



PROJET META-IBD

INDICE BIOLOGIQUE DIATOMÉES :

EVALUATION DE LA MISE EN ROUTINE DE LA METHODE DE
METABARCODING, A L'ECHELLE LOCALE (SUD-OUEST DE LA
FRANCE)

RAPPORT FINAL



SOMMAIRE

1	Contexte et objectifs du projet.....	4
1.1	Contexte	4
1.2	Objectifs.....	5
2	Structuration du projet.....	6
3	Partenaires et implication dans le projet	6
4	Résultats du projet	9
4.1	Mise en routine de la méthode de metabarcoding aux LPL.....	9
4.1.1	Transfert	9
4.1.2	Adaptations aux LPL.....	10
4.1.3	Application de la méthode transférée, par LPL	23
4.1.4	Optimisation de la méthode par LPL	23
4.1.5	Validation de la méthode	24
4.1.6	Conclusion sur le transfert de la méthode.....	24
4.1.7	Axes d'optimisation pour la méthode.....	25
4.2	Evaluation de l'applicabilité de la base de référence Diat Barcode à l'échelle locale (Volet 2) 25	
4.2.1	Comparaison des listes taxonomiques : locale vs Diat.Barcode.....	25
4.2.2	Liste des causes de divergence entre les deux méthodes.....	27
4.2.3	Conclusion sur l'applicabilité de la méthode en local	28
4.2.4	Axes d'optimisation de la méthode pour l'applicabilité en local	28
4.3	Etude comparative méthode de référence vs metabarcoding.....	29
4.3.1	Le protocole d'étude	29
4.3.2	Pertinence de la note IBD.....	30
4.3.3	Corrélation des notes IBD MIC vs ADN.....	32
4.3.4	Sur les classes de qualité	33
4.3.5	Sur la composition des peuplements.....	34
5	Conclusion et perspectives	37
6	Références	38

1 Contexte et objectifs du projet

1.1 Contexte

Afin d'évaluer la qualité écologique d'un cours d'eau, les Diatomées (algues brunes unicellulaires) sont utilisées pour calculer un indice, pouvant ensuite donner une information sur la classe de qualité, 5 classes existant au total. Cet Indice Biologique Diatomées (IBD), s'appliquant exclusivement aux rivières en utilisant les diatomées benthiques, est normalisé dans le document NF T 90-354, détaillant le mode d'échantillonnage, le prétraitement, le traitement et l'analyse de l'échantillon et enfin le calcul de l'indice. Succinctement, l'échantillonnage se fait par grattage de la face supérieure de substrats durs tels que des pierres du cours d'eau, puis un traitement à l'eau oxygénée est réalisé sur l'échantillon afin de rendre apparent le frustule de diatomées, base de l'identification des espèces. Enfin, une partie de l'échantillon traité est fixée sur lame en verre. Une observation à l'aide d'un microscope optique est ensuite entreprise afin d'identifier chaque espèce de Diatomée et d'en évaluer l'abondance relative. L'indice est ensuite calculé et l'interprétation faite quant à la qualité du cours d'eau en question après transformation en écart par rapport à une situation de référence propre à chaque hydro-écorégion (EQR). L'IBD est l'indice diatomique français officiel, reconnu pour l'évaluation des milieux aquatiques dans le cadre de la Directive Cadre européenne et la loi française sur l'eau. Les notes de références et limites de classes d'état écologique sont définies pour cet indice dans chaque hydro-écorégion. Le calcul de l'IBD implique à ce jour la prise en compte de 2530 dénominations de taxons incluant 1456 synonymes, 35 taxons appariés et 212 formes tératogènes (mises en relation avec la présence de pollutions toxiques) (NF T 90-354).

Les indices diatomées utilisés actuellement sont robustes mais présentent quelques inconvénients. En effet, la qualité des données de bioindication dépend notamment de l'analyse de la composition de l'échantillon, dépendante de la résolution taxonomique utilisée et du taux d'erreur d'identification. Les identifications sont réalisées exclusivement par des spécialistes en microscopie optique sur les caractéristiques morphologiques du frustule, ce qui nécessite une forte expertise et une bibliographie conséquente. Avec l'application de la DCE, le nombre d'échantillons à analyser annuellement pour la surveillance des milieux a fortement augmenté alors que le nombre de diatomistes reste, quant à lui, limité, avec même une tendance à la diminution (Kerrmarec, 2012). De plus, le calcul de l'indice peut s'avérer difficile à cause des incertitudes taxonomiques, discipline en continuelle évolution. Il est relativement coûteux car assez chronophage et nécessite surtout une formation des experts relativement longue.

Afin de palier à ces contraintes et de faciliter le suivi de l'état écologique des rivières, le développement d'outils complémentaires voire alternatifs a été entrepris. Ces outils sont basés sur les techniques moléculaires et plus particulièrement, sur le séquençage de l'ADN à travers l'approche de metabarcoding.

Le metabarcoding est une extension du barcoding, technique moléculaire utilisant des marqueurs génétiques dits codes-barres (séquences d'ADN spécifiques) permettant de réaliser l'identification

d'individus dans des échantillons naturels, par du séquençage haut débit. En effet, cette approche permet de réaliser automatiquement l'identification de l'ensemble des espèces contenu dans un échantillon environnemental. L'identification est réalisée en utilisant des programmes bioinformatiques et en se basant sur une bibliothèque de barcodes de référence des différents taxons d'intérêt. Outre l'application taxonomique, les écologues voient dans cette approche de metabarcoding un moyen de pouvoir identifier les espèces présentes dans un milieu alors même que les individus ne sont pas facilement caractérisables. Elle représente donc une véritable révolution dans l'étude de la diversité des milieux naturels et est de plus en plus utilisée dans le milieu de la recherche.

INRAE a entrepris le développement méthodologique de cet outil pour notamment l'évaluation de l'IBD depuis 2010 et de nombreuses études ont permis de l'optimiser et de montrer une bonne corrélation entre les résultats obtenus par metabarcoding par comparaison avec la microscopie (Kermarrec et al, 2013 ; Zimmermann et al, 2015 ; Visco et al, 2015 ; Groendahl et al, 2017 ; Rivera et al, 2018 ; Rimet et al, 2018 ; Vasselon et al, 2018). Cet outil s'avère extrêmement prometteur pour l'évaluation de la qualité des milieux aquatiques, elle est plus rapide, automatisable et son application est abordable par les laboratoires de routine dans le cadre des surveillances des milieux naturels, en particulier lorsque de gros volumes d'échantillons sont à traiter. Dans cette optique, une formation d'initiation à cette approche a d'ailleurs été délivrée, en octobre 2019 par INRAE, aux acteurs concernés, publics ou privés, dans un objectif de déploiement futur sur le territoire.

Ce nouvel outil se place dans un cadre de normalisation : en 2019, un nouveau working group dédié aux méthodes ADN pour l'étude des milieux aquatiques a été constitué au CEN (Comité européen de normalisation). En outre, déjà deux documents ont été acceptés par le CEN en 2018 sous l'impulsion d'INRAE, l'un pour l'échantillonnage, l'autre pour la construction des bases de référence. Un troisième document est en préparation avec un groupe d'une trentaine d'experts européens sur la traçabilité des codes-barres ADN pour tous les organismes aquatiques, également sous l'impulsion d'INRAE.

1.2 Objectifs

En vue de la future normalisation, l'objectif général du projet Meta-IBD est de réaliser un transfert de compétences entre le développeur de l'approche, INRAE, par l'intermédiaire de Scimabio Interface, spécialisée dans le transfert opérationnel science-gestion et plus particulièrement ici dans les approches liées à l'ADN appliquées aux diatomées, et LPL, laboratoire capable de mettre en œuvre cette nouvelle technologie sur le plan matériel et scientifique sur de gros volumes d'échantillons et destinataire final de ce type de développement. Ce transfert, assisté par Artemis, bureau d'études spécialiste en phycologie (études des algues), permettra d'évaluer les conditions de transfert ainsi que l'applicabilité de cette technique auprès de l'utilisateur récepteur, *in situ* à l'échelle locale, comme retour d'expérience. INRAE, dans ce projet, interviendra en tant qu'expert scientifique lors des comités de pilotage.

Les sous-objectifs sont multiples : (i) évaluer finement les conditions de transfert, (ii) évaluer l'applicabilité de la méthode par un laboratoire de routine, (iii) évaluer l'applicabilité de la méthode dans un contexte local géographiquement et (iv) acquérir des nouvelles données de séquençage pour les diatomées. L'ensemble des connaissances acquises lors de ce projet doit permettre d'établir un retour

d'expérience de chaque « corps de métier » en prenant en considération les visions de chacun ce qui conduira à l'établissement de préconisations, de points de vigilance et d'axes d'amélioration.

A long terme, ce projet contribuera à permettre de proposer cette nouvelle analyse aux acteurs chargés de la gestion des milieux naturels, pour l'évaluation des IBD dans le cadre de la surveillance de la qualité de l'eau.

2 Structuration du projet

Le projet Meta-IBD est structuré en 4 volets chacun divisé en tâches.

- Un volet de coordination (C) rassemblant le management, le suivi administratif, la propriété intellectuelle et la veille,
- 3 volets techniques divisés en tâches permettant de répondre aux objectifs :
 - Volet 1 – Mise en routine de la méthode de metabarcoding pour la détermination des indices IBD en laboratoire d'analyse, LPL. Dépôt des données de séquençage sur une plateforme libre d'accès.
 - Volet 2 – Evaluation de l'applicabilité de la base de référence Diat.barcode à l'échelle locale
 - Volet 3 – Etude comparative des résultats de classes de qualité obtenus par la méthode de référence (microscopie) et par la méthode de metabarcoding en conditions de routine, sur le bassin de l'Adour, Gaves et Nives.

3 Partenaires et implication dans le projet

Le projet Meta-IBD associe 3 partenaires scientifiques dont une structure de type EPIC à vocation industrielle, les Laboratoires des Pyrénées et des Landes (LPL) et deux structures privées, Scimabio Interface, consultant en environnement, et Artemis, bureau d'études spécialisé en phycologie. Le tableau suivant présente la description de chaque partenaire ainsi que l'implication de chacun dans le projet.

Partenaire 1	Laboratoires des Pyrénées et des Landes (LPL)
Présentation	Les LPL sont issus de la fusion de 3 laboratoires départementaux (département 64, 65 et 40). Ils ont été regroupés sous le statut d'EPIC en 1998. La structure regroupe 320 personnes au sein de 3 laboratoires (Lagor, Tarbes et Mont-de-Marsan) certifiés par le COFRAC (NF EN ISO/CEI 17025) et BPL (Bonnes Pratiques de Laboratoire).

	<p>Les LPL, au travers de leur activité d'analyse, proposent plus de 2000 paramètres analytiques en environnement et 1300 en biologie. 55000 échantillons environnementaux sont analysés par an et 555000 en biologie. L'Unité R&D des LPL est également engagée dans des projets de recherche d'envergure, en tant que coordinateur et partenaire, et interagit en collaboration avec divers acteurs privés ou publics, notamment sur des sujets de biodiversité et de qualité de l'eau.</p> <p>Le rôle des LPL dans ce projet est de coordonner l'ensemble du projet, de mettre en place l'approche de détermination des IBD par metabarcoding au laboratoire, de réaliser l'intercomparaison avec la méthode de référence sur des échantillons de terrain, réalisée par Artemis, et d'optimiser la méthode pour son application en routine.</p>
Intérêt pour le projet	L'intérêt pour ce projet réside dans la mise en place du metabarcoding au laboratoire, approche en plein déploiement dans divers domaines dont le suivi de la biodiversité, et dans la mise en application d'une méthode alternative et innovante dans le domaine de l'eau, par anticipation des normes à venir.
Partenaire 2	Artémis
Présentation	<p>ARTEMIS est un bureau d'études totalement indépendant (aucun actionnaire extérieur), spécialisé en phycologie (étude des algues) d'eau douce (rivière, plans d'eau) <i>sensu lato</i> : périphyton dont diatomées, phytoplancton dont cyanobactéries, algues filamenteuses. Il met ses compétences au service de tous les organismes publics ou privés pour réaliser des études techniques et/ou de la recherche sur ce compartiment biologique essentiel dans le fonctionnement des écosystèmes aquatiques continentaux.</p> <p>ARTEMIS a été créé en septembre 2014 par Florence PERES, possédant près de 20 ans d'expérience et de nombreuses références dans l'étude des algues et l'application de la DCE (Directive Cadre Européenne sur l'Eau), en France métropolitaine et dans les DOM-TOM. A ce jour, ARTEMIS est composé de trois phycologues très expérimentés, F. PERES, F. GARCIA et V. ROUQUET et d'une technicienne laboratoire A. PAUTE. ARTEMIS est accrédité par le COFRAC depuis le 18 mai 2018 pour les IBD (accréditation n°1-6492, portée disponible sur www.cofrac.fr).</p>
Intérêt pour le projet	Par le biais de ce projet de recherche, ARTEMIS souhaite développer des connaissances nouvelles relatives aux techniques de diagnostic de la qualité écologique des cours d'eau par l'utilisation des bio-indicateurs que sont les diatomées.

	<p>Les diatomées sont des bio-indicateurs très pertinents et retenus pour la mise en œuvre de la DCE dans les états membres. A ce jour, leur analyse est réalisée <i>via</i> la détermination microscopique des différentes espèces qui peuplent une station donnée (méthode « IBD »). La mise en œuvre de l'IBD constitue, pour ARTEMIS, plus de 60% de son chiffre d'affaires. Cette méthode se heurte à des difficultés taxinomiques importantes, en particulier par le manque de personnel qualifié. Les techniques d'analyse ont donc tendance à évoluer en faveur de la reconnaissance génétique de ces organismes. Cette métamorphose de nos métiers va conduire les bureaux d'études, dans les années à venir, à s'adapter et à se former à de nouvelles stratégies de fonctionnement. Face à ce contexte, ARTEMIS a un intérêt certain à prévoir cette transformation afin de mieux l'anticiper, ce qui passe par la participation au projet de recherche présenté ici.</p>
Partenaire 3	SCIMABIO INTERFACE
Présentation	<p>SCIMABIO INTERFACE-Interface est un bureau de recherche appliquée et de conseils pour la conservation de la biodiversité avec un focus principal sur les milieux aquatiques. L'entreprise, créée en août 2014 par Arnaud Caudron et Alexandre Richard (docteurs en écologie aquatique), est actuellement constituée d'une équipe de 10 collaborateurs composée majoritairement de scientifiques.</p> <p>L'activité de SCIMABIO INTERFACE consiste à développer ou co-développer des méthodes et technologies nouvelles intégrant les dernières avancées scientifiques puis à assurer leur transfert opérationnel auprès des acteurs de la gestion sous forme de prestations scientifiques. Pour assurer ce rôle d'interface science-gestion, en parallèle de leurs missions de prestations, ils participent à divers programmes de « recherche appliquée » en écologie aquatique avec plusieurs partenaires scientifiques (INRAE, CNRS, Universités, HEPIA).</p> <p>Depuis plusieurs années, SCIMABIO INTERFACE développe en collaboration avec l'UMR Carrel de nouveaux outils de caractérisation de la biodiversité dans les milieux aquatiques, principalement sur les microorganismes (diatomées, bactéries), basée sur l'ADNe et les technologies de séquençage à haut-débit.</p>
Intérêt pour le projet	L'intérêt pour SCIMABIO INTERFACE dans ce projet réside dans la mise en place des procédures permettant un transfert opérationnel optimisé du metabarcoding ADN diatomées vers les laboratoires qui réaliseront en routine les inventaires de diversité et évaluations d'état écologique des milieux

	<p>aquatique. Via ses activités d'interface science-gestion, l'entreprise pourrait se positionner comme structure de référence pour faciliter le transfert opérationnel de ces approches.</p> <p>De plus ce projet vient en parallèle d'autres activités réalisées par SCIMABIO INTERFACE visant à permettre le transfert opérationnel du metabarcoding ADN diatomées comme l'organisation d'un exercice d'intercalibration méthodologique international (18 laboratoires, 15 pays) et l'animation de formations sur les transferts des outils d'analyses bioinformatiques réalisés en collaboration avec l'UMR Carrel.</p>
--	---

4 Résultats du projet

4.1 Mise en routine de la méthode de metabarcoding aux LPL

4.1.1 Transfert

Scimabio avait pour mission le transfert de la méthode de metabarcoding vers les Laboratoires des Pyrénées et des Landes. Le transfert d'une méthode de recherche vers un laboratoire de routine implique plusieurs étapes afin de s'assurer que cette méthode est mise en œuvre de manière fiable et efficace. L'organisme de transfert joue donc un rôle crucial dans le processus, pour garantir le succès de la transition de la recherche vers l'opérationnel. Ces étapes clés étaient (i) de transmettre les éléments indispensables à la mise en place technique et scientifique au laboratoire incluant les protocoles, les formations techniques et notamment bioinformatiques, le conseil et le suivi tout au long de la mise en place, (ii) d'évaluer les performances de l'analyse transférée, par rapport à des critères établis et attendus, sur la base des éléments fournis par LPL, (iii) et enfin de confirmer le succès du transfert par une intercomparaison basée sur des analyses conduites en parallèle par les deux structures et validée par Scimabio, garant des conclusions quant à la réussite de la mise en routine de la méthode de metabarcoding aux LPL.

Ainsi, au cours de ce transfert de compétence, les LPL ont mis en place en interne les adaptations et procédures utiles à la mise en routine de la méthode et ont également mis à point la partie de séquençage proprement dite (préparation des librairies et séquençage MiSeq (Illumina), non transférée par Scimabio car généralement sous-traitée en plate-forme de séquençage). Les protocoles sont disponibles en libre accès (voir liens internet en fin de document). De manière plus détaillée, les protocoles transférés par Scimabio portaient sur les 2 premières étapes du processus c'est-à-dire la partie extraction d'ADN et amplification du code-barres rbc1 par PCR avec la séquence des amorces permettant cette amplification avant un séquençage « two-step » PCR » sur MiSeq. Pour la seconde partie de la préparation de la librairie (après amplification du code-barres) et la partie séquençage, seules les spécifications techniques ont été fournies. Grâce à ces informations, une mise en place de la méthode au laboratoire a été effectuée sur le plan technique. Elle a consisté à : (i) lister les matériels, équipements et compétences nécessaires, (ii) évaluer la disponibilité de ces éléments au laboratoire et

prendre les dispositions adéquates aux besoins listés, (iii) évaluer l'impact sur les risques biologiques et chimiques au laboratoire (obligatoire lors de la mise en place d'une nouvelle analyse), (iv) adapter les protocoles aux matériels, règles de sécurité et équipements du laboratoire, (v) rédiger les procédures internes dans une optique de prise en main par les techniciens.

Concrètement, Scimabio a transmis à LPL le workflow de l'analyse par metabarcoding, schématisé en Figure 1.

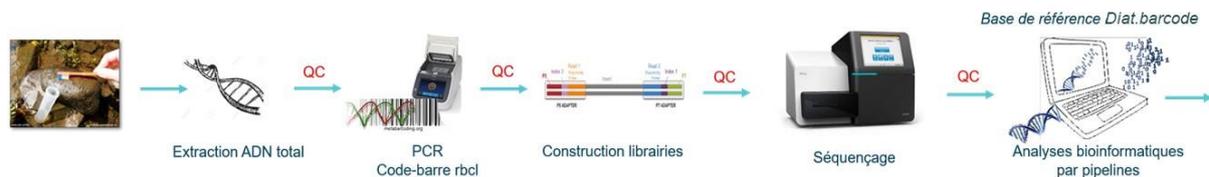


Figure 1. Workflow de la méthode de metabarcoding transmis par Scimabio.

4.1.2 Adaptations aux LPL

4.1.2.1 Partie technique

A partir de ce workflow, trois étapes ont fait l'objet de modifications par LPL (Figure 2).

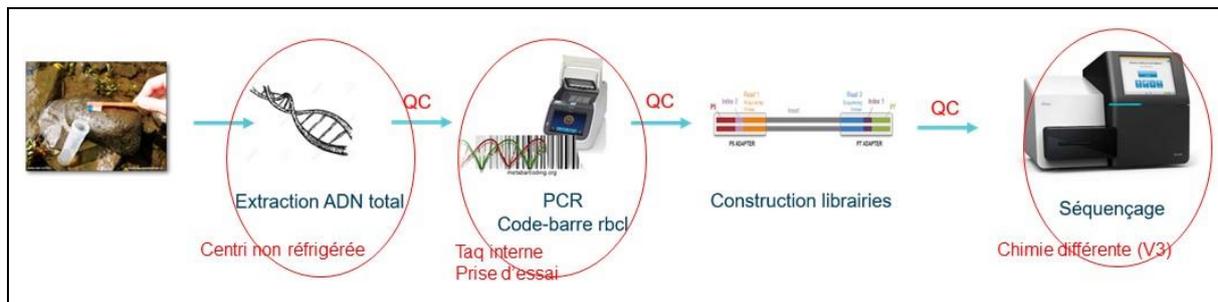


Figure 2. Workflow modifié par LPL et points de modification.

Chaque étape a été testée et validée sur 2 échantillons d'intercomparaison connus (« Lac » et « Rivière ») :

- Au niveau de l'étape d'extraction ADN des biofilms, LPL ne disposait pas de centrifugeuse réfrigérée. Après différents essais, aucun impact sur l'utilisation d'une centrifugation classique à température ambiante n'a été constaté sur les rendements d'extraction et d'amplification. Afin de limiter l'investissement et la multiplication des équipements/références et dans une démarche du laboratoire de simplification des protocoles internes, la centrifugation à température ambiante a été conservée.
- De la même manière, au niveau de l'amplification des codes-barres à partir de l'ADN extrait et purifié, la référence de Taq utilisée au laboratoire en NGS a été conservée. Le protocole d'amplification a donc été adapté en conséquence. Les essais expérimentaux menés sur les échantillons d'intercomparaison ainsi que sur la MOCK (mélange d'ADN dosés, issus de culture pure de diatomées) ont permis de valider ces modifications.

- Au niveau de l'étape de séquençage, si les LPL disposaient des spécifications techniques nécessaires en fin de run pour une analyse de qualité, les choix de méthode de séquençage ont été laissés libres par Scimabio. LPL a donc conservé la stratégie d'une librairie « two step PCR » pour indexer ces échantillons avec des index Illumina adaptés au MiSeq. Puis, en perspective d'une utilisation en routine sur de gros volume d'échantillons, le choix d'une flowcell de chimie V3 600 cycles a été privilégié pour augmenter la profondeur de lecture par échantillon, entraînant une meilleure détection des taxons rares.

