suspendues dans un volume moins important de Chelex (50µL). De plus, l’étape de lyse du protocole originel repose uniquement sur une étape de lyse mécanique (i.e., choc des microbilles de Chelex entre-elles) relativement courte. Nos résultats ont montré que cette étape n’était pas suffisante pour certaines microalgues qui restaient intactes à l’issue de cette lyse (e.g., Figure 6 avec *Haematococcus lacustris*). Pour pallier cela, un choc thermique a été ajouté. Nous préconisons également d’augmenter le temps des étapes d’agitation.



**Figure 6.** Photographie de cellules d'*Haematococcus lacustris*, restées intactes après la lyse mécanique du protocole Chelex.

### III.4.2. Optimisation du protocole d'extraction d'ADN à partir de Sterivex utilisant l'automate MagnetaPure32

Cet automate a été testé en vue d'optimiser l'efficacité de l'extraction d'ADN en permettant un gain de temps et la réduction des coûts, notamment lors du traitement de plusieurs centaines d'échantillons. Des tests ont été réalisés sur des échantillons environnementaux pour comparer l'efficacité de l'automate avec celle du kit NucleoMag Water (MACHEREY-NAGEL) et les méthodes utilisées en routine en 2022 pour les filtres ouverts et les Sterivex. Étant contraint par l'utilisation de Sterivex contenant du tampon de préservation, le nouveau protocole a dû être optimisé pour partir de cette étape et poursuivre l'extraction sur l'automate. Les résultats des tests effectués sur ce dernier ont montré, qu'à partir d'un volume de lysat inférieur, les extraits d'ADN présentaient une qualité supérieure (dosages et vérifications au NanoDrop) et une quantité d'ADN extrait plus importante (quantifications réalisées au QuBit). Les résultats de metabarcoding pour des tests identiques réalisés dans le cadre d'un autre projet montrent que les structures des communautés de diatomées sont similaires pour un même échantillon dont l'extraction a été réalisée avec l'automate et avec le kit NucleoSpin Soil. Des tests de détection de cibles rares par ddPCR (Stilla) ont également été réalisés dans le but de voir à quel point le nouveau protocole testé différait de l'ancien pour extraire de l'ADN présent en faible quantité. Il s'est avéré que les cibles rares étaient mieux détectées avec le protocole d'extraction sur automate qu'avec l'ancien.

## IV. Développements bioinformatiques

### IV.1. Mise à jour de l'application Phytool

La nouvelle version v2 de Phytool (Canino et al. 2021) est disponible sur le site Github de l'UMR CARTEL au lien suivant : <https://github.com/Github-Carrtel>.

Toutefois l'ancien site de Phytool est toujours disponible ([https://caninuzzo.shinyapps.io/phytool\\_v2\\_beta/](https://caninuzzo.shinyapps.io/phytool_v2_beta/)) et propose une redirection vers la nouvelle adresse.

Plusieurs améliorations sont disponibles dans cette nouvelle version :

- La couverture taxonomique a été enrichie. A l'origine, seulement les espèces de microalgues recensées sur Phytobs (Laplace-Treytore et al. 2017) étaient disponibles sur Phytool. La taxonomie relative aux microalgues dans la base de données TAXREFv15 de l'INPA (Gargominy et al. 2022) a été utilisée pour rechercher davantage de microalgues dans les bases de références ADN publiques. Un nouvel onglet a ainsi été mis en place dans le « Taxonomy Browser » pour permettre une distinction entre les microalgues d'eau douce uniquement (Phytobs) et les microalgues pouvant provenir des milieux aquatiques et terrestres (Phytobs et TAXREFv15).
- Les noms des auteurs des taxons ont été ajoutés (suggestions des reviewers de Canino et al. 2021). Par ailleurs, la base de référence Diat.barcode (Rimet et al. 2019) réunissant les barcodes *rbcL* pour les diatomées a également été intégrée dans un onglet dédié.
- La méthode de curation des séquences ajoutées dans Phytool a été modifiée dans la version v2. Lorsque différentes séquences avec un même nom taxonomique sont comparées entre elles, un « cluster » arbitraire est construit avec un seuil de similarité de 80% (i.e., suffisamment large pour exclure uniquement les séquences qui pourraient être considérées comme des outliers et donc comme des séquences mal assignées). Les séquences ayant un même nom taxonomique et qui sont en dehors de ce cluster ne sont pas retenues.
- Concernant les bases de référence (i.e., 16S, 23S, 18S et *rbcL* de Diat.barcode), un outil permettant de visualiser leur contenu de manière interactive a été ajouté.
- Par ailleurs, il est possible de télécharger les séquences en vue de créer une base de référence pour une assignation sur un logiciel dédié (e.g., Mothur qui requiert un format spécial).
- D'autres modifications, d'ordre mineur, ont également été réalisées :
  - L'onglet « Versions » sur l'application permet d'avoir un aperçu des modifications réalisées dans la version 2 de Phytool.
  - Sur la page principale, un lien a été ajouté afin que les personnes intéressées, pour rejoindre le réseau « barcodes » et aider à alimenter les bases de références, puissent prendre contact avec nous.

Enfin, la mise à jour v2 de Phytool a permis d'enrichir de nouvelles séquences, les différentes bases de références des marqueurs d'intérêt (16S, 23S, 18S, *rbcL*).

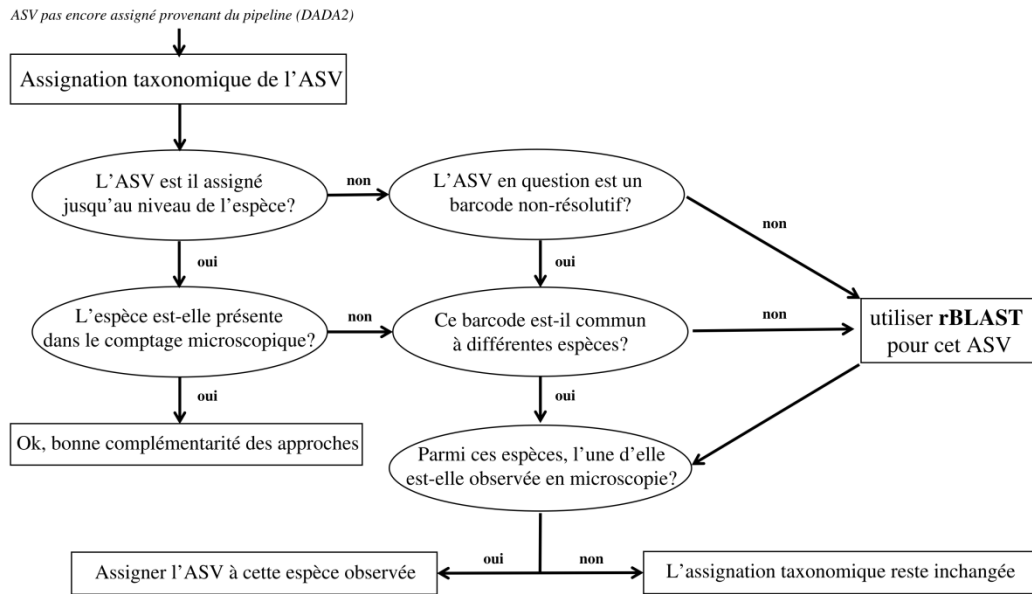
## IV.2. Traitement des données de séquençage

Les données de séquençage (sous la forme de fichiers FASTQ) ont été analysées grâce au pipeline développé en 2021. Il utilise notamment le logiciel cutadapt v3.5 (Martin 2011) qui permet de détecter et de retirer les primers des reads. Le traitement des séquences se poursuit jusqu'à l'obtention des tables d'ASVs (Amplicon Sequence Variant) en utilisant le pack logiciel « open source » DADA2 v1.20 (Callahan et al. 2016). L'assignation taxonomique des ASVs a été réalisée en utilisant la commande *classify.seqs* du logiciel Mothur (Schloss et al. 2009). Le script a été légèrement modifié pour permettre des sorties graphiques supplémentaires comme par exemple i) une visualisation du chevauchement et de l'assemblage des reads curés entre eux, ou ii) une sortie graphique finale permettant d'avoir un aperçu de la diminution du nombre de reads après chacune des étapes du pipeline bioinformatique. Ce script, rédigé en langage de programmation R, est en cours de publication et sera mis en open access sur: <https://github.com/Github-Carrtel>.

## IV.3. Optimisation de l'analyse des données

Comme expliqué précédemment, pour chacun des échantillons environnementaux analysés en metabarcoding, un fichier d'observation microscopique est disponible. La nomenclature taxonomique des comptages microscopiques a été homogénéisée avec Phytool v2 afin de pouvoir comparer les listes obtenues en microscopie aux listes obtenues en metabarcoding. Dans certains cas, l'homogénéisation par Phytool v2 n'a pas suffi car des erreurs étaient présentes dans les fichiers d'entrées de microscopie (e.g., fautes de frappes, majuscules, espaces traînants etc.). Au vu de la récurrence de certaines erreurs, provenant probablement en partie du logiciel d'aide au comptage Phytobs, une liste de ces erreurs a donc été dressée et envoyée au gestionnaire du logiciel Phytobs.

Concernant l'algorithme d'assignation des séquences environnementales, dans certains cas, l'assignation n'allait pas jusqu'à l'espèce et s'arrêtait à des rangs taxonomiques supérieurs comme par exemple l'Ordre. Ce phénomène se produit lorsque l'assignation est ambiguë, du fait de la présence de barcodes proches dans la base de référence de barcodes. Dans ce cas, les barcodes se caractérisent par des noms taxonomiques différents mais un rang taxonomique commun. Afin d'augmenter la comparabilité entre metabarcoding et microscopie nous avons développé une méthode d'assignation semi-manuelle (Figure 7 s'appuyant sur les identifications validées en microscopie pour lever l'ambiguïté. Cette même approche a été utilisée pour identifier les séquences correspondant aux espèces des échantillons synthétiques lorsque le résultat de leur assignation était ambiguë, en s'appuyant sur leur présence connue dans l'échantillon.



**Figure 7.** Présentation schématique du fonctionnement de l'algorithme utilisé pour ré-assigner les séquences environnementales en s'appuyant sur les suivis microscopiques.

## V. Comparaison des inventaires obtenus en microscopie et en metabarcoding

### V.1. Échantillons synthétiques

Les échantillons synthétiques étaient composés de quantités contrôlées d'ADN d'amplicons de 18 espèces différentes. L'hypothèse est qu'il y a, pour une espèce donnée, une relation entre la quantité d'ADN introduite dans l'échantillon synthétique et le nombre de reads obtenus après séquençage. Une corrélation est donc attendue entre la proportion d'ADN introduit et la proportion de reads obtenus sur l'ensemble des espèces constituant l'échantillon synthétique.

Pour le marqueur 16S, 15 espèces sur 18 ont été détectées, dont 3 ont été correctement assignées à l'espèce. Trois espèces n'ont pas été détectées : *H. lacustris*, *P. morum* et *V. magna* (Figure 3). Pour certaines des 15 espèces qui présentaient des barcodes peu résolutifs, l'algorithme de réassignation a été appliqué et une assignation manuelle a été réalisée pour les autres.

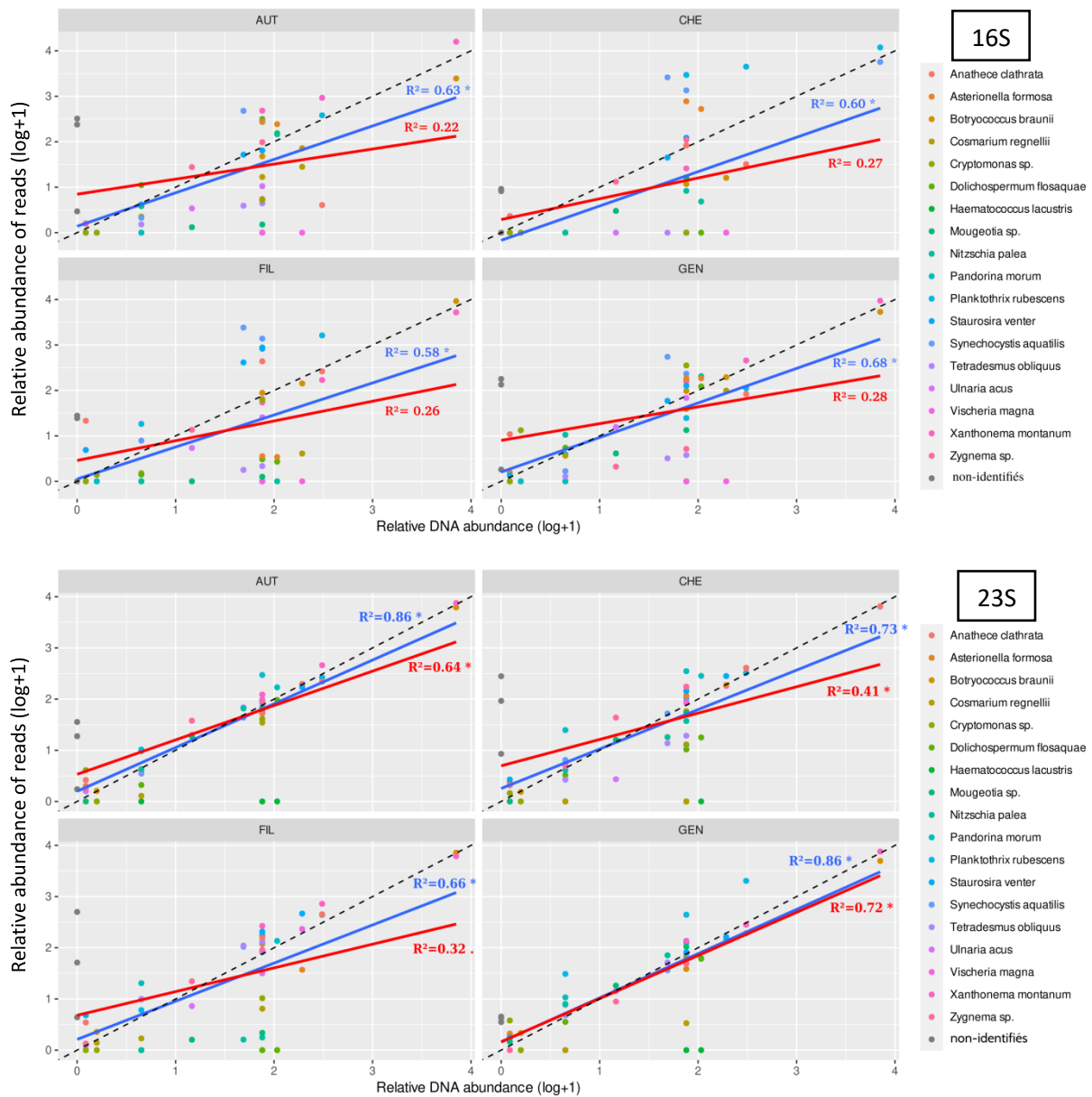
Pour le marqueur 23S, les 18 espèces ont été identifiées, même si certaines (e.g., *H. lacustris*) présentaient une abondance très faible (10 reads seulement sur un échantillon). Parmi ces espèces, 6 ont été correctement assignées jusqu'à l'espèce, les autres ont été réassignées au moyen de notre algorithme de réassignation et/ou d'une réassignation manuelle.

La Figure 8 présente pour l'ensemble des échantillons synthétiques, la quantité relative d'ADN introduite pour une espèce dans l'échantillon en fonction de l'abondance relative des reads obtenus pour cette espèce dans un échantillon donné. Deux droites de corrélation ont été réalisées pour chacune des communautés et chacun des 2 marqueurs :

- En bleu, la droite représente la régression sur l'ensemble des points.
- En rouge, la droite représente la régression lorsque les valeurs extrêmes sont enlevées, en particulier les concentrations les plus fortes en ADN.

L'estimation du lien entre la quantité d'ADN introduite et le nombre de reads est plus forte (i.e., des coefficients de corrélation plus élevés) et plus significative ( $p$ -value < 0.05) pour le marqueur 23S que pour le marqueur 16S, et ceci quelle que soit la méthode d'extraction considérée (i.e., AUT, CHE, FIL, GEN). Un meilleur taux de détection des espèces introduites est également observé avec le marqueur 23S.

En ce qui concerne les abondances relatives de reads pour la condition de mélange équimolaire (Figure 8 : points alignés les uns au-dessus des autres et correspondants à une quantité relative d'ADN environ égale à 2) la dispersion des points avec le 23S est moindre qu'avec le 16S. Cela signifie que les proportions de reads sont mieux corrélées aux proportions d'ADN avec le marqueur 23S. Cela est en particulier influencé par la moindre capacité de détection des espèces observée pour le 16S.



**Figure 8.** Corrélation entre la quantité d'ADN introduite et le nombre de reads obtenus pour les 18 espèces dans les 3 échantillons synthétiques (mélanges équimolaire, gradient et exponentiel) pour le marqueur 16S (haut) et 23S (bas) et pour chacune des 4 méthodes d'extraction utilisées (automate - AUT, Chelex - CHE, filtre - FIL, GenElute - GEN). Les résultats sont exprimés en log des abondances relatives de reads et d'ADN d'amplicon par espèce. En bleu : régression sur l'ensemble des données ; en rouge : régression sur un sous-ensemble des données excluant les valeurs extrêmes ; en pointillés : régression 1:1. Les coefficients de corrélations de Pearson sont indiqués ainsi que la significativité de leur p-value (\* < 0.01 ; . < 0.05).

Plusieurs conclusions peuvent être tirées de ces résultats :

- Premièrement, concernant la capacité de quantification relative des espèces présentes en mélange dans une communauté : la constitution des communautés synthétiques à partir de mélanges d'amplicons (en non d'ADN comme en 2021) a permis une meilleure évaluation de la capacité des 2 marqueurs à produire des données semi-quantitatives en metabarcoding. L'estimation de la quantité d'ADN introduite dans les communautés synthétiques était en effet plus précise car réalisée sur les amplicons purifiés des barcodes d'intérêt, donc très majoritairement sur les organismes cibles (et non sur d'autres

organismes non-cibles comme les bactéries, cf. rapport 2021). On observe une corrélation robuste avec le 23S, en particulier pour les échantillons extraits avec l'automate (AUT).

