Rapport N° 2026-04_ODYA_01F du 16 avril 2026

Inventaire par metabarcoding ADNe des poissons, élasmobranches, crustacés et mollusques sur trois sites de l'Ile Maurice

Client : Odysseo

Oceanarium (Mauritius) Ltd

97117 Port Louis

ARGALY
RECHERCHE & DÉVELOPPEMENT EN BIOTECHNOLOGIE • Code APE : 7211Z • SIRET : 83348520400030
Adresse Postale : CLEANSPACE – 354 Voie Magellan, Alpespace – 73800 Sainte Hélène du Lac • Site Web : www.argaly.com

Résumé

Points-clés du projet :

- Nombre d'échantillons : 144
- Type d'échantillons : filtrations aquatiques (capsules Waterra)
- Groupes taxonomiques ciblés (et marqueurs associés) : Décapodes (Deca01), Élasmobranches (Elas02), Téléostéens (Fish16S), Mollusques (Moll02)
- Approche utilisée : metabarcoding ADNe

Résultats en bref :

- Taxons les plus représentés : *Paratrypaea bouvieri* (Deca01), *Mulloidichthys flavolineatus* (Fish16S), *Pateobatis hortlei* (Elas02), *Tylomelania* (Moll02)
- Nombre de taxons identifiés pour chaque marqueur :

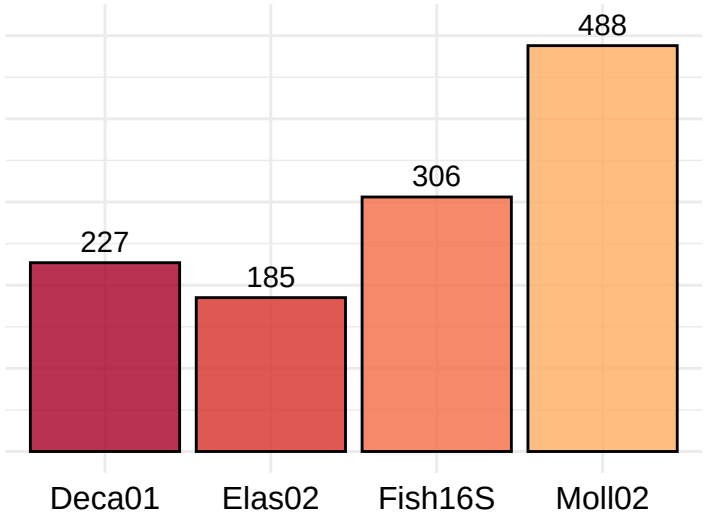


Table des matières

Argaly

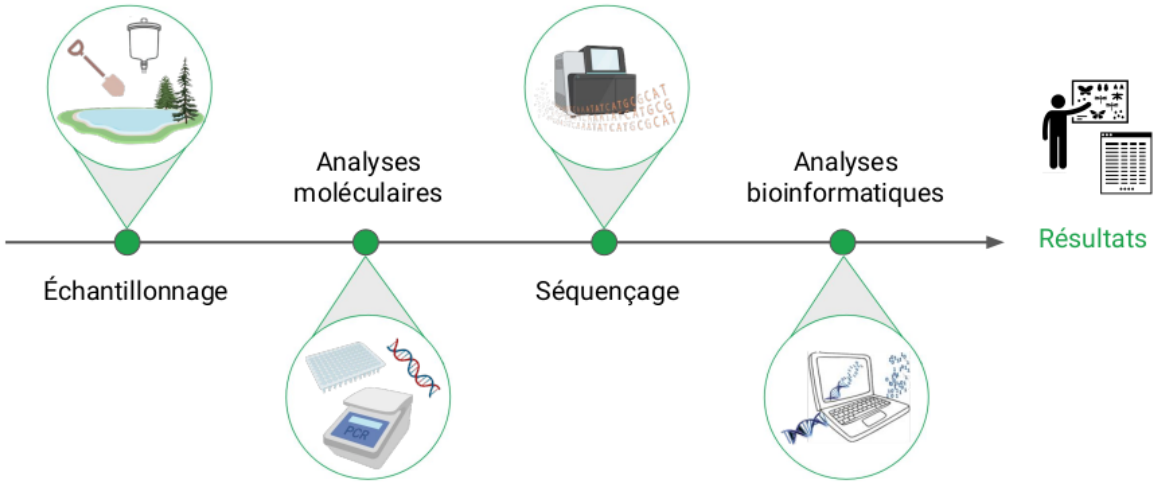
Argaly est une société spécialisée dans les analyses de biodiversité à partir d'ADN environnemental (ADNe), c'est-à-dire l'ADN extrait à partir d'échantillons prélevés dans l'environnement [1]. La mission d'Argaly est de proposer des outils innovants et non-invasifs pour obtenir des informations taxonomiques ou fonctionnelles dans les écosystèmes [1]. Dotés d'un laboratoire d'analyses et d'un pôle bioinformatique, nous dressons des inventaires et détectons des espèces cibles, afin de mieux connaître la biodiversité qui nous entoure, et ainsi mieux la préserver.

Tous les termes du glossaire
sont cliquables en les survolant

Contexte et objectifs de l'étude

Dans le cadre d'un projet de préservation et restauration de la biodiversité marine, Odysseo, premier aquarium de l'Ile Maurice, a sollicité Argaly afin de réaliser des analyses multispécifiques par l'approche ADN environnemental. Trois sites ont fait l'objet de prélèvements aquatiques (mangroves, herbiers, sables) à chaque saison pendant un an, afin d'y caractériser les communautés de poissons, élasmobranches, mollusques et crustacés. Le présent rapport détaille les méthodes utilisées et les résultats obtenus pour cette étude. La **Box 1** ci-dessous présente les étapes de l'analyse metabarcoding de l'ADNe suivie par Argaly.

Box 1 - Étapes de l'analyse metabarcoding de l'ADNe



Après l'échantillonnage de la matrice environnementale d'intérêt, l'ADNe est extrait au laboratoire grâce à un protocole adapté. Des séquences d'ADN spécifiques à un ou plusieurs taxons, nommées "metabarcodes", sont ensuite amplifiées par PCR et séquencées par séquençage haut-débit. À l'issue de plusieurs étapes de traitement bioinformatique, les séquences obtenues sont assignées à leur taxon d'origine. Les étapes effectuées chez Argaly sont détaillées dans la section **Méthodes** de ce rapport.