4.1.2.2 Mise en place de l'intercomparaison technique entre Scimabio et LPL

4.1.2.2.1 Sélection des échantillons calibrés

Afin de valider le transfert des protocoles de laboratoire pour la mise en place du metabarcoding ADN diatomées au sein du LPL, depuis l'extraction ADN jusqu'à la production de données de séquençage, 3 types échantillons calibrés ont été sélectionnés :

- Biofilm issu de Rivière : collecté en Mars 2020 sur la rivière Edian en Haute-Savoie (46.257849°N; 6.725084°E). Permet d'avoir un biofilm avec des assemblages de diatomées
- Biofilm issu de Lac : collecté en Mars 2020 sur le lac Léman en Haute-Savoie à proximité de l'unité de recherche INRAE CARRTEL de Thonon-les-Bains (46.257849°N; 6.725084°E)
- Mélange d'ADN de culture pure : mélange en conditions contrôlées de 12 espèces de diatomées issues de cultures pures de la Thonon Culture Collection (TCC, <https://eng-carrtel-collection.hub.inrae.fr/>) permet de simuler un assemblage naturel de diatomées en contrôlant le nombre d'espèces en mélange et leur abondance relative.

Ces 3 échantillons sont issus d'un exercice d'intercalibration et d'intercomparaison méthodologique mené entre 18 laboratoires issus de 15 pays dans le cadre du réseau européen COST DNAqua-Net (EU COST Action CA15219, <https://dnaqua.net/>). Ils permettent de tester la mise en place du metabarcoding diatomées par LPL sur des échantillons avec des biofilms, des communautés de diatomées distinctes et des niveaux de contrôle différents. Les résultats préliminaires de cet exercice international ont été présentés lors du congrès DNAqua (Vasselon et al. 2021). Cependant, la description et la composition précise des échantillons seront présentées dans l'article scientifique associé à cet exercice (prévu dans « Metabarcoding and Metagenomics » pour 2024), ces informations ne seront donc pas plus détaillées dans ce rapport.

4.1.2.2.2 Mode opératoire de l'intercomparaison

LPL a reçu les échantillons de biofilm de Rivière et de Lac préservés dans de l'éthanol >70% sous forme d'aliquots de 2 mL fournis en tube Eppendorf. Ces échantillons étaient prêts à l'emploi pour l'extraction ADN. Sur ces échantillons, LPL devait réaliser : l'extraction ADN total (Nucleospin Soil Macherey-Nagel), l'amplification PCR ciblant le fragment *rbcl* diatomées pour le metabarcoding ainsi que le séquençage via la technologie Miseq (Illumina, 2 x 300 chimie V3).

LPL a reçu l'échantillon de Mock sous forme d'un extrait ADN dans un tube eppendorff 1,5 mL. L'ADN a été préparé à une concentration de [15 ng.µL⁻¹] afin d'être directement utilisable pour la mise en place de la PCR. Pour cet échantillon, LPL devait mettre en place l'amplification PCR ciblant le fragment *rbcL* diatomées pour le metabarcoding ainsi que le séquençage via la technologie Miseq (Illumina, 2 x 300 chimie V3).

S'agissant d'un exercice d'intercalibration, il a été demandé à LPL de traiter les échantillons avec un seul réplica technique pour évaluer sa déviation par rapport à la composition de référence attendue pour ces 3 échantillons. Les données de metabarcoding obtenues par LPL ont été comparées par Scimabio aux données de metabarcoding obtenues par le laboratoire de référence lors de l'exercice d'intercalibration (Scimabio, INRAE CARRETEL) européen comprenant :

- 3 réplicas techniques pour les 3 échantillons Mock, Rivière, Lac
- 2 run de séquençage Miseq indépendant ci-après nommés RT_Run1 (2 x 250 chimie V2) et RT_Run2 (2 x 250 chimie V2)

L'ensemble des 3 jeux de données a été analysé via le logiciel de bio-informatique DADA2 sous R (https://github.com/fkeck/DADA2_diatoms_pipeline) en utilisant la base de référence génétique Diat.barcode (<https://eng-carrel-collection.hub.inrae.fr/barcoding-databases>) pour l'assignation taxonomique des variants génétiques (ASV). Les inventaires taxonomiques diatomées obtenus ont été normalisés afin de permettre la comparaison des échantillons entre eux en supprimant les taxons ayant une abondance relative <0,1%.

4.1.2.2.3 Comparaison des inventaires de diversité

Les inventaires de diversité diatomées obtenus en metabarcoding par LPL pour les échantillons de la Mock, de Lac et de la Rivière en comparaison des résultats de référence sont présentés dans la Figure 3. Les résultats obtenus permettent de mettre en évidence une bonne correspondance entre les résultats des données du LPL avec les résultats de référence pour les 3 échantillons testés. On observe toutefois des variations dans l'abondance relative de certains taxons comme par exemple *Ulnaria ulna* (13,1% pour LPL vs. 5,6% en moyenne pour la référence) et *Mayamae permitis* (32,7% pour LPL vs. 42,6% en moyenne pour la référence) dans l'échantillon de la Mock. Des fluctuations similaires sont observées pour les échantillons de lac et de rivière et peuvent être expliquées par différents facteurs :

- Variabilité naturelle des échantillons lors de la préparation des aliquots à cause de la variabilité de pipetage,
- Variabilité méthodologique associée à l'utilisation de machine Miseq, avec une chimie et une profondeur de séquençage différentes entre les 3 runs de séquençage. LPL est le seul à avoir utilisé la chimie V3 avec en 2x300 avec une profondeur de séquençage >500 000 séquences par échantillon contre en 10 à 25 000 séquences par échantillon pour les run RT_Run1 et RT_Run2,

- Les échantillons ont potentiellement pu se dégrader légèrement entre l'analyse réalisée par LPL et celle des run RT_Run1 et RT_Run2, plusieurs mois de stockage s'étant écoulés entre ces analyses.

L'absence de réplicas techniques demandés ne permet pas de conclure sur l'origine de cette variabilité. Toutefois, ce n'était pas l'objectif de l'exercice. De plus, ce niveau de variabilité est attendu et n'est pas surprenant.

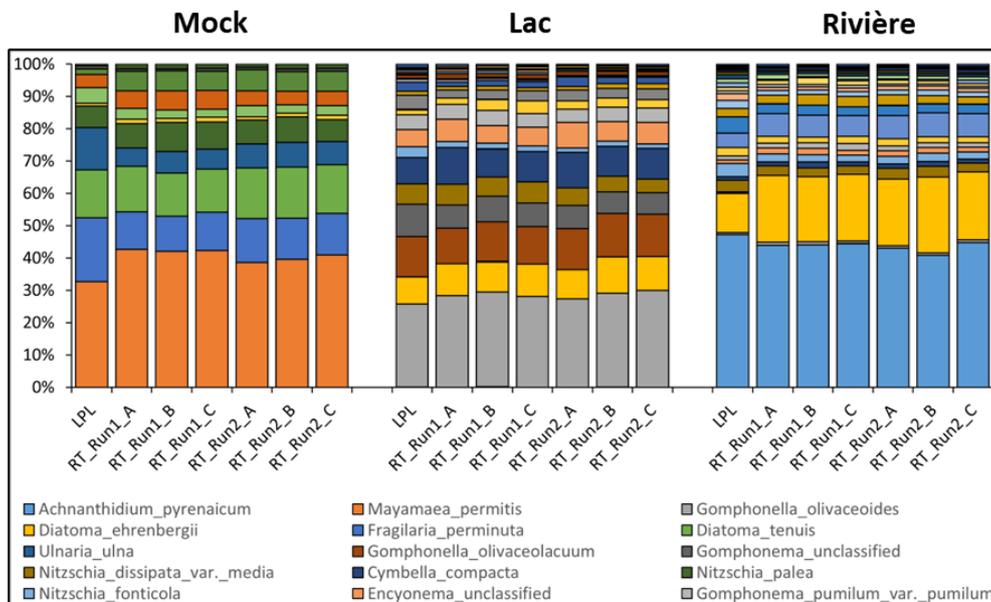


Figure 3. Inventaires taxonomique diatomées obtenus pour les 3 échantillons calibrés par LPL (1 réplica) en comparaison des 2 runs de référence de l'exercice d'intercalibration européen (3 réplicas). Les résultats sont présentés pour les taxons avec une abondance de reads >0,5% et seuls les 15 taxons les plus abondants sont présentés dans la légende.

4.1.2.2.4 Capacité de détection des taxons diatomées

Une analyse plus approfondie permet de comparer la liste des taxons diatomées détectés en metabarcoding par LPL en comparaison des 2 jeux de données de référence et ce pour les échantillons de Mock, Lac et Rivière (Figure 4).

En ce qui concerne l'échantillon Mock, 12 espèces de diatomées connues étaient attendues dans l'assemblage. LPL a obtenu le même résultat que le laboratoire de référence dans les runs RT_Run1 et RT_Run2, à savoir la détection de 11 espèces. La 12ème espèce n'est pas détectée car apportée dans le mélange en très faible quantité, donc avec peu de probabilité d'être détectée en metabarcoding. D'ailleurs, malgré l'importante profondeur de séquençage réalisée par LPL par rapport aux 2 run de référence, cette espèce n'a pas pu être mieux détectée. LPL pourra donc diminuer sa profondeur de séquençage sans altérer la qualité des données générées et la capacité de détection des taxons diatomées.

Pour les échantillons de Lac et de Rivière, correspondant à des assemblages naturels de diatomées, la majorité des espèces sont détectées par LPL et les 2 run de référence. Toutefois, certaines espèces ne sont détectées que dans 2 runs voir 1 seul run (e.g. 4 taxons détectés seulement par LPL dans le Lac, 1 taxon détecté seulement par le run RT_Run1 dans la rivière). Cette variabilité dans la capacité de détection de certains taxons concerne principalement des taxons peu abondants dans l'assemblage (<1% d'abondance relative) et peut s'expliquer par :

- La probabilité de détection des taxons peu abondants : moins un taxon est abondant, plus il sera compliqué de l'amplifier par PCR et de le détecter en metabarcoding. Il faudrait réaliser un grand nombre de répliques techniques (9 à 12) pour valider la détection des événements peu abondants. LPL ayant une plus grande profondeur de séquençage, cela peut expliquer pourquoi il détecte plus de taxons pour les échantillons de Lac et de Rivière.
- Des biais méthodologiques : les étapes d'amplification PCR et de séquençage sont connues pour introduire des erreurs dans les séquences. Bien que minimes et dans la mesure du possible éliminées lors des traitements bioinformatiques, ces dernières peuvent subsister dans les inventaires taxonomiques et aboutir à la détection de taxons erronés.

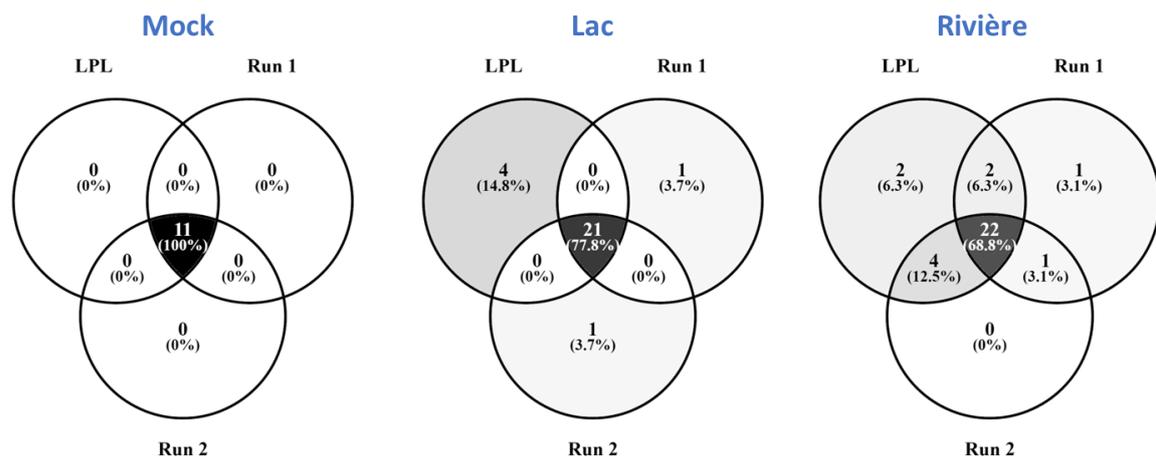


Figure 4. Diagramme de Venn présentant le nombre et le pourcentage de taxons diatomées détectés dans 1 jeu de données, 2 jeux de données et dans les 3 jeux de données (intersection centrale). (Logiciel Venny).

La variabilité observée dans chaque jeu de données est donc attendue et ne concerne que des taxons peu abondants. Les résultats de LPL sont fortement comparables à ceux obtenus par le laboratoire de référence et permet de valider la qualité des résultats obtenus.

4.1.2.2.5 Comparaison des structures de communautés

Afin de comparer les assemblages diatomées obtenus en metabarcoding par LPL aux données de référence, l'indice de dissimilarité Bray-Curtis a été calculé entre chaque échantillon. Ce dernier prenant en compte à la fois la liste des espèces et leur abondance relative, il permet d'évaluer à quel point 2 inventaires sont différents. La projection graphique de ces résultats sous forme de cluster hiérarchique

(Figure 5) ou de PCoA (Figure 6) permet de rapidement voir quels échantillons partagent les structures de communautés les plus proches.

Les résultats obtenus confirment les observations précédentes à savoir que les assemblages obtenus par LPL sont très proches structurellement de ceux obtenus dans les runs RT_Run1 et RT_Run2. Les résultats du cluster hiérarchique permettent de confirmer que certains taxons sont détectés avec des abondances relatives différentes entre LPL et les 2 runs de référence pour les échantillons de Mock, du Lac et de la Rivière (Figure 5).

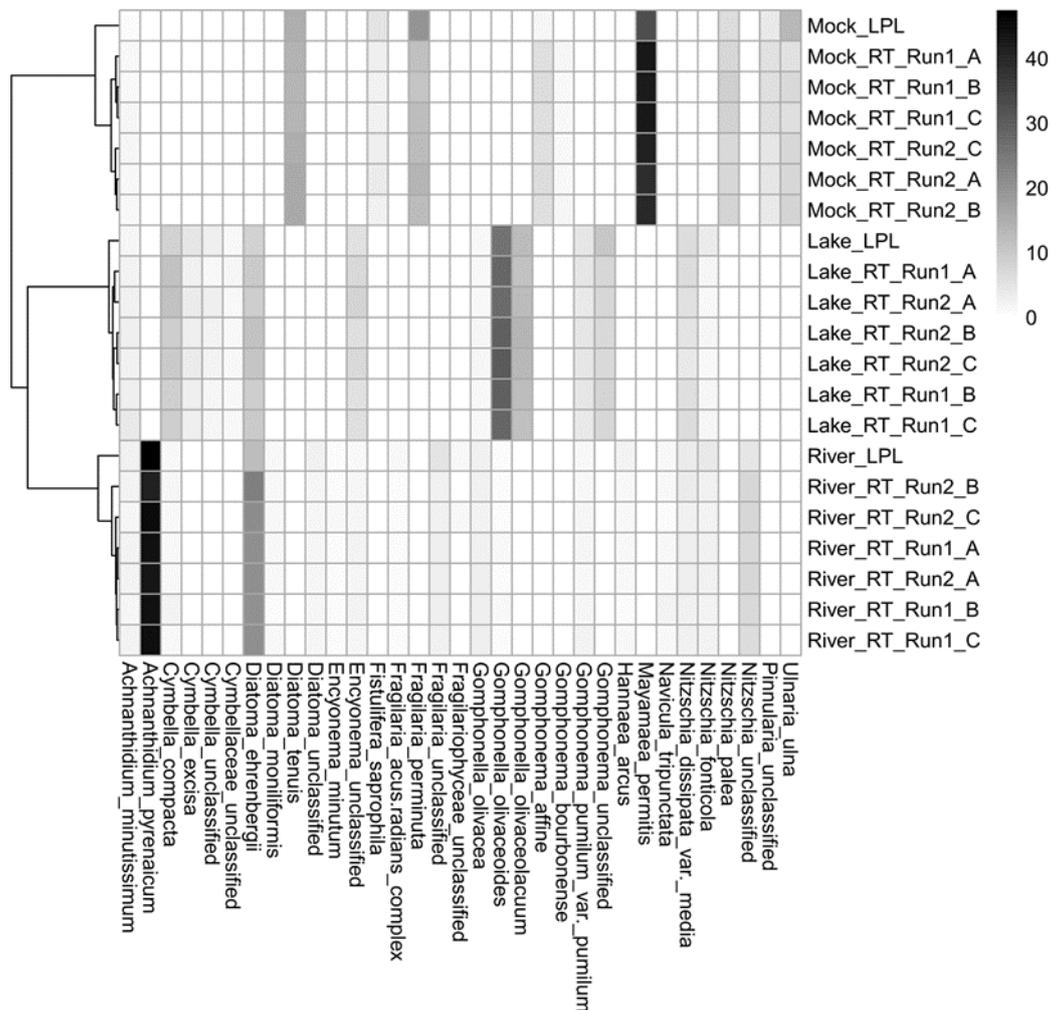


Figure 5. Cluster hiérarchique regroupant les échantillons en fonction de leur dissimilarité de variants génétiques (indice Bray-Curtis). L'intensité de noir renseigne sur l'abondance relative des taxons (% reads). Seuls les taxons >0,5% sont présentés pour plus de clarté.

L'analyse Permanova a montré que les structures de communautés obtenues par LPL, RT_Run1 et RT_Run2 présentent des différences structurelles significatives ($p=0.016$), ce qui explique les regroupements d'échantillons en fonction de leur run de séquençage (Figure 6). Une analyse en SIMPER a permis de confirmer que l'origine de la dissimilarité n'est pas due à des détections de taxons différents, mais bien au fait que les mêmes taxons peuvent être identifiés dans des proportions variables (Figure 6). Dans le cas de la comparaison entre les données du LPL et le run RT_Run1 pour l'échantillon de lac, on voit que 49% de la dissimilarité sont expliqués par 3 taxons (*Diatoma ehrenbergii*,

Achnanthydium minutissimum, *Nitzschia_unclassified*) avec au maximum un écart de 8% dans la quantification.

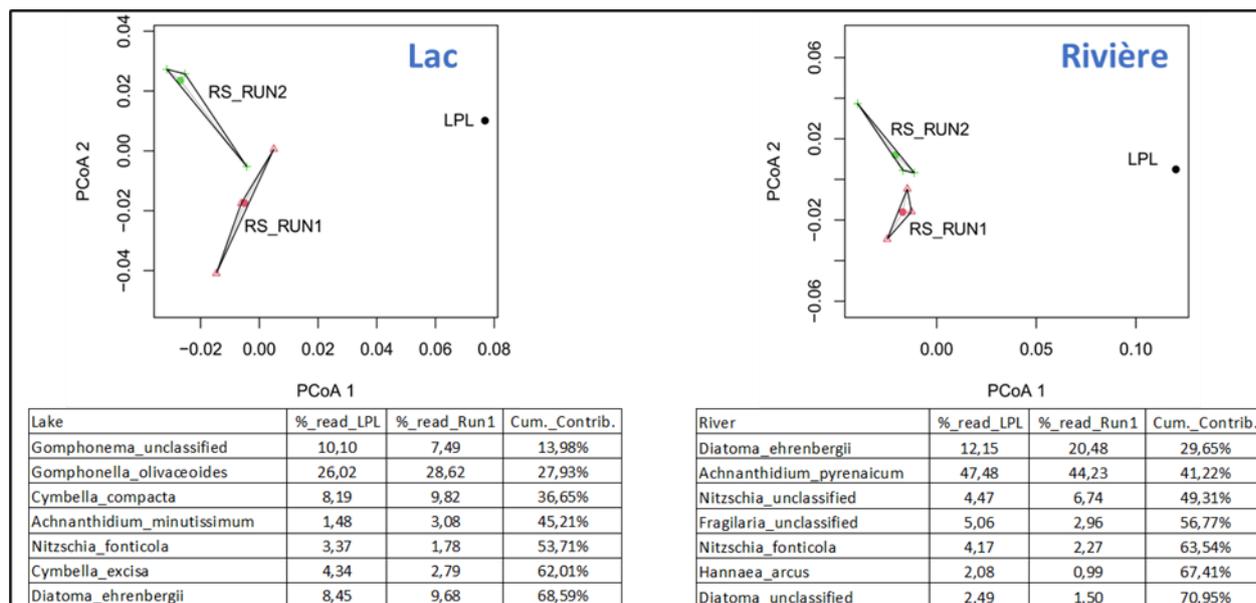


Figure 6. PCoA présentant la dissimilarité (Bray-Curtis) basée sur les variants diatomées pour les échantillons de Lac et Rivière à partir des données du LPL (1 réplica) et des 2 jeux de données de référence (RS_Run1, RS Run2). Les résultats de l'analyse SIMPER entre LPL et le run RS_Run1 sont présentés sous forme de tableau précisant le pourcentage de contribution significative des taxons à la dissimilarité entre les données du LPL et du run RS_Run1.

4.1.2.2.6 Validation du transfert de protocole au LPL

Bien que des différences significatives soient observées entre les résultats du LPL et les données de référence pour les 3 échantillons testés, notamment sur des écarts de quantification de certains taxons, cette variabilité peut être considérée comme normale et attendue. Elle est liée principalement à la variabilité naturelle des échantillons traités, les biais méthodologiques (PCR, chimie de séquençage) et à une profondeur de séquençage différente. Les taxons dominants sont bien identifiés et la variabilité de détection concerne les taxons avec une abondance relative <1%, donc ayant peu de poids dans le calcul de l'IBD. Le transfert des protocoles de biologie moléculaire au LPL peut être considéré comme validé.

Un travail complémentaire devra être réalisé avec plus d'échantillons et de répliques techniques pour étudier les limites de détection et de quantification du metabarcoding diatomées dans le contexte opérationnel du LPL. Un travail supplémentaire pour définir la profondeur de séquençage minimale pour obtenir un inventaire taxonomique diatomées représentatif et robuste pourra aussi être réalisé, la profondeur de séquençage actuelle étant surdimensionnée pour des besoins de bioindication.

4.1.2.3 Partie bioinformatique

Pour la bioinformatique, il a avant tout été nécessaire au personnel LPL de se former aux traitements des données issues de séquençage. La formation dispensée par Scimabio avait pour objectif de :

- Comprendre les différences et le processus de traitement des données de deux approches bioinformatiques, l'une basée sur l'analyse des OTU (unité taxonomique opérationnelle) par Mothur; l'autre sur les ESV (Exact Sequence Variant) par Dada2,
- Se familiariser avec le langage des deux pipelines,
- Comprendre, savoir utiliser et modifier si nécessaire les commandes des scripts.

