



HAL
open science

La reconnaissance des espèces basée sur l'ADN : applications, perspectives et défis en milieu continental terrestre

Hélène Le Borgne, Christophe Bouget

► **To cite this version:**

Hélène Le Borgne, Christophe Bouget. La reconnaissance des espèces basée sur l'ADN : applications, perspectives et défis en milieu continental terrestre. *Naturae*, 2024, 3, pp.31-67. 10.5852/naturae2024a3. hal-04651651

HAL Id: hal-04651651

<https://hal.inrae.fr/hal-04651651>

Submitted on 17 Jul 2024

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

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



Distributed under a Creative Commons Attribution 4.0 International License

La reconnaissance des espèces basée sur l'ADN : applications, perspectives et défis en milieu continental terrestre

Hélène LE BORGNE & Christophe BOUGET



DIRECTEUR DE LA PUBLICATION / *PUBLICATION DIRECTOR*: Gilles Bloch,
Président du Muséum national d'Histoire naturelle

RÉDACTEUR EN CHEF / *EDITOR-IN-CHIEF*: Jean-Philippe Siblet

ASSISTANTE DE RÉDACTION / *ASSISTANT EDITOR*: Sarah Figuet (naturae@mnhn.fr)

MISE EN PAGE / *PAGE LAYOUT*: Sarah Figuet

COMITÉ SCIENTIFIQUE / *SCIENTIFIC BOARD*:

Luc Abbadie (UPMC, Paris)
Luc Barbier (Parc naturel régional des caps et marais d'Opale, Colémbert)
Aurélien Besnard (CEFE, Montpellier)
Vincent Boulet (Expert indépendant flore/végétation, Frugières-le-Pin)
Hervé Brustel (École d'ingénieurs de Purpan, Toulouse)
Patrick De Wever (MNHN, Paris)
Thierry Dutoit (UMR CNRS IMBE, Avignon)
Éric Feunteun (MNHN, Dinard)
Romain Garrouste (MNHN, Paris)
Grégoire Gautier (DRAAF Occitanie, Toulouse)
Olivier Gilg (Réserves naturelles de France, Dijon)
Frédéric Gosselin (Irstea, Nogent-sur-Vernisson)
Patrick Haffner (PatriNat, Paris)
Frédéric Hendoux (MNHN, Paris)
Xavier Houard (OPIE, Guyancourt)
Isabelle Le Viol (MNHN, Concarneau)
Francis Meunier (Conservatoire d'espaces naturels – Hauts-de-France, Amiens)
Serge Muller (MNHN, Paris)
Francis Oliveriau (DREAL Centre, Orléans)
Laurent Poncet (PatriNat, Paris)
Nicolas Poulet (OFB, Vincennes)
Jean-Philippe Siblet (PatriNat, Paris)
Laurent Tillon (ONF, Paris)
Julien Touroult (PatriNat, Paris)

COUVERTURE / *COVER*:

Les barcodes ADN de *Chlorophorus glaucus* (Fabricius, 1781) (anciennement *pilosus* (Forster, 1771)) et *Chlorophorus glabromaculatus* (Goeze, 1777) (Coleoptera, Cerambycidae) sont très proches si bien que ces deux taxa sont considérés par certains experts comme la même espèce. Crédits photos: P. Zagatti.

Naturae est une revue en flux continu publiée par les Publications scientifiques du Muséum, Paris
Naturae is a fast track journal published by the Museum Science Press, Paris

Les Publications scientifiques du Muséum publient aussi / *The Museum Science Press also publish*:
Adansonia, *Zoosystema*, *Anthropozoologica*, *European Journal of Taxonomy*, *Geodiversitas*, *Cryptogamie* sous-sections *Algologie*, *Bryologie*, *Mycologie*, *Comptes Rendus Palevol*.

Diffusion – Publications scientifiques Muséum national d'Histoire naturelle
CP 41 – 57 rue Cuvier F-75231 Paris cedex 05 (France)
Tél.: 33 (0)1 40 79 48 05 / Fax: 33 (0)1 40 79 38 40
diff.pub@mnhn.fr / <https://sciencepress.mnhn.fr>

© Cet article est sous licence Creative Commons Attribution 4.0 International License. (<https://creativecommons.org/licenses/by/4.0/>)
ISSN (électronique / electronic) : 2553-8756

La reconnaissance des espèces basée sur l'ADN : applications, perspectives et défis en milieu continental terrestre

Hélène LE BORGNE
Christophe BOUGET

INRAE – Unité de recherche Écosystèmes forestiers,
Domaine des Barres, F-45290 Nogent-sur-Vernisson (France)
helene.le.borgne.1@gmail.com
christophe.bouget@inrae.fr

Soumis le 12 septembre 2022 | Accepté le 18 juillet 2023 | Publié le 14 février 2024

Le Borgne H. & Bouget C. 2024. — La reconnaissance des espèces basée sur l'ADN : applications, perspectives et défis en milieu continental terrestre. *Naturae* 2024 (3): 31-67. <https://doi.org/10.5852/naturae2024a3>

RÉSUMÉ

Face aux changements globaux actuels, l'enjeu des suivis de la dynamique de la biodiversité est croissant et entraîne une forte demande d'évaluations rapides et détaillées des changements de biodiversité. L'identification moléculaire des espèces est de plus en plus utilisée pour remplacer ou compléter les méthodes de surveillance écologique plus classiques. Le *metabarcoding* est considéré comme un outil d'inventaire, de connaissance de la biologie (prédateurs proies, pollinisateurs, etc.) et même de la découverte de l'histoire d'écosystèmes. Il permet de générer des données sur la biodiversité de manière rapide, précise et fiable, sur un large éventail d'organismes. Ce type de méthodologie est particulièrement intéressant pour les observatoires dépourvus de l'expertise nécessaire pour distinguer les nombreuses espèces de groupes hyper diversifiés comme les insectes ou ceux difficiles à inventorier. La reconnaissance des espèces à partir de l'ADN d'échantillons environnementaux (ADNe), tels que l'eau, les sédiments, le sol, l'air ou diverses matières organiques, a un large champ d'application. Par son caractère non invasif et non destructif, ces approches sont importantes pour l'évaluation déontologique de la biodiversité. Les chercheurs intègrent de plus en plus l'ADNe dans leurs études pour la biosurveillance en raison de sa précision et de sa facilité de déploiement. Dans ce document, nous donnons un aperçu des champs d'application des méthodes basées sur l'ADN pour le suivi de la biodiversité, des méthodes d'acquisition des données, de traitement des données pour la classification des espèces, et nous évoquons les défis inhérents à chacune de ces étapes.

MOTS CLÉS
Surveillance
de la biodiversité,
metabarcoding,
ADN environnemental,
biologie des espèces.

ABSTRACT

Biodiversity and species monitoring with species recognition based on DNA: perspectives and challenges in continental terrestrial environments.

In response to ongoing global changes, the challenge of monitoring the dynamics of biodiversity is growing and leads to a strong demand for rapid and detailed assessments of changes in biodiversity. Molecular identification of species is increasingly used to replace or complement more traditional ecological monitoring methods. Metabarcoding is considered a tool for inventorying, learning about biology (predators, prey, pollinators, etc.) and even discovering the history of ecosystems. It can generate biodiversity data quickly, accurately and reliably on a wide range of organisms. This type of

KEY WORDS
Biodiversity
monitoring,
metabarcoding,
environmental DNA,
species biology.

methodology is particularly interesting for monitoring systems lacking the expertise to distinguish the many species of hyper diverse groups such as insects or those that are difficult to survey. Species recognition based on DNA from environmental samples (eDNA), such as water, sediments, soil, air, or various biotic materials, has a wide range of applications. Because of its non-invasive and non-destructive nature, these approaches are important for the ethical assessment of biodiversity. Researchers are increasingly incorporating eDNA into their biomonitoring studies because of its accuracy and ease of deployment. In this paper, we provide an overview of the scopes of DNA-based methods for biodiversity monitoring, data acquisition methods, data processing for species classification, and discuss the challenges inherent in each of these steps.

ACQUISITION DES DONNÉES MOLÉCULAIRES POUR LES MÉTHODES GÉNÉTIQUES

BARCODING D'UN SPÉCIMEN

L'acide désoxyribonucléique, ou ADN contient des informations génétiques spécifiques à chaque espèce et certains fragments d'ADN sont très conservés au sein d'une même espèce, mais variables entre les espèces. De courts fragments d'ADN composés d'une région hypervariable et appelés code-barres ADN ou marqueurs génétiques, sont utilisés pour discriminer les espèces. Le terme *barcoding* fait ainsi référence à l'identification taxonomique des espèces basée sur le séquençage d'un seul échantillon. Les régions conservées du code-barres ADN permettent de définir des amorces qui serviront à amplifier ces gènes chez la plupart des organismes. Les marqueurs pour le *barcoding* vont être soit des gènes codant des protéines (e.g., Cytochrome c Oxydase I [COI] ou *rbcL*), soit des gènes ribosomiques (e.g., 18S, 16S). Bien que les gènes codant pour des protéines soient généralement uniformes – le code génétique étant redondant – certaines bases vont varier selon les espèces et permettre ainsi une bonne identification des espèces. En couplant le séquençage des spécimens avec une base de données associant des séquences génétiques de référence à des noms de taxa, on est capable de faire une identification fiable et simple d'un spécimen. Le *barcoding* sert également à enrichir ses bases de données de références en se basant sur des spécimens identifiés morphologiquement.

Le *barcoding* peut donc être défini comme le séquençage d'un marqueur génétique spécifique et servant de code-barres normalisé (e.g., région du gène COI du génome mitochondrial pour les animaux) qui est caractérisé par une forte spécificité au sein d'une espèce et une grande variabilité entre les espèces. En effet, l'analyse des divergences de séquence d'ADN mitochondrial permet de capturer les discontinuités biologiques reconnues par les taxonomistes comme des espèces car elles sont ordinairement plus importantes entre les espèces qu'au sein des espèces. Dans le cas des espèces animales, le marqueur génétique choisi comme code-barres ADN standard pour leur discrimination est un fragment d'environ 650 paires de bases du gène mitochondrial codant la COI. Le marqueur COI est sélectionné pour identifier les espèces animales – avec pour certains groupes un complément avec le 16S – en raison de ses petites différences intra-spécifiques et de ses grandes diffé-

rences inter-spécifiques (Hebert *et al.* 2003, 2004 ; Stoeckle & Thaler 2018). Il a été démontré que le COI permet de faire des identifications au niveau des espèces pour de nombreux groupes taxonomiques d'animaux comme les insectes, les oiseaux, les poissons ou encore les Crustacés. Par exemple, une étude a démontré que les codes-barres COI étaient efficaces pour différencier les espèces d'oiseaux nord-américains, tout en identifiant quatre nouvelles espèces probables (Hebert *et al.* 2004). Une autre étude suggère que jusqu'à 28 % des espèces de rhopalocères en Europe pourraient être des espèces cryptiques (i.e. morphologiquement très similaires) et donc totalement nouvelles pour la science, car elles représentent des lignées génétiques distinctes (Dincă *et al.* 2015). Chez les plantes, trois régions du génome du chloroplaste (*rbcL*, *matK* et *trnH-psbA*) ainsi que la région ITS ribosomale nucléaire (ITS1 et ITS2) ont été utilisées comme codes-barres ADN, alors que chez les algues vertes c'est souvent la région *tufA* qui l'est (Kress *et al.* 2015). Le code-barres ADN est également un outil puissant pour l'identification des échantillons de champignons avec les régions ITS1 et ITS2 comme marqueur (Xu 2016).

Les codes-barres ADN sont de plus en plus utilisés dans les études de la biodiversité car ils permettent notamment la discrimination d'un large éventail d'espèces (e.g., oiseaux, chauves-souris, mammifères, divers invertébrés, etc.), l'identification de nouveaux taxa, l'identification de juvéniles (voire d'œufs), ou encore la détection d'espèces cryptiques (Rougerie *et al.* 2009). Par exemple, le *barcoding* a été utilisé dans une étude sur les vers de terre et a aidé à la discrimination des espèces, à l'identification des spécimens fragmentés et des juvéniles qui étaient non identifiables en regardant uniquement la morphologie (Decaens *et al.* 2013). Les grains de pollen présentent des morphologies souvent spécifiques à l'espèce, mais il reste parfois difficile de discriminer les espèces étroitement apparentées morphologiquement. Par conséquent l'utilisation d'outils moléculaires comme les code-barres ADN pour identifier de tels organismes semble avoir un fort potentiel et a été appliqué avec succès à divers groupes taxonomiques, incluant les plantes et leur pollen (Hebert *et al.* 2003 ; Chen *et al.* 2010). Ils permettent également de très grandes avancées dans la connaissance de la biologie des espèces, comme par exemple le régime alimentaire des chauves-souris par *metabarcoding* des fèces (Ingala *et al.* 2021). La méthodologie innovante du

code-barres ADN pourrait aider concernant les problèmes de perte d'expertise taxonomique, car une fois les séquences de référence établies avec l'appui des experts taxonomistes, la technique est autonome. Bien que l'utilisation de codes-barres ADN réduira vraisemblablement la tâche d'identification standard des taxonomistes, ils continueront à jouer un rôle essentiel dans l'analyse des complexes d'espèces soulevés par la génétique, dans la description de la grande partie inconnue de la biodiversité et tout en bénéficiant d'un outil efficace de révision (Rougerie *et al.* 2009). De plus, les limites liées à cette méthodologie telles que la connaissance du stade de développement, l'abondance absolue, la détectabilité pas toujours évidente ou encore les pollutions par d'autres ADN, ne permettent pas encore de se passer d'analyse. Le « code-barres ADN » dépend de l'expertise taxonomique sans laquelle il n'est pas possible d'avoir une base de référence correcte. Cela plaide pour un maintien d'une recherche sur la taxonomie en lien étroit avec le développement des techniques de suivis de la biodiversité par ADN (notamment pour l'ADN environnemental, voir partie « ADN environnemental »).

Actuellement, les codes-barres ADN sont couramment utilisés pour identifier des individus et l'approche a des applications dans la conservation de la biodiversité, la gestion de l'environnement, la biologie des invasions, ou encore l'étude des interactions trophiques. Cette méthodologie est notamment importante pour associer une séquence aux spécimens de référence des collections d'histoire naturelle et le dépôt de leurs séquences uniques dans les bases de données de référence, qui pourront être ensuite utilisés pour attribuer des séquences inconnues à des espèces connues. Les informations taxonomiques associées aux codes-barres standard permettent ainsi d'identifier les regroupements de séquences de *metabarcoding* (i.e. MOTUs – *Molecular Operational Taxonomic Unit* – ou OTUs pour unités taxonomiques opérationnelles moléculaires, voir partie « *Metabarcoding* d'une communauté »). L'approche des code-barres ADN fait des liens entre les écologistes moléculaires et les taxonomistes morphologiques en générant des bases de données de référence basées sur des spécimens vérifiés et conservés (Cristescu 2014).

Les avantages de cette méthode sont nombreux et incluent le fait qu'elle :

- est reproductible (vérifiable) ;
- peut parfois être plus rapide et moins coûteuse que les identifications morphologiques traditionnelles ;
- est accessible tant que des installations de séquençage sont accessibles ;
- peut être utilisée pour les plus petits animaux ;
- fonctionne pour tous les stades de la vie et tous les types de tissus organiques (Rougerie *et al.* 2009 ; Decaens *et al.* 2013 ; Kress *et al.* 2015) ;
- permet d'utiliser les séquences non assignées dans les études écologiques.

Les codes-barres ADN peuvent représenter une ressource cruciale pour discriminer les groupes ayant peu d'expertise taxonomique disponible. Par ailleurs la plupart des étapes de cette méthodologie peuvent être sous-traitées à des laboratoires commerciaux si les chercheurs ou les naturalistes n'ont

ni les compétences techniques, ni les équipements nécessaires pour effectuer des analyses génétiques. Un des inconvénients couramment notés est que l'utilisation d'un seul marqueur mitochondrial de niveau de divergence arbitraire peut occasionner des faux négatifs (e.g., codes-barres ADN identiques dans deux espèces différentes) et des faux positifs (e.g., différents codes-barres ADN entre individus appartenant à la même espèce). Une autre limite est l'impossibilité d'identifier les hybrides (lignée maternelle uniquement). Chez les champignons, une proportion significative de molécules recombinantes a été détectée lors des PCR (Zhou *et al.* 2019). C'est pourquoi il est conseillé d'utiliser les codes-barres ADN avec prudence, en associant plusieurs marqueurs génétiques et en les combinant avec d'autres données telles que la morphologie ou des caractéristiques écologiques, éthologiques et biogéographiques (Rougerie *et al.* 2009).

Ces méthodologies basées sur l'ADN sont intrinsèquement différentes des approches classiques d'identification des espèces. Tout d'abord, elles requièrent des méthodologies d'échantillonnage préservant l'ADN (voir partie « Méthodologies de conservation et archivage des échantillons ») et empêchant les pollutions extérieures. Il y a également plusieurs étapes de laboratoire (i.e. extraction de l'ADN, amplification par PCR et séquençage ; voir partie « Traitement des échantillons en laboratoire »), qui vont nécessiter des connaissances techniques et un choix éclairé à chaque étape pour une détection fiable des espèces (Fig. 1). Finalement, l'interprétation des données obtenues après séquençage requiert certaines compétences spécifiques en bio-informatique (voir partie « Analyse et outils bio-informatiques ») (Liu *et al.* 2020).

METABARCODING D'UNE COMMUNAUTÉ

Récemment, les améliorations techniques et les réductions de prix du séquençage à haut débit ont permis l'extension du *barcoding* à l'identification d'ADN que l'on peut qualifier de multi-spécifique. Le *metabarcoding* élargit ainsi l'identification basée sur l'ADN des spécimens d'espèces aux communautés d'individus appartenant à plusieurs espèces (Ruppert *et al.* 2019 ; Compson *et al.* 2020). Cette méthode d'identification multi-spécifique implique un séquençage parallèle massif d'échantillons d'ADN issu d'un échantillon dit mixte ou composite (i.e. mélange d'organismes plus ou moins entiers collectés sur le terrain ; Annexe 1) ou d'ADN environnemental (voir partie « ADN environnemental »), en utilisant des amorces d'amplification plus ou moins génériques. Le *metabarcoding* peut soit cibler des groupes spécifiques d'organismes (i.e. amorces ciblées), soit cibler plusieurs groupes taxonomiques (i.e. amorces générales ou universelles). Le rythme rapide des progrès et la baisse des coûts laissent envisager que le *metabarcoding* devienne un outil standard pour les écologistes dans le futur (Poland & Rife 2012 ; Deiner *et al.* 2017).

Pour utiliser cette méthode, il faut vérifier qu'il est possible de définir des amorces spécifiques pour le gène choisi, et que la taille du fragment à amplifier (i.e. amplicon) est compatible avec la longueur des séquences obtenue avec la technologie de séquençage choisie (Falentin *et al.* 2019). Le choix du marqueur est essentiel et se décline en fonction des groupes

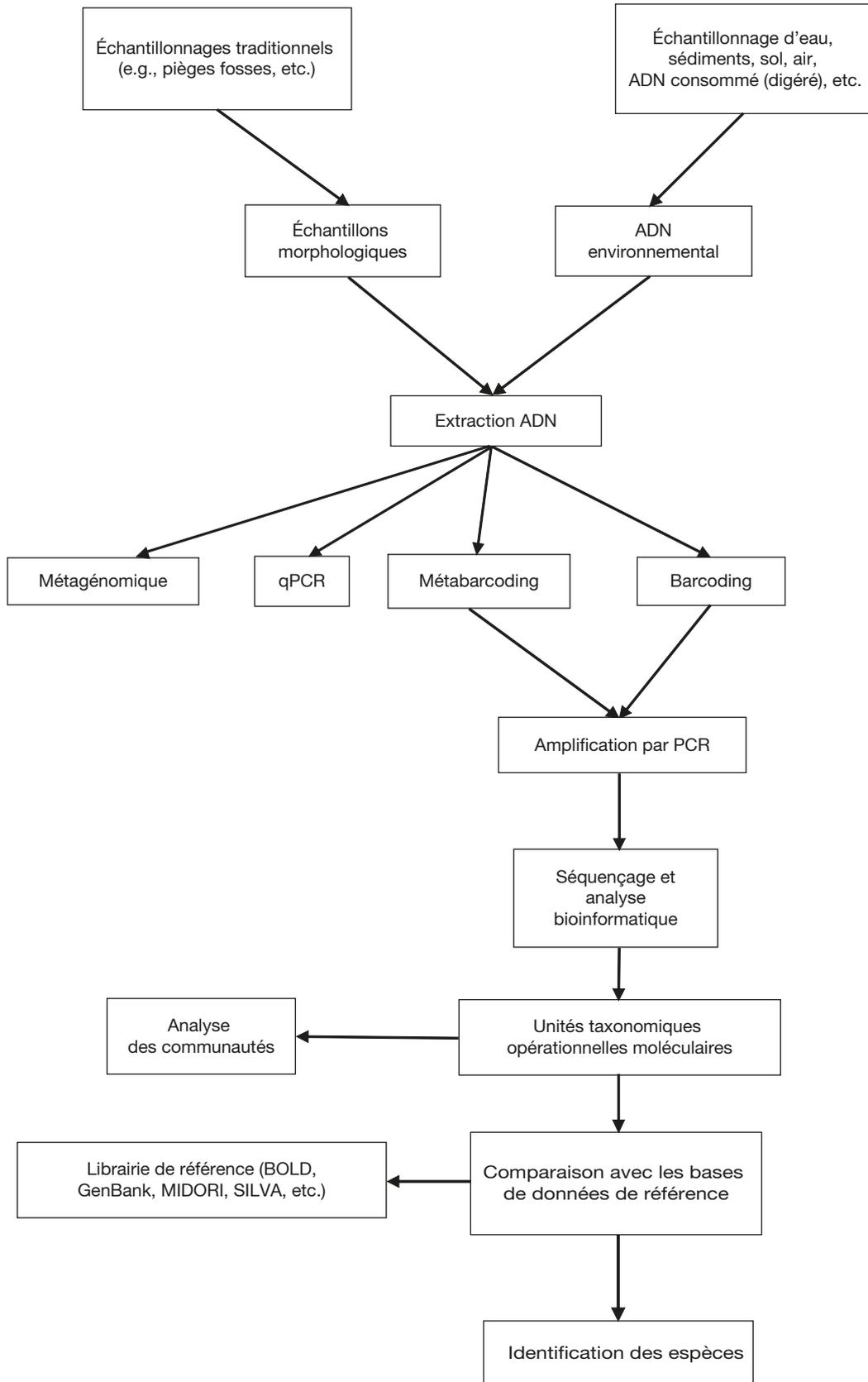


FIG. 1. — Schéma opérationnel sur l'identification des espèces à partir d'échantillons d'ADN, de l'acquisition de données en milieu terrestre (i.e. méthodologies) à la manière dont les données sont ensuite traitées pour la classification des espèces (i.e. gestion des données et analyses). Abréviations : **PCR**, polymérase chain reaction ou amplification en chaîne par polymérase; **qPCR**, PCR quantitative.

d'espèces étudiées (voir partie « *Barcoding* d'un spécimen »). L'ADN peut être dégradé et l'amplification de ces fragments d'ADN plus courts doit être prise en compte pour l'utilisation de codes-barres normalisés. Le fragment amplifié doit avoir une bonne résolution taxonomique et être discriminant, idéalement au niveau de l'espèce. L'efficacité de l'attribution taxonomique va dépendre du degré de complétude de la base de référence contenant les codes-barres, ce qui représente une des grandes difficultés actuelles liées à l'interprétation des données de *metabarcoding*. Lors du choix du code-barres, il est donc important de prendre en compte la disponibilité d'une base de données de référence taxonomique complète et qu'elle soit basée sur plusieurs spécimens morphologiquement vérifiés et conservés pour couvrir la variabilité génétique de chaque espèce. Les estimations de la biodiversité de ces études vont donc dépendre de la résolution du marqueur utilisé, de la qualité des séquences, et finalement des pipelines bio-informatiques utilisés (voir aussi partie « Traitement des données de séquençage et identification des espèces »).

Par ailleurs, il est possible d'utiliser une approche de *metabarcoding* multi-marqueurs afin de pouvoir faire des suivis de biodiversité de plusieurs groupes taxonomiques en combinant les marqueurs les plus utilisés pour chacun des taxa ciblés (e.g., 16S pour les procaryotes, COI et 18S pour les animaux et rcb1 ou matK pour les plantes) (Drummond *et al.* 2015). L'utilisation de deux marqueurs peut parfois fournir des données complémentaires sur le même groupe taxonomique dans l'échantillon. Ainsi, un marqueur conservé avec une couverture taxinomique élevée (e.g., 18S), assurant la détection de tous les groupes principaux du taxon cible, combiné avec un marqueur plus variable (e.g., COI), fournissent des informations à une résolution taxonomique plus élevée, mais au prix d'un certain biais d'amplification (Marquina *et al.* 2019).

Comme énoncé précédemment, les banques de données de référence ne sont pas exhaustives et il n'est pas toujours possible de faire une assignation taxonomique sur les séquences obtenues après séquençage. De plus, l'identification spécifique des MOTUs a tendance à être biaisée en faveur des espèces les mieux connues qui sont généralement les premières référencées dans les bases de données. Même sans aller jusqu'à l'identification de l'espèce, les unités taxonomiques opérationnelles moléculaires sont des lectures correspondant à un groupe taxonomique individualisé. Il est donc tout à fait possible de garder les MOTUs comme des inconnues pour les analyses de communautés et ainsi associer les taxa connus et inconnus dans ces analyses. En effet, certaines analyses écologiques de la richesse et de la composition spécifiques ne nécessitent pas l'identité taxonomique, et l'utilisation des unités taxonomiques opérationnelles moléculaires ou des variants de séquences d'amplicons (ASV, Amplicon Sequence Variants; voir partie « Traitement des données de séquençage et identification des espèces ») reste suffisante. Considérant les avancées rapides au niveau des différentes technologies associées au *metabarcoding*, il est facile de penser que les données de séquences brutes ou les ADN archivés pourront être ré-analysées avec des pipelines bio-informatiques améliorés ou de nouvelles bases de données de référence dans le futur.

Les invertébrés sont parmi les taxa les plus étudiés avec la méthodologie de *metabarcoding* dans les études de suivis de la biodiversité terrestre ou d'eau douce (Liu *et al.* 2020). Ainsi, le *metabarcoding* à base d'échantillons composites nécessite simplement que le personnel soit capable d'effectuer des protocoles de collecte d'échantillons standards (e.g., filets « surbers », pièges Malaise, Winkler ou pièges lumineux) et des techniques de laboratoire. Les spécimens sont ensuite le plus souvent broyés pour former une pâte (e.g., Watts *et al.* 2019) ou homogénéisés à l'aide d'un mixeur (i.e. transformé en « soupe »; Yu *et al.* 2012). Malheureusement, cette procédure empêche ultérieurement un retour au spécimen physique pour un examen morphologique et une vérification. La séparation des spécimens et des matières additionnelles (i.e. sédiments et autres matières inorganiques) n'est généralement pas indispensable. Le passage des études de terrain conventionnelles au *metabarcoding* est souvent facilement réalisable pour les invertébrés. Les échantillons d'ADN d'organismes mixtes peuvent contenir de l'ADN provenant de tissus, de cellules ou encore d'organites d'autres organismes (Deiner *et al.* 2017).

Les analyses d'échantillons composites sont souvent basées sur l'extraction des tissus homogénéisés des échantillons (e.g., Emilson *et al.* 2017). Les spécimens dans ces types d'échantillons vont généralement varier au niveau de la composition en espèces, de l'abondance et de la biomasse, ce qui peut affecter l'évaluation des communautés avec la méthode de *metabarcoding* (Deiner *et al.* 2017). Il est également possible de faire, au préalable, un tri en fonction de la taille pour réduire la distorsion du séquençage par les grands spécimens puisqu'ils ont tendance à libérer plus d'ADN (i.e. ce qui peut conduire à ce que les plus petits spécimens à l'ADN très dilué dans celui des gros spécimens ne soient pas détectés) et améliorer la détection des espèces (Elbrecht *et al.* 2017). Par exemple, les communautés peuvent être homogénéisées en faisant un sous-échantillonnage préalable de parties du corps des organismes les plus grands qui sinon peuvent submerger les séries de séquençage (Creer *et al.* 2016). Elbrecht *et al.* (2017) quant à eux font un tri en plusieurs classes de taille, puis les mettent en commun proportionnellement à leur abondance dans chaque classe de taille (i.e. pour réduire l'influence des grands spécimens dans l'extraction) et obtiennent ainsi une amplification plus équitable des taxa. Cela leur permet de récupérer 30 % de taxa en plus dans les échantillons triés que dans les échantillons non triés, avec une même profondeur de séquençage. La couverture réelle par base représente le nombre exact de fois qu'une base est couverte par une lecture alignée provenant d'un séquençage donné. La redondance de la couverture est appelée profondeur ou profondeur de couverture (Sims *et al.* 2014). La profondeur de séquençage peut donc être définie comme le nombre de fois où une base a été séquencée. Ces étapes de tri des échantillons peuvent prendre beaucoup de temps et être source de contamination croisée entre les échantillons, et ne sont donc pas toujours recommandées. Par exemple, les auteurs ne conseillent pas cette méthode de tri par taille si l'échantillon ne comporte que des spécimens dont la biomasse relative varie peu, ou si la profondeur de séquençage est suffisamment élevée pour

détecter également les taxa petits et rares (ou de façon similaire, s'il n'est pas nécessaire de détecter ces types de taxa).

Cette méthodologie de *metabarcoding* a rapidement été considérée comme une approche intéressante pour les études à grande échelle, notamment lorsqu'il y a de fortes contraintes liées à l'identification traditionnelle des espèces d'un point de vue logistique ou financier (Liu *et al.* 2020). Par exemple, les résultats de Lobo *et al.* (2017) indiquent que la richesse en espèces du macrobenthos des estuaires serait considérablement sous-estimée si seules les méthodes morphologiques étaient utilisées (27 espèces identifiées par la morphologie *vs* 61 détectées par le *metabarcoding*). Le *metabarcoding* est donc de plus en plus utilisé dans les études sur le suivi de la diversité des espèces dans un écosystème, car cela permet une évaluation rapide de la biodiversité en identifiant plusieurs taxa simultanément dans un échantillon. En effet, elle peut rapidement caractériser les espèces présentes dans un échantillon et analyser des centaines d'échantillons en un seul cycle de séquençage, permettant ainsi d'augmenter le nombre d'échantillons analysés tout en réduisant le temps et les coûts associés. Cette approche permet une résolution taxonomique améliorée, des résultats rapides et des méthodes d'échantillonnage non invasives (majoritairement dans le cas de l'ADN environnemental). Au fur et à mesure de l'avancée des technologies de séquençage et de l'augmentation du nombre de laboratoires sous-traitants, de nouvelles possibilités s'offrent à la communauté scientifique pour produire des données génomiques à diverses échelles (i.e. régionale, nationale et mondiale). De plus, ces analyses peuvent facilement être incluses dans les programmes de suivis plus traditionnels et fournir des données complémentaires. Le *metabarcoding* fournit un plus grand volume de données que les méthodes d'échantillonnage conventionnelles (Compson *et al.* 2020). Le *metabarcoding* est utilisé comme méthode efficace pour l'identification et la détection des espèces et est suffisamment développé pour que les chercheurs non spécialistes puissent s'en servir et bénéficier de ses avantages. Par ailleurs, si les spécimens ne sont pas détruits dans le processus d'extraction, l'identification peut ensuite être vérifiée taxonomiquement et de nouvelles espèces auparavant confondues peuvent être décrites.