- Deuxièmement, les résultats de ces échantillons mettent en évidence la sous-représentation ou non-détection de certaines espèces contrôles. Le fait d'avoir en parallèle les données de metabarcoding 16S et le 23S permet d'identifier si cela est dû à un problème d'amplification spécifique au marqueur (quand une espèce est détectée par un marqueur et pas par l'autre) ou d'extraction (quand l'espèce n'est détectée par aucun des 2 marqueurs). Des tests d'amplifications *in silico* peuvent également servir de contrôles supplémentaires. Le cas de *Haematococcus lacustris*, montre la difficulté d'extraire l'ADN de cette souche, qui n'est *in fine* représentée par aucun des deux marqueurs. L'étape de lyse cellulaire pourrait être mise en cause. Cependant, malgré la présence de différents types de lyse des cellules (lyse chimique pour la méthode AUT; lyse enzymatique + choc thermique pour les méthodes GEN et FIL et lyse mécanique + choc thermique pour la méthode CHE) cette espèce n'a pas pu être détectée, ou très peu. Cela pourrait être dû au fait que *H. lacustris* est une espèce aérophile qui peut produire des kystes de résistances qui sont probablement difficiles à lyser.
- Enfin, les 4 méthodes d'extraction d'ADN testées présentent des résultats variables quant à leur potentiel d'extraction des différentes espèces. En considérant uniquement les corrélations significatives, les méthodes « AUT » et « GEN » semblent plus efficaces que les 2 autres méthodes (i.e. « FIL » et « CHE »), permettant une meilleure détection et une meilleure quantification des espèces dans les communautés synthétiques. Pour la méthode « FIL », certaines espèces de microalgues peuvent solidement adhérer aux filtres et/ou former des amas, réduisant l'efficacité de l'extraction d'ADN. Cette hypothèse permettrait d'expliquer pourquoi *Nitzschia palea* a été détectée correctement par les autres méthodes, mais pas sur filtre. Cette diatomée benthique, forme naturellement des amas cellulaires solidement attachés entre eux et également au support. Dans des échantillons environnementaux, de tels amas sont donc susceptibles de se former, en fonction des espèces présentes dans les communautés, l'étape d'agitation permettant de bien détacher les particules du filtre est donc importante. Les étapes de lyse enzymatique doivent également être maintenues (i.e., sur un agitateur rotatif permettant un lessivage complet du filtre). Enfin pour les filtres ouverts, leur découpage en petits morceaux et l'absence de plis permettent de favoriser la disponibilité des microalgues aux étapes de lyse.

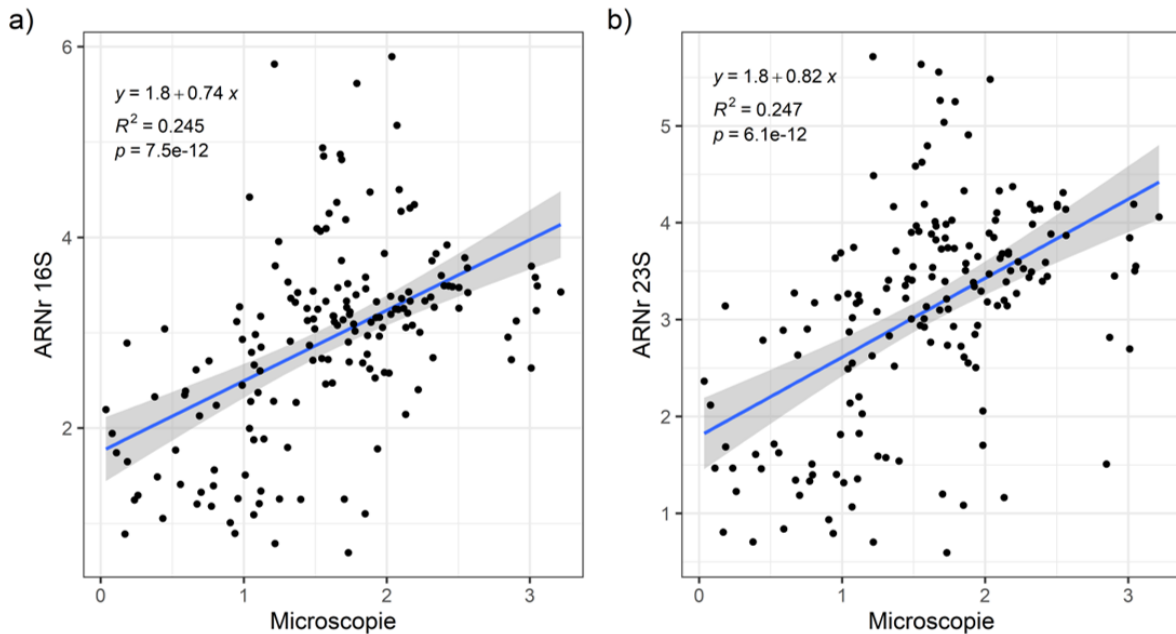
## V.2. Échantillons environnementaux

### V.2.1. Ensemble des échantillons : lacs des DROM et lacs péri-alpins

#### V.2.1.1. Alpha-diversité

Dans un premier temps, la diversité locale à l'échelle de chaque échantillon (i.e., alpha-diversité) a été calculée à partir des données obtenues avec les trois approches (pour rappel : microscopie, 16S et 23S). Une comparaison entre les valeurs d'alpha-diversité obtenues par microscopie et par metabarcoding a été effectuée. Sur l'ensemble des échantillons, les valeurs d'alpha-diversité obtenues en microscopie sont corrélées modérément et significativement avec celles obtenues en moléculaire, et ceci quel que soit le marqueur moléculaire considéré (16S :  $\rho = 0.245$ ,  $p\text{-value} < 0.001$  ; 23S :  $\rho = 0.247$ ,  $p\text{-value} < 0.001$  ; Figure 9). Par ailleurs, les valeurs d'alpha-diversité obtenues en metabarcoding sont globalement supérieures à celles obtenues en microscopie.

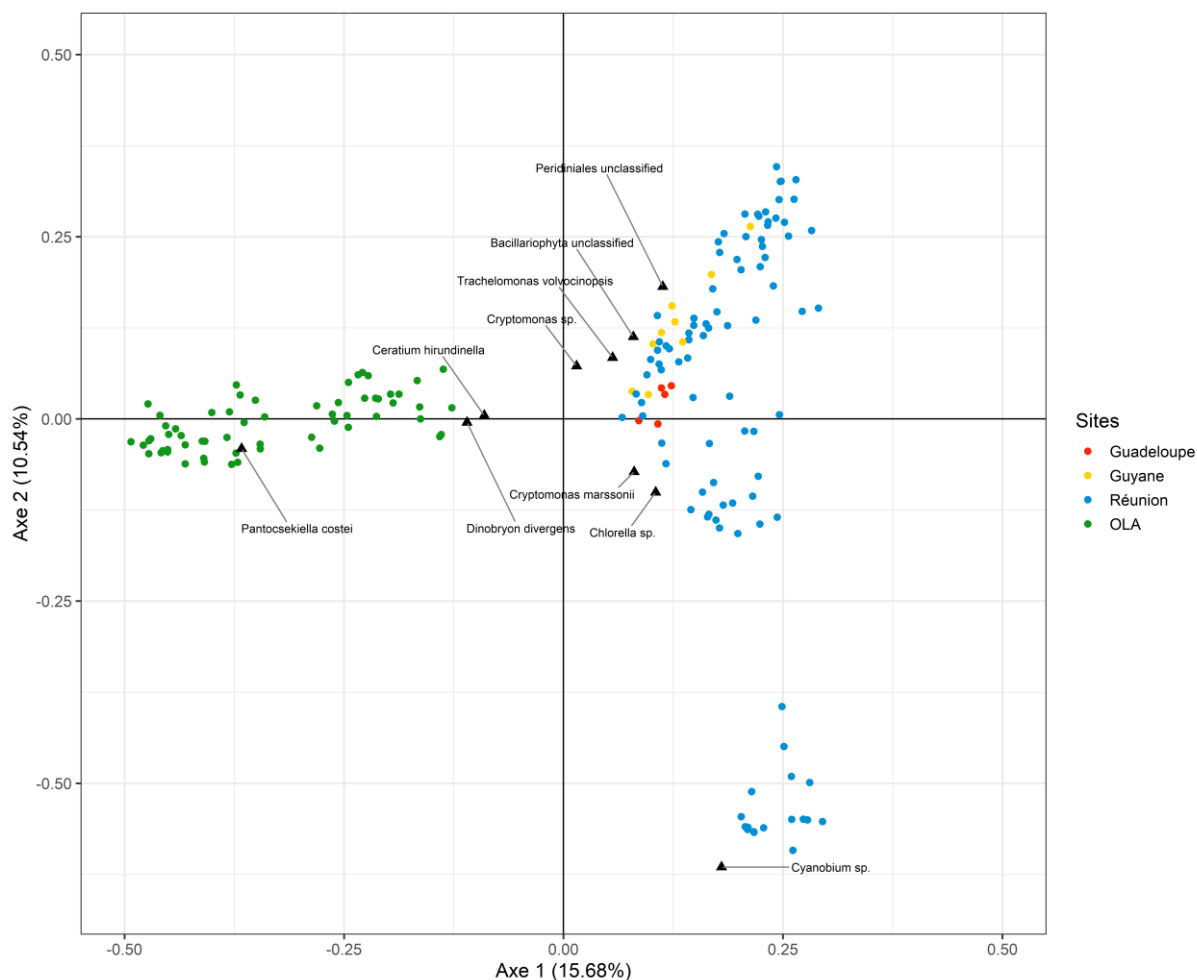




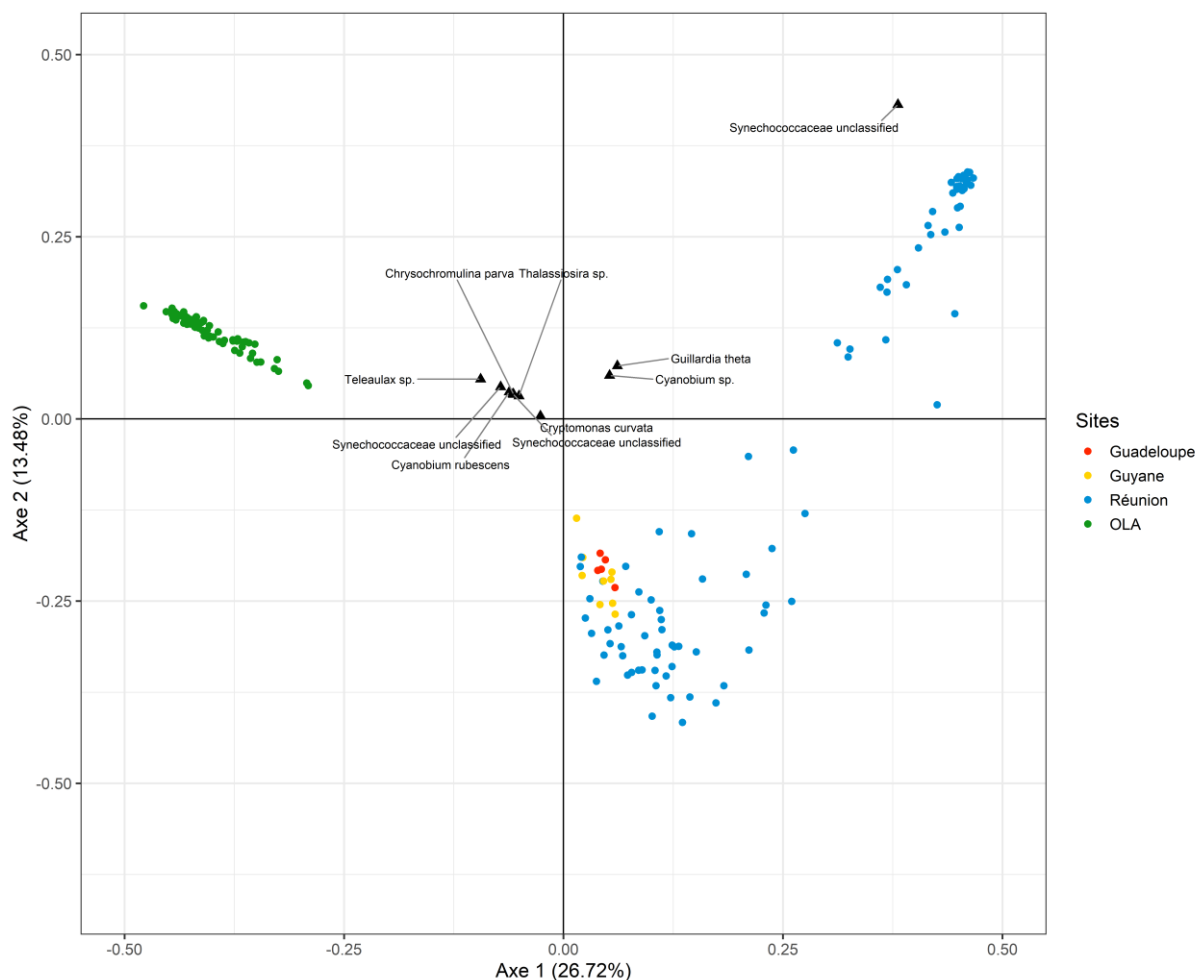
**Figure 9.** Relation entre les valeurs de diversité locale (ou alpha-diversité) calculées à partir des données de microscopie et moléculaires (marqueurs 16S et 23S). a) La relation est estimée entre les données de microscopie et les données du marqueur 16S. b) La relation est estimée en considérant les données de microscopies et les données du marqueur 23S. Chaque point noir correspond à un échantillon des DROM ou de métropole. L'alpha-diversité est exprimée selon l'indice de Shannon. Les relations ont été estimées par le calcul du coefficient de corrélation non paramétrique de Spearman.

### V.2.1.2. Beta-diversité

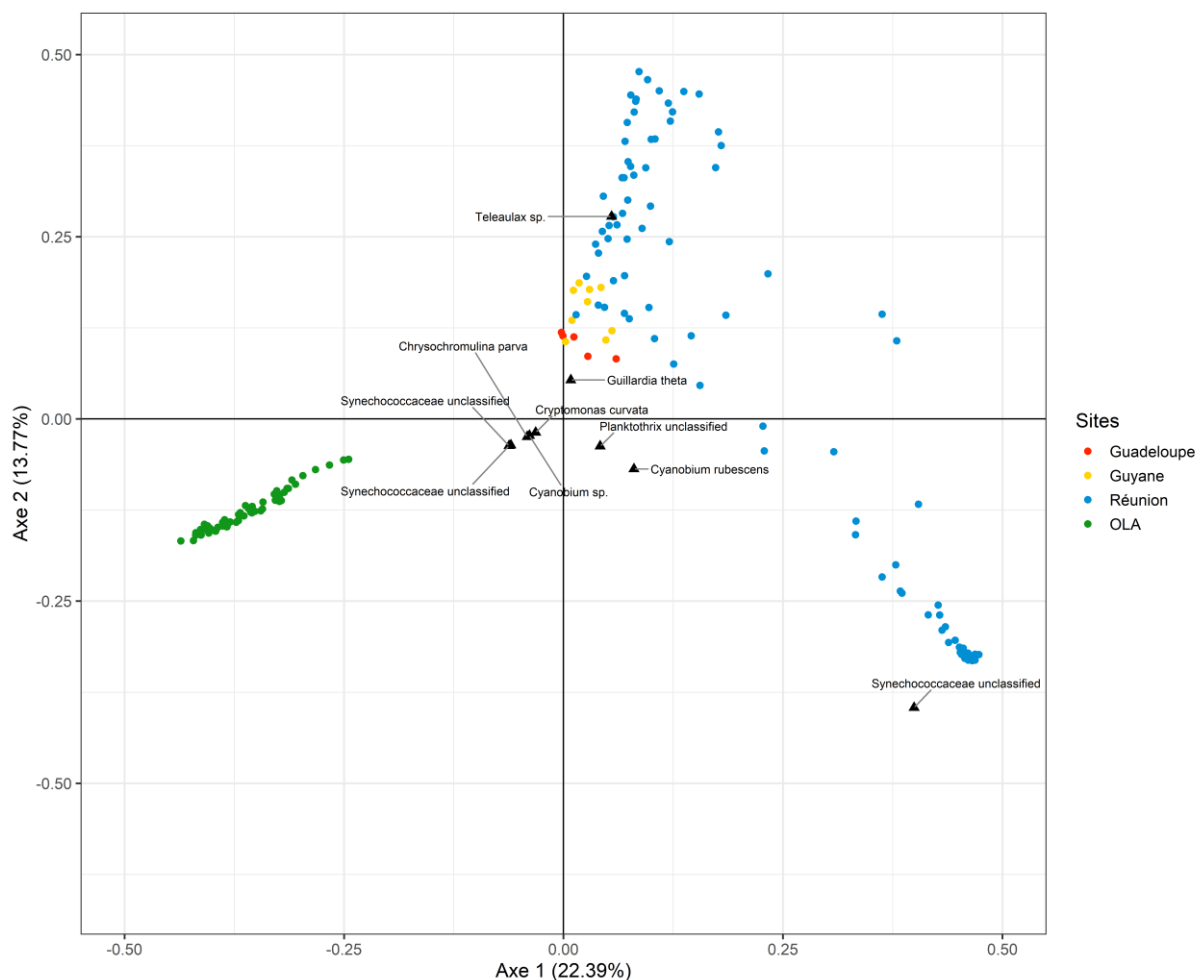
En complément, les structures des communautés (i.e., beta-diversité) obtenues en microscopie ont été comparées à celles obtenues en moléculaire (16S et 23S). Des Analyses en Coordonnées Principales (PCoA) ont été réalisées en se basant sur les listes floristiques, de chaque méthode, exprimées en biomasse relatives d'espèces pour la microscopie et en abondance relative du nombre de reads pour les méthodes moléculaires. L'analyse PCoA pour la microscopie (Figure 10) met en évidence une différence de structure entre les échantillons provenant des DROM et ceux des lacs péri-alpins. Parmi les échantillons des DROM, l'étang de Gol à La Réunion (points bleus dans le cadre inférieur droit du plan factoriel de la PCoA) se démarque des autres échantillons. Cet étang est connecté à l'océan par un grau, suggérant une influence marine dans la structuration des communautés phytoplanctoniques avec des apports d'eau de l'océan vers l'étang et inversement. Cette structuration est retrouvée avec les données moléculaires du marqueur 16S (Figure 11) et du marqueur 23S (Figure 12). Le gradient d'influence marine, notamment pour les sites d'échantillonnage de l'étang de Gol, est également observé. L'identification des 10 taxons les plus abondants met en évidence des taxons communs entre les trois approches appartenant aux genres *Cryptomonas* et *Cyanobium* (Figures 10, 11 et 12).