Ainsi, Scimabio a procédé au transfert des deux approches aux LPL. Les différentes étapes de chaque approche sont décrites dans la Figure 7

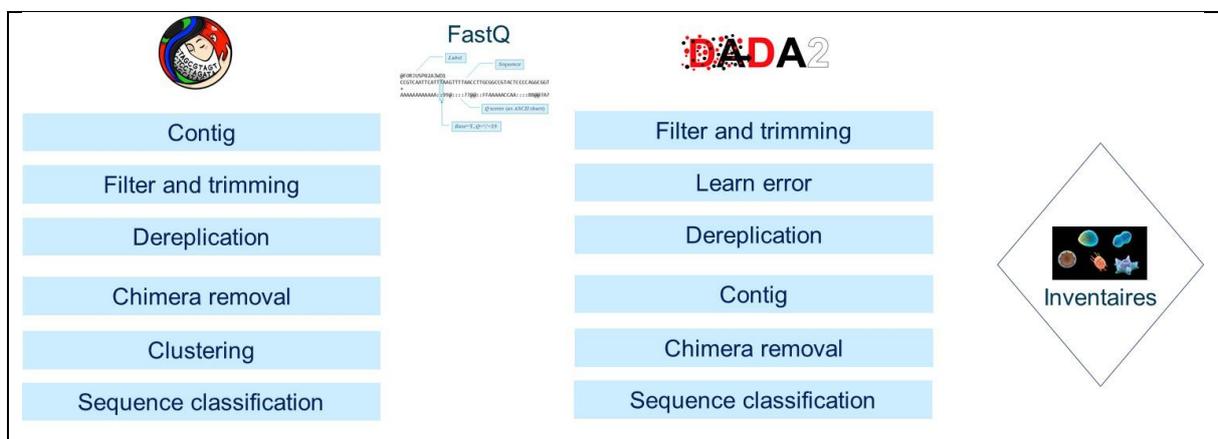


Figure 7. Etapes des deux approches bioinformatiques, Mothur et Dada2, transférées par Scimabio à LPL.

MOTHUR est un logiciel open-source développé pour l'analyse bioinformatique de données de séquençage d'ADN. Il est particulièrement utilisé dans le cadre de projets où l'objectif est de caractériser la diversité génétique dans des échantillons complexes. Il propose une interface en ligne de commande qui permet aux utilisateurs d'automatiser et de personnaliser leurs analyses en fonction de leurs besoins spécifiques. Dans l'approche transférée par Scimabio, il permet notamment de mettre en œuvre des regroupements de séquences sous forme d'OTU (Operational Taxonomic Unit) par la méthode de Furthest Neighbor.

DADA2 est un logiciel bioinformatique utilisé pour le traitement de données de séquençage d'ADN à haut débit. Il se distingue par sa capacité à effectuer un débruitage précis des données de séquençage d'amplicons, c'est à dire qu'il tente de résoudre les erreurs de séquençage et d'inférer les séquences exactes, ce qui correspond à une approche ASV (Amplicon Sequence Variant). Il propose un pipeline intégré qui couvre plusieurs étapes de l'analyse, notamment le filtrage de qualité, la détection d'erreurs, le débruitage, la chiméricité et la création de tables de comptage, de manière. Il est implémenté en langage de programmation R, et son utilisation se fait souvent via des scripts R.

LPL a ainsi implémenté deux approches transférées par Scimabio et a rédigé les procédures afin de pouvoir les mettre en œuvre sur les échantillons tests, l'objectif étant de stabiliser l'une des procédures dans un souci de routine.

Pour cela, les résultats de 9 échantillons terrain fournis par Artemis (Tableau 1), et d'échantillons calibrés (Mock, Lac, Rivière) fournis par Scimabio ainsi que des échantillons contrôles pour la technologie Miseq (NTC, NEGEXTR, DNAPOSE) ont été analysés à l'aide des deux pipelines. L'ensemble des fichiers obtenus a été transmis à Scimabio pour intercomparaison.

Tableau 1. Echantillons environnementaux utilisés lors de l'intercomparaison bioinformatique entre Scimabio et LPL.

	N° ECH	Cours d'eau	Taxons cibles
BV Gave et Nives	235	Bergons	ACLI
	236	Gave de Cauterets	ASBL
	237	Gave de Pau	ACAF + ACHD divers+DGEM
BV Adour	362	Bès 2019	ADCR
	370	Courant de Mimizan	CRCU
	415	Geloux	AASI
	445	Adour	ADMO
	483	Arrêt Darré	ADTH/AZHA
	577	Bès 2020	Eunotia sps

4.1.2.3.1 Comparaison des inventaires taxonomiques

Pour chaque échantillon intégré à l'intercomparaison bioinformatique, 2 inventaires taxonomiques ont été générés par LPL et 2 par Scimabio via l'utilisation des pipelines bioinformatiques sous Mothur et DADA2 (Figure 8). Les résultats préliminaires ont permis de mettre en avant :

- LPL vs Scimabio - Mothur : les inventaires obtenus par LPL et Scimabio sont assez similaires avec l'approche déployée sous Mothur. Toutefois, des variations dans la quantification de certains taxons peuvent être observées comme pour l'échantillon NEGEXTR où *Achnanthes minutissimum* qui représente 83% de l'assemblage pour LPL et 39% pour Scimabio. Cet échantillon étant représenté par un faible nombre de reads (<5 000), les proportions relatives des taxons peuvent rapidement être affectées par la variabilité introduite par Mothur lors des étapes basées sur des probabilités (Détection de chimère, Assignation taxonomique, Clustering). Certains échantillons peuvent aussi montrer des assemblages assez différents, comme *Diatoma ehrenbergii*, représentant environ 40% de l'assemblage de l'échantillon 237 pour LPL et quasi non détecté par Scimabio.
- LPL vs Scimabio – DADA2 : Le même constat est observé pour les inventaires obtenus par LPL et Scimabio en utilisant DADA2. Plus d'écarts d'identification ont été observés, comme par exemple avec l'échantillon 415 pour lequel *Achnanthes unclassified* représente 81% de l'assemblage pour LPL tandis que Scimabio détecte *Adchnanthes minutissimum* avec une abondance relative de 57%. L'intercomparaison a permis de mettre en évidence que des versions de la base Diat.barcode différentes ont été utilisées par LPL et Scimabio du fait de leur mise à jour automatique dans le script DADA2 sous R, expliquant la majorité des écarts d'assignation taxonomique observés.

Mothur vs. DADA2 : Lors de la mise en place de Mothur par LPL et Scimabio ont respectivement produit 341 OTUs (154 taxons) et 426 OTUs (167 taxons), tandis que l'approche DADA2 a permis de produire

1367 ASVs (192 taxons) et 1424 ASVs (165 taxons). L'approche DADA2 est connue pour permettre une identification taxonomique plus fine du fait de l'absence de clustering en OTU qui peut regrouper des séquences génétiques appartenant à la même espèce voire à des espèces différentes mais génétiquement proches. Le nombre d'OTU est donc plus faible que le nombre d'ASVs ce qui se traduit généralement sur le nombre de taxons détectés et la précision de leur assignation. Toutefois, les 2 approches peuvent être utilisées pour réaliser des inventaires taxonomiques diatomées pour des besoins de bioindication.

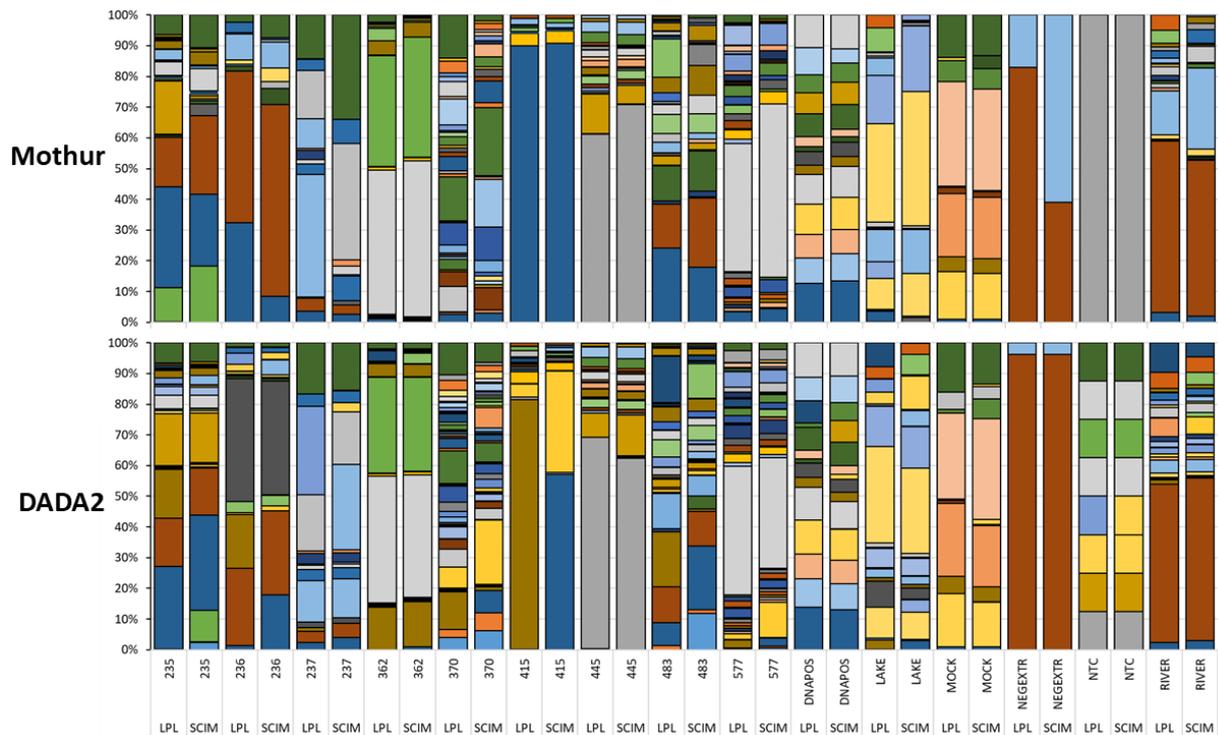


Figure 8. Inventaires taxonomiques diatomées obtenus en metabarcoding pour les 15 échantillons intégrés dans l'exercice d'intercomparaison entre LPL et Scimabio. Les approches bioinformatiques proposées par Mothur et DADA2 ont été comparées. Les résultats sont présentés pour les taxons avec une abondance de reads >0.5%.

4.1.2.3.2 Comparaison des structures de communauté

L'indice de dissimilarité Bray-Curtis a été calculé entre chaque échantillon afin de comparer les assemblages diatomées obtenus en metabarcoding par LPL et Scimabio avec les approches DADA2 et Mothur (Figure 9). Malgré la variabilité observée précédemment, les échantillons se regroupent bien en fonction de leur échantillon d'appartenance, avec peu d'écarts entre inventaires taxonomiques d'un même échantillon à l'exception de l'échantillon NTC, qui correspond à un témoin négatif ne contenant aucun taxon.

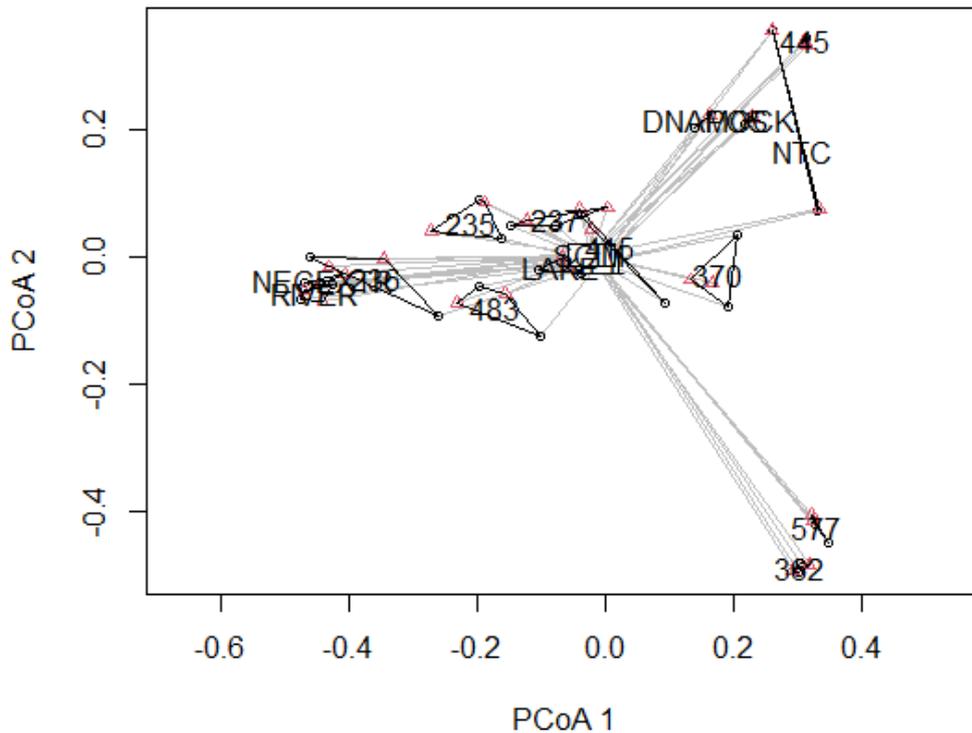


Figure 9. PCoA présentant la dissimilarité (Bray-Curtis) entre inventaires taxonomiques diatomées obtenus par LPL et Scimabio via l'approche Mothur (en Rouge) et DADA2 (en noir).

L'analyse Permanova a été utilisée pour évaluer les facteurs qui expliquent le plus la dissimilarité observée entre échantillons (Tableau 2). Parmi les différentes variables testées, il apparaît un effet significatif du choix de la stratégie bioinformatique (DADA2, Mothur), une différence naturelle dans les assemblages de diatomées des échantillons ainsi qu'un effet de la structure en charge de l'analyse (LPL, Scimabio). Bien qu'un effet de la structure ait été observé, sa significativité reste faible ($p=0,017$) et ne contribue à expliquer que 0,6% de la dissimilarité, tandis que la variabilité naturelle dans l'assemblage des échantillons explique à elle seule 83,8% de la dissimilarité. De plus, la dissimilarité entre LPL et Scimabio est principalement associée à l'utilisation d'une version de la base Diat.barcode différente lors de l'analyse via DADA2. Cette variabilité pourra donc facilement être réduite lors d'un déploiement en routine.

Tableau 2. Résultats de l'analyse Permanova pour déterminer les variables expliquant la dissimilarité dans les assemblages diatomées. Les variables significatives correspondent à la stratégie bioinformatique choisie (DADA2, Mothur), l'effet de la structure en charge de l'analyse (LPL, Scimabio) et l'effet associé à la diversité d'assemblage des échantillons directement (Sample).

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)	Signif.
Bioinf	1	0.5197	0.51969	7.1478	0.02217	0.000999	***
Structure	1	0.1470	0.14701	2.0219	0.00627	0.016983	*
Sample	14	19.6524	1.40375	19.3071	0.83822	0.000999	***
Residuals	43	3.1264	0.07271	0.13335			
Total	59	23.4455	1.00000				

4.1.2.3.3 Effet de la stratégie bioinformatique sur l'IBD

Les inventaires taxonomiques diatomées obtenus par LPL et Scimabio en utilisant Mothur et DADA2 ont été intégrés au logiciel Omnidia (version 6.1.4, base du 15/02/2022) afin de calculer les notes IBD diatomées pour les 11 échantillons naturels (

Tableau 3). L'objectif ici est de déterminer les écarts de notes IBD entre pour les différentes conditions testées et non d'interpréter l'état écologique. Les classes d'état écologiques n'ont donc pas été déterminées.

Tableau 3. Notes IBD calculées à partir des inventaires taxonomiques diatomées obtenus avec DADA2 et Mothur par LPL (ESV_LPL, OTU_LPL) et Scimabio (ESV_SCIM, OTU_SCIM). L'écart de note IBD maximum observé entre les 4 inventaires appelé Delta est aussi indiqué.

Notes IBD	ESV LPL	ESV SCIM	OTU LPL	OTU SCIM	Delta
235	20	20	19,2	20	0,8
236	20	20	20	20	0
237	20	18,4	19,7	20	1,6
362	13,5	13,5	13,4	13,5	0,1
370	14,6	16,6	15,1	15,4	2
415	20	20	20	20	0
445	11,6	10,7	10,6	11,2	1
483	18,1	18,6	18,7	18,4	0,6
577	13,6	13,1	13,4	13,7	0,6
DNAPOS	10,3	10,4	10,7	10,3	0,4
LAKE	20	20	20	19,7	0,3
MOCK	11,6	11,5	12,5	11,6	1
RIVER	20	20	20	20	0

Malgré la variabilité observée précédemment dans la détection et la quantification des taxons en fonction de l'approche bioinformatique utilisée entre LPL et Scimabio, les notes IBD obtenues dans les différentes conditions sont très similaires et fortement corrélées (Figure 10). Les plus gros écarts de notes sont observés pour les échantillons 237 (Delta = 1.6) et 370 (Delta = 2) sur les inventaires LPL et Scimabio obtenus avec l'approche DADA2. Cette variabilité est directement liée à l'utilisation de différentes versions de la base Diat.barcode pour l'assignation taxonomique via l'outil DADA2 et peut donc facilement être corrigé. Ces résultats tendent à confirmer que le calcul de l'IBD est assez robuste à la variabilité introduite par différents choix de stratégie bioinformatique ou par des manipulateurs différents.

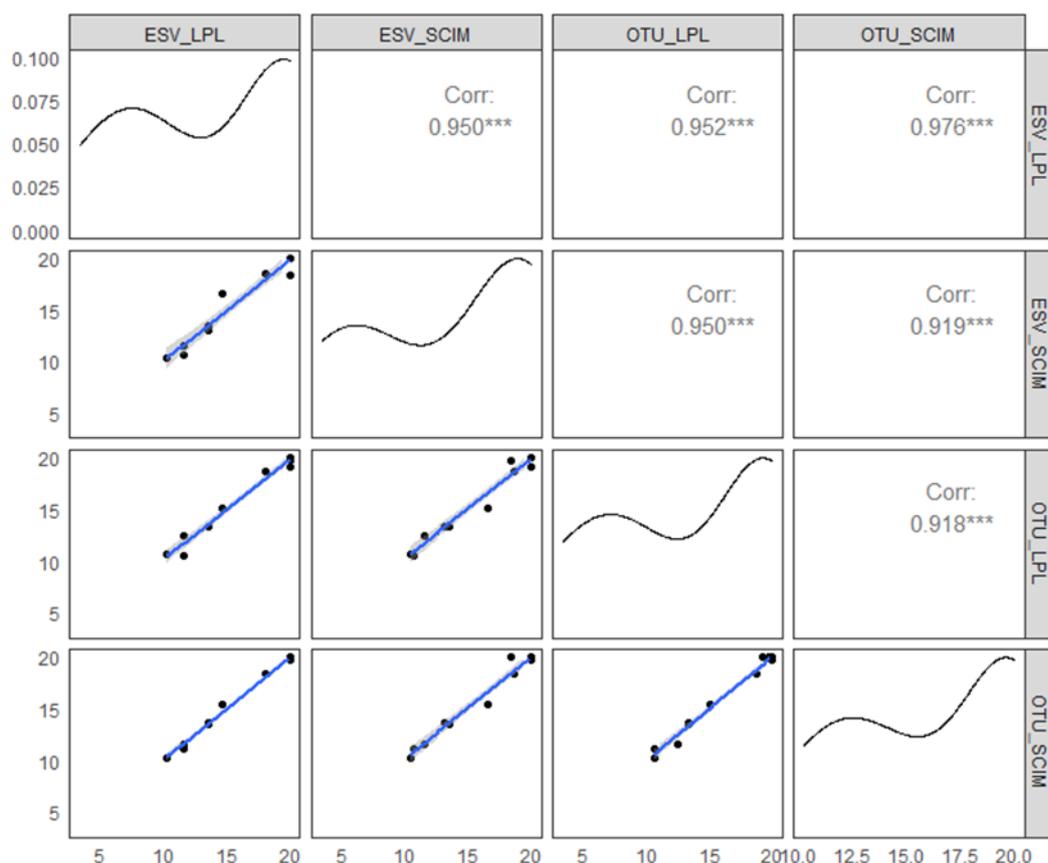


Figure 10. Corrélation 2 à 2 des notes IBD obtenues par LPL et Scimabio avec les approches Mothur et DADA2.

4.1.2.3.4 Validation du transfert de protocole bioinformatique au LPL

Le transfert des protocoles d'analyse bioinformatiques Mothur et DADA2 a pu être réalisé auprès du LPL. Les résultats de l'intercomparaison entre Scimabio et LPL ont montré que de la variabilité peut être introduite en fonction de l'approche utilisée (Mothur, DADA2) avec un effet cumulé de l'opérateur, pouvant affecter la détection et la quantification de certains taxons diatomées. Toutefois, pour des besoins de bioindication et le calcul de notes IBD, cette variabilité dans les inventaires taxonomiques reste négligeable et les écarts de notes IBD acceptables au regard de la variabilité observée lors des exercices d'intercalibration en microscopie. Pour des raisons de facilité d'implémentation dans le contexte du LPL, l'approche via Mothur a été retenue en 1^{er} choix.

Au regard de la forte variabilité observée lors de l'intercomparaison entre LPL et Scimabio, il serait intéressant de mettre des contrôles en interne pour s'assurer de la stabilité d'application de l'outil bioinformatique (contrôle qualité des données par étape d'analyse bioinformatique, version de base utilisée, ...). De plus, face à la grande diversité de logiciels et d'approches bioinformatiques existantes, il semblerait judicieux de prévoir des exercices d'intercalibration méthodologiques récurrents pour s'assurer de l'intercomparabilité de résultats produits par des approches différentes et qu'aucune dérive ne soit observée lorsqu'une même approche est utilisée par des opérateurs différents.

4.1.3 Application de la méthode transférée, par LPL

Après validation par Scimabio des résultats d'intercomparaison et donc de la capacité de LPL à réaliser la méthode, les analyses ont été réalisées en contexte local. L'étude a ainsi porté sur 378 échantillons terrains Gaves, Nive et Adour répartis sur deux années (185 biofilms datant de 2021 et 193 de 2022), prélevés par Artemis. Après analyse, l'ensemble des résultats générés ont été transmis sous forme de listes taxonomiques associées à leurs abondances à Artemis pour le calcul de la note IBD et comparaison des méthodes. Les fichiers fastq correspondant aux données brutes ainsi que les métadonnées ont été rendues publiques et publiées sur Zenodo (<https://zenodo.org/record/7848220>).

4.1.4 Optimisation de la méthode par LPL

Afin de répondre aux exigences d'analyse de routine, LPL a procédé à une optimisation des protocoles sur deux étapes : (i) l'extraction de l'ADN total et (ii) la partie bioinformatique dédiée à l'assignation à la base de référence Diat.Barcode.