Résultats

Les résultats obtenus suite à l'analyse bioinformatique, avec le nombre de séquences par taxon identifié et par échantillon, sont annexés à ce rapport sous forme de tableaux Excel, attachés en **Annexes 1, 2, 3 et 4**. La correspondance des noms d'échantillons est indiquée en **Annexe 5**, et une fiche d'aide à l'interprétation des tableaux de résultats est disponible en **Annexe 6**.

Le **Tableau 1** ci-dessous indique le nombre de lectures de séquençage obtenues par séquençage NGS (données brutes au format fastq), le nombre de lectures obtenues après filtrage, ainsi que le nombre de MOTUs (Molecular Operational Taxonomic Units) et de taxons identifiés après filtrage des **Méthodes**) pour chaque marqueur et à l'échelle du jeu de données.

Suite aux filtres sur le nombre de lectures par réplicat PCR, les échantillons suivants ont été supprimés des jeux de données correspondants, faute d'une profondeur de séquençage suffisante :

- ODYA052 pour Fish16S
- ODYA020, ODYA036, ODYA052, ODYA067, ODYA082, ODYA085, ODYA110 et ODYA114 pour Deca01
- ODYA052 et ODYA082 pour Elas02
- Aucun échantillon supprimé pour Moll02

Tableau 1 – Nombre de lectures, de MOTUs ou de taxons à différentes étapes de l'analyse bioinformatique, pour chaque marqueur utilisé

	Deca01	Elas02	Fish16S	Moll02
Lectures brutes pairées avant filtrage	18600000	18600000	18600000	6600000
Lectures pairées après filtrage	5664428	9915478	9074100	4699350
MOTUs détectés après filtrage	498	210	451	1362
Taxons détectés après filtrage	227	185	306	488

Il est à noter que plusieurs MOTUs peuvent être associés au même taxon, ce qui explique que le nombre de taxons soit inférieur au nombre de MOTUs. C'est par exemple le cas de *Thalamita chaptalii*, identifié sur la base de trois MOTUs disctints dans le jeu de données Deca01 (Deca01_0006 ; Deca01_000886 ; Deca01_00985). Les taxons ne sont également pas tous situés au même rang taxonomique, et l'interprétation écologique du nombre de taxons devra donc être faite avec précaution.

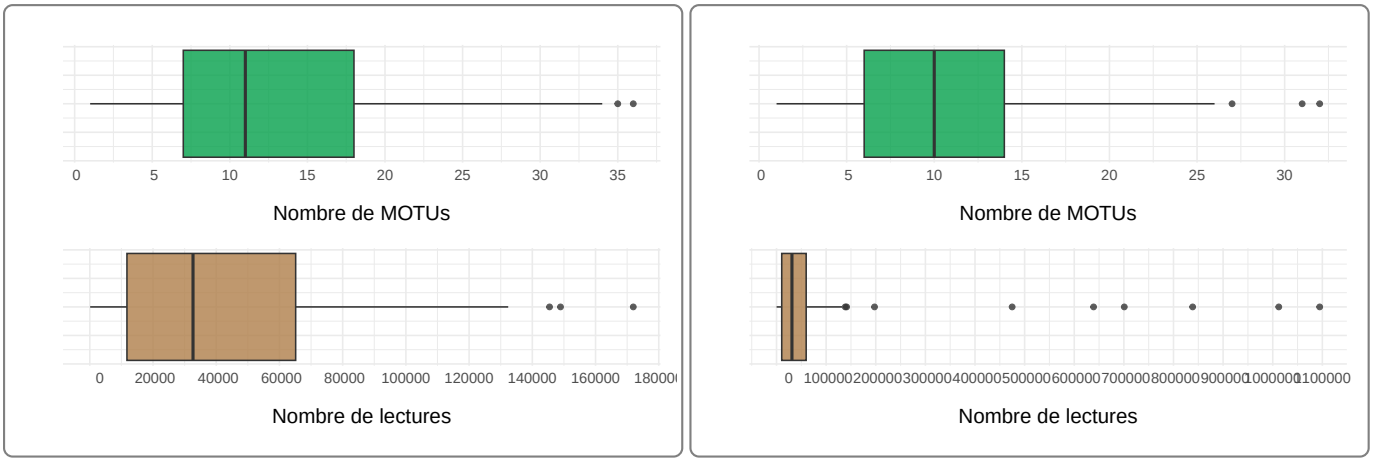
Le paramètre du best ID (score de meilleure identité, noté "Best_identity" dans les résultats) associé à chaque MOTU est notamment important à prendre en compte, puisqu'il

indique si le taxon identifié l'est avec certitude (best ID = 1), ou si le MOTU peut correspondre à une séquence ou un taxon proches absents de la base de référence (best ID < 1).

La représentativité de la base de données de référence, c'est-à-dire sa capacité à représenter de façon homogène l'ensemble des taxons cibles, a également une influence sur le nombre de taxons détectés. En effet, si la base de référence est peu fournie ou biaisée en défaveur de certains taxons, l'assignation taxonomique sera moins efficace à des niveaux taxonomiques précis et aura tendance à générer peu de taxons, et cela à des niveaux grossiers comme la classe, l'ordre ou la famille.

La **Figure 1** ci-dessous informe sur la profondeur de séquençage et le nombre de MOTUs après filtrage pour chaque échantillon, pour chaque marqueur utilisé.

