



Cet article est publié sous la licence Creative Commons (CC BY 4.0). La citation comme l'utilisation de tout ou partie du contenu de cet article doit obligatoirement mentionner les auteurs, l'année de publication, le titre, le nom de la revue, le volume, le numéro de l'article et le DOI..

Analyse du phytoplancton par l'ADN environnemental : nouveaux outils pour évaluer la qualité écologique des plans d'eau

Clélia DURAN^{1,3}, Benjamin ALRIC^{1,3}, Christophe LAPLACE-TREYTURE^{2,3}, Isabelle DOMAIZON^{1,3}, Laurine VIOLLAZ¹, Frédéric RIMET^{1,3}

¹ Université Savoie-Mont Blanc, INRAE, CARTEL, 75 bis avenue de Corzent, 74203 Thonon-les-Bains, France.

² INRAE, EABX, 50 avenue de Verdun, 33612 Cestas-Gazinet, France.

³ Pôle R&D ECLA, France.

Correspondance : Benjamin Alric, benjamin.alric@inrae.fr ; Frédéric Rimet, frédéric.rimet@inrae.fr

Le projet PhytoDOM a évalué des méthodes de bioindication innovantes, fondées sur le métabarcoding ADN du phytoplancton, pour les masses d'eau lacustres en France (métropole et départements et régions d'Outre-Mer). Après cinq années de tests sur différentes stratégies d'analyses moléculaires et outils d'évaluation, ce projet propose des protocoles applicables en routine pour évaluer la qualité écologique des plans d'eau. Cet article en synthétise les principaux résultats et perspectives.

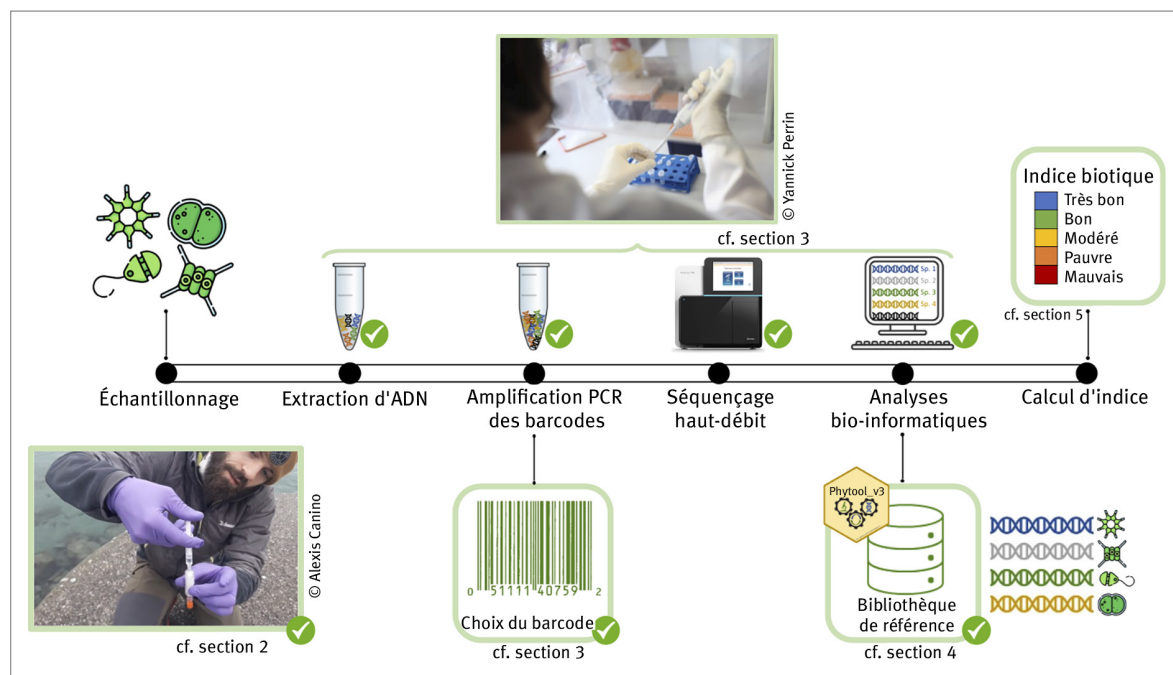
Introduction

Le phytoplancton regroupe les micro-organismes photosynthétiques vivant librement dans la colonne d'eau des plans d'eau. Il est composé d'une grande diversité d'organismes : microalgues eucaryotes et cyanobactéries estimées à plus de soixante-cinq mille espèces (Guiry et Guiry, 2025). Le phytoplancton est un compartiment biologique situé à la base de la chaîne trophique. Il joue un rôle important dans la biomasse, la biodiversité, la dynamique des réseaux trophiques et des processus biogéochimiques des plans d'eau (Bernard et Duperron, 2024). Cette communauté biologique répond rapidement aux perturbations environnementales et est directement affectée par les pressions humaines (Salmaso et Cerasino, 2012). Ces caractéristiques en font un élément biologique clef, dans le cadre de la directive cadre européenne sur l'eau (DCE), pour la biosurveillance des plans d'eau et l'évaluation de leur qualité écologique. En France, un indice biotique, l'indice phytoplancton lacustre (IPLAC), permettant d'évaluer la qualité éco-

logique des plans d'eau a été développé ces dernières années (Laplace-Treyture et Feret, 2016). Il se base sur l'identification et le comptage en microscopie optique des espèces composant le phytoplancton. Les analyses par microscopie sont toutefois relativement longues et coûteuses, nécessitant une expertise pointue en taxonomie, et ne permettent pas la détection de certaines espèces trop petites ou trop similaires morphologiquement (espèces cryptiques).

D'un autre côté, l'approche métabarcoding ADN est une méthode alternative d'identification des êtres vivants qui est basée sur le séquençage de courts fragments d'ADN environnemental (Egea, 2024). Cette approche peut se révéler moins coûteuse, plus rapide, et permet de détecter une partie de la diversité difficile à identifier en microscopie optique. Ainsi, l'utilisation du métabarcoding ADN dans le cadre de la bioindication phytoplanctonique représente une alternative intéressante à l'approche microscopique.

Figure 1 – Schéma des étapes de préparation et d'analyse des échantillons de phytoplancton par métabarcoding ADN pour utilisation en bioindication. La section 2 représente la section «L'échantillonnage du phytoplancton pour l'analyse métabarcoding ADN». La section 3 représente la section «Préparation et analyse de phytoplancton par séquençage métabarcoding». La section 4 représente la section «Identifier les séquences environnementales : la bibliothèque de barcodes ADN de référence Phytool» et la section 5 représente la section «Développement d'indices d'évaluation de la qualité des plans d'eau basés sur l'analyse métabarcoding du phytoplancton».



Le projet PhytoDOM a été déployé afin de tester des méthodes de bioindication basées sur le métabarcoding ADN du phytoplancton, applicables sur l'ensemble des masses d'eau lacustres de France (métropole et départements et régions d'outre-mer). Pour cela, différentes stratégies d'analyses moléculaires ainsi que différents outils d'évaluation basés sur ces analyses ont été testés. Nous présentons dans cet article un résumé du travail réalisé durant les cinq années du projet qui nous a permis de proposer des protocoles et outils qui, à l'avenir, pourraient être appliqués pour l'évaluation en routine par le phytoplancton de la qualité écologique des plans d'eau.