Le protocole d'extraction d'ADN, transmis par Scimabio, bien que tout à fait fonctionnel, s'avère chronophage et coûteux (prix du réactif et du consommable par échantillon plus temps technicien) et peu adapté à de grosses séries d'échantillons et à la mise en place d'un contrôle qualité pour le suivi du processus. La partie extraction a donc été optimisée. Après plusieurs essais de protocoles alternatifs, le protocole retenu ajoute une étape de sonication, une étape de lyse et une purification automatisée sur robot. Les essais avec ce protocole optimisé sur quelques échantillons terrain ont ainsi permis de confirmer plusieurs points :

- Tous les extraits ADN obtenus sont quantifiables (non systématique avec la méthode originale) ce qui valide le bon déroulé de l'extraction et constitue un point de contrôle qualité pour la suite de l'analyse,
- Un abaissement des coûts d'analyse,
- Les échantillons montrent une diversité taxonomique plus large avec un gain des taxons rares,
- Une facilité et rapidité d'exécution adaptées au volume,
- Un gain de reproductibilité.

Concernant la partie bioinformatique, comme évoqué précédemment et pour des raisons de facilité d'implémentation aux LPL, l'approche via Mothur a été retenue avec la création des OTU par Furthest Neighbor. Afin d'optimiser l'assignation taxonomique, il a été décidé, en accord avec Scimabio et suite à divers essais comparatifs, de modifier le pourcentage de similarité utilisé en le passant de 95% à 99%. L'impact de ce changement dans le pipeline est détaillé en paragraphe 4.3.

La partie d'assignation taxonomique des OTUs à la base de référence génétique Diat.Barcode a été modifiée en intégrant une double méthode : i) le classificateur conventionnel RDP qui permet d'assigner rapidement un grand nombre de séquence ; ii) l'approche par BLAST est utilisée en 2^{ème} intention pour les OTUs qui n'ont pas pu être assignées à l'espèce via RDP.

Ces optimisations ont permis d'augmenter la finesse d'assignation taxonomique et d'améliorer l'adéquation entre les résultats issus des deux méthodes (ADN vs microscopie). De plus, cela a également permis d'augmenter le volume d'échantillons analysable à savoir environ 360 échantillons en 6 à 10 jours ouvrés hors calcul IBD. La validation de la méthode (répétabilité, reproductibilité, sensibilité, spécificité) est en cours aux LPL, dans la mesure des possibilités d'accès à des matériaux de référence et à des référentiels de validation. Enfin, afin de finaliser l'analyse jusqu'à l'obtention de la note IBD sur un site, au cours de ce projet, les données générées par séquençage par LPL et transformées en listes taxonomiques ont été systématiquement transmises à Artemis, diatomiste, qui s'est chargé de traduire ces listes en notes IBD et en classes de qualité. Ce mode de fonctionnement en duo, laboratoire générateur de données sur base d'approche de biologie moléculaire et diatomiste, expert en écologie des diatomées, nous a conforté dans l'idée qu'une telle méthode ne pouvait être conduite à bien qu'en associant les deux acteurs.

La méthode de metabarcoding ainsi optimisée a été appliquée et ainsi éprouvée sur des échantillons terrain prélevés par Artemis, qui ont permis la comparaison de ces échantillons. Ces résultats sont discutés en paragraphe 4.3.

4.1.5 Validation de la méthode

En tant que méthode d'analyse pouvant entrer dans un processus d'analyses routine, il est nécessaire de procéder à une validation complète en prenant en compte des critères bien spécifiques, ceci afin de caractériser la méthode sur des points tels que la reproductibilité, la répétabilité, la sensibilité et la fidélité. Ces étapes de validation comprennent une évaluation des résultats après répétitions de l'analyse en variant les conditions techniques et/ou opérateurs (reproductibilité), après répétitions dans les mêmes conditions (répétabilité), ou après dilution de l'échantillon (sensibilité et fidélité). Les premiers essais menés lors du projet Meta-IBD montrent une méthode reproductible et répétable. La validation complète de la méthode est en cours aux LPL, en s'appuyant notamment sur le référentiel technique pour la validation indépendante des outils utilisant l'ADN environnemental (Guigues et al, 2023).

4.1.6 Conclusion sur le transfert de la méthode

a/ Le transfert technique et bioinformatique dans un laboratoire de routine de biologie moléculaire peut se faire sans difficulté, à condition que ce transfert soit réalisé par un organisme compétent et conscient des pré-requis de laboratoire, et que le laboratoire receveur soit équipé et possède les compétences technique et scientifiques adéquates.

b/ La méthode est opérationnelle sur la partie laboratoire. Concernant la partie bioinformatique, elle est également opérationnelle mais peut être encore optimisée notamment par automatisation de certaines tâches d'assignation.

c/ 360 échantillons sont analysables en 6 à 10 jours ouvrés hors calcul IBD.

d/ La validation de la méthode est en cours aux LPL, en s'appuyant notamment sur le référentiel technique pour la validation indépendante des outils utilisant l'ADN environnemental (Guigues et al, 2023).

e/ L'évaluation de l'IBD doit se faire en duo laboratoire/diatomiste pour une bonne pertinence d'après notre expérience dans ce projet.

4.1.7 Axes d'optimisation pour la méthode

a/ La validation complète de la méthode doit être finalisée. Sur ce point, il est essentiel de réfléchir, dans le cadre d'une utilisation routine par les laboratoires dans le futur, à la mise en place d'interlocuteurs référents sur les différentes étapes de maintien des performances à savoir : (i) le transfert de méthode, (ii) la fabrication/stockage et maintien des matériaux de référence, (iii) la mise en place d'essais interlaboratoire, (iv) la mise à jour et l'entretien qualitatif de la base Diat.Barcode dans le temps, notamment en lien avec l'évolution taxonomique, (v) la veille technologique et bioinformatique pour pouvoir répondre à des éventuels changements ou abandon de technologies.

b/ L'assignation par Blast doit être automatisée pour parfaire la méthode et la conditionner en mode routine.

4.2 Evaluation de l'applicabilité de la base de référence Diat Barcode à l'échelle locale (Volet 2)

Cette partie visait à déterminer si la base de référence, créée et implémentée actuellement par INRAE à l'aide de données source variées, était adaptée à son utilisation dans un contexte local géographiquement, ici les bassins Adour, Nives, Gaves, dans le sud-ouest de la France et dans un cadre d'analyses de routine.

Deux étapes ont été réalisées pour évaluer l'applicabilité de la méthode et donc de la base de référence : (i) la comparaison des listes taxonomiques locales avec celle de la base Diat.Barcode et (ii) l'établissement de la liste des causes de divergences d'indices IBD obtenus respectivement par les deux méthodes.

4.2.1 Comparaison des listes taxonomiques : locale vs Diat.Barcode

Une liste de taxons retrouvés dans le bassin du Sud-Ouest a été réalisée par Artemis à partir de ses propres analyses réalisées en 2019 et 2020. Cette liste a été comparée à la liste issue de la base de référence Diat.Barcode, version 10 (voir lien internet en fin de document). Les résultats sont présentés dans le tableau suivant :

Nombre de taxons DB Artemis	632
Nombre de taxons Artemis présents dans DiatBarcode	209
Nombre de taxons Artemis absents dans DiatBarcode	423
Nombre de taxons pris en compte calcul IBD	330
% de taxons communs	33 %
% de taxons manquants	67 % dont 78% entrant dans la note IBD

Le nombre de taxons « locaux » manquants dans la base de référence Diat.Barcode représente un pourcentage de 67% dont 78% entrent dans le calcul de la note IBD ce qui laisse prédire à ce stade des performances moindres dans l'identification des taxons par la méthode ADN au niveau local. Cependant, sur ces 78%, seuls 9% (30 espèces sur 330) montrent une abondance relative supérieure à 5% dans les échantillons. Or, par expérience, les taxons montrant une abondance relative inférieure à 5% contribuent de manière négligeable dans l'indice IBD. La complétion immédiate de la base pour l'optimisation de l'utilisation de l'outil au niveau local, en intégrant les taxons spécifiques aux hydro-écorégions, peut donc se faire sur un nombre relativement faible de taxons et donc plus rapidement. Une complétion plus complète peut intervenir dans un deuxième temps afin d'obtenir une base exhaustive.

4.2.2 Liste des causes de divergence entre les deux méthodes

Cette deuxième étape a été réalisée conjointement par LPL et Artémis, à partir du jeu de données constitué dans le cadre de Meta-IBD en 2021 et 2022, à partir des échantillons traités par Artemis pour le compte de l'Agence de l'Eau Adour Garonne, dans le cadre de la surveillance DCE, et co-traités par LPL, pour la méthode ADN (prélèvement réalisé par Artemis, extraction ADN/amplification rbc1/préparation des librairies et séquençage MiSeq (Illumina)/traitement bioinformatique des séquences/listes taxonomiques par LPL). Les divergences ont été considérées seulement sur les échantillons montrant des écarts d'IBD supérieurs à 2, une différence inférieure pouvant être considérée comme un biais inhérent à la méthode de référence. Les résultats sont consignés dans le Tableau 4.

Tableau 4. Liste des divergences entre les résultats obtenus par microscopie et par metabarcoding, par rapport à cette dernière.

Problématique identifiée	Taxons concernés	Causes et/ou correction
Taxons absents dans Diatbarcode	32 taxons	Implémentation dans Diatbarcode
Taxons non identifiés en ADN mais présents dans Diatbarcode	10 taxons (NDIS, ESLE, ADLA, NDIS, NZSU, APED, AD SH, CSNU, DPSG, ADRI)	Frustules vides ?
Mauvaise identification par ADN	9 taxons (ACHD, ADMI, COCO, ADSB, ACHD, ADMI, NAVI, ADSB, SELL)	Résolution par Blast = 5 taxons Problème persistant = 4 taxons
Non identifiable à l'espèce en ADN	8 taxons (NITZ, AMID/AMPH, NITZ, BACI/NITZ, CAGR, NFIL/NFIC, CATO, AMID)	Pas de distinction entre espèces
Niveau de finesse taxonomique	3 taxons (GPAR/GSPP, NPAD/NPAL, CPLA/CPLI)	Variants non identifiés/non différenciés dans Diatbarcode ou taxons laissés en sp
Divergence d'identification taxonomique	9 taxons (GPAR/GSPP, CPLA/CPLI, ADMI, Ulnaria, DCOF, GMIN/GOMP, RUNI, NCTE, NINC)	Résolution par Blast = 5 taxons Problème persistant = 4 taxons
Surestimation abondance par ADN	18 taxons (NPAL, MVAR, GMMI, NAMC, NESCE, DVUL, NLAN, CPED, UULN, NTPT, GMMI, UULN, MVAR, N LIN, EMIN, DVUL, CTUM, PLEV)	Grandes espèces (>50 µm) et/ou coloniales Facteur de correction ?
Sous-estimation abondance par ADN	17 taxons (KOBG, NAMP, AD SH, ADPY, NCTE, APED, SNIG, ADMI, KOBG, EEXI, NPAE, NAMP, NFON, MP MI, SNIG, CSNU, DPSG)	Petites espèces Facteur de correction ?

La base de référence Diat.Barcode montre des lacunes sur un certain nombre de taxons présents dans la zone d'étude et sur le plan également de la taxonomie proprement dite. Ainsi, pour être rendue applicable dans le cadre d'une analyse de routine, cette base doit répondre à plusieurs critères :

- Prise en compte seulement des taxons présents dans le territoire français, comprenant ceux entrant dans le calcul IBD. Ceci permet de maintenir une diversité dans la base de référence afin de limiter les erreurs d'assignations tout en l'allégeant pour focaliser sur les taxons informatifs,
- Implémentation des taxons manquants, prioritairement ceux montrant des abondances > 5%, au niveau national, avec prise en compte des spécificités locales par hydro-écorégion,
- Affiner la taxonomie sur certains taxons restés en sp,
- Rendre cohérent la taxonomie de la base, créée à l'origine pour des questions et besoins de recherche, avec la taxonomie de routine pour les analyses IBD, plus opérationnelle et permettant de faire certaines simplifications (e.g. regroupement de taxons proches avec écologies similaires).

Un axe intéressant pourrait être de diviser la base Diat.Barcode existante en deux bases différenciées : (i) une base générale, dédiée à la recherche et à l'implémentation de nouveaux taxons sur le plan écologique et en rapport direct avec l'évolution de la taxonomie des diatomées, et (ii) une base dédiée et spécifique à l'utilisation opérationnelle et donc au calcul de l'IBD. Cela impliquerait donc de reprendre la base Diat.Barcode et de la modéliser en fonction de ces critères.

4.2.3 Conclusion sur l'applicabilité de la méthode en local

La base Diat.Barcode n'est pas suffisamment adaptée en l'état à la mise en œuvre de la méthode de metabarcoding sur ADN environnemental, dans un cadre de routine pour l'évaluation de la qualité de l'eau des rivières dans un contexte DCE. Il est cependant nécessaire de garder à l'esprit que la méthode en microscopie montre également des biais qui peut conduire à une divergence des méthodes et donc une moins bonne corrélation des notes IBD en découlant. Ces biais sont décrits par la suite.

4.2.4 Axes d'optimisation de la méthode pour l'applicabilité en local

La base Diat.Barcode doit être divisée et réadaptée spécifiquement pour la rendre opérationnelle en routine. A long terme, pour des besoins de routine et uniquement dans ce cadre, il est nécessaire qu'elle soit maintenue, curée taxonomiquement en lien avec l'évolution taxonomique morphologique, par un organisme dédié.

De plus les connaissances taxonomiques opérationnelles sont mises à jour annuellement pour la microscopie. Lors de ces exercices, la taxonomie est discutée et leur codification sous forme de code à 4 lettres est définie pour simplifier l'intégration des inventaires dans les logiciels de calculs de notes IBD (SEEE, Omnidia). Il serait aussi pertinent que la base Diat.barcode opérationnelle bénéficie de ces

mises à jour annuelles taxonomiques et de codifications des taxons pour faciliter le calcul opérationnel des IBD ADN. Cela permettrait aussi de garder de la cohérence avec la taxonomie utilisée pour les IBD morphologiques.

4.3 Etude comparative méthode de référence vs metabarcoding

4.3.1 Le protocole d'étude

L'étude a été menée à partir de prélèvements réalisés en 2021 et 2022, dans le cadre du réseau de surveillance de l'Agence de l'Eau Adour-Garonne (AEAG), en particulier deux bassins versants proches géographiquement de nos structures, les bassins de l'Adour et des Gaves et Nives. La répartition des échantillons pris en compte est détaillée dans le Tableau 5.

Tableau 5. Nombre et répartition des stations échantillonnées.

	2021	2022
Adour	121	129
Gaves et Nives	65	66
TOTAL	186	195

Au total, 186 et 195 échantillons ont été prélevés en 2021 et 2022, respectivement. Parmi ces prélèvements, quelques rares échantillons n'ont pas donné lieu à des listes taxonomiques en raison du manque de diatomées pour la microscopie et pas assez de séquences d'ADN pour la biologie moléculaire.

Les inventaires d'espèces de diatomées issus de la méthode ADN ont été fournis à Artemis par LPL afin de déterminer l'IBD, *via* une intégration dans le logiciel OMNIDIA pour le calcul de l'indice. En 2021, ont été comparés les inventaires issus d'échantillons traités par LPL de diverses manières :

- Avec des séquences représentatives d'OTU construits à 95% de similarité (OTU 95%), avec et sans transformation par biovolumes,
- Avec des séquences représentatives d'OTU construits à 99% de similarité (OTU 99%), avec et sans transformation par biovolume.

Les résultats obtenus en 2021 ont permis d'identifier la méthode d'analyse ADN la plus performante parmi celles testées et, pour les prélèvements de 2022, l'analyse a été menée avec les inventaires issus du traitement avec des OTU 99% sans transformation par biovolume ; le biovolume n'a pas été retenu en raison de la non-pertinence de la note IBD après transformation (voir plus bas).

Des résultats après traitement par classificateur BLAST ont également été soumis à la comparaison des IBD de manière générale (corrélation) et de la composition des peuplements pour une sélection de stations (celles qui présentaient une grande différence de note notamment).

Le Tableau 6 synthétise les traitements informatiques aux données issues de la méthode ADN réalisés en fonction des années d'échantillonnage.

Tableau 6. Protocole de traitement des données issues de la méthode ADN.

Année de prélèvement	OTU 95% de similarité		OTU 99% de similarité		
	Sans biovol	Avec biovol	Sans biovol	Avec biovol	Sans biovol + BLAST
2021	x	x	x	x	X
2022	x		x		x

4.3.2 Pertinence de la note IBD

Pour rappel, selon la norme N FT 90-354 relative à la mise en application de l'IBD, le pourcentage d'unités diatomiques contributif au calcul de l'IBD doit être pris en compte. Si celui-ci est (i) inférieur ou égal à 25 %, aucune note IBD ne peut être attribuée ; l'IBD est alors non calculable (NC), (ii) compris entre 25 et 50 %, l'IBD doit être émis avec des réserves.

Dans un premier temps, toutes les listes taxonomiques obtenues ont été analysées sur cette base : est-ce que l'IBD obtenu à partir de chacune de ces listes est pertinent ?

Échantillons de 2021

Nous avons comparé la pertinence de la note IBD à partir des deux méthodes, microscopie et ADN, sans transformation en biovolume (avec OTU 95% et OTU 99%), puis après transformation en biovolume (avec OTU 95% et OTU 99%).

Avec la microscopie, sur l'ensemble des relevés, **18 IBD sont émis avec des réserves et 0 non calculable**. Avec l'ADN, **34 sont émis avec des réserves et 1 et non calculable**.

Avec la microscopie, tous les IBD « réservés » sont liés à l'abondance de ADMO, qui n'est pas intégré à l'IBD à ce jour mais le sera à la prochaine révision de cet indice. Une fois cette révision appliquée, il ne restera aucun IBD réservé sur ce même jeu de données.

Avec l'ADN, certains IBD sont réservés également à cause d'ADMO (16) ; il en resterait 18 une fois le profil d'ADMO intégré à l'IBD, et aucun ne serait non calculable.

Avec les OTU à 99% sans biovolume, le même travail a été fait et montre que l'on on a davantage de réussite vis-à-vis de la pertinence de l'IBD (Tableau 7). En revanche, une fois les données transformées en biovolume, **beaucoup plus d'IBD sont émis avec des réserves et non calculables**.

Le bilan sur l'ensemble des échantillons de 2021 est présenté dans le Tableau 7.

Tableau 7. Nombre d'IBD réservés et non calculables pour les différents protocoles d'analyse.

BILAN PERTINENCE IBD en nb de relevés en 2021				
	MICROSCOPIE	ADN sans BIOVOL95	ADN sans BIOVOL99	ADN BIOVOL 95
TOTAL réserve	18	34	27	70
TOTAL Non calculable	0	1	1	77

Pour l'ADN, avec les OTU 99% avec biovolume, les relevés n'ont pas été listés un à un car, à l'instar des résultats obtenus avec les OTU à 95%, il y a de très nombreux IBD réservés et non calculables. Cela est lié au fait que les diatomées non identifiées à l'espèce et laissées au genre comme GOMP, ACHD, AMPH, NITZ prennent beaucoup d'importance en biovolume et se rajoutent à ADMO (Tableau 8).

Tableau 8. Nombre d'IBD réservés et NC avec ou sans ADMO

AVEC ADMO				SANS ADMO			
BILAN PERTINENCE IBD en nb de relevés en 2021				BILAN PERTINENCE IBD en nb de relevés en 2021			
	MICROSCOPIE	ADN sans BIOVOL95	ADN sans BIOVOL99		MICROSCOPIE	ADN sans BIOVOL95	ADN sans BIOVOL99
TOTAL réserve	18	34	26	TOTAL réserve	0	18	10
TOTAL Non calculable	0	1	1	TOTAL Non calculable	0	0	0
BILAN PERTINENCE IBD en % sur 179 relevés				BILAN PERTINENCE IBD en % sur 179 relevés			
	MICROSCOPIE	ADN sans BIOVOL95			MICROSCOPIE	ADN sans BIOVOL95	
TOTAL réserve	10%	19%	15%	TOTAL réserve	0%	10%	6%
TOTAL Non calculable	0%	1%	1%	TOTAL Non calculable	0%	0%	0%

Échantillons de 2022

Le même travail a été réalisé pour les échantillons de 2022 sans transformation des inventaires en biovolume, compte tenu des résultats obtenus en 2021. De même, ont été uniquement réalisées les analyses avec les OTU 99%.

En microscopie, en 2022, **15 IBD sont réservés et 2 non calculables**. Avec l'ADN, on obtient **27 IBD réservés et 4 non calculables**.

Lorsque le profil d'ADMO sera intégré au calcul de l'IBD, on aura plus qu'un seul IBD réservé en microscopie et 0 non calculable. Avec, l'ADN, 9 réservés et 1 non calculable.

La synthèse des résultats obtenus avec l'analyse ADN et OTU à 99% est donnée en Tableau 9.

Tableau 9. Nombre d'IBD réservés et non calculables avec ou sans ADMO, pour les échantillons de 2022.

BILAN PERTINENCE IBD en nb de relevés			SANS ADMO		
	MICROSCOPIE	ADN sans BIOVOL99		MICROSCOPIE	ADN sans BIOVOL99
TOTAL réserve	15	27	TOTAL réserve	1	9
TOTAL Non calculable	2	4	TOTAL Non calculable	0	1
BILAN PERTINENCE IBD en % sur 192 relevés			BILAN PERTINENCE IBD en % sur 192 relevés		
	MICROSCOPIE	ADN sans BIOVOL99		MICROSCOPIE	ADN sans BIOVOL99
TOTAL réserve	8%	14%	TOTAL réserve	1%	5%
TOTAL Non calculable	1%	2%	TOTAL Non calculable	0%	1%

En ce qui concerne la pertinence des notes IBD, les résultats sont comparables entre 2021 et 2022 pour le même type d'analyse.

Cette pertinence est très dépendante de l'abondance d'*Achnanthydium delmontii* (ADMO) (Tableau 9), espèce abondante dans le jeu de données et non prise en compte encore dans l'IBD. La prise en compte de ce taxon est en train d'être faite, ce qui va augmenter les performances au niveau de ce critère de pertinence.

4.3.3 Corrélation des notes IBD MIC vs ADN

La comparaison des notes IBD entre la microscopie (M) et l'ADN est faite, d'une manière générale, à partir de la corrélation entre les notes et évaluée par des coefficients de corrélation. Cela a été fait pour les différents traitements réalisés.

Échantillons de 2021 et 2022

La comparaison des notes IBD entre la microscopie (M) et l'ADN a été faite pour les données avec des OTU 95% et 99%, corrigé ou non par le biovolume en 2021. Le même exercice a été réalisé sur les données de 2022, avec cependant les inventaires issus seulement du traitement avec des OTU 99% sans transformation en biovolume en raison des résultats précédemment obtenus.

Le bilan issu de l'analyse de ces courbes est donné dans le Tableau 10 :

Tableau 10. Bilan de la corrélation des notes IBD obtenues par M et par ADN.

