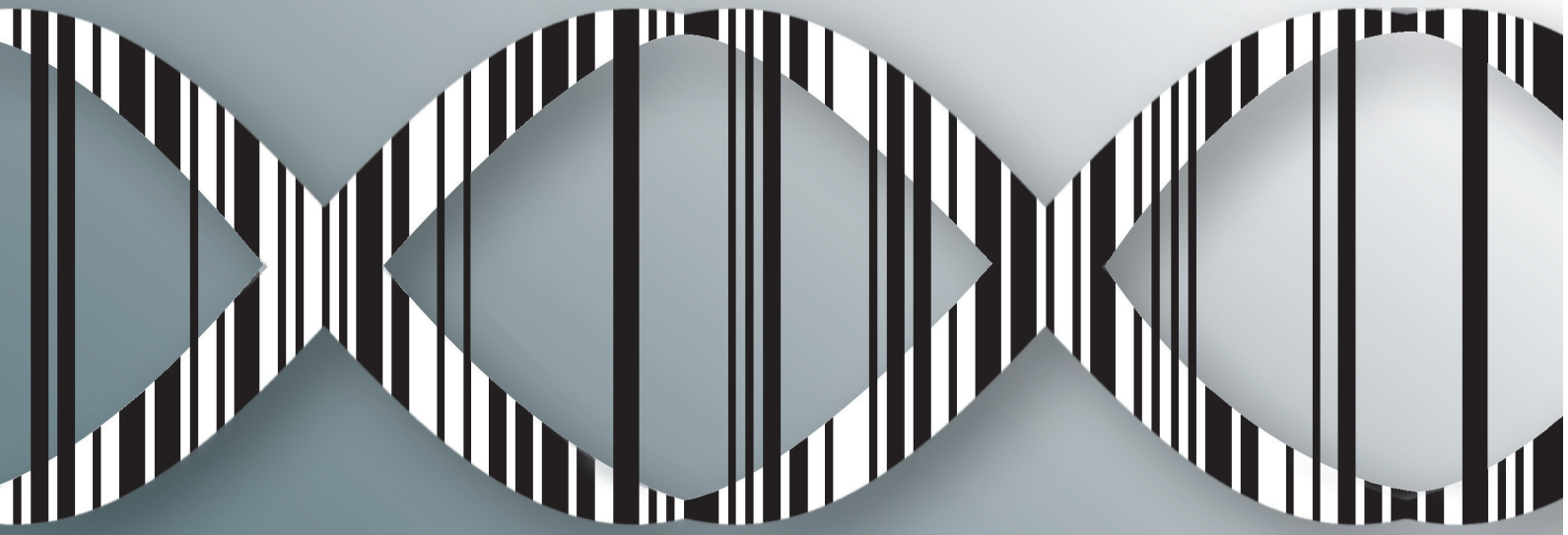


Metabarcoding, une nouvelle façon d'analyser la biodiversité

© DR

La caractérisation de la biodiversité à partir d'ADN issu d'un échantillon d'eau, de sol ou de fèces est maintenant possible grâce aux nouvelles technologies de séquençage. Elle permet d'effectuer des inventaires, de reconstituer des paléo-environnements, de caractériser des régimes alimentaires et offre des perspectives pour la traçabilité par l'ADN et le contrôle d'authenticité des produits agro-alimentaires et cosmétiques.

François Pompanon, Eric Coissac, Pierre Taberlet

Laboratoire d'écologie alpine, CNRS UMR 5553, Université Joseph Fourier, BP 53, 38041 Grenoble cedex 09
francois.pompanon@ujf-grenoble.fr