L'échantillonnage du phytoplancton pour l'analyse par métabarcoding ADN

L'échantillonnage est la première étape de toute analyse de communauté biologique (figure 1). Dans le cadre de l'analyse du phytoplancton par métabarcoding ADN, la méthode d'échantillonnage est identique à celle utilisée pour l'analyse par identification microscopique pour les suivis DCE. Cela consiste à prélever, à l'aide d'un échantillonneur d'eau intégrateur, un échantillon d'eau intégré sur la zone euphotique de la colonne d'eau, qui s'étend de la surface à 2,5 fois la profondeur donnée par le disque de Secchi (norme AFNOR XP T90-719). Par la suite, 250 ml d'eau de ce prélèvement préalablement homogénéisé sont filtrés sur le terrain à l'aide d'une seringue munie d'un filtre fermé (Sterivex®, porosité 0,45 µm). Un tampon de préservation est ajouté dans le filtre qui est ensuite stocké à - 20 °C. Un protocole et une vidéo décrivant la procédure d'échantillonnage à suivre sur le terrain sont disponibles en ligne¹ (Canino et al., 2025).

Ce protocole a été éprouvé sur le terrain lors de plusieurs campagnes de prélèvements entre 2021 et 2023 qui ont été réalisées dans le cadre du projet PhytoDOM par des bureaux d'étude aussi bien en France métropolitaine que dans les départements et régions d'outre-mer. La facilité de réalisation du protocole, et notamment sa rapidité d'exécution (environ cinq minutes) a permis son déploiement sur près de deux cents plans d'eau et rivières et de récolter environ un millier d'échantillons.

Préparation et analyse des échantillons de phytoplancton par séquençage métabarcoding

Plusieurs étapes sont à réaliser dans une analyse de métabarcoding : l'extraction de l'ADN, l'amplification (PCR²) des barcodes ADN, le séquençage et l'analyse bioinformatique (figure 1). Pour chacune d'entre elles, des choix ont été opérés après une série de tests décrits ci-dessous.

Pour l'étape d'extraction de l'ADN, de nombreux kits et protocoles existent sur le marché. Les quatre plus communs ont été testés sur des communautés synthétiques et naturelles de lacs alpins. Le protocole donnant les meilleurs résultats a été sélectionné (*Nucleospin soil Macherey-Nagel*®) et mis en ligne pour un accès et une utilisation libre.

Pour l'étape d'amplification (PCR) d'un barcode ADN, il est nécessaire dans un premier temps de choisir le barcode à amplifier. Un barcode ADN doit rassembler

1. https://www.youtube.com/watch?v=x_to-063Qno

2. Polymerase Chain Reaction.

certaines propriétés indispensables permettant de respecter les quatre règles suivantes (Falentin *et al.*, 2019) :

1. Il faut cibler un gène universel de la communauté étudiée ;

2. Il faut définir une courte région du gène, qui est variable entre les différents taxons de la communauté ;

3. Cette région de gène doit être flanquée de zones hautement conservées pour cibler le barcode d'intérêt ;

4. Le barcode doit avoir une taille (en nombre de paires de base) adaptée aux capacités de la technologie de séquençage utilisée (dans notre cas Illumina MiSeq ©).

L'ARN ribosomal (ARNr), nécessaire à la vie, est présent chez tous les organismes cellulaires (règle 1). Les gènes des ARNr 16S et 23S sont donc des gènes candidats pour l'étude du phytoplancton (Decelle *et al.*, 2015 ; Djemiel *et al.*, 2020) car ils sont présents chez toutes les cyanobactéries et dans les chloroplastes des cellules eucaryotes photosynthétiques. Afin de séquencer un fragment de ces gènes, il est nécessaire de l'amplifier et donc de définir des amorces spécifiques des organismes ciblés (ici les cyanobactéries et les microalgues eucaryotes) en respectant les règles 2, 3 et 4.

La sélection du barcode d'intérêt et du meilleur couple d'amorce pour l'analyse du phytoplancton ont fait l'objet d'une étape importante de développement dont les résultats sont accessibles dans l'article scientifique de Canino *et al.* (2023). Une étude bibliographique ainsi que la bibliothèque de référence Phytool ont permis d'identifier vingt couples d'amorces ciblant les gènes des ARNr 16S et 23S. Des tests bio-informatiques sur des données *in silico* suivis de tests *in vitro* sur des communautés synthétiques (composées à partir de la collection de cultures³ de microalgues de l'UMR CARTEL⁴) ont permis de sélectionner un couple d'amorce spécifique pour chaque gène, permettant ainsi d'amplifier spécifiquement l'ADN des cyanobactéries et des microalgues eucaryotes. Ces travaux ont également montré que les amorces ciblant le gène de l'ARNr 23S présentaient une meilleure précision taxonomique, une meilleure fidélité pour retranscrire les abondances relatives des organismes, et une meilleure spécificité pour le phytoplancton avec moins de bactéries hétérotrophes amplifiées.

Après l'étape d'amplification, les échantillons ont été séquencés avec la technologie Illumina MiSeq, qui est la technologie actuellement la plus communément utilisée pour le métabarcoding ADNe.

La dernière étape est le traitement bio-informatique. Cette étape est essentielle pour le traitement des données issues du séquençage (Falentin *et al.*, 2019). Ce traitement consiste à utiliser un pipeline d'analyse libre d'accès (DADA2, Callahan *et al.*, 2016) et utilisable avec le logiciel R (R Core Team, 2025) pour transformer les données brutes (c'est-à-dire les séquences génétiques) en données quantitatives (c'est-à-dire des tables d'abondances relatives des séquences) utilisables pour le calcul d'indices biotiques. Concrètement, ce pipeline (communément utilisé en métabarcoding et mis à jour régulière-

ment) permet de regrouper les séquences appartenant à un même taxon entre elles et connaître les abondances relatives de chaque séquence dans chaque échantillon. Ce pipeline d'analyse nécessite un paramétrage spécifique au type de données, paramètres que nous avons optimisés et fixés pour le phytoplancton. L'intégralité du processus d'analyses bio-informatiques stabilisé est en accès libre⁵.

Identifier les séquences environnementales : la bibliothèque de barcodes ADN de référence Phytool

Pour identifier taxonomiquement (ex. : nom de genre, d'espèce) les séquences obtenues en sortie du séquençage haut-débit des échantillons environnementaux (nous parlons alors d'assignation taxonomique) une bibliothèque de barcodes ADN de référence est nécessaire (figure 1). Cette bibliothèque est constituée de séquences de référence – ou barcodes de référence – provenant d'organismes préalablement identifiés. Dans le cas des microalgues, les séquences de références proviennent majoritairement de cultures maintenues dans des collections, bien que certaines proviennent également d'échantillons naturels. Une étape de curation régulière de la bibliothèque de référence est essentielle afin de garantir la qualité des séquences ainsi que la disponibilité des métadonnées associées permettant leur réutilisation. Cette étape est réalisée par des experts. Un autre point important est le niveau de complétude de la bibliothèque de référence pour le barcode ADN utilisé. En effet, pour optimiser l'assignation taxonomique de séquences environnementales, il est primordial que la diversité du groupe biologique ciblé soit bien représentée dans la bibliothèque au risque sinon qu'une part importante des séquences environnementales ne soit pas assignée et donc utilisée.

Il existe de nombreuses bibliothèques de barcodes ADN de référence : des généralistes regroupant plusieurs gènes marqueurs ou des spécifiques d'un gène marqueur. Toutefois, il manquait une bibliothèque de référence couvrant le phytoplancton, la plus exhaustive possible et portant sur plusieurs gènes marqueurs. Une bibliothèque de référence, Phytool (Canino *et al.*, 2021) a donc été construite. Cette bibliothèque compile des barcodes de référence des principaux gènes marqueurs utilisés pour caractériser le phytoplancton : les gènes des ARNr 16S, 18S et 23S. Ces barcodes de référence proviennent de la compilation de bases de référence : Silva 138.1 (Quast *et al.*, 2013), PR2 (Guillou *et al.*, 2013), PhytoRef (Decelle *et al.*, 2015), μ -green-db (Djemiel *et al.*, 2020), rassemblant des séquences environnementales et des séquences issues de souches de microalgues cultivées.