		M et ADN 95% non cor biovol	M et ADN 99% non cor biovol		M et ADN 99% non cor biovol+BLAST	M et ADN 95% cor biovol	M et ADN 99% cor biovol
2021	tous relevés	0,5401	0,5793		0,5769	0,6806	0,7186
	bassins séparés		Gaves et Nives 0,6159	Adour 0,5263			
2022	tous relevés		0,6432		0,6709		
	bassins séparés		Gaves et Nives 0,6832	Adour 0,6102			

En 2021, on note que la corrélation des notes IBD est améliorée avec la transformation des inventaires en biovolume surtout avec des OTU de 99%.

Cependant, il subsiste le problème de la pertinence de la note IBD si on applique la règle de la norme comme cela a été vu précédemment.

La surestimation des espèces à gros biovolume en metabarcoding et à petits biovolumes en microscopie tend à générer des écarts de notes IBD orientés dans une même direction car généralement conduits par les mêmes taxons. L'utilisation de la correction par le biovolume permet de corriger de manière artificiel cet écart, ce qui a pour effet d'améliorer la corrélation des notes IBD entre les deux approches. Toutefois, à l'échelle d'une station, l'écart de notes IBD morpho et ADN peut être amélioré lorsque des espèces à gros et petits biovolumes coexistent, mais si la plupart des espèces ont un biovolume équivalent, la correction empire les écarts de notes. De plus, une mise à niveau du facteur de correction de biovolume serait nécessaire pour confirmer la formule de correction sur un plus grand nombre de taxons. Un travail d'application dans des contextes environnementaux variés serait aussi nécessaire pour valider son niveau d'opérationnalité en routine.

En 2021 comme en 2022, la méthode ADN donne de meilleurs résultats avec les OTU 99% comparativement à 95%. Les corrélations sont meilleures dans le bassin versant Gaves et Nives comparativement à Adour. La diversité est moins grande dans ce bassin versant, ce qui peut expliquer ce résultat.

Le traitement BLAST permet d'améliorer la corrélation entre les notes IBD issues de la microscopie et de l'ADN, en particulier en 2022.

4.3.4 Sur les classes de qualité

A l'instar de ce qui a été fait pour la microscopie, les notes IBD issues de l'ADN ont été traduites en classe de qualité en fonction de l'appartenance de chaque station à son hydro-écorégion.

Avec la méthode ADN, les résultats montrent, comparativement à la microscopie, moins de stations dans la classe très bonne et davantage de sites qui entrent dans la classe de mauvaise qualité.

Par exemple, pour le bassin Adour en 2021, on obtient 15 stations dans la classe très bonne contre 25 avec la microscopie et 4 dans la classe mauvaise contre 0 avec la microscopie.

Pour les plus grandes différences de notes (>2 points), une comparaison de la composition des peuplements station par station a été réalisée afin d'expliquer ces divergences.

4.3.5 Sur la composition des peuplements

Les principaux problèmes détectés ont été listés dans le Tableau 4.

Le détail de la comparaison des peuplements pour tous les sites concernés par une différence de note IBD d'au moins deux points est donné en annexes.

Certains problèmes rencontrés sont récurrents et peuvent induire des différences de notation importante :

- **Identification de GPAR en microscopie contre GSPP en ADN, taxon plus polluo-résistant donc « tirant » la note IBD vers le bas ;**

L'identification par l'ADN de GSPP (*Gomphonema saprophilum*) contre GPAR (*G. parvulum*) en microscopie, induit d'importantes différences de notes, en particulier lorsque ce taxon est abondant ; à titre d'exemple dans l'Onesse à l'amont du R. d'Hossegor, on obtient une différence de plus de 11 points à cause de cette espèce dominante.

Il en est de même pour l'Estampon à Nord de St-Gor, station pour laquelle on obtient une différence de note de plus de 4 points en raison de la détermination différente de ce taxon.

Lorsque ce taxon est abondant, la note obtenue avec l'ADN est donc systématiquement plus basse que celle calculée avec la microscopie. Ce problème pourrait être facilement résolu.

- **Surestimation en microscopie de frustules sans doute morts (photos prises sans traitement des frustules) dans certains cas *Achnantheidium minutissimum* (ADMI), *Karayevia oblongella* (KOBG) « tirant la note vers le haut » ; c'est le cas au Rejtons à Tartas, en 2021 comme en 2022.**

Au niveau de cette station, l'ADN détecte peu KOBG ainsi qu'ADMI, alors qu'en microscopie, ces espèces sont identifiées parmi les dominantes ; KOBG est bien présente dans la base Diat.Barcode et LPL la détecte bien si elle est présente ; l'hypothèse la plus probable est que les frustules sont vides donc les cellules sont mortes, comme l'a montré un examen sur le matériel brut (Figure 11). Cette espèce est bien notée par l'IBD ce qui induit une différence importante de note, comme en 2021 (plus de 7 points d'écart).



Figure 11. Cellules vides d'ADMI et KOBG – matériel brut (non traité)

- Surestimation en ADN des grandes espèces et/ou coloniales

Les espèces de grandes taille et/ou coloniales sont logiquement plus représentées en ADN comparativement à la microscopie car elles contiennent davantage d'ADN comparativement aux petites. De plus, lors du comptage les grandes espèces sont moins comptées surtout lorsque beaucoup de petites formes sont présentes dans un champ de comptage. De plus, le traitement de l'échantillon casse les chaînes et les colonies dans la plupart des cas, ce qui implique que l'on ne voit que des cellules isolées ou groupées par deux ou trois seulement (Figure 12).

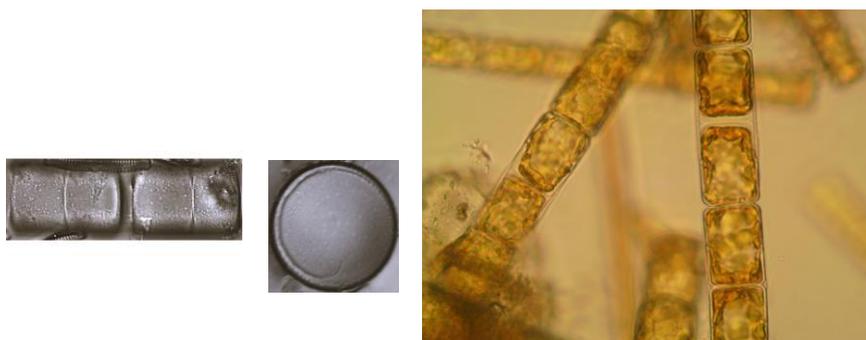


Figure 12. Exemple de *Melosira varians* (MVAR) à gauche matériel traité à droite matériel brut montrant de longues chaînes

Dans notre étude, plusieurs stations sont concernées par ce biais comme le Bès à Villenave (2021), station pour laquelle on note une différence d'IBD de près de 7 points en raison de la surestimation de MVAR (0,2 % en microscopie vs 27% en ADN).

D'autres taxons sont concernés avec systématiquement une surestimation avec l'ADN : *Gomphoneis minuta* (GMMI). Dans l'Echez à Maubourguet notamment, le taxon est nettement dominant avec l'ADN, alors que son abondance relative est faible en microscopie. De la même façon que MVAR, ce taxon est colonial, les grandes espèces formant des arbuscules les cellules étant reliées entre elles par des tubes muqueux.

La liste des taxons provoquant ce type de différence est donnée dans le Tableau 4.

- **Sous-estimation en ADN des petites espèces**

A l'inverse du cas précédent, les petites espèces contiennent moins d'ADN que les grandes et sont sous estimées comparativement à la microscopie. Si elles sont abondantes, la différence de note peut être conséquente comme avec *Sellaphora nigri* (SNIG) ; à titre d'exemple, en 2022, dans le Gave de Pau à l'aval d'Orthez cette espèce représente 67% en microscopie contre 7% avec l'ADN. Assez mal notée par l'IBD, la différence de note est donc importante (6 points de moins avec l'ADN).

La liste des taxons provoquant ce type de différence est donnée dans le Tableau 4.

Il est à noter que les différences de notes IBD liées à la quantité d'ADN dépendant de la taille des diatomées peuvent se coupler à la mortalité éventuelle des frustules.

- **Détection d'ADMI pas toujours évidente en ADN et resté au genre.**

Le traitement par BLAST a permis de détecter des taxons non identifiés à l'espèce sans BLAST, ce qui montre l'intérêt de ce traitement.

Le taxon *Achnanthydium minutissimum* (ADMI) *sensu lato* regroupe un nombre important de formes proches difficilement distinguables en microscopie optique c'est-à-dire en routine ; le profil ADN regroupe ces taxons très proches sous « ADMI » ; néanmoins, avec l'ADN, il est souvent laissé au genre (ACHD) et non déterminé à l'espèce alors qu'il peut être abondant en microscopie. C'est le cas dans le Rejtons à Tartas en 2021 ; en microscopie, on identifie 16% d'ADMI contre 0,5% en ADN mais 2,8% ACHD. Il peut y avoir également l'impact de frustules vides comme cela a été vu précédemment.

Autre exemple, dans le Bès à Villenave, on identifie 16% d'ADMI en microscopie alors qu'il est laissé au niveau du genre en *Achnanthydium* (ACHD) par l'ADN.

Dans certains cas, le traitement BLAST ou la nouvelle extraction a permis de résoudre le problème d'identification d'ADMI. C'est le cas du Gabas au niveau de Ger ou encore dans le Boudigou à la ferme d'Arman ou le traitement BLAST ou la nouvelle extraction a permis d'assigner ADMI.

Les résultats obtenus montrent qu'avec l'amélioration de ses méthodes, LPL peut optimiser la détection des taxons.

- **Comparaison de la richesse taxonomique**

Il ressort également de cette étude une moindre richesse taxonomique à partir de l'ADN, comparativement aux inventaires issus de la microscopie. En moyenne, sur l'ensemble des relevés, on obtient 31 taxons pour la microscopie, contre seulement 12 pour l'ADN en 2022 (Tableau 11).

Ce manque de richesse est lié surtout au manque de taxons dans la base diatomées actuellement utilisée comme le montre le Tableau 4.

Tableau 11. Comparaison de la richesse taxonomique obtenue avec les deux méthodes, MIC et ADN.

	MOY	MAX	MIN
ADN Blast 2022	12	21	1
MIC 2022	31	71	7

5 Conclusion et perspectives

Le projet a montré que la méthode ADN donne des résultats relativement satisfaisants et encore perfectibles en ce qui concerne l'estimation de la qualité biologique du milieu avec toutefois une différence en fonction du jeu de données ; on a vu que les résultats sont nettement meilleurs pour les stations du bassin versant des Gaves et Nives par rapport à l'Adour, que ce soit en 2021 ou en 2022, cela en raison d'une richesse taxonomique moins grande. En 2021, 79% des notes ont une différence inférieure à 2 points et dans 18% des cas, on obtient aucune différence de note ; en 2022, 81% des notes ont une différence inférieure à deux points et 17% n'ont aucune différence. Ce pourcentage sera encore plus élevé avec la prise en compte d'ADMO.

Le projet a permis de mettre en évidence des pistes de travail pour améliorer la méthode et la rendre encore plus performante à l'échelle locale.

Le projet a également permis de se projeter dans un cadre opérationnel et a fait ressortir certains problèmes dans l'application de l'IBD *sensu* l'utilisation de la norme NF-T-90-34, notamment en ce qui concerne le pourcentage d'unités taxonomiques. On a vu que la transformation en biovolume rendait les résultats peu probants avec beaucoup d'IBD non pertinents ou même non calculables. La question se pose alors de l'aspect normatif.

Cette notion de taille est à prendre en considération en effet ; logiquement les grandes espèces sont mieux représentées avec l'ADN :

- Espèces coloniales : *Melosira varians* (MVAR), *Gomphoneis minuta* (GMMI), *Diatoma vulgare* (DVUL)
- Grandes diatomées : *Pleurosira laevis* (PLEV), *Ulnaria ulna* (UULN), *Navicula tripunctata* (NTPT), *N. lanceolata* (NLAN)...

Une solution doit être trouvée/ testée pour prendre en considération cela :

- Facteur de correction empirique ?
- Modification de la façon de compter au microscope avec un sous-comptage des grandes espèces avec un grossissement moindre (x 40 au lieu de x 100) ?

A ce jour, les deux méthodes peuvent être complémentaires et apporter chacune des informations utiles dans le but d'améliorer le diagnostic de l'état écologique des cours d'eau. Chacune peut permettre à l'autre de s'améliorer. Chacune d'elle possède des biais et des atouts qu'il faut essayer d'identifier afin d'analyser leur impact et les utiliser pour améliorer les méthodes.

L'ADN peut sans doute permettre de revoir l'identification morphologique dans le cas d'espèces proches morphologiquement : NPAD/NPAL ; GPAR/GSPP ; GPUM/GBOB.

Concernant le grand « complexe ADMI » : la séparation des taxons est difficile en microscopie. La méthode ADN permet peut-être un moyen de progresser dans l'identification de ces formes.

L'ADN permet certainement une meilleure estimation de la qualité du milieu quand il y a beaucoup de frustules morts. Par exemple dans le Retjons à Tartas ; la note IBD issue de l'ADN est plus en adéquation avec la physicochimie du milieu ; la méthode microscopique peut-elle être améliorée sachant cela, par l'établissement d'un % mort/vivant sur le matériel frais ?

Les différentes pistes d'amélioration sont confrontées à un contexte normatif qui impose, pour la mise en œuvre de l'IBD, de suivre les directives de la norme NFT 90-354, ce qui ne permet pas d'appliquer les différentes mesures testées et applicables à ce jour.

6 Références

Références bibliographiques

Guigues Nathalie, Vitte Isabelle, Betat Sylvie, Sire Lucas, Boissery Pierre, Domaizon Isabelle, Dubut Vincent. 2023. Référentiel technique pour la validation indépendante des outils utilisant l'ADN environnemental. Convention OFB LNE n° 21.0902 du 08/09/21.

Groendahl, Sophie, Maria Kahlert, et Patrick Fink. « The Best of Both Worlds: A Combined Approach for Analyzing Microalgal Diversity via Metabarcoding and Morphology-Based Methods ». Édité par Hideyuki Doi. PLOS ONE 12, n° 2 (24 février 2017): e0172808. <https://doi.org/10.1371/journal.pone.0172808>.

Kermarrec L. 2012. Apport des outils de la biologie moléculaire pour l'utilisation des diatomées comme bioindicateurs de la qualité des écosystèmes aquatiques lotiques et pour l'étude de leur taxonomie. Thèse de doctorat. Université de Grenoble. 297 p.

Kermarrec, L., A. Franc, F. Rimet, P. Chaumeil, J. F. Humbert, et A. Bouchez. 2013. « Next-Generation Sequencing to Inventory Taxonomic Diversity in Eukaryotic Communities: A Test for Freshwater Diatoms ». *Molecular Ecology Resources* 13, no 4: 607-19. <https://doi.org/10.1111/1755-0998.12105>.

NF T 90-354. Qualité de l'eau – Echantillonnage, traitement et analyse de Diatomées benthiques en cours d'eau et canaux. 23 avril 2016. AFNOR. 199 p.

Rimet, Frédéric, Philippe Chaumeil, François Keck, Lenaïg Kermarrec, Valentin Vasselon, Maria Kahlert, Alain Franc, et Agnès Bouchez. « R-Syst::Diatom: An Open-Access and Curated Barcode Database for Diatoms and Freshwater Monitoring ». *Database* 2016 (2016): baw016. <https://doi.org/10.1093/database/baw016>.

Rivera, S. F., V. Vasselon, S. Jacquet, A. Bouchez, D. Ariztegui, et F. Rimet. « Metabarcoding of Lake Benthic Diatoms: From Structure Assemblages to Ecological Assessment ». *Hydrobiologia* 807, n° 1 (février 2018): 37-51. <https://doi.org/10.1007/s10750-017-3381-2>.

Vasselon, Valentin, Agnès Bouchez, Frédéric Rimet, Stéphan Jacquet, Rosa Trobajo, Méline Corniquel, Kálmán Tapolczai, et Isabelle Domaizon. « Avoiding Quantification Bias in Metabarcoding: Application of a Cell Biovolume Correction Factor in Diatom Molecular Biomonitoring ». Édité par Andrew Mahon. *Methods in Ecology and Evolution* 9, n° 4 (avril 2018): 1060-69. <https://doi.org/10.1111/2041-210X.12960>.

Visco, Joana Amorim, Laure Apothéoz-Perret-Gentil, Arielle Cordonier, Philippe Esling, Loïc Pillet, et Jan Pawlowski. « Environmental Monitoring: Inferring the Diatom Index from Next-Generation Sequencing Data ». *Environmental Science & Technology* 49, n° 13 (7 juillet 2015): 7597-7605. <https://doi.org/10.1021/es506158m>.

Zimmermann, Jonas, Gernot Glöckner, Regine Jahn, Neela Enke, et Birgit Gemeinholzer. « Metabarcoding vs. Morphological Identification to Assess Diatom Diversity in Environmental Studies ». *Molecular Ecology Resources* 15, n° 3 (mai 2015): 526-42. <https://doi.org/10.1111/1755-0998.12336>.

Liens internet vers les protocoles

Prélèvements biofilms rivière : <https://dx.doi.org/10.17504/protocols.io.ben6jdhe>

Extraction ADN en utilisant le kit Nucleospin Soil (Macherey-Nagel) :
<https://dx.doi.org/10.17504/protocols.io.bd52i88e>

Préparations des PCR rbcL avec pour objectif un séquençage Miseq en 2 step PCR
<https://dx.doi.org/10.17504/protocols.io.bd94i98w>

Base de référence Diat.Barcode, version 10 : <https://entrepot.recherche.data.gouv.fr/dataverse/inrae>

Séquences obtenues au cours du projet : <https://doi.org/10.5281/zenodo.7848220>

Lien vers congrès

Vasselon et al 2021. <https://doi.org/10.3897/aca.4.e65142>

ANNEXE 1 : Analyse détaillée des différences de notes IBD

Les listes taxonomiques obtenues à partir de l'ADN ont été codées (code 4 lettres) afin d'être compatible avec OMNIDIA. Après intégration dans le logiciel, le calcul des IBD est effectué, exactement comme pour la microscopie.

Échantillons de 2021

Comparaison des IBD M et A (sans correction BIOVOL)

- Comparaison des notes IBD ADN avec OTU 95%

Les différences de notes sont données pour chacun des relevés dans le tableau suivant :

D'une manière générale, dans 73% des cas les notes IBD ont entre 0 et 2 points d'écart ; dans 50 % des cas la différence de note est inférieure à 1.

DIF note	D>10	10≤D≤2	2<D>1	D ≤1	D =0
Nb relevé	1	48	42	69	21
%	1	27	23	38	12

Différence de notes IBD entre MIC et ADN OTU 95%

- Comparaison des notes IBD ADN avec OTU 99%

DIF note	D>10	10≤D≤2	2<D>1	D ≤1	D =0
Nb relevé	1	41	43	78	17
%	1	23	24	43	9

Différence de notes IBD entre MIC et ADN OTU 99%

D'une manière générale, les résultats sont améliorés en passant de 95% à 99% ; dans ce dernier cas on a 77% (contre 73%) de notes IBD qui ont entre 0 et 2 points d'écart ; dans 53 % des cas (contre 50) la différence de note inférieure à 1.

En séparant les relevés Gaves et Nives et Adour on obtient :

- Pour Gaves et Nives

DIF note	D>10	10≤D≤2	2<D>1	D ≤1	D =0
Nb relevé	0	13	16	21	11
%	0	21	26	34	18

Différence de notes IBD entre MIC et ADN 99% - Gaves et Nives

Les résultats montrent que dans 79% des cas la différence de notes est inférieure à 2 points et dans 18 % des cas, il n'y a pas de différence.

- Pour Adour

DIF note	D>10	10≤D≤2	2<D>1	D ≤1	D =0
Nb relevé	1	28	27	57	6
%	1	24	23	48	5

Différence de notes IBD entre MIC et ADN 99% - Adour

Dans 76% des cas, la différence de notes IBD est inférieure à 2 points contre 79% pour Gaves et Nives. Dans 5% de cas, il n'y a pas de différence de notes, contre 18% pour Gaves et Nives.

La méthode ADN donne de meilleurs résultats pour les stations de Gaves et Nives.

Comparaison des notes IBD M et A (avec correction BIOVOL)

- Comparaison des notes IBD ADN avec OTU 95%

DIF note	D>10	10≤D≤2	2<D>1	D ≤1	D =0
Nb relevé	4	51	34	68	23
%	2	28	19	38	13

Différence de notes IBD entre MIC et ADN 95%

D'une manière générale, dans 69% des cas les notes IBD ont entre 0 et 2 points d'écart contre 73% sans transformation en biovolume ; dans 50 % des cas la différence de note est inférieure à 1 comme sans la transformation. On obtient davantage de différence supérieure à 10 et moins de différence égale à 0 avec les biovolumes.

- Comparaison des notes IBD ADN avec OTU 99%

DIF note	D>10	10≤D≤2	2<D>1	D ≤1	D =0
Nb relevé	1	50	35	70	25
%	1	28	19	39	14

Différence de notes IBD entre MIC et ADN 99%

Les résultats sont meilleurs que précédemment (OTU 95%) et équivalents aux résultats obtenus lorsque les inventaires ne sont pas transformés en biovolumes.

Compte tenu de ces premiers résultats, outre le fait que les corrélations soient meilleures après transformation en biovolume, on ne note pas d'amélioration au niveau des différences de notes IBD. De plus, vu le nombre important de notes IBD non pertinentes et non calculables selon la norme, la méthode des biovolumes n'est pas retenue pour l'analyse des résultats suivants.

Les notes IBD ont ensuite été transformées en classe de qualité biologique.

Comparaison des classes de qualité IBD M et ADN avec OTU 95% et 99% (sans correction BIOVOL)

Les résultats sont donnés dans les tableaux suivants :

Stations du BV Gaves de Nives

	NB_M	IBD_A_95_PAS S BIO	IBD_A_99PAS BIO
	23	13	13
	18	20	19
	21	23	23
	3	4	5
	0	1	1

Nombre de relevés dans chaque classe de qualité avec la Microscopie (M), l'ADN OTU 95% et 99% - Gaves et Nives

Stations du BV de l'Adour

	NB_M	IBD_A_95_PAS BIO	IBD_A_99PAS BIO
	25	15	15
	31	42	34
	48	39	48
	14	15	16
	0	7	4

Nombre de relevés dans chaque classe de qualité avec la Microscopie (M), l'ADN OTU 95% et 99% - Adour

Avec l'ADN, on obtient moins de qualité très bonne et davantage de qualité mauvaise comparativement à la microscopie.

Échantillons de 2022

- Comparaison des notes IBD ADN avec OTU 99%

En 2022, 77% des notes présentent une différence inférieure à 2 points et 54% à un point (tous relevés confondus). Ces résultats sont très proches de ceux obtenus en 2021, respectivement 77% et 53%.