La caractérisation de la diversité des espèces vivant au sein des écosystèmes revêt un intérêt scientifique majeur pour comprendre le fonctionnement de ces derniers. Elle devient également un enjeu sociétal puisqu'elle est nécessaire pour mettre en œuvre la conservation voire la restauration de la biodiversité. Depuis toujours, les espèces ont été décrites et caractérisées sur la base de critères morphologiques, qui trouvent notamment leurs limites dans les groupes où ils sont peu accessibles, comme c'est le cas pour bon nombre de micro-organismes, ou peu variables, comme chez les nématodes. Ce sont alors des critères moléculaires qui ont été privilégiés (1). L'utilisation d'un fragment d'ADN, permettant de déterminer rapidement et avec fiabilité l'espèce dont il est issu, s'est développée depuis moins de dix ans sous la dénomination de « code-barres ADN », traduction de l'anglais *DNA barcoding*. Une région du gène mitochondrial codant pour la cytochrome oxydase 1, qui présentait plusieurs propriétés intéressantes, a été définie en 2003 comme le fragment de référence pour la caractérisation des espèces animales (2).

Si les taxonomistes utilisent ce concept de code-barres au sens strict, les écologistes en ont une vision plus large correspondant à l'utilisation de toute technique d'analyse de l'ADN pour l'identification de taxons (3). L'objectif est de pouvoir identifier les espèces présentes dans un milieu alors même que les individus ne sont pas facilement caractérisables. On pense évidemment aux micro-organismes mais beaucoup moins aux macro-organismes animaux ou végétaux, dont on peut détecter la présence dans des échantillons environnementaux grâce aux traces d'ADN qu'ils laissent derrière eux : cadavres, mucus, excréments… Le principe consiste à extraire l'ADN d'un échantillon environnemental (eau, sol, fèces), puis à amplifier par PCR le fragment cible correspondant au code-barres à l'aide d'un couple d'amorces prédéfini. Ces amorces peuvent être spécifiques d'une espèce. Cette approche a permis de démontrer qu'il était possible de détecter la présence d'une espèce invasive comme la grenouille taureau (*Rana catesbeiana*), même à faible densité, à

(1) Floyd R *et al.* (2002) *Mol Ecol* 11, 839-50
(2) Hebert PDN *et al.* (2003) *Proc Biol Sci* 270, S96-9
(3) Valentini A *et al.* (2009) *Trends Ecol Evol* 24, 110-7

partir d'échantillons d'eau de mare **(4)**. À l'inverse, les amorces peuvent amplifier un large spectre d'espèces. On parle alors de *metabarcoding*. Dans ce cas, il faut séquencer les amplicons produits par PCR puis les comparer à une base de référence pour les relier à une espèce donnée.

Définir un bon code-barres

Un bon code-barres ADN est une séquence variable entre espèces mais très conservée au sein d'une même espèce, ce qui lui confère un fort pouvoir discriminant. Cette séquence doit être encadrée par deux zones très conservées d'une espèce à l'autre, pour permettre l'amplification du fragment par PCR chez l'ensemble d'espèces le plus vaste possible, et assurer ainsi une bonne couverture taxonomique **(5)**. Il est important aussi que les amplifications soient fiables, d'où la nécessité que les sites de fixation des amorces PCR soient très conservés, et que le fragment amplifié soit court, permettant ainsi de travailler sur des matrices dégradées. En effet, la dégradation conduit à la fragmentation de l'ADN et les fragments de plus de 150 paires de bases (pb) sont difficilement amplifiables. De plus, pour travailler sur de faibles quantités, l'utilisation de fragments d'ADN mitochondriaux ou chloroplastiques est privilégiée car leur nombre de copies par cellule est 100 à 1 000 fois supérieur à celui de l'ADN nucléaire. Il est également utile que les code-barres ADN soient phylogénétiquement informatifs, c'est-à-dire que le niveau de divergence entre ces séquences de référence reflète le niveau de divergence entre les espèces qui les portent. Cette propriété permet d'assigner des espèces inconnues à un taxon d'ordre supérieur (genre, famille, etc.).