Actuellement, les principales limites du *metabarcoding* sont liées à l'incomplétude des bases de données de référence de code-barres ADN et aux erreurs dans certaines collections utilisées pour la production de ces séquences de références, la nature parfois destructive des méthodes d'échantillonnage, mais aussi aux difficultés liées aux estimations d'abondance. Le fait que cette technique permette uniquement de détecter la présence d'espèces dans les échantillons plutôt que leur abondance absolue et relative est en effet généralement considéré comme un défaut majeur. Plusieurs études ont pourtant démontré des relations positives entre la biomasse des espèces dans l'échantillon initial et le nombre de séquences. Cependant, une méta-analyse récente a révélé la capacité quantitative limitée du *metabarcoding* (Lamb *et al.* 2019). La raison de la variation de ces relations n'est pas entièrement comprise et peut résulter de multiples facteurs sur le terrain et en laboratoire. Par exemple, le nombre de copies d'ADN provenant d'un

individu dans un échantillon environnemental (voir partie «ADN environnemental») peut être influencé par des caractéristiques de l'ADN telles que son origine, son état ou encore son transport. De même, le biais lié aux amorces utilisés en laboratoire peut entraîner une variance accrue de l'abondance des lectures de séquençage par rapport à leur abondance réelle dans un échantillon environnemental (Deiner *et al.* 2017).

Par ailleurs, il existe différents biais inhérents à la technique de *metabarcoding* (erreurs et contamination au niveau de l'échantillonnage, de l'extraction, de la PCR, du séquençage, etc.) qui vont à leur tour générer des erreurs dans les lectures de séquences. En particulier, il faut être conscient de la présence de faux positifs (i.e. la détection d'un signal ADN en l'absence de cet organisme) et de faux négatifs (i.e. non-détection d'un ADN alors que cette espèce est présente dans le site) qui sont des biais majeurs pour les études basées sur les données d'ADN. Il est donc fortement conseillé d'inclure systématiquement des témoins négatifs, des témoins positifs et des répliques et de les traiter avec les échantillons de terrain durant l'ensemble du processus (i.e. pour essayer d'identifier les sources de contaminations par la suite si nécessaire), ainsi que de prendre des précautions lors des manipulations successives de plusieurs échantillons (Falentin *et al.* 2019).

ADN ENVIRONNEMENTAL

L'ADN environnemental (i.e. ADNe) représente tout ADN présent dans l'environnement, et peut provenir de cellules de la peau, des muqueuses, du mucus, de la salive, du sperme, des œufs, des matières fécales, de l'urine, du sang, des racines, des feuilles, des fruits, du pollen ou encore des corps en décomposition. L'ADN environnemental comprend donc à la fois de l'ADN intracellulaire (i.e. ADN trouvé dans les cellules d'organismes vivants ou récemment morts), et de l'ADN extracellulaire. L'ADNe peut être collecté à partir de multiples type d'échantillons environnementaux tels que l'eau, le sol, l'air, divers supports (fleurs, écorces, pelages, etc.), les contenus intestinaux partiellement digérés, les excréments ou encore des liquides conservateurs d'échantillons (Compson *et al.* 2020). L'ADNe est isolé directement à partir d'un matériel environnemental et suppose implicitement que l'organisme entier n'est pas présent dans l'échantillon. Bien que la définition de l'ADN environnemental semble simple, la classification entre les différentes formes d'ADN peut devenir floue puisque l'ADNe peut comprendre à la fois de l'ADN provenant d'organismes entiers de petite taille, et des traces d'organismes de plus grande taille.

La détectabilité de l'ADNe peut dépendre de la libération, de la décomposition et du transport de l'ADN dans l'environnement, et ces processus sont déterminés par des facteurs environnementaux et biologiques. La production dans l'environnement peut varier sensiblement dans le temps et l'espace et dépend largement de l'abondance d'un taxon, ainsi que de ses caractéristiques biologiques et physiologiques. Par exemple, les poissons et les batraciens sont plus susceptibles de libérer d'importantes quantités d'ADN dans leur environnement que les arthropodes, et les Diptères davantage que les Coléoptères, en raison de la perméabilité différentielle de leur tégument

(Pawlowski *et al.* 2020). Il a été démontré que la détectabilité de l'ADNe dans l'eau diminuait considérablement après le retrait de l'organisme source (Dejean *et al.* 2011 ; Thomsen *et al.* 2012). Une espèce peut ainsi produire des quantités d'ADNe variables suivant son métabolisme, son activité ou le nombre d'individus présents (Goldberg *et al.* 2016). La décomposition de l'ADN, quant à elle, va varier en fonction de divers facteurs physicochimiques et biologiques (i.e. température, UV, pH, activité microbienne, organe source d'ADN [e.g., feuilles *vs* racines ; Johnson *et al.* 2023]). Ainsi, les traces génétiques peuvent être conservées dans l'environnement pendant un temps très variable (de plusieurs heures à plusieurs décennies) et l'ADNe peut être échantillonné à partir d'environnements modernes ou anciens (e.g., carottes de sédiments, de glace ou de pergélisol). Par exemple, Giguët-Covex *et al.* (2015) ont utilisé l'ADN des sédiments lacustres pour reconstituer les activités agricoles et les paysages (voir partie « L'ADN ancien »). La décomposition de l'ADNe, sa fragmentation, la désamination des cytosines, va engendrer la diminution du nombre de molécules détectables et ultérieurement entraver le processus d'identification des espèces (Orlando *et al.* 2021). L'ADNe peut notamment se caractériser par un nombre de paires de base très faible du fait de sa dégradation. En conséquence, la dégradation de l'ADN dans l'environnement réduit la portée des études basée sur l'ADNe, car il ne reste souvent que de petits segments de matériel génétique (Ruppert *et al.* 2019). Par ailleurs, on observe des mouvements passifs dans l'environnement de l'ADN intracellulaire et extracellulaire *via* l'eau ou le vent (Pawlowski *et al.* 2020). Il faut donc avoir une bonne connaissance de la phénologie des espèces cibles et des taux de dégradation de l'ADNe dans un environnement afin d'élaborer des stratégies d'échantillonnage efficaces et de pouvoir interpréter les résultats (Goldberg *et al.* 2016). Mauvisseau *et al.* (2022) ont notamment fait une revue de la littérature afin d'améliorer la compréhension générale de la persistance de l'ADNe dans un environnement aquatique.

Bien qu'il soit connu que l'abondance, la persistance et le mouvement de l'ADN dans l'environnement, peuvent être affectés par différents facteurs abiotiques (e.g., température, lumière, oxygène) et biotiques (e.g., richesse en espèces, biomasse relative des espèces, comportement individuel), la dynamique de l'ADNe dans l'environnement reste peu connue (Compson *et al.* 2020). Or, les différentes durées de dégradation et le potentiel de transport de l'ADN a d'importantes ramifications sur l'interprétation de l'échelle d'inférence spatio-temporelle et doivent donc être considérés avec soin lors des conclusions sur la présence de l'espèce dans la communauté. En effet, l'organisme qui a produit l'ADN trouvé dans ces échantillons peut en réalité se situer dans un endroit différent de celui où il a été échantillonné, ou l'ADN a été déposé dans l'environnement plusieurs années auparavant, ou encore l'ADN peut avoir été transporté par une autre espèce plus mobile (Deiner *et al.* 2017 ; Ruppert *et al.* 2019). Ce type de problèmes rend généralement la distinction entre les deux types d'ADN (i.e. communautaire et environnemental) importante, puisque cela fait la différence entre la présence connue d'organismes à un moment et un endroit particuliers

(même si l'ADN intestinal inclut des proies potentiellement exogènes) ou le fait qu'une espèce puisse provenir d'un endroit ou d'un moment différent (Deiner *et al.* 2017 ; Ruppert *et al.* 2019). Il est parfois important de pouvoir différencier l'ADN provenant d'individus vivants de celui provenant de sources extracellulaires. Carini *et al.* (2016) ont estimé qu'en moyenne, 40 % de l'ADN procaryote et fongique provenant d'échantillon de sol était extracellulaire ou provenait de cellules qui n'étaient plus intactes. Cet ADN extracellulaire a augmenté la richesse observée jusqu'à 55 %, ce qui a entraîné une mauvaise estimation de l'abondance relative des taxons. L'ensemble de ces résultats implique que cet « ADN relique » restant dans le sol après la mort des cellules peut masquer les effets du traitement, les schémas spatio-temporels et les relations entre les taxons microbiens et les conditions environnementales, surtout dans les sols pauvres en calcaire.

De nombreuses études sur l'ADNe ont adopté des procédures quantitatives d'échantillonnage sur le terrain (e.g., échantillonnage du sol avec des carottiers, filtration de volumes fixes d'eau). Divers facteurs de terrain peuvent influencer la détectabilité de l'ADN tels que la proximité du site d'échantillonnage avec l'habitat de l'espèce considérée, la quantité de matériel échantillonné (e.g., volume d'eau prélevé), la taille du filtre ou encore la méthode de conservation (Compson *et al.* 2020 ; Pawlowski *et al.* 2020). Il existe quelques protocoles d'échantillonnage standardisés pour l'échantillonnage quantitatif pour les invertébrés du sol (e.g., ISO 23611), ou encore pour les macroinvertébrés aquatiques (Leese *et al.* 2016 ; Liu *et al.* 2020 ; Bruce *et al.* 2021 ; Pawlowski *et al.* 2022). Développer ces approches d'échantillonnage standardisées permettrait une observation cohérente et comparable dans différents environnements et écosystèmes, et maximiserait la puissance des données de *metabarcoding* (Compson *et al.* 2020).

Comme expliqué précédemment, l'inclusion de divers tests de *metabarcoding* avec différents gènes marqueurs permet d'élargir la couverture taxonomique des résultats pour les échantillons. Il est donc relativement facile qu'un échantillon d'ADNe puisse être utilisé à la fois pour une évaluation sur une espèce cible et sur la biodiversité au sens large d'un système. En revanche, en raison d'un compromis dans la spécificité des amorces, il est probable que le *metabarcoding* soit moins sensible à la détection de certaines espèces ou que le taux de détection change en fonction de la richesse des espèces. Cette technique va tout de même permettre de détecter la présence de nombreuses espèces simultanément, y compris des espèces dont on ne soupçonnait pas la présence auparavant (Deiner *et al.* 2017). Par ailleurs, comme pour les échantillons mixtes, l'extrait d'ADN peut facilement être archivé une fois l'échantillon d'ADNe analysé dans le cadre d'un objectif précis, et être séquencé de nouveau ultérieurement à l'aide d'analyses supplémentaires avec d'autres objectifs. L'ADNe permet de collecter des échantillons plus rapidement, plus facilement et dans des endroits plus éloignés et donc d'étendre les efforts d'échantillonnage sur des échelles spatio-temporelles sans précédent (Compson *et al.* 2020). L'ADNe est capable de détecter des espèces rares ou cryptiques, et cette méthodologie semble idéale pour compléter les études plus traditionnelles.

Par exemple, elle peut les compléter en ciblant différentes espèces, en échantillonnant une plus grande diversité ou en augmentant la résolution taxonomique (Deiner *et al.* 2017). Finalement, l'ADNe n'est généralement ni invasif ni destructif pour les organismes, puisqu'il n'est pas nécessaire d'échantillonner les individus (exception faite des études sur les contenus intestinaux, voir partie « Suivis de biodiversité grâce à d'autres sources d'ADNe »). Il permet d'échantillonner tous les organismes (i.e. ne dépend pas des taxa) et a le potentiel de fournir des données à de hautes résolutions spatiale et temporelle. Néanmoins, la méthode n'est pas sans impact indirect compte tenu de l'utilisation de réactifs chimiques et de matériel à usage unique.

De façon générale, afin d'optimiser l'utilisation de la méthode de *metabarcoding*, l'analyse du signal dans l'environnement doit être accompagnée d'une connaissance préalable du site et du groupe taxonomique considéré. Il est également important de considérer les biais inhérents à cette méthodologie et pouvant induire des faux positifs et faux négatifs. Ces types d'erreurs peuvent être liés à la contamination en laboratoire, sur le terrain, à une trop faible quantité d'ADNe présent dans l'échantillon ou encore à la présence d'inhibiteurs environnementaux de l'amplification moléculaire. Par ailleurs, l'ADNe étant constitué d'ADN intra cellulaire et extra cellulaire de longueur variable qui sont répartis de manière hétérogène dans l'environnement, il est possible que ces fragments ne soient pas captés lors des étapes de laboratoires, même s'il n'y a pas d'erreur lors de ces processus (Goldberg *et al.* 2016). Pour pallier au moins partiellement ces biais, il est souvent recommandé d'inclure plusieurs répétitions d'échantillonnage par site (e.g., si une espèce est détectée dans plusieurs répliques du même site, sa présence dans l'environnement est donc plus probable), et tout particulièrement pour les sites où la distribution de l'ADNe est très irrégulière. Il est préférable de faire un échantillonnage représentatif des habitats présents sur la station de façon à optimiser la probabilité de rencontre avec l'ADNe des taxons ciblés. Avec les échantillons d'ADNe, il faut donc garder à l'esprit que même si une espèce n'est pas détectée, cela ne veut pas dire qu'elle n'est pas présente (Ruppert *et al.* 2019). Il est également possible de combiner diverses techniques d'échantillonnage pour obtenir une vision plus exhaustive de la faune locale. Par exemple, Kirse *et al.* (2021) ont utilisé une approche de *metabarcoding* à la fois sur de l'ADNe du sol et sur des échantillons composites (provenant de pièges Malaise) pour étudier les arthropodes. Ainsi, en faisant des séries d'échantillons d'arthropodes composites et d'ADNe du sol dans les mêmes endroits et à différentes périodes, ils ont obtenu des données complémentaires et une image plus complète de la diversité des arthropodes, leur phénologie et leur cycle de vie dans différents types de microhabitats et de forêts. *A contrario* en cours d'eau, la comparaison avec la pêche électrique ont montré que l'ADNe est capable de révéler qualitativement et quantitativement les structures des assemblages de poissons, mais que l'ADNe a intégré un espace plus vaste que le lieu d'échantillonnage (Valentini *et al.* 2016; Pont *et al.* 2018). Les échantillonnages et les protocoles d'ADNe dépendent du type d'écosystème et du groupe taxo-

nomique analysés (e.g., Lear *et al.* 2018). Il existe différents types d'échantillons environnementaux pour lesquels il est possible d'isoler de l'ADN pour les suivis de biodiversités : les milieux liquides, les sédiments, la terre, l'air, les nids et les gîtes ou encore les produits de digestion.

Identification d'espèces à partir des liquides de piégeage ou de conservation

Les spécimens collectés sur le terrain sont majoritairement tués et stockés dans de l'éthanol avant de pouvoir faire des analyses moléculaires. Afin d'éviter certains inconvénients associés au *metabarcoding* des échantillons mixtes, comme la destruction des spécimens ou encore l'étape du tri, des protocoles non destructifs basés sur l'extraction de l'ADN dans des solutions spécifiques à partir de l'agrégat de spécimens, ou les liquides de piégeage ou de conservation des spécimens – généralement l'éthanol – ont été développés. Plusieurs protocoles impliquant des tampons de lyse (plus ou moins agressifs en termes de dégradation des tissus) ont été étudiés pour extraire l'ADN d'un mélange multi-spécifique d'Arthropodes sans broyer les spécimens (Marquina *et al.* 2022; Martoni *et al.* 2022). Zizka *et al.* (2018) ont quant à eux filtré l'éthanol employé pour préserver les échantillons de macro-invertébrés, extrait l'ADN du filtre et ensuite fait les analyses moléculaires. Ces protocoles basés sur l'ADN récupéré à partir d'éthanol de préservation des échantillons va inclure des organismes non cibles, ainsi que des inhibiteurs potentiels de la PCR provenant par exemple de l'éthanol, des sédiments ou matériel de plantes (Martins *et al.* 2019). L'ADN extrait de l'éthanol de conservation ou de piégeage a été utilisé avec succès tout aussi bien pour des séquençages d'organismes terrestres, d'eau douce ou encore d'eau salée (Zizka *et al.* 2018; Martins *et al.* 2019; Zenker *et al.* 2020; Couton *et al.* 2021). Par exemple, Martins *et al.* (2019) ont confirmé que le *metabarcoding* basé sur l'ADN de l'éthanol utilisé pour préserver les échantillons mixtes de macroinvertébrés d'eau douce peut fournir des informations fiables sur la diversité et la composition de ces taxa. Cependant, d'autres études observent des communautés différentes selon les types d'échantillons (éthanol *vs* spécimens) suggérant ainsi que l'extraction d'ADN à partir d'éthanol ne remplace pas l'extraction des échantillons en vrac (Marquina *et al.* 2019; Persaud *et al.* 2021).

Bien que cette technique semble constituer une alternative prometteuse au *metabarcoding* d'échantillons composites lorsqu'on veut conserver les spécimens, de nombreuses questions se posent encore puisqu'elle est encore relativement récente, n'a pas été minutieusement étudiée et la quantité d'ADN extraite par la solution d'éthanol reste inconnue (Persaud *et al.* 2021). La méthode a notamment été utilisée avec succès par Zizka *et al.* (2018) comparativement aux méthodes conventionnelles d'échantillons d'ADNe, mais certains taxa étaient sous-représentés (e.g., ceux de petite taille ou ceux avec un exosquelette épais). Cela suggère que pour un suivi exhaustif de la biodiversité, cette méthode n'est pas suffisante mais serait un bon complément. Par ailleurs, dans une autre étude, le *metabarcoding* à partir de l'éthanol a identifié moins de genres et de familles que la morphométrie conventionnelle

– même si le niveau de résolution était meilleur – pour des macroinvertébrés aquatiques en milieu forestier (Erdozain *et al.* 2019). L'une des principales explications des valeurs de richesse plus faibles est la faible amplification de l'ADN d'individus ayant de faibles masses proportionnelles dans l'échantillon, et cela pourrait avoir des répercussions sur les conclusions et les mesures de gestion découlant des résultats. Bien que l'éthanol conservateur d'échantillons d'insectes (obtenus par des pièges lumineux) puisse être utilisé avec succès, il est également recommandé de stocker les échantillons d'éthanol dans des conditions appropriées pour éviter la dégradation de l'ADN des insectes et augmenter le succès de la PCR (Zenker *et al.* 2020). De plus, l'utilisation des échantillons d'ADNe d'éthanol pour évaluer des communautés de macroinvertébrés ne se fait pas toujours avec succès. Récemment, certains chercheurs ont constaté que la composition de la communauté variait amplement selon la source d'ADN et que peu de taxa étaient partagés entre des échantillons d'ADN provenant de tissus mixtes et d'éthanol (qui contenait moins de séquences cibles), suggérant que l'ADNe à base d'éthanol n'est pas un substitut acceptable (Persaud *et al.* 2021). Certaines études se sont également penchées sur la question de l'influence de la durée d'archivage des échantillons sur les analyses moléculaires à partir de l'ADN extrait de l'éthanol. Par exemple, Martins *et al.* (2019) ont échantillonné des macroinvertébrés d'eau douce et ont sous-échantillonné l'éthanol qui conserve ces échantillons composites de un à 14 jours après stockage pour ensuite extraire l'ADN. Ils ont observé des effets significatifs du jour de sous-échantillonnage sur les résultats de *metabarcoding* et que l'efficacité de l'ADN extrait de l'éthanol de conservation pour identifier les taxa était meilleure dans une période de 7 à 14 jours après échantillonnage.

Par ailleurs, il est possible d'utiliser la méthode de *metabarcoding* dans d'autres types de solutions de conservation des spécimens. En effet, plus récemment, un protocole de *metabarcoding* a été utilisé avec succès sur de l'ADNe provenant d'un fluide de conservation de pièges constitué de solution saline saturée (Young *et al.* 2021). Ce type de fluide a été utilisé car il conserve les spécimens tout en résolvant les problèmes liés à la préservation des caractéristiques morphologiques (i.e. l'éthanol peut fragiliser et dégrader les spécimens d'insectes), au taux d'évaporation élevé de l'éthanol dans les pièges et au placement de pièges non surveillés à forte concentration d'éthanol dans des endroits accessibles au public (i.e. problèmes de sécurité publique). De plus, Milián-García *et al.* (2021) ont également optimisé et validé un protocole rapide et à moindre coût de *metabarcoding* sur des échantillons d'ADNe de pièges à entonnoir Lindgren qui contenaient une solution saline saturée dans le bocal de collecte pour étudier la diversité des insectes et des microorganismes. Il faut tout de même noter qu'il existe, pour le moment, un manque d'information sur la conservation de l'ADNe dans ces solutions salines saturées (efficace au moins jusqu'à quatre semaines pour les deux études citées) et sur les risques de dégradation (par le sel, le pH changé ou par hydrolyse) avec des temps de stockage plus longs.

Suivis de biodiversité en milieu aquatique

L'une des approches les plus fréquemment utilisées pour le *metabarcoding* de l'ADNe est l'échantillonnage d'eau dans les systèmes d'eau douce (Lear *et al.* 2018; Compson *et al.* 2020). Pour les organismes aquatiques difficiles à échantillonner sous l'eau, comme les mollusques bivalves, l'ADNe représente un outil opportun (Prié *et al.* 2021). Il existe deux techniques courantes pour récolter de l'ADNe à partir d'échantillons d'eau. L'eau des échantillons peut être soit filtrée, soit concentrée, et cette décision (tout comme la taille du filtre) dépend des taxa cibles car différents taxa peuvent être isolés plus efficacement de différentes manières (Ruppert *et al.* 2019). La méthodologie appliquée à l'ADNe en milieu aquatique consiste le plus souvent à filtrer une certaine quantité d'eau pour récupérer les molécules d'ADNe car cela permet d'utiliser de plus grandes quantités d'eau. L'utilisation de filtres permet de collecter l'ADN sur la matrice de filtrage, et plusieurs méthodes de filtration existent (e.g., filtration manuelle, par pompe péristaltique ou encore par pompe à vide), mais aucune solution unique optimale n'existe. Comme les techniques de filtration sont influencées par les particules en suspension et que la précipitation est moins sensible à ce problème, cette technique peut être plus avantageuse dans certains cas. Par exemple, cette technique de précipitation est généralement utilisée pour détecter les batraciens des étangs. Durant la précipitation, un mélange de sels et d'éthanol est utilisé pour précipiter l'ADN/l'ARN présent dans l'eau. Ce mélange est ensuite centrifugé pour récolter l'ADN, ce qui limite généralement le volume traité (Pawlowski *et al.* 2020). Cette méthode peut facilement compléter les méthodes conventionnelles en échantillonnant une plus grande diversité. Par exemple, Valentini *et al.* (2016) ont démontré que, pour divers systèmes aquatiques, le nombre d'espèces d'amphibiens détectées à l'aide du *metabarcoding* d'ADNe était égal, voir supérieur au nombre estimé avec des méthodes plus traditionnelles. Par ailleurs, la revue de Takahashi *et al.* (2023) synthétise les procédures liées à l'ADNe aquatique et guide les utilisateurs d'ADNe vers les meilleures pratiques possibles.

Lors de l'échantillonnage d'ADNe provenant de l'eau, il faut savoir que l'échantillonnage de la biodiversité est limité par le volume d'eau traité en raison des limites des membranes filtrantes. Par exemple, la taille des pores d'un filtre peut déterminer le type d'ADNe récolté, et bien que des pores de petites tailles permettent de capturer la plupart des cellules, le volume d'eau qu'on pourra filtrer avant que le filtre ne s'encrasse sera moindre (Pawlowski *et al.* 2020). D'un autre côté, les particules encrassant le filtre sont autant de supports sur lesquels les fragments d'ADNe peuvent s'adsorber et donc augmenter la probabilité de détection car l'adsorption de l'ADN sur des particules peut le protéger de la dégradation et augmenter sa survie (Pedersen *et al.* 2015). Certaines études suggèrent que ces contraintes seraient responsables de l'hétérogénéité observée parmi les échantillons, et que multiplier le nombre de répliques d'échantillon aiderait à la détection précise des espèces. Trois à neuf répétitions d'échantillonnage dans chaque site d'étude sont ainsi recommandées pour une meilleure estimation de la biodiversité (Ficetola *et al.* 2014; Grey

et al. 2018; Beentjes *et al.* 2019). La filtration sur le terrain va diminuer le risque de contamination et faciliter le transport des échantillons, tandis que la filtration en laboratoire va permettre de filtrer de grandes quantités, mais ne pourra s'effectuer que si les échantillons rejoignent le laboratoire rapidement pour éviter la décomposition de l'ADNe (Pawlowski *et al.* 2020). Le transport de l'eau au laboratoire peut également apporter des contraintes logistiques conséquentes, surtout si les volumes sont importants et les sites nombreux. Même si elles présentent des similitudes, il n'existe pas de technique d'échantillonnage ADNe standard s'appliquant à tous les types d'eaux de milieux aquatiques car la méthode va dépendre des propriétés chimiques et physiques des eaux (Pawlowski *et al.* 2020). Par exemple, les étangs sont souvent caractérisés par des flux réduits, des températures variables ou encore une haute turbidité, ce qui va avoir un impact sur la répartition et la dégradation de l'ADN, ainsi que sur les processus de filtration, et ultimement influencera les méthodes d'échantillonnage considérées. De façon similaire, il convient de considérer les groupes taxonomiques ciblés lors de la construction des plans d'échantillonnage (Pawlowski *et al.* 2020).

L'inférence spatio-temporelle pose particulièrement problème dans les systèmes aquatiques car les espèces détectées peuvent être présentes non pas parce qu'elles sont bien présentes, mais parce que leur ADN a été transporté d'un endroit en amont. Il a été démontré que l'ADNe se comporte dans la colonne d'eau comme des fines particules de matière organique et que sa distance de détection dépend principalement des caractéristiques hydrauliques du lit de la rivière. Cependant, l'ADN ne provient pas systématiquement de plusieurs kilomètres en amont et il est possible de caractériser des structures de communautés piscicoles distinctes distantes de quelques kilomètres seulement (Civade *et al.* 2016; Pont *et al.* 2018). L'ARN est souvent considéré comme moins stable que l'ADN (en raison de sa structure et de la présence de groupes hydroxyle augmentant la dégradation chimique abiotique), se dégradant rapidement et ne pouvant pas être détecté en quantités biologiquement significatives. Ainsi, l'utilisation de l'ARN environnemental (ARNe) pour l'identification des espèces reste relativement peu explorée. Néanmoins, des études récentes suggèrent que l'ARNe pourrait refléter la présence ou l'absence d'espèces avec une résolution spatiale et temporelle plus élevée. En effet, l'ARN devrait être produit principalement par des organismes physiologiquement actifs et pourrait donc mieux refléter les assemblages d'organismes métaboliquement actifs. En conséquence, il serait notamment utile pour améliorer la résolution de détection des organismes et l'interprétation des détections positives (Cristescu 2019; Marshall *et al.* 2021; Yates *et al.* 2021). Un risque potentiel est alors de perdre le signal des espèces rares présentes dans l'écosystème. Finalement, cela va dépendre de l'échelle spatiale et donc de la question de l'échelle à laquelle on souhaite détecter la présence des espèces. En effet, le fait que l'eau transporte l'ADN peut également être vu comme une opportunité d'obtenir un échantillonnage à plus large échelle et de détecter plus efficacement les espèces rares. De plus cela permet aussi d'avoir une certaine image de la diversité terrestre, puisque les écosystèmes aquatiques et ter-

restres sont étroitement liés (Coutant *et al.* 2021). La question de l'échelle temporelle (i.e. communauté actuelle *vs* passée) se pose d'autant plus lorsqu'on échantillonne des sédiments. En effet, la durée de conservation de l'ADN est plus longue dans les sédiments que dans l'eau, puisque des particules organiques et inorganiques présentes dans ces sédiments peuvent se lier à l'ADN et le stabiliser. Le matériel nécessaire à la récolte de l'ADNe issu des sédiments va généralement dépendre de la profondeur ainsi que de l'accessibilité des sites d'échantillonnage. La quantité de sédiments échantillonnée, quant à elle, dépend de la méthode d'extraction (e.g., des kits commerciaux) (Pawlowski *et al.* 2020).

Suivis de biodiversité dans le sol

Le sol est un des substrats les plus omniprésents et très riche en biodiversité, contenant diverses bactéries, champignons, protistes et métazoaires. Il est estimé qu'au moins 25 % des espèces animales décrites résident exclusivement dans le sol et la litière, et si le bois mort et les excréments sont pris en compte, la majorité des espèces animales terrestres semble être des habitants du sol (Decaëns *et al.* 2006). De plus, la présence de communautés microbiennes dominées par des cyanobactéries a été rapporté dans la sub-surface continentale en utilisant des preuves moléculaires, microscopiques et métagénomiques (Puate-Sánchez *et al.* 2018). Le sol est un substrat prometteur pour les analyses de *metabarcoding* d'ADNe car il contient des restes d'ADN provenant de nombreux organismes et incluant l'ADN provenant de cellules ou d'organismes multicellulaires vivants, et l'ADN extracellulaire (i.e. provenant de la lyse cellulaire et représentant une proportion importante de l'ADN total du sol) (Taberlet *et al.* 2012b). En effet, l'ADN extrait d'échantillons de sol peut être efficace pour le suivi de la biodiversité des écosystèmes terrestres. Par exemple, la méthode de *metabarcoding* a été utilisée sur de l'ADN extrait du sol et a détecté un éventail de taxa d'invertébrés plus diversifié que par les techniques d'échantillonnage conventionnelles. L'ADN isolé des sols a été notamment utilisé pour caractériser la biodiversité des vers de terre, des invertébrés, des plantes ou encore de certaines espèces de vertébrés (Deiner *et al.* 2017). Les sols forestiers ont aussi utilisé la métagénomique sur des sols forestiers pour découvrir 16 nouveaux virus géants (Schulz *et al.* 2018). Comme l'ADN est probablement plus sensible aux petites quantités de biomasse, une approche basée sur l'ADN du sol peut atténuer certains problèmes rencontrés avec les méthodologies traditionnelles au niveau de la détection de certaines espèces en raison des effets observateurs, des plantes entrant en dormance, ou encore de leurs raretés. Les informations obtenues grâce aux données issues du sol peuvent ainsi compléter les données des méthodologies plus traditionnelles et estimer les composantes de la diversité sur des échelles temporelles plus longues puisque l'ADN est susceptible de s'accumuler dans le sol pendant plusieurs années (Yoccoz *et al.* 2012).

Pour les études des sols, un petit volume de matière fraîche est généralement utilisé dans les protocoles d'extraction d'ADN ciblant les petits organismes, tandis que pour les études sur des plus gros taxa (i.e. et donc pouvant présenter des distributions

inégales et un grand nombre d'espèces rares) il faut traiter des volumes de sol plus importants pour assurer un échantillonnage précis (Creer *et al.* 2016). Comme il existe une forte hétérogénéité entre les échantillons provenant d'un même endroit, certains protocoles ont été proposés afin d'obtenir un échantillon de sol aussi représentatif que possible de la biodiversité locale. Par exemple, Taberlet *et al.* (2012a) ont développé un protocole efficace et conseillent de prélever sur chaque site au moins deux échantillons de sol. Ces échantillons doivent être composés de plusieurs petits échantillons de carottes (20 à 50 g de sol par carottage) collectés et mélangés ensemble, et ayant été eux même échantillonnés au hasard ou sur une grille et incluant du matériel de différentes profondeurs. Les carottes de sol doivent être collectées avec un carottier correctement nettoyé et stérilisé entre chaque site afin de bien enlever tout reste de terre. L'ADN extracellulaire du sol peut être extrait par un tampon de phosphate saturé sans passer par une étape de lyse cellulaire. Pour avoir un exemple de protocole détaillé et standardisé d'échantillonnage d'insectes et de sol, ainsi que des procédures de laboratoires pour l'analyse moléculaire voir Ritter *et al.* (2019).