DIF note	D>10	10≤D≤2	2<D>1	D ≤1	D =0
Nb relevé	0	43	44	85	20
%	0	22	23	44	10

Différence de notes IBD entre MIC et ADN 99%

Si on considère seulement les relevés de Gaves et Nives, on obtient 81% de notes dont la différence est inférieure à 2 points ; dans 17% de cas, il y a 0 différence.

DIF note	D>10	10≤D≤2	2<D>1	D ≤1	D =0
Nb relevé	0	12	12	29	11
%	0	19	19	45	17

Différence de notes IBD entre MIC et ADN 99% - Gaves et Nives

Si on considère seulement des relevés de l'Adour, on obtient 51% de différence de notes inférieure à 2 points ; dans 5% des cas, on obtient 0 point de différence.

DIF note	D>10	10≤D≤2	2<D>1	D ≤1	D =0
Nb relevé	0	31	32	56	9
%	0	16	17	29	5

Différence de notes IBD entre MIC et ADN 99% - Adour

Comme cela a été observé en 2021, la méthode ADN donne de meilleurs résultats au niveau des stations de Gaves et Nives.

Comparaison des classes de qualité IBD M et ADN avec OTU 99%

Les résultats sont donnés dans les tableaux suivants :

Stations du BV Gaves de Nives

Bilan 2022 Gaves et Nives	IBD_M	IBD_A
	18	14
	16	17
	21	27
	8	3
	0	1

Nombre de relevés dans chaque classe de qualité avec la Microscopie (M), l'ADN – Gaves et Nives

Stations du BV de l'Adour

Bilan 2022_ADOUR	IBD_M	IBD_A
	30	16
	34	40
	48	52
	14	15
	1	2

Nombre de relevés dans chaque classe de qualité avec la Microscopie (M), l'ADN – Adour

Comme cela a été observé en 2021, on obtient, avec l'ADN, moins de stations en très bonne qualité biologique et davantage dans la qualité mauvaise.

ANNEXE 2 : Comparaison détaillée de la composition du peuplement entre la microscopie (M) et l'ADN avec OTU 99%

Les échantillons pour lesquels on a obtenu une différence de note IBD supérieure à deux points ont tous été étudiés en détail concernant la composition taxonomique.

La composition taxonomique est comparée entre la microscopie, l'ADN avec OTU 99% sans biovolume et après des traitements spécifiques du LPL, BLAST et une nouvelle extraction. Ces deux derniers traitements ont été analysés seulement sur certaines stations, celles qui présentaient les plus grandes différence d'IBD ou un intérêt particulier.

Dans les tableaux suivants sont portées les abondances relative (en %) des taxons dominants, c'est à dire ceux qui présentent une abondance relative supérieure ou égale à 5%.

Échantillons de 2021

2021253 (Onesse - amont du R. d'Hossegor)

DIF 11.1 / 11,8/8.4	MIC 18	ADN 6.9	ADN BLAST 6.2	NV extraction 9.6
GPARG	33	0	0	1
KOBG	12	3	3	0.5
NEDT	11	0	0	0
ADMI	10	0 mais 3.8 ACHD	3	2.3
EEXI	9	0	0	0.3
GSPP	0	64	74	49
GMIN	0	14	7	1.8
FERI	1.5	11	10	15

La différence de notes est liée à la différence de détermination entre GPARG et GSPP ; GPARG est mieux noté que GSPP par l'IBD.

La nouvelle extraction permet de réduire l'écart mais le problème subsiste ; il s'agit seulement ici d'une identification à l'espèce différente entre les deux méthodes.

La lame a été reprise et des photos et mesures réalisées. pour la microscopie, c'est bien GPARG.

Notons que GMIN est détecté avec OTU 99% alors qu'avec 95% c'était resté en GOMP (inventaire pas présenté ici).

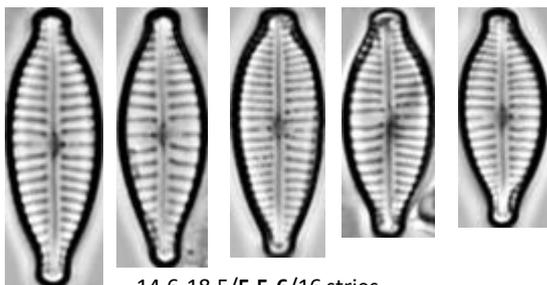
Librairie ADN = 0.198 ng/μl => TROP FAIBLE POUR AVOIR DE BONS RESULTATS => échantillon à écarter

Reads = 233

Que ce soit en Blast ou NV extraction le problème vient de GSPP/GPARG.

J'ai repris la lame, fais des photos et mesures ; Pour nous c'est bien du GPAR (largeur <6). De plus, dans le cortège, ce n'est pas très pollué, ce qui va plus dans le sens de GPAR.

2021253 Ru Hossegor Onesse



14,6-18,5/5,5-6/16 stries

2021078 (Saleys - Salies-de-Béarn)

DIF 9	MIC 12	ADN 3	ADN BLAST 3
NINC	34	0	0
APED	16	0	0
FSAP	6	0	0
RABB	15	1	0
SNIG	5	0	0
PLEV	0.2	92	92
TTAB	TFAS 1.2	31	7

Avec l'ADN, PLEV est hyper dominant ; il s'agit d'une espèce de grande taille de milieu saumâtre.

La conductivité est élevée 1400 µS/cm au niveau de ce site.

En ADN, seulement 3 taxons ont été identifiés, ce qui pose un problème de pertinence ?

Librairie ADN = 0.276 ng/µl => idem que précédemment

Reads = 891

2021279 (Rejtons - Tartas)

DIF 7.6 / 6.7	MIC 15.5	ADN 7.9	ADN BLAST 8.8
SSGE	22	26	33
KOBG	17	1.3	1.3
ADMI	16	0.5 +2.8 ACHD	2.6

SNIG	8	3.3	2.6
NPAL	4	22	0 mais NPAD 26
GOMP	0	9	10.5
PSGI	0	10	11.8

La différence de note est importante (+ de 7 points).

NPAL est surestimée en ADN d'où différence de note ; la qualité passe de vert à rouge ; on obtient beaucoup plus de NPAL (mal noté) avec l'ADN et plus d'ADMI si (bien noté) avec la microscopie ; la chimie montre une qualité mauvaise pour un certain nombre de paramètres, notamment l'ammonium, le Carbone Organique Total ...

Dans ce cas, la note donnée par l'ADN semble plus en adéquation avec la physico-chimie.

On obtient plus de KOBG et ADMI en MIC et moins de NPAL / à l'ADN : gros impact sur la notes car KOBG bien noté et NPAL mal noté par l'IBD.

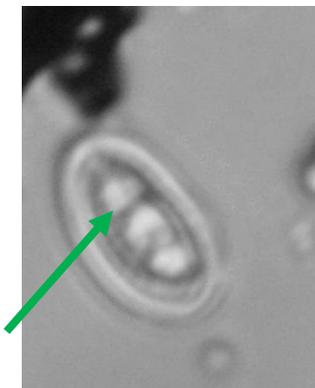
On peut se demander si les KOBG sont morts.

J'ai regardé le matériel brut entre lame et lamelle ; en effet, on voit des cellules vides de KOBG et ADMI.

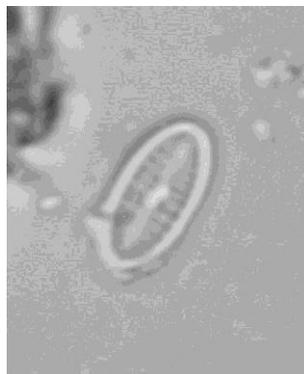
Notons que PSGI détecté avec OTU 99% alors qu'avec 95% non (résultats non présentés).

Librairie ADN = 5.21 ng/ μ l

Reads = 29623



Cellule pleine KOBG



Cellule vide KOBG



cellule vide ADMI en vue connective

Hypothèse : KOBG morts ou extraction n'a pas permis d'extraire cette espèce

2021286 Bès - Villenave

DIF 6.9/6.3	MIC 19	ADN 12.1	BLAST 12.7
KOBG	18	0.5	0
ADMI	16	0 mais ACHD 3.5	4.2
HUCO	5	0	0
MVAR	0.2	27	36
NPAL	1.5	8	0
NPAD	0	0	10

MVAR est surestimé avec l'ADN et KOBG sous-estimé par rapport à la microcopie. MVAR est une grande espèce coloniale donc plus d'ADN. KOBG morts ?

Librairie ADN = 2.53 ng/μl

Reads = 10259

Hypothèse : KOBG morts ou extraction n'a pas permis d'extraire cette espèce

2021351 Echez - Maubourquet

DIF 5.7/5.7	MIC 8.3	ADN 14	BLAST 14
GMMI	0.5	30	32
DVUL	0	16	15
NPAE	14	9	11
MVAR	0.2	9	9
NITZ	0	8	9
NAMP	30	1.5	1.2
ADMO	9	3	2.4
NFON	7	4.6	4.8
GPAR	6	0 MAIS GSPP	0 mais gspp

UULN	0.2	5	4.8
GSPP	0	5	6

GMMI est surestimé en ADN ; il s'agit d'une grande espèce coloniale.

NAMP est sous-estimé.

GPAR et GSPP toujours détecté différemment entre les deux méthodes.

Librairie ADN = 107 ng/μl

Reads = 20294

Hypothèse : NAMP morts ou extraction n'a pas permis d'extraire cette espèce ;

GPAR/GSPP => 98% de similarité entre les séquences appartenant à GPAR et celles GSPP. Diversité de séquences dans GPAR (7 séquences) vs GSPP (3 séquences) qui montrent moins de diversité

En alignement manuel, les OTU sont stricto identiques à GSPP. D'après la génétique, c'est bien GSPP.

Pareil avec BLAST

2021272 Adour - aval de Dax

DIF 5.6/6/1.7	MIC 10.4	ADN 4.8	BLAST 4.4	NV extrac 8.7
EOMI	29	0	0	0
SNIG	13	2.3	1.3	26
NAMC	7	63	72	13
MVAR	0	5	5	8
XXXX BACI	0	7	5	10 BACI

On note la surestimation de NAMC avec l'ADN. Pas d'amélioration avec BLAST mais amélioration avec la nouvelle extraction ; la différence d'IBD que de 1,7 dû à une meilleure détection en abondance de SNIG_

Librairie ADN = 59.5 ng/μl

Reads = 14356

Hypothèses surestimation : effectif plus grand dans l'échantillon que visualisé en microscopie ?? ADN plus proche de la réalité ? Espèces grandes ?

2021069 Bidouze - Bidache

DIF 5	MIC 15.5	ADN 10.5
APED	30	16
ADMO	19	5
ADMI	16	0
AMID	7	0 mais AMPH 18
XXXX	0	30 ACHD avec 95
AMPH	0	18
ACHD	0	7.5
SNIG	6.3	7
PLEV	0	6

Avec l'ADN, ACHD pas identifié en 99% resté en XXX alors qu'avec 95% c'est identifié en ACHD mais pas à l'espèce. ADMI non détecté ni AMID.

Librairie ADN = 0.508 ng/μl

Reads : 3422

Trop faible concentration ADN. Echantillon à écarter

2021361 Lis - niveau de Tarasteix

DIF 4.7	MIC 18.4	ADN 13.7
ADMI	68	0.5
MVAR	6	60
NITZ	0	9
ACHD	0	8.4

ADMI n'est pas détecté avec ADN. MVAR est surestimé par rapport à la microscopie.

Librairie ADN = 12.5 ng/μl

Reads = 23140

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI => ADMI OK en ADN aussi

2021284 Bés - Saint-Yaguen

DIF 4.6	MIC 18.7	ADN 14.1
ADMI	56	0
ADRI	20	0
ACHD	0	42
NLAN	0	13
RSIN	4	9
NTPT	0.3	6
NITZ	0	6
ADSA	0	5

ADRI et ADMI ne sont pas détectés en ADN.

Librairie ADN = 51.2 ng/μl

Reads = 26667

ADMI non détecté en LIBRAIRIE ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI => ADMI OK en ADN aussi

ADRI = 1 seule séquence dans Diatbarcode

2021316 Gabas - Arrien

DIF 4.6	MIC 18.4	ADN 13.8
ADMI	60	0.5
FGRA	7	3
FPEC	6	0 mais FRAG 4
MVAR	5	69
UULN	0	10

ADMI n'est pas détecté et ADN et MVAR est surestimé.

Librairie ADN = 125 ng/μl

Reads = 24630

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI => ADMI OK en ADN aussi

FPEC *Fragilaria pectinalis* absent de Diatbarcode

2021317 Gabas – niveau de Ger

DIF 4.4/4.4/3.2	MIC 17.6	ADN 13.2	BLAST 13.2	NV Extrac 14.4
ADMI (AASI ?)	49	0 mais ACHD	0	15
ENMI	9	3	3.2	4
GPAR	7	18	17	11
CPLI	7	5 CPLA	6 CPLA	2.5 CPLA
ACHD	0	28	20	2
MVAR	0	15	20	26
EMIN	0	8	10	3.7
XXXX	0	6	5 BACI	0.4 BACI

MVAR est surestimé en ADN.

ADMI n'est pas détecté ni en BLAST. On note une amélioration avec la nouvelle extraction : ADMI détecté ! de ce fait différence de note moins importante.

Librairie ADN = 0.187 ng/μl

Reads = 1190

Trop faible concentration ADN. Echantillon à écarter

2021295 Estampon - Nord de St-Gor

DIF 4.4/5.3	MIC 9.4	ADN 5	BLAST 4.1
GPAR	34	7	10
FSAP	27	23	24
ENMI	8	4.1	3.7
ADMI	6	0 mais ACHD 1.5	0
GSPP	0	36	45
MVAR	0	6	6
GOMP	0	5	1.2

On note encore une fois une détection différente de GPAR et GSPP.

Notons que ENMI est détecté avec 99% alors que pas détecté avec 95% (résultats non présentés). On n'observe pas d'amélioration avec BLAST.

Librairie ADN = 179 ng/μl

Reads = 19019

GPAR/GSPP => 98% de similarité entre les séquences appartenant à GPAR et celles GSPP. Diversité de séquences dans GPAR (7 séquences) vs GSPP (3 séquences) qui montrent moins de diversité

En alignement manuel, les OTU sont stricto identiques à GSPP. D'après la génétique, c'est bien GSPP.

2021081 Arrec de Bitole - Lanne en Baretous

DIF 4.4	MIC 19.9	ADN 15.5
ADMO	50	21
ADPY	25	10
NFON	7	3
MVAR	0.5	52

On note une surestimation MVAR avec l'ADN.

Librairie ADN = 17.7 ng/μl

Reads = 18843

2021107 Gave de Pau - Assat

DIF 4.3	MIC 17.5	ADN 13.2
NFON	22	19
ADPY	19	4
ADMI	14	3.3
NPAE	9	11
MVAR	0	31

Avec l'ADN, on note une surestimation de MVAR et une sous-estimation de ADMI et ADPY.

Librairie ADN = 0.583 ng/μl

Reads = 5217

Trop faible concentration ADN. Echantillon à écarter

2021262 Boudigou - la ferme d'Arman

DIF 4.3/3.7	MIC 14	ADN 9.7	ADN BLAST 10.3
ADMI	18	0	3.8
NPAL	7	3.5	0
DISC	1.5 DPSG	7	9
NUPS	2.2 NUWE	8	9
ACHD	0	7	5
PINU	2 pinu trouvé 2%	7	6
SSGE	1.2	6	6
MVAR	0	6	6
DWOL	0	4.7	5
NZSU	0.2	4	5
SELL	0	2.5	5

Avec l'ADN, pas de détection d'ADMI. DISC et NUPE sont abondants comparativement à la microscopie. Plus de DWOL avec Blast et détection d'ADMI .

LIBRAIRIE ADN = 0.71 ng/µl

Reads = 4865

Trop faible concentration ADN. Echantillon à écarter

2021105 Ruisseau Lourrou - niveau de Lucgarier

DIF 4.1	MIC 17.8	ADN 13.7
ADSH	51	1
SNIG	11	0.5
ADMI	10	0 mais ACHD 1.3
NDIS	5	0
MVAR	2.6	73
NITZ	0	11

NDIS pas détecté par l'ADN. MVAR surestimé et ADSH moins abondant mais bien détecté : cellules mortes ?

LIBRAIRIE ADN = 122 ng/μl

Reads = 19225

ACHD = à faire en méthode assignation blast

NDIS = base de données :

>ID_v10_1372__Nitzschia_dissipata_var._media

>ID_v10_1373__Nitzschia_dissipata

>ID_v10_1375__Nitzschia_dissipata_var._media

>ID_v10_1376__Nitzschia_dissipata_var._media

dissipata subs dissipata absente de diatbarcode

2021287 Bès Darengosse - niveau d'Arengosse

DIF 4.1/3.7	MIC 18.5	ADN 14.4	BLAST 14.8
KOBG	24	17	25
GPAR	6	14	17
SNIG	5	7	8.3
ADMI	47	5	9.5
ALBL	7	0	0
ACHD	0	43	32

ADMI moins abondant avec l'ADN. Amélioration avec BLAST par une meilleure détection d'ADMI.

LIBRAIRIE ADN = 38 ng/μl

Reads = 11621

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI => ADMI OK en ADN aussi

2021274 Ouzente - St-Vincent-de-Paul

DIF 3.7	MIC 15.3	ADN 11.6
ADMI	13	0
KOBG	10	7

GBOB	10	20
SNIG	5	2.8
COCO	0.2	18
ACHD	0	15
NESC	2	15

ADMI n'est pas détecté avec l'ADN et NESC surestimé (assez grande diatomée).

LIBRAIRIE ADN = 17.5 ng/μl

Reads = 16651

2021350 Adour – Estirac

DIF 3.6	MIC 12.4	ADN 16
ADMO	40	14
NCTE	15	0
NPAE	12	7
NFON	10	6
GMMI	0	37
DVUL	0	14
MVAR	1	5

Surestimation des coloniales GMMI et DVUL avec l'ADN ; il s'agit d'espèces de grande taille, vivant en colonies donc plus d'ADN.

LIBRAIRIE ADN = 66.5 ng/μl

Reads = 22478

2021251 Ruisseau de Harencin-Onesse-Laharie

DIF 3.2	MIC 12.2	ADN 9
SSGE	44	56
KOBG	17	13
NLAN	0.5	11
ADMI	12	2 mais ADCH 3.5

On rencontre les mêmes dominantes mais à des abondances différentes.

LIBRAIRIE ADN = 34.6 ng/μl

Reads = 16254

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI => ADMI OK en ADN aussi

2021063 Nive d'Arnéguy - Uhart Cize

DIF 3.1	MIC 20	ADN 16.9
ADPY	77	13
GMIN	7	0 mais GOMP1
MVAR	0.5	33
DVUL	0	27

On observe avec l'ADN une surestimation des grandes diatomées coloniales MVAR DVUL. ADPY en revanche est sous-estimé (cellules mortes ?).

LIBRAIRIE ADN = 76.9 ng/μl

Reads = 24326

GMIN/GOMP = GOMP identifié par méthode d'assignation Blast en minutum

2021335 Adour - amont de Riscle

DIF 3.1	MIC 9.9	ADN 13
ADMO	53	34
NCTE	9	0
NPAE	8	9
DVUL	0.2	21
MVAR	0.5	12

On observe encore une surestimation des coloniales MVAR et DVUL avec l'ADN.

LIBRAIRIE ADN = 92.6 ng/μl

Reads = 22406

2021062 Nive - amont du Bastan à Bidarray

DIF 3	MIC 20	ADN 17
ADPY	38	15
ADMO	28	18
NCTE	11	0
DVUL	0	20
NTPT	1.5	11
MVAR	0	10
NFON	2.7	6

Les résultats montrent une surestimation des coloniales MVAR et DVUL et des grandes navicules, NTPT.

LIBRAIRIE ADN = 34.7 ng/μl

Reads = 20651

2021114 Nivelle - aval de St-Pée

DIF 3	MIC 19.7	ADN 16.7
ADMO	50	17
ADPY	25	15
NCTE	6	0
MVAR	0.5	27
DVUL	1	19
GMMI	0.5	8

Les coloniales MVAR, DVUL et GMMI sont surestimées avec l'ADN.

LIBRAIRIE ADN = 130 ng/μl

Reads = 17174

2021259 Courant de Soustons - Magescq

DIF 3	MIC 18.8	ADN 15.8
--------------	----------	----------

KOBG	30	17
ADMI	20	0.3 ET ACHD 5
SRHE	12	0
ESLE	6	0
NLAN	1.2	23
UULN	0	14
ENCY	0	9

On note qu'avec l'ADN, ESLE, ADMI, SRHE ne sont pas détectés. La grande navicule NLAN est surestimée par rapport à la microscopie.

LIBRAIRIE ADN = 30.7 ng/μl

Reads = 17617

SRHE = absent de Diatbarcode

ESLE = 1 seule séquence dans diatbarcode ; par blast, similaire également à minutum => pas de possibilité de différencier silesiacum et minutum

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI => ADMI OK en ADN aussi

2021296 Ruisseau de Larrazieu - niveau de Labastide d'Armagnac

DIF 2.8	MIC 12.9	ADN 10.1
NITZ	0	58
MVAR	0.7	12
NDIS	14	0 mais MDME 3.5
NCRY	10	1.6

NDIS n'est pas détecté par l'ADN.

LIBRAIRIE ADN = 1.02 ng/μl

NDIS = base de données :

>ID_v10_1372__Nitzschia_dissipata_var._media

>ID_v10_1373__Nitzschia_dissipata

>ID_v10_1375__Nitzschia_dissipata_var._media

>ID_v10_1376__Nitzschia_dissipata_var._media

dissipata subs dissipata absente de diatbarcode

2021345 Arros - niveau de Bonnemazon

DIF 2.8	MIC 19.8	ADN 17
ADMO	35	30
ADPY	23	8
ADMI	9	0.8 mais achd 305
CEUG	6	7 mais CPLA
NCTE	5	0
NFON	5	9
NTPT	4.5	19

Avec l'ADN, la grande Navicule NTPT est surestimée tandis que ADMI et ADPY sont sous-estimés comparativement à la microscopie. NCTE n'est pas détecté en abondance (cellule mortes ?).

LIBRAIRIE ADN = 57.9 ng/µl

Reads = 20320

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI => ADMI OK en ADN aussi

CEUG Cocconeis euglypta absente de la base Diatbarcode

NCTE détecté mais à faible abondance (0.03%)

2021076 Bidouze - Uhart Mixe

DIF 2.7/0.6	MIC 17.5	ADN 14.8	BLAST 16.9
ADMI	48	0.3	24
ADMO	20	29	33
NCTE	9	0	0
ADSB	9	0	0

ACHD	0	29	0
CPED	1	7	8
DVUL	0	6	7
NTPT	2	5	6

Avec l'ADN, ADMI et ADSB ne sont pas détectés et laissés en ACHD (niveau du genre). NCTE n'est pas détectée. Les grandes espèces et coloniales sont surestimées : CPED et DVUL

On note une amélioration nette avec BLAST : ADMI détecté d'où la différence de note réduite (0,6 vs 2,7).

