



HAL
open science

Evolution des génomes au travers d'une grande transition écologique

Tristan Lefébure

► **To cite this version:**

Tristan Lefébure. Evolution des génomes au travers d'une grande transition écologique. Biodiversité et Ecologie. Université Claude Bernard Lyon 1, 2025. <tel-05041631>

HAL Id: tel-05041631

<https://hal.science/tel-05041631v1>

Submitted on 24 Sep 2025

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 CC BY-NC 4.0 - Attribution - Non-commercial use - International License

22 janvier, 2025

Habilitation à diriger des recherches

Evolution des génomes au travers d'une grande transition écologique

Tristan Lefébure

Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés (LEHNA),
Université Claude Bernard Lyon 1, CNRS, ENTPE

Soutenue le 25 mars 2025 à Villeurbanne

Jury :

- Lucie BITTNER, Maître de conférences à la Sorbonne Université, examinatrice
- Richard CORDAUX, Directeur de recherche au CNRS, rapporteur
- Nicolas GALTIER, Directeur de recherche au CNRS, rapporteur
- Dominique JOLY, Directrice de recherche au CNRS, examinatrice
- François POMPANON, Professeur des universités à l'Université Grenoble Alpes, rapporteur
- Cristina VIEIRA-HEDDI, Professeure des universités à l'Université Claude Bernard Lyon 1, examinatrice

Table des matières

1. Evolution des génomes au travers d'une grande transition écologique	5
1.1. Construire un modèle génomique en milieu souterrain	5
1.2. Evolution des tailles des génomes	28
1.3. Contraintes en nutriments et composition des protéines	31
1.4. Taux de mutation, traits d'histoire de vie et environnement mutagène	36
1.5. Convergence phénotypique et expression des gènes	42
1.6. Bilan et devenir du modèle Asellidae	50
2. Projets en cours d'émergence : traces ADN et gestion de la ressource en eau	61
2.1. Les acides-nucléiques environnementaux	63
2.2. Mesurer des impacts sans échantillonner les organismes	66
2.3. Dynamique spatio-temporelle des ANe	68
3. La direction de recherche et l'encadrement de doctorant-e	73
4. L'enseignement par la conduite de projets	79
5. CV détaillé	83
6. Remerciements	93
Bibliographie	95

Préambule

J'ai démarré une activité de recherche en 2001 d'abord sur des problématiques de taxonomie moléculaire et de phylogéographie, puis de génomique comparative bactérienne, puis de génomique évolutive en milieu souterrain et enfin de nouveau en écologie moléculaire. Toutes ces activités se sont succédées en se chevauchant plus ou moins longtemps. Bien que l'on puisse intégrer toutes ces activités dans le champ thématique de l'écologie génomique, trouver un fil conducteur n'est pas évident, les résumer en un ensemble cohérent et court est franchement difficile. Ce document n'a pas cette prétention : il se veut relativement court, je l'espère digeste, et se concentrera d'abord sur mes activités de recherche en génomique évolutive en milieu souterrain (2010 à aujourd'hui, chapitre 1). Dans un deuxième chapitre j'aborderai des activités complémentaires en essor qui prennent néanmoins de l'ampleur, à savoir l'étude de l'écologie des ADNe dans les hydrosystèmes. Enfin ce document se terminera par un regard porté sur l'activité de direction de thèse, quelques exemples d'enseignements et enfin par un CV détaillé.

Le choix de ne pas tout aborder est stratégique : ce document doit être lisible. Il ne faut surtout pas y voir une quelconque hiérarchie dans l'importance de mes activités de recherche. J'espère que les personnes (dont des anciens doctorants) avec qui j'ai travaillé mais dont les travaux ne seront pas cités ne le prendront pas mal.

Enfin, en parallèle de mes activités de recherche, je suis enseignant, responsable de formations et de plateforme. Les enseignants-chercheurs ont tendance à segmenter leurs activités de recherche et d'enseignement, mais je pense que souvent elles communiquent et peuvent s'enrichir. C'est globalement mon expérience et là il y aurait aussi beaucoup à dire mais, de nouveau, je ne le ferai pas pour les raisons évoquées précédemment. Rétrospectivement, il en est de même avec mes responsabilités administratives de licence, de master et de plateforme : elles sont en lien avec mes thématiques et ma démarche scientifique. J'espère que même avec une vision partielle, vous y trouverez un peu de cohérence d'ensemble.

1. Evolution des génomes au travers d'une grande transition écologique

Ce chapitre est construit autour des résultats des articles [A32](#), [A33](#), [A34](#) et [A38](#). Il est directement issu des travaux de thèse de Claire Morvan, Clémentine François et Nathanaëlle Saclier dont j'étais co-directeur de thèse. Certains développements méthodologiques et jeux de données sont eux décrits dans les articles [A24](#), [A36](#) et [A45](#). Ces travaux ont été principalement financés par le CNRS (projets Stoichiogenomics, ExoSout, et Stygomics) et l'ANR (projets Convergenomics et NeGA) sur une période de temps allant de 2012 à aujourd'hui. Ces travaux s'insèrent dans un programme de recherche plus large au sein de l'équipe de recherche E3S (Ecologie, Evolution, Ecosystèmes Souterrain) du laboratoire LEHNA sur l'écologie et l'évolution de la famille des Asellidae qui a démarré avec l'ANR DEEP (coordination : Christophe Douady) et les projets européens BioFresh et DARCO (correspondant E3S : Florian Malard).

Je commencerai par présenter rapidement pourquoi et comment un nouveau modèle de génomique évolutive en milieu souterrain a été (et continue d'être) construit. Je présenterai ensuite une sélection de résultats autour de quelques questions générales en biologie : quels processus évolutifs contrôlent la taille de génomes, la qualité et quantité des ressources nutritives influencent-elles la composition des protéines chez les animaux, la radioactivité naturelle et les traits d'histoires de vie influencent-ils les taux de substitutions, et enfin les convergences phénotypiques sont-elles associées à des convergences dans l'expression des gènes ? Je finirai le chapitre avec quelques perspectives ouvertes par ce modèle de transition écologique.

1.1. Construire un modèle génomique en milieu souterrain

1.1.1. Choisir le(s) génome(s)

L'étude des processus évolutifs et de leurs interactions sur l'évolution des génomes s'est traditionnellement faite en utilisant les génomes des organismes modèles. Produire des génomes était coûteux et laborieux. Les biologistes de l'évolution avaient peu de données disponibles, et souvent des données qu'ils ne sélectionnaient pas eux-mêmes. Les choses changent fortement à partir de l'essor du séquençage de deuxième génération dans les années 2000 (Shendure &

Ji 2008). On peut alors commencer à séquencer le génome d'un organisme non plus pour son intérêt médical ou industriel, mais parce que l'organisme qui le porte possède des caractéristiques intéressantes d'un point de vue évolutif ou écologique. J'ai débuté la génomique à cette période : l'université de Cornell venait d'acquérir un séquenceur 454 et un séquenceur Solexa (qui deviendra plus tard Illumina). Nous avons alors séquencé de nombreux génomes d'espèces et de souches bactériennes qui en elles-mêmes n'étaient pas des priorités épidémiologiques mais qui avaient des positions phylogénétiques ou qui occupaient des niches écologiques qui apportaient un regard plus précis sur l'émergence de pathogènes, sur des sauts d'hôtes ou sur l'apparition de nouvelles pathologies.

Pouvoir choisir les taxons possédant des ressources génomiques permet de choisir le cadre phylogénétique de l'étude et de ne pas le subir. Cela permet par exemple de choisir des taxons proches phylogénétiquement mais contrastés pour des traits biologiques ou de par leur environnement. Les longueurs de branches sont bien plus courtes et les événements génomiques qui pourront être observés sur ces branches ont plus de chance d'être associés aux changements de traits étudiés. Cette stratégie d'échantillonnage produit moins de faux positifs même si *in fine* elle reste corrélative. J'ai pu par exemple l'appliquer à l'étude des gains et pertes de gènes ou d'événements de sélection positive sur les patrons de substitution des gènes lors du changement d'hôte (eg. [A11](#), [A13](#), [A17](#)).

Être en capacité de choisir les génomes à séquencer a ouvert la voie à un champs thématique nouveau parfois appelé l'écologie génomique (Fig. 1). Depuis une dizaine d'années on peut plus facilement étudier le lien entre fonctionnement des génomes et écologie des espèces, mais aussi étudier l'évolution des génomes sous différentes conditions écologiques. C'est cette dernière approche que j'ai pu développée en étudiant le passage à la vie souterraine.

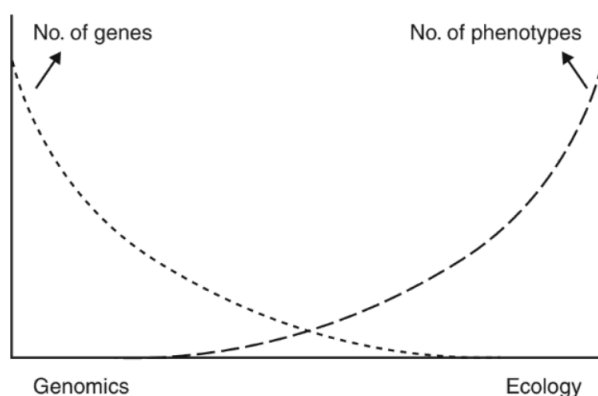


Fig. 1. – Le traditionnel dilemme du biologiste de l'évolution (au moins jusque 2010) : choisir entre l'accès à beaucoup de ressources génétiques ou phénotypiques. L'arrivée du séquençage de 2eme génération résout en partie ce dilemme, permettant l'émergence de l'écologie génomique (qui pourrait être représentée au milieu du graphique). Tiré de Straalen & Roelofs (2011).

¹N.B. Je parlerai souvent de milieux souterrains mais avec un biais évident vers les eaux souterraines. Je pourrais parler directement des eaux souterraines, mais l'écologie des eaux souterraines se comprend difficilement sans prendre en compte les écosystèmes terrestres qui leur sont associés. Par avance, désolé, vous ne trouverez pas beaucoup d'informations sur les écosystèmes souterrains terrestres dans ce document.

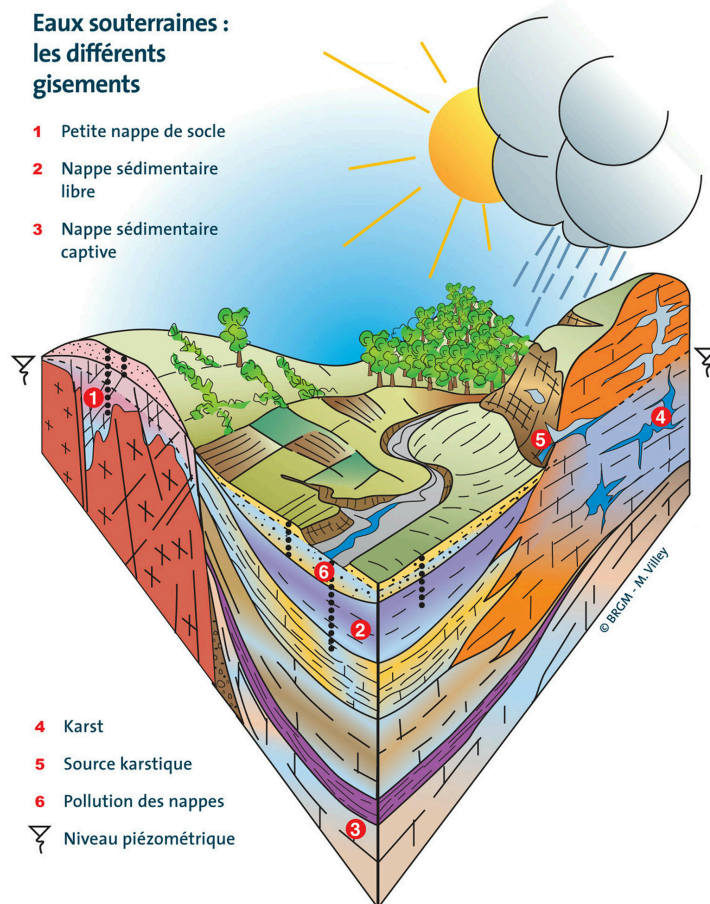


Fig. 2. – Les différents types d'eaux souterraines. Les eaux souterraines se retrouvent dans des roches très variables (consolidées comme les calcaires, ou non consolidées comme les sédiments fluviaux-glaciaires) et en connexion plus ou moins étroite avec les eaux de surface (par exemple une nappe captive est très déconnectée). Illustration du BRGM via Lumni Enseignement.

1.1.2. Choisir le théâtre : la transition souterraine

Les milieux souterrains¹, si on les compare aux milieux de surface, sont d'abord caractérisés par leur oligotrophie. L'absence de lumière fait chuter la production primaire et contraint très fortement les biomasses souterraines. Il existe des exceptions (eg. la grotte de Movile, Chen et al. (2009)), mais les milieux souterrains sont globalement pauvres en ressources nutritives et dépendent fortement des milieux de surface (voir exemple de la nappe de l'Est lyonnais en Fig. 3). L'efficacité des transferts d'énergie depuis les milieux de surface est elle-même contrainte par la connectivité au milieu de surface (Fig. 2) : si la zone non-saturée (zone que les eaux traversent avant d'atteindre la nappe) est peu perméable et épaisse, une grande partie des ressources aura été consommée avant d'atteindre la nappe, et la plupart de la matière organique transférée sera réfractaire et donc difficilement assimilable.

La deuxième caractéristique des milieux souterrains est leur relative stabilité. La plupart des variables environnementales sont stables et ne montrent pas (ou peu) de variations journalières et saisonnières. Même dans des systèmes extrêmement instables comme les rivières glaciaires, les milieux souterrains sont stables (Fig. 4). Il faut néanmoins relativiser cette stabilité, l'évolution du climat impacte par exemple la température moyenne des eaux souterraines, ce sont les variations à courtes échelles temporelles qui sont amorties voir effacées.

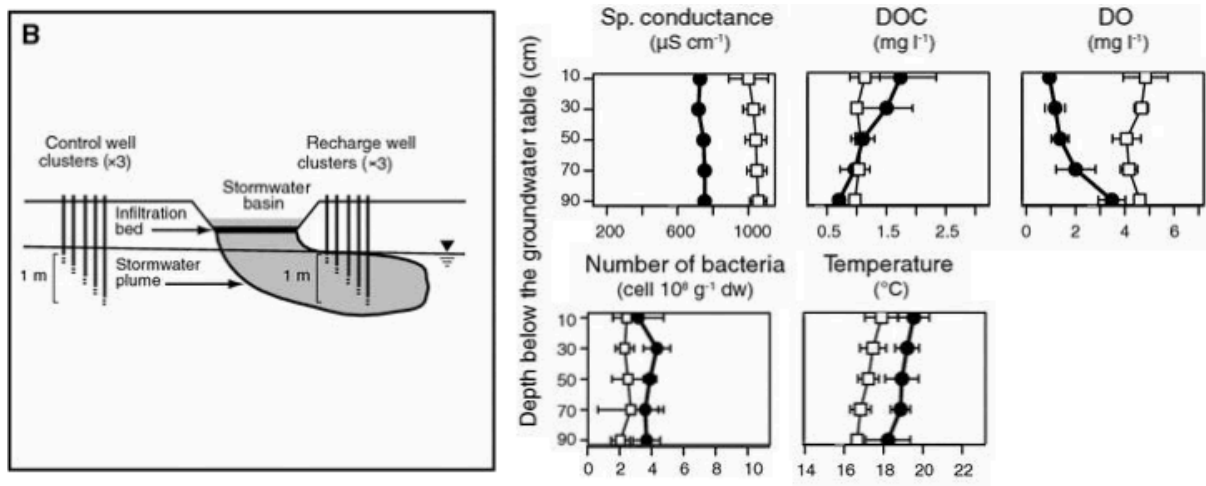


Fig. 3. – Effet de la recharge artificielle en eaux de pluie (simulant une augmentation de connectivité et d’apport en ressources) sur les paramètres d’une nappe d’eau souterraine (ici la nappe de l’Est lyonnais). Les paramètres sont mesurés en amont (condition contrôle, en blanc) et en aval du bassin d’infiltration (condition de forçage, en noir). On observe que la conductivité (directement proportionnelle à l’âge des eaux), la teneur en carbone dissous (DOC, un proxy de la ressource en carbone), la température, la teneur en oxygène et la biomasse microbienne sont impactées par la recharge, démontrant le caractère dépendant des écosystèmes souterrains vis à vis des milieux de surface. En s’infiltrant les eaux de surface plus chaudes et moins minéralisées apportent des ressources en carbone qui stimulent la biomasse microbienne souterraine. La ressource est rapidement consommée et l’effet de forçage s’atténue avec la profondeur (et vraisemblablement avec la distance) pour retrouver les conditions initiales. Modifié de Foulquier et al. (2011).

Il faut aussi renoncer à une vision homogène et simpliste des milieux souterrains : en fonction de leur nature (eg. roches non consolidées versus consolidées) et connexion aux milieux de surface (eg. aquifère profond versus zone hyporhéique), ils peuvent présenter de grandes variations en termes de stabilité, de ressources nutritives, de teneur en oxygène... Il reste néanmoins pertinent de comparer les milieux souterrains aux milieux de surface car le contraste est globalement tellement fort qu’il permet de relativiser les variations entre milieux souterrains.

Ces deux caractéristiques sont supposées avoir de multiples impacts. Tout d’abord, avec moins d’énergie, les écosystèmes souterrains sont supposés supporter une biomasse plus faible avec des réseaux trophiques fortement simplifiés et des tailles de populations plus faibles. Enfin la prédictibilité de ces milieux devrait modifier à la fois les traits d’histoire de vie des organismes mais aussi les taux d’extinction. Certains de ces attendus ont été vérifiés de longue date par exemple sur les changements de traits d’histoire de vie vers des ensembles de traits à l’extrémité du continuum slow-fast (faible fécondité, augmentation de la longévité, ...)², pour d’autres aspects comme les tailles des populations ou les taux d’extinction, les connaissances empiriques sont moins probantes. Ce n’est pas le cas pour le groupe des Asellidae.

1.1.3. Choisir le modèle biologique : les Asellidae

Les milieux souterrains sont un théâtre excellent pour étudier les phénomènes de convergence au sens large (Pipan & Culver 2012), à savoir aussi bien des convergences phénotypiques

²On pourrait néanmoins critiquer un pan important de la littérature qui consiste à comparer quelques taxons souterrains à des taxons de surface phylogénétiquement éloignés, sans aucun contrôle phylogénétique, et qui tire des conclusions hâtives sur l’évolution des traits en milieu souterrain.