**Figure 10.** Analyse en coordonnées principales (PCoA) basée sur une matrice de distance de Bray-Curtis calculée sur les données de microscopie obtenues sur l'ensemble des échantillons provenant des DROM (i.e., Guadeloupe, Guyane et Réunion) et de métropole (i.e., OLA). Chaque point correspond à un échantillon et la couleur à sa provenance : Guadeloupe (rouge), Guyane (jaune), Réunion (bleu), et métropole (vert, lacs OLA). Les échantillons sont projetés sur le plan factoriel formé par les deux premiers axes de l'analyse (% d'inertie expliquée : Axe 1 : 15.58%, Axe 2 : 10.54%). Sur le plan factoriel de l'analyse sont également projetés les 10 taxons les plus abondants dans l'ensemble des échantillons.



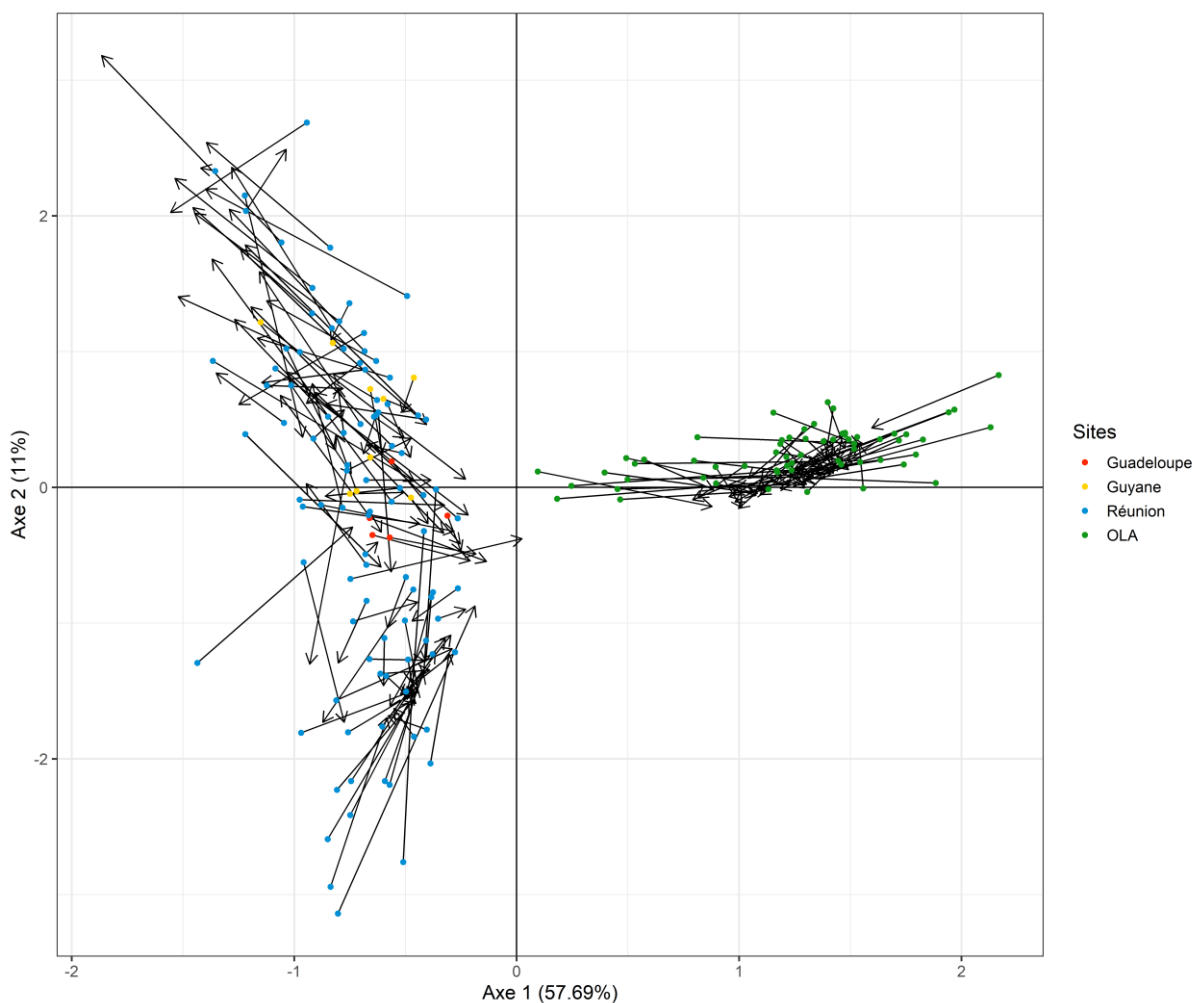
**Figure 11.** Analyse en coordonnées principales (PCoA) basée sur une matrice de distance de Bray-Curtis calculée sur les données moléculaire (marqueur 16S) obtenues sur l'ensemble des échantillons provenant des DROM (i.e., Guadeloupe, Guyane et Réunion) et de métropole (i.e., OLA). Les échantillons sont projetés sur le plan factoriel formé par les deux premiers axes de l'analyse (% d'inertie expliquée : Axe 1 : 26.72%, Axe 2 : 13.48%). Les 10 taxons les plus abondants dans l'ensemble des échantillons sont également projetés sur le plan factoriel de l'analyse. Voir la Figure 10 pour plus de détail sur les légendes.



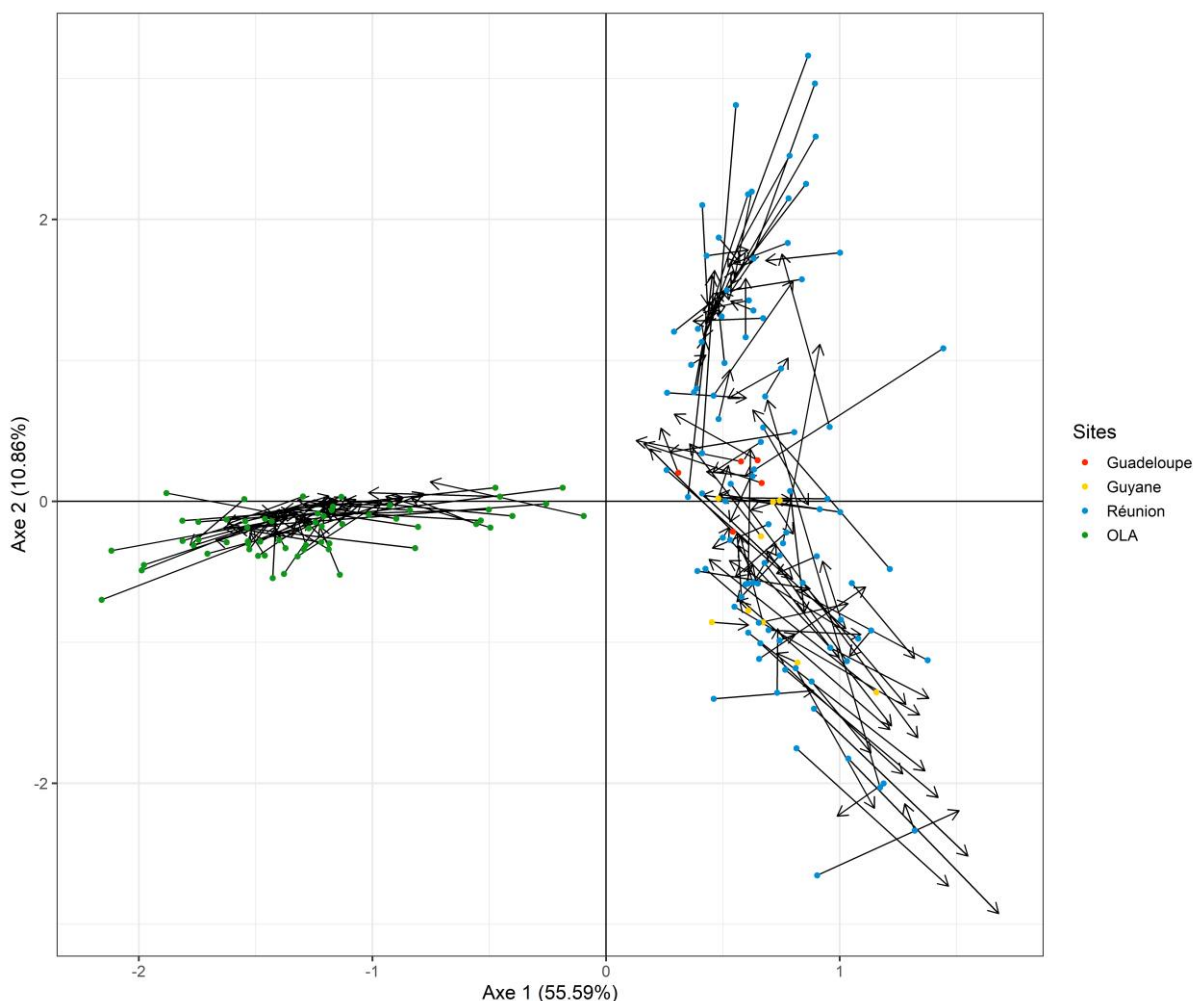
**Figure 12.** Analyse en coordonnées principales (PCoA) basée sur une matrice de distance de Bray-Curtis calculée sur les données moléculaire (marqueur 23S) obtenues sur l'ensemble des échantillons provenant des DROM (i.e., Guadeloupe, Guyane et Réunion) et de métropole (i.e., OLA). Les échantillons sont projetés sur le plan factoriel formé par les deux premiers axes de l'analyse (% d'inertie expliquée : Axe 1 : 22.39%, Axe 2 : 13.77%). Les 10 taxons les plus abondants dans l'ensemble des échantillons sont également projetés sur le plan factoriel de l'analyse. Voir la Figure 10 pour plus de détail sur les légendes.

### V.2.1.3. Co-structure microscopie/metabarcoding

La co-structure (ou la congruence) entre les assemblages phytoplanctoniques obtenus par la microscopie et par les marqueurs moléculaires (i.e., 16S et 23S) a également été évaluée grâce à des analyses de co-inertie. La congruence entre les données de microscopie et celles du marqueur 16S est élevée, avec 68.69% de l'inertie totale expliquée par les deux premiers axes de l'analyse (Axe 1 : 57.69%, Axe 2 : 11.00% ; Figure 13). Un test de Monte-Carlo, montre également un degré de co-variation élevé et significatif entre les deux ensembles de données ( $RV = 0.66$ ,  $p = 0.01$ ). Des patrons similaires sont observés lorsque les données de microscopie sont comparées à celles du marqueur 23S (Figure 14). Dans ce cas, les deux premiers axes de l'analyse de co-inertie expliquent 66.46% de l'inertie totale et le test de Monte-Carlo indique un degré de co-variation élevé et significatif ( $RV = 0.67$ ,  $p$ -value = 0.01). Une congruence élevée est à noter entre les assemblages phytoplanctoniques obtenus par les deux marqueurs moléculaires (i.e., 16S et 23S). L'analyse de co-inertie explique sur ses deux premiers axes 62.67% de l'inertie totale (Axe 1 : 50.10%, Axe 2 : 12.57%) et présente un fort degré significatif de co-variation ( $RV = 0.96$ ,  $p$ -value < 0.001).



**Figure 13.** Analyse de co-inertie décrivant la co-structure entre les assemblages phytoplanctoniques estimés par ACP sur les données de microscopie et moléculaire (16S). Projection des échantillons sur les deux premiers axes de l'analyse. Pour chaque axe, les valeurs entre parenthèses correspondent au pourcentage d'inertie expliquée. La longueur des flèches indique la différence de composition entre les données morphologiques (les points des flèches) et les données moléculaires (les pointes des flèches). La couleur des points correspond à l'origine des échantillons.



**Figure 14.** Analyse de co-inertie décrivant la co-structure entre les assemblages phytoplanctoniques estimés par ACP sur les données de microscopie et moléculaire (23S). Projection des échantillons sur les deux premiers axes de l'analyse. Pour chaque axe, les valeurs entre parenthèses correspondent au pourcentage d'inertie expliquée. La longueur des flèches indique la différence de composition entre les données morphologiques (les points des flèches) et les données moléculaires (les pointes des flèches). La couleur des points correspond à l'origine des échantillons.

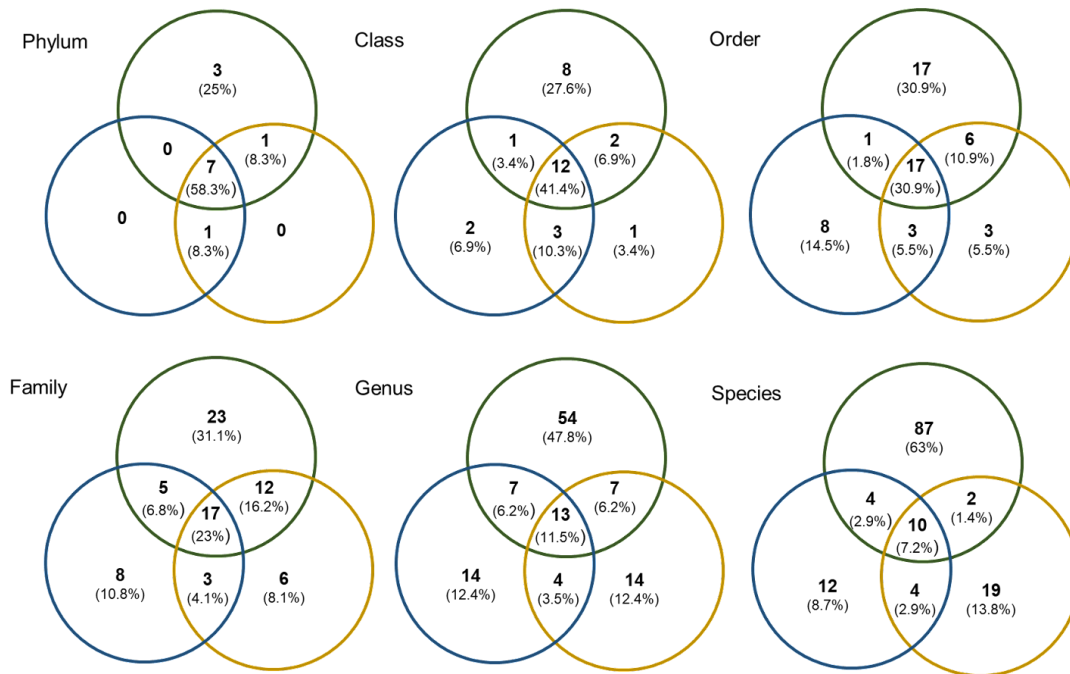
Les analyses de beta-diversité ont mises en évidence des patrons structuraux des communautés phytoplanctoniques distincts entre les échantillons provenant des DROM et de métropole et ceux quel que soit le type d'approche considéré (i.e., microscopie, 16S, 23S ; Figures 10-12). Une analyse spécifique à chaque milieu (i.e., DROM et métropole) a donc été réalisé dans la section suivante.

## V.2.2. Échantillons des Lacs péri-alpins (OLA)

### V.2.2.1. Gap analysis : comparaison des taxons détectés en microscopie et en metabarcoding

La comparaison des taxons des lacs péri-alpins identifiés par les méthodes moléculaires (16S, 23S) et la microscopie est présentée dans la Figure 15. Cette comparaison met en évidence que le nombre de taxons communs détectés par les trois méthodes est élevé pour les niveaux taxonomiques les plus hauts (i.e., phylum, class, ordre) et diminue fortement pour les niveaux les plus fins (i.e., famille, genre, espèce). En particulier, au niveau « espèce » le nombre de taxons communs entre la microscopie et le marqueur 16S est de 10% et entre la microscopie et le marqueur 23S est de 8.6%. La proportion d'espèces identifiées seulement par l'approche moléculaire (16S et 23S) est importante (25%) et

correspond principalement à des taxons de très petite taille (picophytoplancton) difficiles à observer en microscopie et présentant une importante diversité cryptique (e.g., Eustigmatophycées, diversité de pico-cyanobactéries, etc.). La proportion d'espèces identifiées seulement par microscopie est également forte (63%) et met en évidence la vacuité des bases de référence de barcode au niveau de l'espèce. Cette proportion est nettement plus faible pour le genre (47.8%) et la famille (31.1%).



**Figure 15.** Diagrammes de Venn comparant les taxons identifiés par les trois approches (16S : bleu, 23S : jaune et microscopie : vert). Les différents niveaux taxonomiques sont considérés du phylum à l'espèce. Le nombre (en gras) et le pourcentage (%) de taxons sont indiqués.