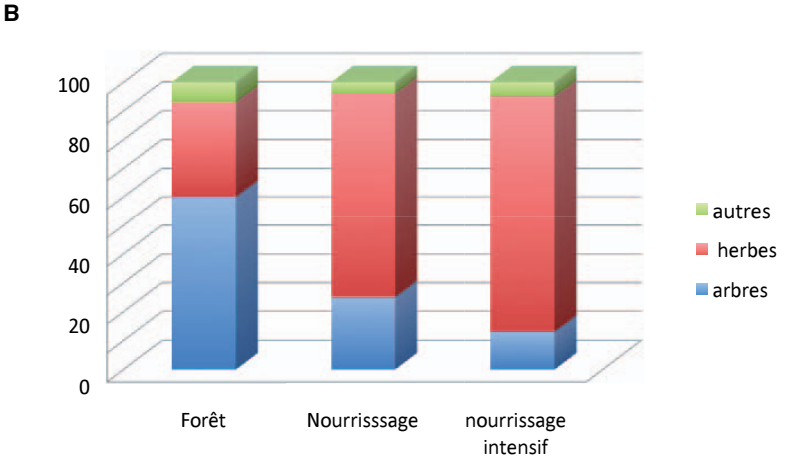
Des outils bio-informatiques ont été développés pour exploiter les bases de données de séquences existantes et définir le code-barres le plus pertinent pour étudier un groupe d'organismes **(figure 1)**. Ils permettent, à partir d'un jeu de séquences de référence prédéfini, de rechercher des zones courtes et variables encadrées par des zones conservées pour choisir les amorces d'amplification les mieux adaptées. Ils sont aussi utilisés pour estimer la qualité des code-barres, en effectuant des PCR in silico pour rechercher toutes les séquences théoriquement amplifiables dans une base de données, et en mesurant le pouvoir discriminant et la couverture taxonomique des code-barres **(6,7)**.

Caractériser la diversité

Un code-barres ADN ainsi défini peut être utilisé pour identifier l'ensemble des taxons présents dans un échantillon environnemental, de la famille à l'espèce. La première étape est donc l'échantillonnage, selon des normes précises afin d'éviter toute contamination. L'étape suivante, l'extraction de l'ADN, suit un protocole adapté au type d'échantillon à étudier. Les ADN extraits servent ensuite de modèle à une amplification par PCR avec les amorces correspondant au code-barres. À cette étape, il est possible de bloquer l'amplification d'une espèce particulière en utilisant une amorce qui vient se positionner spécifiquement sur l'ADN de l'espèce à exclure et empêche l'extension de l'ADN par la polymérase. Cette technique est notamment utilisée pour bloquer l'amplification de l'ADN d'un carnivore dont on étudie le régime alimentaire. Pour ce faire, on a recours à des amorces à large couverture taxonomique amplifiant l'ADN de tous les mammifères. L'amplification préférentielle de l'ADN d'un prédateur présent en grande quantité pourrait en effet masquer la détection de certaines de ses proies.

Après cette étape, chaque produit PCR obtenu est donc un mélange d'amplicons représentatif des ADN des espèces contenus dans l'échantillon de départ. Reste à obtenir la séquence de ces amplicons afin d'identifier les espèces correspondantes. Cela est possible grâce aux nouvelles technologies de séquençage. Elles évitent les étapes de clonage longues et coûteuses qui rendaient jusqu'alors ces expériences inenvisageables dans la pratique. La technologie 454 de Roche Diagnostics, fréquemment utilisée dans les analyses de *metabarcoding*, permet de séquencer un million de molécules d'ADN d'environ 450 pb par expérience. On obtient ainsi, en moyenne, 2 000 séquences par produit PCR lorsque l'on analyse 500 échantillons simultanément. Cette analyse nécessite le mélange équimolaire des produits PCR (multiplexage) avant séquençage des échantillons. Il faut donc garder l'information nécessaire pour attribuer chaque séquence obtenue à son échantillon de départ. Cette information est contenue dans un oligonucléotide, différent pour chaque échantillon, qui joue le rôle d'une étiquette (*tag*) ajoutée aux amorces d'amplification, qui sont, elles, identiques pour tous les échantillons **(8)**. Après le