Ces protocoles d'échantillonnage du sol ont généralement l'avantage de pouvoir être mis en place à tout moment de l'année (i.e. introduisant de la flexibilité dans la planification du terrain), sans connaissances préalable du types d'organismes à échantillonner. Ils réduisent la probabilité d'une identification erronée sur le terrain, et offrent la possibilité de standardiser les échantillonnages et d'analyser de multiples organismes avec un seul type d'échantillonnage (Taberlet *et al.* 2012a; Yoccoz *et al.* 2012).

Comme pour tout ADN, une large portion de l'ADN provenant du sol est susceptible d'être dégradée et donc son utilisation posera certains défis techniques. Il faudra notamment utiliser des marqueurs ADN plus courts pour pouvoir maximiser le nombre de correspondances pour l'identification taxonomique et faire un compromis entre leur pouvoir discriminant et leur capacité à être efficace sur des échantillons d'ADN de sols dégradés (Yoccoz *et al.* 2012). Par exemple, des amorces spécifiques aux mammifères ont été utilisées pour caractériser la diversité des grands vertébrés à partir de courts fragments d'ADN extraits du sol (Andersen *et al.* 2012). Certaines recherches suggèrent que la longueur du fragment amplifiée peut modifier la résolution temporelle d'un échantillon de sol car de longs fragments d'ADN se décomposent rapidement, alors que les courts fragments restent détectables pendant des jours ou des années après la présence de l'espèce (Deiner *et al.* 2017)

Suivis de biodiversité atmosphérique

Les bioaérosols proviennent de tous les types d'environnements (i.e. l'atmosphère, le sol, l'eau douce et les océans) et leur dispersion dans l'air est variable dans le temps et dans l'espace. Les bioaérosols sont composés de particules (dont la taille peut grandement varier) provenant d'organismes vivants (i.e. d'origine microbienne, végétale ou animale). Ils incluent des bactéries, des propagules de dispersion aérienne (spores de bryophytes et de champignons, pollen), des virus, des substances

allergènes, des fibres végétales, ou encore des débris de la peau ou des cheveux (Pumkayo *et al.* 2021). Les échantillons d'air présentent donc un fort potentiel pour les études basées sur le matériel génétique, et l'application des techniques moléculaires sur l'analyse des particules en suspension dans l'air se développe rapidement. Le *metabarcoding* en particulier, a été récemment appliqué dans le monde entier pour décrire la composition des assemblages de bactéries, champignons et plantes en suspension dans l'air (Banchi *et al.* 2020).

Les études sur l'ADN en suspension dans l'air sont concentrées sur la détection, la composition et le suivi des bioaérosols ayant des impacts négatifs sur les écosystèmes agricoles ou la santé humaine (e.g., provoquant des infections ou des réactions allergiques) (Yoo *et al.* 2017). Par exemple, Kraaijeveld *et al.* (2015) ont collecté le pollen dans l'air ambiant à l'aide de techniques classiques, puis ont extrait l'ADN du pollen collecté afin d'identifier les pollens causant des allergies en utilisant le séquençage de l'ADN. La méthode de *metabarcoding* a été utilisée avec succès sur des échantillons d'air filtré pour identifier des pollens et des champignons avec une meilleure résolution taxonomique que les méthodes visuelles (i.e. microscopie) habituelles (e.g., Kraaijeveld *et al.* 2015; Banchi *et al.* 2018). Johnson *et al.* (2021) ont démontré que l'ADN aéroporté permet de détecter plus d'espèces de plantes que la méthode traditionnelle de prospection et pourrait ainsi constituer une méthode sensible et efficace de surveillance des communautés végétales (e.g., suivre les changements dans les communautés végétales dus au changement climatique ou aux perturbations, et contribuer à la surveillance des espèces envahissantes et menacées). Le *metabarcoding* de l'ADN environnemental offre ainsi des informations particulièrement pertinentes dans le cadre du suivi de la diversité et la composition des communautés aériennes (voir Banchi *et al.* 2020). Plus récemment, une étude a démontré le potentiel des échantillons d'air dans l'évaluation de la composition et de la diversité de la communauté d'insectes grâce à l'ADN présent dans les particules en suspension dans l'air. Bien que la récupération du matériel biologique ait été plus difficile en raison des faibles concentrations de biomasse, ils ont réussi à valider la présence de traces biologiques d'origine animale dans les aérosols (Pumkayo *et al.* 2021). Lynggaard *et al.* (2022) ont également testé le fait que l'ADN capturé dans l'air peut offrir une nouvelle méthode d'étude et de surveillance des espèces animales terrestres. En effet, grâce au *metabarcoding* de l'air – et bien qu'une distance plus courte par rapport au dispositif d'échantillonnage de l'air et une biomasse animale plus élevée augmentent la probabilité de détection – ils ont détecté un large éventail de groupes taxonomiques et ainsi ont pu obtenir une caractérisation simple des communautés de vertébrés terrestres. De façon similaire, Clare *et al.* (2022) ont démontré que l'ADN d'animaux terrestres peut être identifié à partir d'échantillons d'air collectés dans des environnements naturels et utilisés pour identifier les espèces et leurs interactions écologiques. Leurs données ont montré que l'ADN dans l'air se concentre autour des zones récemment habitées mais se disperse loin des sources, soulignant le potentiel d'échantillonnage à distance.

La dispersion des champignons se fait dans l'atmosphère et dépend fortement des caractéristiques des taxa, des conditions abiotiques ou encore des phénomènes climatiques (e.g., vent, température, humidité, phénologie, etc.). On observe donc une forte hétérogénéité géographique et temporelle des communautés aériennes fongiques et la distance parcourue par ces particules en suspension dans l'air depuis leurs sources demeure une question régulière (Tordoni *et al.* 2021). Les micro-organismes en suspension dans l'air étant fortement dispersés, cela rend leurs concentrations plus faibles que dans d'autres environnements, tels que le sol ou l'eau. Le choix du dispositif d'échantillonnage et le volume d'air semblent donc représenter des facteurs importants à prendre en compte pour ce type d'étude. Bien que de multiples méthodes d'échantillonnage existent pour collecter les particules en suspension dans l'air, on ne sait toujours pas laquelle est la plus performante pour les analyses moléculaires. Le choix de la méthode d'échantillonnage la plus appropriée est essentiellement déterminé par les objectifs de l'étude. La filtration est l'une des méthodes d'échantillonnage des bioaérosols atmosphériques les plus utilisées et consiste à pomper de l'air à travers un filtre à membrane poreuse pour capturer les bioaérosols. Les champignons et les plantes sont généralement échantillonnés par filtration, et plus rarement par collecte passive ou à l'aide d'échantillonneurs Andersen à aspiration. Par ailleurs, il faut faire attention car les bandes adhésives de certains pièges peuvent contenir des inhibiteurs de PCR (Banchi *et al.* 2020). Enfin, sans recourir à l'air, on peut aussi rechercher l'ADNe dans la neige, piste particulièrement étudiée dans certaines régions du globe (Rosa *et al.* 2020).

Suivis de biodiversité grâce à d'autres sources d'ADNe

Il existe de nombreuses sources supplémentaires d'ADNe telles que les fluides physiologiques (le sang, la salive, la sève), les contenus d'appareil digestif (ADNi ou «*ingested DNA*») et les déjections ou encore le miel. Par exemple, l'utilisation du *metabarcoding* sur le pollen du miel a permis l'identification de sa composition florale et d'étudier les plantes visitées par les abeilles (Hawkins *et al.* 2015). Le *metabarcoding* a également été utilisé sur du salep (i.e. boisson à base d'orchidée) pour comprendre la diversité des espèces et de la provenance des tubercules d'orchidées pour éventuellement remonter la chaîne de commercialisation des espèces menacées jusqu'aux récolteurs et à leurs habitats naturels. De plus, l'analyse moléculaire des os de pelotes de réjection est efficace dans l'identification d'espèces de petits mammifères proies, et a révélé des espèces non détectées auparavant, voire de nouvelles espèces (Rocha *et al.* 2015). Par ailleurs, les mouches se nourrissant sur des carcasses, des plaies ou des fèces représentent une source d'ADN de mammifères et peuvent donner une image relativement impartiale de leur diversité locale (Calvignac-Spencer *et al.* 2013). La classification du matériel génétique extrait à partir des échantillons tels que les matières fécales est parfois problématique; il est tout de même généralement appelé ADNe (Deiner *et al.* 2017). Ces diverses sources d'ADNe ne sont pas toujours utilisées pour évaluer les communautés terrestres, mais offrent d'autres opportunités d'études et tout particulièrement au niveau des réseaux trophiques (e.g., Clare 2014). L'ADN

provenant des déjections est ainsi fréquemment utilisé pour évaluer la composition du régime alimentaire, la dynamique des populations, détecter des espèces d'intérêts ou encore aborder la génétique des populations (Rodgers & Janečka 2013). La salive déposée sur des brindilles broutées a été prélevée comme source d'ADNe et utilisée pour identifier les espèces qui ont consommé la plante (Nichols *et al.* 2012). Il est également possible de collecter l'ADN de la salive des prédateurs sur les coquilles d'œufs d'espèces d'oiseaux nichant au sol et considérés comme en danger, ainsi que sur les carcasses afin d'identifier des espèces de prédateurs (Hopken *et al.* 2016). D'autres auteurs ont étudié les proies consommées par une espèce d'araignée localisée dans diverses forêts (Van Lantman *et al.* 2021). Ils ont donc regardé le régime alimentaire des araignées à l'aide de la méthode de *metabarcoding* et ont pu confirmer que l'espèce était un prédateur généraliste et que la composition des espèces d'arbres affectait à la fois la richesse spécifique des proies et leur taille dans le régime alimentaire de l'araignée.

Fortement lié à l'ADNe, l'ADN «*dérivé d'invertébrés*» (i.e. où le matériel génétique des vertébrés est extrait des invertébrés), est apparu ces dernières années comme une méthode pertinente pour suivre les vertébrés terrestres et augmenter leur détectabilité (Schnell *et al.* 2015). De multiples invertébrés (e.g., sangsues, mouches, etc. ou autrement dit les hémato-phages, coprophages et saprophages) ont été considérés comme des échantillonneurs efficaces de divers vertébrés et utilisés pour détecter et identifier leurs espèces hôtes en séquençant l'ADN contenu dans leurs repas. Par exemple, des sangsues terrestres ont été utilisées comme sources d'échantillons d'ADN, et le *metabarcoding* a permis de détecter des espèces de vertébrés menacées malgré une incertitude sur le cadre spatio-temporel de l'échantillonnage (Schnell *et al.* 2015). L'ADN provenant de sangsue a permis d'obtenir des données complémentaires à d'autres méthodologies comme les pièges caméras, et particulièrement pour des espèces difficiles à détecter (Schnell *et al.* 2012). Une autre étude a également détecté davantage d'espèces de mammifères en analysant l'ADN extrait de mouches qu'avec les pièges photographiques (Lee *et al.* 2016). La plupart des études se concentrent sur les mammifères, mais une étude récente a démontré que ce type d'ADN peut être utilisé pour des suivis de la diversité des grenouilles et permet d'augmenter la détectabilité de certaines espèces par rapport aux échantillonnages traditionnels (Cutajar & Rowley 2020). Les diverses études publiées sur cet ADN provenant de l'ingestion de matériel génétique par les invertébrés semblent démontrer ce potentiel puisqu'elles ont détecté avec succès l'ADN de l'hôte, ce qui particulièrement prometteur dans la détection d'espèces rares.

Une grande majorité des plantes à fleurs sont pollinisées par des insectes et les fleurs sont des ressources pour des milliers d'espèces d'arthropodes. C'est pourquoi des chercheurs se sont intéressés au potentiel des fleurs sauvages en tant que nouvelle source d'ADNe d'arthropodes (Thomsen & Sigsgaard 2019). Ils ont effectué le *metabarcoding* de l'ADNe de fleurs de plusieurs espèces végétales à l'aide de deux ensembles d'amorces génériques, et ont découvert que les arthropodes laissent des

traces d'ADN sur les fleurs avec lesquelles ils interagissent. De plus, leurs données semblaient indiquer qu'une richesse encore plus élevée pourrait être obtenue en utilisant cette approche d'ADNe. Cette nouvelle approche d'ADNe non invasive pourrait permettre d'obtenir une meilleure estimation de la composition et de la distribution des espèces, ou encore d'aider au suivi d'espèces cryptiques, menacées ou envahissantes. Par ailleurs, les données de *metabarcoding* provenant de charges polliniques peuvent aider à construire des réseaux de pollinisateurs (Pornon *et al.* 2017).

Les études évaluant la diversité des arthropodes dans des échantillons mixtes avec la méthode du *metabarcoding* se sont majoritairement concentrées sur l'identification de ces taxa d'arthropodes. Or, un grand nombre d'espèces d'invertébrés se nourrissent de vertébrés et échantillonnent ainsi leur ADN. Une récente étude s'est posée la question de savoir s'il était possible d'optimiser les informations obtenues à partir d'échantillons d'arthropodes mixtes en utilisant des méthodes pour détecter l'ADN de vertébrés dans ces échantillons (i.e. sans cibler un invertébré spécifique se nourrissant sur des vertébrés). Pour cela, Lynggaard *et al.* (2019) ont utilisé deux ensembles d'amorces de *metabarcoding* de vertébrés et de mammifères sur de l'ADN extrait d'échantillons d'arthropodes mixtes collectés à l'aide de pièges Malaise aériens et de pièges Barber au sol. Ils ont ainsi observé que les deux approches pouvaient être combinées et qu'avec un surcroît d'investissement assez faible, ils pouvaient obtenir des informations supplémentaires sur la diversité des vertébrés grâce à des échantillons d'arthropodes mixtes.

De plus, une autre étude suggère que l'application du *metabarcoding* à des échantillons d'insectes peut permettre d'évaluer à la fois la biodiversité des insectes, mais aussi la diversité procaryotique et eucaryotique de ces échantillons (i.e. incluant les organismes présents sur et à l'intérieur des corps d'insectes et toute nourriture restant dans leur système digestif). En effet, les auteurs ont échantillonné des insectes à l'aide de pièges SLAM ou Malaise et ont ensuite comparé les résultats des analyses moléculaires des échantillons composites avec les résultats obtenus à partir d'ADNe du sol provenant des mêmes sites. Ils ont observé une composition taxonomique différente entre les échantillons de sol et d'insectes, ce qui laisse penser que les échantillons composites d'insectes complètent la couverture taxonomique des échantillons de sol (Ritter *et al.* 2019). Une étude récente a utilisé le *metabarcoding* de prélèvements de sciure de bois pour étudier la dynamique temporelle des communautés bactériennes et fongiques et des guildes écologiques microbiennes dans la décomposition de disques de bois expérimentalement placés sur un sol forestier (Viotti *et al.* 2021).

L'ADN ancien

La paléo-génétique (i.e. l'étude de l'ADN ancien ou ADN_a) est l'étude des processus démographiques passés. La recherche sur l'ADN ancien porte sur des molécules d'ADN extrait de matériel biologique ancien dégradé (i.e. dents, os, peaux, tissus naturellement momifiés ou encore écailles de poisson). Ces matériaux peuvent provenir de peaux/squelettes de musées, de matériaux archéologiques ou de restes paléontologiques et

leur âge varie de quelques dizaines d'années à une centaine de milliers d'années (Gugerli *et al.* 2005 ; Leonard 2008). L'extraction de l'ADN_a peut se faire à partir de divers échantillons environnementaux comme les sédiments et d'une variété de substrats d'origine animale tels que des coquilles d'œuf, des plumes ou des coprolithes. L'ADN_a étant dégradé, les études qui l'utilisent se limitent généralement à l'analyse de courtes régions au sein des gènes mitochondriaux (Greal *et al.* 2017).

L'utilisation de l'ADN_a devrait révéler de nouveaux aspects intéressants de la biologie et de l'histoire de nombreuses espèces et populations, et contribuer à leur conservation. En effet, ces données permettent d'approfondir les connaissances sur la biologie et l'écologie des espèces, mais également de formuler des recommandations en matière de conservation et de gestion (Leonard 2008). Par exemple, des données de séquences d'ADN_a d'espèces de Kiwi de Nouvelle-Zélande (*Apteryx* spp. Shaw, 1813) ont permis de mieux comprendre les aires de répartition passées des espèces de kiwi dont les os sont morphologiquement indiscernables (Shepherd & Lambert 2008). Pacioni *et al.* (2015) ont utilisé le Woylie ou Bettong à queue en brosse (*Bettongia penicillata ogilbyi* Gray, 1837) pour étudier dans quelle mesure l'ADN_a peut améliorer la compréhension de la dynamique associée au déclin de la faune. Leurs résultats montrent qu'il existe un précédent pour le rétablissement du flux génétique entre des populations actuellement isolées. Ces résultats ont un impact immédiat sur la manière dont les populations existantes sont gérées puisque cela suggère que des stratégies telles que des corridors naturels et des translocations pourraient aider à rétablir le flux génétique à des niveaux historiques et contribuer au maintien de la diversité génétique restante.

L'ADN sédimentaire ancien (ADN_{seda}) est un outil émergent permettant de reconstituer la biodiversité passée avec une résolution taxonomique élevée (Chen & Ficetola 2020). L'analyse des dépôts sédimentaires lacustres peut notamment reconstituer les communautés fongiques et révéler de potentielles altérations, ainsi que des changements d'écosystèmes sur de longues périodes. Ces changements observés dans ces communautés peuvent fournir des informations sur l'abondance des plantes hôtes des champignons mycorrhiziens ou pathogènes, les changements climatiques et d'autres fonctions écologiques exercées par des groupes fonctionnels spécifiques de champignons (Seeber *et al.* 2022). L'étude de Dommain *et al.* (2020) s'intéresse aux défis et opportunités des approches sédimentaires de l'ADN_{seda} dans un contexte de reconstitution de la biodiversité tropicale en milieu forestier. Ils décrivent la composition taxonomique des sédiments des 2200 dernières années et n'observent globalement pas de déclin de la richesse taxonomique avec l'augmentation de l'âge (i.e. des informations vieilles de plusieurs milliers d'années sur la biodiversité passée peuvent être extraites des sédiments tropicaux). Cependant, ils suggèrent de mener des études génomiques complètes du biote tropical pour que l'ADN sédimentaire soit une méthodologie viable pour les futures études sur la biodiversité tropicale.

L'ADN_a offre ainsi un moyen unique de nous renseigner sur les changements génétiques, les processus écologiques et d'observer l'histoire évolutive des taxons existants et éteints.

L'ADN ancien aviaire a notamment été utilisé pour étudier divers sujets allant de la reconstruction des relations phylogénétiques aux comportements. Ces études ont abordé des questions de phylogénie et biogéographie, de domestication, de taxonomie d'espèces existantes et éteintes, de reconstruction du paléo-environnement, d'archéozoologie, ou encore de gestion de la conservation (Grealy *et al.* 2017). La question de la domestication des animaux – en particulier – est un sujet qui fait couler beaucoup d'encre et intéresse beaucoup les scientifiques. Dans ce contexte, McHugo *et al.* (2019) ont montré comment l'ADN ancien des animaux domestiques et de leurs géniteurs et congénères sauvages a permis une nouvelle vision de leurs origines génétiques et du processus de domestication lui-même. Les auteurs d'une revue de littérature sur les sujets de l'ADNa aviaire, quant à eux soulignent la présence d'un biais dans la focalisation temporelle et l'importance de comprendre les mécanismes à l'origine de l'évolution et de l'extinction des espèces d'oiseaux (Grealy *et al.* 2017). Seerholm *et al.* (2018) ont utilisé l'ADNa provenant d'assemblages d'os sur une grande échelle en Nouvelle-Zélande et ont pu décrire des espèces et des modèles que les approches morphologiques n'avaient pas permis d'identifier. Cette étude offre l'opportunité de mieux comprendre la biodiversité passée et les pratiques de subsistance des Māori, et offre un aperçu des impacts anthropogéniques sur la faune indigène après la première arrivée de l'homme. Par ailleurs, Gugerli *et al.* (2005) ont résumé les recherches faites sur l'ADNa des plantes, en examinant les conditions nécessaires à la préservation de l'ADN dans les tissus végétaux et en indiquant les domaines de recherche prometteurs.

CONSIDÉRATIONS TECHNIQUES POUR L'ÉCHANTILLONNAGE DE MATÉRIEL GÉNÉTIQUE

L'ADN collecté doit être préservé, extrait, amplifié et séquencé, pour permettre la classification des espèces par assignation taxonomique à l'aide des bases de données de référence. Pour avoir plus de détails sur les méthodes d'extraction, de stockage, d'amplification et de séquençage de l'ADN à partir d'échantillons environnementaux, voir la revue de Lear *et al.* (2018). Une attention toute particulière aux procédures d'échantillonnage est nécessaire pour éviter la contamination de l'ADN et s'assurer de sa bonne conservation. Les méthodologies de laboratoire ont tendance à être bien documentées, car la plupart utilisent des kits commerciaux d'extraction d'ADN standardisés et des ensembles d'amorces déjà publiés (Compton *et al.* 2020). L'efficacité de la détection de l'ADN va dépendre des procédés moléculaires employés, et l'activité de laboratoire du *metabarcoding* montre une complexité qui doit être prise en compte dans les études de suivis de biodiversité (e.g., kits d'extraction d'ADN, choix des amorces PCR, etc.). Ce sont des étapes cruciales, car elles peuvent toutes être susceptibles d'introduire des biais (e.g., biais d'extraction, d'amplification par PCR) ou des erreurs dans les séquences introduites lors des PCR ou du séquençage.

DÉCONTAMINATION ET MANIPULATION DES ÉCHANTILLONS D'ADN

Un défi majeur de l'utilisation des méthodes de *metabarcoding* de l'ADN est d'éviter la contamination croisée entre sites lors des prélèvements d'échantillons sur le terrain (la contamination par de l'ADN humain est moins problématique, car facile à éliminer de la liste des séquences), ainsi que d'avoir des techniques de décontamination qui sont compatibles avec le matériel génétique. Tout équipement et matériel entrant en contact avec l'échantillon d'ADN doit être dépourvu d'ADN. Pour cela, il est possible soit d'avoir recours à de l'équipement stérilisé à usage unique (e.g., flacons, tubes, bouteilles, filtres à eau), soit de décontaminer le matériel au préalable. Par exemple, l'usage de gants stériles renouvelés pendant l'échantillonnage sur le terrain est fortement conseillé voire indispensable, ainsi que l'utilisation de sacs et flacons stériles pour le transfert des échantillons. Néanmoins, il est également recommandé de faire des contrôles négatifs sur le terrain et durant les étapes d'analyse pour s'assurer de la validité d'un échantillonnage et éventuellement identifier et quantifier de potentielles contaminations (Goldberg *et al.* 2016; Deiner *et al.* 2017).

L'utilisation de matériels à usage unique peut vite augmenter les coûts et les scientifiques souhaiteront réutiliser le matériel pour les diminuer. Pour cela, la décontamination est souvent la solution envisagée. Le personnel doit être particulièrement prudent s'il veut réutiliser l'équipement pour l'échantillonnage, car la sensibilité à détecter des traces d'ADN lors du séquençage crée des risques de contamination non négligeables par transfert d'ADN. Le plus couramment, l'équipement est lavé entre chaque échantillon afin de minimiser la contamination et maintenir l'indépendance, ce qui est particulièrement important pour les études d'ADNe (Liu *et al.* 2020). Ces traitements variés peuvent être relativement simples tels un lavage avec de l'eau de javel ou de l'eau savonneuse. Tremper l'équipement dans de l'eau de javel à 10 % pendant au moins 10 min semble une méthode simple et efficace pour la décontamination de l'ADN, surtout si un traitement supplémentaire de rinçage à l'éthanol ou à l'eau stérilisée pour éliminer l'eau de javel restante est utilisé sur le terrain (Liu *et al.* 2020). Comme il est préférable d'éviter l'utilisation de grandes quantités d'eau de javel sur le terrain, certains produits de décontamination commerciaux (e.g., DNA away, DNA-ExitusPlus) – dont le prix est souvent élevé – existent (Pawlowski *et al.* 2020). Par exemple, DNA-ExitusPlus garantit une destruction rapide et efficace des acides nucléiques sans propriétés corrosives ou toxiques (Esser *et al.* 2006). Il est également possible d'avoir des procédures plus compliquées et strictes, constituées d'étapes où il est demandé de laver successivement à l'eau savonneuse, rincer à l'eau distillée, puis à l'éthanol, faire un traitement pour éliminer les ARN, DNases et ADN (e.g., ELIMINase) et exposer le matériel à l'irradiation UV (i.e. lumière du soleil) pendant un jour avant utilisation. De plus, les matériaux sont à nouveau rincés avec de l'éthanol et de l'eau courante prélevée sur le site immédiatement avant échantillonnage pour éliminer toute trace résiduelle de solution décontaminante pouvant dégrader l'ADN (Erdozain *et al.* 2019).

MÉTHODOLOGIES DE CONSERVATION ET ARCHIVAGE DES ÉCHANTILLONS

La décomposition de l'ADN juste après échantillonnage et pendant les étapes de traitement en laboratoire doit être prise en compte. Les échantillons doivent donc être stockés et manipulés de manière à garantir au mieux la conservation de l'ADN. Il est important d'optimiser cette conservation qui aura un impact sur la qualité de l'ADN. De plus, l'archivage de ce matériel génétique est essentiel car il permet d'autres ré-analyses ultérieures (Compson *et al.* 2020). Ainsi, une bonne conservation des échantillons permettra par la suite de les réutiliser pour de nouvelles questions de recherches ou des questions auxquelles nous ne pouvons pas actuellement répondre (limitations liées à la technologie), en changeant d'amorces ou en les comparant avec d'autres données par exemple.

Il faut bien évidemment éviter les conservateurs génériques qui ne conviennent pas à la conservation de l'ADN (e.g., formaldéhyde, éthylène glycol, éthanol trop dilué [$< 70\%$]). L'utilisation d'alcools transformés et de formol doit être évitée car ils dénaturent les acides nucléiques (Creer *et al.* 2016). La conservation immédiate des échantillons est essentielle pour éviter la dégradation, sous l'influence de variables environnementales telles que la température, le pH ou encore l'exposition à la lumière. Il existe différents types de conservation adéquates, comme le séchage, la congélation à -20°C , la conservation dans de l'éthanol à 100 % ou un tampon de lyse cellulaire (Creer *et al.* 2016; Goldberg *et al.* 2016). La température étant un facteur clé de la préservation de l'ADN, les températures ambiantes ne sont à tolérer que pour les stockages à court terme. La plupart des études stockent les échantillons en dessous de la température ambiante pour réduire la dégradation de l'ADN (conditions de stockage de -20 ou -80°C en laboratoire) (Liu *et al.* 2020).

Les protocoles ayant pour objectif d'empêcher ou de limiter la dégradation de l'ADN pendant l'échantillonnage et le stockage sont divers et changent selon la source (e.g., terrestre, eau douce, sédiments), l'origine (i.e. échantillons mixtes ou ADN) ou la durée du stockage des échantillons. Diverses méthodes de stockage sont envisageables, comme le refroidissement à 4 , -20 ou -80°C , le séchage, la lyophilisation ou encore l'ajout de tampons de conservation (Lear *et al.* 2018). Dans le cas des échantillons mixtes où le refroidissement des échantillons sur le terrain est difficile, il est possible de les transporter et de les stocker à des températures plus élevées (i.e. *c.* 5°C ou à température ambiante), si l'échantillon est directement mis dans une solution de conservation ou un tampon qui stabilise l'ADN. Cependant, il est généralement recommandé d'au minimum les placer dans un réfrigérateur ou d'abaisser la température de la pièce de stockage pour ralentir la décomposition de l'ADN (Pawlowski *et al.* 2020). En dépit de l'empreinte environnementale non négligeable, il est également recommandé de transférer les échantillons dans un congélateur à -20°C dans les 48 heures pour un stockage à court terme (e.g., quelques semaines ou mois) et dans des congélateurs à -80°C pour un stockage à plus long terme (e.g., plusieurs mois ou années) (Lear *et al.* 2018).

Il est particulièrement important d'optimiser la préservation des spécimens dans les échantillonnages par piégeage qui accumulent des arthropodes sur de longues périodes de temps sur le terrain. Le type d'agent de conservation et les conditions de transport et de stockage vont être tous deux importants. La plupart des études utilisent de l'éthanol à $> 95\%$ de qualité moléculaire comme agent de conservation pour piéger les invertébrés terrestres. Le propylène glycol de qualité alimentaire est également une alternative intéressante à considérer dans ces cas, en raison de sa non-toxicité, de son ininflammabilité et de ses faibles taux d'évaporation (Liu *et al.* 2020), mais on ne connaît pas sa capacité à conserver l'ADN libre et sa viscosité est un problème dans les étapes de filtration, où l'ADN s'attache sur des membranes filtrantes par polarité. En fonction des conditions climatiques et du choix de l'agent de conservation, ces types de pièges peuvent quand même être exploités sur le terrain pendant plusieurs jours à quelques semaines, avant le prélèvement et le traitement des échantillons (Barsoum *et al.* 2019; Dopheide *et al.* 2019).

TRAITEMENT DES ÉCHANTILLONS EN LABORATOIRE

Une étude standard utilisant du matériel génétique comprend trois étapes au cours desquelles sont traités les échantillons mixtes ou d'ADN : l'extraction d'ADN (i.e. isolation des molécules d'ADN des échantillons), l'amplification en chaîne par polymérase (i.e. PCR, où de nombreuses copies d'une région génomique cible sont créées suite à une série de réactions enzymatiques) et le séquençage à haut débit ou séquençage de nouvelle génération (i.e. HTS, « High Throughput Sequencing », qui produit des millions de séquences). Il peut exister divers biais et risques de contaminations durant chaque étape. En plus des contrôles de terrain discutés précédemment, il est important d'avoir une procédure de laboratoire qui comprend des contrôles positifs, des contrôles négatifs et des répliques à toutes les étapes (Deiner *et al.* 2017). Un point essentiel pour limiter les sources de contaminations en laboratoire est la sectorisation de chacune des étapes (idéalement au moins trois salles dédiées à chacune des étapes : préparation des mélanges réactionnels de PCR; extraction de l'ADN à partir des échantillons; analyse des amplifions et préparation des librairies HTS le cas échéant). Cela permettra notamment de limiter la contamination des échantillons et des réactifs de PCR par des aérosols d'amplifions.

Extraction

En tout premier lieu, pour extraire l'ADN d'un échantillon, il faut passer par une étape de lyse du matériel échantillonné afin d'isoler l'ADN présent dans les cellules ou les organelles. La lyse peut s'effectuer soit à l'aide de composants chimiques (i.e. présents dans les tampons de lyse), soit par rupture mécanique des tissus (e.g., en ajoutant des billes à l'échantillon et en le secouant à l'aide d'un vortex ou d'un homogénéiseur à billes). Par la suite, il est nécessaire d'isoler l'ADN en retirant de l'échantillon tous les composants organiques et inorganiques autres que l'ADN. Parfois, cette étape implique l'élimination des inhibiteurs de PCR contenus dans certains échantillons environnementaux afin d'assurer une bonne amplification

ultérieure. Ces étapes se font en isolant et purifiant l'ADN sur une membrane de silice, à l'aide de billes magnétiques ou par précipitation (Pawlowski *et al.* 2020). Comme l'utilisation de différentes méthodes d'extraction peut mener à des assemblages d'espèces différents, l'emploi d'un seul protocole est requis au cours d'un projet, afin de garantir la reproductibilité et la comparabilité des résultats (Pawlowski *et al.* 2020).