(a) Deca01 **(b)** Elas02



(c) Fish16S **(d)** Moll02

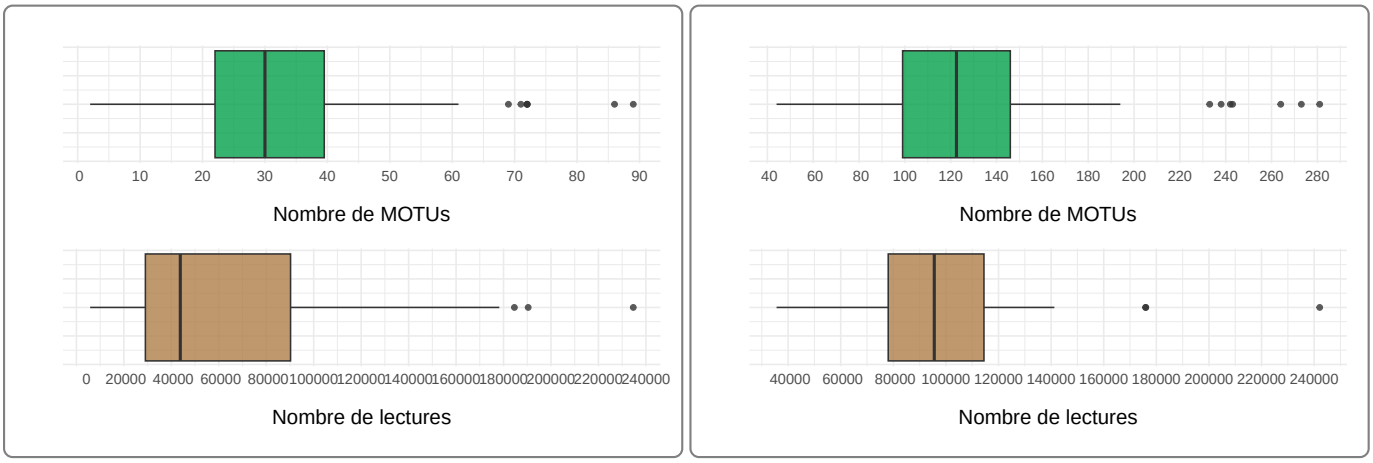


Figure 1 – Distribution du nombre de MOTUs et du nombre de lectures par échantillon pour les marqueurs Deca01 (a), Elas02 (b), Fish16S (c) et Moll02 (d)

Les profondeurs de séquençage obtenues pour chaque jeu de données sont suffisantes d'après notre expérience pour caractériser la diversité des taxons cibles à partir de prélèvements aquatiques ADNe. Les échantillons pour lesquels un nombre limité de lectures a été obtenu, voire aucune lecture après filtrage bioinformatique, ne devaient contenir que peu ou pas de traces d'ADN ciblées par les marqueurs de metabarcoding utilisés. C'est par exemple le cas de l'échantillon ODYA052, qui est systématiquement éliminé lors des étapes de filtrage pour les trois marqueurs analysés.

Le marqueur Moll02 a permis d'identifier entre 40 et 280 MOTUs par échantillon, contre 1 à 35 MOTUs maximum par échantillon pour Deca01 et Elas02. Ces observations s'expliquent par la couverture taxonomique plus ou moins restreinte des marqueurs, Elas02 amplifiant majoritairement des poissons osseux et cartilagineux à biomasse limitée, tandis que Moll02 amplifie des mollusques mais aussi des diatomées, polychètes, insectes, porifères ou encore rotifères. Même si ces taxons ne font pas initialement partie des groupes préférentiellement ciblés par le marqueur Moll02, ils peuvent présenter des séquences d'ADN suffisamment proches pour être amplifiées par les amorces de PCR.

Comme le montre la **Figure 2** ci-dessous, la grande majorité des informations taxonomiques obtenues par les analyses ADNe proviennent d'un unique marqueur, soulignant leur pertinence et leur complémentarité pour inventorier les communautés cibles. On observe tout de même une certaine redondance entre Fish16S et Elas02, qui ont détecté 89 taxons communs. Ceux-ci ne sont toutefois pas systématiquement identifiés dans les mêmes échantillons d'un marqueur à l'autre, comme *Cymolutes praetextatus*, détecté dans huit échantillons par Elas02 contre 34 par Fish16S.

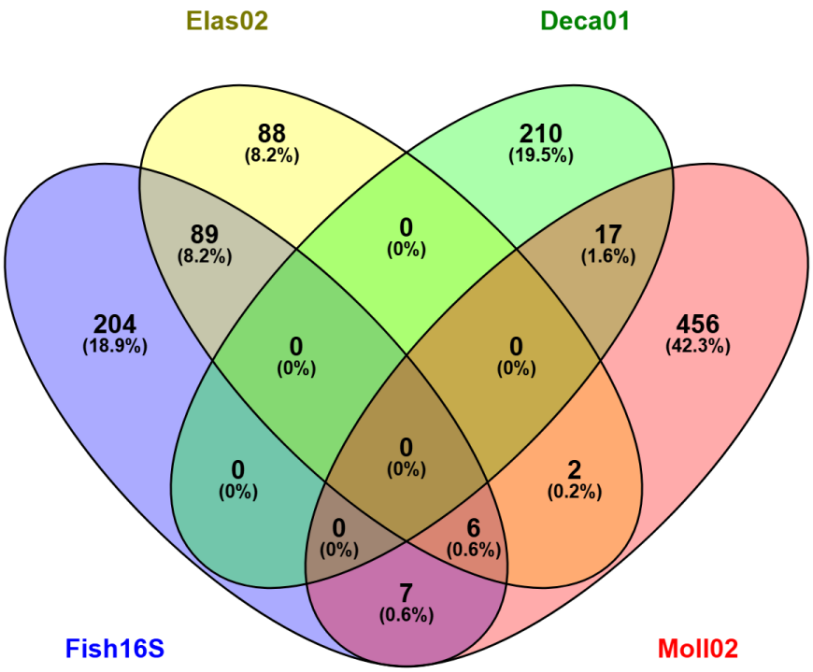


Figure 2 – Diagramme de Venn présentant le nombre de taxons identifiés par chaque marqueur de metabarcoding

Le marqueur Deca01 a permis d'identifier 498 MOTUs, dont 88% sont assignés à des crustacés décapodes, le reste étant représenté par des stomatopodes, cnidaires, insectes et quelques rares amphipodes et isopodes. Il est possible de préciser l'assignation taxonomique de certains MOTUs en comparant leur séquence *a posteriori* avec la base de référence via l'outil d'alignement en ligne BLAST. Par exemple, le MOTU Deca01_00012 peut probablement correspondre à *Corallianassa coutierei* (100% d'identité de séquence).