© R. KOWALCZYK



Régime alimentaire du Bison d'Europe en hiver dans la forêt de Biatowieza Primeval (Pologne)
A. Bison européen *Bison bonasus*
B. Variations selon l'habitat et l'apport de rations de foin (nourrissage) **(14)**

(4) Ficetola F *et al.* (2008) *Biol letters* 4, 423-5
(5) Taberlet *et al.* (2007) *Nucleic Acids Res* 35, doi:10.1093/nar/gkl938
(6) Ficetola F *et al.* (2010) *BMC genomics* 11, 434
(7) www.grenoble.prabi.fr/trac/OBITools

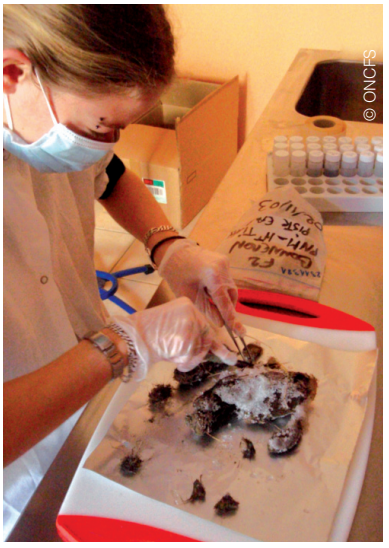
séquençage, les séquences sont triées par échantillon grâce aux *tags*, puis assignées à des taxons par comparaison avec des séquences de référence. Là encore, l'outil bioinformatique est indispensable pour trier les données, constituer les bases de référence, assigner les séquences aux taxons via ces bases, définir des listes de *tags* et gérer les erreurs de séquençage.

La qualité de la base de référence et son exhaustivité par rapport à la zone d'étude conditionne bien entendu la qualité de l'assignation des séquences produites aux espèces. Lors de l'analyse de la diversité en espèces animales et végétales, des bases de données peuvent être générées sans trop de difficulté, l'information phylogénétique contenue dans les code-barres permettant généralement d'identifier le genre ou la famille correspondant à une espèce non référencée. Sans identifier les espèces, on peut utiliser les différentes séquences obtenues pour définir des unités opérationnelles, ou MOTUS (*Molecular Operationnal Taxonomic Units*), à partir desquelles il est également possible de quantifier la biodiversité des échantillons. C'est généralement le cas lorsque l'on travaille sur des micro-organismes dont la plupart des espèces ne sont ni décrites ni cultivables.

De nombreux champs d'application

Ces nouveaux outils biotechnologiques et bio-informatiques offrent des alternatives aux techniques souvent beaucoup plus lourdes à mettre en œuvre qui étaient jusqu'à présent utilisées pour décrire la biodiversité. De nouvelles perspectives s'ouvrent ainsi pour étudier le fonctionnement et l'évolution des écosystèmes, terrestres et aquatiques, avec pour unique prérequis la connaissance des communautés d'espèces interagissant en leur sein. Décrire la biodiversité à partir d'échantillons de terre en utilisant le *metabarcoding* se révèle, entre autres, utile lorsque les individus sont difficiles à trouver et à identifier morphologiquement, comme c'est le cas pour de nombreuses espèces de la faune du sol dont la fonction au sein de l'écosystème est essentielle (vers de terre, insectes, collemboles, etc.). *Le metabarcoding* peut aussi se substituer aux relevés botaniques classiquement utilisés, notamment dans les milieux où la diversité est extrêmement élevée, les forêts tropicales humides par exemple. On estime que l'Amazonie renferme 11 000 espèces d'arbres dont la moitié présente un risque d'extinction **(9)**. Or les méthodes botaniques classiques ne permettent pas de toutes les identifier, ce qui conduit les chercheurs à ignorer involontairement jusqu'à 20 % des genres présents dans leurs études **(10)** ! Le biais d'échantillonnage qui en résulte peut avoir de fortes conséquences sur les résultats. Une situation qui plaide en faveur de l'utilisation de codes-barres ADN pour identifier ces espèces menacées **(11)**.