Phytool propose également un outil d'homogénéisation taxonomique pour des données de séquençage ou des tableaux d'inventaires par microscopie optique. L'application Phytool est en accès libre en ligne⁶, où l'interface graphique permet de consulter et télécharger l'intégra-

3. Thonon Culture Collection (TCC) : <https://doi.org/10.15454/UQEMVW>

4. Le Centre alpin de recherche sur les réseaux trophiques et les écosystèmes limniques est une unité mixte de recherche INRAE-Université de Savoie Mont Blanc.

5. https://raw.githubusercontent.com/benalric/2023_Nicolosi-Gelis-STOTEN/main/DADA2_pipeline_phytoplankton_16S_23S.html

6. https://github-carrel.shinyapps.io/phytool_v2/

lité de la bibliothèque et de réaliser une harmonisation taxonomique. Récemment, une mise à jour de la bibliothèque de référence, hébergée dans Phytool, a été réalisée en annotant les barcodes de référence avec des traits bio-écologiques. Ceci permet maintenant d'assigner aux séquences environnementales des traits fonctionnels (Tapolczai *et al.*, 2025). La bibliothèque de référence compile actuellement plus de 30 500 séquences appartenant à plus de 26 400 taxons. Elle sera complétée par de nouvelles séquences et espèces dans le cadre de nouveaux projets européens (ex. : RefBar).

Développement d'indices d'évaluation de la qualité des plans d'eau basés sur l'analyse métabarcoding du phytoplancton

Nos travaux de recherche ont conduit vers trois stratégies pour le développement d'indices d'évaluation de la qualité écologique des plans d'eau basés sur l'analyse du phytoplancton par métabarcoding ADN (figure 2), à savoir :

1. Tester une métrique déjà existante, habituellement développée et utilisée pour les identifications par microscopie optique ;
2. Développer une nouvelle métrique biotique en se basant sur les séquences d'ADN sans identification taxonomique préalable ;

3. Créer un indice en se basant sur la topologie des réseaux écologiques.

Utilisation d'une métrique déjà existante : est-ce pertinent ?

L'IPLAC, indice phytoplancton lacustre (Laplace-Treytore et Feret, 2016), est l'indice d'évaluation de la qualité des plans d'eau basé sur le phytoplancton et développé en France afin de répondre à la DCE. C'est un indice multi-métrique composé, sous la forme d'une moyenne pondérée, i) d'une métrique de biomasse algale (MBA) basée sur la teneur en chlorophylle a et ii) d'une métrique de composition spécifique des échantillons (MCS). Cette métrique de composition spécifique est basée sur la formule de Zelinka et Marvan (1961) qui prend en compte une liste fermée d'espèces retrouvées dans les échantillons par identification microscopique, leurs abondances ainsi que leurs profils écologiques. Les profils écologiques peuvent être assimilés à la dynamique en abondance des espèces le long d'un gradient environnemental (ex. : le niveau en phosphore). Ces profils écologiques se caractérisent par un optimum et une gamme de tolérance autour de cet optimum (figure 3).

Nos efforts se sont centrés sur la métrique MCS, puisque la nature semi-quantitative (c'est-à-dire abondance relative) des données de métabarcoding ADN rend impossible l'estimation de la biomasse algale. La com-

Figure 2 – Illustration de la construction des métriques et indice d'évaluation de la qualité des plans d'eau basé sur le métabarcoding ADN du phytoplancton.

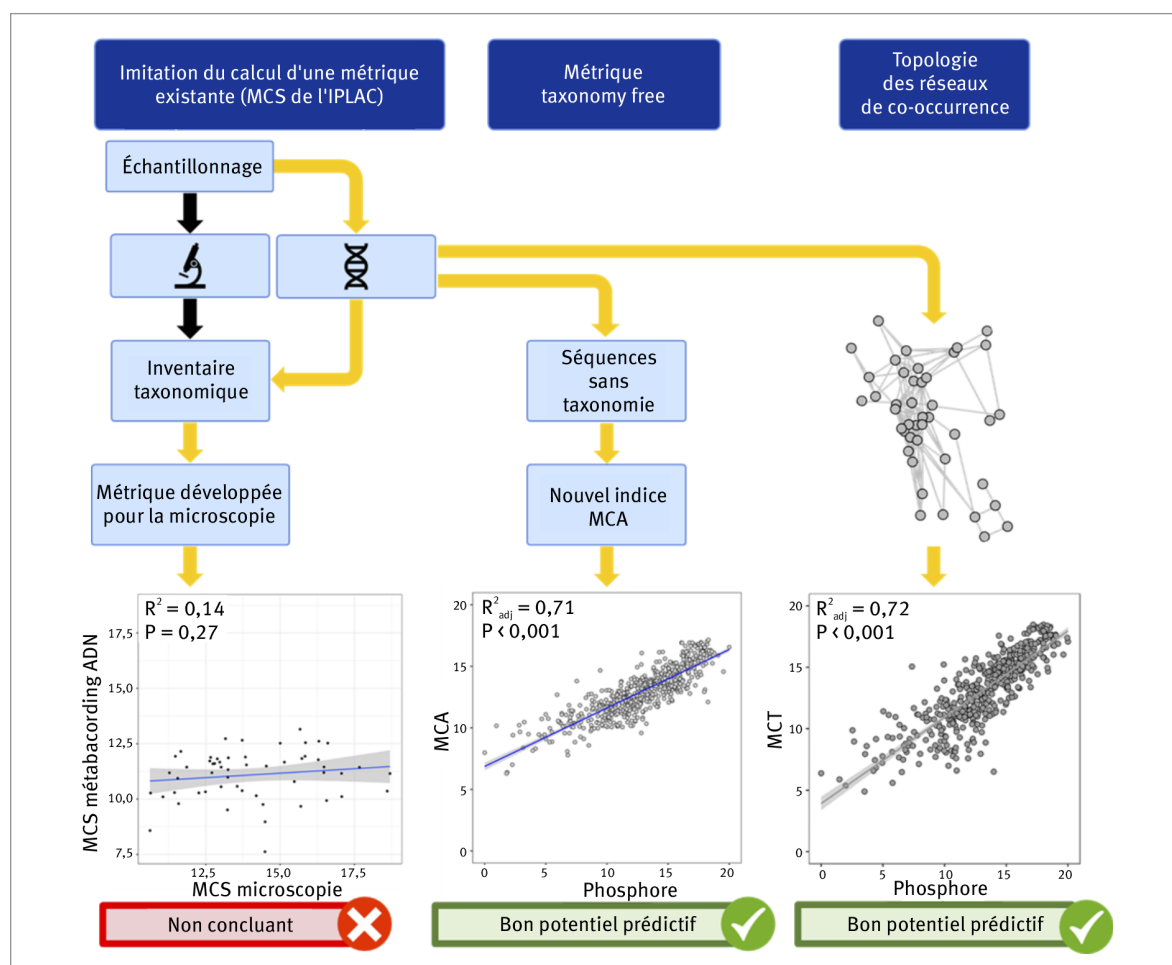
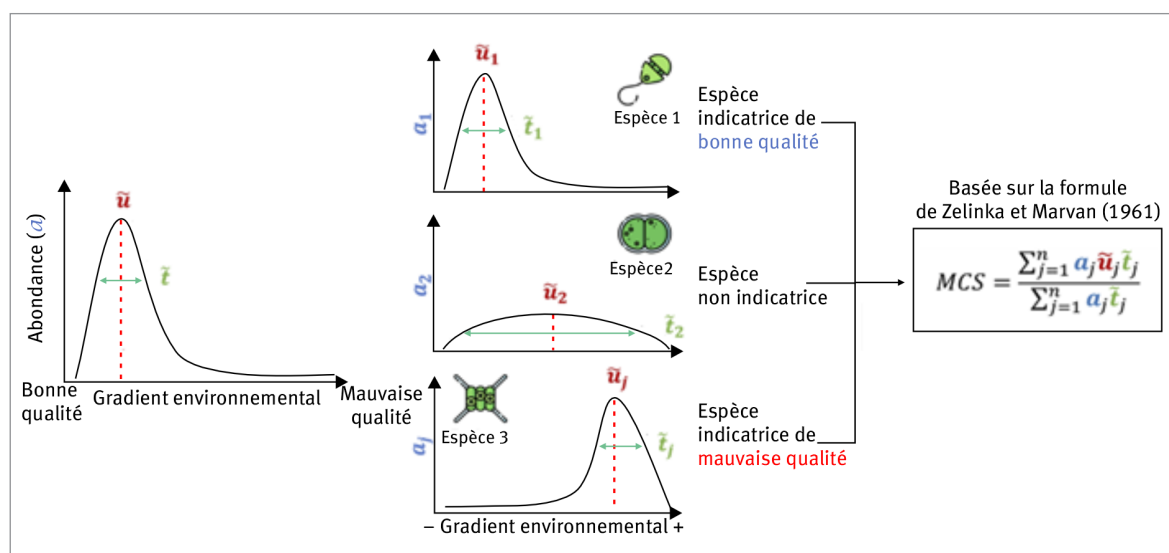