Le marqueur Elas02 a mené à la détection de 210 MOTUs, dont seulement huit sont assignés à des chondrichtyens. Parmi eux, une vérification avec BLAST indique plus probablement les assignations suivantes :

— Elas02_00001 : *Pateobatis fai* (99% d'homologie)
— Elas02_00012 : *Torpedo sinuspersici* ou *Torpedo marmorata* (93% d'homologie)
— Elas02_00020 : *Rhynchobatus djiddensis* (100% d'homologie)
— Elas02_00085 : *Himantura leoparda* (91% d'homologie)

Parmi les autres taxons identifiés par ce marqueur figure une majorité de poissons oisseux (86% du jeu de données) et quelques mammifères et oiseaux, ainsi que la tortue verte (*Chelonia mydas*). D'après BLAST, le MOTU Elas02_00031 peut très probablement correspondre

au cerf de Java (*Rusa timorensis*, 99% d'identité de séquence), introduit au XVIIème siècle à l'Ile Maurice pour la chasse, et le MOTU Elas02_00114 à *Numenius phaeopus* (100% d'identité). La détection d'ADNe issu d'organismes terrestres dans les prélèvements aquatiques peut s'expliquer par un échantillonnage en site côtier propice à la visite des espèces identifiées (chèvre, chien, pachyure musquée).

Un total de 451 MOTUs a été identifié par le marqueur Fish16S, dont 95% correspondent à des poissons actinoptérygiens, le reste étant représenté par des mollusques gastéropodes et des échinodermes. L'assignation de certains MOTUs peut être précisée avec BLAST :

- Fish16S_00002 : *Moolgarda crenilabis* ou *Moolgarda seheli* (100% d'homologie)
- Fish16S_00012 : *Moolgarda seheli* ou Crenimugil sp. (100% d'homologie)
- Fish16S_00014 : *Plotosus canius* (98%) ou *Plotosus lineatus* (96%)
- Fish16S_00016 : *Vanderhorstia ornatissima* (100%)
- Fish16S_00019 : *Parupeneus ciliatus* (97%)
- Fish16S_00070 : *Lethrinus obsoletus* (98%)

Le marqueur Moll02 a permis d'identifier 1362 MOTUs dont 33% sont assignés à des mollusques (gastéropodes en majorité, bivalves, céphalopodes et Polyplacophora en minorité). Les MOTUs suivants peuvent être assignés d'après BLAST à :

- Moll02_00005 : *Cerithium rostratum* (100% d'identité de séquence)
- Moll02_00019 : *Stylocheilus longicauda* (100%)
- Moll02_00081 : *Fulvia australis* (100%)
- Moll02_00094 : *Tubulophilinopsis pilsbryi* (98%)
- Moll02_00110 : *Dolabrifera dolabrifera* (100%)

Méthodes

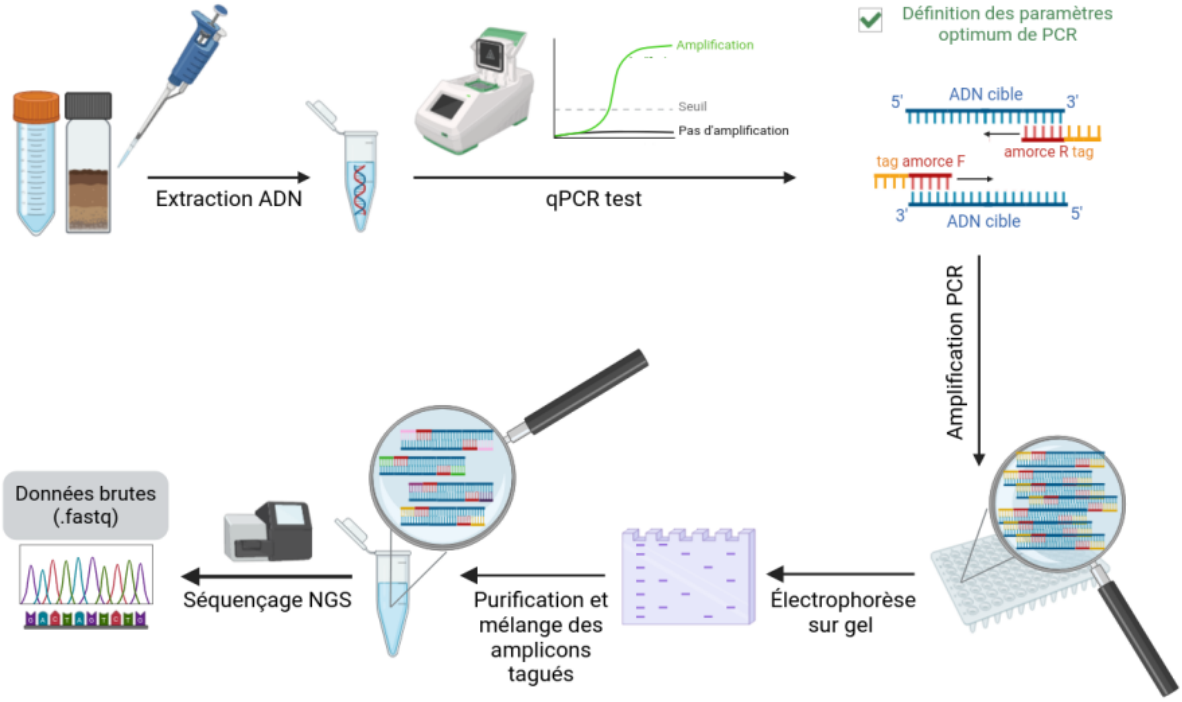
Échantillonnage

Cent quarante-quatre kits de filtration aquatique, comprenant des capsules Waterra 0.45μm 600cm2, des paires de gants et des tubes 50mL de tampon de préservation de l'ADN, ont été envoyés par Argaly à Odysseo le 19 décembre 2024. Une formation aux filtrations d'eau a été dispensée par visioconférence le 24 janvier 2025 et les prélèvements aquatiques ont été réalisés en avril, juillet puis novembre 2025 (collecte de 48 échantillons par expédition), selon le protocole décrit en **Annexe 7**. Quarante-huit échantillons d'eau ont été réceptionnés au laboratoire Argaly le 07/04/2025, puis le 10/09/2025, et enfin le 02/01/2026, pour un total de 144 échantillons. Les filtres ont été nommés "ODYA001 à ODYA144" (**Annexe 5**).