LIBRAIRIE ADN = 2.51 ng/μl

Reads = 14439

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI => ADMI OK en ADN aussi

NCTE => unclassified en ADN, identifiés en cryptotenella en méthode assignation Blast

ADSB => en blast, c'est Achnanthidium_subatomus qui sort sur certains OTU mais pas straubianum

2021067 Ardanavy - niveau de Briscous

DIF 2.6	MIC 14	ADN 11.4
SNIG	17	1
ADMI	14	0 mais ACHD 2.6
APED	13	1.3
ADRU	10	0
EOMI	9	0
AMID	5	0 mais AMPH 2
UULN	0	51
XXXX	0	12

UULN est surestimée avec ADN ; il s'agit d'une espèce de grande taille.

Les petites sont en revanche sous-estimées : SNIG, APED...

On obtient 12% de non identifié en ADN et ADRU n'est pas détecté.

LIBRAIRIE ADN = 5.92 ng/μl

Reads = 12746

ADRU = Achnantheidium druartii absent de la base Diatbarcode

EOMI = Eolimna minima absent de la base Diatbarcode

AMID = même en blast, ça reste unclassified car espèces trop proches

12% de non identifié = à voir avec la méthode d'assignation blast

2021294 Ruisseau de Tauzie - niveau de Retjons

DIF 2.6	MIC 16.2	ADN 13.6
ADMI	40	0.5
SIDE	11	0
KOBG	9	2
APED	7	1.8
MVAR	0.5	44
NLAN	1	20
ACHD	0	9

Les grandes espèces et coloniales sont surestimées : NLAN, MVAR.

ADMI et SIDE ne sont pas détectés.

LIBRAIRIE ADN = 87.4 ng/μl

Reads = 19743

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI => ADMI OK en ADN aussi

SIDE = Simonsenia delognei absent de Diatbarcode

2021277 Luzou - Bégaar

DIF 2.5	MIC 17.5	ADN 20
KOBG	23	20
ADMI	12	11
SSGE	8	0
EEXI	7	0

NPAL	6	0
SNIG	5	0
FGRA	2	48
NITZ	0	10
ENMI	0.5	6

Dans ce cas, ADMI est bien détecté en ADN mais NPAL non.

FGRA est surestimée.

LIBRAIRIE ADN = 0.137 ng/μl

Reads = 501

Trop faible concentration ADN. Echantillon à écarter

2021276 Midouze - niveau de Bégaar

DIF 2.4	MIC 14.9	ADN 12.5
CPLI	44	0 mis en CPLA
GPAR	8	0
CEUG	8	0 mis en CPLA
NGRE	6	1.7
MVAR	0.7	39
UULN	0	11
NLAN	1.5	11
NITZ	0	8
CPLA	0	7
CTPU	0.2	6

Les grandes espèces et coloniales sont surestimées par l'a méthode ADN : UULN, MVAR.

LIBRAIRIE ADN = 2 ng/μl

Reads = 17706

CPLI = identifié en CPLA en ADN

GPAR détecté en ADN mais faible abondance

2021344 Arrêt-Darré - Lhez

DIF 2.4	MIC 15.3	ADN 12.9
CEUG	53	0 mais CPLA
ADMO	16	0
ADMI	12	0
UULN	0	48
CPLA	1.7	29
CPED	0.5	16
MVAR	0.7	6

Les grandes espèces et coloniales sont surestimées : UULN, MVAR.

LIBRAIRIE ADN = 0.11 ng/μl

Reads = 334

Trop faible concentration ADN. Echantillon à écarter

2021056 Nive - Sud-Est de Cambo-les-Bains

DIF 2.4	MIC 20	ADN 17.6
ADPY	36	13
ADMO	34	12
NFON	13	8
DVUL	0	45
GMMI	0.5	14

Les grandes espèces et les espèces coloniales sont surestimées : DVUL, GMMI.

LIBRAIRIE ADN = 94.4 ng/μl

Reads = 24523

2021267 Luy de Béarn - Amou

DIF 2.3	MIC 11.8	ADN 14.1
ADMO	64	40
GMIN	6	0 mais GOMP1.5
MVAR	0.2	14
DVUL	0	12
NTPT	2.9	12

Les grandes espèces et les espèces coloniales sont surestimées : NTPT , DVUL, MVAR.

LIBRAIRIE ADN = 47.9 ng/μl

Reads = 19512

GMIN/GOMP = GOMP identifié par méthode d'assignation Blast en minutum

2021313 Gabas - niveau de Montaut

DIF 2.2	MIC 11.7	ADN 13.9
ADMO	40	13
NCTE	10	0
FSAP	9	3.8
GMIN	7	0 mais GOMP2.5
DVUL	1	51
MVAR	0	14

Les espèces coloniales sont surestimées : DVUL MVAR.

NCTE n'est pas détecté.

LIBRAIRIE ADN = 24.3 ng/μl

Reads = 20613

NCTE => unclassified en ADN, identifiés en cryptotenella en méthode assignation Blast

GMIN/GOMP = GOMP identifié par méthode d'assignation Blast en minutum

2021058 Lakako Erreka - Ossès

DIF 2.2	MIC 13.6	ADN 11.4
ADMO	43	25
NCTE	12	0
NFON	11	18
NAMP	7	1
DVUL	0.5	15
GSPP	0 mais GPAR 3.5	11
MVAR	0.5	9

Les espèces coloniales sont surestimées : DVUL MVAR. NCTE n'est pas détecté.

On rencontre toujours de même problème d'identification entre GSPP et GPAR.

LIBRAIRIE ADN = 28 ng/μl

Reads = 19947

NCTE => unclassified en ADN, identifiés en cryptotenella en méthode assignation Blast

GPAR/GSPP => Blast = la moitié des OTU sont indifférenciable entre parvulum et saprophilum. Identification correcte sur le reste. C'est bien saprophilum en majorité et quelques OTU en parvulum. GPAR détecté mais à faible abondance.

2021101 Gave de Pau - Labastide Cérézacq

DIF 2.2	MIC 12.5	ADN 14.7
ADMO	38	12
NFON	28	16
SNIG	6	0
MPMI	5	0.5
DVUL	0.2	31
MVAR	0.7	29

Les espèces coloniales sont surestimées : DVUL MVAR

LIBRAIRIE ADN = 3.96 ng/μl

Reads = 22864

SNIG détecté en ADN mais à très faible abondance

2021270 Luy de France - Amou

DIF 2.1	MIC 12.5	ADN 14.6
ADMO	51	44
CEUG	12	10 CPLA
SNIG	12	0
NITZ	0	13
MVAR	0	9

Les espèces coloniales sont surestimées : MVAR.

NITZ n'est pas identifié à l'espèce.

LIBRAIRIE ADN = 66.8 ng/μl

Reads = 21817

SNIG détecté en ADN mais à faible abondance (0.17%)

NITZ => blast ne change rien, on ne peut pas identifier à l'espèce

2021288 Midouze - Campagne

DIF 2 / 1.6	MIC 15	ADN 13	BLAST 13.4
APED	27	16	18
NCTE	16	0	0
GMIN	14	0 mais 5.6 GOMP	5
RABB	9	6	7
CEUG	6	8 CPLA	79 CPLA
SNIG	5	7	7
NTPT	0.5	11	13

MVAR	0	8	9
NLAN	0.7	7	8
GBOB	0	6	7
GPRI	0 3.4 GPUM	1.3	1.3

Les grandes espèces (NTPT, NLAN) et espèces coloniales sont surestimées : MVAR.

NCTE n'est pas détecté. GMIN est détecté avec BLAST.

GBOB/GPRI : on note une différence d'identification entre MIC et ADN pour le dominant.

LIBRAIRIE ADN = 112 ng/μl

Reads = 21651

NCTE => unclassified en ADN, identifiés en cryptotenella en méthode assignation Blast donc OK

GMIN/GOMP = GOMP identifié par méthode d'assignation Blast en minutum

CEUG Cocconeis euglypta absente de la base Diatbarcode

2021336 Arros - Tasque

DIF 2	MIC 11.3	ADN 13.3 NC
ADMO	61	65
ADLA	9	0 mais ACHD 6.5
SNIG	7	0.5
NCTE	6	0

Avec l'ADN, NCTE et ADLA ne sont pas détectés.

LIBRAIRIE ADN = 64.8 ng/μl

Reads = 17825

NCTE => unclassified en ADN, identifiés en cryptotenella en méthode assignation Blast donc OK

2021363 Souy - Lagarde

DIF 2	MIC 11.9	ADN 9.9
--------------	----------	---------

GPAR	23	0.3
GMIN	13	0 mais GOMP 3.8
NPAE	9	7
NCTE	7	0
ADMO	6	2.3
ADPY	6	3.5
GSPP	0	25
MVAR	0.5	16
GMMI	0	11
NFON	3.4	5
DVUL	0	5

Les espèces coloniales sont surestimées : MVAR GMMI. On note toujours le même problème d'identification entre GPAR et GSPP.

LIBRAIRIE ADN = 99.8 ng/μl

Reads = 12500

GPAR/GSPP/GOMP => Blast = certains OTU sont indifférenciable entre parvulum et saprophilum. Identification correcte sur le reste. C'est bien saprophilum en majorité et quelques OTU en parvulum. GPAR détecté mais à faible abondance.

Analyse détaillée après BLAST sur quelques échantillons supplémentaires

2021068 Joyeuse - Bardos

DIF 1.3/0.8	MIC 10.7	ADN 12	ADN BLAST 11.5
SNIG	24	9	9
NINC	22	0	19
MPMI	6	4.3	4.1
NCTE	5	0	0
GMIN	5	0	1.4
XXXX	0	17	0
NSTS	0	9	11
NTPT	0	7	8
NITZ	0	6	7

Avec BLAST, on note la détection de NINC et GMIN, non détectés sans BLAST.

2021264 Midour - Couloumé-Mondebat

DIF 0.6/0.7	MIC 15.5	ADN 14.9	ADN BLAST 16.2
APED	25	7	8
RABB	17	2.6	2.6
AMID	11	0 mais amph 5	0 mais amp 3.8
ADMI	6	3.1	5
GCUN	5	0	0
XXX	0	38	0
CPLA	0 mais ceug 0.2 et cpla 1	7	8
NAVI	0	6	6
RUNI	4.2	0	45

Le traitement BLAST permet de détecter RUNI.

Échantillons de 2022

2022335 Rejtons à Tartas

DIF 7,4/6.9	MIC 18.9	ADN 11.5	BLAST 12
ADMI	24	0 mais ACHD 2.3	0 mais ACHD 2.3
KOBG	23	1.2	1.2
EEXI	5.7	0	0
SSGE	4	8	8
PSGI	0	22	22
ACOP	0.5	16	16
NCRY	0	12	13
GPAR	1.7	9	9
NPAL	1	8	8
SLAC	0	6	6
GOMP	0	6	6

Comme cela a été observé en 2021, ADMI et KOBG sont sous-estimés en ADN ; la principale hypothèse est que les frustules sont vides. La même chose est observée pour EEXI.

Ci-dessous, un cellule vide d'Eunotia prise sur le brut.



LIBRAIRIE ADN = 2.14 ng/μl

Reads = 7535

ACHD => peuvent être identifiés par Blast

2022130 Saleys - Salies-de-Béarn

DIF 6.3/6.3	MIC 9.2	ADN 2.9	BLAST 2.9
NINC	19	0	0
RABB	16	0	0
TFAS	15	12 TTAB	12 TTAB
NZSU	15	0	0
APED	8	0	0
CEUG	5	0	0
PLEV	1	86	86

Il n'y a que 3 espèces d'identifiées en ADN, comme cela a été observé en 2021. PLEV, de grande taille compose l'essentiel du peuplement avec l'ADN. Le traitement BLAST me change rien.

LIBRAIRIE ADN = 47.6 ng/μl

Reads = 7535

Que 3 espèces => 82.6% d'abondance rien que sur PLEV

TFAS => absent de Diatbarcode

NZSU présent dans Diatbarcode

APED présent dans Diatbarcode

CEUG Cocconeis euglypta absente de la base Diatbarcode

2022143 Gave de Pau – aval d'Orthez

DIF 6,2/5.7	MIC 11,7	ADN 17.9	BLAST 17.4
SNIG	67	7	6.8
ADMO	7	23	23

DTEN	0	15	15
APED	2	8	9
NITZ	1.5 NSIT	8	8
CPLA	1.7 CEUG	7	6.8

SNIG est nettement sous représenté en ADN ; il s'agit d'une diatomées de petite taille. On note une très légère amélioration avec BLAST.

LIBRAIRIE ADN = 48.2 ng/μl

Reads = 14803

NITZ => blast ne change rien, on ne peut pas identifier à l'espèce

CEUG Cocconeis euglypta absente de la base Diatbarcode

2022342 Bès - Villenave

DIF 5.9/5.6	MIC 18.5	ADN 12.6	BLAST 12.9
KOBG	22	5,3	5.3
SRHE	14	0 mais SELL 1.3	0 SELL 1.3
ADMI	11	0	0
MVAR	0.5	27	28
SNIG	1.5	12	12
NAVI	0	9	9
STRS	0	8	8
NESC	0	6	7

MVAR est bien plus représenté en ADN ; il s'agit d'une grande espèce coloniale.

SRHE pas détecté en ADN mais du SELL (genre Sellaphora) est détecté : même chose ?

On note plus de SNIG en ADN / MIC alors que d'habitude c'est le contraire ; SRHE proche de SNIG ?

ADMI n'est pas détecté ni en BLAST.

LIBRAIRIE ADN = 3.89 ng/μl

Reads = 13942

SRHE absent de Diatbarcode

2022309 Palue - aval de Castets

DIF 5.5/5.3	MIC 16,1	ADN 10,6	BLAST 10.8
SNIG	18	2.4	2.4
ADSH	16	0	0
ADMI	14	0	2.4
CPLI	12	0 pas séparé de cpla	0
GANG	5	0 mais gomp 2.4	0 mais gomp 1.2
CPLA	0	18	18
UULN	0	17	17
ENCY	ENVE+ESLE	13	13
NLAN	1.5	12	12
GPAP	2.2	9	11

UULN montre une plus grande abondance en ADN ; elle n'a pas été trouvée en MIC : il s'agit d'une espèce de grande taille qu'il est possible de rater » lors de comptage.

ENCY est non détecté à l'espèce en ADN ni avec BLAST.

CPLA / CPLI ne sont pas séparés en ADN alors que CPLI est mieux noté par l'IBD.

ADMI est mieux détecté en BLAST mais pas ADSH.

LIBRAIRIE ADN = 26.4 ng/μl

Reads = 26656

ENCY absent de Diatbarcode

GANG absent de Diatbarcode

CPLA/CPLI =CPLI identifié en CPLA en ADN

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI =>
ADMI OK en ADN aussi

2022329 Ruisseau de Cabanes - St-Paul-les-Dax

DIF 5	MIC 19,4	ADN 14,4
ADMI	64	0
KOBG	14	10
PSAT	8	0
PDAU	8	0
ACHD	0	36
UULN	0	15
ACHN	0	11

UULN, grande espèce, est surestimée avec l'ADN. ADMI n'est pas détecté mais laissé au genre, ACHD ; PSAT et PDAU ne sont pas détectés.

LIBRAIRIE ADN = 27.8 ng/µl

Reads = 27143

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI =>
ADMI OK en ADN aussi

ACHD OK en blast = ADMI

PSAT absent de Diatbarcode

PDAU absent de Diatbarcode

2022429 Adour - Maubourquet

DIF 4,9	MIC pas calculé	ADN 14.7*
ADMO	83	71

UULN	0	7.6
DVUL	0	6.3
NPAE	0	6.3

IBD pas pertinent à cause de l'abondance d'ADMO. Cette dernière est bien pris en compte par les deux méthodes.

Pas pertinent

LIBRAIRIE ADN = 98.3 ng/μl

Reads = 42914

Pas de problème ADMO est détecté en ADN

2022412 Echez - Maubourguet

DIF 4,8	MIC 10.7	ADN 15.5
NAMP	32	1.2
NFON	18	3.7
ADMO	11	2.5
NPAE	7	1.2
GMMI	0.2	48
UULN	0	18
NITZ	0	7
DVUL	0.5	7

Avec l'ADN les grandes espèces et/ou coloniales sont surestimées : GGMI, UULN.

NPAE NAMP NFON détecté en ADN mais moins abondantes sont sous-estimées

LIBRAIRIE ADN = 77 ng/μl

Reads = 31035

2022428 Echez - amont des Angles

DIF 4,8	MIC 20	ADN 15.2
ADPY	64	39
GPAR	7	0
ADMI	6	2.4
GSPP	0	26
NTPT	3	12
DVUL	0.5	9

La différence de notes est liée à la détermination différente entre GPAR et GSPP.

LIBRAIRIE ADN = 28.9 ng/μl

Reads = 38624

GPAR/GSPP => saprophyllum en ADN

2022307 Onesse - amont du R. d'Hossegor

DIF 4.4	MIC 19.5	ADN 15.1
ADMI	43	0
KOBG	29	6.5
GPAR	12	0
GEXL	2	0
FERI	1.7	35
GSPP	0	33
ACHD	0	16

La différence de notes est liée à la détermination différente entre GPAR et GSPP.

ADMI n'est pas détecté.

LIBRAIRIE ADN = 2.18 ng/μl

Reads = 28379

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI =>
ADMI OK en ADN aussi

GPAR/GSPP => saprophyllum en ADN

GEXL absent de diatbarcode

ACHD OK en blast = ADMI

2022378 Gabas - niveau de Ger

DIF 4.4	MIC 17.6	ADN 13.2
ADMI	65	0 mais ACHD 3.6
MVAR	5.1	64
NLIN	0	11
UULN	0	6

MVAR, grande espèce coloniale est surestimée avec ADN tout comme les grandes espèces NLIN et UULN.

LIBRAIRIE ADN = 25.8 ng/μl

Reads = 39612

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI =>
ADMI OK en ADN aussi

2022316 Boudigau - la ferme Darman

DIF 4.4	MIC 12.4	ADN 8
NILA	12	0
ADMI	7	1.3+achd +adsa1.3
NZSU	6	34
CSNU	6	0
NPAL	6	11

DPSG	5	0
EUNO	0 mais ebil3	7.6
EMIN	0	7.6
CPLA	2.2 ceug+coco+cpli	6.3

On observe pas de détection des mêmes dominantes entre l'ADN et la MIC. EUNO et NZSU sont plus abondantes en ADN ; en MIC, NILA est abondante et n'est pas détectée en ADN.

LIBRAIRIE ADN = 8.71 ng/μl

Reads = 22820

NILA absent de diatbarcode

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI => ADMI OK en ADN aussi

CSNU, DPSG present dans Diatbarcode

CPLA/CPLI =CPLI identifié en CPLA en ADN

2022330 Ouzente - St-Vincent-de-Paul

DIF 4	MIC 13.3	ADN 9.3
ADMI	22	0 mais ACHD 6
GPRI	12	0 GBOB identifié
NESC	7	23
SNIG	6	2.6
CEUG	6	0 mais CPLA 2.6
SSGE	6	6
GBOB	1	19
NAVI	0	11
NITZ	0	6
GPAP	3	6

On a identifié GPRI en MIC mais aussi GBOB mais moins abondant (1%) alors qu'en ADN, GBOB est dominant et il n'y a pas de GPRI.

Ces deux espèces sont assez proches morphologiquement ; il n'est pas toujours facile de les distinguer en microscopie en routine.

LIBRAIRIE ADN = 57.5 ng/μl

Reads = 30380

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI => ADMI OK en ADN aussi

CEUG absent de diatbarcode

2022402 Arros - aval de Montégut

DIF 3.9	MIC 19.4	ADN 15.5
ADMO	47	37
CAFF	20	0
DVUL	1	44

On note plus de DVUL en ADN ; il s'agit d'une grande espèce coloniale.

CAFF pas détecté mort ? ou gaine gêne l'extraction ?

LIBRAIRIE ADN = 27.6 ng/μl

Reads = 21972

CAFF absent de diatbarcode

2022110 Lakako-Erreka - Ossès

DIF 3.8	MIC 10.6	ADN 14.4
NAMP	45	2.4
ADMO	25	19

NCTE	12	0
NANT	7	7
DVUL	0.2	40
GMMI	0.5	16

DVUL et GMMI, grandes espèces coloniales sont surestimées en ADN.

NAMP est sous estimé.

LIBRAIRIE ADN = 84.6 ng/μl

Reads = 29499

NCTE => unclassified en ADN, identifiés en cryptotenella en méthode assignation Blast donc OK

2022351 Ruisseau de Tauzie - niveau de Retjons

DIF 3.8	MIC 13.6	ADN 17.4
ADMI	47	0 mais ACHD 6
ADSH	10	0
APED	7	0
NLAN	5	16
SIDE	5	0
MVAR	1.2	73

MVAR est surestimée avec ADN tandis que ADMI et ADSH ne sont pas détectés.

LIBRAIRIE ADN = 33.1 ng/μl

Reads = 38622

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI => ADMI OK en ADN aussi

ADSH present dans diatbarcode

SIDE absent de diatbarcode

2022377 Gabas - Arrien

DIF 3.7	MIC 18.7	ADN 15
ADMI	75	12 + 27 ACHD
MVAR	1.7	41

MVAR est surestimée avec l'ADN.

LIBRAIRIE ADN = 19.9 ng/μl

Reads = 32598

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI => ADMI OK en ADN aussi

2022380 Adour - Bordères

DIF /0.7	MIC 9.4	ADN pas calculé	BLAST Reserve 10.1
ADMO	36	67	67
SNIG	19	0	0
CEUG	6	0	0
DCOF	6	0 mais DIAM 7.8	8
NPAL	5	2.6	0 mais 2.6 NPAD
MVAR	0.7	11	12

MVAR est surestimée avec ADN

Bonne amélioration avec BLAST ; plus que 0.7 point d'écart et IBD réservé et plus NC par la détection de DCOF notamment.

LIBRAIRIE ADN = 11.2 ng/μl

Reads = 34441

Diatbarcode = il y a une séquence Diadesmis_sp

2022150 Baysole - niveau de Lasseube

DIF 3.4	MIC 20	ADN 16.6
ADMI	17	15
ACLI	17	0 mais ACHD 1.3
AMID	13	0 mais AMPH 5
APED	12	15
ADPY	10	2.5
GSCI	1.7	32
NTPT	2.7	14
NITZ	0	6

GSCI, espèce de grande taille est surestimée avec l'ADN.

LIBRAIRIE ADN = 22.2 ng/μl

Reads = 23995

AMID présent dans la base => pas de distinction possible entre pediculus et indistincta quelque soit la méthode d'assignation

2022121 Bidouze - niveau de Bidache

DIF 3.2	MIC 12.7	ADN nc
NITZ aff NZAG	35	0
CAGR	27	0
APED	12	0
CATO	6	11
BACI		70 = mon NITZ indéterminé ?