L'extraction d'ADN est souvent réalisée avec des kits d'extraction d'ADN commerciaux (voir Lear *et al.* 2018 pour plus de détails). Par exemple, l'utilisation de la filtration et du kit DNeasy Blood & Tissue (Qiagen, Hilden, Allemagne) est recommandée pour les échantillons d'eau dans les enquêtes sur la biodiversité des eucaryotes et ce kit Qiagen DNeasy Blood & Tissue est aussi largement utilisé pour les invertébrés terrestres composites (Liu *et al.* 2020). Le choix du kit dépend donc du type d'échantillons mais aussi du type d'organismes considéré (e.g., l'abondance des acides gras chez les insectes et annélides suscite le recours à des kits particuliers).

Exemples de kits commerciaux :

– utilisés pour les échantillons mixtes: le kit Machery-Nagel NucleoSpin Tissue extraction (Machery-Nagel GmbH & Co. KG, Düren, Germany), le DNeasy® Blood and Tissue Kit (QIAGEN GmbH, Hilden, Germany), le Bokun Insect DNA Extraction Magnetic Bead Kit (Changchun Bokun Biotech Co., Changchun, Jilin), le kit NucleoMag® ;

– pour les échantillons sol : NucleoSpin® Soil (Machery-Nagel, Düren, Germany), Precellys® Soil DNA Kit (Bertin Technologies, Montigny-le-Bretonneux, France), PowerLyzer® Soil DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA), PowerSoil® Soil DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA), PowerMax® Soil DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA), E.Z.N.A. Mag-Bind® Soil (Omega bio-tek, Norcross, GA, USA), FastDNA™ SPIN Kit for Soil (MP Biomedicals, Solon, OH, USA) ;

– pour les échantillons aériens l'ADN est extrait en utilisant généralement des kits comme le PowerSoil DNA isolation Kit (Qiagen, anciennement MoBio), ou le Machery-Nagel NucleoSpin Food Kit, il y a aussi le ZR Fungal/Bacterial DNA MicroPrep™ Kit (Zymo Research). Pour le pollen il existe le QIAamp DNA Mini Kit, le Qiagen DNeasy Plant Mini Kit ;

– pour les échantillons d'eau: le kit VigiDNA RW, le DNeasy PowerWater Kit (Qiagen), les kits Blood and Tissue ;

Par ailleurs, il existe un kit d'extraction d'ADN universelle modulaire qui fournit une méthode d'extraction adaptable pour différents types d'échantillons (e.g., échantillons de tissus, de sol ou d'eau) et peu coûteuse (Sellers *et al.* 2018).

Tsuji *et al.* (2019) ont fait une revue des méthodes courantes pour récolter et extraire de l'ADNe provenant de milieux aquatiques. Pour récupérer l'ADNe à partir d'échantillons d'eau, il faut généralement passer par une étape de filtration, précipitation à l'éthanol et centrifugation. L'utilisation de la précipitation à l'éthanol convient le plus souvent aux échantillons d'eau de petit volume et pour les zones où l'accès à un équipement de filtration est difficile. L'utilisation d'éthanol non dénaturé est préférable pour les précipitations mais peut être limitant d'un point de vue logistique ou financier (Liu *et al.* 2020).

Amplification par PCR

L'amplification en chaîne par polymérase (PCR) qui permet de générer de nombreuses copies du fragment d'ADN cible repose sur l'utilisation de fragments courts appelés « amorces » qui sont capables de s'hybrider de façon spécifique, grâce à la complémentarité des bases sur l'un des brins d'ADN. Comme l'ADN cible est généralement en quantité faible ou dégradé dans les échantillons – particulièrement dans les échantillons d'ADNe – il faut amplifier la région cible grâce à une PCR pour générer suffisamment de matériel pour le séquençage. Grâce à la méthodologie PCR, le nombre de fragments d'ADN amplifiés (i.e. aussi appelés amplicons) croît de manière exponentielle. La PCR est constituée de trois étapes principales (chacune durant en général de 30 à 60 secondes), qui seront répétées sur 15 à 40 cycles. L'ADN est d'abord dénaturé par séparation des deux brins de l'hélice d'ADN grâce à de hautes températures (95°C). Les amorces introduites vont ensuite être hybridées à la région cible de l'ADN à une température qui dépendra des amorces et du degré de spécificité requis (i.e. entre 45 et 60°C). À une température dépendante de l'enzyme utilisée, autour de 72°C, une polymérase est enfin utilisée pour l'élongation de la région de code-barres (Pawlowski *et al.* 2020).

Il existe un compromis lié à la longueur du code-barres amplifié, entre des fragments d'ADN plus longs permettant une meilleure résolution taxonomique et des fragments plus courts augmentant la probabilité d'être séquencés (Liu *et al.* 2020). Par ailleurs, les fragments plus courts peuvent persister plus longtemps dans l'environnement et donc modifier l'échelle spatio-temporelle de l'inférence faite à partir d'un échantillon d'ADNe (Deiner *et al.* 2017). Il faut donc s'assurer que le code-barres ne soit pas trop court et présente assez de variations pour distinguer les espèces proches, mais pas trop long pour pouvoir répondre aux exigences techniques du séquençage (variable selon la technologie) (Pawlowski *et al.* 2020). Ceci est particulièrement important dans le cas de l'ADNe, où les amorces doivent être suffisamment courtes pour amplifier les échantillons dégradés, identiques au sein d'une espèce mais variables d'une espèce à l'autre, et flanquées de régions assez conservées pour amplifier plusieurs espèces sans perdre la spécificité du groupe cible (Ruppert *et al.* 2019).

La sélection des amorces est un élément essentiel pour le succès de la détection et l'identification des espèces ; ce choix aura notamment des conséquences sur la capacité à distinguer les taxa. Il existe deux types d'amorces PCR dites sens et anti-sens, qui sont des séquences d'ADN s'hybridant à la région du code-barres ADN comme points de départ de la synthèse des brins d'ADN complémentaires. Elles sont taxon-dépendantes et déterminent le fragment d'ADN à amplifier. Idéalement, les amorces sont assez génériques pour amplifier toutes les espèces appartenant au groupe cible, et assez spécifiques pour amplifier seulement celles-ci (Creer *et al.* 2016 ; Deiner *et al.* 2017 ; Ruppert *et al.* 2019). La région variable cible doit donc être choisie pour pouvoir discriminer les espèces attendues dans l'écosystème, et maximiser les chances d'assignation taxonomique au moment de l'analyse bio-informatique (Falentin *et al.* 2019). Dans le cas de la détection

d'espèces cibles, les amorces doivent être bien spécifiques afin d'éviter les faux positifs lors de la détection. Bien que le choix de sites de liaison d'amorce conservés de manière évolutive augmente la couverture taxonomique du marqueur ADN, des bases dégénérées peuvent être utilisées pour augmenter la couverture pour les taxa où les sites de liaison d'amorce ne sont pas conservés. Cela améliore généralement la récupération des taxa, mais augmente aussi les biais d'amplification entre les taxa car il y a compétition d'appariement pour tous les variants d'amorces générés.

La base de données des amorces du système de base de données Barcode of Life (BOLD, http://boldsystems.org/index.php/Public_Primer_PrimerSearch, dernière consultation le 22 juin 2023), fournit une collection d'amorces publiées. Les chercheurs peuvent déterminer les amorces candidates en recherchant dans la base de données des taxa étudiés, mais les évaluations *in silico* et *in vivo* sont également encouragées pour la validation, en particulier pour les groupes taxonomiques moins étudiés.

Les opérateurs vont généralement utiliser un sous-échantillon de l'ADN total extrait pour chaque PCR, ce qui explique que certaines espèces puissent manquer dans une réaction PCR (i.e. faux négatif). Plusieurs répétitions de PCR (i.e. répliques) sont généralement effectuées pour chaque échantillon d'ADN. En effet, plusieurs répliques de PCR augmentent la probabilité de détection des espèces cibles, de récupérer des taxa rares et ainsi de réduire les détections de faux négatifs (Ficetola *et al.* 2014; Dopheide *et al.* 2019). Cependant, le nombre de répétitions utilisées diffère souvent d'une étude à l'autre et dépend de divers facteurs tels que les objectifs de recherche, le choix de l'amorce, les contraintes de coût, la profondeur de séquençage, ou la plateforme de séquençage (Ruppert *et al.* 2019). Il est considéré qu'un minimum de deux répliques PCR pour les études de *metabarcoding* est nécessaire pour une meilleure estimation de la biodiversité (Pawlowski *et al.* 2020). Si l'objectif d'étude est de détecter une espèce spécifique, il convient d'augmenter le nombre de répliques. On peut également envisager d'ajouter davantage d'ADN à chaque réaction et d'effectuer moins de répliques (Pawlowski *et al.* 2020), tout en étant conscient qu'un excès d'ADN peut aussi inhiber une PCR.

L'amplification par PCR peut avoir de nombreuses sources de biais, tels que des substitutions et insertions introduites par l'enzyme polymérase, des substitutions causées par les changements de température durant les cycles de la PCR. Par conséquent, il est recommandé d'utiliser des ADN polymérases haute-fidélité pour réduire les taux d'erreur de PCR (Liu *et al.* 2020). On peut également observer des variations dans la capacité de liaison des amorces. En effet, l'hybridation des amorces peut être influencée par d'éventuels mésappariements avec la matrice ou la polymérase peut patiner davantage sur certaines régions du fait de la séquence ou de structures secondaires. Ainsi, certaines différences d'efficacité d'amplification sont difficiles à prédire ou à corriger (Falentin *et al.* 2019). Le choix des amorces a également le potentiel de biaiser les résultats en amplifiant certaines séquences plus que d'autres, ce qui peut être résolu par l'utilisation de plusieurs ensembles

d'amorces ayant pour cibles différentes régions variables dans la même réaction PCR, bien que cela puisse être plus coûteux et plus long (Ruppert *et al.* 2019). Cependant, cette approche peut également diminuer l'efficacité de la réaction PCR en engendrant certains biais si elle n'est pas convenablement réalisée (Pawlowski *et al.* 2020). Il faut noter par exemple que les amorces ont des températures de fusion spécifiques et qu'une mauvaise température engendre soit une dégradation ou des appariements non spécifiques (il est rare d'avoir plusieurs couples ayant des températures optimales identiques).

De plus, l'utilisation de marqueurs multiples ou alternatifs n'améliore pas nécessairement la détection des espèces (Elbrecht *et al.* 2019). Il faut savoir que les amorces n'amplifient pas tout l'ADN de la même façon dans l'échantillon, conduisant à un biais dans la proportion d'abondance de l'ADN des différentes espèces (Elbrecht *et al.* 2017; Piñol *et al.* 2019). Lors de l'exécution d'une PCR avant HTS, les amorces peuvent être marquées avec de courtes séquences nucléotidiques (« tag ») pour identifier de manière unique leur source lors d'un processus appelé multiplexage, ce qui peut apporter une source de biais dans les résultats (Ruppert *et al.* 2019).

En outre, on peut observer la formation de séquences dites chimériques (i.e. assemblages artéfactuels de plusieurs séquences) au cours de la PCR. Lors de l'élongation d'une séquence, il est possible que celle-ci se détache de sa matrice puis qu'elle soit utilisée comme amorce et se ré-hybride avec un autre brin matrice dans les cycles suivant, engendrant ainsi une chimère. Cette chimère, constituée de la séquence de plusieurs brins matrices ne provenant d'aucun organisme de l'échantillon, est donc présente dans la librairie de séquençage (i.e. ensemble des fragments d'ADN que l'on veut séquencer) et sera séquencée. Bien que leur abondance reste plus faible que les séquences réelles, elles peuvent conduire à la création de séquences chimériques numériquement abondantes (Falentin *et al.* 2019). Il y a maintenant des outils bio-informatiques performants qui excluent ces chimères, et le « twin-tagging » est un multiplexage d'échantillons efficace pour les filtrer avant même les programmes spécifiques (i.e. la même séquence de marquage a été attachée aux amorces avant et arrière d'une PCR, c'est ce « twin-tagging » qui permettra la détection et l'élimination des séquences avec saut de marquage, qui produisent des marquages non jumelés) (Yang *et al.* 2021). Bohmann *et al.* (2022) présentent quant à eux, une vue d'ensemble des trois principaux flux de travail pour le marquage spécifique des échantillons et la préparation des librairies dans les études de *metabarcoding* (avec les plateformes de séquençage Illumina).

Séquençage

Avant le séquençage, les répliques PCR vont être purifiées, quantifiées et regroupées pour préparer la librairie. Le multiplexage d'échantillons – ou séquençage multiplex – permet de regrouper et séquencer plusieurs échantillons (ou librairies) simultanément au cours d'une seule analyse, tout en créant des opportunités d'erreurs et de biais. Le multiplexage des échantillons est généralement réalisé grâce à l'incorporation d'indices nucléotidiques spécifiques à l'échantillon et d'adap-

tateurs de séquençage lors de l'amplification par PCR, ce qui permettra le démultiplexage (i.e. l'identification des échantillons) après le séquençage (Deiner *et al.* 2017 ; Pawlowski *et al.* 2020). Les séquences des fragments sont appelées des « reads » (ou lectures), et sont sauvegardées dans un fichier compressé (e.g., « .fasq.gz »).

Depuis le début du séquençage à haut débit, l'utilisation du *metabarcoding* comme outil de détection de la biodiversité a suscité un immense intérêt. Les plates-formes de séquençage à haut débit permettent désormais le séquençage rapide de l'ADN à partir de divers types d'échantillons. Les HTS permettent de séquencer plusieurs milliers de séquences différentes à la fois. Avec les progrès technologiques, les séquenceurs ont des capacités de séquençage, de nombre d'échantillons séquencés simultanément et de profondeur (Annexe 1) toujours plus élevées. Ces progrès créent de nouvelles opportunités pour les études de *metabarcoding* et la quantité de données a augmenté. Des erreurs telles que des substitutions, insertions ou délétions peuvent être introduites pendant le séquençage et conduisent à des erreurs de lecture. La technologie Illumina est connue pour introduire des séquences inattendues lors du séquençage d'échantillons témoins de composition connue qui ont été identifiées comme provenant d'autres bibliothèques séquencées dans le même « run » et seraient donc mal attribuées (Falentin *et al.* 2019).

Les plates-formes de séquençage à haute capacité permettent une évaluation rapide et complète sur le plan taxonomique de pratiquement tous les organismes d'un échantillon en utilisant une approche uniforme (Compson *et al.* 2020). Plusieurs plateformes sont disponibles pour cette étape de séquençage, comme le séquenceur ultra-portable MinION (Oxford nanopore technology) proposé à un tarif modéré (voir liste de prix sur le site : <https://store.nanoporetech.com/eu/priceList.html>, dernière consultation le 20 décembre 2023). Cet outil a une petite taille, une alimentation via un ordinateur portable et ne nécessite qu'un faible investissement en équipement (une connexion internet et la capacité de faire l'extraction, la PCR, et les étapes de préparation de la bibliothèque qui requièrent de la glace), ce qui permet le séquençage sur site (Runtuwene *et al.* 2019). Pour utiliser efficacement l'ADNe, les étapes d'extractions et de préparation de la bibliothèque nécessitent tout de même des investissements spécifiques dus aux problèmes de contamination. Cependant, la plate-forme de séquençage haut débit Illumina MiSeq (San Diego, Californie) est actuellement la plus populaire pour les études de *metabarcoding*. Elle offre une profondeur de séquençage raisonnable et de faibles taux d'erreur à un coût abordable (Liu *et al.* 2020). En effet, cette technologie présente un taux d'erreur inférieur à 0,4 %, (i.e. majoritairement dû à des substitutions), et les dernières améliorations ont permis d'obtenir des fragments de taille relativement importante avec un coût beaucoup plus faible (Falentin *et al.* 2019). Le MiSeq est capable de produire jusqu'à 30M de lectures de 2x300 pb. Un des avantages de ce séquençage est que la bio-informatique MiSeq peut être sous-traitée à des sociétés commerciales ou à des laboratoires dédiés à un coût raisonnable, et l'output des fichiers peut être directement traité par un opérateur ordinaire. Actuellement,

il n'y a pas de consensus sur la profondeur de séquençage nécessaire pour récupérer tous les taxa dans un échantillon donné. L'augmentation de la profondeur de séquençage peut augmenter dans une certaine mesure le taux de détection des taxa de faible abondance, mais elle est toujours soumise aux limites des instruments de séquençage et des coûts (Braukmann *et al.* 2019).

D'autres technologies de séquençage haut-débit existent avec différents niveaux de profondeurs de séquençage ou de débits comme le PGM Ion Torrent (séquences allant jusqu'à 400 pb), celle d'Oxford Nanopore ou de Pacific BioSciences (amplicons de plusieurs kb), mais avec un taux d'erreur encore plus élevé que la technique Illumina MiSeq présentée ci-dessus, i.e. jusqu'à 20 % pour l'ONT (Oxford Nanopore Technology ; Low & Tammi (2017) ; Creer *et al.* 2016 ; Falentin *et al.* 2019). Cependant, Braukmann *et al.* (2019) ont démontré des performances similaires pour trois plateformes (MiSeq, Ion Torrent PGM et Ion Torrent S5) dans la récupération d'espèces, bien que MiSeq soit généralement recommandé en raison de son taux d'erreur plus faible et de ses procédures bio-informatiques bien établies. Certaines technologies offrent le séquençage en paires (i.e. « paired-end »), telle que la technologie Illumina. Cela permet aux utilisateurs de pouvoir séquencer à partir des deux extrémités d'un fragment et de générer des données de séquence de meilleure qualité.

Par ailleurs, il est recommandé de soumettre les fichiers bruts des séquençages (i.e. fichiers FASTQ) avec des métadonnées standardisées au « Sequence Read Archive » (SRA) du NCBI (<https://www.ncbi.nlm.nih.gov/sra>, dernière consultation le 22 juin 2023), à l'European Nucleotide Archive (ENA, <https://www.ebi.ac.uk/ena/browser/home>, dernière consultation le 22 juin 2023), ou à d'autres bases de données nationales publiques avant publication. En France par exemple, le Pôle national de Données de Biodiversité (PNDB) produit, gère et analyse des données de biodiversité. En particulier, il fournit un accès aux jeux de données et de métadonnées, et facilite le partage des pratiques, des données et leur réutilisation avec les autres communautés de recherche (<https://www.pndb.fr/>, dernière consultation le 22 juin 2023). Cette pratique d'archivage dans des bases de données accessibles au public est courante dans les études génomiques, car elle permet de réanalyser ultérieurement les données avec de nouveaux outils informatiques (Creer *et al.* 2016 ; Deiner *et al.* 2017).

AUTRES APPROCHES MÉTHODOLOGIQUES

Abondance et quantification : qPCR et ddPCR

Une critique contre l'utilisation des approches génomiques est que les abondances relatives de lecture d'ADN séquencé ne sont pas fortement corrélées avec l'abondance ou la biomasse des espèces dans l'échantillon initial (voir partie « *Metabarcoding* d'une communauté »). Lorsque les études cherchent à quantifier la concentration initiale d'ADN dans un échantillon, elles peuvent avoir recours à la PCR quantitative en temps réel (qPCR) ou à la PCR digitale en micro-compartiments (ddPCR). Par exemple, la méta-analyse de Yates *et al.* (2019), basée sur des résultats de qPCR et ddPCR, a montré

que malgré une corrélation généralement plus faible dans les environnements naturels comparés aux études de laboratoire contrôlées, la concentration d'ADNe expliquait souvent une variation substantielle de l'abondance.

La qPCR ressemble à la PCR puisqu'elle consiste en une amorce spécifique à une espèce, mais l'addition d'un marqueur fluorescent assure l'émission de lumière fluorescente lorsque les produits PCR s'accumulent. La qPCR nécessite une machine quantifiant la fluorescence du produit PCR. Pour réaliser une qPCR, on peut soit utiliser un marqueur fluorescent, soit utiliser une sonde marquée par fluorescence. Pour réaliser une quantification absolue de l'ADN, il faut effectuer une série de dilutions avec un échantillon à la concentration d'ADN cible connue, qui vont permettre de déterminer la limite de détection (LOD, i.e. quantité minimale de séquence d'ADN cible pouvant être détectée dans l'échantillon) ainsi que la limite de quantification (LOQ, i.e. concentration minimale fournissant des niveaux acceptables de précision et d'exactitude pour la quantification) (Pawlowski *et al.* 2020). La sonde nucléotidique marquée qui doit se lier à la séquence cible doit correspondre à la séquence spécifique à l'espèce. Comme la quantification implique de faire correspondre les trois éléments (la sonde ainsi que les amorces sens et antisens), le test est plus difficile mais présente une spécificité et une sensibilité accrues (Pawlowski *et al.* 2020). La qPCR semble ainsi avoir un seuil de détection plus bas que la PCR en raison de sa sensibilité généralement plus élevée (Thomsen *et al.* 2012). Les approches qPCR utilisant une sonde sont couramment utilisées pour la détection d'espèces cibles (Biggs *et al.* 2015 ; Mauvisseau *et al.* 2018). En effet, l'approche qPCR va cibler de petits ensembles d'espèces d'intérêt et ne repose pas sur le séquençage de l'ADN tandis que l'approche de *metabarcoding* tire parti d'instruments de séquençage d'ADN à haut débit pour identifier simultanément des organismes dans divers groupes taxonomiques (Compson *et al.* 2020). La qPCR présente des avantages par rapport aux méthodes de détection traditionnelles basées sur la PCR, tels qu'une sensibilité et une spécificité plus élevées, une détection plus rapide, et la capacité de fournir des résultats quantitatifs (Yoo *et al.* 2017). La qPCR a par exemple été utilisée pour détecter une espèce de souris sauteuse (*Zapus hudsonius luteus* Miller, 1911) du Nouveau-Mexique sur du matériel végétal (Lyman *et al.* 2022). Cette étude a démontré que l'ADNe des mammifères peut persister sur la végétation des nids longtemps après la présence de l'animal, et pointe le potentiel de l'ADNe des plantes dans la détection d'espèces terrestres rares ou menacées. Keller *et al.* (2022) ont modélisé la détection par qPCR via l'ADNe et les données des pièges traditionnels pour estimer la densité du Crabe vert européen (*Carcinus maenas* Linnaeus, 1758). Ils documentent ainsi l'ADNe du crabe vert au-delà du front d'invasion précédemment connu et démontre que l'ADNe réduit l'incertitude sur les sites pour lesquels les données sont limitées.

Comme pour la qPCR, la PCR digitale (i.e. aussi appelée PCR digitale en micro-compartiments, ddPCR ou dPCR) est basée sur l'utilisation de marqueurs ou de sondes fluorescentes, mais la méthode de quantification est fondée sur une technique

différente. La ddPCR fractionne un échantillon individuel en milliers de partitions, où chaque compartiment contient – ou non – un brin d'ADN matrice et fonctionne comme un microréacteur PCR (Pawlowski *et al.* 2020). La PCR digitale permet d'amplifier chaque ADN cible, s'il est présent, dans un compartiment indépendant. Chaque subdivision produit ainsi un signal positif ou négatif selon la présence de la molécule d'intérêt, et l'analyse de la fluorescence permet de détecter ceux qui contiennent un ADN cible. En réalisant l'amplification de molécules d'ADN dans des compartiments indépendants, la dPCR en systèmes microfluidiques permet de déterminer la proportion d'amplifications PCR positives avec de hauts niveaux de sensibilités et de spécificités (Pawlowski *et al.* 2020). Cette approche permet de quantifier l'ADN sans établir de courbe standard avec les échantillons et certaines études suggèrent que la qPCR serait moins précise et plus onéreuse que la ddPCR (Hunter *et al.* 2017 ; Mauvisseau *et al.* 2019 ; Pawlowski *et al.* 2020). Par exemple, une étude a constaté que les erreurs d'analyse étaient plus faibles avec la dPCR que la qPCR et qu'elle fournissait des résultats plus précis pour l'abondance et la biomasse de l'espèce cible (Doi *et al.* 2015).

Métagénomique

La métagénomique consiste à fragmenter tous les ADN présents dans un échantillon mixte en courts fragments et à en séquencer la totalité grâce à un séquenceur haut débit (i.e. « Shotgun sequencing »). Les nombreuses séquences obtenues seront ensuite ré-assemblées en génome mitochondrial complet ou partiel via des pipelines bio-informatiques afin de reconstruire les génomes d'origine. Cette approche présente ainsi une alternative pour la caractérisation de communautés complexes telles que les échantillons mixtes (Crampton-Platt *et al.* 2016). Elle a notamment été utilisée pour produire des séquences génomiques mitochondriales (i.e. mitogénomique) d'assemblages de coléoptères, permettant l'analyse simultanée de la diversité des espèces, de la structure phylogénétique et des facteurs de diversification des communautés de coléoptères du sol (Andújar *et al.* 2015). De plus, Arribas *et al.* (2016) ont utilisé une combinaison de pipelines de séquençage par PCR (*metabarcoding*) et de métagénomique pour caractériser la diversité des sols dans un contexte phylogénétique et écologique communautaire. Tang *et al.* (2015) ont montré que la métagénomique et le reséquençage des génomes mitochondriaux (i.e. mitogénomique) peuvent être appliqués avec succès à des échantillons en vrac d'abeilles sauvages. Les auteurs concluent que la mitogénomique permet d'automatiser l'identification taxonomique, même pour les espèces cryptiques, et de suivre les changements dans la richesse et de la répartition des espèces. La comparaison de la similarité des données métagénomiques mitochondriales des collemboles avec une base de données mondiale de (Cicconardi *et al.* 2017) a permis de révéler que près de 30 % des espèces insulaires échantillonnées sont génétiquement identiques, ou presque, à des individus échantillonnés dans des régions géographiques souvent très éloignées du monde. Ces schémas de parenté de l'ADNmt impliquent des introductions

d'espèces par l'homme, avec des estimations minimales de la proportion d'espèces introduites sur les îles échantillonnées allant de 45 % à 88 %.

Contrairement aux protocoles de *metabarcoding*, la métagénomique ne repose pas sur l'amplification de séquences génomiques spécifiques. Elle se fait sans PCR, évitant ainsi les potentiels biais introduits par les procédures de la PCR (Creer *et al.* 2016). Le principal facteur limitant l'efficacité de la métagénomique est lié à la profondeur de lecture qui doit être adéquate pour obtenir des résultats précis (ce qui peut être difficile à obtenir à partir d'échantillons complexes comme le sol). L'augmentation de la puissance de séquençage nécessaire pour obtenir cette profondeur de lecture entraîne souvent des surcoûts importants qui peuvent être prohibitifs (Francioli *et al.* 2021). L'analyse des ensembles de données métagénomiques conduit à des bases de données plus conséquentes que celles produites par les approches de *metabarcoding*, ce qui entraîne des analyses plus longues à réaliser, et requiert une plus grande puissance de calcul et d'expertise. La métagénomique peut également être utilisée pour séquencer tous les fragments d'ADNe présents dans un échantillon sans se concentrer sur un groupe taxonomique, un locus ou une longueur de fragment spécifique.

Srivathsan *et al.* (2015) ont comparé les performances de la métagénomique et du *metabarcoding* dans un contexte d'étude de régime alimentaire à partir de fèces de primates. Ils ont observé qu'avec la métagénomique, les taxons dominants étaient identifiés avec une plus grande résolution, tandis que les données de *metabarcoding* permettaient une meilleure couverture pour les taxons rares dans le régime alimentaire. Chua *et al.* (2021) quant à eux, indiquent que la métagénomique a le potentiel de devenir la technique de référence pour la reconstruction du régime alimentaire dans le futur, en particulier lorsqu'une base de données de référence localisée n'est pas disponible ou lorsque des informations supplémentaires sont nécessaires (e.g., sur les parasites, la génétique de l'hôte et le microbiome). Cependant, ils soulignent plusieurs limitations possibles dans l'utilisation de la métagénomique dans les études sur les régimes alimentaires :

- le coût ;
- les défis bioinformatiques ;
- le type de question de recherche ;
- la disponibilité d'échantillons fécaux frais ;
- l'exhaustivité de la base de données de référence utilisée.

Dans l'étude de Becker & Pushkareva (2023), la diversité des bactéries du sol a été analysée *via metabarcoding* et métagénomique à partir d'échantillons d'ADN provenant de sites de l'Arctique et de l'Antarctique. Ils ont trouvé que presque toutes les unités taxonomiques opérationnelles trouvées dans les analyses de *metabarcoding* étaient retrouvées dans les analyses métagénomiques, mais qu'à l'inverse la métagénomique avait permis d'identifier un grand nombre d'OTU supplémentaires qui étaient absentes des analyses de *metabarcoding*. Ils ont également observé des différences dans l'abondance des OTUs entre les deux méthodes. Les auteurs concluent que les analyses métagénomiques sont mieux adaptées que le *metabarcoding* pour étudier la biodiversité des échantillons de sol terrestre.

TRAITEMENT DES DONNÉES DE SÉQUENÇAGE ET IDENTIFICATION DES ESPÈCES

ANALYSE ET OUTILS BIO-INFORMATIQUES

Suite au séquençage, les lectures sont sauvegardées dans un fichier « .fastq » dans lequel elles sont associées avec leurs scores de qualité phred (qui évaluent la confiance du séquençage). Les données de séquence brutes devront être traitées en plusieurs étapes bio-informatiques. En premier lieu, il faudra faire une sélection des séquences selon leur qualité et nettoyer les séquences afin d'enlever le bruit de fond apporté par de potentiels biais méthodologiques. Dans un second temps, les séquences sont regroupées en clusters en fonction de leur similarité et assemblées en unités taxonomiques opérationnelles (Deiner *et al.* 2017). Une classification des séquences à plusieurs niveaux taxonomiques hiérarchiques : royaume, phylum, classe, ordre, famille, genre, etc. sera effectuée grâce à la comparaison de ces OTUs à des séquences connues provenant de bases de données de référence. Les résultats vont ainsi consister en des listes de tous les taxa détectés dans un échantillon donné. Grâce à un séquençage ultra-profond et à de nouveaux outils bio-informatiques, le *metabarcoding* peut générer une liste taxonomique des espèces présentes dans un échantillon assez rapidement (Compson *et al.* 2020). Les analyses bio-informatiques des données de séquençage peuvent toutefois être assez complexes et décourageantes pour les utilisateurs.

Une des premières étapes de prétraitement des données de séquençage consiste au « rognage » des séquences pour enlever divers éléments comme les adaptateurs de séquençages si on part des fichiers bruts, les tags ou encore les amorces. Par la suite, il y a un « filtering » afin d'éliminer les lectures qui sont trop courtes (i.e. information insuffisante pour l'assignation taxonomique) ou trop longues (i.e. pouvant être des chimères), qui contiennent des bases inconnues, n'ont pas d'amorces ou avec des erreurs dans les amorces (Deiner *et al.* 2017). Lors de ce filtre, les séquences de mauvaise qualité (i.e. avec des scores Phred faibles) sont également supprimées. Les scores Phred calculés par les plateformes de séquençage peuvent également servir à déterminer les probabilités d'erreur maximales et être utilisés dans le filtrage de qualité. La sélection d'un niveau de filtrage d'erreur maximal de 1 ou 0,5 (plus l'erreur considérée est faible, plus la limite est stricte) est courante dans certaines études sur les macro-organismes (Deiner *et al.* 2017). Certaines technologies, comme l'Illumina, utilisent un séquençage en paires (i.e. « paired-end ») où le séquençage se fait dans les deux sens. Comme la comparaison des séquences permet d'aligner les parties qui se recouvrent partiellement ou chevauchantes, il est alors possible de relier les séquences qui se chevauchent en enchainements appelés *contigs* et de les fusionner. De plus, plusieurs échantillons sont généralement regroupés au cours d'un séquençage et chaque échantillon est caractérisé par un court fragment d'ADN (i.e. avec un « tag »). Le démultiplexage est une étape qui inclut l'identification des séquences (à l'aide de ces tags pour pouvoir associer les lectures à l'échantillon d'origine) et leur séparation en échantillons d'origine, et le *trimming* (i.e. rognage) de tous

les tags et adaptateurs. La déréduplication est le processus dans lequel les séquences filtrées par qualité sont regroupées en un ensemble de lectures uniques et qui permet ainsi de garder un seul exemplaire d'une séquence (Falentin *et al.* 2019). Le débruitage consiste à corriger les erreurs de séquençage à l'aide de modèles d'erreur (e.g., <https://github.com/biocore/deblur>, dernière consultation le 22 juin 2023).