Analyses moléculaires

La **Box 2** suivante décrit les étapes de l'analyse moléculaire réalisées en laboratoire, allant de l'extraction d'ADNe au séquençage haut-débit.

Box 2 - Étapes de l'analyse moléculaire



Après réception des échantillons au laboratoire, la première étape est l'extraction d'ADN, qui consiste à isoler l'ADN contenu dans la matrice environnementale. Une PCR quantitative (qPCR) est ensuite réalisée afin de déterminer la dilution optimale des ADN à utiliser ainsi que le nombre de cycles PCR à effectuer pour l'étape d'amplification. Cette dernière permet ensuite de multiplier les fragments d'ADN de la région ciblée grâce à un couple d'amorces spécifiques. Une électrophorèse sur gel d'agarose permet de vérifier le succès de l'amplification avant la purification des amplicons obtenus. La séquence nucléotidique de chaque amplicon est enfin déterminée lors d'un run de séquençage adapté à la taille du marqueur ciblé.

Extraction d'ADN

L'ADN des 144 échantillons aquatiques a été extrait dans un laboratoire dédié à la manipulation d'ADN environnemental. Les filtres ont d'abord été agités vigoureusement pendant une minute pour resuspendre l'ADN dans le tampon de préservation [2]. Les échantillons ont ensuite été transférés en tubes puis centrifugés 15 minutes à 15 000g. L'ADN des culots a été extrait selon le protocole commercial NucleoSpin eDNA Water (Macherey Nagel). Un contrôle d'extraction a été ajouté lors de chaque session d'extraction afin d'examiner les potentielles contaminations lors de cette étape.

Amplification, purification et séquençage

Des qPCR ont été effectuées sur plusieurs extraits d'ADN sélectionnés aléatoirement, afin de déterminer la dilution optimale des ADN et le nombre de cycles PCR à effectuer pour chaque couple d'amorces (voir résultats de ces tests dans le **Tableau 2**).

Chaque ADN a ensuite été amplifié en huit réplicats PCR pour tous les marqueurs utilisés. Chaque réplicat PCR a été identifié de manière unique par une combinaison de deux tags de huit bases accolées en 5' à chaque amorce de PCR. Ces tags servent à assigner les séquences au réplicat PCR correspondant pendant l'analyse bioinformatique. Après amplification, tous les produits PCR ont été mélangés par marqueur puis purifiés avec le kit de purification MinElute (Qiagen GmbH). La construction des librairies ainsi que le séquençage haut-débit ont ensuite été effectués par la société Fasteris (Genève, Suisse). Les librairies de séquençage ont été préparées suivant le protocole Metafast destiné à limiter les artéfacts de séquençage, et ont ensuite été séquencées dans un run Illumina (**Tableau 2**).

Tableau 2 – Caractéristiques des marqueurs et du séquençage associé

Marqueur	Deca01	Elas02	Fish16S	Moll02
Référence	[3]	[1]	[4]	Bonin, pers. comm.
Dilution des ADNs extraits	1:5	1:5	1:5	1:5
Nombre de cycles PCR	42	40	48	36
Température d'hybridation	52°C	59°C	55°C	53°C
Taille médiane amplicon sans amorces (bp) (min-max)	156 (15 - 335)	182 (170 - 185)	63 (12 - 508)	73,5 bp (55 - 986)
Type de run Illumina	NovaSeq 2*250 bp	NovaSeq 2*250 bp	NovaSeq 2*250 bp	NovaSeq 2*250 bp

Contrôles qualité

Différents contrôles ont été introduits à chaque étape du protocole permettant de détecter les éventuelles contaminations pour une meilleure interprétation des résultats. Pour les

marqueurs Deca01, Elas02 et Fish16S, les contrôles réalisés pour chaque réplicat PCR sont : trois ou quatre contrôles négatifs d'extraction, trois contrôles négatifs PCR, deux contrôles positifs, et au total 128 contrôles bioinformatiques. Pour le marqueur Moll02, les contrôles réalisés pour chaque réplicat PCR sont : cinq contrôles négatifs d'extraction, trois contrôles négatifs PCR, deux contrôles positifs, et au total 64 contrôles bioinformatiques.

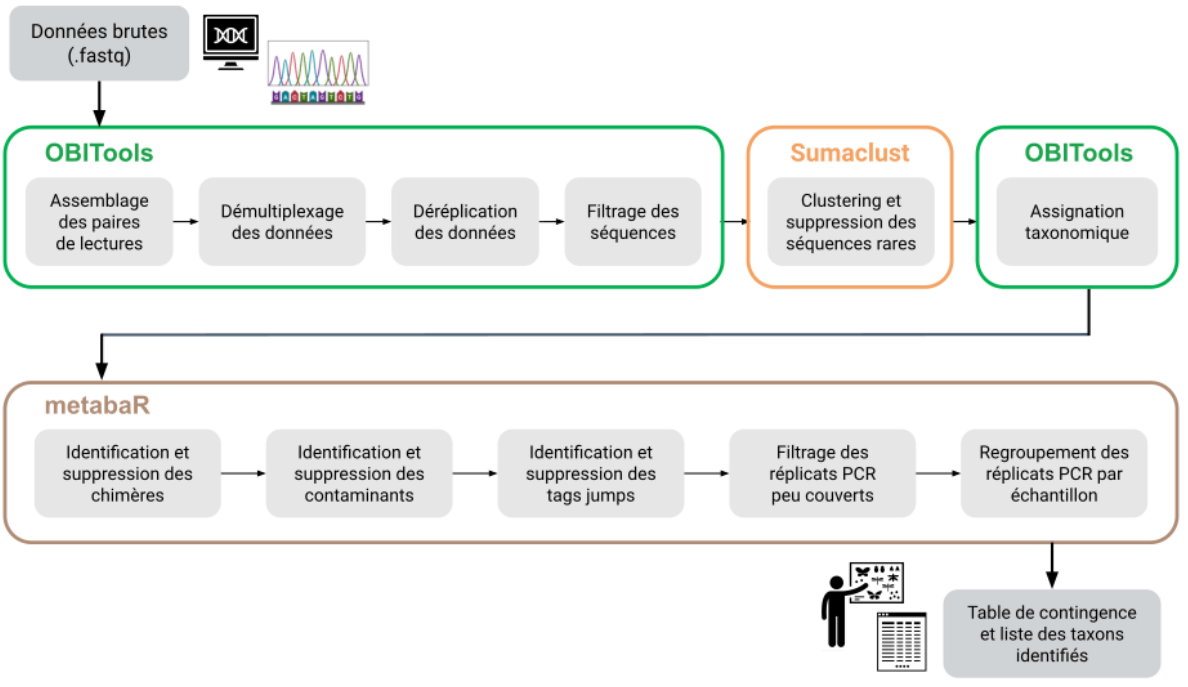
Les contrôles bioinformatiques correspondent à des combinaisons d'amorces de PCR tag-guées n'existant pas dans l'expérience mais suivies bioinformatiquement. Ils permettent d'évaluer le niveau de tag jumps [5], c'est-à-dire le niveau de séquences chimériques pro-duites au moment de la fabrication de la librairie de séquençage.