Figure 3 – Caractérisation des profils écologiques des espèces utilisées pour calculer la métrique MCS de l'indice IPLAC. Pour le profil écologique de chaque espèce le long d'un gradient environnemental, un optimum $\tilde{u}$ et une gamme de tolérance $\tilde{r}$ autour de celui-ci sont calculés. L'optimum $\tilde{u}$ correspond au centroïde de la niche écologique et la gamme de tolérance $\tilde{r}$ correspond à la valence écologique ou à la largeur de la niche écologique. L'espèce 1 de phytoplancton présente un optimum dans les environnements les moins pollués et une faible tolérance. Cette espèce est donc indicatrice d'environnements non pollués. L'espèce 2 a un optimum dans les environnements avec un niveau de pollution intermédiaire mais a une très grande tolérance. Cette espèce apparaît sur l'ensemble du gradient de pollution et n'est donc pas une espèce indicatrice des niveaux de pollution. L'espèce 3 présente un optimum dans les milieux les plus pollués et une faible tolérance, ce qui suggère que cette espèce est tolérante à la pollution et indicatrice des milieux pollués. Une fois estimés, l'optimum et la tolérance de chaque espèce sont utilisés avec l'abondance des espèces pour construire une métrique (MCS) basée sur la formule de Zelinka et Marvan.



paraison des métriques MCS obtenues avec des données de microscopie et des données de métabarcoding, pour quatre grands lacs alpins de métropole (Aiguebelette, Annecy, Bourget et Léman) échantillonnés en 2021, a montré les limites de l'approche (Nicolosi Gelis *et al.*, 2024). Pour cette comparaison, la métrique MCS métabarcoding était calquée sur la métrique MCS microscopie de la première version de l'IPLAC, en utilisant uniquement les séquences avec une assignation taxonomique à l'espèce mais basés sur les abondances relatives contrairement à la MCS microscopie. Aucune relation significative entre les métriques MCS calculées selon les deux approches n'a été observée, suggérant que l'utilisation d'indice, développé à la base pour des données de microscopie, avec des données de métabarcoding n'est pas pertinent dans le cas du phytoplancton (figure 2).

Trois explications peuvent être données à ces différences d'évaluation de la qualité obtenues par la MCS basée sur la microscopie et celle basée sur le métabarcoding :

1. Les faibles taux d'assignation taxonomique à l'espèce avec l'analyse par métabarcoding. Cela s'explique par un manque de séquences de référence pour certaines espèces dans la bibliothèque Phytool. Parmi ces espèces, certaines sont bioindicateurs dans la métrique MCS (et donc l'IPLAC) et ne sont donc pas détectées dans les échantillons. La valeur de la métrique MCS calculée à partir des données de métabarcoding est par conséquent imprécise lorsque ces espèces sont abondantes dans l'échantillon ;

2. Le faible recouvrement de diversité d'organismes entre les deux approches. La diversité mesurée est systématiquement plus élevée avec l'approche par métabarcoding

que par microscopie. L'analyse par métabarcoding permet de détecter de nombreuses espèces difficiles voire impossibles à identifier en microscopie (ex. : espèces cryptiques ou de petite taille) ;

3. Les profondeurs de séquençage actuelles permettent l'obtention de dizaines de milliers de séquences par échantillons alors que les analyses microscopiques prennent en compte seulement quatre cent objets algaux par échantillons (Laplace-Treytore et Feret, 2016). Ceci permet la détection d'espèces rares en métabarcoding, ce qui entraîne également des différences d'évaluations de la qualité avec la métrique MCS entre microscopie et métabarcoding.

Malgré ces différences, il existe une congruence significative entre les structures de communautés obtenues par identification morphologique et métabarcoding. Cela indique que les communautés évaluées par métabarcoding ou microscopie répondent à des gradients environnementaux similaires. Ces résultats nous ont conduit à aller vers d'autres approches, qui s'affranchissent de la taxonomie en se basant directement sur les séquences détectées.

Développement d'une métrique biotique taxonomy free

Une nouvelle métrique *taxonomy free*, basée sur le calcul des profils écologiques des séquences et que nous avons appelée « métrique de composition d'amplicon » (MCA), a été développée (Viollaz *et al.*, 2025). Le principe d'une métrique biotique taxonomy free est d'utiliser l'intégralité des données de séquençage en s'affranchissant de leur assignation taxonomique. Cela permet de