Il existe deux principes de déchimérisation qui permettent d'éliminer les séquences dites chimériques (Falentin *et al.* 2019). Il est possible de détecter une chimère en alignant les lectures obtenues avec une base de données de référence et d'éliminer celles qui génèrent des alignements partiels sur deux (ou plus) brins appartenant à des taxa différents (i.e. déchimérisation sur référence). L'inconvénient est qu'elle n'est pas très bien adaptée pour des espèces d'écosystèmes peu connus dont la plupart des séquences ne sont pas dans les bases de données. La deuxième possibilité est de détecter les chimères en alignant chaque lecture sur les autres lectures de l'échantillon, puis d'écarter celles qui génèrent des alignements partiels (i.e. sur une partie de la longueur) avec plusieurs autres lectures différentes (i.e. déchimérisation de novo qui suppose que les séquences « parentales » de la chimère sont présentes dans l'échantillon mais en plus grande quantité que la chimère elle-même).

Les contrôles sont essentiels dans les analyses basées sur le séquençage de l'ADN dû au grand nombre d'étapes de laboratoire et d'informatique entre l'échantillonnage et les résultats finaux, et au grand nombre d'échantillons traités en parallèle. Les contrôles négatifs sont des réactions sans ADN qui peuvent être incorporés pendant les étapes d'extraction, de PCR et de préparation de bibliothèques. Les contrôles positifs peuvent être composés d'un seul ou d'un mélange d'extraits d'ADN provenant de taxa connus (idéalement de l'ADN provenant de taxa qui ne sont pas censés être présents dans le lieu d'échantillonnage), appelés communautés fictives (Alberdi *et al.* 2019). Les contrôles expérimentaux peuvent alors jouer un rôle majeur dans la détermination des seuils, car ils permettent d'estimer la proportion d'artefacts dans un ensemble de données. Un nombre croissant d'études ont utilisé le séquençage de contrôles positifs pour établir un niveau seuil et les répliques techniques peuvent également être utilisées pour évaluer la cohérence (e.g., De Barba *et al.* 2014; Port *et al.* 2016; Corse *et al.* 2017, 2019; Deiner *et al.* 2017; Galan *et al.* 2018). En effet, afin de contrôler la contamination en laboratoire ou celle des clusters sur l'instrument de mesure, certaines études utilisent des méthodes de filtrage pour réduire la prévalence des artefacts et éliminer les séquences de faible abondance (i.e. MSCTs, « Minimum sequence copy thresholds »). Le seuil devra être choisi avec soin car les seuils bas ne parviendront pas à éliminer les artefacts (i.e. faux positifs) tandis que les seuils élevés risquent d'éliminer trop de données (i.e. faux négatifs). Normalement, les contrôles négatifs ne devraient pas contenir d'ADN et les contrôles positifs devraient contenir que de l'ADN provenant de taxa sélectionnés. C'est toutefois rarement le cas (e.g., contamination, mauvaise assignation), et ces lectures inattendues facilitent la détermination des seuils optimaux pour le nettoyage des données (Drake *et al.* 2022). Par exemple, l'utilisation d'un niveau d'erreur défini

par un contrôle positif permet d'identifier l'abondance des séquences dans l'échantillon de contrôle qui appartiennent à des taxa non ciblés et peuvent résulter d'erreurs telles que la contamination (Deiner *et al.* 2017). Villsen *et al.* (2022) ont ainsi déterminé plusieurs seuils d'abondance et de fréquence à partir des données des contrôles négatifs et positifs pour réduire les faux positifs dans les échantillons de matières fécales. Cependant, il est important de noter que l'application de ces seuils et l'utilisation de contrôles pour l'évaluation des seuils restent ambiguës et non standardisées, et on observe de nombreuses études employant des méthodologies et des seuils totalement distincts (Drake *et al.* 2022). Si des avancées dans la standardisation des protocoles sur le terrain et au laboratoire a fait l'objet de plusieurs travaux et a été discutée par plusieurs groupes de travail nationaux mais aussi européen, la standardisation des méthodes de filtration des données reste encore peu explorée. Les lectures sont alors regroupées en cluster avec un fort pourcentage d'identité et toutes les lectures du même cluster seront par la suite assignées taxonomiquement.

Les séquences peuvent être comparées et leurs relations seront exprimées en pourcentage de similarité ou de divergence entre les bases nucléiques qui les composent. De ce fait, les séquences ayant un pourcentage de divergence inférieur – ou de similarité supérieur – à un seuil donné se verront regroupées au sein d'une même unité taxonomique opérationnelle. Par exemple, un critère de similitude de 97 % est communément utilisé pour différencier deux espèces (Falentin *et al.* 2019). Par ailleurs, les MOTUs peuvent être utilisés dans les suivis de biodiversité sans passer par l'étape d'assignation taxonomique. Suite à leur création, un filtre basé sur l'abondance des MOTUs est généralement appliqué afin d'éliminer celles qui ont une faible abondance ou qui n'apparaissent qu'une seule fois (i.e. *singletons*) dans les données, car ce sont potentiellement des taxa rares, des faux positifs, dus à une faible contamination ou des chimères non supprimées préalablement (Deiner *et al.* 2017). Par exemple, Bokulich *et al.* (2013) conseillent de supprimer les OTUs dont l'abondance dans les données est inférieure à 0,005 %. Par ailleurs, il est possible d'appliquer un filtre de prévalence et de ne sélectionner que les OTUs présents dans X % des répliques, pour ne conserver que ceux dont la présence est répétée. De façon similaire, si on dispose de répliques techniques, il est possible de déterminer le seuil à partir duquel les MOTUs identifiées sont liées aux biais techniques, et retirer ceux dont l'abondance est inférieure à ce seuil dans tous les échantillons. Par conséquent, les filtres sont considérés comme une étape importante car ils réduisent le nombre de faux MOTUs, facilitent l'interprétation biologique des résultats et accélèrent les traitements ultérieurs (Falentin *et al.* 2019).

Les variants de séquences d'amplicon (ASV, Amplicon Sequence Variants) permettent d'éviter l'étape de clustering puisque le but est de corriger le bruit introduit dans les fragments séquencés plutôt que d'agglomérer les séquences proches. Étant donné que les lectures appartiennent initialement toutes au même groupe, si la diversité des lectures au sein de ce groupe est plus grande que la diversité attendue uniquement en présence d'erreurs de séquençage, ce groupe sera divisé en sous-groupes plus homogènes, et cela jusqu'à ce que chaque

sous-groupe soit homogène (i.e. correspondant à une lecture dont les variations sont engendrées par les erreurs de séquençage). Ces approches permettent d'obtenir des MOTUs très fines, mais sont très sensibles à des biais systématiques lors du séquençage (Falentin *et al.* 2019).

Finalement, l'assignation taxonomique se fait en comparant la similarité des MOTUs ou directement avec les lectures restantes après filtrage de qualité à des bases de données de séquences de référence publiques ou personnelles. En effet, la base de données de référence peut être constituée de séquences accessibles au public ou de séquences de référence conçues par l'étude selon le taxon et le marqueur utilisé (Deiner *et al.* 2017). Ces méthodes d'assignation se font souvent grâce à l'alignement de séquences. Nous n'en discutons que brièvement ici, mais voir l'étude de Deiner *et al.* (2017) qui résume les diverses approches utilisées dans cet aspect important du pipeline bio-informatique et fournit diverses références. Habituellement, cela se fait *via* les interfaces web BOLD, BLAST ou MIDORI. BLAST (i.e. Basic Local Alignment Search Tool) repose sur la base de données du NCBI et où le paramétrage concerne les seuils minimaux de pourcentage d'identité et de séquence couverte que l'on autorise pour l'assignation taxonomique (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>, dernière consultation le 22 juin 2023). Si la base de données de référence contient les séquences des espèces, alors les séquences provenant d'échantillons peuvent être identifiées au niveau de l'espèce. En raison du caractère incomplet des actuelles bases de données de référence, l'identification ne peut être réalisée dans certains cas qu'à des niveaux taxonomiques supérieurs (e.g., l'attribution se fait au niveau de la famille ou de la classe), ou il est également possible pour la séquence d'être affectée à une OTU particulière.

Parmi les algorithmes disponibles pour analyser les données de *metabarcoding*, la plupart sont proposés avec une interface de ligne de commande, tandis que d'autres disposent d'une interface graphique plus conviviale. Cela implique d'avoir un minimum de connaissances en programmation de base telles que la capacité d'utiliser, manipuler ou écrire des programmes en langage Shell bash (GNU) ou Python pour analyser de gros volumes de données, et une certaine familiarité avec R (Creer *et al.* 2016). Il est recommandé aux débutants de consulter des experts en bio-informatique pour éviter les erreurs. Un pipeline bio-informatique générique pour les données de *metabarcoding* se compose de cinq étapes principales : demultiplexage des échantillons, fusion des lectures de paires, filtrage de la qualité, curation OTU et attribution taxonomique (Liu *et al.* 2020). Il est donc possible de devoir faire appel à plusieurs programmes bio-informatiques distincts pour analyser des données brutes provenant du séquençage. Il existe de nombreux exemples de programmes et ressources informatiques (Tableau 1) pouvant être utilisés dans les diverses tâches liées à l'identification des espèces (voir aussi Mathon *et al.* 2021).

L'équipement technique, l'expertise en laboratoire et informatique nécessaire à l'utilisation de ces types de méthodes en font des approches qui restent hors de portée pour de nombreux chercheurs sans partenariats avec des laboratoires.

Cependant, avec le développement du nombre de laboratoires génomiques commerciaux, il est possible de sous-traiter différentes étapes telles que la partie bio-informatique (Compson *et al.* 2020 ; Liu *et al.* 2020).

BASES DE DONNÉES DE RÉFÉRENCE

Une étape importante pour le succès des méthodes de *metabarcoding* dans les suivis de biodiversité est l'assignation taxonomique et donc l'existence de bases de données de référence qui associent des données génétiques à des spécimens identifiés. Ces bases de données vont permettre d'identifier les spécimens inconnus obtenus à partir du séquençage de marqueurs par comparaison de séquences. Un des obstacles majeurs à l'utilisation de la méthodologie du *metabarcoding* dans les études de suivis de biodiversité est l'incomplétude des bases de données de référence. Par exemple, les lacunes de la bibliothèque de référence du biote aquatique actuellement utilisée pour la biosurveillance en Europe varient d'un groupe taxonomique à l'autre et d'un pays à l'autre (Weigand *et al.* 2019 ; Specchia *et al.* 2020). De même, les analyses d'exhaustivité se sont avérées très faibles (3-37 %) pour des espèces d'abeilles des États-Unis en utilisant les enregistrements de spécimens d'abeilles actuellement disponibles (Cheshire *et al.* 2023). La couverture des bases de données peut être à la fois incomplète et inégale, et présente des variations basées sur le groupe taxonomique, la région géographique, ou encore sur la couverture génétique (e.g., COI, 16S, 18S) (Compson *et al.* 2020). Diverses initiatives nationales et internationales permettent de faire avancer la recherche sur les codes-barres ADN à travers des projets tels que l'International Barcode of Life (iBOL, <https://ibol.org/>, dernière consultation le 22 juin 2023), ainsi que des versions régionales et nationales qui vont aider à la création de bases de données de codes-barres ADN. En effet, ce type de projet va générer de nombreuses séquences qui pourront être ensuite compilées sur une plateforme bio-informatique (i.e. Barcode of Life Data System). Ainsi, plusieurs bases de données internationales et nationales existent en ligne et continuent à croître régulièrement (Tableau 2). Il est important de faire attention à la qualité des séquences dans les bases de données, notamment pour celles qui ne sont pas nettoyées (e.g., Genbank et MIDORI) puisque cela peut entraîner des erreurs d'assignation taxonomique.

MÉTHODES GÉNÉTIQUES

ET PROGRAMMES DE SCIENCES CITOYENNES

Le recours à des volontaires pour aider et participer aux suivis de la biodiversité est en augmentation dans le monde et peut apporter une contribution précieuse aux suivis à long terme. L'ADNe peut être considéré comme un outil écologique et de conservation très utile lorsqu'il est utilisé en combinaison avec d'autres sources de données provenant, par exemple, des sciences citoyennes. Cela permet d'échantillonner de vastes zones géographiques avec un minimum d'efforts tout en étant non invasif, ce qui est une caractéristique essentielle pour le suivi et la conservation des espèces (ou des écosystèmes) menacés (Pascher *et al.* 2022). Bien que les compétences requises pour réaliser les étapes de laboratoire

TABLEAU 1. — Exemples de programmes et ressources informatiques pouvant être utilisés dans les diverses étapes liées à l'identification des espèces basée sur le *metabarcoding*. Les sites internet ont été consultés pour la dernière fois le 22 juin 2023.

Nom du programme et ressource	Type et fonction
MBRAVE – Multiplex Barcode Research and Visualisation Environment https://www.mbrave.net/	Plateforme prenant en charge le stockage, la validation, l'analyse et la publication de projets hautement multiplexés basés sur des instruments de séquençage haut débit.
DAMe https://github.com/MLZM-lab/DAMe	Boîte à outils (ensemble de scripts python) conçue pour effectuer les premières étapes de l'analyse d'ensembles de données très complexes générés par des méthodes de <i>metabarcoding</i> qui utilisent des amplicons à double marquage, des répliques PCR, le multiplexage, le regroupement d'échantillons et le séquençage avec des technologies HTS.
Galaxie https://training.galaxyproject.org/training-material/topics/introduction/tutorials/galaxy-intro-101/tutorial.html?utm_source=redirect&utm_medium=learn&utm_campaign=galaxyhub	Plateforme Web open source pour la recherche biomédicale à forte intensité de données. L'interface contient des outils modulaires qui peuvent être facilement déployés dans un flux de travail individuel par des utilisateurs sans expérience en programmation.
Qiime2studio – q2studio https://docs.qiime2.org/2023.5/	Package d'analyse permettant aux chercheurs de commencer une analyse avec des données brutes de séquence d'ADN et de terminer avec des chiffres et des résultats statistiques. La communauté d'utilisateurs de QIIME est active et de nombreux tutoriels existent (http://qiime.org/tutorials/index.html).
slim https://trtrcd.github.io/SLIM/	Application Web dont l'objectif est de faciliter l'accès à des outils bioinformatiques de pointe aux non-spécialistes (i.e. passe des données brutes du <i>metabarcoding</i> à une matrice OTU annotée). Elle simplifie la création et le déploiement d'un pipeline de traitement et fournit une boîte à outils intégrative permettant le traitement des données étape par étape.
Frogs – Find Rapidly Otus with Galaxy Solution http://frogs.toulouse.inra.fr/	Pipeline convivial pris en charge par Galaxy qui analyse de grands ensembles de données de séquençage d'amplicons d'ADN. Il prend en charge la fusion, le nettoyage, la déplication de séquences et la visualisation des résultats statistiques en aval. Il produit une table des OTUs avec leurs affiliations taxonomiques.
DADA2 https://benjjneb.github.io/dada2/	Package dada2 est un pipeline open source qui fournit un ensemble complet d'outils permettant de prendre des données brutes de séquençage d'amplicon jusqu'à une table de caractéristiques représentant la composition de l'échantillon, et a l'avantage de proposer des ASV. Dada2 est un package R qui est utilisable depuis qiime2 ou directement sous R. https://benjjneb.github.io/dada2/tutorial.html
CutAdapt https://cutadapt.readthedocs.io/en/stable/index.html	Il permet le nettoyage des données en trouvant et supprimant divers types de séquences indésirables des lectures de séquençage à haut débit.
Mothur https://mothur.org/	Logiciel open source développé pour répondre aux besoins bio-informatiques de la communauté et analyser les données moléculaires utilisées par les écologistes microbiens. Mothur offre la possibilité de passer des séquences brutes à la génération d'outils de visualisation pour décrire la diversité α et β .
USEARCH	Outil d'analyse de séquence et propose de nombreux algorithmes de recherche et de clustering qui sont souvent plus rapides que BLAST et gratuit pour la plupart des utilisateurs (i.e. les licences d'utilisation de USEARCH 32 bits sont offertes à tous, https://www.drive5.com/usearch/) UPARSE (Edgar 2013) est une méthode pour générer des clusters (OTUs) à partir de lectures de séquençage de nouvelle génération de gènes marqueurs grâce à l'algorithme UPARSE-OTU, qui est implémenté en tant que commande cluster_otus dans USEARCH (https://drive5.com/uparse/)
MG-RAST – Metagenomic Rapid Annotations using Subsystems Technology https://www.mg-rast.org/	Serveur web open-source qui propose une analyse phylogénétique et fonctionnelle des données métagénomiques.
DNA Subway https://learning.cyverse.org/dna_subway_guide/ https://dnasubway.cyverse.org/#	Plateforme de bio-informatique éducative développée par CyVerse. Il regroupe des outils bio-informatiques de recherche, permet d'annoter des séquences d'ADN, d'effectuer des analyses phylogénétiques et d'analyser les données de séquençage et avec une interface facile à utiliser.
Obitools https://git.metabarcoding.org/obitools/obitools/wikis/home	Package avec un ensemble de programmes spécialement conçus pour analyser les données NGS dans un contexte de <i>metabarcoding</i> d'ADN, capable de filtrer et éditer des séquences tout en tenant compte des annotations taxonomiques (Boyer <i>et al.</i> 2016). Il est distribué sous forme de logiciel open source.
JAMP (Just Another Metabarcoding Pipeline) https://github.com/VascoElbrecht/JAMP	Pipeline de <i>metabarcoding</i> modulaire, intégrant différentes fonctions de USEARCH, VSEARCH, CUTADAPT et d'autres programmes. Le pipeline est exécuté en tant que package R et génère automatiquement les dossiers nécessaires et les statistiques récapitulatives.
Barque https://github.com/enormandeaubarque	Pipeline d'analyse de <i>metabarcoding</i> d'ADNe qui annote les lectures, au lieu d'unité taxonomique opérationnelle (OTU), en utilisant des bases de données de codes à barres de haute qualité. Barque peut également produire des OTUs, qui sont ensuite annotées à l'aide d'une base de données.

TABLEAU 1. — Suite.

Nom du programme et ressource	Type et fonction
Phyloseq https://github.com/joey711/phyloseq	Package qui permet d'importer, stocker, analyser et afficher graphiquement des données de séquençage phylogénétique complexes qui ont déjà été regroupées en OTUs.
Swarm https://github.com/torognes/swarm	Méthode de clustering robuste et rapide pour les études basées sur les amplicons. Il fournit un algorithme de clustering qui gère des ensembles massifs d'amplicons.
PEMA https://github.com/hariszaf/pema	Ensemble d'outils d'analyse de <i>metabarcoding</i> qui nécessitent peu d'efforts lors de la configuration, de l'exécution et de la personnalisation en fonction des besoins des chercheurs. Basé sur des outils tiers, il effectue un prétraitement des lectures, un regroupement des MOTUs, une inférence des variants de séquence d'amplicon (i.e. ASV) et une attribution de taxonomie pour l'ARN ribosomal 16S et 18S, ainsi que des données de gènes marqueurs ITS et COI (Zafeiropoulos <i>et al.</i> 2020). Enfin, PEMA prend en charge l'analyse écologique en aval des profils récupérés, facilitée par le package phyloseq R.
ONTrack https://github.com/MaestriSi/ONTrack	Pipeline de barcoding rapide et précis basé sur les lectures de séquences provenant du séquenceur MinION. Ce pipeline est capable de fournir des séquences consensus précises en ~ 15 minutes par échantillon sur un ordinateur portable standard (Maestri <i>et al.</i> 2019). De plus, un pipeline de prétraitement est fourni pour que toute l'analyse bio-informatique soit simple et directe.
ASHURE https://github.com/BBaloglu/ASHURE	Pipeline Python conçu pour le traitement des données, la construction de consensus, le regroupement et l'affectation taxinomique des lectures provenant de dispositifs de séquençage à nanopores comme le séquenceur MinION (Baloglu <i>et al.</i> 2021)

TABLEAU 2. — Bases de données de référence existantes et pouvant être utilisées dans l'identification des espèces basée sur le *metabarcoding*. Les sites internet ont été consultés pour la dernière fois le 22 juin 2023.

Nom de la base de données	Ressource
GenBank	Référentiel général de données de séquence (Sayers <i>et al.</i> 2019). www.ncbi.nlm.nih.gov/genbank/
Barcode of Life Data Systems – BOLD	Référentiel de données de séquences orienté sur les groupes taxonomiques (Ratnasingham & Hebert 2007). http://www.barcodinglife.org/
MIDORI	Base de données de référence de séquences d'ADN, qui peut être utilisée pour les attributions taxonomiques de toutes les séquences d'ADN mitochondrial eucaryote. https://reference-midori.info/index.html
SILVA	Une des banques de référence d'ARNr 16S et 18S les plus riches. La base de données LTP (All-Species Living Tree Project) est une sous-partie « nettoyée » de Silva. Les séquences répertoriées correspondent uniquement à des souches types d'espèces bien classifiées d'Archées et de Bactéries. https://www.arb-silva.de/
RDP – Ribosomal Database Project	C'est à la fois une banque de données et un ensemble d'outils. Cette banque est pour l'instant non nettoyée et regroupe plus de trois millions de séquences, la rendant lourde à manipuler. https://bio.tools/rdp
Greengenes	Base réputée pour sa rigueur, avec des séquences vérifiées manuellement et ne contenant aucune chimère, sa représentativité pour tous les phyla est très bonne. https://greengenes.secondgenome.com/
EZBioCloud	Base de données regroupant des séquences 16S curées ou extraites de génomes complets provenant du NCBI et du JGI (Yoon <i>et al.</i> 2017). Les fichiers de séquence et de taxonomie sont disponibles sur demande après enregistrement et gratuits pour les académiques. https://www.ezbiocloud.net/?bannerId=6
UNITE	Base de référence, assez complète et diversifiée pour les champignons, qui contient un peu plus de 800 000 séquences d'ITS. Les fichiers préformatés pour Qiime, Mothur et Usearch sont mis à disposition sur le site web. https://unite.ut.ee/index.php
Diat.barcode	Bibliothèque de codes-barres pour diatomées et autres algues principalement dédiée au marqueur rbcL des diatomées. https://www6.inrae.fr/carrtel-collection/Barcoding-database/Description ; https://github.com/fkeck/diatbarcode

et de bio-informatique du *metabarcoding* sont souvent trop complexes pour les scientifiques citoyens, ils deviendront certainement une grande composante de l'échantillonnage sur le terrain (Compson *et al.* 2020). La science citoyenne peut être utilisée comme une source de collecte d'échantillons pour l'analyse de l'ADNe afin d'évaluer la diversité des espèces

dans une zone particulière pour contribuer au suivi et à la conservation des espèces. Lavin 2022 a montré que les contaminations des échantillons d'ADNe collectés le long de deux rivières par les citoyens étaient similaires aux contaminations des échantillons collectés par des professionnels. Des projets nationaux – incluant des professionnels et des non-profession-

nels – sont notamment en train d'émerger dans divers pays, tels que les États-Unis (CaleDNA Initiative, <https://ucedna.com/>, dernière consultation le 22 juin 2023) et le Canada (Sequencing the Rivers for Environmental Assessment and Monitoring, STREAM, <https://stream-dna.com/>, dernière consultation le 22 juin 2023), pour collecter des échantillons de *metabarcoding* d'ADNe. Dans le cadre de CaleDNA et grâce aux résultats provenant des divers échantillons (sol, sédiments et eau) partagés en ligne, il est possible de découvrir les patrons de biodiversité de la Californie qui ne recourent que légèrement ceux des suivis plus traditionnels (Meyer *et al.* 2021). Srivathsan *et al.* (2021) dévoilent quant à eux un flux de travail de *barcoding* utilisant le séquenceur portable MinION et un nouveau logiciel, ONTbarcode (aide aux défis bio-informatiques posés par les lectures MinION). Cette méthodologie ne sollicite qu'un équipement de laboratoire minimal et a l'avantage de présenter un apprentissage rapide (en quelques jours), une réduction des coûts de séquençage (< 10 cents par codes-barres), et d'obtenir rapidement la séquence. Ils décrivent notamment comment les amplicons marqués peuvent être obtenus et séquencés en temps réel sur le terrain, dans des laboratoires de biodiversité, dans le cadre de science citoyenne, ou encore dans des écoles. Les auteurs concluent que cette méthodologie favorisera probablement l'utilisation généralisée de cet outil pour les études de suivi de la biodiversité pour les agences gouvernementales, les universités, les musées et les écoles. L'intervention des sciences participatives peut être grandement bénéfique en impliquant les citoyens, en les éduquant sur ces sujets et en diffusant les connaissances scientifiques au grand public.

Les études basées sur les analyses moléculaires ouvrent de nouvelles opportunités très diverses pour des programmes de science citoyenne et permettent de faire intervenir des bénévoles au niveau de la récolte d'échantillons ou du processus d'identification. Par exemple, des échantillons de coléoptères aquatiques ont été collectés au Monténégro grâce à des scientifiques citoyens, des étudiants et des taxonomistes, qui ont été séquencés par la suite (Freitag *et al.* 2021). Un programme de science citoyenne (i.e. « Conservation Scats ») a notamment été créé pour collecter des échantillons d'excréments pour décrire les régimes alimentaires et l'utilisation de l'habitat de divers méso-carnivores avec une approche de *metabarcoding* ADN (Smith *et al.* 2018). Par ailleurs, des piquants perdus par une espèce de porc-épic ont été collectés par des observateurs citoyens comme sources d'ADN aisément disponibles pour contribuer à une mise à jour de la carte de répartition de ce rongeur (Mori *et al.* 2017). Un autre projet de science citoyenne (i.e. le « Peninsular Malaysia Butterfly Count ») a fait intervenir des participants pour participer au décompte d'espèces de papillons et échantillonner – de manière non létale – des pattes de spécimens afin d'identifier les espèces par code-barres ADN et établir une base de référence pour de potentiels futurs suivis de changements de communautés (Wilson *et al.* 2015). Une approche de science citoyenne avec un protocole simple (i.e. photos, collecte de spécimens et données de végétation) a également permis d'évaluer l'abondance et l'occurrence de limaces invasives dans les jardins autrichiens

et d'identifier les facteurs environnementaux sous-jacents à leur occurrence (Dörler *et al.* 2018). Les échantillons d'eau nécessaires à l'analyse d'ADNe ont l'avantage d'être relativement simples et rapides à collecter. L'étude de Biggs *et al.* (2015) a ainsi montré que des bénévoles ont pu collecter avec succès des échantillons d'ADNe aquatiques pour détecter la présence ou l'absence de tritons, après une formation limitée.

En plus des utilisations typiques de taxonomie et de suivis de biodiversité, les code-barres ADN peuvent être un outil important pour enseigner et apprendre aux étudiants la méthode et la recherche scientifique. Par exemple, des étudiants ont été impliqués dans l'évaluation de la biodiversité du parc marin de New York avec la méthode de codes-barres ADN (Marizzi *et al.* 2018). Ces scientifiques citoyens ont été rapidement formés pour identifier les espèces à l'aide de codes-barres ADN, et en effectuant des recherches NCBI GenBank BLAST, les étudiants ont identifié taxonomiquement de nombreux échantillons et ont publié de nouveaux codes-barres ADN sur GenBank. Dans une autre recherche récente, des étudiants ont contribué à une meilleure compréhension des impacts d'une espèce envahissante, à travers l'étude de son régime alimentaire basée sur les code-barres ADN des proies (Eble & Pecore 2019). Par ailleurs, un projet de science citoyenne éducatif basé sur le code-barres ADN de spécimens de reptiles provenant de collections de tissus congelés de Musées a été mis en œuvre pour générer des données scientifiques solides (e.g., nouvelles séquences) qui pourraient être analysées en toute confiance (Chiovitti *et al.* 2019).

CONCLUSION

Un des avantages majeurs d'une approche fondée sur l'ADN est de pouvoir identifier et surveiller un large éventail d'organismes, des microbes aux plantes et animaux. Ces nouvelles approches offrent notamment des perspectives de suivis pour des groupes jusqu'alors délaissés par les méthodes plus traditionnelles, qui sont souvent très diversifiés et posent des problèmes lors de l'identification morphologique. On citera également toutes les perspectives en biologie des espèces, des communautés, etc. Par ailleurs, un avantage essentiel à toutes ces approches basées sur l'ADNe est le caractère non invasif et non destructif de l'échantillonnage pour de nombreux organismes. Cela s'avère particulièrement critique d'un point de vue éthique et pour la conservation si l'on veut étudier des espèces rares. Il est très intéressant de pouvoir ré-analyser des échantillons ultérieurement pour des vérifications ou répondre à de nouvelles questions grâce aux progrès des techniques. Les chercheurs peuvent implémenter diverses méthodologies complémentaires sur le terrain afin d'améliorer les suivis de la biodiversité. Récemment, des chercheurs ont proposé un réseau de stations majoritairement autonomes composées de plusieurs capteurs automatisés afin de faire un suivi de la biodiversité (i.e. « AMMOD » ou Automated Multisensor stations for Monitoring of species Diversity; Wägele *et al.* 2022). Ce réseau combine des technologies de pointe avec l'informatique de la biodiversité (i.e. reconnaissance des espèces à l'aide d'images, de l'acoustique et du *metabarcoding*) en utilisant des pièges à caméra pour les mammifères et les

TABLEAU 3. — Estimation des coûts liés à l'identification des espèces à partir de l'ADN (les chiffres fournis ici sont présentés comme des ordres de grandeur). Abréviations : **éch.**, échantillon.

	Invertébrés (matériel frais)		Tous groupes Barcoding (tous marqueurs)	Invertébrés (matériel ADN dégradé / collections)		Invertébrés (insectes piège Malaise)
	Barcoding (COI)			Barcoding (COI)		
	Coût si interne (MNHN)	Coût par échantillon si sous-traitance (GUELPH)	Coût si interne (MNHN)	Coût si interne (MNHN)	Coût par échantillon si sous-traitance (GUELPH)	Coût si interne (MNHN)
	Par éch. Individuel (SANGER)	Par lots de 95 éch. (SANGER)	Par éch. Individuel (NGS – Illumina, par lots de 4560 éch. ou plus)	Par éch. Individuel (SANGER)	Construction de librairies (SANGER)	Par éch. Individuel (NGS – Illumina, pour lot de 32 éch.) Profondeur de séquençage forte dans le cas d'un lot de 32 pièges Malaise; grouper davantage d'échantillons (64) semble encore raisonnable pour ce type d'échantillon, ce qui diminue par deux le coût du séquençage.
Action	4 €	2,5 €	2,50 €	5,70 €	12,54 €	12 € (trois réplicats)
Gestion des données brutes	Extraction ADN (kits, produits, etc.)	0,50 €	0,50 €	0,50 €	6 € (banque NGS, 3 réplicats PCR) - Double indexage	
	PCR (et matériel associé)	4,50 €	0,40 €	4,50 €	57 €	
	Séquençage (et matériel associé)	5 min / éch.	5 min / éch.	5 min / éch.	5 min / éch.	
	Données de référence (création)	2 min / éch.	2 min / éch.	2 min / éch.	15 min / éch. (estimé à 1j pour le démultiplexage d'un run de séquençage)	
	Coût de l'étape bio-informatique	0,5 € (stockage ADN, hors coût congélation)	0,5 € (stockage ADN, hors coût congélation)	0,5 € (stockage ADN, hors coût congélation)	0,5 € (stockage ADN, hors coût congélation)	
Archivage	Final des échantillons (ADN, voucher)	Expérimentations en laboratoire	Expérimentations en laboratoire	Expérimentations en laboratoire	Expérimentations en laboratoire	
Autre	Divers coûts liés à l'acquisition ou à la gestion des données	Contrôle et validation des résultats	Contrôle et validation des résultats	Contrôle et validation des résultats	Contrôle et validation des résultats	Contrôle et validation des résultats

TABLEAU 4. — Exemples d'études sur l'identification d'espèces à partir de l'ADN.