L'identification des 10 taxons les plus abondants détectés pour les trois méthodes met en évidence que les taxons enregistrés en microscopie sont différents de ceux enregistrés en moléculaire, à l'exception de *Plagioselmis nannoplantica* (correspondant à *Teleaulax sp.*, pour le 16S), qui a été enregistré en microscopie et en moléculaire (Tableau 4). Les résultats montrent i) que le picophytoplancton est sous-estimé en microscopie (*Cyanobium*, *Synechococcaeae*) et ii) que certains taxons sont identifiés à des niveaux grossiers en moléculaire alors qu'ils sont identifiés au niveau spécifique en morphologie (e.g., *Mediophyceae* en moléculaire et *Pantocsekiella costei* en microscopie), ce qui traduit la vacuité de la base de références de barcodes.

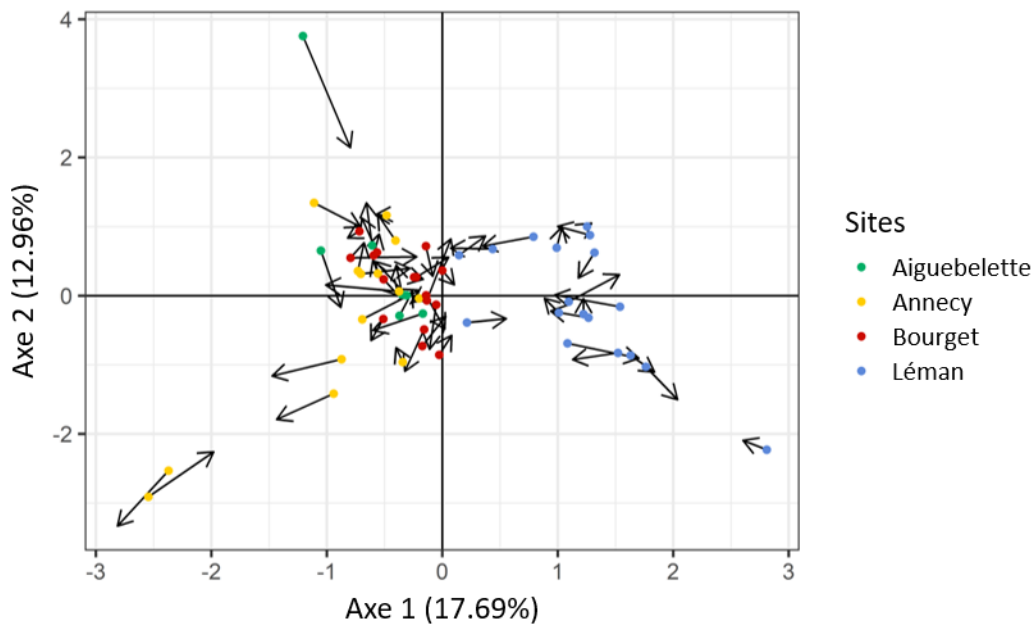
**Tableau 4.** Comparaison des dix taxons les plus abondants détectés par les approches morphologiques et moléculaires.

Taxon (Microscopie)	Ab (%)	Taxon (23S)	Ab (%)	Taxon (16S)	Ab (%)
<i>Pantocsekiella costei</i>	27.08	<i>Cyanobium rubescens</i>	12.79	<i>Synechococcaceae unclassified</i>	19.25
<i>Dinobryon divergens</i>	9.30	<i>Plagioselmis nannoplanctica</i>	9.45	<i>Teleaulax sp.</i>	18.12
<i>Ceratium hirundinella</i>	6.90	<i>Mediophyceae unclassified</i>	8.63	<i>Cyanobium rubescens</i>	11.61
<i>Plagioselmis lacustris</i>	5.57	<i>Cyanobium unclassified</i>	8.59	<i>Cryptomonas curvata</i>	8.58
<i>Fragilaria crotonensis</i>	5.48	<i>Synechococcaceae unclassified</i>	7.82	<i>Thalassiosira sp.</i>	7.99
<i>Mougeotia gracillima</i>	4.62	<i>Ochrophyta unclassified</i>	7.41	<i>Chrysochromulina parva</i>	7.04
<i>Plagioselmis nannoplanctica</i>	4.57	<i>Chromulinaceae unclassified</i>	4.63	<i>Cyanobium sp.</i>	5.17
<i>Cryptomonas sp.</i>	3.48	<i>Cryptophyceae unclassified</i>	3.42	<i>unknown unclassified</i>	4.38
<i>Diatoma tenuis</i>	3.35	<i>unknown unclassified</i>	2.87	<i>Cyanobium unclassified</i>	2.33
<i>Dinobryon sociale var. americanum</i>	3.05	<i>Chrysochromulina parva</i>	2.58	<i>Planktothrix rubescens</i>	1.77

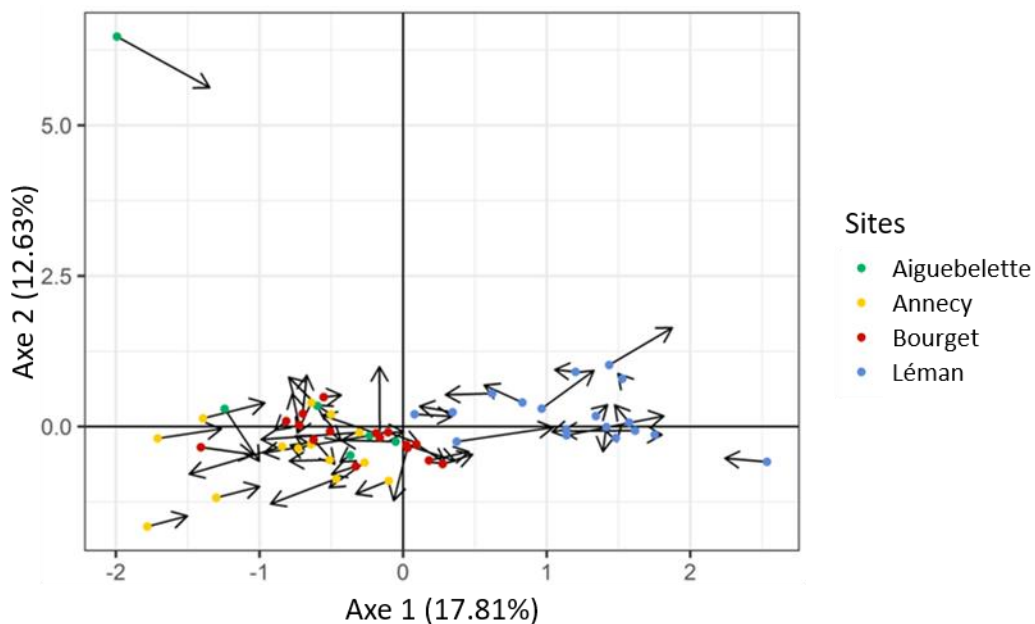
#### V.2.2.2. Co-structure microscopie/metabarcoding

Afin de comparer les structures (ou beta-diversité) obtenues en microscopie et en moléculaire (16S, 23S) dans les lacs péri-alpins, des analyses de co-inertie ont été réalisées (Figures 16 et 17). Les listes floristiques de chaque méthode sont exprimées en biomasse relative d'espèces pour la microscopie et en abondance relative du nombre de reads pour les méthodes moléculaires. Pour le marqueur de 16S, l'analyse de co-inertie met en évidence une congruence sur les deux premiers axes, représentant 30.65% de l'inertie totale projetée (Axe 1 : 17.69%, Axe 2 : 12.96%). De plus, le test de Monte-Carlo montre un degré significatif de co-variation entre les deux assemblages (RV = 0.62,  $p < 0.001$ , Figure 16). Une congruence a également été observée entre les assemblages basés sur la microscopie et le marqueur moléculaire 23S, avec 30.44% de l'inertie totale expliquée par les deux premiers axes (Axe 1 : 17.81%, Axe 2 : 12.63%), et une co-variation significative (RV = 0.65,  $p < 0.001$ , Figure 17). La co-structure entre les assemblages phytoplanctoniques obtenus par les deux marqueurs moléculaires (i.e., 16S et 23S) a également été testé. Une congruence forte est observée, avec les deux premiers axes expliquant 76.1% de l'inertie totale projetée (Axe 1 : 63.81%, Axe 2 : 12.29%), et une co-variation significative de 0.79 (RV,  $p < 0.001$ ).





**Figure 16.** Analyse de co-inertie entre les assemblages phytoplanctoniques estimés sur les données de microscopie et moléculaire avec le marqueur 16S. La longueur des flèches indique la différence de composition entre les données morphologiques (points des flèches) et les données moléculaires (pointes des flèches). Chaque point correspond à un échantillon et sa couleur à sa provenance : Aiguebelette (vert), Annecy (jaune), Bourget (rouge), et Léman (bleu).

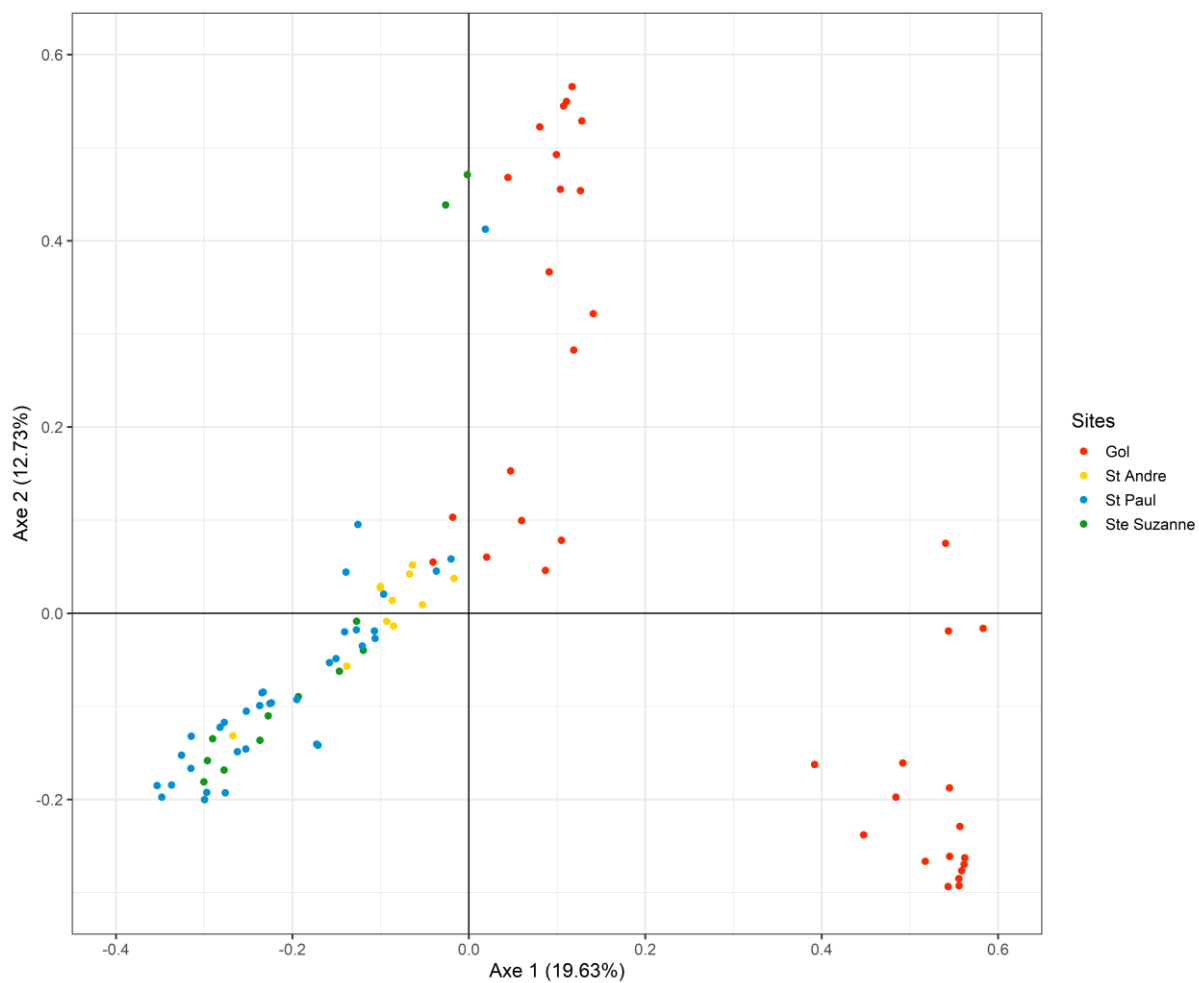


**Figure 17.** Analyse de co-inertie entre les assemblages phytoplanctoniques estimés par les données de microscopie et moléculaire avec le marqueur 23S. La longueur des flèches indique la différence de composition entre les données morphologiques (points des flèches) et les données moléculaires (pointes des flèches). Chaque point correspond à un échantillon et sa couleur à sa provenance : Aiguebelette (vert), Annecy (jaune), Bourget (rouge), et Léman (bleu).

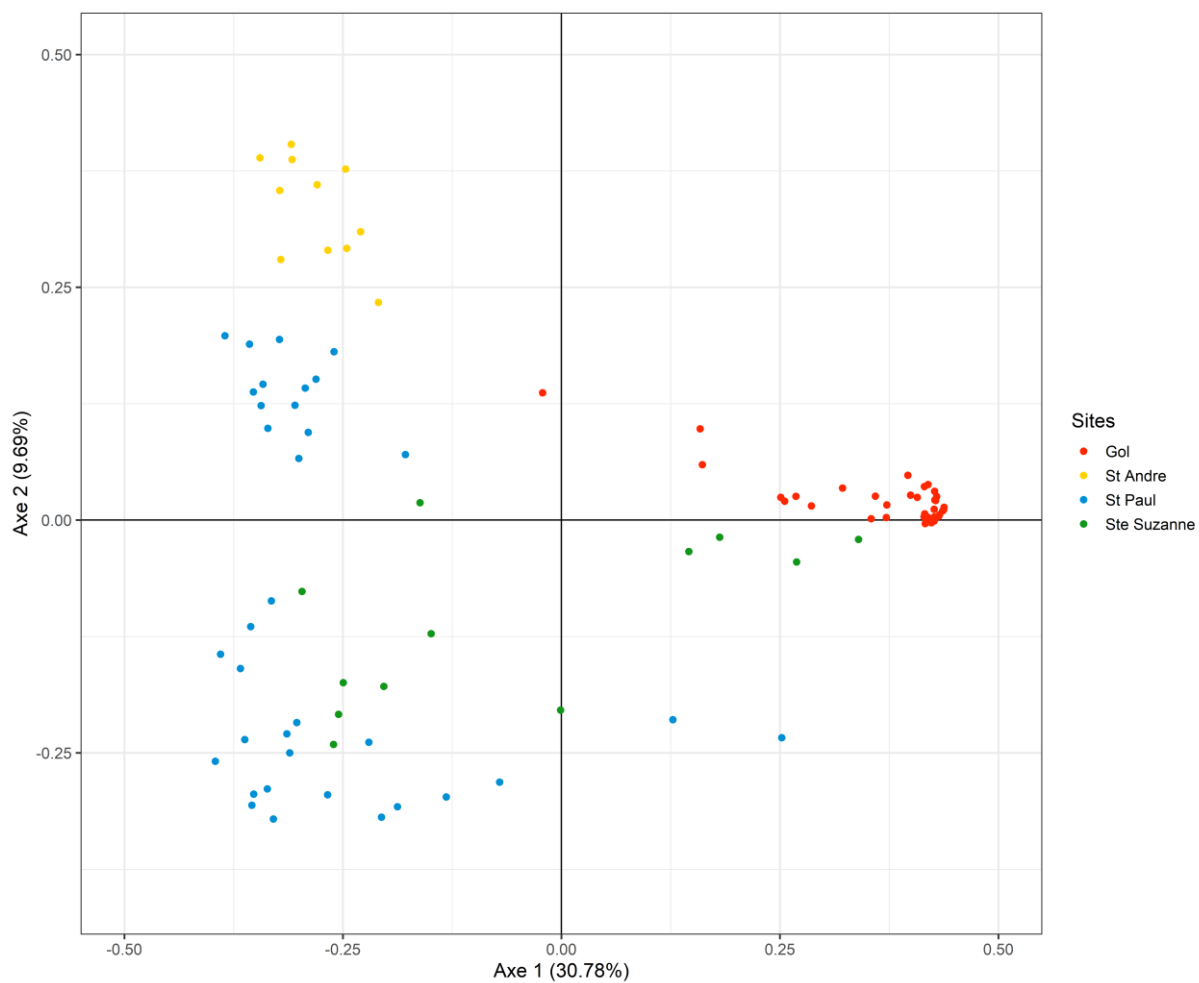
### V.2.3. Échantillons des Lacs des DROM