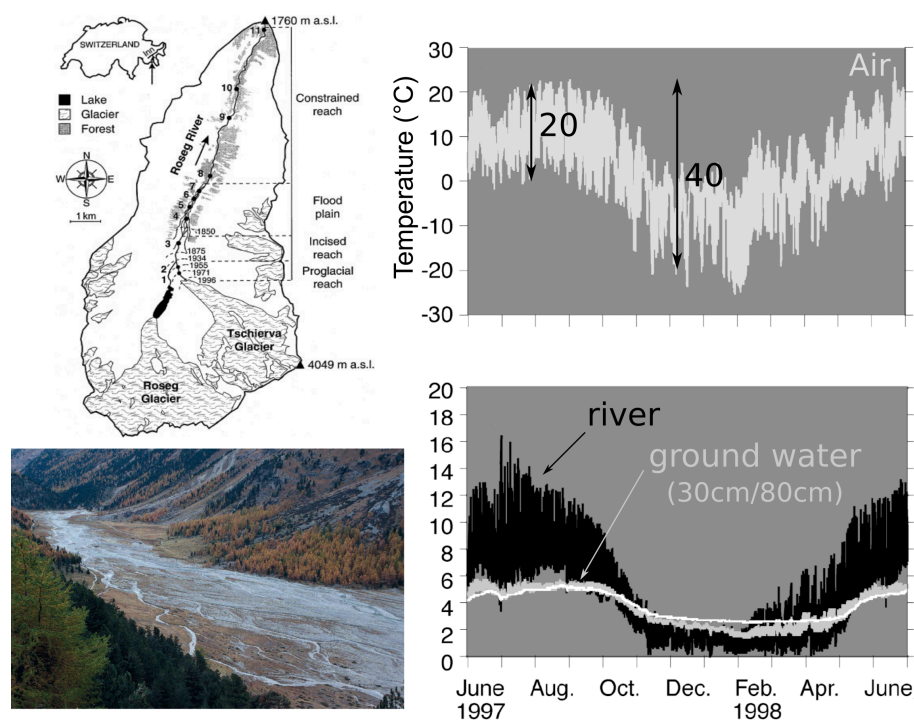


Fig. 4. – Variation quotidienne de la température dans la rivière glacière du Val Roseg (Suisse) pendant une année. La température est mesurée dans l'air (en haut à droite), puis dans l'eau (en bas à droite) de la rivière (en noir), ou à différentes profondeurs dans la zone hyporhéique (à 30 et 80 cm de profondeur). La zone de sédiment noyé qui accompagne et alimente ou est alimentée par une rivière est appelée zone hyporhéique. C'est un des milieux souterrains les plus connectés au milieu de surface (et donc potentiellement un des milieux souterrains les plus instables). A cette altitude la température de l'air montre d'énormes variations journalière (20°C) et saisonnière (40°C). La température des eaux de surface reproduit ces variations avec néanmoins l'effet isolant du manteau neigeux qui amortit en partie les variations journalières en hiver. A 30 cm, et d'autant plus à 80cm, de profondeur dans la zone hyporhéique, une grande partie de ces variations de température est lissée (variations saisonnières) voire gommée (variations journalières). Tiré de Malard et al. (2001) et Malard et al. (2003).

que des convergences d'habitats (obscurité, oligotrophie, stabilité). Néanmoins étudier ce phénomène requiert un élément essentiel supplémentaire, l'accès à un groupe d'organismes ayant (1) colonisé les eaux souterraines à de multiples reprises (et donc formant des réplicats indépendants), (2) au sein duquel il existe toujours un grand nombre d'espèces de surface et souterraines, (3) présentant des évidences fortes de convergences phénotypiques, (4) pour lequel l'histoire évolutive est connue et datée, (5) dont la taxonomie est stable et l'identification des espèces est relativement aisée, (6) et pour lequel l'échantillonnage des organismes ou l'accès à des collections est relativement facile. Ne cherchez pas trop, parmi le pool de modèles biologiques souterrains qui est en expansion rapide (Fig. 5, Mammola et al. (2021)), il n'y a aujourd'hui qu'un seul groupe qui présente l'ensemble de ces caractéristiques.

Histoire évolutive et paires d'espèces contrastées.

La famille des Asellidae (Latreille, 1802) comporte plus de 400 espèces et sous-espèces distribuées dans l'hémisphère Nord. A peu près $\frac{1}{3}$ des espèces sont de surface et $\frac{2}{3}$ sont souterraines (Saclier et al. (2024), Fig. 8). La plupart ont des petites aires de distribution avec une étendue linéaire maximale moyenne de 41 km pour les taxons délimités moléculairement (Eme et al. 2018). A partir de ces connaissances phylogénétiques, il est possible de délimiter 34 paires d'espèces de surface et souterraines formant 34 réplicats phylogénétiquement indépendants

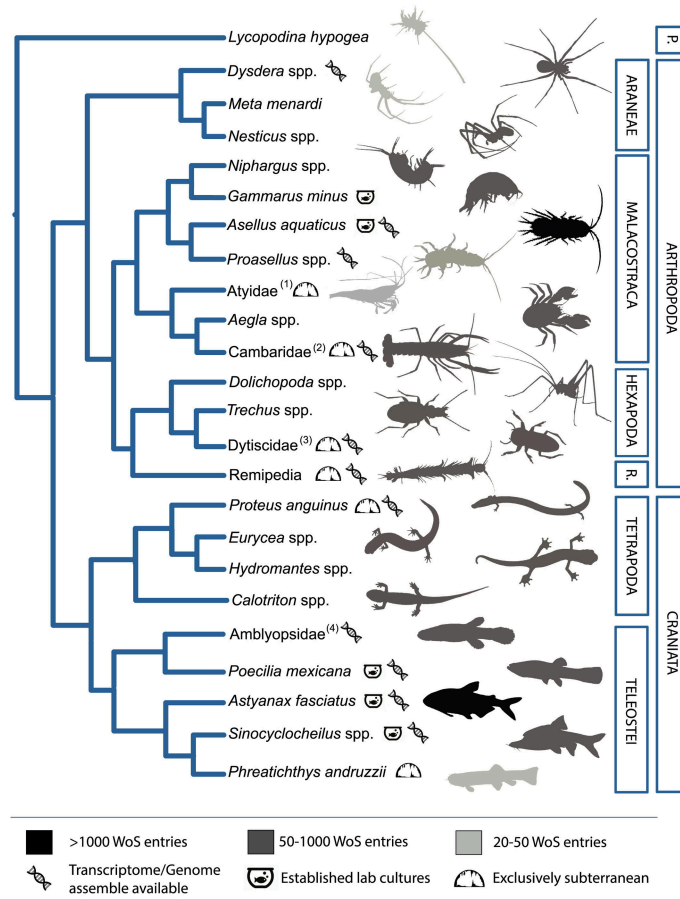


Fig. 5. – Modèles biologiques souterrains définis sur la base du nombre de publications en 2021. Tiré de Mammola et al. (2021).

de contrastes en habitat. La notion de paires phylogénétiquement indépendantes est définie par Felsenstein (2004) ainsi:

An alternative to the contrasts method is to look at pairs of species, chosen from the tree in such a way that the paths between them do not overlap. Looking at pairs of related species is an old method that dates back to at least the work of Salisbury (1942). I have extended it (Felsenstein, 1985a) to allow pairs that are connected by nonintersecting paths on the tree. Figure 25.7 [Fig. 7] shows the choice of such pairs. We can make a sign test of whether two characters change in the same direction (+) or opposite directions

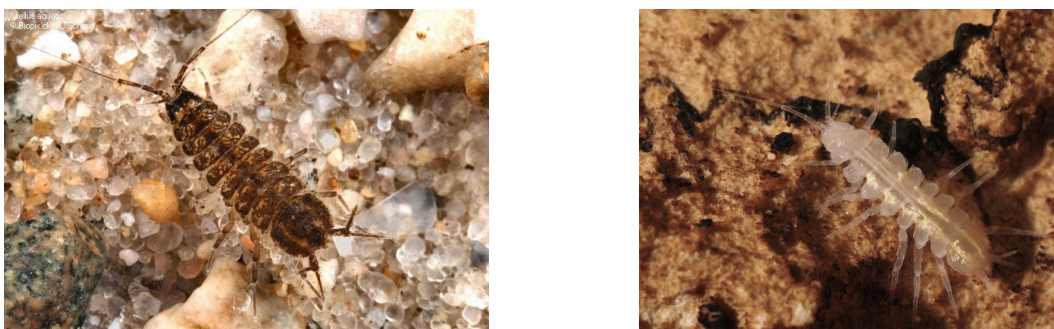


Fig. 6. – Asellidae de surface (*Asellus aquaticus*, à gauche) et souterrain (*Proasellus valdensis*, à droite). Crédit photo : R Lepennec

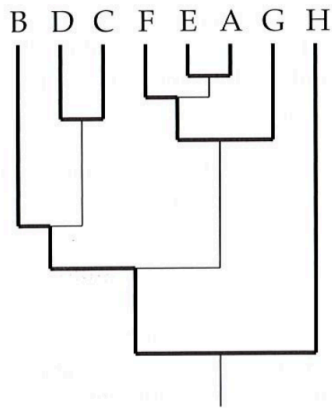


Fig. 7. – Paires d'espèces phylogénétiquement indépendantes selon Felsenstein. La légende originale est « Paires d'espèces choisies de manière à ce que les chemins entre les deux membres de chaque paire ne se chevauchent pas. Les chemins entre les paires d'espèces sont indiqués par les lignes foncées ». On remarquera que les paires n'ont pas besoin d'être monophylétiques. Tiré de Felsenstein (2004)

(-) between members of the pairs. If there is No correlation in the evolution of the two characters, our expectation is that these two outcomes would be equally frequent. So a simple sign test suffices.

– Felsenstein 2004, page 443

Contrairement à des études phylogénétiques comparatives classiques où un groupe d'espèces est échantillonné et pour lesquelles on étudie la relation entre plusieurs traits, l'approches par paires d'espèces contrastées apportent plusieurs avantages statistiques :

1. Le design de l'étude est construit autour d'un nombre de paires connu. La sensibilité et la spécificité de l'étude sont directement dépendantes de ce nombre de paires.
2. Les paires forment des réplicats qui sont analogues aux réplicats dans un design expérimental classique. On peut alors utiliser des approches statistiques très simples comparant le nombre de paires présentant un patron à comparer à une distribution nulle attendue si ce patron est indépendant du contraste testé (voir par exemple les applications à l'étude des taux d'évolution dans Lanfear et al. (2010)).
3. Les approches phylogénétiques comparatives (eg. la régression phylogénétique) ou la modélisation de traits sur les phylogénies restent applicables.

Le syndrome de troglomorphie.

Lors de la colonisation du milieu souterrain, les Asellidae convergent vers un ensemble de traits et caractéristiques :

- diminution voire perte des yeux (Fig. 9)
- perte de la pigmentation (Fig. 9)
- un effacement du dimorphisme sexuel (Saclier et al. 2024)
- la perte du comportement et des structures morphologiques associées au gardiennage des femelles par les mâles
- une diminution de la taille corporelle dans les habitats souterrains à faible porosité (Saclier et al. 2024)
- un allongement du cycle de vie et en particulier une augmentation de la longévité (Fig. 10)
- ...

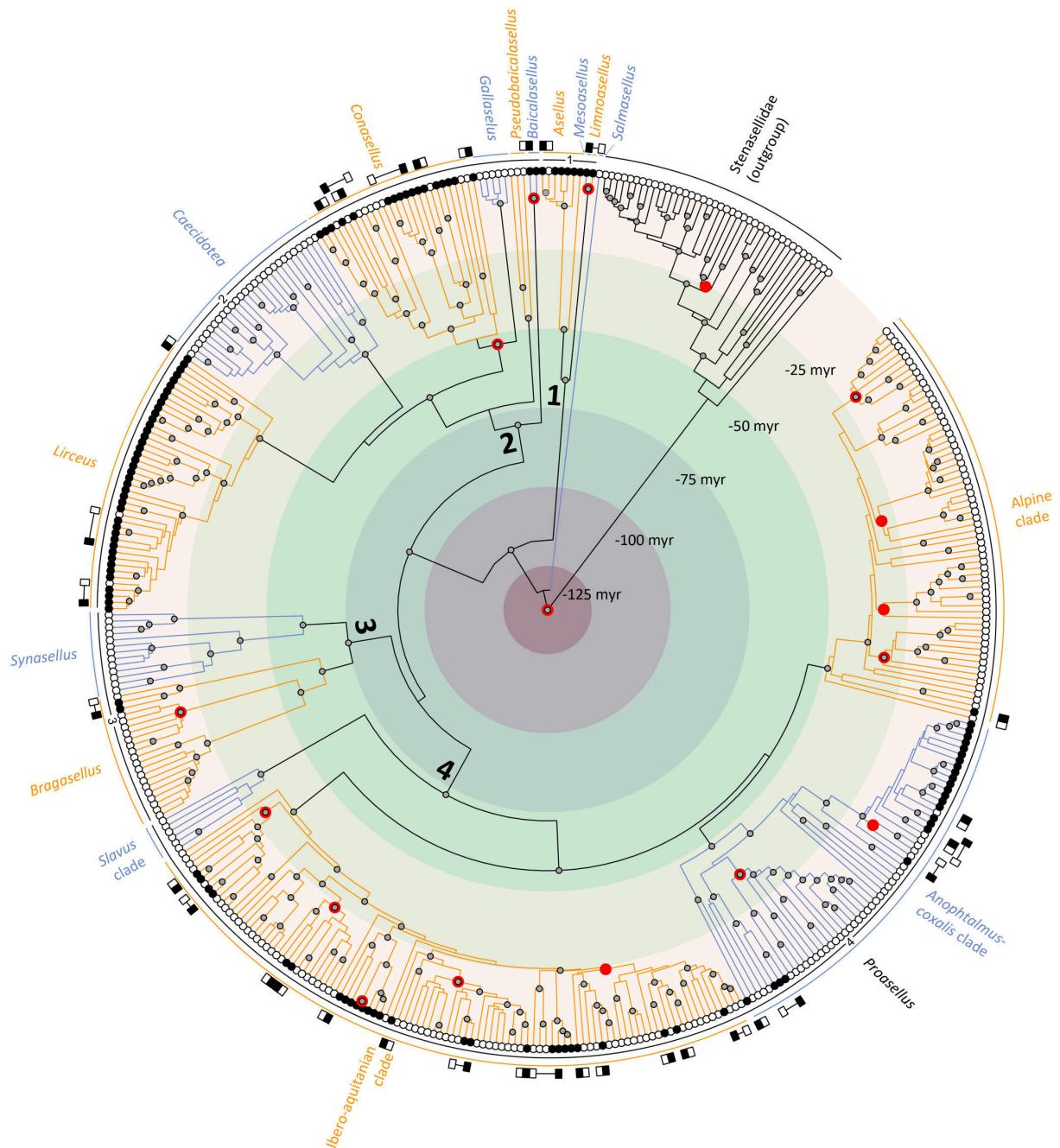


Fig. 8. – Chronogramme des Asellidae. L’arbre est enraciné en utilisant les Stenassellidae comme groupe externe. Les nœuds terminaux sont des unités taxonomiques moléculaires (MOTUs) tels que délimitées par la méthode du seuil fixe de COI (Lefébure et al. 2006). Les nœuds terminaux blancs sont des MOTUs sans yeux et dépigmentés ; les nœuds terminaux noirs sont des MOTUs oculés et/ou pigmentés. Les anneaux de couleur indiquent le temps. Les points rouges et gris indiquent les points de calibration paléobiogéographique et les supports de nœuds avec des probabilités postérieures $> 0,9$, respectivement. Les carrés noirs et blancs de l’anneau extérieur montrent des paires indépendantes d’espèces d’Asellidae de surface (noir) et souterrains (blanc). Les légendes indiquent les genres et les principaux groupes d’espèces au sein de la famille des Asellidae et du genre *Proasellus*. Tiré de Saclier et al. (2024)

Cette liste n’est pas exhaustive et certaines de ces évolutions sont encore décrites sur quelques taxons seulement ou alors seulement à partir d’études populationnelles. La plupart des études populationnelles sont issues de l’espèce *Asellus aquaticus* qui est un modèle en plein développement. Un nombre grandissant d’études l’utilisent pour comprendre le déterminisme

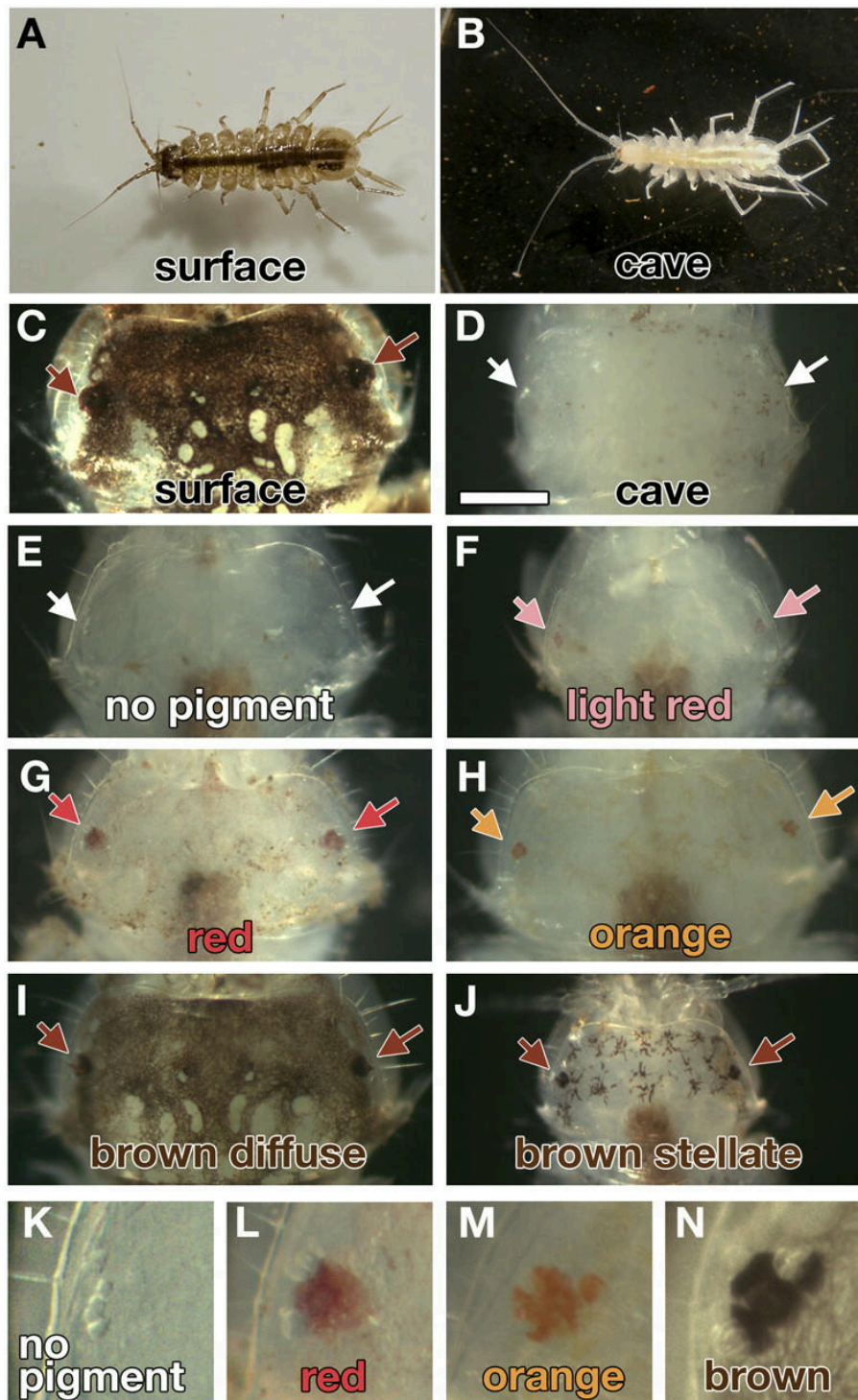


Fig. 9. – Différences de pigmentation et de structure oculaire chez des individus de grotte, de surface et issus de rétrocroisements d'*A. aquaticus*. (A) Mâle de surface provenant des eaux de surface de Planina Polje. (B) Mâle de grotte provenant de la grotte de Planina (canal de Pivka). Têtes des animaux de surface (C), de grotte (D) et rétrocroisés (E-J). Les flèches indiquent les taches oculaires. (E et K) Pas de pigmentation des yeux ou de la tête. (F) Faible œil rouge et pas de pigmentation de la tête. (G et L) Œil rouge et pigmentation de la tête. (H et M) Pigmentation orange de l'œil et de la tête. (I et N) Pigmentation brune de l'œil et pigmentation brune diffuse. (J) Pigmentation brune de l'œil et pigmentation brune stellaire. (Barre d'échelle : 0,25 mm ; s'applique à C-J). L'intestin des individus (E-H et J) est brun car ils mangent des feuilles en décomposition. Tiré de Protas et al. (2011)

génétique de la pigmentation et de la vision (Fiser et al. 2024; Protas et al. 2023; 2011), mais

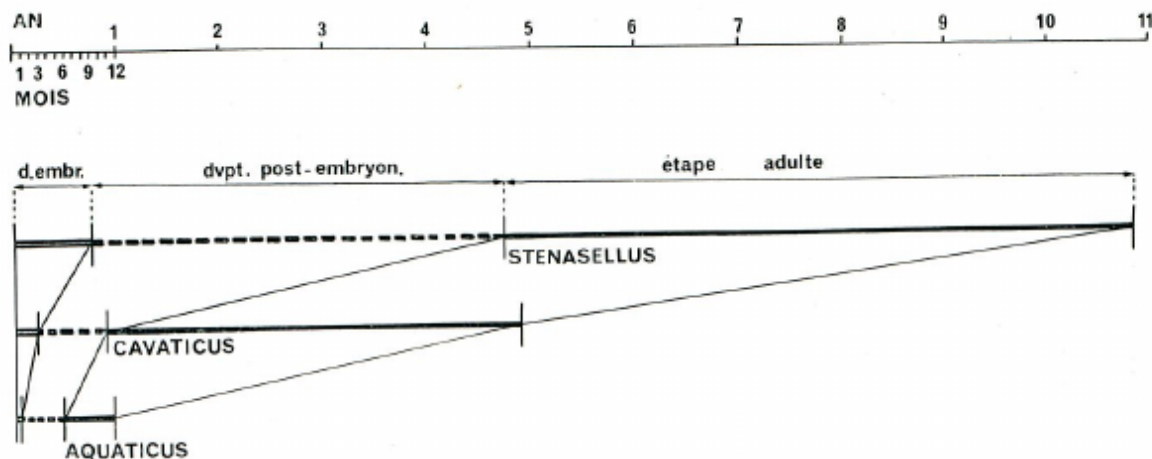


FIG. 43. — Durées relatives comparées des trois périodes principales du cycle vital (développement embryonnaire, développement post-embryonnaire et étape adulte) chez *Stenasellus*, *P. cavaticus* et *A. aquaticus*.