Les contrôles positifs correspondent à oligonucléotides synthétiques et à des extraits d'ADNe dont l'amplification a été positive lors de projets précédemment traités par Argaly. Le succès des amplifications et des purifications a été vérifié par électrophorèse capillaire (TapeStation 4200, Agilent Technologies).

Analyses bioinformatiques

La **Box 3** ci-dessous représente les grandes étapes du processus d'analyse bioinformatique des séquences et de leur filtrage.

Box 3 - Étapes de l'analyse bioinformatique



Analyse des séquences

Les fichiers fastq bruts contenant les lectures pairées ont d'abord été sous-échantillonnés à l'aide du programme *seqtk* à hauteur de 1,5 fois la profondeur attendue pour chaque librairie de séquençage, soit environ 9 millions de lectures pour les librairies Deca01, Elas02 et Fish16S, et environ 6 millions de lectures pour Moll02. Ce choix a été effectué en raison d'une profondeur très importante obtenue à l'issue de ce type de séquençage (NovaSeq Illumina), et permet de garantir la qualité des résultats puisqu'une profondeur trop élevée entraîne l'apparition de nombreux artéfacts comme les tag jumps, et ne contribue pas à la détection de nouveaux MOTUs.

Les données ont ensuite été analysées grâce à la suite de programmes OBITools 4 [6] et à l'outil de clustering SumaClust [7] selon la procédure séquentielle suivante (cf. **Box 3**) :

- Contrôle qualité des lectures avec Falco [8]
- Assemblage des lectures paired-end (programme *obipairing*, option –min-identity=0.8) ;
- Démultiplexage des données, soit l'assignation au réplicat PCR d'origine sur la base de la combinaison de tags en 5' des amorces de PCR (programme *obimultiplex*) ;
- Déréplication des lectures assemblées (programme *obiuniq*) ;
- Filtrage basique des lectures de mauvaise qualité (avec au moins un N dans la séquence), des lectures observées une seule fois dans le jeu de données, et des lectures dont la longueur n'appartient pas à l'intervalle de longueur observé *in silico* (programme *obigrep* ; cf. **Tableau 2**) ;
- Clustering des séquences similaires à 97% (programme SumaClust) ;
- Sélection des séquences les plus abondantes comme centres de cluster, et addition des abondances des séquences d'un même cluster (script python interne) ;
- Sélection des clusters avec au moins 10 lectures dans au moins un réplicat PCR (script python interne) ;
- Asssignation taxonomique à l'aide d'une base de référence de séquences (programme *ecotag*).

La base de référence est fabriquée à partir des séquences disponibles publiquement sur le site de GenBank release 264 à l'aide de la procédure suivante :

- PCR *in silico* avec les amorces de PCR, en autorisant trois mésappariements au maximum par amorce (programme *obipcr*) ;
- Déréplication des séquences obtenues (programme *obiuniq*) ;
- Filtrage des séquences obtenues pour ne conserver que les séquences assignées au moins au niveau de la famille (programme *obigrep*).

Certaines étapes sont accompagnées d'un schéma en **Annexe 8**.

Filtrage des séquences

Le package R metabaR [9] a été utilisé pour supprimer du jeu de données les séquences artéfactuelles présentes en faible abondance dans les données de metabarcoding, mais qui peuvent influer sur les conclusions écologiques que nous pouvons en tirer [10].

Plus spécifiquement, ont été éliminés :

- Les MOTUs trop éloignés des séquences de la base de référence (seuil best_identity $<$ 0.85 pour les marqueurs Deca01 et Moll02, 0.90 pour les marqueurs Elas02 et Fish16S), car ce sont des chimères potentielles ;
- Les MOTUs dont l'abondance est maximale dans au moins un contrôle négatif, car ce sont des contaminants potentiels (méthode "max" de la fonction *contaslayer*) ;
- Les MOTUs avec une fréquence relative $<$ 3% au sein d'un réplicat PCR car ce sont vraisemblablement des artéfacts générés au moment de la fabrication de la librairie de séquençage comme les "tag jumps" ; [5]) (fonction *tagjumpslayer*) ;
- Les réplicats PCR avec une couverture de séquençage $<$ 100 séquences.

Les réplicats PCR restants ont été agrégés par échantillon. Enfin, les MOTUs observés moins de 10 fois dans un échantillon ont été recodés comme absents de cet échantillon.

Références

[1] Taberlet P, Bonin A, Zinger L, Coissac E (2018). Environmental DNA : For Biodiversity Research and Monitoring. Oxford University PressOxford.

[2] Longmire JL, Maltbie M, Baker RJ (1997). Use of "Lysis Buffer" in DNA Isolation and Its Implication for Museum Collections, volume 163 of *Occasional Papers*. Museum of Texas Tech University, Lubbock, TX.