#### V.2.3.1. *Beta-diversité*

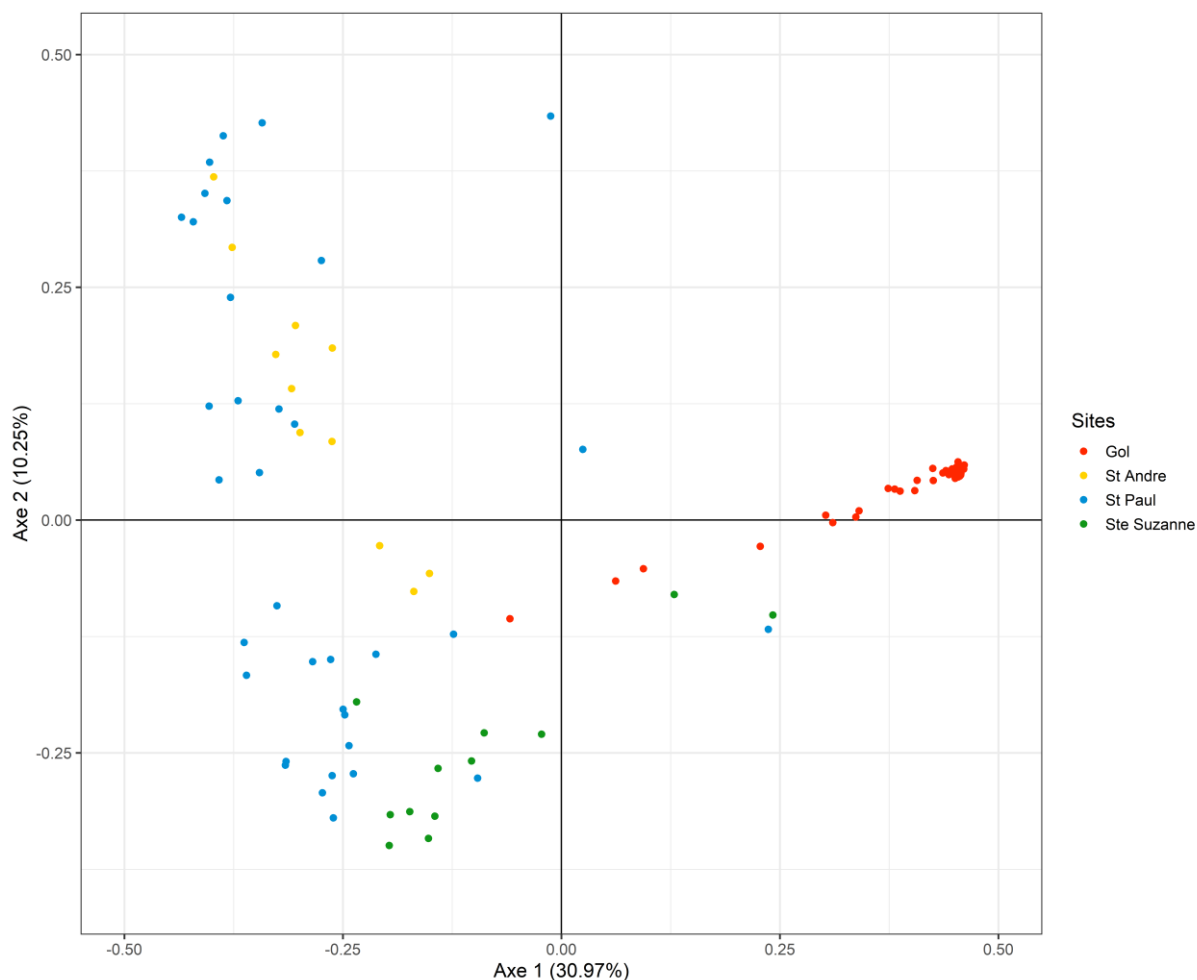
Pour les lacs des DROM, les structures des communautés (beta-diversité) obtenues en microscopie ont été comparées à celles obtenues en moléculaire (16S et 23S) grâce à des PCoA. Les listes floristiques de chaque méthode sont exprimées en biomasse relatives d'espèces pour la microscopie et en abondance relative du nombre de reads pour les méthodes moléculaires. Pour les échantillons provenant des DROM, avec la microscopie, on observe toujours une différence de composition des assemblages phytoplanctoniques de certains échantillons de l'étang de Gol (La Réunion) par rapport aux autres échantillons (Figure 18). Alors qu'une partie des échantillons de l'étang de Gol sont projetés dans la continuité des échantillons des 3 autres lacs Réunionnais (i.e., les points rouges présentant des valeurs positives sur l'axe 1 de la PCoA), l'autre partie des échantillons (correspondant aux échantillons du site le plus influencé par les apports marins) sont plus éloignés dans le plan factoriel. Ces résultats confirment les patrons observés en microscopie sur l'ensemble des échantillons (Figure 10) et indiquent une influence des apports marins dans la structuration des assemblages de phytoplancton dans l'étang de Gol. Ce patron de structuration des assemblages phytoplanctoniques est également retrouvé avec les marqueurs 16S (Figure 19) et 23S (Figure 20), mais de façon moins marquée. Par ailleurs, la dispersion plus large des échantillons pour les 3 autres lacs Réunionnais (i.e., Sainte Anne, Saint Paul et Sainte Suzanne) montre que les données moléculaires permettent d'accéder à des assemblages plus complexes et diversifiés qu'en microscopie.



**Figure 18.** Analyse en coordonnées principales (PCoA) basée sur une matrice de distance de Bray-Curtis calculée sur les données de microscopie obtenues à partir des échantillons prélevés dans les 4 sites de la Réunion : Gol (rouge), Saint Andre (jaune), Saint Paul (bleu), Sainte Suzanne (vert).



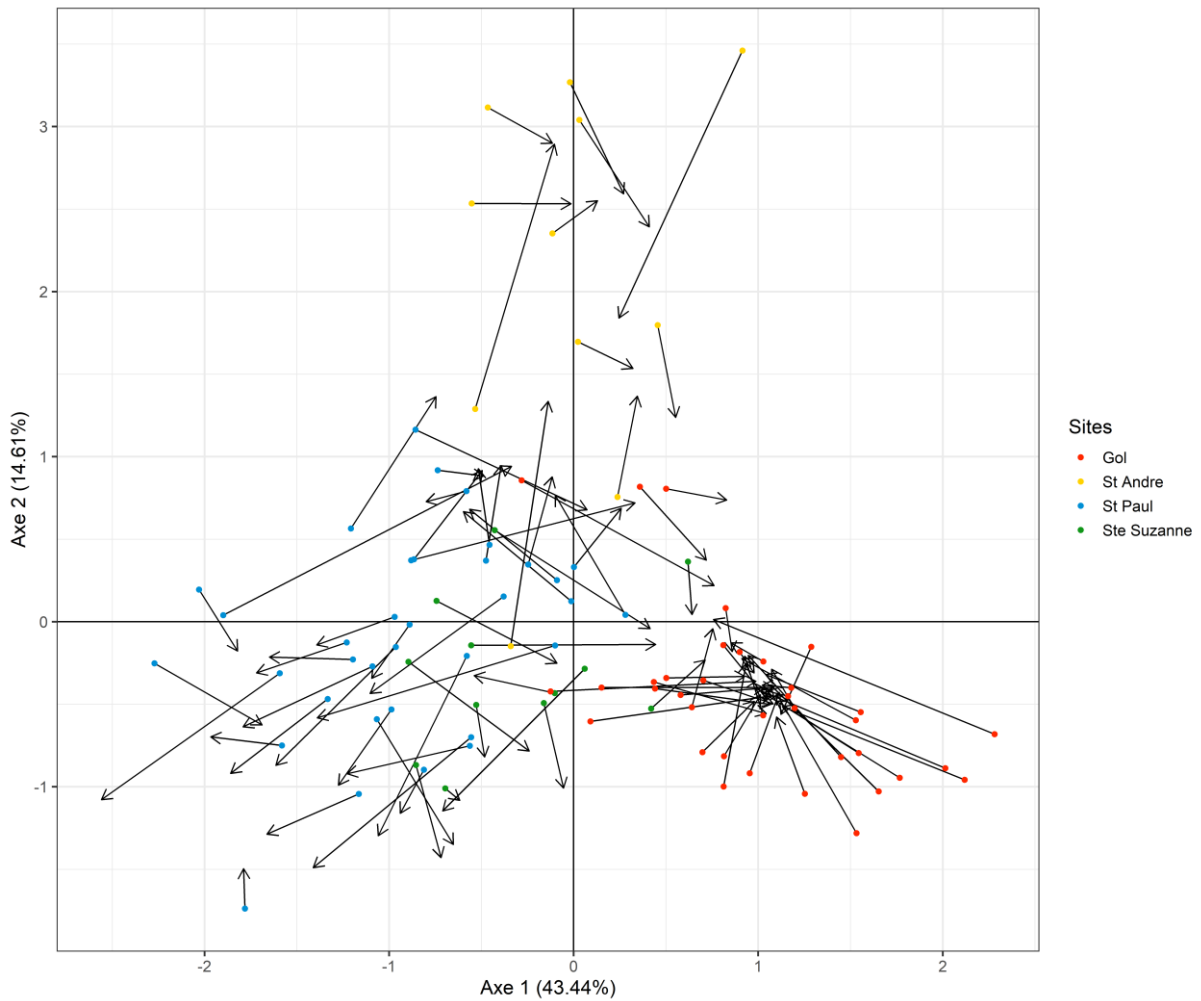
**Figure 19.** Analyse en coordonnées principales (PCoA) basée sur une matrice de distance de Bray-Curtis calculée sur les données moléculaires (marqueur 16S) obtenues à partir des échantillons prélevés dans les 4 sites de la Réunion : Gol (rouge), Saint Andre (jaune), Saint Paul (bleu), Sainte Suzanne (vert).



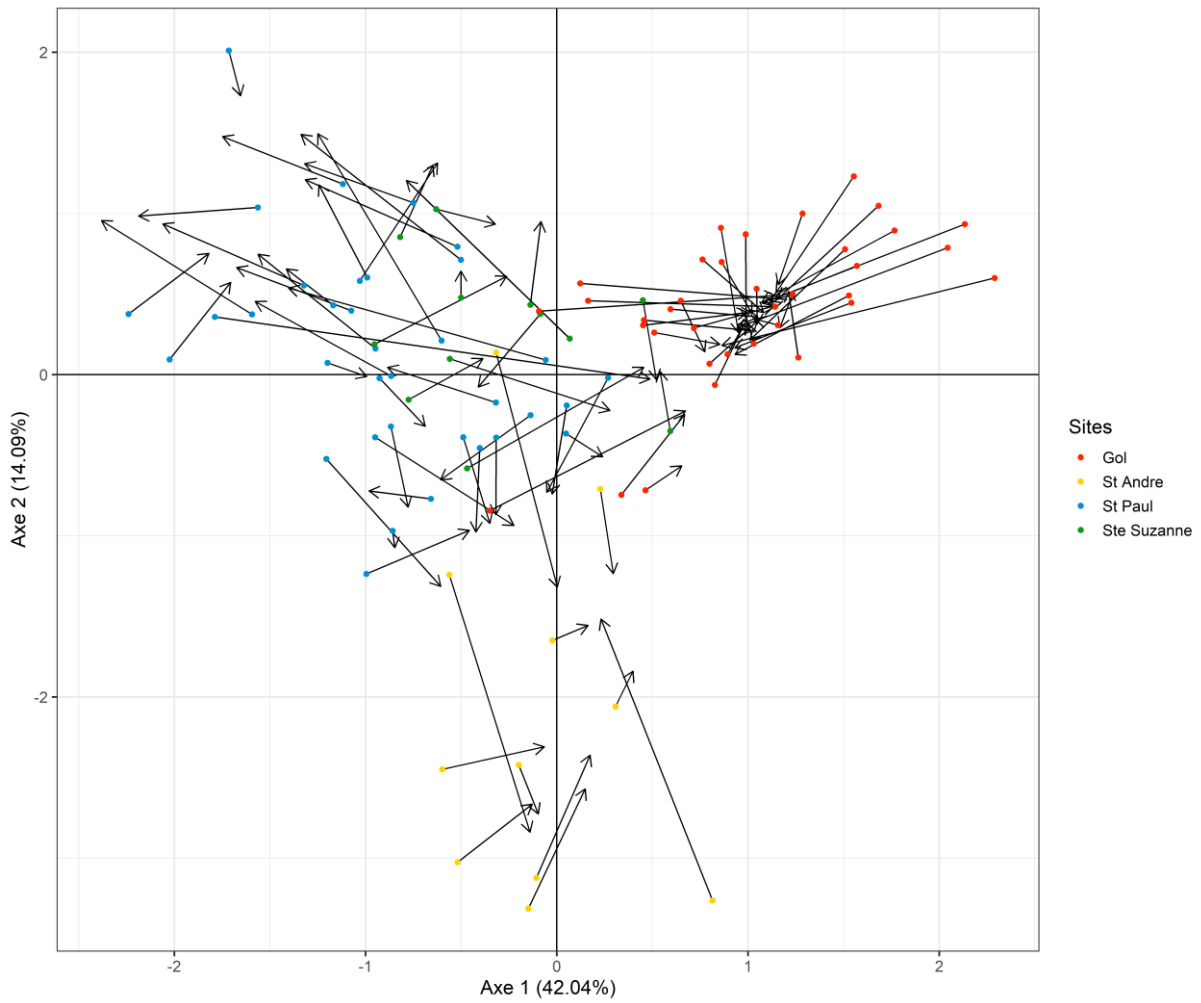
**Figure 20.** Analyse en coordonnées principales (PCoA) basée sur une matrice de distance de Bray-Curtis calculée sur les données moléculaires (marqueur 23S) obtenues à partir des échantillons prélevés dans les 4 sites de la Réunion : Gol (rouge), Saint Andre (jaune), Saint Paul (bleu), Sainte Suzanne (vert).

### V.2.3.2. Co-structure microscopie/metabarcoding

La co-structure (ou la congruence) entre les assemblages phytoplanctoniques obtenus par microscopie et les marqueurs moléculaires (i.e., 16S et 23S) pour les plans d'eau des DROM a également été évaluée par analyse de co-inertie. La congruence entre les données de microscopie et celles du marqueur 16S est modérée, avec 58.05% de l'inertie totale projetée expliquée par les deux premiers axes (Axe 1 : 43.444%, Axe 2 : 14.61%). En outre, le test de Monte-Carlo, montre un degré modéré de co-variation entre les deux ensembles de données ( $RV = 0,57$ ,  $p = 0.01$ , Figure 21). La congruence entre les données de microscopie et celles du marqueur 23S est également modérée, avec 56.13% de l'inertie totale projetée expliquée par les deux premiers axes de l'analyse (Axe 1 : 42.04%, Axe 2 : 14.09%), et un degré de co-variation modéré entre les deux ensembles de données estimé par le test de Monte-Carlo ( $RV = 0,57$ ,  $p 0.01$ , Figure 22). La co-structure entre les assemblages phytoplanctoniques obtenus par les deux marqueurs moléculaires (i.e., 16S et 23S) a également été testé. Une congruence modérée est également observée, avec les deux premiers axes expliquant 56.56% de l'inertie totale projetée (Axe 1 : 44.07%, Axe 2 : 12.49%), et une co-variation significative de 0.95 ( $RV$ ,  $p < 0.001$ ).



**Figure 21.** Analyse de co-inertie décrivant la co-structure entre les assemblages phytoplanctoniques estimés par des ACP sur les données de microscopie et moléculaires (16S). La longueur des flèches indique la différence de composition entre les données morphologiques (les points des flèches) et les données moléculaires (les pointes des flèches). Chaque point correspond à un échantillon et la couleur à sa provenance : Gol (rouge), Saint Andre (jaune), Saint Paul (bleu), et Sainte Suzanne (vert).



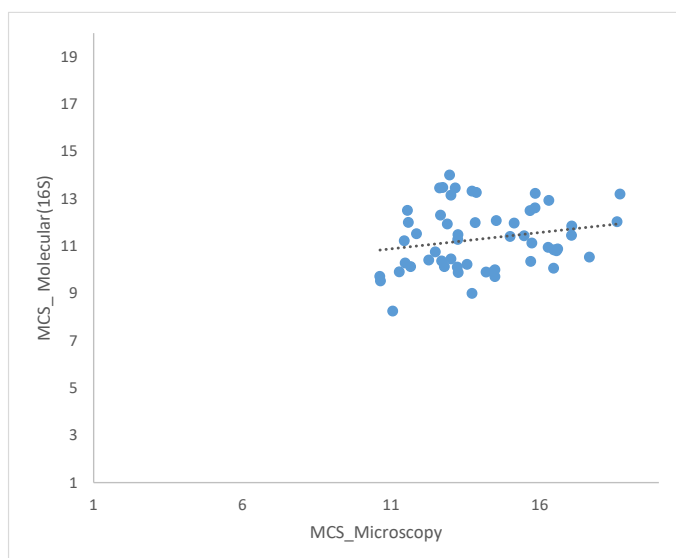
**Figure 22.** Analyse de co-inertie décrivant la co-structure entre les assemblages phytoplanctoniques estimés par des ACP sur les données de microscopie et moléculaires (23S). La longueur des flèches indique la différence de composition entre les données morphologiques (les points des flèches) et les données moléculaires (les pointes des flèches). Chaque point correspond à un échantillon et la couleur à sa provenance : Gol (rouge), Saint Andre (jaune), Saint Paul (bleu), et Sainte Suzanne (vert).

## VI. Tests d'application d'une métrique de bio-indication : MCS de l'IPLAC

Pour le test d'application d'une métrique de bioindication, nous avons choisi d'utiliser la MCS (Métrique de Composition Spécifique) de l'IPLAC (Laplace-Treytore & Feret 2016). Cette métrique est calculée sur la base de la formule de Zelinka et Marvan (1961), qui utilise le biovolume des espèces dans l'échantillon et leur profils écologiques spécifiques tels que définis dans l'IPLAC.

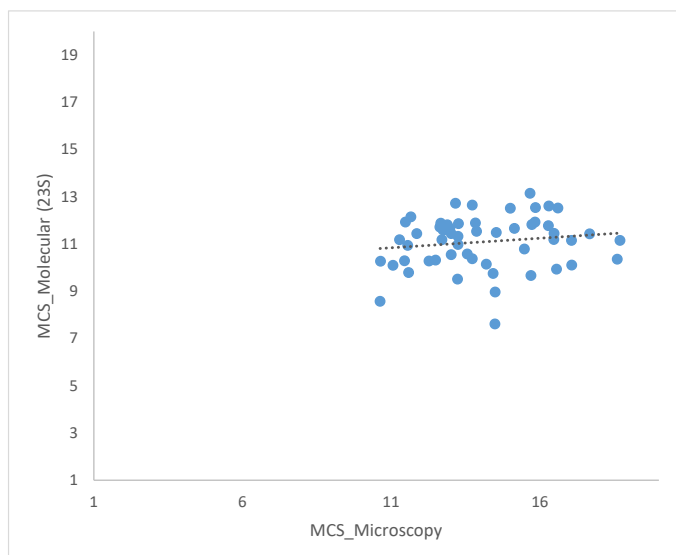
Nous avons donc calculé la MCS pour les listes floristiques des lacs OLA obtenues en microscopie et l'approche moléculaire 16S et 23S. Pour ces calculs, les listes floristiques de chaque méthode ont été exprimées pour chaque échantillon en biomasse relatives d'espèces pour la microscopie, et en abondance relative de reads d'espèces pour les méthodes moléculaires, ce qui est une modification de la méthode définie pour la MCS basée sur les biovolumes absolus. Cette transposition de la métrique pourra expliquer des différences ultérieures.