En ADN comme en MIC la dominante n'est pas déterminée à l'espèce.

Avec ADN, CAGR n'est pas distinguée de CATO.

En BLAST NC également

LIBRAIRIE ADN = 12 ng/μl

Reads =22370

CAGR et CATO présent dans diatbarcode => pas de distinction des variants dans diatbarcode

BACI en blast => quelques OTU Nitzschia_inconspicua mais non majoritaires (abondance relative 0.048%). Les majoritaires sont non identifiés.

2022392 Lees - Baleix

DIF 3.2	MIC 14.7	ADN 11.5
NDIS	18	0
ADMI	18	0
NGER	14	0
NPAL	7	7
NITZ	0	47
NROS	0.5	22
MVAR	2.4	11

Avec l'ADN, NDIS n'est pas détecté ; NGER et NROS ne sont pas séparées et MVAR est surestimée.

LIBRAIRIE ADN = 5.97 ng/µl

Reads = 31770

NDIS = pas de distinction de NITZ entre variants dans Diatabarcode

NDIS = base de données :

>ID_v10_1372__Nitzschia_dissipata_var._media

>ID_v10_1373__Nitzschia_dissipata

>ID_v10_1375__Nitzschia_dissipata_var._media

>ID_v10_1376__Nitzschia_dissipata_var._media

dissipata subs dissipata absente de diatbarcode

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI => ADMI OK en ADN aussi

NGER absent de diatbarcode

NITZ indifférentiable en ADN

2022322 Luy de Béarn - Amou

DIF 3.1	MIC 9.9	ADN 13
SNIG	20	0
NAMP	18	2.7
ADMO	17	12
NFON	6	4.1
GMIN	5	5
UULN	0	18
CTUM	2.5	15
EPRO	0	8
MVAR	0	8
CLAN	0	5
NITZ	0	5

Avec ADN, on observe une plus grande abondance des grandes espèces, CTUM, UULN et moins grande abondance de la petite SNIG.

LIBRAIRIE ADN = 22.7 ng/μl

Reads = 35961

SNIG détecté mais à très faible abondance

NITZ indifférentiable en ADN

2022417 Arrêt - niveau de Tournay

DIF 3.1	MIC 19.4	ADN 16.3
ADMO	44	45
ADMI	21	0

ADPY	10	4
NCTE	5	0
UULN	0	17
ACHD	0	6.5
CLAN	0	6

Avec l'ADN, ADMI n'est pas détecté et UULN est surestimé.

LIBRAIRIE ADN = 70.6 ng/μl

Reads = 45558

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI => ADMI OK en ADN aussi

NCTE => unclassified en ADN, identifiés en cryptotenella en méthode assignation Blast donc OK

2022119 Ardanavy - niveau de Briscous

DIF 3	MIC 14.2	ADN 11.2
APED	14	6
EOMI	11	0
ADMI	11	0
AMID	9	0 MAIS 6 AMPH
ADRU	8	0
SNIG	7.5	6
ADSB	7	0
UULN	0.5	26
ACHD	0	14
BACI	0	11
TTAB	1.3 TFAS	6

Avec l'ADN, ADRU n'est pas détecté, UULN est surestimée.

LIBRAIRIE ADN = 29.2 ng/μl

Reads = 26527

EOMI absent de diatbarcode

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI => ADMI OK en ADN aussi ; ACHD => minutissimum et subatomus

ADSB => absent même en blast des ACHD

AMID présent dans la base => pas de distinction possible entre pediculus et indistincta quelque soit la méthode d'assignation

ADRU = absent de diatbarcode

2022383 Ruisseau de Gioulé - niveau de Lussagnet

DIF 3	MIC 11.2	ADN 14.2
NPAL	14	10
EOMI	12.7	0
NSOC	7.8	0
GSCI	2	57
NITZ	0	11

Avec l'ADN, la grande espèce GSCI est surestimée.

NSOC n'est pas détecté.

LIBRAIRIE ADN = 27.7 ng/μl

Reads = 30636

EOMI absent de diatbarcode

NSOC absent de diatbarcode

2022142 Gave de Pau - amont du Gave d'Oloron

DIF 2.9	MIC 11.2	ADN 14.1
---------	----------	----------

ADMO	43	57
SNIG	25	2.8
CEUG	9	6.9 CPLA
CPED	0	10
DVUL	0	0

SNIG est sous-estimé par ADN tandis que DVUL est surestimé.

LIBRAIRIE ADN = 21.7 ng/μl

Reads = 32727

CEUG absent de diatbarcode

2022366 Ruisseau de la Moulie - niveau de Maupas

DIF 2.9	MIC 11.2	ADN 14.1
CEUG	19	49 CPLA
HUCO	16	0
MAAT	14	0
LFRQ	7	0 mais Lven 4.8
NSIA	6	0
NAVI	0	14

HUCO et MAAT ne sont pas détectés par l'ADN.

LIBRAIRIE ADN = 24.8 ng/μl

Reads = 6999

LFRQ absent de diatbarcode

HUCO absent de diatbarcode

MAAT absent de diatbarcode (variant atomus)

NSIA absent de diatbarcode

NAVI Blast => sp et quelques Otu en symmetrica

2022411 Adour - Estirac

DIF 2.9	MIC 12.9	ADN 15.8
ADMO	49	20
NPAAE	12	6.3
NFON	9	5
GMMI	0	30
DVUL	0	20

Les grandes espèces coloniales sont surestimées avec ADN : GMMI, DVUL

LIBRAIRIE ADN = 74.9 ng/μl

Reads = 33988

2022422 Lis au niveau de niveau de Tarasteix

DIF 2.9	MIC 16.5	ADN 13.6
ADMI	36	0 mais ACHD 1.3
CPLI	15	0
MVAR	8.7	82
NDIS	5.5	0 mais 12.5 NITZ
NITZ	0	12

Pas de détection de NDIS CPLI ADMI avec l'ADN et MVAR est surestimée.

LIBRAIRIE ADN = 17.2 ng/μl

Reads = 39623

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI => ADMI OK en ADN aussi ; ACHD => minutissimum et subatomus

CPLA/CPLI =CPLI identifié en CPLA en ADN

NDIS = pas de distinction de NITZ entre variants dans Diatabarcode

NDIS = base de données :

>ID_v10_1372__Nitzschia_dissipata_var._media

>ID_v10_1373__Nitzschia_dissipata

>ID_v10_1375__Nitzschia_dissipata_var._media

>ID_v10_1376__Nitzschia_dissipata_var._media

dissipata subs dissipata absente de diatbarcode

NITZ indifférentiable en ADN

2022353 Ruisseau de Larrazieu - niveau de Labastide-d'Armagnac

DIF 2.8	MIC 12.5	ADN 9.7
UACU	10	13
ADMI	8.5	8
MING	8	0
EOMI	7	0
HCAP	6	0
ULNARIA	0	43
NFIL	1 mais NFIC	9
AUGR	0.3	6

ULNARIA n'est pas identifié en ADN et surestimé ; en MIC on a trouvé UULN et UACU.

LIBRAIRIE ADN = 8.97 ng/μl

Reads = 33509

MING absent de diatbarcode

EOMI absent de diatbarcode

ULNARIA = 9 séquences de Ulnaria_ulna et 1 sequence de Ulnaria_goulardii dans diatbarcode

NFIL/NFIC => pas de différenciation entre variants

2022304 Onesse - St-Julien-en-Born

DIF 2.8	MIC 10.7	ADN 7.9
SNIG	70	23
SEAT	7	0
SSGE	6	9
SELL	0	47

SNIG est surestimé en MIC.

LIBRAIRIE ADN = 2.17 ng/μl

Reads = 18827

SEAT absent de diatbarcode

SELL Blast = Sellafora nigri (SNIG)

2022406 Arros - niveau de Bonnemazon

DIF 2.7	MIC 17.1*	ADN 19.8 nc	BLAST NC
ADMO	69	76	77
ADPY	8	9	9
ADMI	8	0	2.5
NTPT	0	6	6

ADMI n'est pas détecté en ADN. Avec BLAST ADM I est détecté.

Notons que l'IBD est non calculable à cause d'ADMO.

LIBRAIRIE ADN = 29.4 ng/μl

Reads = 43518

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADM I =>
ADMI OK en ADN aussi ; ACHD => minutissimum et subatomus

2022158 Ruisseau Lourrou - niveau de Lucgarier

DIF 2.6	MIC 16	ADN 13.4
ADSH	24	0
ADMI	13	0 mais ACHD 2.6
SNIG	8	0
MVAR	8	51
NDIS	7	0
APED	6	5
NITZ	0	17
NLIN	0	14

Des espèces sont non détectées en ADN : NDIS , ADSH, ADMI. MVAR est surestimée. NLIN également.

LIBRAIRIE ADN = 43.2 ng/μl

Reads = 30575

NDIS = pas de distinction de NITZ entre variants dans Diatabarcode

NDIS = base de données :

>ID_v10_1372__Nitzschia_dissipata_var._media

>ID_v10_1373__Nitzschia_dissipata

>ID_v10_1375__Nitzschia_dissipata_var._media

>ID_v10_1376__Nitzschia_dissipata_var._media

dissipata subs dissipata absente de diatbarcode

ADSH present dans diatbarcode mais non détecté en DNA

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI => ADMI OK en ADN aussi ; ACHD => minutissimum

SELL Blast = Sellafora nigri (SNIG) + SNIG détecté à très faible abondance

NITZ indifférentiable en ADN

2022379 Bahus - niveau de Larrivière

DIF 2.6	MIC 12.6	ADN 15.2
NCTE	8	0
NTPT	2.2	5
ADMO	23	5
DVUL	1.9	53
MVAR	1.9	23
NITZ	0	6
GMIN	7	0
CSNU	7	1.3
NDIS	7	0
CEUG	5	0

CEUG, NDIS GMIN ne sont pas détectés par l'ADN. MVAR et DVUL , espèces coloniales sont surestimées.

LIBRAIRIE ADN = 54.5 ng/μl

Reads = 39657

NCTE => unclassified en ADN, identifiés en cryptotenella en méthode assignation Blast donc OK

NITZ indifférentiable en ADN

GMIN/GOMP = GOMP identifié par méthode d'assignation Blast en minutum

NDIS = pas de distinction de NITZ entre variants dans Diatabarcode

NDIS = base de données :

>ID_v10_1372__Nitzschia_dissipata_var._media

>ID_v10_1373__Nitzschia_dissipata

>ID_v10_1375__Nitzschia_dissipata_var._media

>ID_v10_1376__Nitzschia_dissipata_var._media

dissipata subs dissipata absente de diatbarcode

CEUG absent de diatbarcode

2022320 Grand Arrigan - Estibeaux

DIF 2.3	MIC 11.5	ADN 13.8
SNIG	16	0
NCTE	13	0
CEUG	7	0
NSIA	6	0 mais NROS 1.3
MPMI	6	0
DVUL	5	42
GMIN	5	0 mais 1.3 GOMP
MVAR	2.5	28
UULN	3	23

En ADN, on note une surestimation des coloniales MVAR DVUL et le la grande diatomée, UULN.

LIBRAIRIE ADN = 93.7 ng/μl

Reads = 41260

SELL Blast = Sellafora nigri (SNIG) ; SNIG Ok mais à très faible abondance

CEUG absent de diatbarcode

NCTE => unclassified en ADN, identifiés en cryptotenella en méthode assignation Blast donc OK

NSIA absent de diatbarcode

MPMI détecté mais à très faible abondance

GMIN/GOMP = GOMP identifié par méthode d'assignation Blast en minutum

2022129 Gaves-Réunis - Peyrehorade

DIF 2.2	MIC 15.9	ADN 13.7
ADMI	48	0 mais ACHD 11

CNLP	11	0
SNIG	10	1.3
ADEU	0.7	36
BACI	0	16
ACHD	0	11
NROS	0.5	6
NFON	1.9	6

On observe une différence d'identification entre la MIC et ADN pour ADMI/ ADEU. La lame a été reprise. La majorité pour la MIC est ADMI si avec quelques ADEU dans l'inventaire pour les formes plus rhombiques mais moins nombreuses sur la lame : peut être ADMI mortes ? ou toutes les formes sont ADEU ? et diagnose à revoir pour la largeur des cellules qui pourraient donc être moins larges que ce que dit l'auteur (3.2-4.8)?

LIBRAIRIE ADN = 32.8 ng/μl

Reads = 19019

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI => ADMI OK en ADN aussi ; ACHD => minutissimum

CNLP absent de diatbarcode

2022372 Ruisseau de Marrein - Gouts

DIF 2.2	MIC 13.6	ADN 15.8
GPRI	26	1.2
CPLI	17	1.2 CPLA
ADMI	14	0
APED	6	1.2
ADSH	5	0
NDIS	5	0
MVAR	2.7	75

NITZ

0

17

NDIS n'est pas détecté et MVAR est surestimé avec l'ADN.

LIBRAIRIE ADN = 19.5 ng/μl

Reads = 22624

CPLA/CPLI =CPLI identifié en CPLA en ADN

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI =>
ADMI OK en ADN aussi ; ACHD => minutissimum

ADSH present dans diatbarcode mais non détecté en DNA

NDIS = base de données :

>ID_v10_1372__Nitzschia_dissipata_var._media

>ID_v10_1373__Nitzschia_dissipata

>ID_v10_1375__Nitzschia_dissipata_var._media

>ID_v10_1376__Nitzschia_dissipata_var._media

dissipata subs dissipata absente de diatbarcode

NITZ indifférentiable en ADN

2022163 Ouzom - Igon

DIF 2.2	MIC 18.9	ADN 16.7
ADMO	50	50
ADPY	16	7
CEUG	14	10 CPLA
ADMI	5	0 mais ACHD 2.5
MVAR	0.5	15
CPED	0.2	6

MVAR est surestimée avec l'ADN.

LIBRAIRIE ADN = 34.9 ng/μl

Reads = 39531

CEUG absent de diatbarcode

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI =>
ADMI OK en ADN aussi ; ACHD => minutissimum

2022416 Chella - Chelle-Debat

DIF 2.1	MIC 14.1	ADN 12
ADMO	55	28
APED	8	0
ADMI	6	0
MVAR	1	43
NPAL	0.7	11
DVUL	0	5

Avec l'ADN, MVAR est surestimée tout comme NPAL. Il s'agit de diatomées de grande taille.

LIBRAIRIE ADN = 27.8 ng/μl

Reads = 40694

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI =>
ADMI OK en ADN aussi ; ACHD => minutissimum

2022104 Nive - l'amont d'Ustaritz

DIF 2	MIC 13.7	ADN 15.7
ADMO	75	42

CEUG	5	1.2 CPLA
NFON	5	5
GMMI	1.5	6
DVUL	1	40

Avec l'ADN, DVUL est surestimée.

LIBRAIRIE ADN = 21.9 ng/μl

Reads = 41698

CEUG absent de diatbarcode

2022120 Joyeuse - Bardos

DIF 2	MIC 13.5	ADN 15.5
SNIG	18	2.9
RABB	17	2.9
GMIN	17	0 mais GOMP 9
NCTE	12	0
CEUG	9	0 mais CPLA 2.9
DVUL	0	26
MVAR	0	17
GMMI	0	14
NTPT	2	13

Avec l'ADN on note la surestimation des diatomées coloniales : MVAR, GMMI, DVUL.

LIBRAIRIE ADN = 24.3 ng/μl

Reads = 32780

GMIN/GOMP = GOMP identifié par méthode d'assignation Blast en minutum

NCTE => unclassified en ADN, identifiés en cryptotenella en méthode assignation Blast donc OK

CEUG absent de diatbarcode

2022373 Gaillou - niveau de Souprosse

DIF 2	MIC 16.8	ADN 14.8
KOBG	32	21
FVUL	0	16
NDME	0.7	11
NLIN	0	8
UULN	0	8
NITZ	0	7
ACHD	0	6
ADMI	24	0
SNIG	7	0
NZSU	6	0

ADMI n'est pas détecté en ADN. FVUL n'a pas été vu en MIC.

LIBRAIRIE ADN = 3.16 ng/μl

Reads = 14095

ADMI non détecté en ADN = assignation Wang donne unclassified ; Blast permet d'assigner à ADMI => ADMI OK en ADN aussi ; ACHD => minutissimum

SELL Blast = Sellafora nigri (SNIG) ; SNIG Ok mais à très faible abondance

Sur quelques échantillons, la composition du peuplement a été analysée en détail après traitement BLAST.

2022105 Latsa - Larressore

DIF 1.1/1.1	MIC 14.4	ADN 15.5	ADN BLAST 15.5
CEUG	70	0	0
ADMO	17	25	25
CPLA	0	53	53

GOMP	0	6	0
RUNI	0.2	0	6

Avec l'ADN, on note l'amélioration du résultat après traitement BLAST par détection de RUNI.

2022126 Bidouze - Biscay

DIF 0.8/0.7	MIC 16	ADN 15.2	ADN BLAST 15.3
APED	31	28	29
ADMO	29	26	25
ADMI	11	2.7	4
NCTE	8	0	0
DVUL	0	11	11
MVAR	1.2	9	9
NTPT	0.7	8	9
GMIN	2.9	0	7

Avec l'ADN, on note l'amélioration du résultat après traitement BLAST par détection de GMIN alors que sans BLAST il est laissé au niveau du genre, en GOMP.

NCTE n'est pas détecté.

2022165 Gave de Pau - amont de Lourdes

DIF 0.2/0.7	MIC 15.2	ADN 15.4	ADN BLAST 15.9
NFON	37	38	38
ADMI	20	2.6	3.8
NPAE	7	11	11
ESLE	3.7	19	19
NPAL	2.9	10	0
ADPY	3.4	6	6.4

NPAD	0.5	0	10
------	-----	---	----

Avec BLAST, NPAD détecté plutôt que NPAL.

2022319 Midour - niveau de Couloume-Mondebat

DIF 0.1/0.4	MIC 15.6	ADN 15.7	ADN BLAST 16.1
APED	22	1	1
AMID	10	0	0
RABB	7	1	1
ADMI	6	1	1
RUNI	6	0	41
GCUN	6	0	0
NCTE	6	0	0
NDIS	6	0	0
GOMP	0	42	1
NITZ	0	18	17
GOLV (=goli)	1 GOLI	12	12

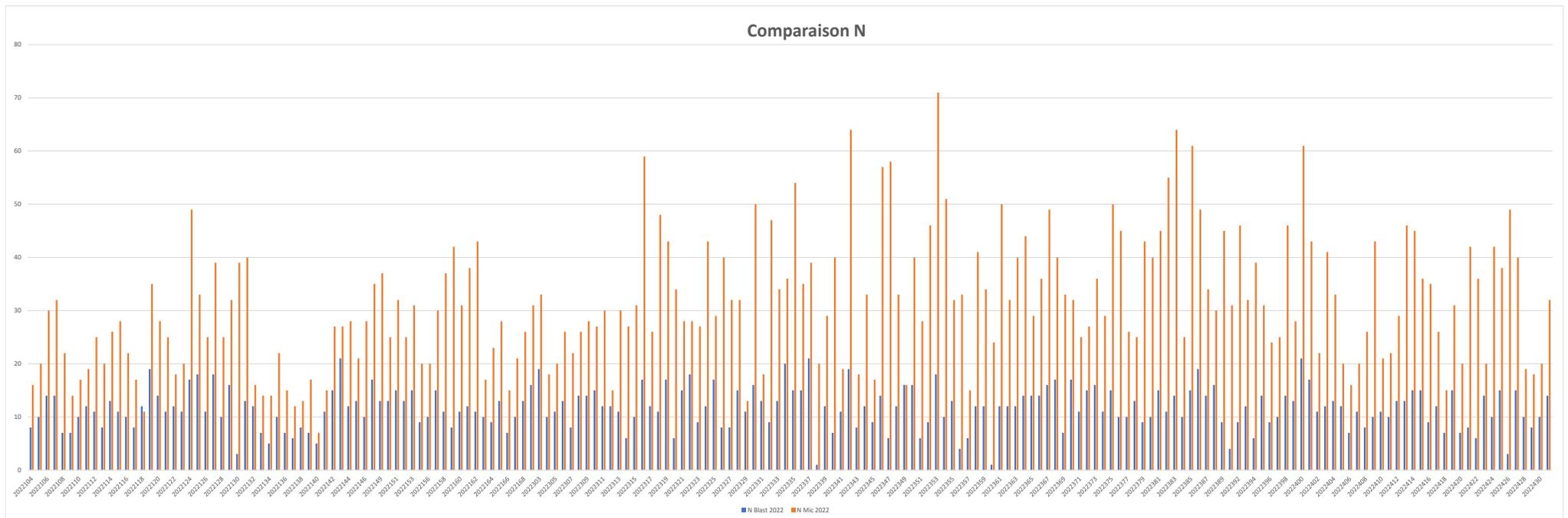
Avec BLAST, RUNI est détectée mais toujours pas NCTE ni NDIS.

Comparaison du nombre d'espèces inventoriées (N) ou richesse taxonomique

Cela a été fait pour la méthode qui a marché le mieux, c'est-à-dire ADN 99% avec BLAST en 2022 ;

Comparaison du N avec la microscopie : figure suivante. Le nombre d'espèces inventoriées est toujours nettement inférieur avec la méthode ADN. En moyenne, sur l'ensemble des relevés, on obtient 31 taxons pour la microscopie, contre seulement 12 pour l'ADN :

	MOY	MAX	MIN
ADN Blast 2022	12	21	1
MIC 2022	31	71	7



Richesse taxonomique sur l'ensemble des relevés en 2022

Titre du rapport	Rapport Final - Projet META-IBD	
Auteurs du rapport	Isabelle Vitte, Florence Peres, Valentin Vasselon, Adeline Jouanillou,,	
Période couverte par le rapport	Mai 2021	Septembre 2023
Titre du projet	Indice Biologique Diatomées : Evaluation de la mise en routine de la méthode de Metabarcoding, à l'échelle locale (sud-ouest de la France)	
Acronyme du projet	META-IBD	
Durée du projet	1 ^{er} mai 2021 – 31 septembre 2023	
Coordinateur de projet	Laboratoires des Pyrénées et des Landes	Madame Isabelle VITTE i.vitte@labopl.com
Partenaire financier	Agence de l'Eau Adour Garonne	Monsieur Jean-Pierre REBILLARD jean-pierre.rebillard@eau-adour-garonne.fr Monsieur Lomig LEBORGNE lomig.leborgne@eau-adour-garonne.fr
Partenaires scientifiques	LPL	Madame Isabelle VITTE i.vitte@labopl.com
	Scimabio Interface	Monsieur Valentin VASSELON valentin.vasselon@Scimabio Interface-interface.fr
	Artemis	Madame Florence PERES florence.peres@artemis.ovh