[3] Komai T, Gotoh RO, Sado T, Miya M (2019). Development of a New Set of PCR Primers for eDNA Metabarcoding Decapod Crustaceans. *Metabarcoding and Metagenomics*, **3**.

[4] Shaw JL, Clarke LJ, Wedderburn SD, Barnes TC, Weyrich LS, Cooper A (2016). Comparison of Environmental DNA Metabarcoding and Conventional Fish Survey Methods in a River System. *Biological Conservation*, **197**, 131–138.

[5] Schnell IB, Bohmann K, Gilbert MTP (2015). Tag Jumps Illuminated – Reducing Sequence-to-sample Misidentifications in Metabarcoding Studies. *Molecular Ecology Resources*, **15**(6), 1289–1303.

[6] Boyer F, Mercier C, Bonin A, Le Bras Y, Taberlet P, Coissac E (2016). OBITools : A Unix-inspired Software Package for DNA Metabarcoding. *Molecular Ecology Resources*, **16**(1), 176–182.

[7] Mercier C, Boyer F, Bonin A, Coissac E (2013). SUMATRA and SUMACLUST : Fast and Exact Comparison and Clustering of Sequences. *SeqBio*, 27–29.

[8] Brandine GdS, Smith AD (2021). Falco : High-Speed FastQC Emulation for Quality Control of Sequencing Data.

[9] Zinger L, Lionnet C, Benoiston AS, Donald J, Mercier C, Boyer F (2021). metabaR : An R Package for the Evaluation and Improvement of DNA Metabarcoding Data Quality. *Methods in Ecology and Evolution*, **12**(4), 586–592.

[10] Calderón-Sanou I, Münkemüller T, Boyer F, Zinger L, Thuiller W (2020). From Environmental DNA Sequences to Ecological Conclusions : How Strong Is the Influence of Methodological Choices ? *Journal of Biogeography*, **47**(1), 193–206.

Annexes

Certaines annexes sont en pièces attachées de ce rapport.

Annexe 1 - Tableau de contingence obtenu après filtrage avec le marqueur Deca01.

Annexe 2 - Tableau de contingence obtenu après filtrage avec le marqueur Elas02.

Annexe 3 - Tableau de contingence obtenu après filtrage avec le marqueur Fish16S.

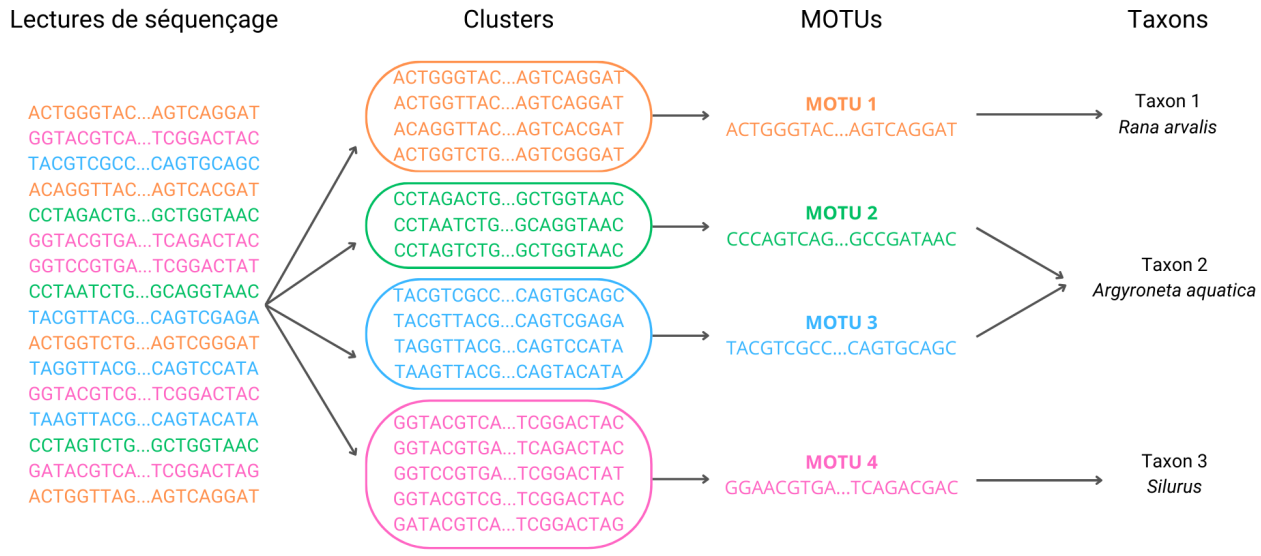
Annexe 4 - Tableau de contingence obtenu après filtrage avec le marqueur Moll02.

Annexe 5 - Correspondance des noms d'échantillons.

Annexe 6 - Tableau explicatif des résultats obtenus pour un marqueur.

Annexe 7 - Protocole de prélèvement aquatique.

Annexe 8 - Schéma simplifié du processus bioinformatique d'obtention des taxons à partir des lectures de séquençage.



Les lectures de séquençage les plus similaires sont regroupées en clusters. La lecture de séquençage la plus abondante du cluster est ensuite choisie comme MOTU (Molecular Operational Taxonomic Unit) représentatif du cluster. L'assignation taxonomique associe enfin chaque MOTU à un taxon par comparaison avec une base de référence.

Glossaire

ADN environnemental Mélange complexe de fragments d'ADN issus de différents organismes contenu dans un échantillon environnemental (eau, sol, sédiment, fécès etc.). Ces fragments peuvent être d'origine intracellulaire, s'ils proviennent de cellules intactes présentes dans le milieu, ou extracellulaire s'ils résultent de la mort d'une cellule et de la décomposition des structures cellulaires qui s'ensuit. 3, 4, 9, 10, 12, 17

Amorce de PCR Oligonucléotide d'une vingtaine de bases permettant d'amplifier un fragment d'ADN par PCR en combinaison avec une deuxième amorce de PCR. 11, 12, 13, 17

Amplicon Copie d'un fragment d'ADN obtenue par amplification PCR et encadrée par les deux amorces. 10, 17

Base de données de référence Ensemble de séquences issues de taxons connus servant à l'assignation taxonomique des MOTUs. Ces séquences peuvent être issues de bases de données publiques comme GenBank, ou provenir du séquençage génomique de specimens identifiés par un taxonomiste. 6, 13, 14, 16, 17