Fig. 10. – Cycles de vie de trois Asellidae différents. Tiré de Henry (1976).

aussi pour aborder des questions qui dépassent la question souterraine comme le lien entre le microbiome digestif et la décomposition de la ligno-cellulose, le contrôle du sexe par des endosymbiotes, les relations hôtes-parasites, les phénomènes de renforcement lors de la spéciation, la sélection sexuelle... (Lafuente et al. 2021). Un petit groupe international de chercheurs³ est en cours de constitution autour de ce modèle. Au sein du complexe d'espèces *Asellus aquaticus*, on retrouve des populations souterraines et de surface, certaines présentant des évolutions de traits plus ou moins marquées, tout en gardant des possibilités de croisement entre individus de populations différentes. On retrouve alors des conditions analogues de celles du téléoste *Astyanax mexicanus* (Protas et al. 2007), conditions idéales pour le développement d'un modèle en génétique et en évo-dévo.

Taille de population et proxy.

La réduction des tailles de population suite à la colonisation du milieu souterrain est sans doute un attendu fréquent mais aussi une des caractéristiques les plus mal cernées. Il y a quelques données disparates et parfois contradictoires sur les tailles des populations souterraines en général (Vörös et al. 2019; Trajano et al. 2010; Buhay & Crandall 2005; Jugovic et al. 2015), mais presque rien sur les Asellidae (Capderrey et al. 2013). Une analyse quantitative et comparative manque cruellement. Chez les Asellidae nous avons abordé ce problème au travers de deux proxy des tailles efficaces de populations : le polymorphisme nucléotidique neutre mesuré à l'échelle du transcriptome et les patrons de substitutions non-synonymes et synonymes.

Polymorphisme. Dans le premier cas, on utilise la relation linéaire entre le taux de mutation populationnel (θ) et le produit du taux de mutation (μ) et de la taille efficace des populations (N_e) : $\theta = 4N_e\mu$. Un estimateur courant de θ est la diversité nucléotidique neutre π_S qui est calculée en mesurant le nombre moyen de polymorphisme entre sites synonymes tirés aléatoirement entre deux chromosomes homologues d'une population. Ce paramètre nécessite le séquençage indépendant des génomes de plusieurs individus d'une population. Un autre estimateur de θ , θ_W , utilise le nombre de sites synonymes qui ségrégent dans une popula-

³<https://www.asellus.org/>

tion : $\hat{\theta}_w = \frac{p_S}{\sum_{i=1}^{2n-1} \frac{1}{i}}$ avec p_S la fréquence des sites synonymes polymorphes et n le nombre d'individus. θ_W a l'avantage de pouvoir être calculé sur des données poolées, ce qui était notre cas pour les premiers transcriptomes que nous avons séquencé (les ARN de plusieurs individus étaient co-extraits, voir la section 1.1.4.1).

Patron de substitutions. En parallèle, à partir des mêmes données de transcriptomes, nous avons calculé le ratio entre les taux de substitution synonymes et non-synonymes (d_N/d_S). Dans le contexte de la théorie de l'évolution quasi-neutre (Ohta 1973; 1992), on peut définir qu'il existe trois grandes catégories de substitutions non-synonymes⁴ (Fig. 11) :

1. des mutations non-synonymes fortement délétères qui sont rapidement purgées par la sélection naturelle. Plus les protéines sont contraintes, plus ce pool de mutation augmente, et plus le d_N et le d_N/d_S baisse.
2. des mutations non-synonymes neutres qui vont ségréger puis éventuellement disparaître ou se fixer dans les populations suivant une dynamique qui sera la même que celle des mutations synonymes. Plus ces mutations sont fréquentes, plus le d_N/d_S augmente.
3. des mutations faiblement délétères⁵ dont la dynamique va dépendre de la taille efficace des populations (N_e). On parle de barrière de la dérive sous laquelle les phénomènes stochastiques dominent, et les mutations faiblement délétères deviennent effectivement neutres (« effectively neutral »), c'est à dire qu'elles se comportent comme des mutations neutres (même probabilité de fixation $\frac{1}{2N_e}$, même temps moyen de fixation $4N_e$...). Cette barrière de la dérive est fixée par le coefficient de sélection (s) et N_e , et se situe autour de $\sim |N_e \cdot s| = 1$. Cette notion est généralisable à tous types de traits ayant un déterminisme génétique et va contraindre l'optimisation de ce trait (Fig. 12). Elle a d'abord été appliquée à l'étude des taux de mutations (Lynch et al. 2016), nous l'appliquerons plus tard aux tailles de génomes et aux biais d'usage des codons synonymes.

Quand une espèce voit sa taille de population être fortement contrainte, la proportion de mutations effectivement neutres augmente. Si la distribution des effets de fitness reste la même, alors le d_N/d_S est censé augmenter. Cet attendu se retrouve dans les modèles d'évolution de séquences quasi-neutres (Nielsen & Yang 2003):

$$\frac{d_N}{d_S} = \frac{2N_e s}{1 - e^{-2N_e s}}$$

Le d_N/d_S est mesuré sur les branches d'une phylogénie en utilisant un modèle d'évolution de codons appliqué à un alignement de séquences codantes. Le d_N/d_S est donc relatif à un clade et un ensemble de gènes. Il sera donc un paramètre qui pourra nous informer sur le N_e moyen sur une période de temps pouvant aller jusqu'à plusieurs millions d'années. Enfin, si on utilise différents sets de gènes, il est attendu qu'on obtienne des d_N/d_S différents pour la même branche. C'est donc un proxy de N_e qui est très intégratif et qui ne peut être comparé qu'avec des d_N/d_S mesurés sur le même set de gènes. Les approches de polymorphismes comme le θ_w sont elles moins dépendantes des régions génomiques étudiées et sont a priori sans cadre temporel. Elles reflètent la dynamique du polymorphisme actuel, donc dans une

⁴mutations dans les séquences codantes changeant la séquence des protéines

⁵la même logique s'applique aux mutations faiblement avantageuses

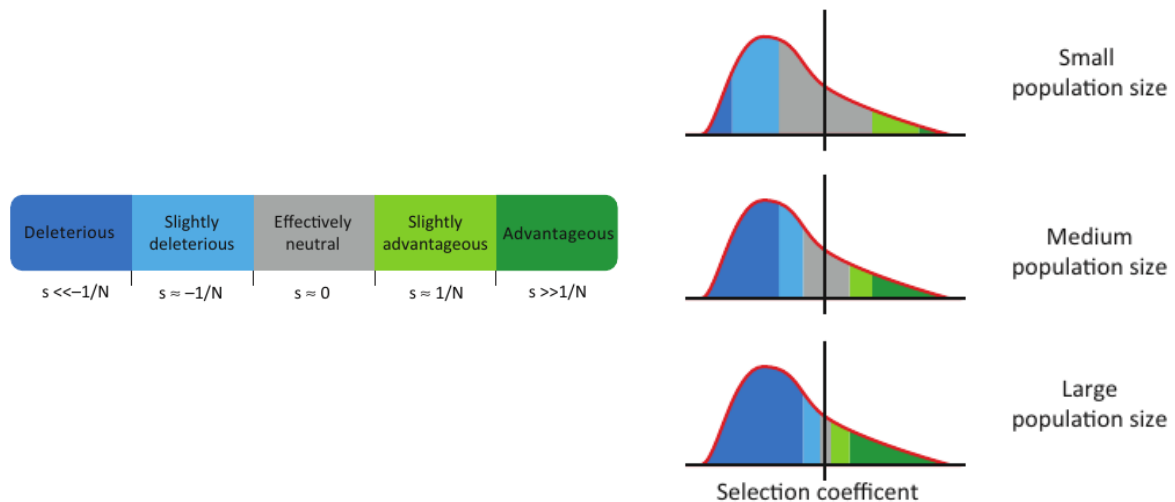


Fig. 11. – Influence de la taille efficace des populations sur la proportion de mutations effectivement neutres. La figure de gauche définit les grandes catégories de mutations en fonction de leur effet sur la fitness et la taille de population. Le graphe de droite représente la distribution des effets de fitness et comment les mutations effectivement neutres envahissent ce spectre quand les tailles de population diminuent. Tiré de Lanfear et al. (2014)

fenêtre temporelle bien plus réduite (depuis le temps de coalescence moyen, $2.N_e$ générations, à aujourd'hui).

Proxy de N_e chez les Asellidae. Chez les Asellidae, à partir de données de transcriptomes pour 22 paires d'espèces de surface et souterraines, nous avons montré une tendance non significative à une baisse du polymorphisme (PGLS, valeur de $p = 0.091$) mais une augmentation significative du d_N/d_S en milieu souterrain (PGLS, valeur de $p = 0.015$, Fig. 13). Huit paires montrent une augmentation du d_N/d_S chez l'espèce souterraine alors qu'une seule paire montre le patron opposé.

Le faible signal observé dans les données de polymorphisme et l'absence de corrélation avec le d_N/d_S peuvent avoir plusieurs causes:

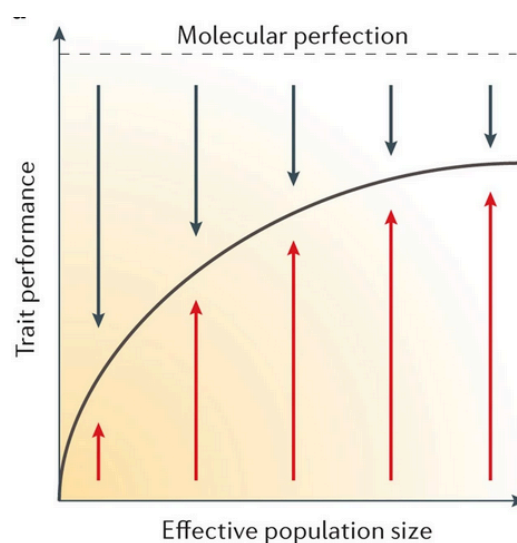


Fig. 12. – Barrière de la dérive et optimisation d'un trait. Les flèches rouges représentent la sélection naturelle, les flèches grises la dérive, la ligne noire la barrière de la dérive. Tiré de Lynch et al. (2016)

- les données de polymorphisme sont plus bruitées : le polymorphisme chez les Asellidés est faible (entre 1‰ et 7‰) et la méthode utilisée (données RNA-seq poolées) est peut-être trop peu précise dans ce cadre.
- θ_w et d_N/d_S ne mesurent pas la même chose : le d_N/d_S mesure l'efficacité de la sélection, et est donc un proxy indirect de N_e pour un clade sur un long pas de temps. θ_w est lui le reflet de la dynamique démographique des populations sur les dernières $2 \cdot N_e$ générations. On parle parfois d'estimateurs de N_e à court et long-terme⁶ qui pourraient être décorrélés si l'estimateur à court-terme est bruité par des événements stochastiques à forte fréquence temporelle.

Depuis que le patron d'augmentation du d_N/d_S chez les espèces souterraines a été observé dans les transcriptomes à partir de 2015, il n'a cessé de se consolider, que ce soit en rajoutant des nouveaux gènes, des nouvelles espèces ou en utilisant des méthodes de calculs plus performantes. La diminution de l'efficacité de la sélection liée à un déplacement de la barrière de la dérive chez les Asellidae souterrains est une observation solide qui nous permet d'utiliser les Asellidae comme un modèle génomique de baisse de N_e sur le long-terme. Le lien au polymorphisme demeure lui plus obscur et nécessite plus d'études (et surtout des données de polymorphismes à l'échelle génomiques).

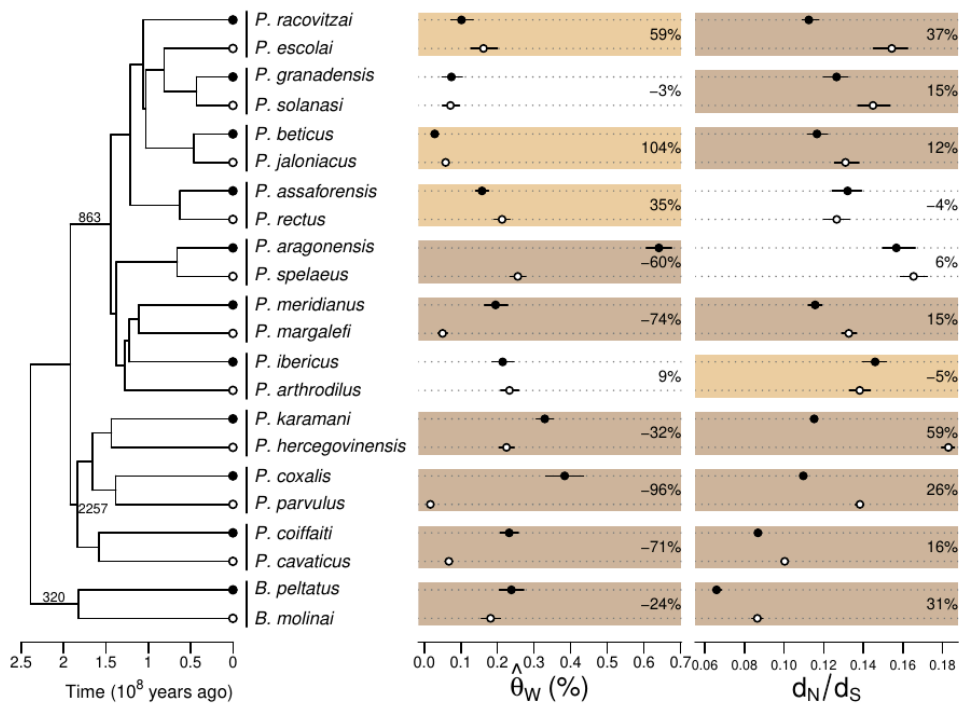


Fig. 13. – Mesure de deux proxy de la taille efficace des populations (N_e) pour 22 espèces d'Asellidae formant 11 paires d'espèces contrastées. Les espèces de surface sont représentées par des ronds noirs, les espèces souterraines par des ronds blancs. P: *Proasellus*, B: *Bragasellus*. Les chiffres sur les branches indiquent le nombre de gènes utilisés pour calculer les d_N/d_S . Quand au sein d'une paire une différence significative est observée dans le sens d'une baisse de N_e , un cartouche foncé est dessiné, un cartouche clair pour la relation inverse, ou rien s'il n'y a pas de différence significative. Tiré de Lefébure et al. (2017)

⁶La définition classique de N_e n'a pas de dimension temporelle, cette différenciation en deux N_e est sujette à discussion... On pourrait aussi argumenter que N_e est un concept flou supportant de multiples définitions toutes plus ou moins liées au modèle théorique (et simpliste) de Wright-Fisher dans lequel l'intensité de la dérive est mesurée par N_e .

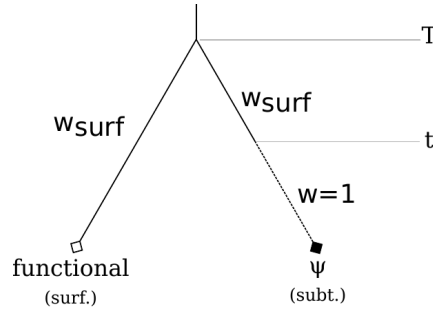


Fig. 14. – Suite à un évènement de spéciation (au temps T), une espèce colonise le milieu souterrain (au temps t), pour former une paire d'espèce souterraine/surface.

Dater la colonisation.

L'analyse par paire d'espèces contrastées permet d'isoler et de tester simplement l'impact d'une transition environnementale, mais elle ne résout que partiellement la problématique du timing de cette transition. En effet au sein d'une paire, on peut différencier le temps de spéciation (T) du temps de colonisation (t , Fig. 14). Alors que t est borné par T ($t \leq T$), la colonisation peut être très récente et dans ce cas la plupart de l'histoire évolutive qui sera comparée sera entre espèce du même habitat. Pour contrôler cet effet, il faudrait connaître (t) ou bien la proportion de la branche qui est réellement souterraine ($\frac{t}{T}$).