Le *metabarcoding* environnemental s'avère tout particulièrement utile lorsqu'il s'agit de reconstituer des paléo-environnements et que les espèces recherchées ont disparu. La reconstitution se fait classiquement à partir de l'étude de macrofossiles et de pollens, qui biaise l'échantillonnage et est lourde à mettre en œuvre pour une faible résolution taxonomique. Des études de *metabarcoding* d'échantillons de permafrost datant de plus de 20 000 ans ont démontré une bien meilleure résolution que les relevés polliniques **(12)**.



Méthode d'analyse classique (collecte de reste non digérés dans les fèces) pour la détermination des régimes alimentaires

Le *metabarcoding* se révèle également plus efficace et plus résolutif que les méthodes traditionnelles pour étudier les régimes alimentaires à partir des fèces et des contenus stomacaux. Longue et fastidieuse, l'identification au microscope de fragments de cuticule de plante chez les herbivores ou de restes de proies chez les carnivores n'apporte que des informations très partielles. L'analyse de la composition en alcanes de la cuticule des plantes n'est, elle, performante que pour un nombre de cas limité pour lesquels on compare le profil des déjections à des profils de référence correspondant à quelques mélanges de plantes susceptibles d'avoir été consommées. Quant aux analyses en spectroscopie proche infrarouge, si elles donnent une idée de la composition en nutriments des espèces consommées (azote total, fibres, amidon…), elles ne permettent pas en revanche de les identifier. Le *metabarcoding* utilisé pour reconstituer les régimes alimentaires se pose donc en technique complémentaire, voire alternative, à toutes ces méthodes. Son potentiel pour déterminer les espèces consommées a été démontré chez des animaux d'élevage (vache, mouton) **(13)** et pour une grande variété d'animaux sauvages, des mammifères aux mollusques en passant par les oiseaux et les insectes **(8)**. Chez le bison d'Europe, cette technique a, par exemple, été utilisée pour voir l'influence d'un apport de fourrage sur le régime alimentaire en hiver **(figure p. 31)** **(14)**.

Conclusion

L'approche *metabarcoding* connaît un essor grâce à l'utilisation des nouvelles techniques de séquençage et de la bioinformatique. C'est maintenant une alternative efficace pour décrire la biodiversité à partir d'échantillons environnementaux dans les nombreux cas où les méthodes classiques s'avèrent peu résolutives et fastidieuses. Elle offre également de nouvelles perspectives avec la possibilité de combiner différents codes-barres pour réaliser des études intégrées à partir d'un même échantillon. On peut ainsi concevoir l'analyse simultanée du microbiote intestinal ou ruminal, du cortège parasitaire et du régime alimentaire d'une espèce grâce à ses excréments.

Les champs d'application ne se limitent d'ailleurs pas aux études écologiques et à l'analyse de la biodiversité. Ils s'étendent à l'analyse de tout type de matrice complexe ou transformée contenant de l'ADN, même dégradé et en faible quantité. On pense bien entendu à la médecine légale mais aussi au contrôle d'authenticité et à la traçabilité ADN. Des techniques qui commencent à être utilisées en agro-alimentaire, pour étudier la composition des plats cuisinés ou des fluides de rinçage dans des chaines de production, et en cosmétique, pour analyser la composition d'un produit ou contrôler les matières premières qui le constituent. ●

(8) Valentini A *et al.* (2009) *Mol Ecol Res* 9, 51-60
(9) Hubbell SP *et al.* (2008) *Proc Natl Acad Sci USA* 105, 11498-504
(10) Pitman NCA *et al.* (2008) *Biotropica* 40, 525-35
(11) Gonzalez MA *et al.* (2009) *PLoS ONE* 4, e7483
(12) Sønstebø JH *et al.* (2010) *Mol Ecol Res* 10, 1009-18
(13) Pegard A *et al.* (2009) *J Agric Food Chem* 57, 5700-6
(14) Rafal Kowalczyk *et al.* (2011) *Forest Ecol Manage*, doi:10.1016/j.foreco.2010.11.026