La comparaison entre les méthodes est présentée aux Figures 23 (microscopie et 16S) et 24 (microscopie et 23S).



**Figure 23.** Comparaison des valeurs de MCS obtenues avec la microscopie et l'approche moléculaire 16S. Aucune corrélation significative n'a été obtenue.





**Figure 24.** Comparaison des valeurs de MCS obtenues avec la microscopie et l'approche moléculaire 23S. Aucune corrélation significative n'a été obtenue.

Les valeurs entre les méthodes microscopiques et moléculaires ne sont pas corrélées significativement et les méthodes moléculaires 16S et 23S donnent des valeurs éloignées de la microscopie. Plusieurs raisons peuvent être mises en avant pour expliquer ces différences. En premier lieu, la transformation des données en abondance relative de reads et en biovolume relatif par espèce impacte fortement le calcul et la pertinence de l'utilisation de la métrique. Ensuite, le nombre de taxons pris en compte pour le calcul de la MCS pour les méthodes moléculaires est faible. En effet, sur un total de 35 taxons assignés à l'espèce dans la liste des taxons à partir du marqueur moléculaire 23S, 14 (40%) présentent des valeurs pour le calcul de MCS. De plus, sur un total de 30 taxons assignés à l'espèce dans la liste des taxons du marqueur moléculaire 16S, seulement 13 (43%) présentent des valeurs pour le calcul de MCS. Ces faibles nombres sont à comparer à l'approche microscopique où sur un total de 104 espèces identifiées dans l'inventaire, 54 (52%) présentent des valeurs pour le calcul du MCS. Par conséquent, i) les taxons dominants dans les méthodes moléculaires et morphologique ne sont pas les mêmes (cf. raisons données dans la section V.2.1) et ii) certains taxons ne sont pas identifiés ou seulement à des niveaux élevés (e.g., la famille) en méthode moléculaire car la base de référence ne présente pas de barcodes de taxons proches. Au final, les taxons les plus indicateurs, i.e., avec des cotes spécifiques élevées, ne sont pas retrouvés par les méthodes moléculaires (absence de barcodes correspondants). Les notes de MCS plafonnent alors à 13-14 avec ces méthodes alors qu'elles montent jusqu'à 18 avec les observations microscopiques.

Pour toutes ces raisons, l'utilisation de métriques de bioindication existantes et développées pour l'approche microscopique comme la MCS de l'IPLAC ne semble pas être la meilleure stratégie à suivre en vue du développement de métriques de bioindication basées sur des données moléculaires tant que ces dernières seront d'un niveau de détermination trop faible et incomplet.

## VII. Références

- Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, & S. P. Holmes, 2016. DADA2: High resolution sample inference from Illumina amplicon data. *Nature methods* 13: 581–583.
- Canino, A., A. Bouchez, C. Laplace-Treyture, I. Domaizon, & F. Rimet, 2021. Phytool, a ShinyApp to homogenise taxonomy of freshwater microalgae from DNA barcodes and microscopic observations. *Metabarcoding and Metagenomics Pensoft Publishers* 5: e74096.
- Decelle J., S. Romac, R.F. Stern, E.M. Bendif, A. Zingone, S. Audic, M.D. Guiry, L. Guillou, D. Tessier, F. Le Gall, P. Gourvil, A.L. Dos Santos, I. Probert, D. Vaultot, C. de Vargas, & R. Christen. 2015. PhytoREF: a reference database of the plastidial 16S rRNA gene of photosynthetic eukaryotes with curated taxonomy. *Molecular Ecology Resources* 15: 1435–1445.
- Djemiel C., D. Plassard, S. Terrat, O. Crouzet, J. Sauze, S. Mondy, V. Nowak, L. Wingate, J. Oggée, & P.A. Maron. 2020.  $\mu$ green-db: a reference database for the 23S rRNA gene of eukaryotic plastids and cyanobacteria. *Scientific Reports* 10: 5915.
- Gargominy, O., Terceirie, S., Régnier, C., Ramage, T., Dupont, P., Daszkiewicz, P. & Poncet, L. 2022. TAXREF, référentiel taxonomique pour la France : méthodologie, mise en œuvre et diffusion. Rapport PatriNat (OFB-CNRS-MNHN), Muséum national d’Histoire naturelle, Paris. 47 pp.
- Guillou L., D. Bachar, S. Audic, D. Bass, C. Berney, L. Bittner, C. Boutte G. Burgaud, C. de Vargas, J. Decelle, J. del Campo, J.R. Dolan, M. Dunthorn, B. Edvardsen, M. Holzmann, W.H.C.F. Kooistra, E. Lara, N. Le Bescot, R. Logares, F. Mahé, R. Massana, M. Montresor, R. Morard, F. Not, J. Pawlowski, I. Probert, A.L. Sauvadet, R. Siano, T. Stoeck, D. Vaultot, P. Zimmermann, & R. Christen. 2013. The Protist Ribosomal Reference database (PR<sup>2</sup>): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research* 41: D597–D604.
- Laplace-Treyture C, Hadoux E, Plaire M, Esmieu P, Dubertrand A, Crampe F (2017) PHYTOBS v3. 0: Phytoplankton counting tool in laboratory and calculation of IPLAC. Version 3.0. Java application.
- Laplace-Treyture, C., & T. Feret, 2016. Performance of the Phytoplankton Index for Lakes (IPLAC): A multimetric phytoplankton index to assess the ecological status of water bodies in France. *Ecological Indicators* 69: 686–698.
- Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17: 10–12.
- Quast C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, & F.O. Glöckner. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acid Research* 41: D590–D596.
- Rimet F., C. Chardon, L. Lainé, A. Bouchez, S. Jacquet, I. Domaizon, & J. Guillard. 2018. Thonon Culture Collection – TCC – a freshwater microalgae collection. <https://doi.org/10.15454/UQEMVW>.
- Rimet, F., E. Gusev, M. Kahlert, M. G. Kelly, M. Kulikovskiy, Y. Maltsev, D. G. Mann, M. Pfannkuchen, R. Trobajo, V. Vasselon, J. Zimmermann, & A. Bouchez, 2019. Diat.barcode, an open-access curated barcode library for diatoms. *Scientific Reports* 9: 1–12.
- Schloss, P. D., S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn, & C. F. Weber,

2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology* 75: 7537–7541.

Vautier, M., C. Chardon, C. Capelli, R. Kurmayer, N. Salmaso, & I. Domaizon, 2021. Plankton DNA extraction from Sterivex filter units. protocols.io. <https://www.protocols.io/view/plankton-dna-extraction-from-sterivex-filter-units-bvgzn3x6>.

Walsh, P.S., D.A. Metzger, & R. Higuchi, 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* 10: 206–513. Zelinka, M., & P. Marvan, 1961. Zur Prazisierung der biologischen Klassifikation der Reinheit fliessender Gewasser. *Arch.Hydrobiol.* 57: 389–407.

## VIII. Annexes

### Annexe 1 : Protocole d'extraction d'ADN sur filtre ouvert

# EXTRACTION D'ADN

## A partir de filtres découpés conservés sans tampon

Rédigé par Alexis CANINO (✉ alexis.canino@inrae.fr)

Dernière modification: 24 mai 2022

### Objectif:

Ce protocole décrit l'étape d'extraction d'ADN à partir d'un filtre (porosité 0.45 µm) préalablement découpé en morceaux et conservé au congélateur sans tampon. Il est appliqué dans ce contexte pour récupérer l'ADN du phytoplancton d'un échantillon d'eau filtrée. Il nécessite l'utilisation du kit NucleoSpin Soil de MACHERY-NAGEL (plus efficace que les kits "Water" pour éliminer les inhibiteurs de PCR). Ce protocole peut être réalisé pour traiter 24 échantillons, les temps indicatifs donnés sont propres à ce nombre d'échantillons. Le protocole exposé est une adaptation de celui développé par VAUTIER *et al.*, 2021.



## PROTOCOLE

### ETAPE 1 [Temps nécessaire env. 5 minute(s)]

#### Décongélation brève

- Décongeler les tubes rapidement (environ 5 minutes) à température ambiante et à l'abri de la lumière
- Pendant ce temps, allumer un incubateur à 37°C et un autre à 55°C
- Ajouter 1mL de tampon de lyse dans chacun des tubes

### ETAPE 2 [Temps nécessaire 10 minute(s)]

#### Retirer les particules du filtre

- Vortexer horizontalement les tubes pendant 10' à intensité maximale

### ETAPE 3 [Temps nécessaire env. 180 minute(s)]

#### Lyse cellulaire et digestion des débris cellulaires

- Ajouter 20 µL de lysozyme dans chaque tube
- Incuber à 37°C pendant 45 min les tubes sur un agitateur rotatif (placer également la solution de SDS dans l'incubateur)
- Ajouter 25 µL de protéinaseK et 50 µL de SDS 20% dans chaque tube
- Incuber à 55°C pendant 2 heures les tubes sur un agitateur rotatif (placer également la solution de buffer SE dans l'incubateur)

### ETAPE 4 [Temps nécessaire env. 15 minute(s)]

#### Transferts des contenus

- Pipeter la solution des tubes (ne pas récupérer les morceaux de filtres) et la transférer dans un nouveau tube 2mL
- Ajouter ensuite 200 µL de buffer SL3 dans chaque tube
- Incuber les tubes à 4°C (réfrigérateur) pendant 5 min
- Préparer pendant ce temps n colonnes rouges et n tubes eppendorf 2mL (n étant le nombre d'échantillons)

Remarque : Annoter (si nécessaire) ou placer de manière à éviter toute confusion

### ETAPE 5 [Temps nécessaire 10-15 minute(s)]

#### Elimination des inhibiteurs

- Centrifugation des tubes 2mL : 2 min à 11000xg
- Prélever 650 µL de surnageant maximum et les déposer sur les colonnes rouges respectives
- Centrifugation des colonnes rouges : 1 min à 11000xg
- Répéter les 2 opérations précédentes jusqu'à passer tous les surnageants sur les colonnes rouges associées (attention à ne pas croiser les échantillons !)

Remarque : A la fin de l'opération, vous devriez avoir à nouveau n tubes de 2mL contenant les surnageants filtrés (maximum 1.1mL par tube) au travers de la colonne. Les conserver et jeter les colonnes rouges.

## EQUIPEMENT

### Laboratoire:

- Hotte (avec UV)
- Incubateurs thermiques
- Centrifugeuse pour tubes 2mL
- Micropipettes (lot 10 à 1000 µL)
- Vortex horizontal\* Agitateur rotatif\*
- \*adaptateur tubes 2mL

### Consommables pour n échantillons :

- Gants
- Tubes eppendorf 2mL (2n)
- Tubes eppendorf 1.5mL (n)
- Tips 100 µL
- Tips 1000 µL
- Tubes de collection 2mL (n)\*<sup>MN</sup>
- "Colonnes rouges" (n)\*<sup>MN</sup>
- "Colonnes vertes" (n)\*<sup>MN</sup>

### Réactifs pour n échantillons :

- Lysozyme (n\*20 µL)
- ProtéinaseK (n\*25 µL)
- SDS 20% (n\*50 µL)
- Buffer SL3 (n\*400 µL)\*<sup>MN</sup>
- Buffer SB (n\*1000 µL)\*<sup>MN</sup>
- Buffer SW1 (n\*550 µL)\*<sup>MN</sup>
- Buffer SW2 (n\*1300 µL)\*<sup>MN</sup>
- Buffer SE (n\*30 à 50 µL)\*<sup>MN</sup>

\*<sup>MN</sup> inclus dans le kit NucleoSpin Soil de MACHERY-NAGEL

## PRÉCAUTIONS

Respecter les consignes de bases en biologie moléculaire, porter une blouse et des gants. Jeter les déchets chimiques dans un contenant prévu à cet effet par le laboratoire.

- Ajouter 250 µL de buffer SB par tubes de 2mL
- Vortex 5 secondes et centrifuger brièvement

**ETAPE 6** [Temps nécessaire 10-15 minute(s)]

#### Adsorption de l'ADN

- Préparer les colonnes vertes, les annoter et les placer sur des tubes de collection
- Prélever 650 µL de solution maximum et les déposer sur les colonnes vertes respectives
- Centrifugation des colonnes vertes : 1 min à 11000xg
- Vider le tube de collection dans un contenant prévu à cet effet, replacer la colonne verte sur le tube de collection et répéter les 2 opérations précédentes jusqu'à passer tout le contenu des tubes (attention à ne pas croiser les échantillons !)
- A la fin de l'opération, **conserver les n colonnes vertes (sur lesquelles sont adsorbés les ADN) et les tubes de collections, qui seront utilisés pour les lavages.**

**ETAPE 7** [Temps nécessaire env. 7 minute(s)]

#### Lavages des colonnes

- Ajouter 500 µL de buffer SB
- Vortexer brièvement et centrifuger 30 sec. à 11000xg
- Vider le tube de collecte
- Ajouter 550 µL de buffer SW1
- Vortexer brièvement et centrifuger 30 sec. à 11000xg
- Vider le tube de collecte
- Ajouter 650 µL de buffer SW2
- Vortexer brièvement et centrifuger 30 sec. à 11000xg
- Vider le tube de collecte
- Refaire le dernier lavage au buffer SW une seconde fois
- Centrifugation finale : 2 min à 11000xg pour sécher les colonnes

**ETAPE 8** [Temps nécessaire env. 4 minute(s)]

#### Elution de l'ADN

- Placer la colonne verte sur le tube final (i.e. celui qui contiendra l'extrait d'ADN et sera conservé au frais) : un tube eppendorf 1.5mL
- Ajouter 30 à 50 µL de buffer SE (incubé à 55°)
- Ne pas fermer le bouchon et laisser incuber à température ambiante 1 min 30 sec.
- Centrifugation : 30 sec. à 11000xg
- Jeter la colonne verte
- Conserver à -20°C

## RÉFÉRENCES ET CONTACTS

Marine Vautier, Cécile Chardon, Isabelle Domaizon 2021. Fish eDNA: DNA extraction from water samples filtered through Sterivex filter unit and preserved in buffer. protocols.io <https://dx.doi.org/10.17504/protocols.io.bfk8jkw>

CANINO Alexis (✉ [alexis.canino@inrae.fr](mailto:alexis.canino@inrae.fr))  
RIMET Frédéric (✉ [frederic.rimet@inrae.fr](mailto:frederic.rimet@inrae.fr))

## Annexe 2 : Protocole d'extraction d'ADN sur filtre Sterivex

# EXTRACTION D'ADN

A partir de Sterivex conservés avec du tampon de lyse

Rédigé par Alexis CANINO (✉ alexis.canino@inrae.fr)

Dernière modification 24 mai 2022

### Objectif:

Ce protocole décrit l'étape d'extraction d'ADN à partir d'un filtre encapsulé de type Sterivex (porosité utilisé ici : 0.45 µm, conservé dans du tampon de lyse. Il est appliqué dans ce contexte pour récupérer l'ADN du phytoplancton d'un échantillon d'eau filtré. Il nécessite l'utilisation du kit NucleoSpin Soil de MACHEREY-NAGEL (plus efficace que les kits "Water" pour éliminer les inhibiteurs de PCR). Ce protocole peut être réalisé pour traiter 12 Sterivex, les temps indicatifs donnés sont propres à ce nombre d'échantillons. Le protocole exposé s'appuie sur celui développé par VAUTIER et al., 2021.



## PROTOCOLE

**ETAPE 1** [Temps nécessaire env. 20 minute(s)]

### Décongélation du Sterivex

Dans le cas où les Sterivex ont été conservés au congélateur

- Laisser les Sterivex environ 20 minutes à température ambiante à l'abri de la lumière
- Pendant ce temps, allumer un incubateur à 37°C et un autre à 55°C

**ETAPE 2** [Temps nécessaire 5 minute(s)]

### Retirer les particules du filtres

- Vortexer horizontalement les Sterivex pendant 5' à intensité maximale

**ETAPE 3** [Temps nécessaire env. 180 minute(s)]

### Lyse cellulaire et digestion des débris cellulaires

- Ajouter 40 µL de lysozyme dans chaque Sterivex
- Incuber à 37°C pendant 45 min les Sterivex sur un agitateur rotatif (placer également la solution de SDS dans l'incubateur)
- Ajouter 50 µL de protéinaseK et 100 µL de SDS 20% dans chaque Sterivex
- Incuber à 55°C pendant 2 heures les Sterivex sur un agitateur rotatif (placer également la solution de buffer SE dans l'incubateur)

**ETAPE 4** [Temps nécessaire env. 15 minute(s)]

### Récupération et transferts des contenus

- Aspirer le contenu des Sterivex à l'aide d'une petite seringue 5mL
- Répartir le contenu d'un Sterivex dans 2 tubes eppendorf 2mL (soit environ 1.1mL par tube)
- Ajouter ensuite 200 µL de buffer SL3 par tube
- Incuber les tubes à 4°C (réfrigérateur) pendant 5 min
- Préparer pendant ce temps n colonnes rouges et 2n tubes eppendorf 2mL (n étant le nombre d'échantillons)

Remarque: Annoter (si nécessaire) ou placer de manière à éviter toute confusion

**ETAPE 5** [Temps nécessaire 10-15 minute(s)]

### Élimination des inhibiteurs

- Centrifugation des tubes 2mL : 2 min à 11000xg
- Prélever 650 µL de surnageant maximum et les déposer sur les colonnes rouges respectives
- Centrifugation des colonnes rouges : 1 min à 11000xg
- Répéter les 2 opérations précédentes jusqu'à passer tous les surnageants sur les colonnes rouges associées (attention à ne pas croiser les échantillons !)