Il est possible d'estimer t au travers de traits régressifs, à savoir des traits qui régressent voire disparaissent en milieu souterrain. Un gène codant associé à un trait régressif, s'il ne possède plus de fonction, va évoluer sous un régime de sélection neutre et devenir un pseudo-gène. Il va alors accumuler des mutations qui auparavant étaient délétères comme des mutations non-sens (indel, codons stop) mais aussi un ensemble de mutations non-synonymes. La quantité de ces marques de non-fonctionnalisation sera donc directement proportionnelle à t . Certaines marques comme les délétions sont néanmoins très destructives et peuvent conduire à l'effacement de marques précédentes voire à la disparition totale du pseudo-gène. Je me suis alors concentré sur une méthode utilisant l'accumulation de substitutions non-synonymes et un modèle simple de mélange de deux d_N/d_S : le d_N/d_S du gène fonctionnel des organismes de surface (ω_{surf}) et le d_N/d_S d'un pseudo-gène ($\omega_1 = 1$). On mesure ensuite un d_N/d_S sur la branche souterraine (ω_{subt}) et qui peut être défini ainsi (Fig. 14):

$$\omega_{subt} = \omega_{surf} \frac{T-t}{T} + \omega_1 \frac{t}{T}$$

d'où on peut extraire :

$$t = T \times \frac{\omega_{subt} - \omega_{surf}}{1 - \omega_{surf}}$$

On peut également calculer la proportion de la branche étant souterraine :

$$\frac{t}{T} = \frac{\omega_{subt} - \omega_{surf}}{1 - \omega_{surf}}$$

Cette approche a l'avantage de la simplicité, elle nécessite néanmoins:

- un d_N/d_S de surface stable

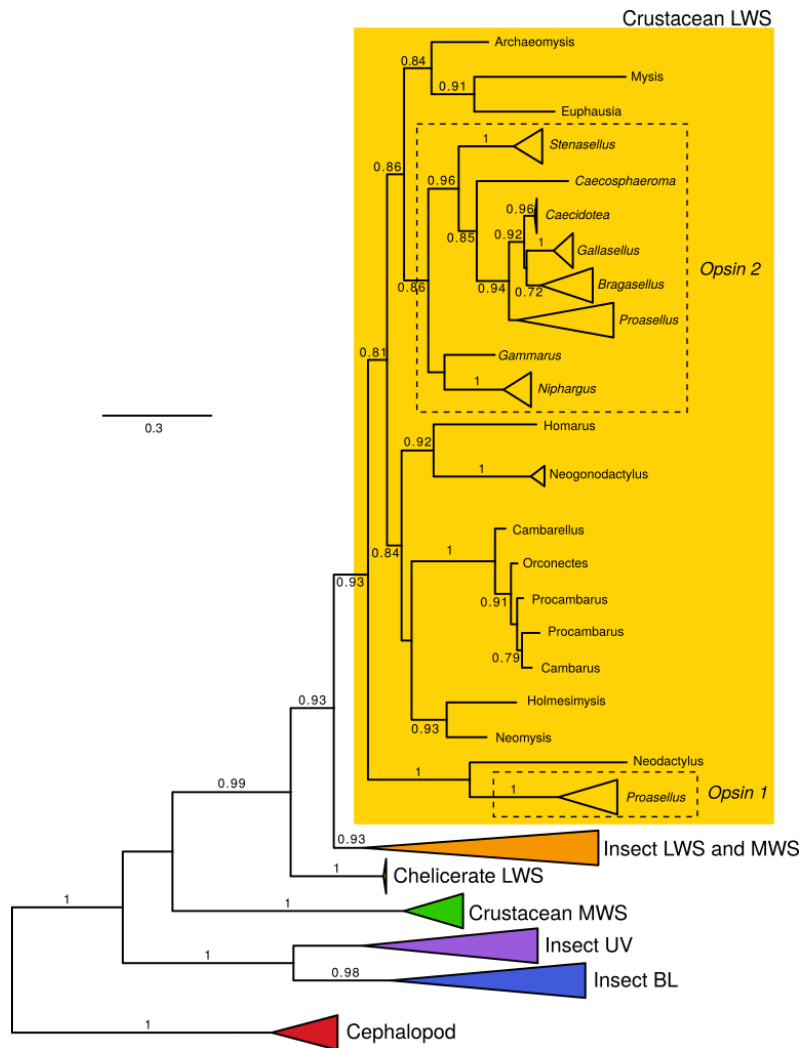


Fig. 15. – Reconstruction de l'évolution de la famille des gènes de l'opsine au sein des arthropodes en utilisant les Céphalopodes comme groupe externe (données de Porter et al. (2007)) et en incluant les données produites au laboratoire pour les Asellidae (opsine₁ et opsine₂, en italique). Les genres monophylétiques ont été réduits. Les groupes d'opsines de Porter sont représentés (nom taxonomique et sensibilité à la longueur d'onde) (LWS : sensible aux grandes longueurs d'onde, MWS : sensible aux longueurs d'onde moyennes, BL : longueur d'onde bleue, UV : sensible aux UV). Arbre ML, WAG+G+I, seules les valeurs aLRT > 0,7 sont représentées.

- un d_S stable, donc un taux de mutation stable⁷
- un gène se pseudogénéisant de façon convergente chez tous les Asellidae lors de la colonisation du milieu souterrain

Plusieurs gènes candidats à la pseudogénéisation ont été cherchés en lien avec le syndrome de troglomorphie : gènes liés à la pigmentation, aux rythmes circadiens, à la photoréception... Après plusieurs tentatives seuls les gènes de la famille des opsines ont montré des évidences de non-fonctionnalisation. D'abord produits à partir de séquençage Sanger, puis à partir de transcriptomes et enfin de génomes, nous connaissons aujourd'hui seulement 2 copies de l'opsine (appelées par la suite opsine₁ et opsine₂). Ces deux copies se retrouvent au sein des

⁷Nous verrons au Chapitre 1.4 que ce n'est pas vrai, le taux de mutation diminue chez les organismes souterrains, le calcul du d_N/d_S de la branche va donc donner plus de poids à la partie de surface, et cette méthode va sous-estimer le temps de colonisation. Il faudrait revoir la méthode pour rajouter deux d_S , un de surface et un souterrain.

opsines de crustacés sensibles aux grandes longueurs d'ondes (Fig. 15). La répartition taxonomique de ces deux copies est complexe : certains genres comme *Proasellus* possèdent les deux, alors que d'autres ne possèdent qu'une copie (la copie 2 pour *Bragasellus*).

Une recherche de ces deux copies a été effectuée chez un grand nombre d'Asellidae regroupés en trois types de phénotypes : les Asellidae oculés, microphthalmes (structures oculaires très réduites) et anophtalmes (absence de structure oculaire). La relation est sans appel : les espèces anophtalmes présentent souvent des séquences non-fonctionnelles (présence de mutations non-sens). De la même manière, quand une opsine n'a pas pu être caractérisée pour une espèce, elle est souvent anophtalme, suggérant que de nombreuses espèces souterraines ont totalement perdu leurs opsines. De nouveau cela dépend des genres : les genres ne possédant que l'opsine₂, comme le genre *Bragasellus* gardent une opsine₂ fonctionnelle même quand l'espèce est anophtalme. Les traces de non-fonctionnalisation sont fréquentes, multiples et rarement partagées entre espèces (voir par exemple Fig. 17 pour l'opsine₂). Cela soutient un scénario dans lequel les espèces non-fonctionnalisent leur(s) opsine(s) indépendamment et ont donc colonisé le milieu souterrain indépendamment. Il y a quelques exceptions comme les différentes unités de *P. beroni* ou de *P. ebrensis* qui partagent des mutations non-fonctionnalisantes, mais cela reste marginal et souvent associé à des espèces proches et cryptiques⁸. En plus d'arbore des mutations non-sens, les espèces souterraines montrent également des changements forts dans les d_N/d_S des opsines (Fig. 17), en particulier pour les branches terminales menant aux espèces souterraines anophtalmes. Les conditions étaient donc réunies pour appliquer la méthode de datation de la colonisation décrite en amont de cette section.

Reste que nous avons donc deux gènes de l'opsine avec des différences notables : l'opsine₁ est principalement présente chez le genre *Proasellus* et se non-fonctionnalise chez les anophtalmes, alors que l'opsine₂ est plus ubiquiste mais ne se non-fonctionnalise pas dans tous les groupes. De plus l'opsine₁ se non-fonctionnalise moins vite que l'opsine₂ (Fig. 18), et en consé-

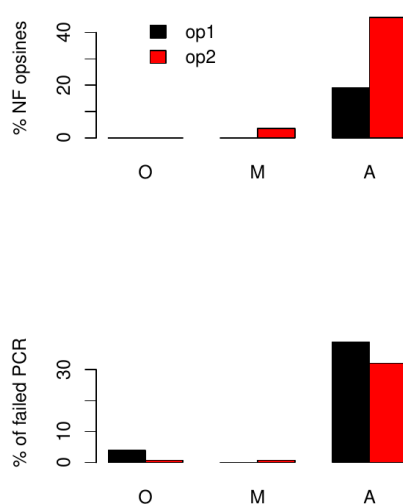


Fig. 16. – Taux de non-fonctionnalisation (NF, en haut) et d'échec de caractérisation (en bas) pour 3 catégories de phénotypes oculaires : Oculé (O), microphthalmal (M) et anophtalme (A). L'opsine₁ est noire, l'opsine₂ en rouge.

⁸d'un point de vue macro-évolutif, cette observation a d'énormes implications. Si la plupart des espèces souterraines le sont devenues indépendamment, cela implique une dynamique de diversification très particulière : les espèces souterraines auraient un taux de spéciation et d'extinction faible, alors que cela serait l'inverse pour les espèces de surface. Une modélisation de ce processus reste à être implémentée...

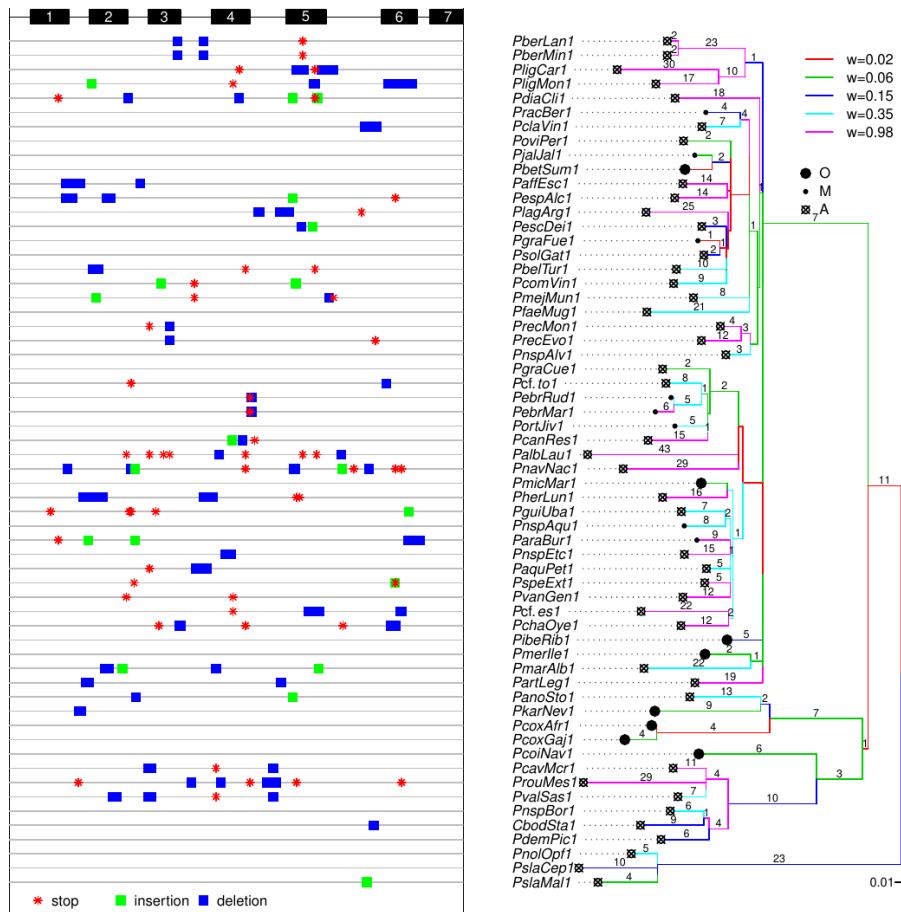


Fig. 17. – Représentation d'un alignement d'opsine₂ et des traces de non-fonctionnalisation pour le genre *Proasellus* (à gauche). Les boîtes noires représentent les domaines transmembranaires de l'opsine qui sont les parties les plus conservées. Les symboles rouges, verts et bleus représentent des mutations non-sens et sont indicateurs de non-fonctionnalisation. A droite, on retrouve une reconstruction de l'histoire de l'opsine₂. La couleur des branches indique le d_N/d_S correspondant tel que défini par mapping des substitutions puis clustering des branches (Dutheil et al. 2012). On passe donc de branches extrêmement contraintes ($d_N/d_S = 0.02$; 98% des mutations non-synonymes sont rejetées) à des branches évoluant sous un régime quasi neutre ($d_N/d_S = 0.98$). Les branches avec des valeurs intermédiaires indiquent probablement un changement de régime sur la branche en question, i.e. la colonisation du milieu souterrain. Les symboles en bout de branche indiquent le phénotype de la structure oculaire (O, M, A), voir la Fig. 16 pour leur signification.

quence elle est retrouvée chez plus de taxons anophtalmes. Les deux copies ont probablement des fonctions différentes et une relation à la colonisation du milieu souterrain qui est clade-dépendante⁹. Pour maximiser le nombre de datations, j'ai finalement choisi d'utiliser l'opsine₁ comme marqueur de datation. L'utilisation de l'opsine₂ aurait donné un jeu de données plus petit, mais des datations probablement plus anciennes. In fine, bien que la datation absolue soit un exercice très périlleux, je pense que dans un cadre comparatif les valeurs relatives de ces datations nous permettent de différencier des « vieilles » et des « jeunes » espèces souterraines. Appliqué aux 22 espèces du jeu de données transcriptomiques (voir la section 1.1.4.1), on sait maintenant que les temps de colonisation au sein de nos 11 paires d'espèces sont très hétérogènes (Fig. 19) : nos répliqués de transition écologique sont plutôt des pseudo-répliqués le long d'un gradient de temps de colonisation.

⁹L'origine de ces différences reste encore aujourd'hui inconnue

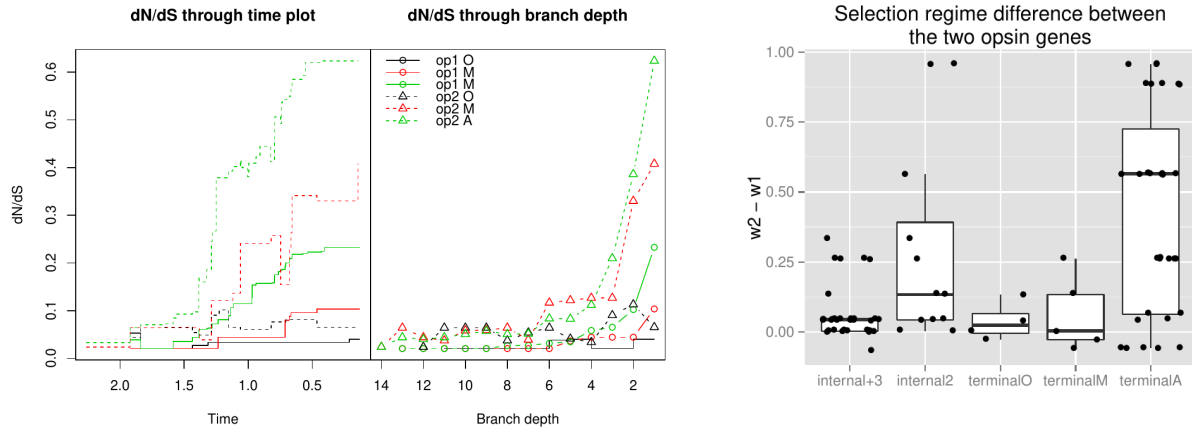


Fig. 18. – Moyenne des d_N/d_S dans le temps (en centaine de millions d’années) ou en remontant dans la profondeur de l’arbre (branche terminale = 1, racine de l’arbre = 14), en partant de trois types de feuilles (phénotype du système oculaire : O, M, A) et pour les arbres des opsine₁ et opsine₂ (panels de gauche). A droite sont comparés les d_N/d_S des opsine₁ et opsine₂ ($\omega_2 - \omega_1$) pour les branches que les 2 arbres de gènes possèdent en commun. En plus d’observer que les d_N/d_S des branches menant à des taxons anophtalmes sont bien plus élevés, on remarque également que l’opsine₂ commence à se non-fonctionnaliser plus tôt (vers 13 millions d’année ou sur des branches de 3^{ème} ordre, contre 11 millions d’années ou des branches de 2^{ème} ordre).

1.1.4. Produire les données

Choisir son modèle génomique à quelques inconvénients : il faut tout faire soi-même, du terrain à la bioinformatique en passant par la biologie moléculaire. Sous-traiter une partie du travail s’avère difficile. Les centres de séquençage sont souvent accoutumés aux mêmes types d’échantillons et ratent parfois des runs entiers quand on leur fournit des extractions exotiques... Depuis 2012, grâce au travail acharné de Lara Konecny puis avec l’aide de Clément-

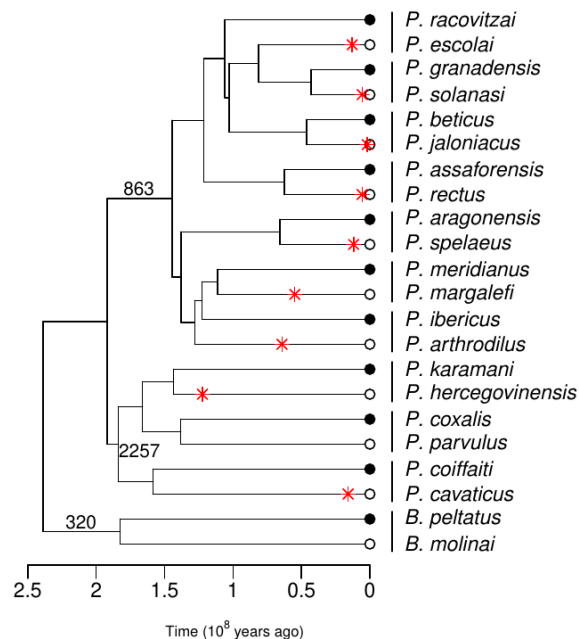


Fig. 19. – Estimation de la date de colonisation (étoile rouge) à partir de l’approche de d_N/d_S appliquée aux opsine1. L’estimation n’est pas possible pour les espèces du genre *Bragasellus* et pour l’espèce *P. parvulus* qui a perdu l’opsine1.

tine François, nous avons développé deux approches de production de données génomiques : l'approche RNA-seq (2011-2020) puis l'approche long-read Nanopore (2020-2024).

L'approche RNA-seq.

A partir de 2011, nous nous lançons dans l'acquisition de données génomiques pour les Asellidae (aucune donnée n'étant disponible à ce moment-là), mais en choisissant une stratégie de réduction génomique « naturelle » : le RNA-seq. En séquençant les ADN complémentaires des ARNm, on réduit grandement la difficulté de séquençage d'un génome entier : l'effort de séquençage à produire est bien plus faible et l'étape d'assemblage est moins complexe. En plus de permettre l'assemblage d'un grand nombre de séquences codantes, cette méthode permet d'estimer l'expression de ces séquences sans données supplémentaires. Une dizaine de couples d'espèces souterraines/surface sont sélectionnés, et des missions d'échantillonnages avec congélation sur le terrain sont organisées un peu partout en Europe. Les bibliothèques Illumina sont fabriquées au laboratoire, d'abord à partir de pool d'organismes, puis sur un seul organisme. Un premier pipeline d'assemblage et de définition des familles de gènes est construit et servira dans de multiples analyses (Fig. 20).

Les approches de transcriptomiques *de novo* ont néanmoins un défaut majeur : les gènes faiblement exprimés peuvent ne pas être assemblés car séquencés à trop faible profondeur¹⁰. Ces gènes réellement présents disparaissent néanmoins du jeu de données. On ne sait alors pas si un gène absent d'un jeu de données transcriptomique est absent du génome, non-exprimé ou exprimé mais non-assemblé. Quand on étudie la composition des répertoires de gènes ou bien l'expression des gènes, cela pose un réel problème. Dans le contexte des paires d'espèces, le problème peut être en partie résolu en alignant les lectures d'une espèce sur le transcriptome de l'autre pour potentiellement découvrir des gènes artificiellement perdus (exprimés mais non-assemblés). Il est même possible d'estimer leur expression en prenant en compte la divergence moyenne entre espèces : en effet aligner les lectures d'un taxon sur un autre va conduire à ne pas aligner certaines lectures qui ont accumulé trop de substitutions. La plupart des logiciels d'alignement de lectures utilisent un seuil fixe de différence, seuil qui

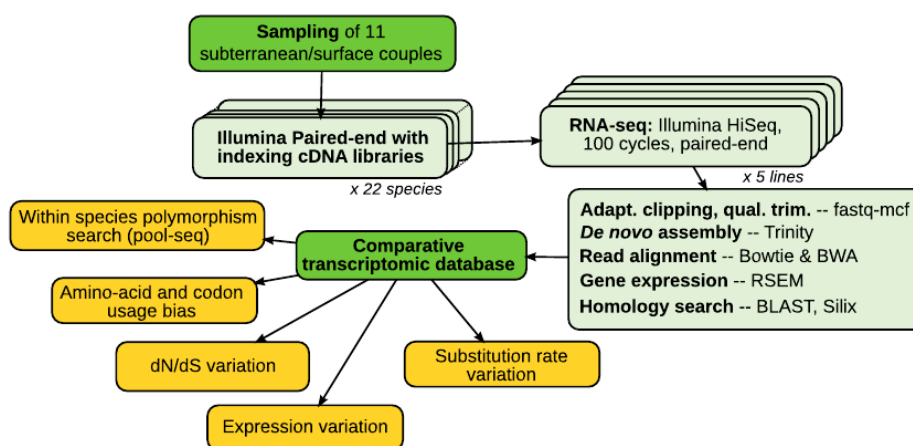


Fig. 20. – Pipeline de production de données RNA-seq, d'assemblage de transcriptomes, puis de définition de familles de gènes.