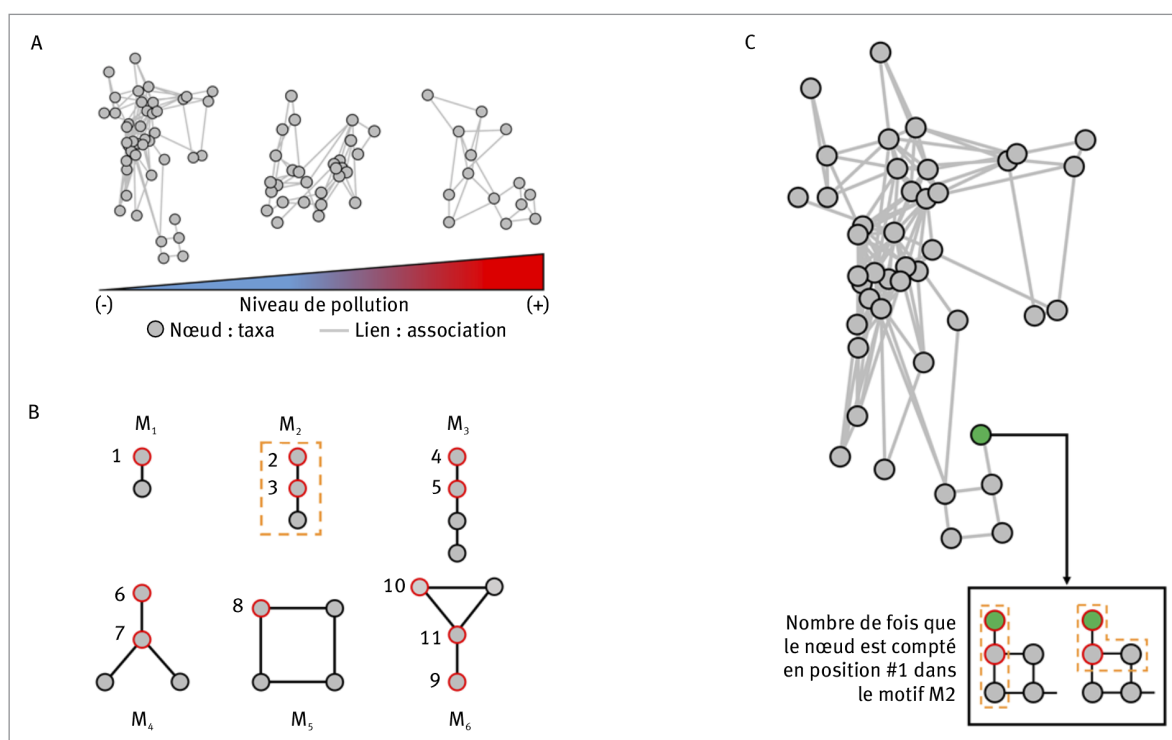
considérer dans le calcul de la métrique, des séquences présentant un profil écologique indicateur mais qui seraient non utilisées dans l'approche initiale du fait de leur absence d'assignation taxonomique à l'espèce.

La métrique MCA a été développée grâce à un jeu de données regroupant 599 échantillons prélevés entre 2022 et 2023 dans 186 plans d'eau métropolitains répartis sur l'ensemble du territoire des agences de l'eau (Adour-Garonne, Artois-Picardie, Loire-Bretagne, Rhin-Meuse, Rhône-Méditerranée-Corse, Seine-Normandie). Les échantillonnages ont été réalisés par des bureaux d'étude en charge des suivis phytoplanctoniques mandatés par les agences de l'eau pour les suivis DCE. Ces échantillonnages ont été synchronisés aux campagnes DCE de mesure des paramètres physico-chimiques, ce qui nous a permis de déterminer un gradient environnemental principal le long duquel les échantillons se répartissent. Ce gradient environnemental est représenté par la concentration en phosphore dans les plans d'eau, ce qui est un indicateur d'eutrophisation lié à la pression humaine.

Pour tester la métrique MCA, le jeu de données global a été scindé en deux sous-jeux de manière aléatoire selon un ratio 75 %-25 %. Cela signifie que 75 % du jeu

de données global a été utilisé comme jeu de données d'apprentissage et les 25 % restant comme jeu de données test. Le jeu de données d'apprentissage a permis de définir l'optimum et la tolérance de chaque séquence par rapport au gradient de phosphore en se basant sur leurs abondances relatives. Celui-ci a également permis d'apprendre au modèle la relation entre les valeurs de la métrique MCA et les valeurs le long du gradient environnemental. Le jeu de données test a permis, quant à lui, de tester la robustesse du modèle à prédire la position de nouveaux échantillons le long du gradient environnemental avec la métrique MCA. Les analyses ont montré une relation linéaire significative ($R^2 = 0,71$, $p < 0,001$) entre les valeurs le long du gradient environnemental et les valeurs de MCA (figure 2). Ceci indique que la métrique *taxonomy free* MCA reflète avec fidélité la concentration en phosphore dans les plans d'eau. Il est également intéressant de noter que parmi les séquences présentant un profil écologique, 40 % n'avaient pas d'assignation taxonomique jusqu'à l'espèce et s'avéraient donc inexploitable par la MCS. S'affranchir de la taxonomie permet ainsi d'éviter la perte d'informations et d'avoir une évaluation robuste du niveau d'eutrophisation des plans d'eau.

Figure 4 – Illustration de la méthode utilisée pour caractériser les propriétés structurales des réseaux écologiques à travers l'estimation du rôle topologiques des nœuds. (A) Simplification de la structure des réseaux écologiques avec l'intensification du niveau de pollution (phosphore). (B) Représentation des six motifs (ou structures) récurrents issus de la méthode des graphlets. Ces motifs sont des éléments constitutifs des réseaux écologiques et peuvent donc être identifiés et quantifiés dans chacun d'eux. Chaque motif est composé de nœuds reliés par des arêtes (ou liens). En fonction de sa position dans un motif, un nœud peut avoir différents rôles. Les nœuds numérotés et cerclés de rouge correspondent aux onze nœuds présentant un rôle topologique non redondant dans les réseaux. (C) Pour chaque nœud dans un réseau est estimé le nombre de fois où celui-ci est compté dans une des onze positions. Par exemple, le nœud vert est trouvé dans deux configurations représentant le motif M2. Dans ces deux configurations, ce nœud occupe la position 2. Pour le rôle topologique correspondant à la position, la valeur de 2 est attribuée au nœud vert.



Développement d'un indice basé sur la topologie des réseaux écologiques

Les réseaux écologiques sont des outils basés sur la co-occurrence d'organismes, permettant d'obtenir une représentation graphique d'une communauté biologique. Dans ce type de représentations, un nœud correspond à un taxon et une arête correspond au lien (co-occurrence) entre deux nœuds (figure 4A). La signification écologique des liens est la représentation des associations entre les taxons, correspondant à des interactions biotiques potentielles (ex. : compétition, prédation). Les propriétés structurelles (ou topologie) des réseaux écologiques fait référence à l'agencement des nœuds : combien de nœuds sont connectés les uns aux autres, avec qui et comment. Il a été montré que la topologie des réseaux écologiques varie le long de gradients environnementaux (ex. : Codello *et al.*, 2022), avec une simplification de la structure des réseaux lorsque les pressions environnementales s'intensifient (figure 4A). Il a semblé donc pertinent de tester l'utilité de la topologie des réseaux écologiques pour construire un indice biotique. La théorie des graphes fournit un large éventail de méthodes permettant de qualifier et de quantifier la topologie des réseaux. En particulier, la méthode dite des graphlets permet de caractériser les réseaux à l'aide de structures (ou motifs) récurrentes définies comme des éléments constitutifs (figure 4B). À partir de ces structures récurrentes (identifiées de M1 à M6 dans la figure 4B), onze positions traduisant le rôle des nœuds peuvent être calculées. Pour cela, nous avons identifié dans combien de structures récurrentes est retrouvé chaque nœud (ici les séquences d'ADN). Ensuite, nous avons estimé pour chaque nœud combien de fois celui-ci est retrouvé à l'une de ces onze positions (figure 4C). La caractérisation des nœuds par ces onze positions permet ainsi d'évaluer leur rôle topologique. Ces informations sont ensuite mises en regard du gradient anthropique (dans notre cas, la concentration en phosphore) afin d'identifier les nœuds qui présentent un rôle topologique indicateur d'une qualité écologique donnée, de la même manière que le calcul des profils écologiques (figure 3). Les informations obtenues des onze positions ont ensuite été utilisées pour construire des métriques que nous avons appelé « métriques de composition topologique » (MCT).