## EQUIPEMENT

Laboratoire :

- Hotte (avec UV)
- Incubateurs thermiques
- Centrifugeuse pour tubes 2mL
- Micropipettes (lot 10 à 1000 µL)
- Vortex horizontal\* Agitateur rotatif\*

\*adaptateur tubes 15mL

Consommables pour n échantillons :

- Gants Tubes eppendorf 2mL (4n)
- Tubes eppendorf 1.5mL (n)
- Seringue 5mL (n) Tips 100 µL
- Tips 1000µL
- Tubes de collection 2mL (n)\*<sup>MN</sup>
- "Colonnes rouges" (n)\*<sup>MN</sup>
- "Colonnes vertes" (n)\*<sup>MN</sup>

Réactifs pour n échantillons :

- Lysozyme (n\*40 µL) [0.05g/L]
- ProtéinaseK (n\*50 µL) [20mg/mL]
- SDS 20% (n\*100 µL)
- Buffer SL3 (n\*400 µL)\*<sup>MN</sup>
- Buffer SB (n\*1000 µL)\*<sup>MN</sup>
- Buffer SW1 (n\*550 µL)\*<sup>MN</sup>
- Buffer SW2 (n\*1300 µL)\*<sup>MN</sup>
- Buffer SE (n\*30 à 50 µL)\*<sup>MN</sup>
- Ethanol 96° qualité biomol (pour SW2)

\*<sup>MN</sup> inclus dans le kit NucleoSpin Soil de MACHEREY-NAGEL

## PRÉCAUTIONS

Respecter les consignes de bases en biologie moléculaire, porter une blouse et des gants. Jeter les déchets chimiques dans un contenant prévu à cet effet par le laboratoire.

**Remarque :** A la fin de l'opération, vous devriez avoir à nouveau 2n tubes de 2mL contenant les surnageants filtrés (maximum 1.1mL par tube) au travers de la colonne. Les conserver et jeter les colonnes rouges.

- Ajouter 250 µL de buffer SB par tubes de 2mL
- Vortex 5 secondes et centrifuger brièvement

---

#### ETAPE 6

[Temps nécessaire 10-15 minute(s)]

##### Adsorption de l'ADN

- Préparer les colonnes vertes, les annoter et les placer sur des tubes de collection
- Prélever 650 µL de solution maximum et les déposer sur les colonnes vertes respectives
- Centrifugation des colonnes vertes : 1 min à 11000xg
- Vider le tube de collection dans un contenant prévu à cet effet, replacer la colonne verte sur le tube de collection et répéter les 2 opérations précédentes jusqu'à passer tout le contenu des tubes (attention à ne pas croiser les échantillons !)
- A la fin de l'opération, conserver les n colonnes vertes (sur lesquelles sont adsorbés les ADN) et les tubes de collections, qui seront utilisés pour les lavages.

---

#### ETAPE 7

[Temps nécessaire env. 7 minute(s)]

##### Lavages des colonnes

- Ajouter 500 µL de buffer SB
- Vortexer brièvement et centrifuger 30 sec. à 11000xg
- Vider le tube de collecte
- Ajouter 550 µL de buffer SW1
- Vortexer brièvement et centrifuger 30 sec. à 11000xg
- Vider le tube de collecte
- Ajouter 650 µL de buffer SW2
- Vortexer brièvement et centrifuger 30 sec. à 11000xg
- Vider le tube de collecte
- Refaire le dernier lavage au buffer SW2 une seconde fois
- Centrifugation finale : 2 min à 11000xg pour sécher les colonnes

---

#### ETAPE 8

[Temps nécessaire env. 4 minute(s)]

##### Elution de l'ADN

- Placer la colonne verte sur le tube final (*i.e.* celui qui contiendra l'extrait d'ADN et sera conservé au frais) : un tube eppendorf 1.5mL
- Ajouter de buffer SE (incubé à 55°)
- Ne pas fermer le bouchon et laisser incuber à température ambiante 1 min 30 sec.
- Centrifugation : 30 sec. à 11000xg
- Jeter la colonne verte
- Conserver à -20°C

## RÉFÉRENCES ET CONTACTS

Marine Vautier, Cécile Chardon, Isabelle Domaizon 2021. Fish eDNA: DNA extraction from water samples filtered through Sterivex filter unit and preserved in buffer. protocols.io <https://dx.doi.org/10.17504/protocols.io.bfk8jkzw>

CANINO Alexis (✉ [alexis.canino@inrae.fr](mailto:alexis.canino@inrae.fr))  
RIMET Frédéric (✉ [frederic.rimet@inrae.fr](mailto:frederic.rimet@inrae.fr))

## Annexe 3 : Protocole d'extraction d'ADN Genelute

# EXTRACTION D'ADN

## A partir d'un prélèvement d'une culture pures de microalgues

Rédigé par Alexis CANINO ✉ alexis.canino@inrae.fr

Dernière modification 23 mai 2022

### Objectif:

Ce protocole décrit l'étape d'extraction d'ADN à partir d'un prélèvement de 2mL d'une culture pure de microalgues (ex. TCC). Ce dernier est facile à mettre en place et peu coûteux, cependant son efficacité peut varier en fonction des souches de microalgues (et selon leur densité, structure/paroie cellulaire, taille des cellules etc.). Si les souches ne sont pas axéniques (ce qui est souvent le cas) l'ADN extrait peut alors être principalement bactérien. S'assurer donc de l'efficacité du protocole sur la souche et garder ces lignes en considération. Le temps estimé est donné pour 10 échantillons. Ce protocole peut être appliqué pour des échantillons environnementaux également, mais il ne sera pas aussi efficace que d'autres pour retirer les inhibiteurs de PCR.

**CARTELL**  
CENTRE ALPIN DE RECHERCHE  
SUR LES RESEAUX PROPRIETAIRES  
ET LES SYSTEMES LIMNIQUES

**INRAE**



## PROTOCOLE

### ETAPE 1

[Temps nécessaire (...) minute(s)]

#### Prélèvement (ou décongélation des échantillons déjà prélevés)

Dans le cas où les échantillons sont directement prélevés des cultures :

- Sous hotte : agiter le flacon de culture, l'ouvrir et prélever environ 2mL de culture dans un tube 2mL

Dans le cas où les échantillons extraits ont été conservés au congélateur :

- Décongeler les tubes environ 20 minutes à température ambiante et à l'abri de la lumière

Pendant ce temps, allumer un incubateur à 37°C et un autre à 55°C

### ETAPE 2

[Temps nécessaire 30 minute(s)]

#### Préparation du culot

- Centrifuger le(s) tube(s) 30 min. à 13000rpm et à 4°C

Pendant ce temps placer la solution de SDS dans l'incubateur à 37°C

### ETAPE 3

[Temps nécessaire 10 minute(s)]

#### Récupération du culot

- Retirer le surnageant
- Ajouter 300 µL de tampon d'éluion (TE)
- Ajouter 200 µL de tampon de lyse (TL)

**Remarque :** Resuspendre le culot en faisant des aspirations/refoulements lors de l'ajout des tampons

### ETAPE 4

[Temps nécessaire 17 à 37 minute(s)]

#### Lyse cellulaire

Choc thermique :

- Placer les tubes 15 min. à -80°C
- Placer ensuite les tubes 2 min. à 55°C

Lyse ultrasonique (facultatif, pour cellules résistantes) :

- Placer les tubes 20 min. dans un bain à ultrasons

### ETAPE 5

[Temps nécessaire env. 95 minute(s)]

#### Digestion des débris cellulaires

- Ajouter 50 µL de SDS 10% dans chaque tube
- Ajouter 10 µL de protéinaseK dans chaque tube
- Incuber à 37°C pendant 1 heure les tubes sur un agitateur rotatif
- Incuber ensuite à 55°C pendant 20 min. les tubes sur un agitateur rotatif
- Centrifugation : 3min. à 13000rpm à 4°C

## EQUIPEMENT

### Laboratoire :

- Hotte (avec UV)
- Incubateurs thermiques
- Centrifugeuse (réfrigérante) pour tubes 2mL
- Micropipettes (lot 10 à 1000 µL)
- SpeedVac ou cloche à vide
- Agitateur rotatif\*
- Bain à ultrasons (facultatif)

\*adaptateur tubes 2mL

### Consommables pour n échantillons :

- Gants
- Tubes eppendorf 2mL (2n)
- Tips 10 µL
- Tips 100 µL
- Tips 1000 µL

### Réactifs pour n échantillons :

- ProtéinaseK (n\*10 µL) [20mg/mL]
- SDS 10% (n\*50 µL)
- Tampon d'éluion (TE) (n\*330 µL)
- Tampon de lyse (TL) (n\*200 µL)
- Acétate de Sodium (n\*50 µL) [3M, pH 5.2]
- GenElute<sup>TM</sup>-LPA (n\*1 µL) [25 µg/µL]
- Isopropanol\* qualité biomol. (n\*500 µL)
- Ethanol 80% qualité biomol. (n\*300 µL)

\*A conserver à -20°C

## PRÉCAUTIONS

Respecter les consignes de bases en biologie moléculaire, porter une blouse et des gants. Jeter les déchets chimiques dans un contenant prévu à cet effet par le laboratoire.

Avant de commencer le protocole, s'assurer d'avoir l'ensemble des réactifs nécessaires !



- Transférer les surnageants dans de nouveaux tubes 2mL

---

#### ETAPE 6 [Temps nécessaire env. 15 minute(s)]

##### Précipitation de l'ADN

- Ajouter 50 µL d'acétate de sodium
- Ajouter 1 µL de GenElute
- Laisser 5 à 10 min. à température ambiante
- Ajouter 500 µL d'isopropanol froid (stocké à -20°C)

Remarque: A ce stade il est possible de prendre une pause (et même de reprendre le jour suivant si nécessaire) en laissant les tubes au réfrigérateur

---

#### ETAPE 7 [Temps nécessaire env. 35 minute(s)]

##### Lavage des culots

- Centrifugation : 10 min à 12000xg à 4°C
- Retirer le surnageant
- Laver le culot avec 200 µL d'éthanol biomol. 80%
- Centrifugation : 10 min à 12000xg à 4°C
- Retirer le surnageant
- Répéter les 3 opérations précédentes (éthanol, centrifugation, élimination du surnageant)

---

#### ETAPE 8 [Temps nécessaire 20 minute(s)]

##### Séchage des culots

Remarque: Préférer utiliser un SpeedVac si possible, sinon cette étape peut se faire dans une cloche sous vide.

- Placer les tubes 20 minutes au SpeedVac ou dans une cloche sous vide pour évaporer les restes d'éthanol

---

#### ETAPE 9 [Temps nécessaire 65 minute(s)]

##### Elution des culots

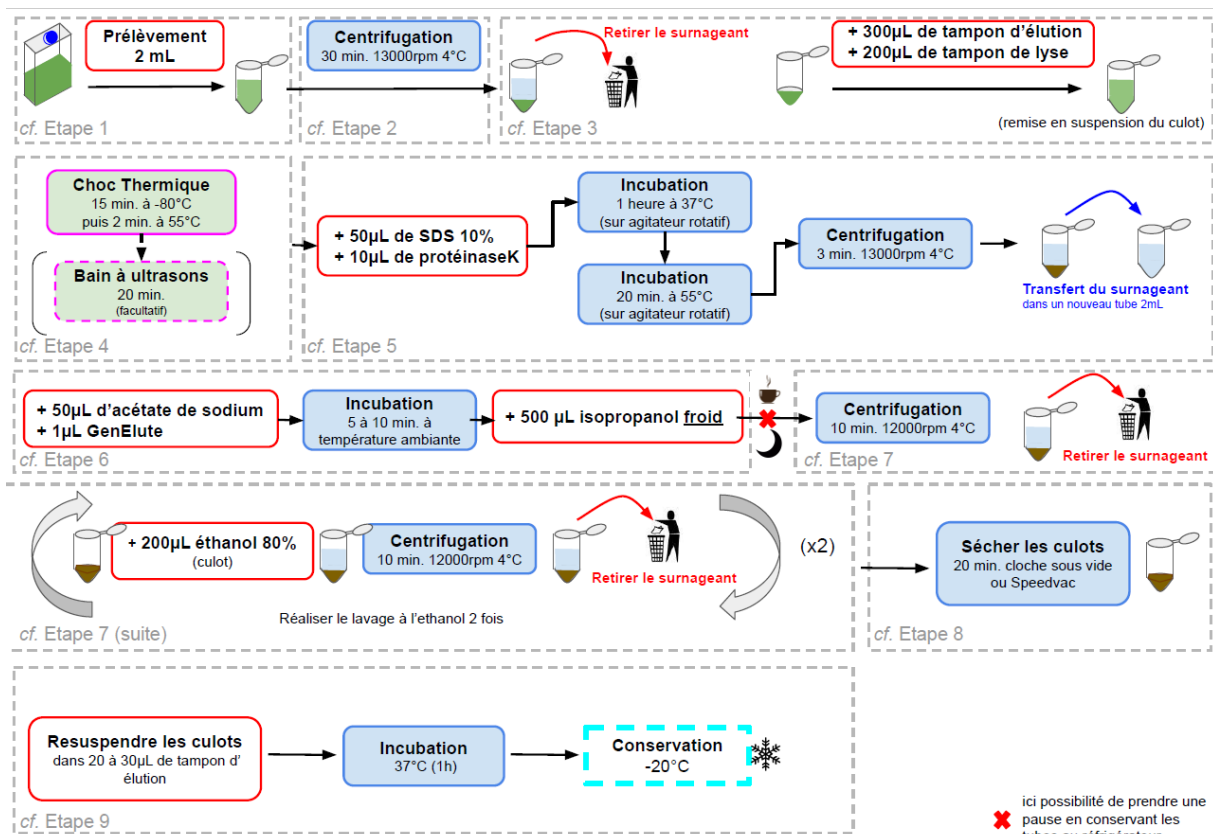
- Resuspendre les culots dans 20 à 30 µL de tampon d'éluion (TE) en fonction de la taille du culot et de la concentration désirée de l'extrait
- Incuber 1 heure à 37°C (à l'obscurité)
- Conserver les extraits à -20°C

## RÉFÉRENCES ET CONTACTS

Le protocole de base est brièvement expliqué ici :

Chonova T, Keck F, Labanowski J et al. (2016) Separate treatment of hospital and urban wastewaters : A real scale comparison of effluents and their effect on microbial communities. Science of The Total Environment, 542, 965-975.

CANINO Alexis (✉ alexis.canino@inrae.fr)  
RIMET Frédéric (✉ frederic.rimet@inrae.fr)



# EXTRACTION D'ADN

## A partir d'un isolement de cellule(s) de microalgues

Rédigé par Alexis CANINO ✉ alexis.canino@inrae.fr

Dernière modification 24 mai 2022

### Objectif:

Ce protocole décrit l'extraction d'ADN à partir d'un isolement de une ou quelques cellules de microalgues, préalablement réalisé au microscope. Il a été mis en place par Walsh et al., 1991 et est réutilisé aussi par Richlen & Barber, 2005 pour des dinoflagellés. Une légère optimisation a été apportée, afin d'augmenter l'efficacité de la lyse cellulaire, qui peut être limitée pour de petites cellules avec des parois résistantes. Ce protocole est intéressant pour ceux qui souhaitent faire des approches 'single-cell PCR'. Des variantes ont été réalisées pour appliquer ce dernier sur des diatomées préservées dans du lugol (Hamilton et al., 2015).

## PROTOCOLE

### ETAPE 1 [Temps nécessaire - minute(s)]

#### Préparation des alicots de Chelex 10% (à réaliser en amont)

Sous hotte stérile :

- Peser 1g. de Chelex 100 et déposer dans un tube falcon (15mL) stérile
- Dissoudre en ajoutant 10mL d'eau stérile qualité biomol.
- Bien agiter (secouer ou vortexer)
- Réaliser des alicots de 50 µL dans des tubes PCR

Remarque : lors de la préparation des alicots de 50 µL de Chelex 10% s'assurer de toujours bien agiter la solution mère à chaque pipetage pour avoir des alicots homogènes. Le Chelex étant une résine constituée de petites billes, celles-ci ont tendance à rapidement sédimenter.