¹⁰en RNA-seq la profondeur de séquençage est directement dépendante du niveau d'expression et de l'effort de séquençage

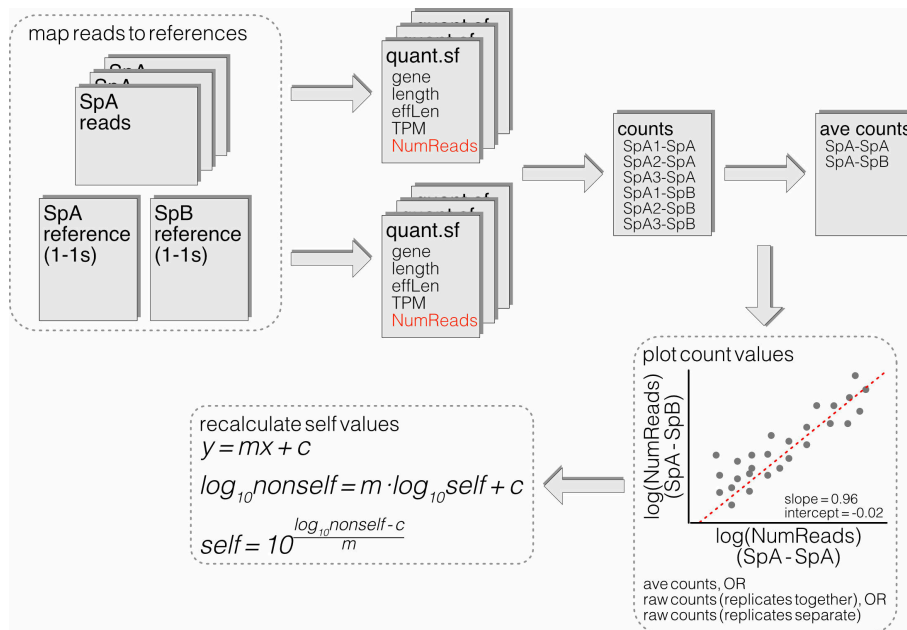


Fig. 21. – Méthode d'estimation de l'expression de gènes non-assemblés. Pour l'espèce A (SpA), nous disposons de lectures RNA-seq pour trois échantillons (1, 2, 3) et d'un transcriptome de référence, mais également d'un transcriptome de référence pour une espèce sœur B (SpB). Les lectures de l'espèce A sont alignées sur le transcriptome de l'espèce A et B, le moyenne est faite entre échantillons, puis les log des comptages sur l'espèce A et B sont comparés à l'aide d'un modèle linéaire sur l'ensemble des gènes simple-copie partagés entre les 2 espèces. Les paramètres de ce modèle sont ensuite utilisés pour estimer l'expression des gènes de l'espèce A (self) qui sont absents de l'assemblage de l'espèce A mais présent dans celui de l'espèce B (non-self). Grice et al in prep.

aura d'autant plus de chance d'être atteint que les deux espèces sont divergentes. On attend alors une relation simple entre le nombre moyen de lectures non-alignées sur une espèce sœur et la divergence entre l'espèce servant de référence et celle d'où sont issues les lectures. Pour chaque paire de taxons, cette approche a été utilisée pour les études d'expression de gènes et nous permet de construire des matrices d'expression de gènes comparatives avec moins de « faux zéro ». La méthode fonctionne en estimant d'abord la relation entre l'expression des gènes mesurée sur le transcriptome d'une espèce cible et sur celui d'une espèce sœur sur l'ensemble des gènes communs à ces deux espèces, puis cette relation est utilisée pour estimer l'expression des gènes non-assemblés dans l'espèce cible (Fig. 21).

Cette approche a été testée sur un jeu de données à 11 paires d'espèces en utilisant les vraies séquences codantes, et donc les vraies divergences entre taxons, mais en effaçant les variations d'expression de gènes entre espèces d'une même paire. Des expériences de RNA-seq et de comptage ont alors été simulées pour chaque espèce (avec 4 individus par espèce, le tout répliqué 100 fois). Enfin j'ai artificiellement rajouté des « faux zéro » aux jeux de données en les distribuant aléatoirement et à des fréquences variables (de 1% à 40% de « faux zéro »), puis j'ai analysé le jeu de donnée pour voir si la présence de faux zéro créait des gènes différentiellement exprimés et si la méthode présentée ci-dessus permettait de réduire le nombre de faux positifs (gènes estimés comme différentiellement exprimés alors que ce n'est pas le cas). Comme attendu la présence de « faux zéro » génère beaucoup de faux positifs d'expression différentielle, mais l'alignement sur l'espèce sœur puis la correction des comptages pour prendre en compte l'effet de divergence diminue très fortement ce taux de faux positifs. La

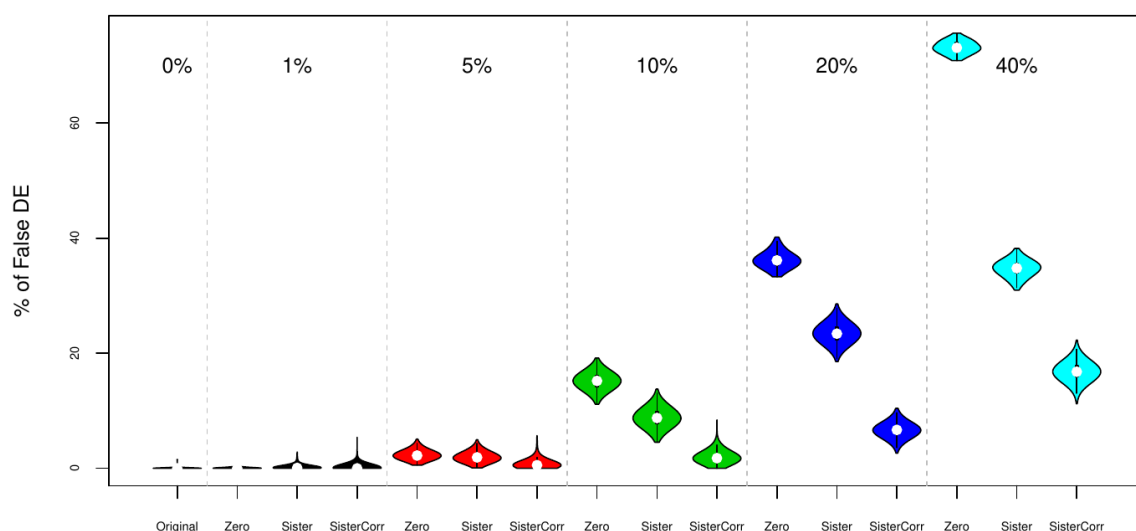


Fig. 22. – Simulation d'un jeu de données RNA-seq à 11 paires d'espèces sans expression différentielle mais en rajoutant différents taux de « faux zéro » (0, 1, 5, 10, 20 et 40%). Chaque simulation comprend 3 catégories : inclusion de faux zéro puis alignement sur le transcriptome de la même espèce (« zéro »), comptages issus de l'alignement sur l'espèce soeur sans (« Sister ») et avec correction pour la divergence (« SisterCorr »). Pour chaque catégorie une recherche de gènes différentiellement exprimés entre les espèces de surface et souterraines est effectuée avec DESeq2 et tous les résultats significatifs sont comptés comme des faux positifs. Grice et al in prep.

méthode semble suffisamment efficace pour être utilisée en routine jusqu'à 10% de « faux zéro », ce qui correspond à notre attendu dans les données de RNA-seq d'Asellidae.

L'approche long-read.

A partir de 2019, nous changeons de stratégie de séquençage et nous nous lançons dans le séquençage longues lectures Nanopore et PacBio. La mise au point est longue et pénible, mais à partir de 2023, nous commençons à obtenir des runs sur séquenceur MinION et PromethION de qualité qui permettent une production soutenue de génomes d'Asellidae. Un premier pipeline d'assemblage a été testé et optimisé (Fig. 23), puis une version simplifiée est aujourd'hui utilisée en routine. Des tests de couvertures croissantes nous ont permis de déterminer qu'une profondeur de séquençage minimale de 14X de données Nanopore (équivalente à 2 flowcells Minion de bonne qualité) permet d'obtenir un assemblage de qualité correcte (Fig. 24). La technologie Nanopore est en constante évolution, que ce soit les séquenceurs, la chimie des bibliothèques, les flowcells et les basecallers¹¹. Nous venons depuis quelques mois de passer sur des flowcells Promethion (entre 50Gb et 100Gb brutes par flowcell) et à un basecaller (dorado) permettant d'obtenir des lectures de qualité bien supérieure ($Q \geq 30$) à ce que nous obtenions il y a seulement 1 an (Fig. 25).

Les progrès dans l'acquisition de données génomiques sont importants mais il reste encore beaucoup de défis:

- nous sommes encore très loin d'obtenir des assemblages chromosomiques (télomère à télomère et diploïde). Pour atteindre cet objectif il faudrait augmenter très fortement la profondeur de séquençage et diversifier les types de données (Hi-C, Ultra-long reads, données trio ...). Il n'est pas possible pour un laboratoire d'écologie d'envisager des développements sur toutes ces technologies, et il n'est financièrement pas possible de sous-

¹¹logiciel permettant de transformer le signal électrique d'un séquençage Nanopore en séquence ADN.

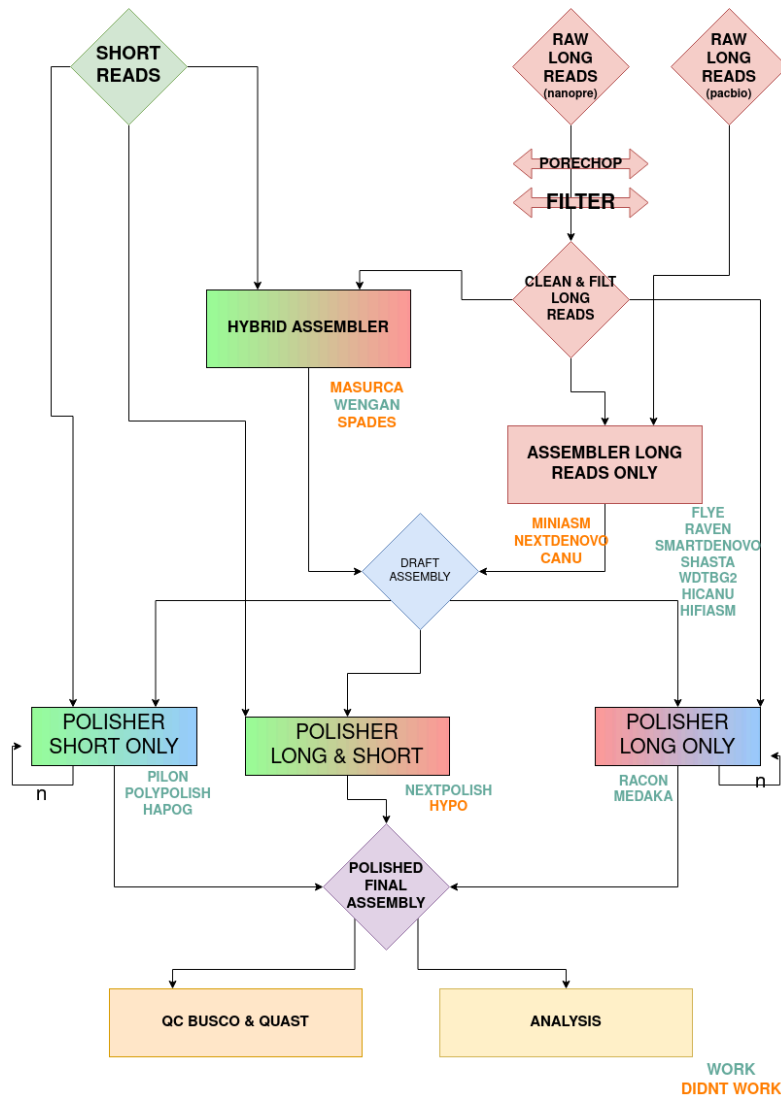


Fig. 23. – Pipeline développé et multiples solutions testées pour les données Nanopore et PacBio d’Asellidae. In fine, notre pipeline de routine est beaucoup plus simple : production de lectures Nanopore seules, enlèvement des adaptateurs (porechop) et des lectures de faible qualité, assemblage (flye), puis contrôle qualité (BUSCO et QUASt). Les étapes de polishing ne sont pas apparues nécessaires. Tiré de Debaecker et al, in prep.

traiter l’assemblage de beaucoup de génomes d’organismes non-modèle (quand bien même on trouverait un prestataire en capacité de le faire). Pour le moment, nous avons opté pour la stratégie d’obtenir des génomes pour une vingtaine d’espèces de qualité suffisante pour la plupart des recherches en génomique évolutive (chapitre suivant).

- la technologie Nanopore est très instable : entre le début du projet la préparation des librairies, la qualité et la quantité des données, les formats et les logiciels de basecalling ont fortement évolués. En conséquence, même en ayant comme objectif de produire un jeu de données homogène produit par un pipeline bioinformatique défini une seule fois, ce n’est déjà plus vraiment le cas... Il est difficile de ne pas tomber dans une perpétuelle mise à jour des procédures qui parfois apporte de vrais gains qualitatifs mais souvent participe juste à un effet de mode.
- il faut maintenant stocker un volume de données brutes énorme (un bon run de PromethION produit 1 To de données brutes) ce qui pose des problèmes informatiques non négligeables...

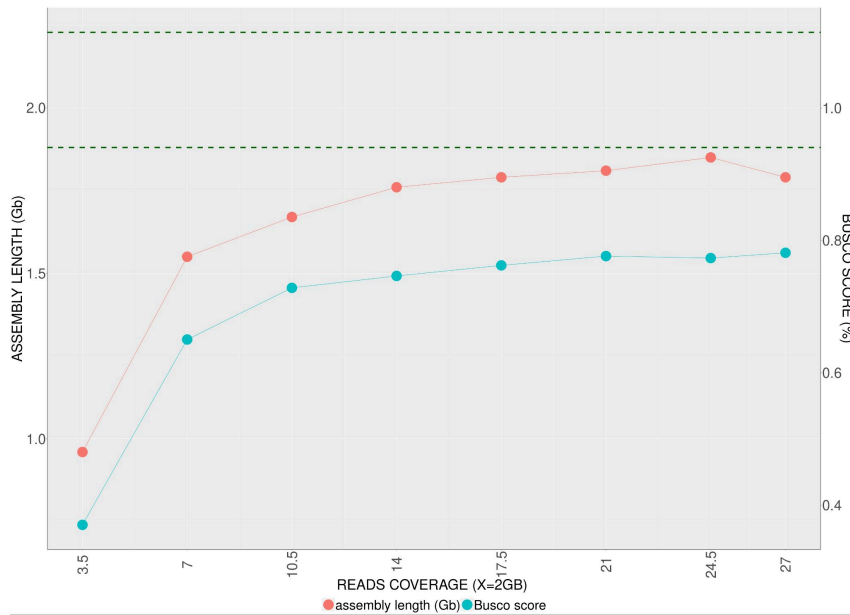


Fig. 24. – Relation entre la couverture en données Nanopore et la qualité de l’assemblage (taille de l’assemblage, en rouge) et scores BUSCO (% de gènes reconstruits complètement, en vert). La taille attendue du génome par mesure cytométrique est entre les deux lignes en pointillés. Tiré de Debaecker et al, in prep.

- les défis sont maintenant principalement bioinformatiques : il faut annoter des éléments répétés totalement inconnus et annoter les génomes avec peu de connaissances transférables depuis des génomes d’organismes modèles.

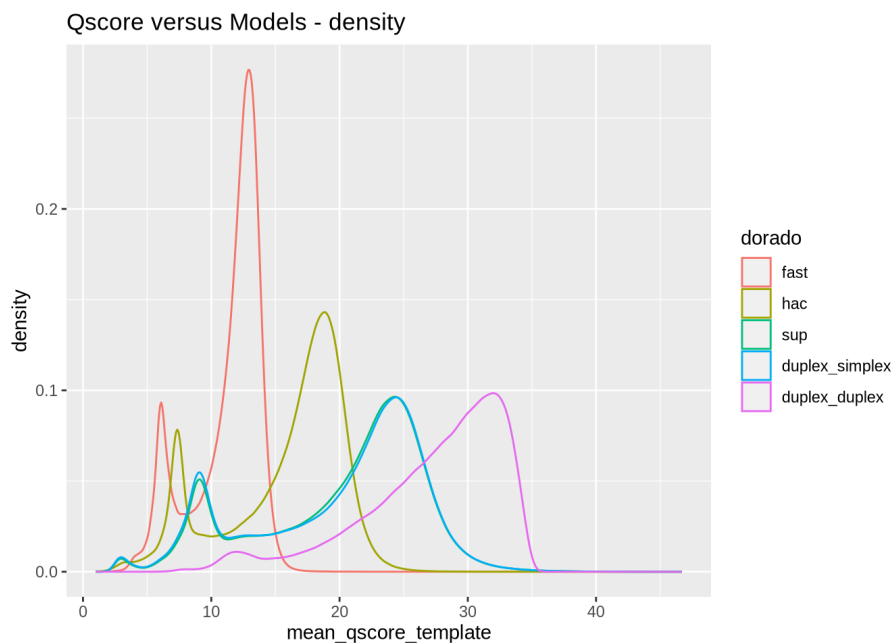


Fig. 25. – Influence du mode (fast, hac, sup et duplex) de basecalling sur la qualité moyenne (score PRHED) des lectures d’un run Promethion. L’espèce est *P. ibericus* pour un run de 65Gb.

1.2. Evolution des tailles des génomes

La caractéristique la plus simple qui définit un génome - sa taille - reste une énigme : pourquoi cette caractéristique du génome varie-t-elle autant d'un taxon à l'autre et quelles sont les forces à l'origine de cette énorme variabilité ? Deux points de vue opposés ont d'abord émergé : les tailles du génome ont une forte influence sur la fitness des organismes et évoluent dans un sens ou l'autre sous l'effet de la sélection naturelle (hypothèse adaptationniste), ou bien elles sont non-adaptatives et résultent de l'interaction entre mutation et dérive (voir par exemple la revue récente de Blommaert (2020)).

L'hypothèse adaptationniste s'appuie sur d'innombrables preuves de covariation entre la taille du génome et des traits phénotypiques tels que la taille du corps, le métabolisme de base, le temps de génération ou le taux de croissance. Sur la base de ces nombreuses corrélations, une hypothèse générale a été proposée, appelée l'hypothèse du « bulk » (Gregory 2005), qui intègre la taille du génome comme un trait central en interaction avec de nombreux traits phénotypiques. La taille du génome serait une sorte de super-trait à la frontière entre génotype et phénotype et qui contrôlerait le taux de croissance cellulaire et le métabolisme, qui ont eux-mêmes un impact sur d'autres traits dans un réseau d'interactions complexes. La taille du génome, en plus de l'information génétique qu'elle porte, est donc censée jouer un rôle central dans la définition de la physiologie et de l'écologie d'un organisme et est donc étroitement contrôlée par la sélection naturelle. Il est intéressant de noter qu'avec la découverte que la taille du génome dépend principalement de la quantité d'ADN répété présent, cette hypothèse est restée valable dans le sens où, bien que la principale force de mutation générant rapidement des variations de la taille du génome soit le taux d'insertion et de suppression d'éléments répétés, c'est en fin de compte la sélection naturelle qui sélectionnera les organismes ayant des génomes plus petits ou plus grands en fonction de leurs fitness respectives. Alors que les variations sont principalement générées par l'activité des éléments transposables et que la direction de l'évolution est contrôlée par la sélection naturelle, la taille du génome joue un rôle central entre ces deux forces.