Clustering Action de regrouper des séquences en clusters sur la base d'un seuil minimal de similarité, le seuil de clustering. Les séquences très similaires peuvent provenir d'erreurs de PCR (ou moins probablement de séquençage) ou de polymorphismes naturels au sein du taxon d'intérêt. 13, 17

Contaminant Organisme, ADN exogène ou artéfact introduit dans l'échantillon pendant le processus d'analyse (terrain, extraction ADN, PCR etc.) et pouvant générer un faux positif. 17

Contrôle négatif d'extraction Échantillon vide extrait comme un échantillon classique, et permettant de contrôler que l'étape d'extraction n'introduit pas de contaminant. 12, 17

Contrôle négatif de terrain Échantillon permettant de contrôler que l'étape d'échantillonnage n'introduit pas de contaminant. Il peut s'agir d'une capsule de filtration utilisée pour filtrer de l'eau stérile sur le terrain, ou d'un flacon de collecte de sol laissé ouvert le temps de l'échantillonnage. 17

Contrôle négatif PCR Échantillon contenant de l'eau stérile à la place de l'extrait d'ADN et permettant de contrôler que l'étape de PCR n'introduit pas de contaminant.. 12, 17

Contrôle positif ADN ou mélange d'ADN ayant la particularité de pouvoir être amplifié avec certitude par les amorces de PCR utilisées. Il permet de s'assurer que la réaction d'amplification a bien fonctionné. Il peut provenir de tissu(s), d'une communauté de composition connue, ou il peut s'agir d'un oligonucléotide synthétique incluant les séquences cibles des amorces de PCR. 12, 17

Démultiplexage Opération bioinformatique consistant à réattribuer chaque séquence à son réplicat PCR d'origine, sur la base de la combinaison de tags accolés en 5' des amorces de PCR. 17

Fastq Format des fichiers bruts obtenus à l'issue d'un run de séquençage, contenant les lectures de séquençage et le score de qualité associé à chaque base. Ce score indique la probabilité que la base soit erronée. 5, 13, 17

Flow cell Lame de verre couverte d'oligonucléotides complémentaires aux adaptateurs de séquençage. 17

Inhibiteur Molécule interférant avec le fonctionnement optimal des processus moléculaires tels que la PCR. 17

Lecture de séquençage Séquence nucléotidique obtenue par séquençage haut-débit d'un fragment d'ADN. 5, 16, 17

Lectures pairées Couple de lectures provenant du séquençage des deux extrémités d'un même amplicon (séquençage paired-end). 5, 13, 17

Librairie de séquençage Mélange d'amplicons auquel sont ajoutés des adapteurs P3 et P5 nécessaires à la fixation sur la flow cell et au séquençage Illumina. Ces adaptateurs comportent également un tag spécifique à chaque librairie permettant de mélanger plusieurs librairies sur la même flow cell, et de réattribuer ensuite les séquences à la librairie correspondante. 11, 12, 13, 14, 17

Marqueur de metabarcoding Courte région génomique permettant de discriminer les différents taxons d'un groupe taxonomique donné, avec une résolution plus ou moins précise selon le marqueur. La discrimination des taxons se fait sur la base des séquences observées pour cette région. 8, 17

Metabarcoding Approche consistant à identifier différents taxons à partir d'un mélange complexe d'ADN, sur la base des séquences détectées pour un marqueur donné. 17

MOTU (Molecular Operational Taxonomic Unit) Séquence unique observée pour un marqueur de metabarcoding qu'il est possible d'associer à un taxon. 5, 6, 7, 8, 9, 13, 14, 16, 17

PCR (Réaction de Polymérisation en Chaîne) Réaction enzymatique permettant de produire de nombreuses copies d'un fragment d'ADN. Elle nécessite l'utilisation de deux amorces encadrant le fragment à amplifier. 4, 17

Profondeur de séquençage Nombre de lectures de séquençage observées pour un marqueur de metabarcoding. 5, 17

qPCR (PCR quantitative) PCR permettant de suivre l'amplification de l'ADN cible en temps réel par mesure d'un signal fluorescent. Dans certains cas, le nombre de copies d'ADN présentes dans l'échantillon peut être estimé. 10, 11, 17

Run de séquençage Séquençage d'une ou plusieurs librairies sur une flow cell Illumina.. 10, 17

Réplicat PCR Répétition indépendante d'une réaction de PCR à partir d'un même échantillon, permettant d'évaluer la reproductibilité de l'amplification. 5, 11, 12, 13, 14, 17

Score Best ID Valeur reflétant l'homologie de séquence entre un MOTU et la séquence la plus proche dans la base de référence. Un score de 1 représente une identité de séquence absolue, tandis que qu'un score de 0,98 correspond à une divergence des séquences de 2%. 5, 6, 17

Séquençage haut-débit ou NGS (Next-Generation Sequencing) Technologie permettant de déterminer en quelques heures la séquence nucléotidique de milliers ou de millions de molécules d'ADN simultanément. 4, 9, 11, 17

Tag Oligonucléotide de huit bases accolé en 5' d'une amorce de PCR. En combinaison avec le tag de l'autre amorce, il permet d'identifier de façon unique les amplicons amplifiés par ces amorces et de les attribuer au réplicat PCR correspondant. 11, 17

Tag jump Artéfact technique généré au moment de la fabrication d'une librairie de séquençage Illumina et produisant une chimère de deux séquences proches qui porte une nouvelle combinaison de tags. Après démultiplexage, cette chimère peut être attribuée de manière erronée à un échantillon, conduisant à un faux positif. 13, 14, 17

Taxon Groupe d'organismes partageant la même information à un rang taxonomique donné (espèce, genre, famille, etc.). Par exemple, le taxon *Drosophila* regroupe les organismes appartenant au genre "Drosophila". Il contient, entre autres, les individus appartenant aux espèces *Drosophila melanogaster* et *Drosophila obscurans*. Le taxon *Drosophila melanogaster*, quant à lui, contient tous les individus appartenant à l'espèce *Drosophila melanogaster*. 4, 5, 6, 16, 17