Groupe taxonomique	Exemple de cas d'étude
Vertébrés	Ushio <i>et al.</i> 2017 ; Abrams <i>et al.</i> 2019 ; Drinkwater <i>et al.</i> 2019, 2021 ; Fahmy <i>et al.</i> 2020 ; Furlan <i>et al.</i> 2020 ; Sales <i>et al.</i> 2020 ; Lyet <i>et al.</i> 2021 ; Broadhurst <i>et al.</i> 2021 ; Mena <i>et al.</i> 2021 ; Van der Heyde <i>et al.</i> 2021 ; Lynggaard <i>et al.</i> 2022 ; Newton <i>et al.</i> 2022 ; Seeber & Epp 2022 ; Srivathsan <i>et al.</i> 2022 ; McDonald <i>et al.</i> 2023 ; Métris & Métris 2023
Amphibiens et reptiles	Lopes <i>et al.</i> 2017, 2021 ; Sasso <i>et al.</i> 2017 ; Raemy & Ursenbacher 2018 ; Ficetola <i>et al.</i> 2019 ; Tingley <i>et al.</i> 2019 ; Kačergytė <i>et al.</i> 2021 ; Li <i>et al.</i> 2021, 2022 ; Nordstrom <i>et al.</i> 2022 ; Leonhardt <i>et al.</i> 2023
Arthropodes	Oliverio <i>et al.</i> 2018 ; Barsoum <i>et al.</i> 2019 ; Creedy <i>et al.</i> 2019 ; Fernandes <i>et al.</i> 2019 ; Thomsen & Sigsgaard 2019 ; Hardulak <i>et al.</i> 2020 ; Kirse <i>et al.</i> 2021 ; Van Lantman <i>et al.</i> 2021 ; Hermans <i>et al.</i> 2022 ; Keller <i>et al.</i> 2022 ; Krehenwinkel <i>et al.</i> 2022 ; Roger <i>et al.</i> 2022 ; Graham <i>et al.</i> 2023 ; Holmquist <i>et al.</i> 2023 ; Nell <i>et al.</i> 2023
Plantes	Suchan <i>et al.</i> 2019 ; Drummond <i>et al.</i> 2021 ; Ji <i>et al.</i> 2021 ; Johnson <i>et al.</i> 2021 ; Leontidou <i>et al.</i> 2021 ; Varotto <i>et al.</i> 2021 ; Banerjee <i>et al.</i> 2022 ; Ibrahim <i>et al.</i> 2022 ; Arstingstall <i>et al.</i> 2023 ; Câmara <i>et al.</i> 2023 ; Vasar <i>et al.</i> 2023.

petits invertébrés, mais aussi des échantillonneurs autonomes pour les insectes, le pollen et les spores, des enregistreurs audio pour les animaux qui vocalisent, des capteurs pour les composés organiques volatils émis par les plantes. En revanche, un des problèmes de ces approches est le fait que les bases de données de références sont encore très incomplètes et recèlent des erreurs d'assignation taxonomiques, notamment pour certains groupes taxonomiques hyperdiversifiés. Il est aussi important de noter la complexité des étapes bio-informatiques de ces approches, leur rapidité d'évolution et la complexité du traitement de l'ADNe *sensu stricto* (i.e. les infrastructures nécessaires pour éviter les problèmes de contamination ne sont pas accessibles à tous). Comme pour la reconnaissance des espèces basée sur les images ou l'acoustique, nous assistons actuellement à un développement méthodologique avec un rythme de publication soutenu ; et il n'est pas facile de trouver des standards à suivre.

Pour beaucoup, les méthodologies de reconnaissance basées sur l'ADN – en particulier l'ADNe – *via* le *metabarcoding* sont en train de transformer et révolutionner les suivis de la biodiversité et offrent de nouvelles opportunités de recherches appliquées et de conservation. Avec le développement des technologies génétiques, les champs d'application des études sur l'ADNe ne feront qu'augmenter pour les questions de recherche et de gestion associées à la surveillance des espèces et des écosystèmes (Pawlowski *et al.* 2020 ; Keller *et al.* 2022 ; Pascher *et al.* 2022). Les coûts liés à la mise en place de ce type de méthodologie évoluent rapidement et sont de plus en plus abordables (voir les estimations faites dans le Tableau 3). Au cours des dernières années, de nombreuses études ont utilisé le *metabarcoding* de l'ADN ou de l'ADNe de divers substrats pour l'évaluation de la biodiversité terrestre dans des contextes variés (Tableau 4). On peut s'attendre à ce que la quantité et la richesse des informations fournies avec ces approches augmentent avec les améliorations des technologies de séquençage et d'échantillonnage et les évolutions des traitements bio-informatiques. Pour vraiment valoriser les informations obtenues avec ces approches, il est recommandé de rendre les données d'ADN plus accessibles (i.e. plateformes ouvertes en accès libre, qui facilite la réutilisation), de les conserver dans des formats cohérents (Berry *et al.* 2021) et les chercheurs travaillent sur leur har-

monisation dès l'acquisition des données de *metabarcoding* (e.g., Arribas *et al.* 2022).

Par ailleurs, sans une compréhension approfondie des forces et des faiblesses des méthodologies de l'ADNe (e.g., comportement de l'ADNe), il reste difficile de faire des déductions correctes à partir des données générées (Van der Heyde *et al.* 2022). Malgré le potentiel largement reconnu de l'ADNe, les limitations techniques et des problèmes conceptuels font encore obstacle à sa large application dans les programmes de surveillance (Cordier *et al.* 2021). Les gestionnaires pourraient ainsi être hésitants à mettre en place un protocole ou à utiliser les données d'ADNe en raison de l'incertitude entourant l'interprétation des données. C'est pourquoi, l'utilisation de l'ADNe pour détecter divers taxa peut être envisagée comme complément à des méthodes de suivis plus traditionnelles (e.g., capture de spécimens vivants, pose de pièges et piégeage par caméra). La possibilité de combiner la collecte d'ADNe avec des méthodes de surveillance traditionnelles permettrait ainsi de faire des déductions plus fiables sur les sites pour lesquels les données sont limitées (Keller *et al.* 2022).

Remerciements

Ce travail s'inscrit dans le cadre du projet Passifor2, financé par le ministère de l'Écologie. Les auteurs remercient chaleureusement Antoine Levêque pour son aide tout au long de la rédaction de ce document, Rodolphe Rougerie pour son aide dans l'estimation des coûts liés à l'identification des espèces à partir d'images, ainsi que Lucas Sire pour sa relecture attentive et ses commentaires très constructifs. Ils remercient également Francis Olivereau, Vincent Dubut et le troisième rapporteur pour leurs commentaires et leur révision attentive en vue d'améliorer leur manuscrit.

RÉFÉRENCES

- ABRAMS J. F., HÖRIG L. A., BROZOVIC R., AXTNER J., CRAMPTON-PLATT A., MOHAMED A., WONG S. T., SOLLMANN R., YU D. W. & WILTING A. 2019. — Shifting up a gear with iDNA: from mammal detection events to standardised surveys. *Journal of Applied Ecology* 56 (7): 1637-1648. <https://doi.org/10.1111/1365-2664.13411>

- ALBERDI A., AIZPURUA O., BOHMANN K., GOPALAKRISHNAN S., LYNGGAARD C., NIELSEN M. & GILBERT M. T. P. 2019. — Promises and pitfalls of using high-throughput sequencing for diet analysis. *Molecular Ecology Resources* 19 (2): 327-348. <https://doi.org/10.1111/1755-0998.12960>
- ANDERSEN K., BIRD K. L., RASMUSSEN M., HAILE J., BREUNING-MADSEN H., KJÆR K. H., ORLANDO L., GILBERT M. T. P. & WILLERSLEV E. 2012. — Meta-barcoding of 'dirty' DNA from soil reflects vertebrate biodiversity. *Molecular Ecology* 21 (8): 1966-1979. <https://doi.org/10.1111/j.1365-294X.2011.05261.x>
- ANDÚJAR C., ARRIBAS P., RUZICKA F., CRAMPTON-PLATT A., TIMMERMANS M. J. T. N. & VÖGLER A. P. 2015. — Phylogenetic community ecology of soil biodiversity using mitochondrial metagenomics. *Molecular Ecology* 24 (14): 3603-3617. <https://doi.org/10.1111/mec.13195>
- ARRIBAS P., ANDÚJAR C., HOPKINS K., SHEPHERD M. & VÖGLER A. P. 2016. — Metabarcoding and mitochondrial metagenomics of endogean arthropods to unveil the mesofauna of the soil. *Methods in Ecology and Evolution* 7 (9): 1071-1081. <https://doi.org/10.1111/2041-210X.12557>
- ARRIBAS P., ANDÚJAR C., BOHMANN K., DEWAARD J. R., ECONOMO E. P., ELBRECHT V., GEISEN S., GOBERNA M., KREHENWINKEL H., NOVOTNY V., ZINGER L., CREEDY T. J., MERAMVELIOTAKIS E., NOGUERALES V., OVERCAST I., MORLON H., PAPADOPOULOU A., VÖGLER A. P. & EMERSON B. C. 2022. — Toward global integration of biodiversity big data: a harmonized metabarcode data generation module for terrestrial arthropods. *GigaScience* 11: giac065. <https://doi.org/10.1093/gigascience/giac065>
- ARSTINGSTALL K. A., DEBANO S. J., LI X., WOOSTER D. E., ROWLAND M. M., BURROWS S. & FROST K. 2023. — Investigating the use of pollen DNA metabarcoding to quantify bee foraging and effects of threshold selection. *PLOS ONE* 18 (4): e0282715. <https://doi.org/10.1371/journal.pone.0282715>
- BALOGLU B., CHEN Z., ELBRECHT V., BRAUKMANN T., MACDONALD S. & STEINKE D. 2021. — A workflow for accurate metabarcoding using nanopore MinION sequencing. *Methods in Ecology and Evolution* 12 (5): 794-804. <https://doi.org/10.1111/2041-210X.13561>
- BANCHI E., AMETRANO C. G., STANKOVIĆ D., VERARDO P., MORETTI O., GABRIELLI F., LAZZARIN S., BORNEY M. F., TASSAN F., TRETACH M., PALLAVICINI A. & MUGGIA L. 2018. — DNA metabarcoding uncovers fungal diversity of mixed airborne samples in Italy. *PLOS ONE* 13 (3): e0194489. <https://doi.org/10.1371/journal.pone.0194489>
- BANCHI E., PALLAVICINI A. & MUGGIA L. 2020. — Relevance of plant and fungal DNA metabarcoding in aerobiology. *Aerobiologia* 36 (1): 9-23. <https://doi.org/10.1007/s10453-019-09574-2>
- BANERJEE P., STEWART K. A., DEY G., ANTOGNAZZA C. M., SHARMA R. K., MAITY J. P., SAHA S., DOI H., DE VERE N., CHAN M. W. Y., LIN P.-Y., CHAO H.-C. & CHEN C.-Y. 2022. — Environmental DNA analysis as an emerging non-destructive method for plant biodiversity monitoring: a review. *AoB PLANTS: plac031*. <https://doi.org/10.1093/aobpla/plac031>
- BARBA M. DE, MIQUEL C., BOYER F., MERCIER C., RIOUX D., COISSAC E. & TABERLET P. 2014. — DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Molecular Ecology Resources* 14 (2): 306-323. <https://doi.org/10.1111/1755-0998.12188>
- BARSOUM N., BRUCE C., FORSTER J., JI Y.-Q. & YU D. W. 2019. — The devil is in the detail: metabarcoding of arthropods provides a sensitive measure of biodiversity response to forest stand composition compared with surrogate measures of biodiversity. *Ecological Indicators* 101: 313-323. <https://doi.org/10.1016/j.ecolind.2019.01.023>
- BECKER B. & PUSHKAREVA E. 2023. — Metagenomics provides a deeper assessment of the diversity of bacterial communities in polar soils than metabarcoding. *Genes* 14 (4): 812. <https://doi.org/10.3390/genes14040812>
- BEENTJES K. K., SPEKSNIJDER A. G. C. L., SCHILTHUIZEN M., HOOGEVEEN M. & VAN DER HOORN B. B. 2019. — The effects of spatial and temporal replicate sampling on eDNA metabarcoding. *PeerJ* 7: e7335. <https://doi.org/10.7717/peerj.7335>
- BERRY O., JARMAN S., BISSETT A., HOPE M., PAEPER C., BESSEY C., SCHWARTZ M. K., HALE J. & BUNCE M. 2021. — Making environmental DNA (eDNA) biodiversity records globally accessible. *Environmental DNA* 3 (4): 699-705. <https://doi.org/10.1002/edn3.173>
- BIGGS J., EWALD N., VALENTINI A., GABORIAUD C., DEJEAN T., GRIFFITHS R. A., FOSTER J., WILKINSON J. W., ARNELL A., BROTHERTON P., WILLIAMS P. & DUNN F. 2015. — Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (*Triturus cristatus*). Special issue: Environmental dna: a powerful new tool for biological conservation. *Biological Conservation* 183: 19-28. <https://doi.org/10.1016/j.biocon.2014.11.029>
- BOHMANN K., ELBRECHT V., CARØE C., BISTA I., LEESE F., BUNCE M., YU D. W., SEYMOUR M., DUMBRELL A. J. & CREER S. 2022. — Strategies for sample labelling and library preparation in DNA metabarcoding studies. *Molecular Ecology Resources* 22 (4): 1231-1246. <https://doi.org/10.1111/1755-0998.13512>
- BOKULICH N. A., SUBRAMANIAN S., FAITH J. J., GEVERS D., GORDON J. I., KNIGHT R., MILLS D. A. & CAPORASO J. G. 2013. — Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods* 10 (1): 57-59. <https://doi.org/10.1038/nmeth.2276>
- 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. <https://doi.org/10.1111/1755-0998.12428>
- BRAUKMANN T. W. A., IVANOVA N. V., PROSSER S. W. J., ELBRECHT V., STEINKE D., RATNASINGHAM S., WAARD J. R. DE, SONES J. E., ZAKHAROV E. V. & HEBERT P. D. N. 2019. — Metabarcoding a diverse arthropod mock community. *Molecular Ecology Resources* 19 (3): 711-727. <https://doi.org/10.1111/1755-0998.13008>
- BROADHURST H. A., GREGORY L. M., BLEAKLEY E. K., PERKINS J. C., LAVIN J. V., BOLTON P., BROWETT S. S., HOWE C. V., SINGLETON N., TANSLEY D., SALES N. G. & MCDEVITT A. D. 2021. — Mapping differences in mammalian distributions and diversity using environmental DNA from rivers. *Science of the Total Environment* 801: 149724. <https://doi.org/10.1016/j.scitotenv.2021.149724>
- BRUCE K., BLACKMAN R., BOURLAT S. J., HELLSTRÖM A. M., BAKKER J., BISTA I., BOHMANN K., BOUCHEZ A., BRYS R., CLARK K., ELBRECHT V., FAZI S., FONSECA V., HÄNFLING B., LEESE F., MÄCHLER E., MAHON A. R., MEISSNER K., PANKSEP K., PAWLOWSKI J., SCHMIDT YAÑEZ P., SEYMOUR M., THALINGER B., VALENTINI A., WOODCOCK P., TRAUGOTT M., VASSELON V. & DEINER K. 2021. — *A Practical Guide to DNA-Based Methods for Biodiversity Assessment*. Pensoft Publishers, Sofia, 90 p. <https://doi.org/10.3897/ab.e68634>
- CALVIGNAC-SPENCER S., MERKEL K., KUTZNER N., KÜHL H., BOESCH C., KAPPELER P. M., METZGER S., SCHUBERT G. & LEENDERTZ F. H. 2013. — Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. *Molecular Ecology* 22 (4): 915-924. <https://doi.org/10.1111/mec.12183>
- CÂMARA P. E. A. S., LOPES F. A. C., BONES F. L. V., RODRIGUES L. A. C., CARVALHO-SILVA M., STECH M., CONVEY P. & ROSA L. H. 2023. — Investigating aerial diversity of non-fungal eukaryotes across a 40 degrees latitudinal transect using DNA metabarcoding. *Austral Ecology* 48 (6): 1178-1194. <https://doi.org/10.1111/aec.13332>
- CARINI P., MARSDEN P. J., LEFF J. W., MORGAN E. E., STRICKLAND M. S. & FIERER N. 2016. — Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nature Microbiology* 2 (3): 1-6. <https://doi.org/10.1038/nmicrobiol.2016.242>

- CHEN S., YAO H., HAN J., LIU C., SONG J., SHI L., ZHU Y., MA X., GAO T., PANG X., LUO K., LI Y., LI X., JIA X., LIN Y. & LEON C. 2010. — Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLOS ONE* 5 (1): e8613. <https://doi.org/10.1371/journal.pone.0008613>
- CHEN W. & FICETOLA G. F. 2020. — Numerical methods for sedimentary-ancient-DNA-based study on past biodiversity and ecosystem functioning. *Environmental DNA* 2 (2): 115-129. <https://doi.org/10.1002/edn3.79>
- CHESHIRE P. R., FISCHER E. E., DOWDY N. J., GRISWOLD T. L., HUGHES A. C., ORR M. C., ASCHER J. S., GUZMAN L. M., HUNG K.-L. J., COBB N. S. & MCCABE L. M. 2023. — Completeness analysis for over 3000 United States bee species identifies persistent data gap. *Ecography* 2023 (5): e06584. <https://doi.org/10.1111/ecog.06584>
- CHIOVITTI A., THORPE F., GORMAN C., CUXSON J. L., ROBEVSKA G., SZWED C., DUNCAN J. C., VANYAI H.K., CROSS J., SIEMERING K. R. & SUMNER J. 2019. — A citizen science model for implementing statewide educational DNA barcoding. *PLOS ONE* 14 (1): e0208604. <https://doi.org/10.1371/journal.pone.0208604>
- CHUA P. Y. S., CRAMPTON-PLATT A., LAMMERS Y., ALSOS I. G., BOESSENKOOL S. & BOHMANN K. 2021. — Metagenomics: a viable tool for reconstructing herbivore diet. *Molecular Ecology Resources* 21 (7): 2249-2263. <https://doi.org/10.1111/1755-0998.13425>
- CICCONARDI F., BORGES P. A. V., STRASBERG D., OROMÍ P., LÓPEZ H., PÉREZ-DELGADO A.J., CASQUET J., CAUJAPÉ-CASTELLS J., FERNÁNDEZ-PALACIOS J. M., THÉBAUD C. & EMERSON B. C. 2017. — MtDNA metagenomics reveals large-scale invasion of belowground arthropod communities by introduced species. *Molecular Ecology* 26 (12): 3104-3115. <https://doi.org/10.1111/mec.14037>
- CIVADE R., DEJEAN T., VALENTINI A., ROSET N., RAYMOND J.-C., BONIN A., TABERLET P. & PONT D. 2016. — Spatial representativeness of environmental DNA metabarcoding signal for fish biodiversity assessment in a natural freshwater system. *PLOS ONE* 11 (6): e0157366. <https://doi.org/10.1371/journal.pone.0157366>
- CLARE E. L. 2014. — Molecular detection of trophic interactions: emerging trends, distinct advantages, significant considerations and conservation applications. *Evolutionary Applications* 7 (9): 1144-1157. <https://doi.org/10.1111/eva.12225>
- CLARE E. L., ECONOMOU C.K., BENNETT F. J., DYER C. E., ADAMS K., MCROBIE B., DRINKWATER R. & LITTLEFAIR J. E. 2022. — Measuring biodiversity from DNA in the air. *Current Biology* 32 (3): 693-700.e5. <https://doi.org/10.1016/j.cub.2021.11.064>
- COMPSON Z. G., MCCLENAGHAN B., SINGER G. A. C., FAHNER N. A. & HAJIBABAEI M. 2020. — Metabarcoding from microbes to mammals: comprehensive bioassessment on a global scale. *Frontiers in Ecology and Evolution* 8: 581838. <https://doi.org/10.3389/fevo.2020.581835>
- CORDIER T., ALONSO-SÁEZ L., APOTHÉLOZ-PERRET-GENTIL L., AYLAGE S., BOHAN D. A., BOUCHEZ A., CHARITON A., CREER S., FRÜHE L., KECK F., KEELEY N., LAROCHE O., LEESE F., POCHON X., STOECK T., PAWLOWSKI J. & LANZÉN A. 2021. — Ecosystems monitoring powered by environmental genomics: a review of current strategies with an implementation roadmap. *Molecular Ecology* 30 (13): 2937-2958. <https://doi.org/10.1111/mec.15472>
- CORSE E., MEGLÉCZ E., ARCHAMBAUD G., ARDISSON M., MARTIN J.-F., TOUGARD C., CHAPPAZ R. & DUBUT V. 2017. — A from-benchtop-to-desktop workflow for validating HTS data and for taxonomic identification in diet metabarcoding studies. *Molecular Ecology Resources* 17 (6): e146-e159. <https://doi.org/10.1111/1755-0998.12703>
- CORSE E., TOUGARD C., ARCHAMBAUD-SUARD G., AGNÈSE J.-F., MESSU MANDENG F. D., BILONG BILONG C. F., DUNEAU D., ZINGER L., CHAPPAZ R., XU C. C. Y., MEGLÉCZ E. & DUBUT V. 2019. — One-locus-several-primers: a strategy to improve the taxonomic and haplotypic coverage in diet metabarcoding studies. *Ecology and Evolution* 9 (8): 4603-4620. <https://doi.org/10.1002/ece3.5063>
- COUTANT O., RICHARD-HANSEN C., DE THOISY B., DECOTTE J.-B., VALENTINI A., DEJEAN T., VIGOUROUX R., MURIENNE J. & BROUSSE S. 2021. — Amazonian mammal monitoring using aquatic environmental DNA. *Molecular Ecology Resources* 21 (6): 1875-1888. <https://doi.org/10.1111/1755-0998.13393>
- COUTON M., BAUD A., DAGUIN-THIÉBAUT C., CORRE E., COMTET T. & VIARD F. 2021. — High-throughput sequencing on preservative ethanol is effective at jointly examining intraspecific and taxonomic diversity, although bioinformatics pipelines do not perform equally. *Ecology and Evolution* 11 (10): 5533-5546. <https://doi.org/10.1002/ece3.7453>
- CRAMPTON-PLATT A., YU D. W., ZHOU X. & VOGLER A. P. 2016. — Mitochondrial metagenomics: letting the genes out of the bottle. *GigaScience* 5 (1): 15. <https://doi.org/10.1186/s13742-016-0120-y>
- CREEDY T. J., NG W. S. & VOGLER A. P. 2019. — Toward accurate species-level metabarcoding of arthropod communities from the tropical forest canopy. *Ecology and Evolution* 9 (6): 3105-3116. <https://doi.org/10.1002/ece3.4839>
- CREER S., DEINER K., FREY S., PORAZINSKA D., TABERLET P., THOMAS W.K., POTTER C. & BIK H.M. 2016. — The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution* 7 (9): 1008-1018. <https://doi.org/10.1111/2041-210X.12574>
- CRISTESCU M. E. 2019. — Can environmental RNA revolutionize biodiversity science? *Trends in Ecology & Evolution* 34 (8): 694-697. <https://doi.org/10.1016/j.tree.2019.05.003>
- CRISTESCU M. E. 2014. — From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends in Ecology & Evolution* 29 (10): 566-571. <https://doi.org/10.1016/j.tree.2014.08.001>
- CUTAJAR T. P. & ROWLEY J. J. L. 2020. — Surveying frogs from the bellies of their parasites: invertebrate-derived DNA as a novel survey method for frogs. *Global Ecology and Conservation* 22: e00978. <https://doi.org/10.1016/j.gecco.2020.e00978>
- DECAËNS T., JIMÉNEZ J. J., GIOIA C., MEASEY G. J. & LAVELLE P. 2006. — The values of soil animals for conservation biology. *ICSZ. European Journal of Soil Biology* 42 (suppl. 1): S23-S38. <https://doi.org/10.1016/j.ejsobi.2006.07.001>
- DECAËNS T., PORCO D., ROUGERIE R., BROWN G. G. & JAMES S. W. 2013. — Potential of DNA barcoding for earthworm research in taxonomy and ecology. *Applied Soil Ecology* 65: 35-42. <https://doi.org/10.1016/j.apsoil.2013.01.001>
- DEINER K., BIK H. M., MÄCHLER E., SEYMOUR M., LACOURSIÈRE-ROUSSEL A., ALTERMATT F., CREER S., BISTA I., LODGE D. M., VERE N. DE, PFRENDER M. E. & BERNATCHEZ L. 2017. — Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *Molecular Ecology* 26 (21): 5872-5895. <https://doi.org/10.1111/mec.14350>
- DEJEAN T., VALENTINI A., DUPARC A., PELLIER-CUIT S., POMPANON F., TABERLET P. & MIAUD C. 2011. — Persistence of Environmental DNA in Freshwater Ecosystems. *PLOS ONE* 6 (8): e23398. <https://doi.org/10.1371/journal.pone.0023398>
- DINCĂ V., MONTAGUD S., TALAVERA G., HERNÁNDEZ-ROLDÁN J., MUNGUIRA M. L., GARCÍA-BARROS E., HEBERT P. D. N. & VILA R. 2015. — DNA barcode reference library for Iberian butterflies enables a continental-scale preview of potential cryptic diversity. *Scientific Reports* 5 (1): 12395. <https://doi.org/10.1038/srep12395>
- DOI H., UCHII K., TAKAHARA T., MATSUHASHI S., YAMANAKA H. & MINAMOTO T. 2015. — Use of droplet digital PCR for estimation of fish abundance and biomass in environmental DNA surveys. *PLOS ONE* 10 (3): e0122763. <https://doi.org/10.1371/journal.pone.0122763>
- DOMMAIN R., ANDAMA M., MCDONOUGH M. M., PRADO N. A., GOLDHAMMER T., POTTS R., MALDONADO J. E., NKU-