Parallèlement, la découverte du rôle important joué par les éléments transposables et de leur impact délétère fréquent sur le fonctionnement d'un génome, a conduit à l'émergence d'une hypothèse totalement différente où la sélection positive ne joue aucun rôle. Une première hypothèse neutraliste postule que la taille du génome n'a que peu ou pas d'impact sur la fitness, et que la taille du génome est principalement sous le contrôle du taux d'insertion et de délétion (Petrov 2002). Combiné à des mécanismes tels que la recombinaison ectopique ou l'inactivation d'éléments transposables par l'hôte, ce modèle peut générer des trajectoires évolutives complexes de la taille du génome (Bourgeois & Boissinot 2019). Pour tenir compte des nombreuses covariations documentées entre la taille du génome et les traits phénotypiques, une autre hypothèse non adaptative a été proposée dans les années 2000, nommée la « mutational-hazard hypothesis » (MHH) (Lynch 2007a; 2007b). Cette hypothèse intègre la taille effective de la population (N_e) comme un paramètre central qui module l'intensité de la dérive génétique et génère *in fine* une covariation non causale entre la taille du génome et de nombreux traits phénotypiques. Comme vu dans le contexte du d_N/d_S (voir la section 1.1.3.3.2), l'efficacité de la sélection dépend de la dérive génétique, en particulier pour les mutations

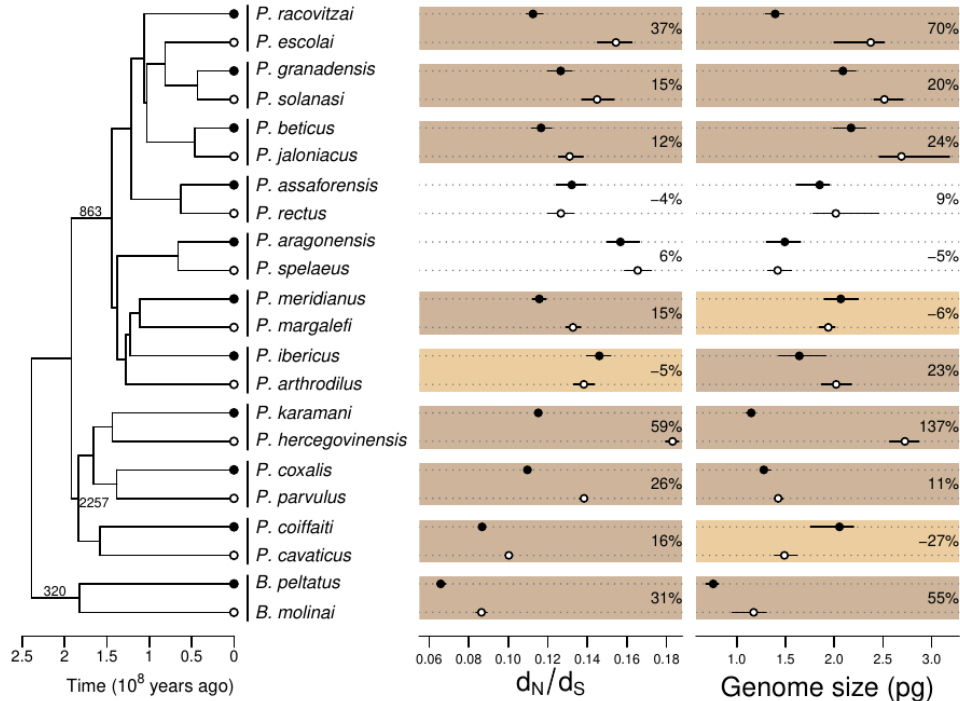


Fig. 26. – Evolution de l’efficacité de la sélection, approximée par le d_N/d_S , et la taille du génome pour 11 paires d’espèces de surface ou souterraines. La figure reprend les codes couleurs de la Fig. 13. Tiré de Lefébure et al. (2017).

légèrement délétères et légèrement avantageuses qui peuvent devenir effectivement neutres dans les petites populations (Ohta 1992). La MHH postule que de nombreuses caractéristiques génomiques sont légèrement délétères mais sont plus susceptibles d’être fixées dans de petites populations où elles deviennent effectivement neutres. Cela inclut les insertions d’éléments transposables qui devraient donc être plus fréquentes dans les petites populations. À long terme, les populations de petite taille devraient donc évoluer vers des génomes plus grands. Étant donné que la taille de la population est également corrélée à de nombreux traits tels que le temps de génération ou la taille corporelle, la longévité ... la MHH suggère que la corrélation observée entre la taille du génome et de nombreux traits phénotypiques n’est pas causale et résulte de l’effet confondant de N_e . Bien qu’elle soit ancrée dans la théorie de la génétique des populations, la MHH demeure difficile à tester et manque de soutien empirique statistiquement robuste.

L’une des caractéristiques les plus importantes de la plupart des habitats souterrains est la faible disponibilité des ressources trophiques (Venarsky et al. 2014). Lorsque l’on considère deux taxons phylogénétiquement proches, l’un vivant en surface et l’autre dans le milieu souterrain, ce dernier est donc plus susceptible d’avoir une densité de population plus faible et un N_e plus petit. Il est probable que l’on trouve de nombreux contre-exemples, mais si l’on étudie de nombreuses paires de taxons, on peut émettre l’hypothèse qu’en moyenne, les taxons souterrains auront une N_e plus faible. Si cette hypothèse est correcte et si de nombreuses paires d’espèces de surface/souterraines étroitement apparentées peuvent être examinées, nous disposons alors d’un modèle original et puissant pour tester l’influence de N_e sur l’évolution de la taille du génome. A partir des 11 paires d’espèces surface/souterrain d’Asellidae pour lesquelles nous avons des évidences de réduction de N_e et d’efficacité de

la sélection au travers du d_N/d_S (Fig. 13), nous avons estimé la taille de leurs génomes par cytométrie et avons constaté qu'en moyenne, les asellidés souterrains ont des génomes plus grands (25 % d'augmentation en moyenne, PGLS valeur de $p=0.014$, Fig. 26, Lefébure et al. (2017)). Le plus grand changement de taille de génome a été observé entre l'espèce de surface *Proasellus karamani* (1,1 Gb) et l'espèce souterraine *Proasellus hercegovinensis* (2,7 Gb). Alors que la taille du corps et le taux de croissance ne sont pas corrélés avec la taille du génome, le d_N/d_S est lui significativement corrélé avec la taille du génome (PGLS, valeur de $p=0.009$). Une augmentation analogue de la taille du génome a également été observée dans la même étude chez deux paires de décapodes de surface/souterrains et une paire de gastéropodes, ce qui suggère qu'il pourrait s'agir d'une tendance générale chez les taxons souterrains.

L'estimation des temps de colonisation conforte globalement ces résultats : les « vieilles » espèces souterraines ont tendance à avoir des génomes plus gros (PGLS valeur de $p=0.019$), le plus grand écart de taille de génomes au sein d'une paire étant observé chez la paire la plus « ancienne » (*P. karamani*, *P. hercegovinensis*).

La MHH sous-entend un processus d'accumulation lente et graduelle d'événements faiblement délétères. Le séquençage à faible profondeur des génomes nous a permis de reconstruire les

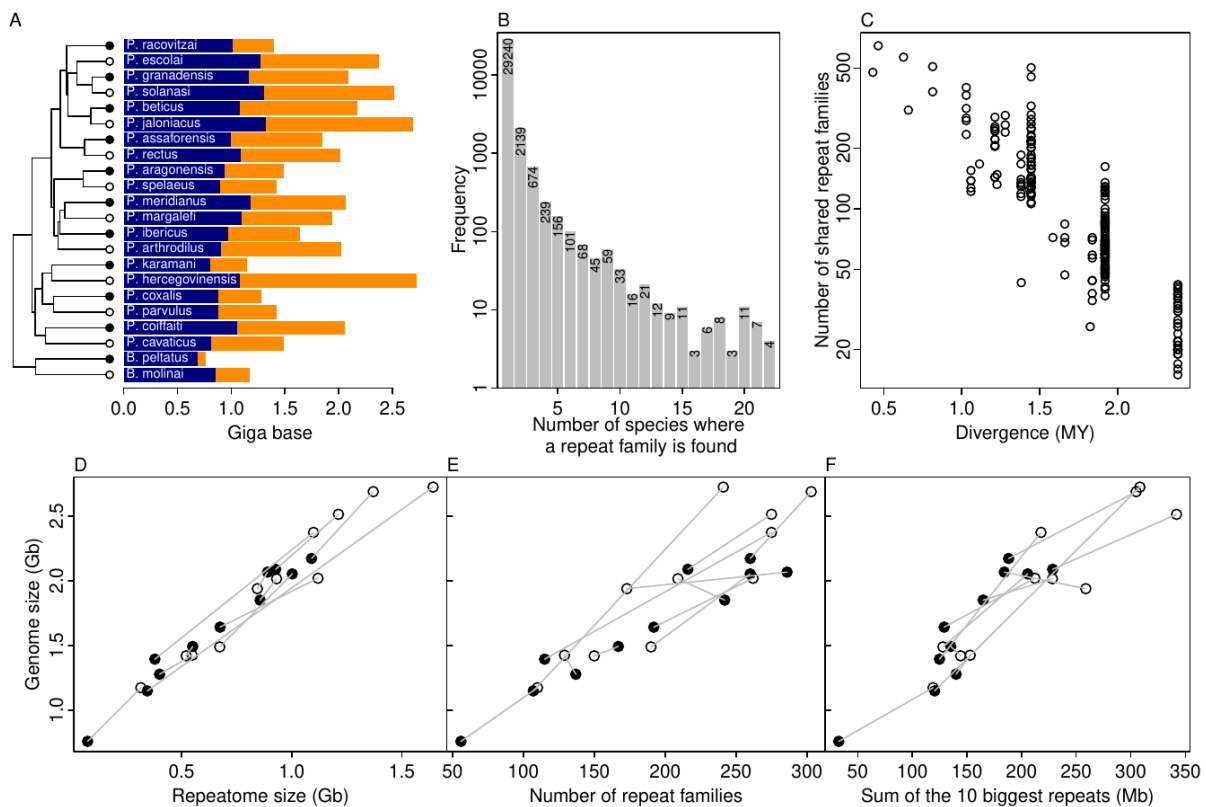


Fig. 27. – Evolution du contenu en éléments répétés chez 11 paires d'Asellidae de surface et souterrains. A : taille du génome et sa fraction répétée (en jaune) et non-répétée (bleue). B : spectre de fréquence des familles de gènes d'éléments répétés. C : relation entre nombre d'éléments répétés partagés et divergence. Relation entre la taille du génome et la taille du répétome (D), le nombre de familles de gènes (E) et la taille cumulée des 10 familles d'éléments répétés les plus fréquentes. Les observations issues d'une même paire d'espèces sont reliées par une droite, avec les espèces souterraines en noire et de surface en blanc. C'est la pente moyenne de ces droites qui décrit le mieux la relation entre 2 variables dans un cadre phylogénétique structuré en paires. Tiré de Lefébure et al. (2017).

éléments répétés des génomes de 11 paires d'espèces¹². Les variations de tailles de génomes sont presque entièrement expliquées ($r^2 = 0.96$) par la teneur en éléments répétés des génomes (Fig. 27). Peu de familles d'éléments sont partagées entre espèces, suggérant un important turn-over alimenté par de fréquentes invasions et pertes d'éléments transposables. Les plus gros génomes ont plus de familles d'éléments mais aussi des familles d'éléments plus invasives. Tous ces éléments sont en accord avec les attendus de la MHH, les Asellidés souterrains apportent le premier soutien empirique solide à cette hypothèse.

1.3. Contraintes en nutriments et composition des protéines

En écologie fonctionnelle, l'étude des flux de nutriments et le transfert de ces nutriments dans les chaînes trophiques sous l'angle des ratios molaires s'appelle la stœchiométrie écologique (Sturner & Elser 2017). La disponibilité de certains éléments et leurs abondances relatives (par exemple le ratio C/N) structurent fortement les communautés (principalement C, N, P et S). Les organismes développent des mécanismes pour assurer leur homéostasie (un maintien de la composition élémentaire des consommateurs), mais il est aussi observé que certains organismes modifient leur composition. Un des exemples les plus parlants est la composition élémentaire des protéines dont la fonction est l'acquisition dudit élément : une protéine permettant l'acquisition d'un élément sera pauvre en cet élément afin de pouvoir assurer sa production en cas de carence. Les premières évidences de ce type de contrainte sur la composition de protéines ont été observées chez les protéines associées à l'assimilation du soufre qui sont globalement plus pauvres en soufre que l'ensemble du protéome chez la levure et *E. coli* (Fig. 28, Baudouin-Cornu et al. (2001)). D'un point de vue évolutif, on s'attend donc à ce que la disponibilité en éléments influe sur le devenir de mutations impactant l'usage de ces éléments.

Inspiré par ces travaux, un nouveau cadre théorique a été développé sur l'influence de la disponibilité des nutriments sur la composition élémentaire du génome, du transcriptome et du protéome : la stœchiogénomique (Bragg & Wagner 2009; Elser et al. 2011). Les 20 acides aminés et les quatre nucléotides diffèrent par le nombre de nutriments qu'ils contiennent (voir par exemple la Fig. 29 pour l'usage en azote des différents nucléotides et acides aminés). Ainsi, des changements dans la composition des ARN ou des protéines peuvent modifier les besoins en nutriments et impacter la fitness d'un organisme confronté à des limitations environnementales. Cette hypothèse s'applique conceptuellement à tout nutriment, mais elle est particulièrement pertinente pour l'azote (N), car il représente une grande proportion des macromolécules biologiques (Sturner & Elser 2017). Ainsi, une faible disponibilité en azote pourrait théoriquement sélectionner des changements dans la composition des protéines ou

¹²En séquençant un génome à faible profondeur, par exemple à 0.01X, les parties non-répétées du génome ne seront pas assemblées, alors que les régions fortement répétées, par exemple à plus de 1000 copies, pourront être assemblées, décrites, et quantifiées.

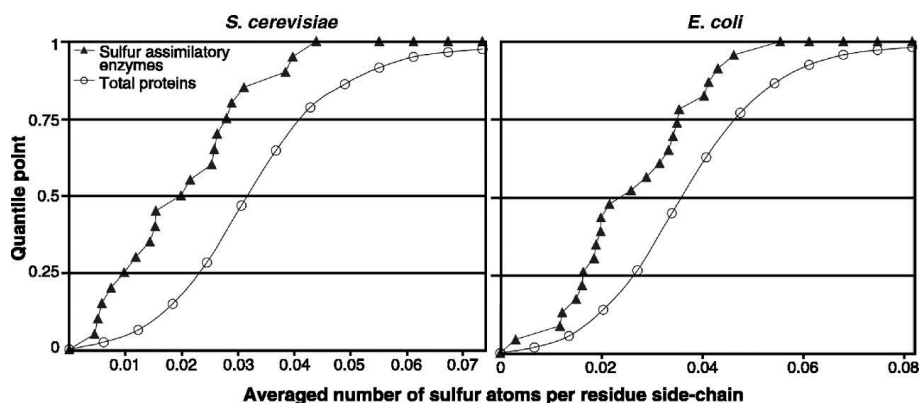


Fig. 28. – Distribution des quantiles de l'utilisation du soufre chez *S. cerevisiae* et *E. coli*. Pour les deux espèces, le nombre moyen d'atomes de soufre trouvés dans les chaînes latérales des résidus pour chaque protéine a été calculé pour l'ensemble du protéome et pour le set de gènes dont la fonction est associée à l'assimilation du soufre. Tiré de Baudouin-Cornu et al. (2001).

de l'ARN par l'utilisation préférentielle d'acides aminés ou de nucléotides pauvres en azote. De tels mécanismes d'économie d'azote ont été observés de manière fiable chez des micro-organismes marins évoluant dans des environnements pauvres en azote (Luo et al. 2015) mais les évidences chez des eucaryotes multicellulaires sont rares et discutables (Acquisti et al. 2009; Günther et al. 2013).

Comme les eaux souterraines ont généralement une disponibilité extrêmement faible en nutriments, y compris en N (Venarsky et al. 2014; Francois et al. 2016), les organismes souterrains pourraient avoir sélectionné des mutations économes en N depuis leur transition vers les eaux souterraines. Nous avons testé cette hypothèse en analysant les transcriptomes (et les protéomes correspondants) de 13 paires d'espèces de surface et souterraines (Francois et al. 2016). L'usage en N par le protéome et le transcriptome pondéré par l'expression des gènes n'est pas différent entre espèces souterraines et de surface et n'est pas lié à disponibilité en N dans l'environnement. De même le nombre moyen d'atomes N par chaîne latérale d'acide-aminé (NARSC) n'augmente pas chez les espèces souterraines (Fig. 30 a). Les gènes les plus

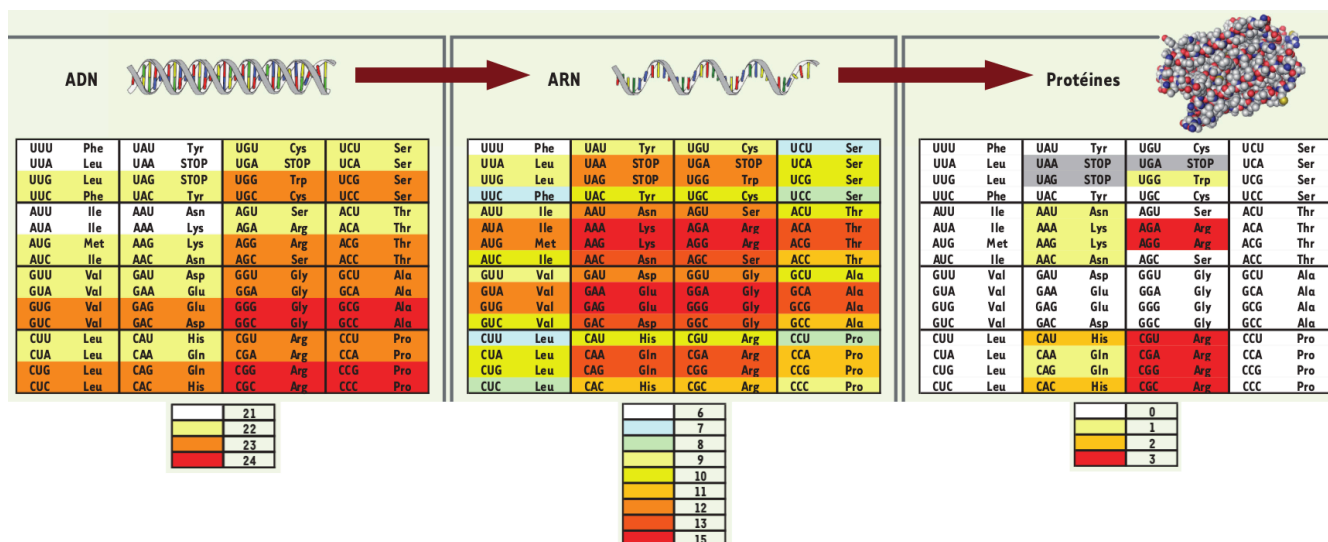


Fig. 29. – Usage en azote des différentes bases et acides aminés composant l'ADN, l'ARN et les protéines. Le code couleur représente le nombre d'atomes d'azote par unité. Tiré de Baudouin-Cornu (2008).