Le même jeu de données que pour la métrique MCA a été utilisé pour le développement de l'indice topologique, avec comme particularité que celui-ci a été scindé en deux sous-jeux de manière aléatoire selon un ratio 80 %-20 % (apprentissage/test).

En combinant le métabarcoding ADN (*taxonomy free*), l'analyse de réseau, la modélisation des niches écologiques et un algorithme d'apprentissage automatique supervisé, nous avons développé un indice topologique basé sur les onze métriques MCT et nous avons démontré son efficacité pour prédire la concentration en phosphore dans les plans d'eau, avec une proportion élevée (72 %) de la variance du gradient environnemental expliqué par les onze métriques MCT (figure 2).

L'intérêt d'utiliser la topologie des réseaux écologiques est de pouvoir développer des indicateurs qui permettent non seulement de mesurer les effets des paramètres environnementaux sur la composition taxonomique des

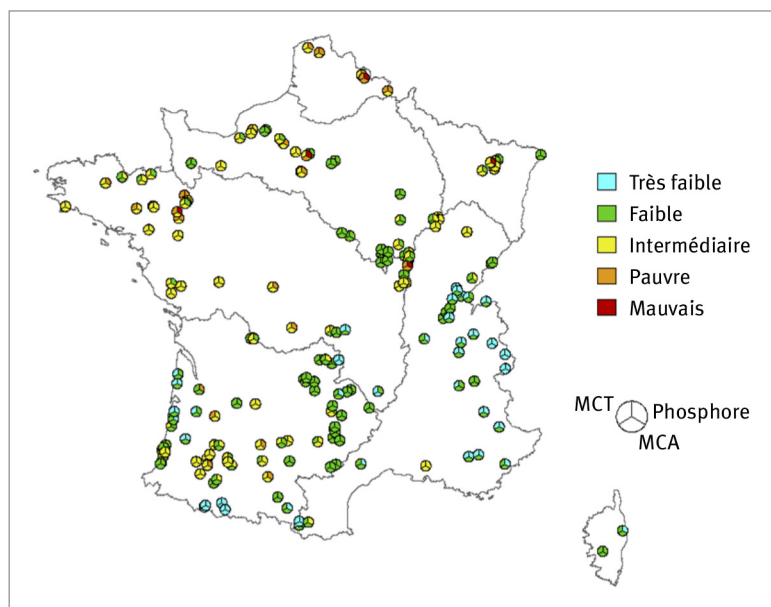
communautés biologiques, mais aussi sur les interactions biotiques qui influencent la stabilité et la complexité des écosystèmes.

Conclusions et perspectives

Les travaux de recherche menés dans le cadre du projet PhytoDOM ont permis de développer une méthode d'analyse du niveau d'eutrophisation des plans d'eau par l'analyse du phytoplancton en métabarcoding ADNe. Cette méthode d'analyse se caractérise par des protocoles d'échantillonnage, la sélection et conception d'amorce PCR, la mise en place d'un pipeline bioinformatique, et le développement de métriques et indice biotiques. L'application de la métrique MCS de la première version de l'indice IPLAC avec des données de métabarcoding ADNe a été testée et s'est révélée inadaptée dans ce contexte. À l'inverse les approches s'affranchissant de la taxonomie, la métrique biotique *taxonomy free* (MCA) et l'indice de topologie des réseaux (MCT), ont montré des résultats prometteurs. En effet, leur bonne concordance avec le gradient environnemental de phosphore est observable sur la figure 5, montrant que les outils développés permettent de mesurer le niveau d'eutrophisation des plans d'eau de manière robuste. La métrique MCA et l'indice biotique MCT ont été testés pour le gradient principal de pression anthropique auquel répond le phytoplancton (à savoir le phosphore), mais cela pourrait être calculé pour d'autres gradients anthropiques pertinents.

Au-delà de ces résultats, plusieurs perspectives qui pourraient faire suite au projet PhytoDOM peuvent être discutées, à savoir i) le positionnement des nouveaux outils de biosurveillance que nous proposons par rapport à ceux déjà en place ; ii) l'état de maturité et la transférabilité des outils développés et iii) les équipements, expertises et coûts nécessaires à la mise en place d'outils de biosurveillance métabarcoding en routine.

Figure 5 – Carte de France métropolitaine sur laquelle sont représentés les scores des plans d'eau étudiés pour le gradient environnemental de phosphore, la métrique MCA *taxonomy free* et l'indice MCT de topologie de réseaux écologiques.



Tout d'abord, la métrique MCA et l'indice MCT développés dans le cadre du projet PhytoDOM, pourraient être intégrés dans de futurs outils de biosurveillance ne visant pas nécessairement à remplacer ceux déjà en place. Il est en effet possible d'utiliser les approches décrites ici pour obtenir des indices biotiques renseignant d'autres gradients que le phosphore. De plus les outils de biosurveillance basés sur le métabarcoding et notamment l'indice MCT n'apportent pas les mêmes informations écologiques que l'IPLAC. Une complémentarité des deux approches pourrait donc être envisagée et faire l'objet de futures études. Par ailleurs, l'utilisation du métabarcoding permet de développer rapidement des outils de biosurveillance et notamment

dans les régions où les connaissances taxonomiques sur les communautés phytoplanctoniques sont rares ou difficiles à obtenir. À titre d'exemple, la Serbie ne possède actuellement aucun cadre réglementaire permettant de mettre en place un programme de biosurveillance de ses masses d'eau basé sur le phytoplancton. C'est dans ce cadre que le projet européen BIOLAWEB dans lequel nous avons collaboré avec des chercheurs serbes et norvégiens a permis de développer des outils de biosurveillance basés sur le métabarcoding ADNe du phytoplancton pour des plans d'eau saumâtres en utilisant une approche *taxonomy free*. Il pourrait être envisageable d'utiliser ces outils dans des lacs en France dépourvus d'outils de biosurveillance.

Encadré 1 – Glossaire

ADN : Acide DésoxyriboNucléique. Molécule biologique support de l'information génétique et donc de l'ensemble des caractères héréditaire d'un individu.

ADN environnemental : désigne l'ADN obtenu d'échantillons environnementaux.

Affiliation taxonomique : désigne le fait d'attribuer une dénomination taxonomique (ex. : ordre, genre ou espèce) à une séquence provenant d'échantillons environnementaux en la comparant à des séquences présentes dans des bibliothèques de références appartenant à des organismes connus.

Amorces : courtes séquences ADN simple brin (une dizaine de nucléotides) servant de site de reconnaissance et d'ancrage pour l'ADN polymérase, enzyme permettant l'amplification spécifique d'un fragment ADN d'intérêt lors de la PCR. Les amorces vont par paire, et sont des séquences complémentaires à la région ADN d'intérêt et la délimite physiquement en l'encadrant.

Amplicons/produits de PCR : fragments d'ADN obtenus par amplification à la suite d'une PCR.

ARNr : Acide RiboNucléique ribosomal. Molécule au centre de l'activité des ribosomes responsables de la production des protéines. On distingue principalement la petite (16S chez les procaryotes, 18S chez les eucaryotes) et la grande (23S chez les procaryotes, 28S chez les eucaryotes) sous-unité de l'ARNr dont les gènes sont particulièrement étudiés en métabarcoding.

Barcode : courte séquence d'ADN présente sur un gène universel d'une communauté cible permettant l'identification d'une grande diversité d'organismes.

Analyses (bio)moléculaire : se réfère à l'étude des molécules cellulaires, principalement l'ADN et l'ARN.

Chloroplastes : structures spécialisées des cellules eucaryotes photosynthétiques où se déroule la photosynthèse.