- RUNUNGI J. B. & CAMPANA M. G. 2020. — The challenges of reconstructing tropical biodiversity with sedimentary ancient DNA: a 2200-year-long metagenomic record from Bwindi Impenetrable Forest, Uganda. *Frontiers in Ecology and Evolution* 8 (218): 1-26. <https://doi.org/10.3389/fevo.2020.00218>
- DOPHEIDE A., TOOMAN L. K., GROSSER S., AGABITI B., RHODE B., XIE D., STEVENS M. I., NELSON N., BUCKLEY T. R., DRUMMOND A. J. & NEWCOMB R. D. 2019. — Estimating the biodiversity of terrestrial invertebrates on a forested island using DNA barcodes and metabarcoding data. *Ecological Applications* 29 (4): e01877. <https://doi.org/10.1002/eap.1877>
- DÖRLER D., KROPF M., LAHA G. & ZALLER J. G. 2018. — Occurrence of the invasive Spanish slug in gardens: can a citizen science approach help deciphering underlying factors? *BMC Ecology* 18 (1): 23. <https://doi.org/10.1186/s12898-018-0179-7>
- DRAKE L. E., CUFF J. P., YOUNG R. E., MARCHBANK A., CHADWICK E. A. & SYMONDSON W. O. C. 2022. — An assessment of minimum sequence copy thresholds for identifying and reducing the prevalence of artefacts in dietary metabarcoding data. *Methods in Ecology and Evolution* 13 (3): 694-710. <https://doi.org/10.1111/2041-210X.13780>
- DRINKWATER R., SCHNELL I. B., BOHMAN K., BERNARD H., VERON G., CLARE E., GILBERT M. T. P. & ROSSITER S. J. 2019. — Using metabarcoding to compare the suitability of two blood-feeding leech species for sampling mammalian diversity in North Borneo. *Molecular Ecology Resources* 19 (1): 105-117. <https://doi.org/10.1111/1755-0998.12943>
- DRINKWATER R., JUCKER T., POTTER J. H. T., SWINFIELD T., COOMES D. A., SLADE E. M., GILBERT M. T. P., LEWIS O. T., BERNARD H., STRUEBIG M. J., CLARE E. L. & ROSSITER S. J. 2021. — Leech blood-meal invertebrate-derived DNA reveals differences in Bornean mammal diversity across habitats. *Molecular Ecology* 30 (13): 3299-3312. <https://doi.org/10.1111/mec.15724>
- DRUMMOND A. J., NEWCOMB R. D., BUCKLEY T. R., XIE D., DOPHEIDE A., POTTER B. C., HELED J., ROSS H. A., TOOMAN L., GROSSER S., PARK D., DEMETRAS N. J., STEVENS M. I., RUSSELL J. C., ANDERSON S. H., CARTER A. & NELSON N. 2015. — Evaluating a multigene environmental DNA approach for biodiversity assessment. *GigaScience* 4 (1): s13742-015-0086-1. <https://doi.org/10.1186/s13742-015-0086-1>
- DRUMMOND J. A., LARSON E. R., LI Y., LODGE D. M., GANTZ C. A., PPRENDER M. E., RENSHAW M. A., CORREA A. M. S. & EGAN S. P. 2021. — Diversity metrics are robust to differences in sampling location and depth for environmental DNA of plants in small temperate lakes. *Frontiers in Environmental Science* 9: 88. <https://doi.org/10.3389/fenvs.2021.617924>
- EBLE J. & PECORE J. 2019. — “Invasive Aliens”: a student citizen-science activity using DNA barcoding to investigate concepts in ecology & molecular biology. *The American Biology Teacher* 81 (3): 169-174. <https://doi.org/10.1525/abt.2019.81.3.169>
- EDGAR R. C. 2013. — UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* 10 (10): 996-998. <https://doi.org/10.1038/nmeth.2604>
- ELBRECHT V., PEINERT B. & LEESE F. 2017. — Sorting things out: assessing effects of unequal specimen biomass on DNA metabarcoding. *Ecology and Evolution* 7 (17): 6918-6926. <https://doi.org/10.1002/ece3.3192>
- ELBRECHT V., BRAUKMANN T. W. A., IVANOVA N. V., PROSSER S. W. J., HAJIBABAEI M., WRIGHT M., ZAKHAROV E. V., HEBERT P. D. N. & STEINKE D. 2019. — Validation of COI metabarcoding primers for terrestrial arthropods. *PeerJ* 7: e7745. <https://doi.org/10.7717/peerj.7745>
- EMILSON C. E., THOMPSON D. G., VENIER L. A., PORTER T. M., SWYSTUN T., CHARTRAND D., CAPELL S. & HAJIBABAEI M. 2017. — DNA metabarcoding and morphological macroinvertebrate metrics reveal the same changes in boreal watersheds across an environmental gradient. *Scientific Reports* 7 (1): 12777. <https://doi.org/10.1038/s41598-017-13157-x>
- ERDOZAIN M., THOMPSON D. G., PORTER T. M., KIDD K. A., KREUTZWEISER D. P., SIBLEY P. K., SWYSTUN T., CHARTRAND D. & HAJIBABAEI M. 2019. — Metabarcoding of storage ethanol vs conventional morphometric identification in relation to the use of stream macroinvertebrates as ecological indicators in forest management. *Ecological Indicators* 101: 173-184. <https://doi.org/10.1016/j.ecolind.2019.01.014>
- ESSER K.-H., MARX W. & LISOWSKY T. 2006. — DNA decontamination: DNA-ExitusPlus in comparison with conventional reagents. *Nature Methods – NAT METHODS* 3: 1-2. <https://doi.org/10.1038/nmeth853>
- FAHMY M., WILLIAMS K. M., TESSLER M., WEISKOPF S. R., HEKALA E. & SIDDALL M. E. 2020. — Multilocus metabarcoding of terrestrial leech bloodmeal iDNA increases species richness uncovered in surveys of vertebrate host biodiversity. *Journal of Parasitology* 106 (6): 843-853. <https://doi.org/10.1645/19-189>
- FALENTIN H., AUER L., MARIADASSOU M., PASCAL G., RUÉ O., DUGAT-BONY E., DELBES C., NICOLAS A., RIFA E., MONDY S., LE BOULCH M., CAUQUIL L., HERNANDEZ RAQUET G., TER-RAT S. & ABRAHAM A.-L. 2019. — Guide pratique à destination des biologistes, bioinformaticiens et statisticiens qui souhaitent s'initier aux analyses métabarcoding. *Cahiers des Techniques de l'INRA* 2019 (97): 1-23
- FERNANDES K., VAN DER HEYDE M., COGHLAN M., WARDELL-JOHNSON G., BUNCE M., HARRIS R. & NEVILL P. 2019. — Invertebrate DNA metabarcoding reveals changes in communities across mine site restoration chronosequences. *Restoration Ecology* 27 (5): 1177-1186. <https://doi.org/10.1111/rec.12976>
- FICETOLA G. F., PANSU J., BONIN A., COISSAC E., GIGUET-COVEX C., BARBA M. D., GIELLY L., LOPES C. M., BOYER F., POMPANON F., RAYÉ G. & TABERLET P. 2014. — Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources* 15 (3): 543-556. <https://doi.org/10.1111/1755-0998.12338>
- FICETOLA G. F., MANENTI R. & TABERLET P. 2019. — Environmental DNA and metabarcoding for the study of amphibians and reptiles: species distribution, the microbiome, and much more. *Amphibia-Reptilia* 40 (2): 129-148. <https://doi.org/10.1163/15685381-20191194>
- FRANCIOLI D., LENTENDU G., LEWIN S. & KOLB S. 2021. — DNA metabarcoding for the characterization of terrestrial microbiota – pitfalls and solutions. *Microorganisms* 9 (2): 361. <https://doi.org/10.3390/microorganisms9020361>
- FREITAG H., DE VRIES R., PATERNO M., MAESTRI S., DELLEDONNE M., THOMPSON C. G., LAMED H., LAMBERT R., FOX M. F., GONZALEZ M. C., DELOCADO E. D., SABBORDO M.R., PANGANTIHON C. V. & NJUNJIĆ I. 2021. — *Hydraena* (s.str.) *dinarica*, new species (Coleoptera: Hydraenidae) along with further records of *Hydraena* spp. from Durmitor National Park, Montenegro and comments on the DNA barcoding problem with the genus. *Biodiversity Data Journal* 9: e59892. <https://doi.org/10.3389/BDJ.9.e59892>
- FURLAN E. M., DAVIS J. & DUNCAN R. P. 2020. — Identifying error and accurately interpreting environmental DNA metabarcoding results: a case study to detect vertebrates at arid zone waterholes. *Molecular Ecology Resources* 20 (5): 1259-1276. <https://doi.org/10.1111/1755-0998.13170>
- GALAN M., PONS J.-B., TOURNAYRE O., PIERRE É., LEUCHTMANN M., PONTIER D. & CHARBONNEL N. 2018. — Metabarcoding for the parallel identification of several hundred predators and their prey: application to bat species diet analysis. *Molecular Ecology Resources* 18 (3): 474-489. <https://doi.org/10.1111/1755-0998.12749>
- GIGUET-COVEX C., POULENARD J., GIELLY L., BAJARD M., FOUINAT L., FICETOLA G.-F., DAVID F., PANSU J., WALSH K., MOCCI F., TABERLET P., CHOLER P. & ARNAUD F. 2015. — L'histoire des activités agricoles et des paysages révélée par un outil en plein essor: l'ADN sédimentaire lacustre. Collection EDYTEM. *Cahiers de géographie* 18 (1): 81-102. <https://doi.org/10.3406/edyte.2015.1352>

- GOLDBERG C. S., TURNER C. R., DEINER K., KLYMUS K. E., THOMSEN P. F., MURPHY M. A., SPEAR S. F., MCKEE A., OYLER-MCCANCE S. J., CORNMANN R. S., LARAMIE M. B., MAHON A. R., LANCE R. F., PILLIOD D. S., STRICKLER K. M., WAITS L. P., FRIEMER A. K., TAKAHARA T., HERDER J. E. & TABERLET P. 2016. — Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution* 7 (11): 1299-1307. <https://doi.org/10.1111/2041-210X.12595>
- GRAHAM N. R., KREHENWINKEL H., LIM J. Y., STANICZENKO P., CALLAGHAN J., ANDERSEN J. C., GRUNER D. S. & GILLESPIE R. G. 2023. — Ecological network structure in response to community assembly processes over evolutionary time. *Molecular Ecology* 32 (23): 6489-6506. <https://doi.org/10.1111/mec.16873>
- GREALLY A., RAWLENCE N. J. & BUNCE M. 2017. — Time to spread your wings: a review of the avian ancient DNA field. *Genes* 8 (7): 184. <https://doi.org/10.3390/genes8070184>
- GREY E. K., BERNATCHEZ L., CASSEY P., DEINER K., DEVENNEY M., HOWLAND K. L., LACOURSIÈRE-ROUSSEL A., LEONG S. C. Y., LI Y., OLDS B., PFRENDER M. E., PROWSE T. A. A., RENSHAW M. A. & LODGE D. M. 2018. — Effects of sampling effort on biodiversity patterns estimated from environmental DNA metabarcoding surveys. *Scientific Reports* 8 (1): 8843. <https://doi.org/10.1038/s41598-018-27048-2>
- GUGERLI F., PARDUCCI L. & PETTIT R. J. 2005. — Ancient plant DNA: review and prospects. *New Phytologist* 166 (2): 409-418. <https://doi.org/10.1111/j.1469-8137.2005.01360.x>
- HARDULAK L. A., MORINIÈRE J., HAUSMANN A., HENDRICH L., SCHMIDT S., DOCZKAL D., MÜLLER J., HEBERT P. D. N. & HASZPRUNAR G. 2020. — DNA metabarcoding for biodiversity monitoring in a national park: screening for invasive and pest species. *Molecular Ecology Resources* 20 (6): 1542-1557. <https://doi.org/10.1111/1755-0998.13212>
- HAWKINS J., VERE N. DE, GRIFFITH A., FORD C. R., ALLAINGUILLAUME J., HEGARTY M. J., BAILLIE L. & ADAMS-GROOM B. 2015. — Using DNA metabarcoding to identify the floral composition of honey: a new tool for investigating honey bee foraging preferences. *PLOS ONE* 10 (8): e0134735. <https://doi.org/10.1371/journal.pone.0134735>
- HEBERT P. D. N., RATNASINGHAM S. & DE WAARD J. R. 2003. — Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270 (suppl. 1): S96-S99. <https://doi.org/10.1098/rsbl.2003.0025>
- HEBERT P. D. N., STOECKLE M. Y., ZEMLAK T. S. & FRANCIS C. M. 2004. — Identification of birds through DNA barcodes. *PLOS Biology* 2 (10): e312. <https://doi.org/10.1371/journal.pbio.0020312>
- HERMANS S. M., LEAR G., BUCKLEY T. R. & BUCKLEY H. L. 2022. — Environmental DNA sampling detects between-habitat variation in soil arthropod communities, but is a poor indicator of fine-scale spatial and seasonal variation. *Ecological Indicators* 140: 109040. <https://doi.org/10.1016/j.ecolind.2022.109040>
- HOLMQUIST A. J., ADAMS S. A. & GILLESPIE R. G. 2023. — Invasion by an ecosystem engineer changes biotic interactions between native and non-native taxa. *Ecology and Evolution* 13 (2): e9820. <https://doi.org/10.1002/ece3.9820>
- HOPKEN M. W., ORNING E. K., YOUNG J. K. & PIAGGIO A. J. 2016. — Molecular forensics in avian conservation: a DNA-based approach for identifying mammalian predators of ground-nesting birds and eggs. *BMC Research Notes* 9 (14): 1-9. <https://doi.org/10.1186/s13104-015-1797-1>
- HUNTER M. E., DORAZIO R. M., BUTTERFIELD J. S. S., MEIGSFRIEND G., NICO L. G. & FERRANTE J. A. 2017. — Detection limits of quantitative and digital PCR assays and their influence in presence-absence surveys of environmental DNA. *Molecular Ecology Resources* 17 (2): 221-229. <https://doi.org/10.1111/1755-0998.12619>
- IBRAHIM A., HÖCKENDORFF S., SCHLEHECK D., EPP L., VAN KLEUNEN M. & MEYER A. 2022. — Vegetation changes over the last centuries in the Lower Lake Constance region reconstructed from sediment-core environmental DNA. *Environmental DNA* 4 (4): 830-845. <https://doi.org/10.1002/edn3.292>
- INGALA M. R., SIMMONS N. B., WULTSCH C., KRAMPIS K., PROVOST K. L. & PERKINS S. L. 2021. — Molecular diet analysis of neotropical bats based on fecal DNA metabarcoding. *Ecology and Evolution* 11 (12): 7474-7491. <https://doi.org/10.1002/ece3.7579>
- JIFU, YAN L., YAN S., QIN T., SHEN J. & ZHA J. 2021. — Estimating aquatic plant diversity and distribution in rivers from Jingjinji region, China, using environmental DNA metabarcoding and a traditional survey method. *Environmental Research* 199: 111348. <https://doi.org/10.1016/j.envres.2021.111348>
- JOHNSON M. D., FOKAR M., COX R. D. & BARNES M. A. 2021. — Airborne environmental DNA metabarcoding detects more diversity, with less sampling effort, than a traditional plant community survey. *BMC Ecology and Evolution* 21 (1): 218. <https://doi.org/10.1186/s12862-021-01947-x>
- JOHNSON M. D., FREELAND J. R., PARDUCCI L., EVANS D. M., MEYER R. S., MOLANO-FLORES B. & DAVIS M. A. 2023. — Environmental DNA as an emerging tool in botanical research. *American Journal of Botany* 110 (2): e16120. <https://doi.org/10.1002/ajb2.16120>
- KÁČERGÝTĚ I., PETERSSON E., ARLT D., HELLSTRÖM M., KNAPE J., SPENS J., ŽMIHORSKI M. & PÄRT T. 2021. — Environmental DNA metabarcoding elucidates patterns of fish colonisation and co-occurrences with amphibians in temperate wetlands created for biodiversity. *Freshwater Biology* 66 (10): 1915-1929. <https://doi.org/10.1111/fwb.13800>
- KELLER A. G., GRASON E. W., McDONALD P. S., RAMÓN-LACA A. & KELLY R. P. 2022. — Tracking an invasion front with environmental DNA. *Ecological Applications* 32 (4): e2561. <https://doi.org/10.1002/eap.2561>
- KIRSE A., BOURLAT S. J., LANGEN K. & FONSECA V. G. 2021. — Metabarcoding Malaise traps and soil eDNA reveals seasonal and local arthropod diversity shifts. *Scientific Reports* 11 (1): 10498. <https://doi.org/10.1038/s41598-021-89950-6>
- KRAAIJEVELD K., WEGER L. A. DE, GARCÍA M. V., BUERMANS H., FRANK J., HIEMSTRA P. S. & DEN DUNNEN J. T. 2015. — Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. *Molecular Ecology Resources* 15 (1): 8-16. <https://doi.org/10.1111/1755-0998.12288>
- KREHENWINKEL H., WEBER S., BROEKMANN R., MELCHER A., HANS J., WOLF R., HOCHKIRCH A., KENNEDY S. R., KOSCHORRECK J., KUENZEL S., MÜLLER C., RETZLAFF R., TEUBNER D., SCHANZER S., KLEIN R., PAULUS M., UDELHOVEN T. & VEITH M. 2022. — Environmental DNA from archived leaves reveals widespread temporal turnover and biotic homogenization in forest arthropod communities. *eLife* 11: e78521. <https://doi.org/10.7554/eLife.78521>
- KRESS W. J., GARCÍA-ROBLEDO C., URIARTE M. & ERICKSON D. L. 2015. — DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology & Evolution* 30 (1): 25-35. <https://doi.org/10.1016/j.tree.2014.10.008>
- LAMB P. D., HUNTER E., PINNEGAR J. K., CREER S., DAVIES R. G. & TAYLOR M. I. 2019. — How quantitative is metabarcoding: a meta-analytical approach. *Molecular Ecology* 28 (2): 420-430. <https://doi.org/10.1111/mec.14920>
- LANTMAN I. M. VAN S., VESTERINEN E. J., HERTZOG L. R., MARTEL A., VERHEYEN K., LENS L. & BONTE D. 2021. — Body size and tree species composition determine variation in prey consumption in a forest-inhabiting generalist predator. *Ecology and Evolution* 11 (12): 8295-8309. <https://doi.org/10.1002/ece3.7659>
- LAVIN J. 2022. — *Environmental DNA metabarcoding and citizen science as a cost-effective and rapid tool for monitoring terrestrial mammalian species*. University of Salford, 105 p.
- LEAR G., DICKIE I., BANKS J., BOYER S., BUCKLEY H.L., BUCKLEY T. R., CRUICKSHANK R., DOPHEIDE A., HANDLEY K. M., HER-

- MANS S., KAMKE J., LEE C. K., MACDIARMID R., MORALES S. E., ORLOVICH D. A., SMISSEN R., WOOD J. & HOLDAWAY R. 2018. — Methods for the extraction, storage, amplification and sequencing of DNA from environmental samples. *New Zealand Journal of Ecology* 42 (1): 10-50A
- LEE P.-S., GAN H. M., CLEMENTS G. R. & WILSON J.-J. 2016. — Field calibration of blowfly-derived DNA against traditional methods for assessing mammal diversity in tropical forests. *Genome* 59 (11): 1008-1022. <https://doi.org/10.1139/gen-2015-0193>
- LEESE F., ALTERMATT F., BOUCHEZ A., EKREM T., HERING D., MEISSNER K., MERGEN P., PAWLOWSKI J., PIGGOTT J., RIMET F., STEINKE D., TABERLET P., WEIGAND A., ABARENKOV K., BEJA P., BERVOETS L., BJÖRNSDÓTTIR S., BOETS P., BOGGERO A., BONES A., BORJA Á., BRUCE K., BURSIĆ V., CARLSSON J., ČIAMPOR F., ČIAMPOROVÁ-ZATOVIČOVÁ Z., COISSAC E., COSTA F., COSTACHE M., CREER S., CSABAI Z., DEINER K., DELVALLS Á., DRAKARE S., DUARTE S., ELERŠEK T., FAZI S., FIŠER C., FLOT J.-F., FONSECA V., FONTANETO D., GRABOWSKI M., GRAF W., GUÐBRANDSSON J., HELLSTRÖM M., HERSHKOVITZ Y., HOLLINGSWORTH P., JAPOSHVILI B., JONES J., KAHLERT M., STROIL B.K., KASAPIDIS P., KELLY M., KELLY-QUINN M., KESKIN E., KÖLJALG U., LJUBEŠIĆ Z., MAČEK I., MÄCHLER E., MAHON A., MAREČKOVÁ M., MEJDANDŽIĆ M., MIRCHEVA G., MONTAGNA M., MORITZ C., MULK V., NAUMOSKI A., NAVODARU I., PADISÁK J., PÁLSSON S., PANKSEK K., PENEV L., PETRUSEK A., PFANNKUCHEN M., PRIMER C., RINKEVICH B., ROTTER A., SCHMIDT-KLOIBER A., SEGURADO P., SPEKSNIJDER A., STOEVE P., STRAND M., ŠULČIUS S., SUNDBERG P., TRAUOGOTT M., TSIGENOPOULOS C., TURON X., VALENTINI A., HOORN B. VAN DER, VÁRBÍRÓ G., HADJILYRA M.V., VIGURI J., VITONYTÉ I., VOGLER A., VRÁLSTAD T., WÄGELE W., WENNE R., WINDING A., WOODWARD G., ZEGURA B. & ZIMMERMANN J. 2016. — DNAqua-Net: developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. *Research Ideas and Outcomes* 2: e11321. <https://doi.org/10.3897/rio.2.e11321>
- LEONARD J. A. 2008. — Ancient DNA applications for wildlife conservation. *Molecular Ecology* 17 (19): 4186-4196. <https://doi.org/10.1111/j.1365-294X.2008.03891.x>
- LEONHARDT F., KELLER A., ARRANZ AVECÉS C. & ERNST R. 2023. — From alien species to alien communities: host- and habitat-associated microbiomes in an alien amphibian. *Microbial Ecology* 86: 2373-2385. <https://doi.org/10.1007/s00248-023-02227-5>
- LEONTIDOU K., VOKOU D., SANDIONIGI A., BRUNO A., LAZARINA M., DE GROEVE J., LI M., VAROTTO C., GIRARDI M., CASIRAGHI M. & CRISTOFORI A. 2021. — Plant biodiversity assessment through pollen DNA metabarcoding in Natura 2000 habitats (Italian Alps). *Scientific Reports* 11 (1): 18226. <https://doi.org/10.1038/s41598-021-97619-3>
- LI W., SONG T., HOU X., QIN M., XU C. & LI Y. 2021. — Application of eDNA metabarcoding for detecting anura on a Tropical Island. *Diversity* 13 (9): 440. <https://doi.org/10.3390/d13090440>
- LI W., QIN M., HOU X., ZHANG J., WANG S., LI Y., LUO Z., DENG T., SONG T., XU C., LIU X., WANG X. & LI Y. 2022. — Application of eDNA metabarcoding for detecting anura in North China. *Asian Herpetological Research* 13 (4): 224-232. <https://doi.org/10.16373/j.cnki.ahr.220021>
- LIU M., CLARKE L. J., BAKER S. C., JORDAN G. J. & C. P. 2020. — A practical guide to DNA metabarcoding for entomological ecologists. *Ecological Entomology* 45 (3): 373-385. <https://doi.org/10.1111/een.12831>
- LOBO J., SHOKRALLA S., COSTA M. H., HAJIBABAEI M. & COSTA F. O. 2017. — DNA metabarcoding for high-throughput monitoring of estuarine macrobenthic communities. *Scientific Reports* 7 (1): 15618. <https://doi.org/10.1038/s41598-017-15823-6>
- LOPES C. M., SASSO T., VALENTINI A., DEJEAN T., MARTINS M., ZAMUDIO K. R. & HADDAD C. F. B. 2017. — eDNA metabarcoding: a promising method for anuran surveys in highly diverse tropical forests. *Molecular Ecology Resources* 17 (5): 904-914. <https://doi.org/10.1111/1755-0998.12643>
- LOPES C. M., BAËTA D., VALENTINI A., LYRA M. L., SABBAG A. F., GASPARINI J. L., DEJEAN T., HADDAD C. F. B. & ZAMUDIO K. R. 2021. — Lost and found: frogs in a biodiversity hotspot rediscovered with environmental DNA. *Molecular Ecology* 30 (13): 3289-3298. <https://doi.org/10.1111/mec.15594>
- LOW L. & TAMMI M. T. 2017. — Introduction to next generation sequencing technologies, in LOW L. & TAMMI M. T. (éds), *Bioinformatics. A Practical Handbook of Next Generation Sequencing and Its Applications*. World Scientific, Singapour: 1-21 https://doi.org/10.1142/9789813144750_0001
- LYET A., PELLISSIER L., VALENTINI A., DEJEAN T., HEHMEYER A. & NAIDOO R. 2021. — eDNA sampled from stream networks correlates with camera trap detection rates of terrestrial mammals. *Scientific Reports* 11 (1): 11362. <https://doi.org/10.1038/s41598-021-90598-5>
- LYMAN J. A., SANCHEZ D. E., HERSHAUER S. N., SOBEK C. J., CHAMBERS C. L., ZAHRATKA J. & WALKER F. M. 2022. — Mammalian eDNA on herbaceous vegetation? Validating a qPCR assay for detection of an endangered rodent. *Environmental DNA* 11: 11362. <https://doi.org/10.1002/edn3.331>
- LYNGGAARD C., BERTELSEN M. F., JENSEN C. V., JOHNSON M. S., FRØSLEV T. G., OLSEN M. T. & BOHMANN K. 2022. — Airborne environmental DNA for terrestrial vertebrate community monitoring. *Current Biology* 32 (3): 701-707.e5. <https://doi.org/10.1016/j.cub.2021.12.014>
- LYNGGAARD C., NIELSEN M., SANTOS-BAY L., GASTAUER M., OLIVEIRA G. & BOHMANN K. 2019. — Vertebrate diversity revealed by metabarcoding of bulk arthropod samples from tropical forests. *Environmental DNA* 1 (4): 329-341. <https://doi.org/10.1002/edn3.34>
- MAESTRI S., COSENTINO E., PATERNO M., FREITAG H., GARCÉS J. M., MARCOLUNGO L., ALFANO M., NJUNJIĆ I., SCHILTHUIZEN M., SLIK F., MENEGON M., ROSSATO M. & DELLEDONNE M. 2019. — A rapid and accurate MinION-Based workflow for tracking species biodiversity in the field. *Genes* 10 (6): 1-11. <https://doi.org/10.3390/genes10060468>
- MARIZZI C., FLORIO A., LEE M., KHALFAN M., GHIBAN C., NASH B., DOREY J., MCKENZIE S., MAZZA C., CELLINI F., BARIA C., BEPAT R., COSENTINO L., DVORAK A., GACEVIC A., GUZMAN-MOUMTZIS C., HELLER F., HOLT N. A., HORENSTEIN J., JORALEMON V., KAUR M., KAUR T., KHAN A., KUPPAN J., LAVERTY S., LOCK C., PENA M., PETRYCHYN I., PUTHENKALAM I., RAM D., RAMOS A., SCOCA N., SIN R., GONZALEZ I., THAKUR A., USMANOV H., HAN K., WU A., ZHU T. & MICKLOS D.A. 2018. — DNA barcoding Brooklyn (New York): a first assessment of biodiversity in Marine Park by citizen scientists. *PLOS ONE* 13 (7): e0199015. <https://doi.org/10.1371/journal.pone.0199015>
- MARQUINA D., ESPARZA-SALAS R., ROSLIN T. & RONQUIST F. 2019. — Establishing arthropod community composition using metabarcoding: Surprising inconsistencies between soil samples and preservative ethanol and homogenate from Malaise trap catches. *Molecular Ecology Resources* 19 (6): 1516-1530. <https://doi.org/10.1111/1755-0998.13071>
- MARQUINA D., ROSLIN T., ŁUKASIK P. & RONQUIST F. 2022. — Evaluation of non-destructive DNA extraction protocols for insect metabarcoding: gentler and shorter is better. *Metabarcoding and Metagenomics* 6: e78871. <https://doi.org/10.3897/mbmg.6.78871>
- MARSHALL N. T., VANDERPLOEG H. A. & CHAGANTI S. R. 2021. — Environmental (e)RNA advances the reliability of eDNA by predicting its age. *Scientific Reports* 11 (1): 2769. <https://doi.org/10.1038/s41598-021-82205-4>
- MARTINS F. M. S., GALHARDO M., FILIPE A. F., TEIXEIRA A., PINHEIRO P., PAUPÉRIO J., ALVES P. C. & BEJA P. 2019. — Have the cake and eat it: Optimizing nondestructive DNA metabarcoding of macroinvertebrate samples for freshwater biomonitoring.

- ring. *Molecular Ecology Resources* 19 (4): 863-876. <https://doi.org/10.1111/1755-0998.13012>
- MARTONI F., PIPER A. M., RODONI B. C. & BLACKET M. J. 2022. — Disentangling bias for non-destructive insect metabarcoding. *PeerJ* 10: e12981. <https://doi.org/10.7717/peerj.12981>
- MATHON L., VALENTINI A., GUÉRIN P.-E., NORMANDEAU E., NOEL C., LIONNET C., BOULANGER E., THUILLER W., BERNATCHEZ L., MOUILLOT D., DEJEAN T. & MANEL S. 2021. — Benchmarking bioinformatic tools for fast and accurate eDNA metabarcoding species identification. *Molecular Ecology Resources* 21 (7): 2565-2579. <https://doi.org/10.1111/1755-0998.13430>
- MAUVISSEAU Q., COIGNET A., DELAUNAY C., PINET F., BOUCHON D. & SOUTY-GROSSET C. 2018. — Environmental DNA as an efficient tool for detecting invasive crayfishes in freshwater ponds. *Hydrobiologia* 805 (1): 163-175. <https://doi.org/10.1007/s10750-017-3288-y>
- MAUVISSEAU Q., DAVY-BOWKER J., BULLING M., BRYNS R., NEYRINCK S., TROTH C. & SWEET M. 2019. — Combining ddPCR and environmental DNA to improve detection capabilities of a critically endangered freshwater invertebrate. *Scientific Reports* 9 (1): 14064. <https://doi.org/10.1038/s41598-019-50571-9>
- MAUVISSEAU Q., HARPER L. R., SANDER M., HANNER R. H., KLEYER H. & DEINER K. 2022. — The multiple states of environmental dna and what is known about their persistence in aquatic environments. *Environmental Science & Technology* 56 (9): 5322-5333. <https://doi.org/10.1021/acs.est.1c07638>
- MCDONALD R., BATEMAN P. W., COOPER C., VAN DER HEYDE M., MOUSAVI-DERAZMAHALLEH M., HEDGES B. A., GUZIK M. T. & NEVILL P. 2023. — Detection of vertebrates from natural and artificial inland water bodies in a semi-arid habitat using eDNA from filtered, swept, and sediment samples. *Ecology and Evolution* 13 (4): e10014. <https://doi.org/10.1002/ece3.10014>
- McHUGO G. P., DOVER M. J. & MacHUGH D. E. 2019. — Unlocking the origins and biology of domestic animals using ancient DNA and paleogenomics. *BMC Biology* 17 (1): 98. <https://doi.org/10.1186/s12915-019-0724-7>
- MENA J. L., YAGUI H., TEJEDA V., BONIFAZ E., BELLEMAIN E., VALENTINI A., TOBLER M. W., SÁNCHEZ-VENDIZÚ P. & LYET A. 2021. — Environmental DNA metabarcoding as a useful tool for evaluating terrestrial mammal diversity in tropical forests. *Ecological Applications* 31 (5): e02335. <https://doi.org/10.1002/eap.2335>
- MÉTRIS K. L. & MÉTRIS J. 2023. — Aircraft surveys for air eDNA: probing biodiversity in the sky. *PeerJ* 11: e15171. <https://doi.org/10.7717/peerj.15171>
- MEYER R., RAMOS M., LIN M., SCHWEIZER T., GOLD Z., RAMOS D., SHIRAZI S., KANDLIKAR G., KWAN W., CURD E., FREISE A., PARKER J., SEXTON J., WETZER R., PENTCHEFF N., WALL A., PIPES L., GARCIA-VEDRENNE A., MEJIA M., MOORE T., ORLAND C., BALLARE K., WORTH A., BERAUT E., ARONSON E., NIELSEN R., LEWIN H., BARBER P., WALL J., KRAFT N., SHAPIRO B. & WAYNE R. 2021. — The CALeDNA program: Citizen scientists and researchers inventory California's biodiversity. *California Agriculture* 75 (1): 20-32. <https://doi.org/10.3733/ca.2021a0001>
- MILIÁN-GARCÍA Y., YOUNG R., MADDEN M., BULLAS-APPLETON E. & HANNER R. H. 2021. — Optimization and validation of a cost-effective protocol for biosurveillance of invasive alien species. *Ecology and Evolution* 11 (5): 1999-2014. <https://doi.org/10.1002/ece3.7139>
- MORI E., BAERI A., SFORZI A., VITALE A. & GALIMBERTI A. 2017. — From accidental citizen-science observations to genetic confirmation: how to spot new hidden invaders. *Hystrix, the Italian Journal of Mammalogy* 28 (2): 284-287. <https://doi.org/10.4404/hystrix-28.2-12421>
- NELL C. S., PRATT R., BURGER J., PRESTON K. L., TRESEDER K. K., KAMADA D., MOORE K. & MOONEY K. A. 2023. — Consequences of arthropod community structure for an at-risk insectivorous bird. *PLOS ONE* 18 (2): e0281081. <https://doi.org/10.1371/journal.pone.0281081>
- NEWTON J. P., BATEMAN P. W., HEYDENRYCH M. J., MOUSAVI-DERAZMAHALLEH M. & NEVILL P. 2022. — Home is where the hollow is: revealing vertebrate tree hollow user biodiversity with eDNA metabarcoding. *Environmental DNA* 4 (5): 1078-1091. <https://doi.org/10.1002/edn3.313>
- NICHOLS R. V., KÖNIGSSON H., DANELL K. & SPONG G. 2012. — Browsed twig environmental DNA: diagnostic PCR to identify ungulate species. *Molecular Ecology Resources* 12 (6): 983-989. <https://doi.org/10.1111/j.1755-0998.2012.03172.x>
- NORDSTROM B., MITCHELL N., BYRNE M. & JARMAN S. 2022. — A review of applications of environmental DNA for reptile conservation and management. *Ecology and Evolution* 12 (6): e8995. <https://doi.org/10.1002/ece3.8995>
- OLIVERIO A.M., GAN H., WICKINGS K. & FIERER N. 2018. — A DNA metabarcoding approach to characterize soil arthropod communities. *Soil Biology and Biochemistry* 125: 37-43. <https://doi.org/10.1016/j.soilbio.2018.06.026>
- ORLANDO L., ALLABY R., SKOGLUND P., DER SARKISSIAN C., STOCKHAMMER P. W., ÁVILA-ARCOS M. C., FU Q., KRAUSE J., WILLERSLEV E., STONE A. C. & WARINER C. 2021. — Ancient DNA analysis. *Nature Reviews Methods Primers* 1 (1): 1-26. <https://doi.org/10.1038/s43586-020-00011-0>
- PACIONI C., HUNT H., ALLENTOFT M. E., VAUGHAN T. G., WAYNE A. F., BAYNES A., HAOUCHAR D., DORTCH J. & BUNCE M. 2015. — Genetic diversity loss in a biodiversity hotspot: ancient DNA quantifies genetic decline and former connectivity in a critically endangered marsupial. *Molecular Ecology* 24 (23): 5813-5828. <https://doi.org/10.1111/mec.13430>
- PASCHER K., ŠVARA V. & JUNGMEIER M. 2022. — Environmental DNA-based methods in biodiversity monitoring of protected areas: application range, limitations, and needs. *Diversity* 14 (6): 463. <https://doi.org/10.3390/d14060463>
- PAWLOWSKI J., APOTHÉLOZ-PERRET-GENTIL L., MÄCHLER E. & ALTERMATT F. 2020. — *Environmental DNA applications For Biomonitoring and Bioassessment in Aquatic Ecosystems*. Federal Office for the Environment (FOEN), Bern, 71 p. <https://doi.org/10.5167/UZH-187800>
- PAWLOWSKI J., BRUCE K., PANKSEP K., AGUIRRE F. I., AMALFITANO S., APOTHÉLOZ-PERRET-GENTIL L., BAUSSANT T., BOUCHEZ A., CARUGATI L., CERMAKOVA K., CORDIER T., CORINALDESI C., COSTA F. O., DANOVARO R., DELL'ANNO A., DUARTE S., EISENDLE U., FERRARI B. J. D., FRONTALINI F., FRÜHE L., HAEGERBAEUMER A., KISAND V., KROLICKA A., LANZÉN A., LEESE F., LEJZEROWICZ F., LYAUTEY E., MAČEK I., SAGOVA-MAREČKOVÁ M., PEARMAN J. K., POCHON X., STOECK T., VIVIEN R., WEIGAND A. & FAZI S. 2022. — Environmental DNA metabarcoding for benthic monitoring: a review of sediment sampling and DNA extraction methods. *Science of The Total Environment* 818: 151783. <https://doi.org/10.1016/j.scitotenv.2021.151783>
- PEDERSEN M. W., OVERBALLE-PETERSEN S., ERMINI L., SARKISSIAN C. D., HAILE J., HELLSTROM M., SPENS J., THOMSEN P. F., BOHMANN K., CAPPELLINI E., SCHNELL I. B., WALES N. A., CARØE C., CAMPOS P. F., SCHMIDT A. M. Z., GILBERT M. T. P., HANSEN A. J., ORLANDO L. & WILLERSLEV E. 2015. — Ancient and modern environmental DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370 (1660): 20130383. <https://doi.org/10.1098/rstb.2013.0383>
- PERSAUD S. F., COTTENIE K. & GLEASON J. E. 2021. — Ethanol eDNA reveals unique community composition of aquatic macroinvertebrates compared to bulk tissue metabarcoding in a biomonitoring sampling scheme. *Diversity* 13 (1): 34. <https://doi.org/10.3390/d13010034>
- PIÑOL J., SENAR M. A. & SYMONDSON W. O. C. 2019. — The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Molecular Ecology* 28 (2): 407-419. <https://doi.org/10.1111/mec.14776>