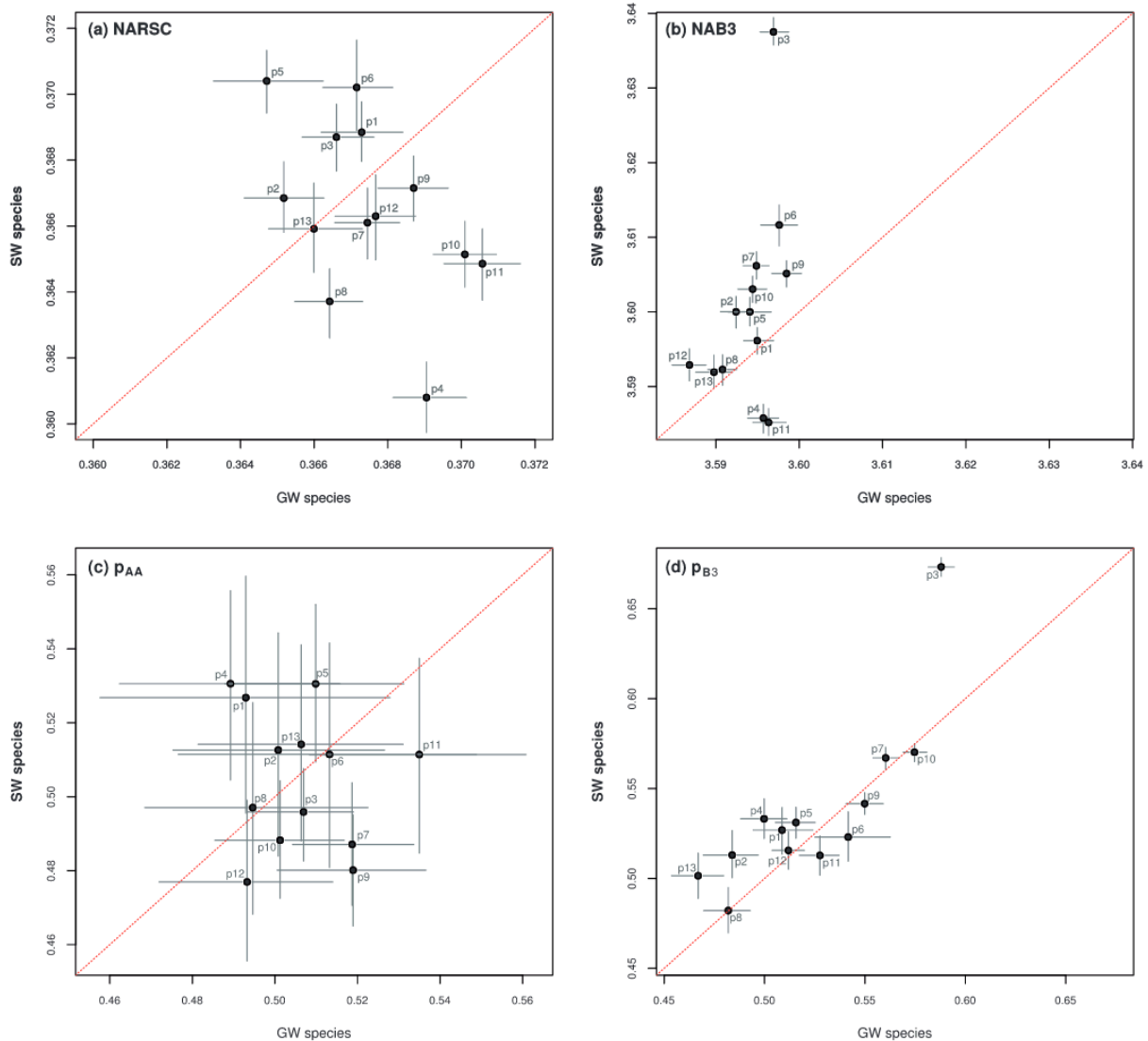


Fig. 30. – Comparaison de différents proxy de l’usage en azote entre paires d’espèces de surface et souterraines au sein du protéome et du transcriptome. Chaque point correspond à une paire d’espèces surface/souterrain. (a) NARSC : nombre moyen d’atomes N par chaîne latérale de résidu. (b) NAB3 : nombre moyen d’atomes N par base en troisième position du codon. (c) pAA : proportion de substitutions coûteuses en N dans les protéines. (d) pB3 : proportion de substitutions coûteuses en N en troisième position de codon. Les axes x et y représentent les valeurs pour les espèces de surface et souterraines, respectivement. Sous l’hypothèse d’une sélection pour l’économie en azote plus forte chez les organismes souterrains, les points devraient donc se situer au-dessus de la ligne bissectrice (ligne pointillée). Les intervalles de confiance sont issus de bootstrap (95% de la distribution).
Tiré de Francois et al. (2016).

exprimés montrent une baisse du NARSC mais indépendamment de l’environnement des espèces considérées. Néanmoins, le nombre moyen d’atomes N par base en troisième position des codons (NAB3) diminue chez les espèces souterraines (Fig. 30 b).

Au lieu de regarder des budgets globaux en azote, nous avons conduit des analyses restreintes sur des familles de gènes orthologues en calculant la proportion de substitutions qui augmente le coût en N pour les lignées souterraines et de surface sur les protéines (pAA) et la troisième position des codons des ARNm (pB3). La tendance à accumuler des mutations chères en N pour le protéome (pAA) n’est pas différente entre espèces souterraines et de surface, même en limitant l’analyse aux gènes les plus fortement exprimés (Fig. 30 c), mais elle est de nouveau

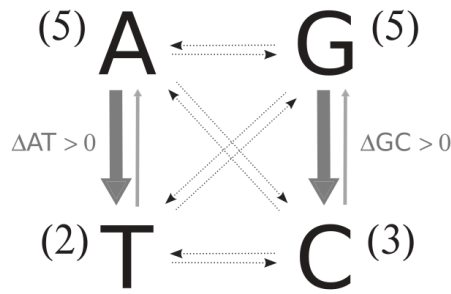


Fig. 31. – Asymétries des taux de substitution dans le cadre de la théorie stœchiogénomique d'économies en azote. Les substitutions qui nous intéressent (c'est-à-dire celles qui sont sensibles à la sélection stœchiogénomique mais insensibles à la gBGC) sont représentées par des flèches en traits pleins. Les nombres entre parenthèses indiquent le coût N de chaque nucléotide (en nombre d'atomes N). La largeur des flèches est proportionnelle au taux de substitution attendu dans l'hypothèse d'économies stœchiogénomiques en N. Les asymétries sont calculées ainsi :

$$\Delta AT = q_{A \rightarrow T} - q_{T \rightarrow A}$$

$$\Delta GC = q_{G \rightarrow C} - q_{C \rightarrow G}$$

où $q_{X \rightarrow Y}$ est le taux de substitution de X à Y. Tiré de Francois et al. (2016).

significative pour les troisièmes positions des codons (Fig. 30 d). Cette observation est compatible avec une économie d'azote et une utilisation plus faible de l'azote en troisième position de codons chez les espèces souterraines. Cependant, les espèces souterraines avec des faibles pAA présentent aussi des faibles taux de GC en troisième position (GC_3) : l'accumulation de mutations moins coûteuses en N pourrait n'être que le sous-produit de mécanismes non-adaptatifs favorisant la fixation de mutations AT chez les espèces souterraines (e.g. biais mutationnel ou gBGC¹³).

Pour tester explicitement la présence de sélection « stœchiogénomique », nous avons développé un test basé sur l'asymétrie de fixation (Fig. 31). En effet la sélection pour des économies en N devrait laisser une empreinte de substitution asymétrique à l'échelle du transcriptome. Entre deux substitutions complémentaires, la sélection stœchiogénomique devrait favoriser la substitution réduisant le coût en N par base sur le brin codant (par exemple, $A \rightarrow T$ par rapport à $T \rightarrow A$, et $G \rightarrow C$ par rapport à $C \rightarrow G$), générant des asymétries de brin dans les taux de substitution (par exemple, $q_{A \rightarrow T} > q_{T \rightarrow A}$). Un processus non-adaptatif agira de la même façon sur les 2 brins (et donc $q_{A \rightarrow T} = q_{T \rightarrow A}$). Nous avons estimé le niveau d'asymétrie (ΔAT et ΔGC) en troisièmes positions de codons pour ces deux classes de transitions pour les 26 espèces en ignorant les autres types de substitutions qui sont particulièrement sensibles aux processus non-adaptatifs comme la gBGC et donc au GC_3 des espèces (on parle de mutation GC-conservative). Si la sélection opère, on s'attend à ce que ces deux mesures augmentent chez les espèces limitées en azote. Bien que toutes les espèces montrent des patrons de substitution asymétriques¹⁴, ces patrons ne sont pas différents entre espèces souterraines et de surface.

¹³La gBGC pour « GC-biased gene conversion » est le biais de conversion génique vers GC. C'est une tendance à la fixation des allèles G ou C induite par la machinerie de recombinaison. Ce processus ne crée pas de mutation mais augmente la fréquence des certaines mutations de façon directionnelle. Comme la sélection positive, son efficacité dépend de N_e . Elle peut impacter fortement l'évolution des génomes, comme dans la lignée humaine, ou avoir un rôle plus ténu, comme chez la drosophile (Duret & Galtier 2009).

¹⁴Malgré ce qui a été dit plus tôt, l'asymétrie de brin dans des séquences codantes peut être générée par un processus neutre. Par exemple, un biais mutationnel brin-spécifique peut être généré par le processus même

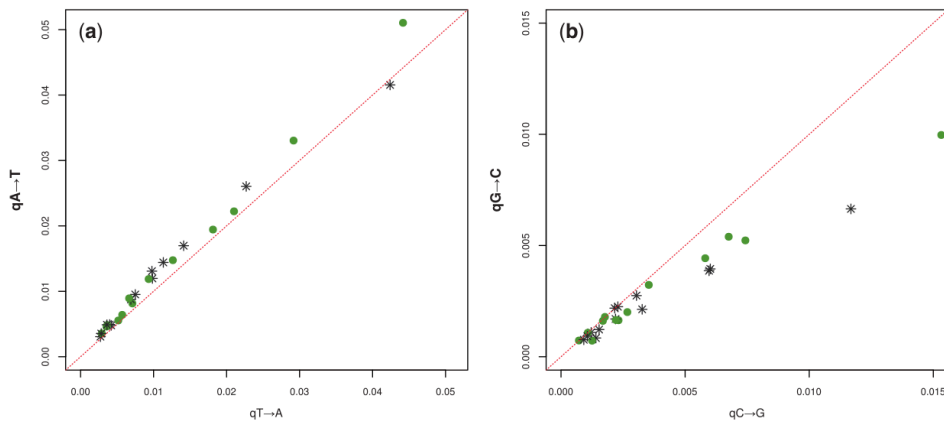


Fig. 32. – Graphiques de parité entre taux de substitutions complémentaires. (a) Taux de substitution de A vers T ($q_{A\rightarrow T}$) et de T vers A ($q_{T\rightarrow A}$). (b) Taux de substitution de G à C ($q_{G\rightarrow C}$) et de C à G ($q_{C\rightarrow G}$). Les espèces souterraines et les espèces de surface, au total 26 espèces, sont représentées par des astérisques et des cercles, respectivement. L'axe des y, indiqué en gras, correspond à la substitution qui devrait se produire plus fréquemment que sa complémentaire (axe des x) selon la prédiction stœchiogénomique de l'asymétrie des brins. La sélection stœchiogénomique devrait amener les espèces au-dessus de la ligne bissectrice (ligne pointillée), cette déviation étant plus forte chez les espèces limitées en azote. Bien que des biais soient observés, ils ne sont pas corrélés à l'environnement des organismes. Tiré de Francois et al. (2016).

Les différences d'usage en N aux troisième position des codons ne peuvent donc pas être expliquées par de la sélection « stœchiogénomique » mais sont plus cohérentes avec l'action de processus influant aussi sur le GC₃ comme la gBGC. De par la diminution de leur N_e , l'efficacité de la gBGC pourrait baisser chez les espèces souterraines et le GC₃ baisser légèrement. Une analyse ciblant le processus de gBGC est nécessaire pour valider cette hypothèse.

Deux hypothèses pourraient expliquer pourquoi les mutations permettant d'économiser des atomes d'azote n'ont pas été sélectionnées dans le transcriptome et le protéome des isopodes souterrains évoluant depuis des milliers ou des millions d'années dans des conditions de disponibilité extrêmement faible en azote. Premièrement, même lorsqu'elle se produit dans un transcrit ou une protéine fortement exprimée, une mutation ponctuelle d'économie de N pourrait avoir un effet adaptatif trop faible pour être efficacement sélectionnée. Cela est particulièrement vrai pour les petites populations, telles que les isopodes souterrains, caractérisées par une faible efficacité de la sélection naturelle (voir la discussion sur la barrière de la dérive, Fig. 12). Deuxièmement, d'autres adaptations permettant d'économiser les nutriments peuvent avoir été sélectionnées plus rapidement et plus efficacement après la colonisation des eaux souterraines. Par exemple, les isopodes souterrains ont des taux de croissance plus faibles que leurs homologues de surface (Lefébure et al. 2017). Une première réponse adaptative aux travers de modification de traits d'histoire de vie pourrait avoir suffisamment atténué les contraintes stœchiométriques pour que la sélection « stœchiogénomique » soit non-opérante (cette fois l'optimisation serait contrainte par s et pas seulement par N_e). Globalement ces résultats remettent en cause l'applicabilité des hypothèses « stœchiogénomiques » aux animaux voire à l'ensemble des organismes pluricellulaires eucaryotes.

de transcription : lors de la transcription le brin codant est moins protégé et est plus sujet à la désamination ce qui conduit à un déséquilibre entre les transitions complémentaires C→T and G→A induisant $q_{C\rightarrow T} > q_{G\rightarrow A}$ (Mugal et al. 2010). Dans notre cas ce processus n'est pas en cause car nous n'avons pas travaillé sur ces mutations, mais d'autres processus neutres sont possibles.

1.4. Taux de mutation, traits d'histoire de vie et environnement mutagène

Le taux de mutation varie fortement entre taxons (Fig. 33) : de 10^{-11} mutations par site par génération chez certains organismes unicellulaires à environ 10^{-8} chez les primates (Michael Lynch et al. 2023; Wang & Obbard 2023; Bergeron et al. 2023). L'origine de ces variations alimente de multiples débats autour du rôle des traits d'histoire de vie, de N_e et des conditions environnementales. Le modèle Asellidae permet d'étudier ces différentes facettes : (i) lors de la transition souterraine les traits d'histoire de vie sont fortement impactés ainsi que N_e , (ii) le milieu souterrain présente des caractéristiques (absence d'UV, faible dispersion...) qui permettent de tester plus facilement l'impact de facteurs mutagènes environnementaux.

L'estimation des taux de mutations peut se faire au travers de l'étude du taux d'évolution moléculaire ou taux de substitution (K). C'est la vitesse à laquelle les séquences d'ADN accumulent des substitutions¹⁵. Ce taux dépend du taux de mutation (μ) et de la probabilité qu'une mutation se propage dans la population (appelée probabilité de fixation p). Ainsi, le taux de substitution est le taux de mutation multiplié par la probabilité de fixation ($K = \mu \times p$). Pour les mutations neutres, le taux de substitution est égal au taux de mutation car la probabilité de fixation par dérive ($\frac{1}{2N_e}$) compense la probabilité d'apparition de nouvelles mutations ($2N_e\mu$) : $K = 2N_e\mu \times \frac{1}{2N_e} = \mu$ (Kimura 1983). Une façon relativement simple d'estimer le taux de mutation consiste donc à collecter un vaste ensemble de données comparatives et à estimer le taux de substitution pour une classe de substitutions considérées comme n'ayant pas d'impact sur la fitness. Les substitutions synonymes dans les gènes codant pour les protéines constituent une classe de substitutions couramment utilisée¹⁶. Alternativement, on peut aujourd'hui estimer plus directement le taux de mutation à partir de pedigrees ou bien à partir d'expériences d'accumulation de mutations dans des lignées maintenues expérimentalement en laboratoire. Ces dernières méthodes ont l'avantage de fournir une estimation directe du taux de mutation par génération, alors que les approches phylogénétiques fournissent un taux de mutation par branche ou par unité de temps si on dispose de points de calibration ou d'un chronogramme. La quantité de données nécessaires est néanmoins bien plus simple à obtenir, surtout pour de nombreuses espèces non-modèles. C'est donc l'approche que nous avons adoptée. Par la suite, on parlera donc de vitesse d'évolution moléculaire ou de taux de mutation par unité de temps (et non pas par génération) mesuré à partir de taux de substitutions synonymes.

¹⁵Une substitution est une mutation qui a « réussi », elle est présente chez tous les individus de l'entité considérée sans autre polymorphisme à ce site. Le type de mutation n'a pas d'importance (SNP, indel, ...).

¹⁶On verra cependant au Chapitre 1.6.2 que pour les gènes les plus fortement exprimés, certains taxons présentent de forts biais d'usage des codons synonymes. Dans ce contexte précis et pour certaines positions, les substitutions synonymes ne sont pas neutres et sont sous le contrôle de la sélection traductionnelle. On considère néanmoins que cet impact est limité, en partie compensé par les méthodes de calcul des taux de substitutions synonymes (par exemple par l'estimation des fréquences de codons par une matrice F34), ou bien que ce biais est commun à tous les taxons d'un jeu de données et aura donc peu d'impact dans une approche comparative.

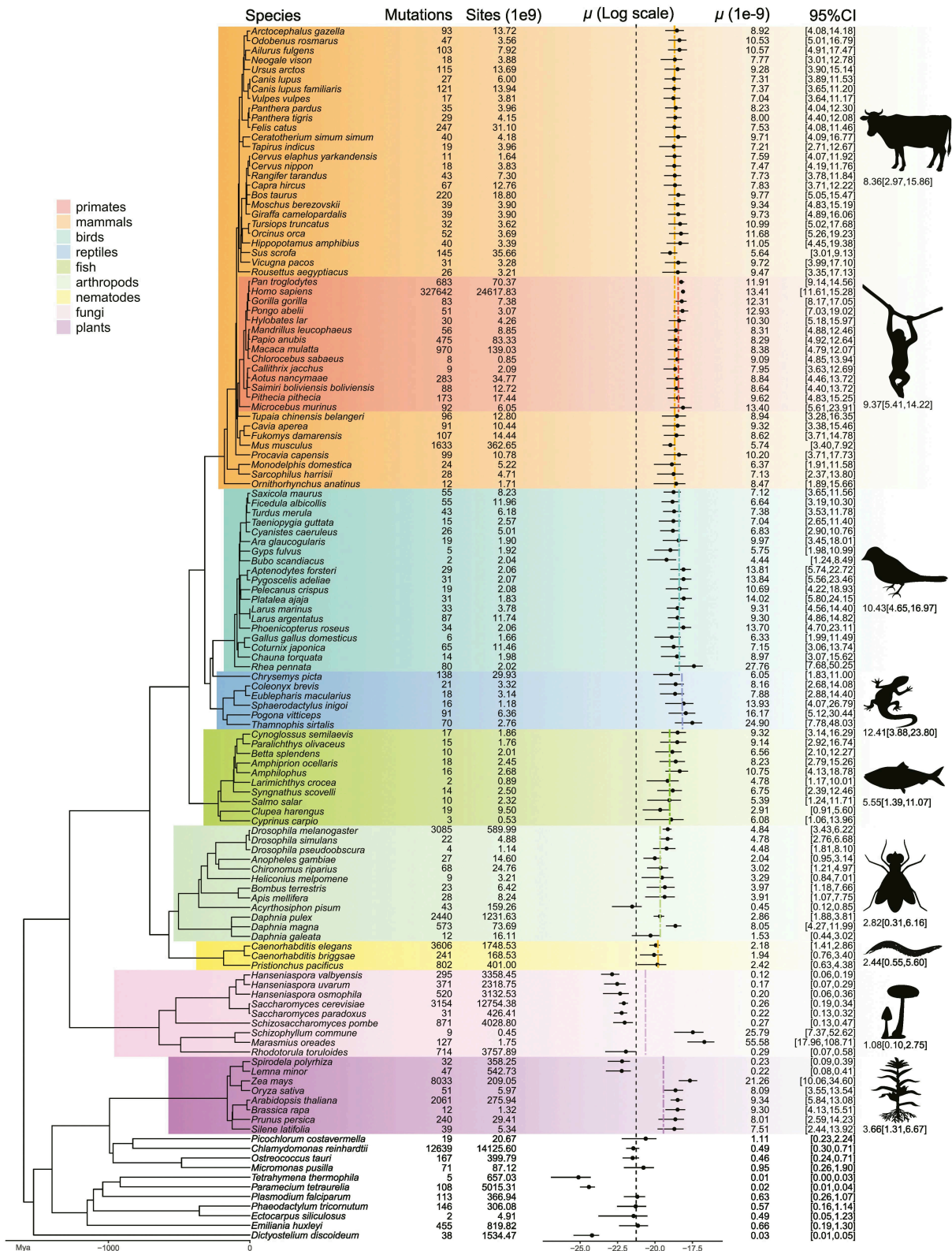


Fig. 33. – Reconstruction des taux de mutations ponctuelles pour les eucaryotes à partir des données de la littérature. La ligne verticale en pointillés indique le taux de mutation de l'ancêtre commun de l'arbre. Tiré de Wang & Obbard (2023).