Disque de Secchi : dispositif permettant d'évaluer la profondeur de pénétration verticale de lumière dans la colonne d'eau. Cette méthode permet ainsi de mesurer la transparence/turbidité d'une colonne d'eau et ainsi d'évaluer la zone euphotique d'un milieu aquatique.

Espèce bioindicatrice : espèce présentant un profil écologique pouvant indiquer la présence et l'intensité d'une perturbation environnementale ou son absence.

Identification microscopique : identification d'un organisme comme appartenant à une espèce déjà décrite dans la littérature en se basant sur ses caractéristiques morphologiques observées par microscopie optique.

Indice de bioindication/Indice biotique : indice permettant l'évaluation de la qualité écologique d'environnements (présence et intensité d'une perturbation ou absence de perturbation) en se basant sur les communautés peuplant ces environnements (ici le phytoplancton).

Métabarcoding : méthode d'analyse d'ADN par séquençage haut débit d'un court fragment de gène « barcode » permettant d'étudier la diversité d'une communauté cible dans un échantillon environnemental.

Métrique : grandeur calculée décrivant une caractéristique biologique, écologique ou fonctionnelle d'une communauté variant en fonction de la qualité écologique d'environnements et pouvant être utilisée seule ou combinée à d'autres métriques dans le cadre d'indices biotiques.

PCR : *Polymérase Chain Reaction*/Réaction de polymérase en chaîne. Technique permettant l'amplification d'un fragment de matériel génétique cible.

Pipeline bioinformatique : succession d'étapes et de commandes pour le traitement de données (ici de séquençage haut débit de l'ADN) à travers divers logiciels et algorithmes.

Optimum écologique : valeur d'un facteur environnemental (ex. : concentration en phosphore) pour laquelle la réponse d'un taxon est optimale. Cet optimum correspond à la niche écologique du taxon.

Profil écologique : profil d'une espèce prenant en compte sa préférence et sa tolérance écologique. Un profil écologique intéressant en bioindication montre une préférence écologique claire et une faible tolérance écologique.

Séquençage haut-débit : le séquençage de l'ADN permet la détermination de l'ordre d'enchaînement des nucléotides pour un fragment d'ADN cible. Les techniques de séquençage haut-débit permettent l'obtention rapide de séquences de millions de fragments ADN en parallèle.

Tampon de préservation : solution permettant le maintien et l'intégrité des cellules dans le temps.

Taxonomie/Taxon : désigne la classification des êtres vivants en différents rangs (taxon) d'après leurs caractéristiques communes. Les principaux taxa sont le domaine, le phylum, la classe, l'ordre, la famille, le genre et l'espèce (du plus large au plus précis) mais il peut exister des rangs intermédiaires.

Tolérance écologique : capacité d'un taxon à tolérer des variations de facteurs environnementaux (ex. : concentration en phosphore). Cette tolérance correspond à la largeur de la niche écologique.

Zone euphotique : qualifie la couche superficielle des masses d'eau dans laquelle la photosynthèse est possible grâce à l'intensité de la lumière solaire.

Pour ce qui est de la transférabilité, selon l'échelle de maturité technologique (TRL ; *Technology Readiness Level*), notre méthode d'analyse, allant de l'échantillonnage jusqu'aux analyses bio-informatiques, peut être classée au niveau 7-8 de maturité. Cela signifie que la démonstration de la méthode en environnement opérationnel a été validée et qu'elle est prête à fonctionner. Pour ce qui est des indices de bioindication développés (MCA, MCT), le niveau de maturité est moins avancé (niveau TRL = 6). Afin d'améliorer leur niveau de maturité, la prochaine étape serait d'évaluer la fiabilité de la métrique MCA et de l'indice MCT sur l'ensemble des plans d'eau du réseau DCE (métropole et départements et régions d'outre-mer), de comparer leurs résultats avec l'indice en vigueur (et notamment la version révisée de l'IPLAC) et les évaluations avec les autres éléments biologiques applicables. Enfin, si les décideurs en ressentent le besoin, un transfert opérationnel de la méthode vers les gestionnaires et les laboratoires d'analyse pourrait être envisagé. Ceci peut être fait en a) normalisant les différentes étapes techniques, b) réalisant des formations et/ou tutorats auprès des gestionnaires et laboratoires d'analyse et c) procédant à des exercices d'inter-comparaison entre laboratoires. Ce type de transfert est en partie déjà en cours pour l'échantillonnage, puisque les laboratoires d'analyses mandatés pour l'application de la DCE ont réalisé de nombreux échantillonnages en vue d'analyses métabarcoding phytoplancton.

Concernant l'utilisation de la méthode que nous avons développée, celle-ci nécessite l'acquisition de données de métabarcoding ADN et donc l'accès à des laboratoires de biologie moléculaires et des séquenceurs haut-débit. Plusieurs laboratoires départementaux et vétérinaires

(ex. : Laboratoire des Pyrénées et des Landes, Terana), ainsi que des laboratoires d'analyses privés (ex. : Eurofins, Pasteur), dont certains sont déjà prestataires des agences et offices de l'eau pour les prélèvements DCE, sont équipés de ce type de matériel, ce qui leur permet de réaliser les analyses moléculaires et bio-informatiques. Des organismes publics peuvent également faire de la prestation de service pour ce genre d'analyses (ex. : la plateforme Montpellier GenomiX - CNRS, INSERM ; la plateforme GeT-PlaGe - INRAE). L'acquisition de ce type d'outil sous-tend l'existence d'une expertise, en interne, en biologie moléculaire et bio-informatique. Ces expertises sont de plus en plus fréquentes, apportant à la fois une certaine sécurité dans la mise en place d'outils de biosurveillance basés sur le métabarcoding et une prise en charge rapide des analyses.

Enfin, la question des coûts, économiques, environnementaux et sociaux, de ce type d'analyse peuvent être évoquée. Il n'existe pas à l'heure actuelle d'études robustes estimant précisément ces trois types de coûts pour l'analyse métabarcoding ADN du phytoplancton. Les analyses à réaliser pour l'approche métabarcoding sont moins chronophages que celles de l'approche en microscopie optique, ce qui entraîne des coûts salariaux et donc économiques moins importants. La quantité de consommable nécessaire pour les analyses métabarcoding en revanche entraîne des coûts environnementaux plus élevés comme cela a été montré pour les analyses de diatomées (voir Monnier dans Rimet *et al.*, 2025). Enfin, les coûts sociaux sont plus difficiles à estimer, puisqu'ils dépendront notamment des prestataires et de leur engagement comme par exemple dans la démarche RSE (responsabilité sociétale des entreprises). ■

Encadré 2 – Ressources mises à disposition pour l'application de l'analyse métabarcoding ADN du phytoplancton en bioindication.

<https://carrtel-collection.hub.inrae.fr/barcoding-databases/phytool>

Protocoles :

Canino, A., Alric, B., Bouchez, A., Domaizon, I., Duran, C., Laplace-Treytore, C., & Rimet, F. (2025). Protocoles d'échantillonnage, extraction d'ADN et amplification de barcode du phytoplancton en vue d'un séquençage métabarcoding—PhytoDOM.