### ETAPE 2 [Temps nécessaire - minute(s)]

#### Isolation de(s) cellule(s) (à réaliser en amont)

Sous un microscope, et à l'aide d'un micromanipulateur, de pipettes pasteur travaillées ou de cônes :

- Isoler une ou plusieurs cellules
- Réaliser des lavages de celle(s)-ci avec de l'eau stérile (qualité biomol.) par aspirations/refoulements
- Placer les cellules dans les alicots préalablement préparés de solution de Chelex à 10%

### ETAPE 3 [Temps nécessaire env. 30 minute(s)]

#### Lyse cellulaire

- Vortexer le(s) tube(s) 5 sec.
- Centrifuger le(s) tube(s) 10 à 15 sec.
- Placer les tubes 5 min. à -80°C
- Placer les tubes 20 min. à 95°C (dans un thermocycleur)
- Vortexer le(s) tube(s) 10 sec.
- Centrifuger brièvement

### ETAPE 4a [Temps nécessaire - minute(s)]

#### Stockage

- Stocker au réfrigérateur (4°C) à court terme
- Stocker au congélateur (-20°C) pour un stockage à plus long terme

### ETAPE 4b [Temps nécessaire - minute(s)]

#### Utilisation immédiate pour du single-cell PCR

Juste avant de déposer votre extrait dans votre mélange réactif de PCR :

- Vortexer le(s) tube(s) 5 sec.
- Centrifuger le(s) tube(s) 10 à 15 sec.
- Pipeter le surnageant (éviter les billes au fond du tube)



## EQUIPEMENT

### Laboratoire:

- Hotte (avec UV)
- Vortex de paillasse
- Centrifugeuse de paillasse pour tubes PCR
- Micropipettes (lot 10 à 100 µL)
- Micromanipulateur (facultatif)
- Thermocycleur
- Accès à un congélateur -80°C
- Balance de laboratoire

### Consommables:

- Gants
- Tubes PCR (200 µL)
- Tube/Corning falcon 15mL
- Tips 10 µL
- Tips 100 µL
- Pipettes pasteur en verre (facultatif)

### Réactifs :

- H<sub>2</sub>O qualité biomol.
- Chelex 100 (BIORAD) 10%

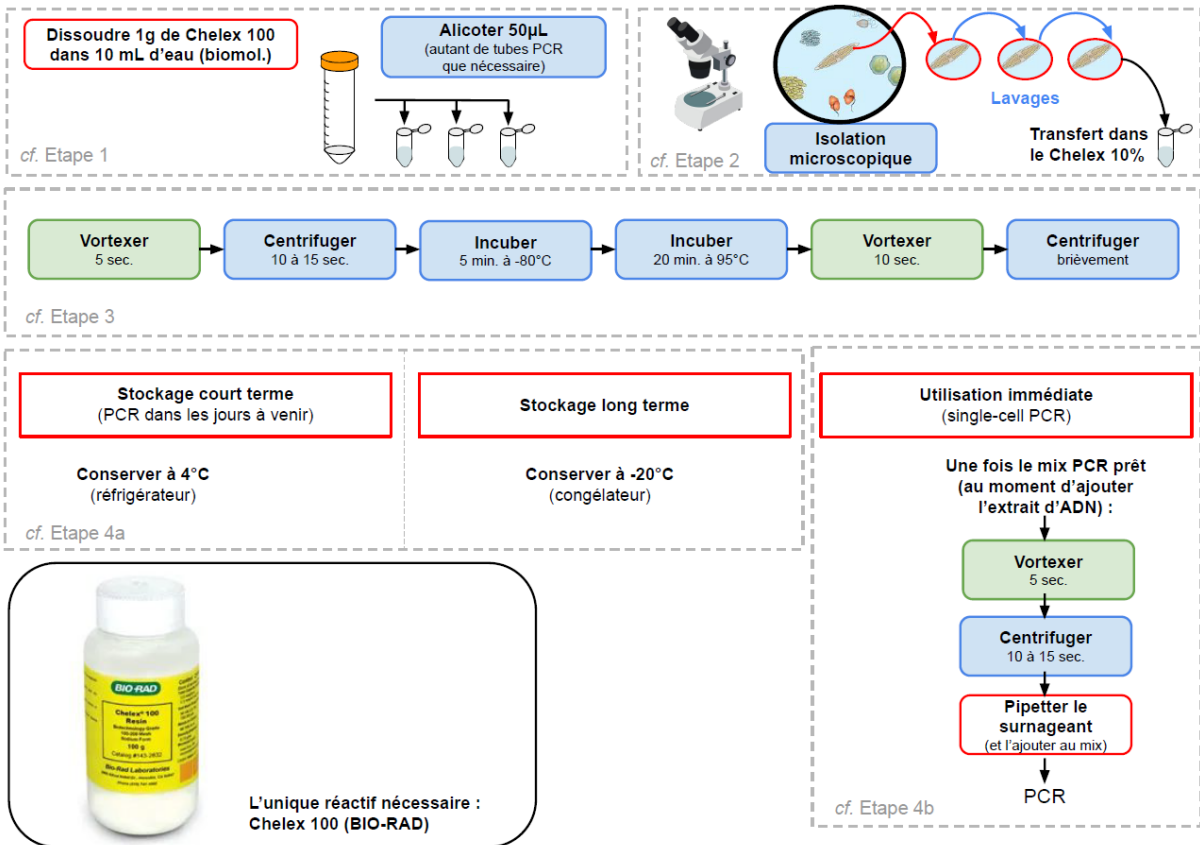
## PRÉCAUTIONS

Respecter les consignes de bases en biologie moléculaire, porter une blouse et des gants. Jeter les déchets chimiques dans un contenant prévu à cet effet par le laboratoire.

## RÉFÉRENCES ET CONTACTS

- Hamilton, Paul B., Keely E. Lefebvre, and Roger D. Bull. "Single cell PCR amplification of diatoms using fresh and preserved samples." *Frontiers in microbiology* 6 (2015): 1084.
- Richlen, M. L., and P. H. Barber. "A technique for the rapid extraction of microalgal DNA from single live and preserved cells." *Molecular Ecology Notes* 5.3 (2005): 688-691.
- Walsh, P. Sean, David A. Metzger, and Russell Higuchi. "Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material." *Biotechniques* 54.3 (2013): 134-139.

CANINO Alexis ✉ alexis.canino@inrae.fr  
RIMET Frédéric ✉ frederic.rimet@inrae.fr



## Annexe 5 : Protocole de filtration sur filtre ouvert

# PROTOCOLE DE FILTRATION

## Filtration d'échantillons d'eau pour collecter l'ADNe phytoplancton

Rédigé par Alexis CANINO (✉ alexis.canino@inrae.fr)

Dernière modification 17 mai 2022

### Objectif:

Ce protocole décrit la filtration en laboratoire (sur filtres ouverts) d'échantillons d'eau en vue de leur analyse en metabarcoding pour le phytoplancton. Il peut également être réalisé sur le terrain à condition d'avoir le matériel nécessaire et de respecter les précautions.

### PROTOCOLE À RÉALISER EN LABORATOIRE

#### ETAPE 1 [Temps nécessaire env. 1 minute(s)]

##### Préparation du matériel

Prendre l'échantillon d'eau à filtrer (prélèvement intégré demandé par la DCE pour le suivi microscopique du phytoplancton par exemple)

Prendre une unité de filtration adaptée au diamètre du filtre à utiliser et une pompe péristaltique (avec les raccords nécessaires)

#### ETAPE 2 [Temps nécessaire env. 5 minute(s)]

##### Stérilisation du matériel nécessaire

- Porter le matériel de protection adapté
- Nettoyer les parties supérieures de l'unité de filtration au peroxyde d'hydrogène ( $H_2O_2$ ). Le collecteur du filtrat, en bas de l'unité n'a donc pas besoin d'être nettoyé.
- Faire tremper 1 minute les ustensiles (pinces ciseaux) dans l' $H_2O_2$
- Rincer l'ensemble du matériel nettoyé à l'eau MilliQ stérile et relier l'unité de filtration à la pompe

#### ETAPE 3 [Temps nécessaire env. 3 minute(s)]

##### Filtration de l'échantillon d'eau

- Utiliser les pinces pour placer le filtre proprement sur l'unité de filtration
- Agiter la bouteille contenant l'échantillon et déposer son contenu dans l'unité de filtration
- Allumer la pompe, prendre garde à ce que l'aiguille du manomètre ne dépasse pas 50kPa

#### ETAPE 4 [Temps nécessaire env. 2 minute(s)]

##### Découpe du filtre

- Découper le filtre en 8 (ou plus) morceaux sur l'unité de filtration
- Utiliser la pince pour mettre ces morceaux dans un tube 2mL stérile

#### ETAPE 5 [Temps nécessaire env. 1 minute(s)]

##### Conservation

- Stocker le tube contenant le filtre découpé au congélateur

#### Remarque(s) :

Si vous devez filtrer plusieurs échantillons à la suite, vous pouvez reconduire ce protocole à partir de l'ETAPE 2. Se débarrasser du  $H_2O_2$  dans un bidon prévu à cet effet.



### EQUIPEMENT

- Gants
- Unité de filtration
- Filtre (ref) 0.45  $\mu$ m
- Pinces à épiler
- Petit ciseaux
- Pompe péristaltique
- Marqueur indélébile
- Tube eppendorf 2mL


### PRODUIT(S) CHIMIQUE(S)

- Péroxyde d'hydrogène 35% ( $H_2O_2$ )

### EQUIPEMENT(S) DE PROTECTION

- Gants
- Lunettes de protection
- Blouse

### PRÉCAUTIONS

-  **Produit chimique**  
 $H_2O_2$  à manier avec prudence !

### RÉFÉRENCES ET CONTACTS

Pour toutes questions :

- CANINO Alexis (✉ alexis.canino@inrae.fr)
- RIMET Frédéric (✉ frederic.rimet@inrae.fr)

# PROTOCOLE D'ÉCHANTILLONNAGE

## Echantillonnage de l'ADNe phytoplancton

Rédigé par Alexis CANINO ✉ alexis.canino@inrae.fr

Dernière modification 17 mai 2022

### Objectif:

Ce protocole décrit un prélèvement ADN du phytoplancton en évitant le risque de contamination. Ce dernier s'appuie sur celui de Domaizon et al, 2019\*.



## PROTOCOLE RÉALISABLE SUR LE TERRAIN

### ETAPE 1 [Temps nécessaire (...) minute(s)]

Réaliser le prélèvement intégré demandé par la DCE pour le suivi microscopique du phytoplancton

### ETAPE 2 [Temps nécessaire env. 1 minute(s)]

#### Préparation du matériel nécessaire

- Sortir une seringue 60mL et un Sterivex de leur emballage, placer le Sterivex ouvert dans une boîte de prélèvement
- Placer à proximité les bouchons Sterivex, le tube avec le tampon de lyse (TL) et la seringue 5mL

### ETAPE 3 [Temps nécessaire 2 à 5 minute(s)]

#### Filtration de l'échantillon d'eau

- Utiliser la seringue pour prélever 50mL de l'échantillon dans la bouteille de collecte
- Placer le bout de la seringue en haut du Sterivex, visser et filtrer l'intégralité
- Répéter les opérations précédentes jusqu'à filtrer 250mL de l'échantillon (si possible)
- Si le filtre colmate et que la filtration devient difficile, s'arrêter et noter le volume total filtré
- A la fin de la filtration, vider la seringue puis la remplir d'air et la placer sur le Sterivex afin de chasser l'eau dans celui-ci
- Prendre un bouchon et boucher le bas du Sterivex

### ETAPE 4 [Temps nécessaire env. 1 minute(s)]

#### Ajout du tampon de lyse

- Sortir la petite seringue (5mL) et prendre le tube contenant le tampon de lyse (TL)
- Prélever 2mL maximum du tampon
- Insérer le bout de la petite seringue dans le haut du Sterivex et pousser pour remplir le Sterivex de tampon
- Prendre le 2ème bouchon et boucher le haut

### ETAPE 5 [Temps nécessaire env. 1 minute(s)]

#### Conservation

- Placer le Sterivex dans un sachet ziplock, annoter celui-ci avec les informations relatives à l'échantillon (nom, lieu, date, volume filtré)
- Conserver à l'abri du soleil et de la chaleur, dans une glacière, jusqu'au retour au laboratoire où les échantillons doivent être placés au réfrigérateur dans l'attente de leur envoi

## EQUIPEMENT

- Gants (facultatif)
- Boîte (facultatif)
- Seringue 60mL (x1)
- Seringue 5mL (x1)
- Sterivex (x1)
- Bouchons Sterivex (x2)
- Tampon de Lyse (2mL)
- Sachet ziplock (x1)
- Marqueur indélébile

## PRODUIT(S) CHIMIQUE(S)

- Tampon de lyse (Sucrose, Tris, EDTA)

## RÉFÉRENCES ET CONTACTS

\*Isabelle Domaizon, Rainer Kurmayer, Camilla Capelli, Cécile Chardon, Peter Hufnagl, Marine Vautier, Nico Salmaso 2019. Lake plankton sample collection from the field for downstream molecular analysis. protocols.io  
<https://dx.doi.org/10.17504/protocols.io.xn6fmhe>

CANINO Alexis ✉ alexis.canino@inrae.fr  
RIMET Frédéric ✉ frederic.rimet@inrae.fr

## TUTORIEL VIDÉO



## Annexe 7 : Protocole d'amplification par PCR en vue d'un séquençage par la plateforme PGTB

# AMPLIFICATION D'ADN

## PCR en vue d'un séquençage à la PGTB

Rédigé par Alexis CANINO (✉ alexis.canino@inrae.fr)

Dernière modification 24 mai 2022

### Objectif:

Ce protocole est celui recommandé par la Plateforme Génome Transcriptome de Bordeaux (PGTB) pour l'amplification de barcodes ADN destinés à un séquençage MiSeq. Ce dernier a été rédigé par Emilie Chancerel, et est résumé et adapté ici à l'utilisation des couples d'amorces CYA359F/CYA781R (v3-v4 16S) et 108F/108R (UPA 23S) pour le metabarcoding phytoplancton.

## PROTOCOLE

### ETAPE 1 [Temps nécessaire - minute(s)]

#### Prérequis

En vue d'une amplification destinée au séquençage :

- Commander les amorces nécessaires en intégrant les séquences universelles de PGTB
- Réaliser des dilutions des amorces 'mères' : les alicots à 1 µM seront utilisés dans les mélanges pour la PCR

Séquences universelles à intégrer (en violet) :

Forward :

5' *TCTGTCGGCAGCGTCAGATGTGTATAAGAGACAG*XXXX(amorce Forward)XXX3'

Reverse :

5' *GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG*XXXX(amorce reverse)XXX3'

### ETAPE 2 [Temps nécessaire - minute(s)]

#### Préparation du mélange réactionnel

Mélange réactionnel par échantillon pour un volume final de 25 µL

- 2X KAPA HiFi HotStart Ready Mix : 12.5 µL
- Amorce Forward [1 µM] : 5 µL
- Amorce Reverse [1 µM] : 5 µL
- ADN [5 ng/µL] : 2.5 µL

Remarque : penser à ajouter un témoin positif et négatif.

### ETAPE 3 [Temps nécessaire env. 55 minute(s)]

#### PCR

Température (°C)	Durée	Cycles	Étapes
95	3 min.	-	activation/dénaturation initiale
95	30 sec.	30	dénaturation
58	30 sec.		hybridation
72	30 sec.	-	élongation
72	5 min.	-	élongation finale
10	∞	-	conservation

## ANNEXE : COUPLES D'AMORCES

### 16S (v3-v4):

CYA359F : 5'-GGGGAATYTTCCGCAATGGG-3'

CYA781R : 5'-GACTACWGGGTATCTAATCCWTT-3'

### 23S (UPA):

108F : 5'-ACAGWAAGACCCTATGAAGCTT-3'

108R : 5'-CCTGTTATCCCTAGAGTAACTT-3'



## EQUIPEMENT

### Laboratoire :

Hotte (avec UV)

Micropipettes (lot 10 à 100 µL)

Thermocycleur

### Consommables :

Gants Barrettes tubes PCR (200 µL)

Bouchons barrettes tubes PCR

Tips 10 µL

Tips 100 µL

### Réactifs :

H<sub>2</sub>O MilliQ qualité biomol.

Amorces (avec queues moléculaires PGTB)

2X KAPA HiFi HotStart Ready Mix

## PRÉCAUTIONS



**2X KAPA HiFi HotStart Ready Mix** contient du *tetramethylammonium chloride*. Cette substance est classée mortelle par ingestion, toxique par contact cutané, engendrant un risque avéré d'effets graves pour les organes et est toxique pour les organismes aquatiques. Le Mix KAPA doit donc être manipulé avec précaution

Respecter les consignes de bases en biologie moléculaire, porter une blouse et des gants. Jeter les déchets chimiques dans un contenant prévu à cet effet par le laboratoire.

## RÉFÉRENCES ET CONTACTS

### PGTB:

CHANCEREL Emilie (✉ emilie.chancerel@inrae.fr)

INRAE (UMR-CARTEL) :

CANINO Alexis (✉ alexis.canino@inrae.fr)

RIMET Frédéric (✉ frederic.rimet@inrae.fr)

**Annexe 8: Souches utilisées et concentrations d'ADN introduites pour constituer les communautés synthétiques.**

Numéro TCC	Nom d'espèce	16S			23S		
		«EQ»	«EX»	«GR»	«EQ»	«EX»	«GR»
TCC300	<i>Anathece clathrata</i>	20	0.1	50	20	0.1**	50
TCC362	<i>Asterionella formosa</i>	20	0.1	30	20	0.1	40
TCC57	<i>Botryococcus braunii</i>	20	50*	40	20	50**	1
TCC56	<i>Cosmarium regnellii</i>	20	1	40	20	1	1
TCC826	<i>Cryptomonas sp.</i>	20	0.1	1	20	0.1	1
TCC79	<i>Dolichospermum flosaquae</i>	20	1	30	20	1	30
TCC8	<i>Haematococcus lacustris</i>	20	1	1	20	1	30
TCC814	<i>Mougeotia sp.</i>	20	0.1	10	20	0.1	10
TCC139-1	<i>Nitzschia palea</i>	20	1	30	20	1	20
TCC9	<i>Pandorina morum</i>	20	1	1	20	1	30
TCC14	<i>Planktothrix rubescens</i>	20	1*	50	20	1**	50
TCC691	<i>Stausosira venter</i>	20	0.1	20	20	0.1	40
TCC80	<i>Synechocystis aquatilis</i>	20	1*	20	20	1	20
TCC116	<i>Tetradesmus obliquus</i>	20	1	20	20	1	20
TCC365	<i>Ulnaria acus</i>	20	0.1	10	20	0.1	10
TCC345	<i>Vischeria magna</i>	20	0.1	40	20	0.1	40
TCC165	<i>Xanthonema montanum</i>	20	50*	50	20	50**	50
TCC815	<i>Zygnema sp.</i>	20	1	10	20	0.1	10

dans le cas de l'extraction Chelex, la quantité d'ADN mesurée était faible et les volumes constituaient alors un facteur limitant pour établir les différents mélanges ci-dessus. Certaines souches ont été inversées afin d'obtenir les plus fortes concentrations d'ADN des communautés « EX » :

\*(16S) *B. braunii* et *X. montanum* ne constituent pas les souches abondantes, mais à la place se sont *P. rubescens* et *S. aquatilis*.

\*\* (23S) *B. braunii* et *X. montanum* ne constituent pas les souches abondantes, mais à la place se sont *A. clathrata* et *P. rubescens*.