- POLAND J. A. & RIFE T. W. 2012. — Genotyping-by-sequencing for plant breeding and genetics. *The Plant Genome* 5 (3): 92-102. <https://doi.org/10.3835/plantgenome2012.05.0005>
- PONT D., ROCLE M., VALENTINI A., CIVADE R., JEAN P., MAIRE A., ROSET N., SCHABUSS M., ZORNIG H. & DEJEAN T. 2018. — Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. *Scientific Reports* 8 (1): 10361. <https://doi.org/10.1038/s41598-018-28424-8>
- PORNON A., ANDALO C., BURRUS M. & ESCARAVAGE N. 2017. — DNA metabarcoding data unveils invisible pollination networks. *Scientific Reports* 7 (1): 16828. <https://doi.org/10.1038/s41598-017-16785-5>
- PORT J. A., O'DONNELL J. L., ROMERO-MARACCINI O. C., LEARY P. R., LITVIN S. Y., NICKOLS K. J., YAMAHARA K. M. & KELLY R. P. 2016. — Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology* 25 (2): 527-541. <https://doi.org/10.1111/mec.13481>
- PRIÉ V., VALENTINI A., LOPES-LIMA M., FROUFE E., ROCLE M., POULET N., TABERLET P. & DEJEAN T. 2021. — Environmental DNA metabarcoding for freshwater bivalves biodiversity assessment: methods and results for the Western Palearctic (European sub-region). *Hydrobiologia* 848 (12): 2931-2950. <https://doi.org/10.1007/s10750-020-04260-8>
- PUNTE-SÁNCHEZ F., ARCE-RODRÍGUEZ A., OGGERIN M., GARCÍA-VILLADANGOS M., MORENO-PAZ M., BLANCO Y., RODRÍGUEZ N., BIRD L., LINCOLN S. A., TORNOS F., PRIETO-BALLESTEROS O., FREEMAN K. H., PIEPER D. H., TIMMIS K.N., AMILS R. & PARRO V. 2018. — Viable cyanobacteria in the deep continental subsurface. *Proceedings of the National Academy of Sciences* 115 (42): 10702-10707. <https://doi.org/10.1073/pnas.1808176115>
- PUMKAE0 P., TAKAHASHI J. & IWAHASHI H. 2021. — Detection and monitoring of insect traces in bioaerosols. *PeerJ* 9: e10862. <https://doi.org/10.7717/peerj.10862>
- RAEMY M. & URSENBACHER S. 2018. — Detection of the European pond turtle (*Emys orbicularis*) by environmental DNA: is eDNA adequate for reptiles? *Amphibia-Reptilia* 39 (2): 135-143. <https://doi.org/10.1163/15685381-17000025>
- RATNASINGHAM S. & HEBERT P. D. N. 2007. — BOLD: the Barcode of Life Data System (<http://www.barcodinglife.org>). *Molecular Ecology Notes* 7 (3): 355-364. <https://doi.org/10.1111/j.1471-8286.2007.01678.x>
- RITTER C. D., HÄGGQVIST S., KARLSSON D., SÄÄKSJÄRVI I. E., MUASYA A. M., NILSSON R.H. & ANTONELLI A. 2019. — Biodiversity assessments in the 21 st century: the potential of insect traps to complement environmental samples for estimating eukaryotic and prokaryotic diversity using high-throughput DNA metabarcoding. *Genome* 62 (3): 147-159. <https://doi.org/10.1139/gen-2018-0096>
- ROCHA R. G., JUSTINO J., LEITE Y. L. R. & COSTA L. P. 2015. — DNA from owl pellet bones uncovers hidden biodiversity. *Systematics and Biodiversity* 13 (4): 403-412. <https://doi.org/10.1080/14772000.2015.1044048>
- RODGERS T. W. & JANEČKA J. E. 2013. — Applications and techniques for non-invasive faecal genetics research in felid conservation. *European Journal of Wildlife Research* 59 (1): 1-16. <https://doi.org/10.1007/s10344-012-0675-6>
- ROGER F., GHANAVI H. R., DANIELSSON N., WAHLBERG N., LÖNDAHL J., PETTERSSON L. B., ANDERSSON G. K. S., BOKE OLÉN N. & CLOUGH Y. 2022. — Airborne environmental DNA metabarcoding for the monitoring of terrestrial insects – A proof of concept from the field. *Environmental DNA* 4 (4): 790-807. <https://doi.org/10.1002/edn3.290>
- ROSA L. H., PINTO O. H. B., ŠANTL-TEMKIV T., CONVEY P., CARVALHO-SILVA M., ROSA C. A. & CÂMARA P. E. A. S. 2020. — DNA metabarcoding of fungal diversity in air and snow of Livingston Island, South Shetland Islands, Antarctica. *Scientific Reports* 10 (1): 21793. <https://doi.org/10.1038/s41598-020-78630-6>
- ROUGERIE R., DECAËNS T., DEHARVENG L., PORCO D., JAMES S. W., CHANG C.-H., RICHARD B., POTAPOV M., SUHARDJONO Y. & HEBERT P. D. N. 2009. — DNA barcodes for soil animal taxonomy. *Pesquisa Agropecuária Brasileira* 44 (8): 789-802. <https://doi.org/10.1590/S0100-204X2009000800002>
- RUNTUWENE L. R., TUDA J. S. B., MONGAN A. E. & SUZUKI Y. 2019. — On-Site MinION sequencing, in SUZUKI Y. (éd.), *Single Molecule and Single Cell Sequencing*. Springer, Singapore: 143-150. https://doi.org/10.1007/978-981-13-6037-4_10
- RUPPERT K. M., KLINE R. J. & RAHMAN M. S. 2019. — Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: a systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation* 17: e00547. <https://doi.org/10.1016/j.gecco.2019.e00547>
- SALES N. G., KAIZER M. DA C., COSCIA I., PERKINS J. C., HIGHLANDS A., BOUBLI J. P., MAGNUSSON W. E., DA SILVA M. N. F., BENVENUTO C. & MCDEVITT A. D. 2020. — Assessing the potential of environmental DNA metabarcoding for monitoring Neotropical mammals: a case study in the Amazon and Atlantic forest, Brazil. *Mammal Review* 50 (3): 221-225. <https://doi.org/10.1111/mam.12183>
- SASSO T., LOPES C. M., VALENTINI A., DEJEAN T., ZAMUDIO K. R., HADDAD C. F. B. & MARTINS M. 2017. — Environmental DNA characterization of amphibian communities in the Brazilian Atlantic forest: potential application for conservation of a rich and threatened fauna. *Biological Conservation* 215: 225-232. <https://doi.org/10.1016/j.biocon.2017.09.015>
- SAYERS E. W., CAVANAUGH M., CLARK K., OSTELL J., PRUITT K. D. & KARSCH-MIZRACHI I. 2019. — GenBank. *Nucleic Acids Research* 47 (D1): D94-D99. <https://doi.org/10.1093/nar/gky989>
- SCHNELL I. B., THOMSEN P. F., WILKINSON N., RASMUSSEN M., JENSEN L. R. D., WILLERSLEV E., BERTELSEN M. F. & THOMAS M. GILBERT P. 2012. — Screening mammal biodiversity using DNA from leeches. *Current Biology* 22 (8): R262-R263. <https://doi.org/10.1016/j.cub.2012.02.058>
- SCHNELL I. B., SOLLMANN R., CALVIGNAC-SPENCER S., SIDDALL M. E., YU D. W., WILTING A. & THOMAS M. GILBERT P. 2015. — iDNA from terrestrial haematophagous leeches as a wildlife surveying and monitoring tool – prospects, pitfalls and avenues to be developed. *Frontiers in Zoology* 12 (1): 24. <https://doi.org/10.1186/s12983-015-0115-z>
- SCHULZ F., ALTEIO L., GOUDEAU D., RYAN E. M., YU F. B., MALMSTROM R. R., BLANCHARD J. & WOYKE T. 2018. — Hidden diversity of soil giant viruses. *Nature Communications* 9 (1): 4881. <https://doi.org/10.1038/s41467-018-07335-2>
- SEEBER P. A. & EPP L. S. 2022. — Environmental DNA and metagenomics of terrestrial mammals as keystone taxa of recent and past ecosystems. *Mammal Review* 52 (4): 538-553. <https://doi.org/10.1111/mam.12302>
- SEEBER P. A., VON HIPPEL B., KAUSERUD H., LÖBER U., STOFF-LEICHSENRING K. R., HERZSCHUH U. & EPP L. S. 2022. — Evaluation of lake sedimentary ancient DNA metabarcoding to assess fungal biodiversity in Arctic paleoecosystems. *Environmental DNA* 4 (5): 1150-1163. <https://doi.org/10.1002/edn3.315>
- SEERSHOLM F. V., COLE T. L., GREALY A., RAWLENCE N. J., GREIG K., KNAPP M., STAT M., HANSEN A. J., EASTON L. J., SHEPHERD L., TENNYSON A. J. D., SCOFIELD R. P., WALTER R. & BUNCE M. 2018. — Subsistence practices, past biodiversity, and anthropogenic impacts revealed by New Zealand-wide ancient DNA survey. *Proceedings of the National Academy of Sciences* 115 (30): 7771-7776. <https://doi.org/10.1073/pnas.1803573115>
- SELLERS G. S., MURI C. D., GÓMEZ A. & HÄNFLING B. 2018. — Mu-DNA: a modular universal DNA extraction method adaptable for a wide range of sample types. *Metabarcoding and Metagenomics* 2: e24556. <https://doi.org/10.3897/mbmg.2.24556>
- SHEPHERD L. D. & LAMBERT D. M. 2008. — Ancient DNA and conservation: lessons from the endangered kiwi of New Zealand. *Molecular Ecology* 17 (9): 2174-2184. <https://doi.org/10.1111/j.1365-294X.2008.03749.x>

- SIMS D., SUDBERY I., ILOTT N. E., HEGER A. & PONTING C. P. 2014. — Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics* 15 (2): 121-132. <https://doi.org/10.1038/nrg3642>
- SMITH J. A., THOMAS A. C., LEVI T., WANG Y. & WILMERS C. C. 2018. — Human activity reduces niche partitioning among three widespread mesocarnivores. *Oikos* 127 (6): 890-901. <https://doi.org/10.1111/oik.04592>
- SMITH J. A., THOMAS A. C., LEVI T., WANG Y. & WILMERS C. C. 2018. — Human activity reduces niche partitioning among three widespread mesocarnivores. *Oikos* 127 (6): 890-901. <https://doi.org/10.1111/oik.04592>
- SPECCHIA V., TZAFESTA E., MARINI G., SCARCELLA S., D'ATTIS S. & PINNA M. 2020. — Gap analysis for DNA barcode reference libraries for aquatic macroinvertebrate species in the apulia region (Southeast of Italy). *Journal of Marine Science and Engineering* 8 (7): 538. <https://doi.org/10.3390/jmse8070538>
- SRIVATHSAN A., SHA J. C. M., VOGLER A. P. & MEIER R. 2015. — Comparing the effectiveness of metagenomics and metabarcoding for diet analysis of a leaf-feeding monkey (*Pygathrix nemaeus*). *Molecular Ecology Resources* 15 (2): 250-261. <https://doi.org/10.1111/1755-0998.12302>
- SRIVATHSAN A., LEE L., KATOH K., HARTOP E., KUTTY S. N., WONG J., YEO D. & MEIER R. 2021. — ONTbarcoder and MinION barcodes aid biodiversity discovery and identification by everyone, for everyone. *BMC Biology* 19 (1): 217. <https://doi.org/10.1186/s12915-021-01141-x>
- SRIVATHSAN A., LOH R. K., ONG E. J., LEE L., ANG Y., KUTTY S. N. & MEIER R. 2022. — Network analysis with either Illumina or MinION reveals that detecting vertebrate species requires metabarcoding of iDNA from a diverse fly community. *Molecular Ecology* 32 (23): 6418-6435. <https://doi.org/10.1111/mec.16767>
- STOECKLE M. Y. & THALER D. S. 2018. — Why should mitochondria define species? *Human Evolution* 33 (1-2): 1-30. <https://doi.org/10.14673/HE2018121037>
- SUCHAN T., TALAVERA G., SÁEZ L., RONIKIER M. & VILA R. 2019. — Pollen metabarcoding as a tool for tracking long-distance insect migrations. *Molecular Ecology Resources* 19 (1): 149-162. <https://doi.org/10.1111/1755-0998.12948>
- TABERLET P., COISSAC E., HAJIBABAEI M. & RIESEBERG L. H. 2012a. — Environmental DNA. *Molecular Ecology* 21 (8): 1789-1793. <https://doi.org/10.1111/j.1365-294X.2012.05542.x>
- TABERLET P., PRUD'HOMME S. M., CAMPIONE E., ROY J., MIQUEL C., SHEHZAD W., GIELLY L., RIOUX D., CHOLER P., CLÉMENT J.-C., MELODELIMA C., POMPANON F. & COISSAC E. 2012b. — Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. *Molecular Ecology* 21 (8): 1816-1820. <https://doi.org/10.1111/j.1365-294X.2011.05317.x>
- TAKAHASHI M., SACCO M., KESTEL J. H., NESTER G., CAMPBELL M. A., VAN DER HEYDE M., HEYDENRYCH M. J., JUSZKIEWICZ D. J., NEVILL P., DAWKINS K. L., BESSEY C., FERNANDES K., MILLER H., POWER M., MOUSAVI-DERAZMAHALLEH M., NEWTON J. P., WHITE N. E., RICHARDS Z. T. & ALLENTOFT M. E. 2023. — Aquatic environmental DNA: a review of the macro-organismal biomonitoring revolution. *Science of the Total Environment* 873: 162322. <https://doi.org/10.1016/j.scitotenv.2023.162322>
- TANG M., HARDMAN C. J., JI Y., MENG G., LIU S., TAN M., YANG S., MOSS E. D., WANG J., YANG C., BRUCE C., NEVARD T., POTTS S. G., ZHOU X. & YU D. W. 2015. — High-throughput monitoring of wild bee diversity and abundance via mitogenomics. *Methods in Ecology and Evolution* 6 (9): 1034-1043. <https://doi.org/10.1111/2041-210X.12416>
- THOMSEN P. F., KIELGAST J., IVERSEN L. L., WIUF C., RASMUSSEN M., THOMAS M., GILBERT P., ORLANDO L. & WILLERSLEV E. 2012. — Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology* 21 (11): 2565-2573. <https://doi.org/10.1111/j.1365-294X.2011.05418.x>
- THOMSEN P. F. & SIGSGAARD E. E. 2019. — Environmental DNA metabarcoding of wild flowers reveals diverse communities of terrestrial arthropods. *Ecology and Evolution* 9 (4): 1665-1679. <https://doi.org/10.1002/ece3.4809>
- TINGLEY R., GREENLEES M., OERTEL S., VAN ROOYEN A. R. & WEEKS A. R. 2019. — Environmental DNA sampling as a surveillance tool for cane toad *Rhinella marina* introductions on offshore islands. *Biological Invasions* 21 (1): 1-6. <https://doi.org/10.1007/s10530-018-1810-4>
- TORDONI E., AMETRANO C. G., BANCHI E., ONGARO S., PALLAVICINI A., BACARO G. & MUGGIA L. 2021. — Integrated eDNA metabarcoding and morphological analyses assess spatio-temporal patterns of airborne fungal spores. *Ecological Indicators* 121: 107032. <https://doi.org/10.1016/j.ecolind.2020.107032>
- TSUJI S., TAKAHARA T., DOI H., SHIBATA N. & YAMANAKA H. 2019. — The detection of aquatic macroorganisms using environmental DNA analysis – a review of methods for collection, extraction, and detection. *Environmental DNA* 1 (2): 99-108. <https://doi.org/10.1002/edn3.21>
- USHIO M., FUKUDA H., INOUE T., MAKOTO K., KISHIDA O., SATO K., MURATA K., NIKAIKO M., SADO T., SATO Y., TAKE-SHITA M., IWASAKI W., YAMANAKA H., KONDOH M. & MIYA M. 2017. — Environmental DNA enables detection of terrestrial mammals from forest pond water. *Molecular Ecology Resources* 17 (6): e63-e75. <https://doi.org/10.1111/1755-0998.12690>
- VALENTINI A., TABERLET P., MIAUD C., CIVADE R., HERDER J., THOMSEN P. F., BELLEMMAIN E., BESNARD A., COISSAC E., BOYER F., GABORIAUD C., JEAN P., POULET N., ROSET N., COPP G. H., GENIEZ P., PONT D., ARGILLIER C., BAUDOIN J.-M., PEROUX T., CRIVELLI A. J., OLIVIER A., ACQUEBERGE M., BRUN M. L., MØLLER P. R., WILLERSLEV E. & DEJEAN T. 2016. — Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology* 25 (4): 929-942. <https://doi.org/10.1111/mec.13428>
- VAN DER HEYDE M., BATEMAN P. W., BUNCE M., WARDELL-JOHNSON G., WHITE N. E. & NEVILL P. 2021. — Scat DNA provides important data for effective monitoring of mammal and bird biodiversity. *Biodiversity and Conservation* 30 (12): 3585-3602. <https://doi.org/10.1007/s10531-021-02264-x>
- VAN DER HEYDE M., BUNCE M. & NEVILL P. 2022. — Key factors to consider in the use of environmental DNA metabarcoding to monitor terrestrial ecological restoration. *Science of the Total Environment* 848: 157617. <https://doi.org/10.1016/j.scitotenv.2022.157617>
- VAROTTO C., PINDO M., BERTONI E., CASAROTTO C., CAMIN F., GIRARDI M., MAGGI V. & CRISTOFORI A. 2021. — A pilot study of eDNA metabarcoding to estimate plant biodiversity by an alpine glacier core (Adamello glacier, North Italy). *Scientific Reports* 11 (1): 1208. <https://doi.org/10.1038/s41598-020-79738-5>
- VASAR M., DAVISON J., MOORA M., SEPP S.-K., ANSLAN S., AL-QURAIISHY S., BAHRAM M., BUENO C. G., CANTERO J. J., FABIANO E. C., DECOCQ G., DRENKHAN R., FRASER L., OJA J., GARIBAY-ORIJEL R., HIIESALU I., KOOREM K., MUCINA L., ÖPIK M., PÖLME S., PÄRTEL M., PHOSRI C., SEMCHENKO M., VAHTER T., DOLEŽAL J., PALACIOS A. M. V., TEDERSOO L. & ZOBEL M. 2023. — Metabarcoding of soil environmental DNA to estimate plant diversity globally. *Frontiers in Plant Science* 14: 1-9. <https://doi.org/10.3389/fpls.2023.1106617>
- VILLEN K., CORSE E., MEGLÉCZ E., ARCHAMBAUD-SUARD G., VIGNES H., ERESKOVSKY A. V., CHAPPAZ R. & DUBUT V. 2022. — DNA metabarcoding suggests adaptive seasonal variation of individual trophic traits in a critically endangered fish. *Molecular Ecology* 31 (22): 5889-5908. <https://doi.org/10.1111/mec.16698>
- VIOTTI C., BACH C., MAILLARD F., ZIEGLER-DEVIN I., MIESZKIN S. & BUÉE M. 2021. — Sapwood and heartwood affect differentially bacterial and fungal community structure and successional dynamics during *Quercus petraea* decomposition. *Environmental Microbiology* 23 (10): 6177-6193. <https://doi.org/10.1111/1462-2920.15522>

- WÄGELE J. W., BODESHEIM P., BOURLAT S. J., DENZLER J., DIEPENBROEK M., FONSECA V., FROMMOLT K.-H., GEIGER M. F., GEMEINHOLZER B., GLÖCKNER F. O., HAUCKE T., KIRSE A., KÖLPIN A., KOSTADINOV I., KÜHL H. S., KURTH F., LASSECK M., LIEDEKE S., LOSCH F., MÜLLER S., PETROVSKAYA N., PIOTROWSKI K., RADIG B., SCHERBER C., SCHOPPMANN L., SCHULZ J., STEINHAGE V., TSCHAN G. F., VAUTZ W., VELOTTO D., WEIGEND M. & WILDERMANN S. 2022. — Towards a multisensor station for automated biodiversity monitoring. *Basic and Applied Ecology* 59: 105-138. <https://doi.org/10.1016/j.baae.2022.01.003>
- WATTS C., DOPHEIDE A., HOLDAWAY R., DAVIS C., WOOD J., THORNBURROW D. & DICKIE I. A. 2019. — DNA metabarcoding as a tool for invertebrate community monitoring: a case study comparison with conventional techniques. *Austral Entomology* 58 (3): 675-686. <https://doi.org/10.1111/aen.12384>
- WEIGAND H., BEERMANN A. J., CIAMPOR F., COSTA F.O., CSABAI Z., DUARTE S., GEIGER M. F., GRABOWSKI M., RIMET F., RULIK B., STRAND M., SZUCSICH N., WEIGAND A. M., WILLASSEN E., WYLER S. A., BOUCHEZ A., BORJA A., ČIAMPOROVÁ-ZAŤOVIČOVÁ Z., FERREIRA S., DIJKSTRA K.-D. B., EISENDLE U., FREYHOF J., GADAWSKI P., GRAF W., HAEGERBAEUMER A., VAN DER HOORN B. B., JAPOSHVILI B., KERESZTES L., KESKIN E., LEESE F., MACHER J. N., MAMOS T., PAZ G., PEŠIĆ V., PFANNKUCHEN D. M., PFANNKUCHEN M. A., PRICE B.W., RINKEVICH B., TEIXEIRA M. A. L., VÁRBÍRÓ G. & EKREM T. 2019. — DNA barcode reference libraries for the monitoring of aquatic biota in Europe: gap-analysis and recommendations for future work. *Science of the Total Environment* 678: 499-524. <https://doi.org/10.1016/j.scitotenv.2019.04.247>
- WILSON J.-J., JISMING-SEE S.-W., BRANDON-MONG G.-J., LIM A.-H., LIM V.-C., LEE P.-S. & SING K.-W. 2015. — Citizen Science: the first peninsular Malaysia butterfly count. *Biodiversity Data Journal* 3: e7159. <https://doi.org/10.3897/BDJ.3.e7159>
- XU J. 2016. — Fungal DNA barcoding 1. *Genome* 59 (11): 913-932. <https://doi.org/10.1139/gen-2016-0046>
- YANG C., BOHMAN K., WANG X., CAI W., WALES N., DING Z., GOPALAKRISHNAN S. & YU D. W. 2021. — Biodiversity Soup II: a bulk-sample metabarcoding pipeline emphasizing error reduction. *Methods in Ecology and Evolution* 12 (7): 1252-1264. <https://doi.org/10.1111/2041-210X.13602>
- YATES M. C., FRASER D. J. & DERRY A. M. 2019. — Meta-analysis supports further refinement of eDNA for monitoring aquatic species-specific abundance in nature. *Environmental DNA* 1 (1): 5-13. <https://doi.org/10.1002/edn3.7>
- YATES M. C., DERRY A. M. & CRISTESCU M. E. 2021. — Environmental RNA: a revolution in ecological resolution? *Trends in Ecology & Evolution* 36 (7): 601-609. <https://doi.org/10.1016/j.tree.2021.03.001>
- YOCOZ N. G., BRÄTHEN K. A., GIELLY L., HAILE J., EDWARDS M. E., GOSLAR T., STEDINGK H. V., BRYSTING A. K., COISSAC E., POMPANON F., SØNSTEBØ J. H., MIQUEL C., VALENTINI A., BELLO F. D., CHAVE J., THUILLER W., WINCKER P., CRUAUD C., GAVORY F., RASMUSSEN M., THOMAS M. GILBERT P., ORLANDO L., BROCHMANN C., WILLERSLEV E. & TABERLET P. 2012. — DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology* 21 (15): 3647-3655. <https://doi.org/10.1111/j.1365-294X.2012.05545.x>
- YOO K., LEE T. K., CHOI E. J., YANG J., SHUKLA S. K., HWANG S. & PARK J. 2017. — Molecular approaches for the detection and monitoring of microbial communities in bioaerosols: a review. *Journal of Environmental Sciences* 51: 234-247. <https://doi.org/10.1016/j.jes.2016.07.002>
- YOON S.-H., HA S.-M., KWON S., LIM J., KIM Y., SEO H. & CHUN J. 2017. — Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *International Journal of Systematic and Evolutionary Microbiology* 67 (5): 1613-1617. <https://doi.org/10.1099/ijsem.0.001755>
- YOUNG R. G., MILIÁN-GARCÍA Y., YU J., BULLAS-APPLETON E. & HANNER R. H. 2021. — Biosurveillance for invasive insect pest species using an environmental DNA metabarcoding approach and a high salt trap collection fluid. *Ecology and Evolution* 11 (4): 1558-1569. <https://doi.org/10.1002/ece3.7113>
- YU D. W., JI Y., EMERSON B. C., WANG X., YE C., YANG C. & DING Z. 2012. — Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution* 3 (4): 613-623. <https://doi.org/10.1111/j.2041-210X.2012.00198.x>
- ZAFEIROPOULOS H., VIET H. Q., VASILEIADOU K., POTIRAKIS A., ARVANITIDIS C., TOPALIS P., PAVLOUDI C. & PAFILIS E. 2020. — PEMA: a flexible Pipeline for Environmental DNA Metabarcoding Analysis of the 16S/18S ribosomal RNA, ITS, and COI marker genes. *GigaScience* 9 (3): 1-12. <https://doi.org/10.1093/gigascience/giaa022>
- ZENKER M. M., SPECHT A. & FONSECA V. G. 2020. — Assessing insect biodiversity with automatic light traps in Brazil: pearls and pitfalls of metabarcoding samples in preservative ethanol. *Ecology and Evolution* 10 (5): 2352-2366. <https://doi.org/10.1002/ece3.6042>
- ZHOU J.-L., XU J., JIAO A.-G., YANG L., CHEN J., CALLAC P., LIU Y. & WANG S.-X. 2019. — Patterns of PCR amplification artifacts of the fungal barcode marker in a hybrid mushroom. *Frontiers in Microbiology* 10 (2686): 1-12. <https://doi.org/10.3389/fmicb.2019.02686>
- ZIZKA V. M. A., LEESE F., PEINERT B. & GEIGER M. F. 2018. — DNA metabarcoding from sample fixative as a quick and voucher-preserving biodiversity assessment method. *Genome* 62: 122-136. <https://doi.org/10.1139/gen-2018-0048>

Soumis le 12 septembre 2022;
 accepté le 18 juillet 2023;
 publié le 14 février 2024.

ANNEXE 1. — Glossaire (d'après Creer *et al.* 2016; Deiner *et al.* 2017; Liu *et al.* 2020)

ADN communautaire : l'ADN extrait d'un mélange d'organismes représentant plusieurs espèces de spécimens collectés ensemble (bulk sample ou échantillon mixte) et séparés de l'échantillon environnemental.

ADN environnemental (ADNe) : ADN extrait d'échantillons environnementaux, tels que le sol, l'eau ou l'air, sans isolement préalable des organismes cibles, qui ont relâché de l'ADN dans leur environnement.

Amplification par PCR : utilisation de l'amplification en chaîne par polymérase ou PCR pour créer de nombreuses copies d'un fragment d'ADN cible avec une enzyme polymérase. Le fragment est ciblé à l'aide d'une paire d'amorces (avant et arrière) qui se lient au fragment flanquant d'ADN d'intérêt.

Analyse bioinformatique : application d'approches informatiques pour le traitement de données biologiques complexes.

Barcoding ADN : identification des espèces grâce à un marqueur ADN consensuel couramment utilisé (e.g., le gène mitochondrial du cytochrome c oxydase I [COI] est le code-barres ADN standard pour identifier les espèces animales).

Chimère : artefact ADN constitué de deux séquences (ou plus) combinées au cours de l'étape d'amplification PCR.

Échantillon mixte ou composite (i.e. bulk sample) : échantillon contenant plusieurs spécimens collectés ensemble et non triés.

Marqueur ADN : tout gène ou fragment d'ADN ciblé dans le séquençage pour identifier une espèce, un individu ou un génotype.

Metabarcoding ADN : méthode d'identification taxonomique de plusieurs espèces, extraites à partir d'un échantillon d'ADN en vrac des spécimens du mélange d'espèces ou d'ADN environnemental, qui ont été amplifiées par PCR et séquencées sur une plateforme à haut débit.

Profondeur de séquençage : le nombre de séquences d'ADN récupérées (généralement par échantillon) lors du séquençage à haut débit. La profondeur de séquençage est le nombre de lectures uniques qui incluent un nucléotide donné dans la séquence ou nombre de « reads » pour chaque position.

Read ou lecture : succession de paires bases lues par le séquenceur correspondant à tout ou à une partie d'un fragment d'ADN séquencé (la longueur varie selon la technologie).

Séquençage à haut débit ou séquençage de nouvelle génération : technique de séquençage permettant l'analyse simultanée de millions de fragments d'ADN.

Séquençage d'amplicons : séquençage ciblé de fragments (i.e. gène marqueur) d'ADN amplifiés.

Unité taxonomique opérationnelle moléculaire (MOTU) : les MOTUs sont souvent utilisées comme indicateur des espèces taxonomiques dans les études de *metabarcoding* lorsqu'il n'est pas possible de déterminer avec précision les limites des espèces. Les groupes sont identifiés grâce à l'utilisation d'algorithmes de regroupement (clustering) fondé sur un pourcentage prédéfini de similitude des séquences ADN de clusters et de la similitude des séquences ADN en pourcentage prédéfinie.