<https://zenodo.org/records/16942527>

Bibliothèque Phytool :

V2 : https://github-carrtel.shinyapps.io/phytool_v2/

V3 : <https://entrepot.recherche.data.gouv.fr/dataset.xhtml?persistentId=doi:10.57745/TIRNFD>

Pipelines bioinfo :

https://github.com/Github-Carrtel/2023_Nicolosi-Gelis_et_al?tab=readme-ov-file

Encadré 3 – En savoir plus

Hering, D., Borja, A., Jones, J. I., Pont, D., Boets, P., Bouchez, A., Bruce, K., Drakare, S., Hänfling, B., Kahlert, M., Leese, F., Meissner, K., Mergen, P., Reyjol, Y., Segurado, P., Vogler, A., & Kelly, M. (2018). Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. *Water Research*, 138, 192-205. <https://doi.org/10.1016/j.watres.2018.03.003>

Monnier, O., Basilico, L., Reyjol, Y., & Ximénès, M. C. (2016). La bioindication en outre-mer : Situation et perspectives dans le contexte de la directive cadre sur l'eau. ONEMA. https://oai-gem.ofb.fr/exl-php/document-affiche/ofb_recherche_oai/OUVRE_DOC/60048?fic=PUBLI/R10/80.pdf

Pawlowski, J., Apothéloz-Perret-Gentil, L., & Altermatt, F. (2020). Environmental DNA : What's behind the term? Clarifying the terminology and recommendations for its future use in biomonitoring. *Molecular Ecology*, 29(22), 4258-4264. <https://doi.org/10.1111/mec.15643>

Poulet, N., & Basilico, L. (2019). L'ADN environnemental pour l'étude de la biodiversité : État de l'art et perspectives pour la gestion. Agence Française pour la biodiversité. <https://ofb.gouv.fr/doc/adn-environnemental-pour-etude-de-la-biodiversite-etat-de-art-et-perspectives-pour-la-gestion>

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, 183, 4-18. <https://doi.org/10.1016/j.biocon.2014.11.019>

RÉFÉRENCES

- Bernard, C., & Duperron, S. (2024). *Le phytoplancton, un monde méconnu qui fait vivre la planète*. Site Institutionnel du Museum National d'Histoire Naturelle. <https://www.mnhn.fr/fr/le-phytoplancton-un-monde-meconnu-qui-fait-vivre-la-planete>
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2 : High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581-583. <https://doi.org/10.1038/nmeth.3869>
- Canino, A., Alric, B., Bouchez, A., Domaizon, I., Duran, C., LAPLACE-TREYTURE, C., & Rimet, F. (2025). Protocoles d'échantillonnage, extraction d'ADN et amplification de barcode du phytoplancton en vue d'un séquençage métabarcoding—PhytoDOM. <https://zenodo.org/records/16942527>
- Canino, A., Bouchez, A., Laplace-Treyture, C., Domaizon, I., & Rimet, F. (2021). Phytool, a ShinyApp to homogenise taxonomy of freshwater microalgae from DNA barcodes and microscopic observations. *Metabarcoding and Metagenomics*, 5, e74096. <https://doi.org/10.3897/mbmg.5.74096>
- Canino, A., Lemonnier, C., Alric, B., Bouchez, A., Domaizon, I., Laplace-Treyture, C., & Rimet, F. (2023). Which barcode to decipher freshwater microalgal assemblages? Tests on mock communities. *International Journal of Limnology*, 59, 8. <https://doi.org/10.1051/limn/2023008>
- Codello, A., Hose, G. C., & Chariton, A. (2022). Microbial co-occurrence networks as a biomonitoring tool for aquatic environments : A review. *Marine And Freshwater Research*, 74(5), 409-422. <https://doi.org/10.1071/MF22045>
- Decelle, J., Romac, S., Stern, R. F., Bendif, E. M., Zingone, A., Audic, S., Guiry, M. D., Guillou, L., Tessier, D., Gall, F. L., Gourvil, P., Santos, A. L. D., Probert, I., Vaulot, D., Vargas, C. de, & Christen, R. (2015). PhytoREF: A reference database of the plastidial 16S rRNA gene of photosynthetic eukaryotes with curated taxonomy. *Molecular Ecology Resources*, 15(6), 1435-1445. <https://doi.org/10.1111/1755-0998.12401>
- Djemiel, C., Plassard, D., Terrat, S., Crouzet, O., Sauze, J., Mondy, S., Nowak, V., Wingate, L., Ogée, J., & Maron, P.-A. (2020). µgreen-db: A reference database for the 23S rRNA gene of eukaryotic plastids and cyanobacteria. *Scientific Reports*, 10(1), 5915. <https://doi.org/10.1038/s41598-020-62555-1>
- Egea, E. (2024). *Métabarcoding—DiagnoPict*. Site DiagnoPict INRAE OFB. <https://diagnopict.hub.inrae.fr/techniques/metabarcoding>
- Falentin, H., Auer, L., Mariadassou, M., Pascal, G., Rué, O., Dugat-Bony, E., Delbès, C., Nicolas, A., Rifa, E., Mondy, S., Le Boulch, M., Cauquil, L., Hernandez Raquet, G., Terrat, S., & Abraham, A.-L. (2019). Guide pratique à destination des biologistes, bioinformaticiens et statisticiens qui souhaitent s'initier aux analyses métabarcoding. *Cahier des Techniques de l'INRA*, 97, 46-69. <https://hal.science/hal-02311421/>
- Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H. C. F., Lara, E., Le Bescot, N., Logares, R., ... Christen, R. (2013). The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, 41(D1), 597-604. <https://doi.org/10.1093/nar/gks1160>
- Guiry, M.D. & Guiry, G.M. (2025). *AlgaeBase*. World-wide electronic publication, University of Galway. <https://www.algaebase.org>
- Laplace-Treyture, C., & Feret, T. (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. <https://doi.org/10.1016/j.ecolind.2016.05.025>
- Nicolosi Gelis, M. M., Canino, A., Bouchez, A., Domaizon, I., Laplace-Treyture, C., Rimet, F., & Alric, B. (2024). Assessing the relevance of DNA metabarcoding compared to morphological identification for lake phytoplankton monitoring. *Science of The Total Environment*, 914, 169774. <https://doi.org/10.1016/j.scitotenv.2023.169774>
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glockner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1), D590-D596. <https://doi.org/10.1093/nar/gks1219>
- R Core Team. (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rimet, F., Lemonnier, C., Alric, B., Beja, P., Bittner, L., Bylemans, J., Leese, F., Logares, R., Meissner, K., Not, F., Orsini, L., Paix, B., Rodríguez-Ezpeleta, N., Siano, R., Thalinger, B., Tromas, N., Vasselon, V., & Domaizon, I. (2025). Omics to Study and Manage Aquatic Environments : A Snapshot From the AquaEcOmics Meeting (Evian-les-Bains, 2025). *Molecular Ecology*, 34(17), e70041. <https://doi.org/10.1111/mec.70041>
- Salmaso, N., & Cerasino, L. (2012). Long-term trends and fine year-to-year tuning of phytoplankton in large lakes are ruled by eutrophication and atmospheric modes of variability. *Hydrobiologia*, 698(1), 17-28. <https://doi.org/10.1007/s10750-012-1068-2>
- Tapolczai, K., Rimet, F., Ćirić, M., Ballot, A., Laplace-Treyture, C., & Alric, B. (2025). A novel framework for phytoplankton biomonitoring: Trait assignment of 23S rRNA sequences. *Ecological Indicators*, 173, 113361. <https://doi.org/10.1016/j.ecolind.2025.113361>
- Viollaz, L., Rimet, F., Alric, B. (2025). Environmental DNA approach for lake phytoplankton monitoring. *ACS ES&T Water*, 5, 6603-6611. <https://doi.org/10.1021/acsestwater.5c00669>
- Zelinka, M., & Marvan, P. (1961). Zur Präzisierung der biologischen Klassifikation der Reinheit fließender Gewässer. *Archiv für Hydrobiologie*, 57, 389-407.