1.4.1. Taux de mutation et traits biologiques

De nombreux traits d'histoire de vie (ou plus généralement de traits biologiques) sont corrélés et semblent influencer le taux d'évolution moléculaire. Historiquement les plus cités sont :

- *Temps de génération.* Chaque fois que l'ADN est dupliqué pendant la production de gamètes, des erreurs de réplication se produisent, entraînant de nouvelles mutations. L'hypothèse du temps de génération (Li & Tanimura 1987) propose simplement que chez les espèces dont le temps de génération est plus long, le génome est copié moins souvent, ce qui entraîne moins de mutations par unité de temps. Il a aussi été observé des corrélations entre le taux de mutation par génération et le temps de génération chez les vertébrés (ainsi que l'âge à la maturité et la fécondité (Bergeron et al. 2023)) mais aussi plus généralement chez les eucaryotes (Wang & Obbard 2023) : les espèces avec des temps de génération longs transmettent plus de mutations à leur descendants. On a donc un effet en apparence contradictoire : le temps de génération est négativement corrélé au taux de mutation par unité de temps, mais positivement corrélé au taux de mutation par génération.
- *Investissement parental.* L'investissement parental pourrait avoir un impact sur l'équilibre entre la sélection et la dérive sur le taux de mutation (Britten 1986) : une espèce produisant peu de descendants devrait transmettre moins de mutations à sa progéniture afin de limiter le transfert de mutations délétères qui auraient un impact significatif sur sa fitness.
- *Métabolisme.* Le métabolisme de l'oxygène génère des espèces réactives de l'oxygène (ROS) qui sont mutagènes. L'hypothèse du taux métabolique (Martin et al. 1992) propose que les espèces ayant un taux métabolique plus élevé génèrent plus de ROS et ont par conséquent un taux de mutation plus élevé. Cependant, ces corrélations sont régulièrement remises en cause (Lanfear et al. 2007) et la quantité de ROS dépend également de l'efficacité de la chaîne respiratoire mitochondriale et leur impact mutationnel dépend aussi de l'efficacité des mécanismes de protection (Galtier et al. 2009).
- *Longévité.* Il a été suggéré que le lien entre les ROS et le taux de substitution mitochondrial ne dépend pas du taux métabolique mais du trade-off entre le coût des mécanismes évitant les mutations et le coût des mutations délétères (Nabholz et al. 2008). Le point d'équilibre entre ces deux coûts dépendrait de la longévité de l'organisme. Comme les ROS peuvent accélérer le vieillissement, la sélection naturelle conduirait à une diminution de la production de ROS ou à une amélioration de la protection des cellules contre les ROS chez les espèces longévives.
- N_e . Comme indiqué plus haut, la relation N_e et taux de mutation est la première évidence empirique supportant la « drift-barrier hypothesis » (Lynch et al. 2016). Cette relation est très évidente à grande échelle évolutive en intégrant tous les domaines du vivants, mais perd en intensité sur des échelles phylogénétiques plus petites (Michael Lynch et al. 2023; Wang & Obbard 2023; Bergeron et al. 2023).

Bien que la littérature sur le lien entre le taux de substitution et les traits d'histoire de vie soit très abondante, il n'y a pas de consensus sur l'influence réelle de tel ou tel traits sur les taux de mutation, à l'exception de l'impact quasi-mécanique du temps de génération. La question est difficile car la plupart des traits co-varient. Néanmoins, la majorité des tests empiriques effectués se concentrent sur les mammifères, souvent sur des gènes mitochondriaux en petit nombre¹⁷.

¹⁷Depuis 2020, plusieurs études génomiques ou sur de très nombreux marqueurs nucléaires ont vu le jour et bien au-delà des mammifères.

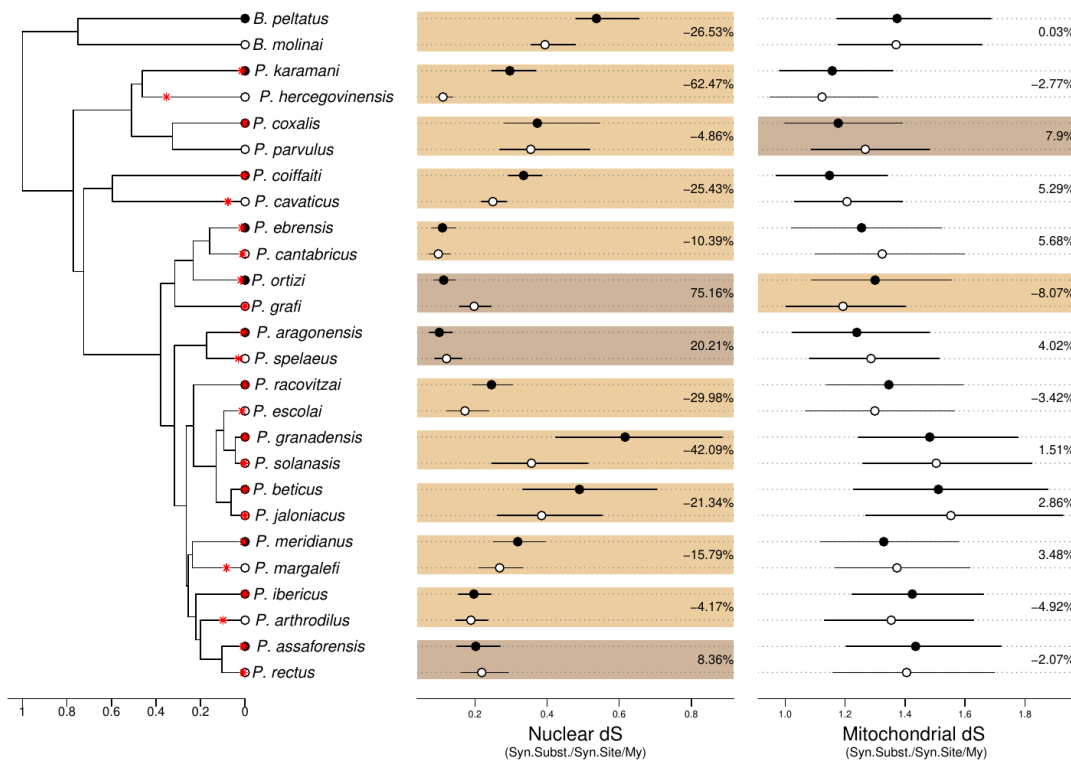


Fig. 34. – Taux de substitutions synonymes nucléaires (à gauche) et mitochondriaux (à droite). La figure reprend les codes des figures Fig. 19 et Fig. 26. Tiré de Saclier et al. (2018).

En calculant les taux de substitution nucléaires et mitochondriaux sur un grand nombre de gènes (382 et 13, respectivement) pour 26 espèces d'Asellidae formant 13 paires d'espèces souterraine/surface, nous avons montré que les Asellidae souterrains ont un taux d'évolution moléculaire nucléaire ralenti comme attendu sous l'action simple du temps de génération (Fig. 34, Saclier et al. (2018)). La relation observée va donc aussi à l'inverse de l'effet attendu de N_e . Cependant, le taux d'évolution moléculaire mitochondrial n'est pas affecté par la colonisation des eaux souterraines et les importantes modifications du temps de génération et de longévité qui lui sont associées. Il s'agit d'un résultat totalement inattendu qui remet en question notre vision de l'impact des traits d'histoire de vie sur le taux de mutation des mitochondries. Chez les arthropodes les taux de mutation mitochondrial et nucléaire de la lignée germinale pourraient avoir des déterminismes totalement différents.

1.4.2. Taux de mutation et environnement mutagène

Le taux de mutation peut également être influencé par des agents mutagènes environnementaux, en particulier les radiations ionisantes naturelles comme le rayonnement ultraviolet (UV) et le rayonnement tellurique. Leur effet à forte dose sur la lignée somatique est établi (Dubrova et al. 1996), mais leurs effets multi-générationnels à faible dose sont peu connus. L'étude des radiations naturelles sur le taux de mutation est particulièrement difficile : les études expérimentales à long-terme sont impraticables sur des organismes multicellulaires et les études comparatives sont difficiles à concevoir car il est difficile de trouver des lignées ayant

évoluées sous des conditions stables de rayonnements contrastés pendant de très longues périodes de temps. Grâce à leur capacité de dispersion limitée (Eme et al. 2018), les espèces souterraines offrent une excellente occasion d'étudier l'impact de l'environnement ionisant sur le taux de mutation. En effet, une population souterraine est susceptible d'avoir été exposée au même rayonnement pendant de nombreuses générations, ce qui est très peu probable pour une espèce de surface. En outre, contrairement aux espèces de surface, les espèces souterraines ne sont pas exposées au rayonnement UV, ce qui permet (i) d'étudier l'impact du rayonnement UV sur le taux de mutation en comparant les espèces de surface et souterraines, et (ii) d'écarter le rayonnement UV comme facteur de confusion pour étudier l'impact de la radioactivité naturelle tellurique sur le taux de mutation.

Le rayonnement UV laisse une signature mutationnelle particulière : il provoque une photo-excitation des dimères de pyrimidines (C et T), qui tordent les molécules d'ADN. La base endommagée est remplacée par deux thymines (Brash 2015), ce qui entraîne une mutation vers un dinucléotide TpT. En recherchant des variations dans la fréquence des dinucléotides de pyrimidines, aucune preuve d'un impact différentiel des UV sur le taux de mutation entre les Asellidae de surface et souterrains n'a été identifiée (Saclier et al. 2018). Ce résultat suggère que les espèces de surface disposent de mécanismes de protection pour éviter les mutations dues au rayonnement UV dans les cellules germinales.

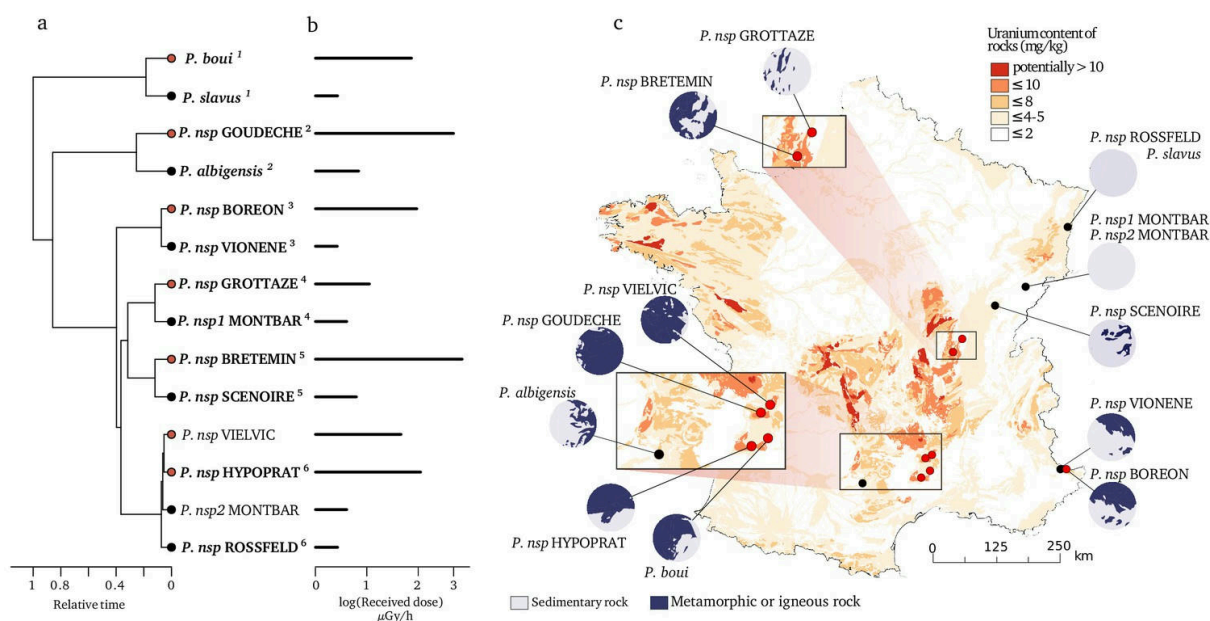


Fig. 35. – Espèces et localités sélectionnées pour étudier l'impact de la radioactivité naturelle sur le taux et le spectre de mutations. 14 espèces avec des expositions contrastées à la radioactivité naturelle ont été sélectionnées (points noirs : faible exposition, points rouges : forte exposition). A partir de leurs relations phylogénétiques (a), nous avons ensuite sélectionné 6 paires d'espèces monophylétiques pour comparer leur spectre mutationnel (Vielvic et Montbar ont été exclus en raison de leur phylogénie non résolue, les paires d'espèces sont indiquées avec des numéros). La dose effective de radioactivité a été mesurée à partir des sédiments échantillonnés dans chaque station. Pour chaque site l'affleurement de roches sédimentaires faibles en radioactivité et de roches magmatiques ou métamorphiques riches en radioactivité dans un rayon de 15km autour du point d'échantillonnage ($\lambda 15$) est représenté avec un cercle sur le côté de la carte. Alors que les mesures effectuées sur les sédiments issus des sites échantillonnés sont représentatifs des expositions reçues par les individus échantillonnés lors de l'étude, le $\lambda 15$ se veut un proxy plus intégratif et robuste à la variabilité spatiale (variation du rayonnement dans le paysage) et temporelle (dispersion des organismes). Tiré de Saclier et al. (2020).

Différentes formations géologiques sont caractérisées par différents niveaux de rayonnement radioactif. Par exemple, les formations sédimentaires comme les roches calcaires émettent moins de radiations que les roches magmatiques et métamorphiques. En étudiant des espèces souterraines endémiques étroitement apparentées vivant dans des environnements géologiques contrastés, il est alors possible de tester l'influence de la radioactivité de la roche mère sur le taux de mutation mais aussi le spectre mutationnel. A partir de 6 paires d'espèces souterraines vivant dans des environnements contrastés en radioactivité naturelle (Fig. 35), nous avons observé que les taux de mutation mitochondrial et nucléaire augmentent respectivement de 60% et 30% dans les environnements plus mutagènes (Saclier et al. 2020). Les résultats sont robustes à la façon de mesurer la radioactivité : mesure du rayonnement α sur les sédiments, estimation de la dose reçue par les organismes, ou bien la proportion de roches métamorphiques ou magmatiques dans un rayon de 15 km (λ_{15}), ce dernier facteur étant le meilleur prédicteur du taux de mutation ($r^2_{nucléaire} = 0.476$, $r^2_{mitochondrial} = 0.566$, Fig. 36).

A partir de données de polymorphisme, nous avons également constaté que la radioactivité de la roche-mère modifie non seulement le taux de mutation mais augmente également la probabilité des mutations G vers T (Fig. 37). Cette mutation est une caractéristique du stress oxydatif

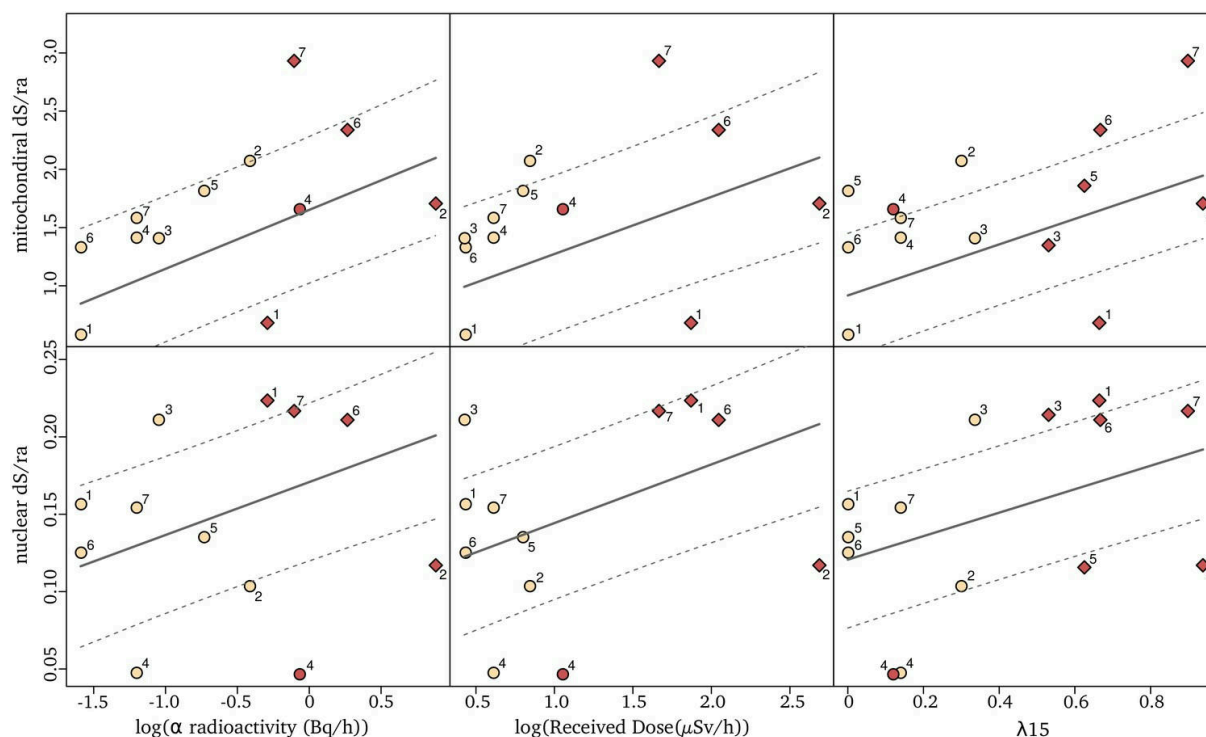


Fig. 36. – Relations entre le taux de substitution synonyme par unité de temps arbitraire (dS/ra, ra: root age) et la radioactivité mesurée soit comme la radioactivité α du sédiment (à gauche), la dose reçue (au milieu) ou la proportion de roches magmatiques et métamorphiques dans un rayon de 15 km autour des sites d'échantillonnage, λ_{15} (à droite). Chaque point représente une espèce avec le dS/ra nucléaire en haut et mitochondrial en bas. L'ajustement d'un modèle pGLS est indiqué par une ligne continue (et son intervalle de confiance par des lignes pointillées). Les deux espèces d'une paire sont étiquetées avec un numéro comme dans la Fig. 35 (le numéro sept correspond aux deux espèces ne formant pas une paire), avec en rouge l'espèce de la paire avec la radioactivité α la plus élevée. Les espèces issues de formations sédimentaires ($\lambda_{15} < 50\%$) sont représentées par des cercles et les espèces issues des formations magmatiques/métamorphiques ($\lambda_{15} > 50\%$) par un losange. Tiré de Saclier et al. (2020).

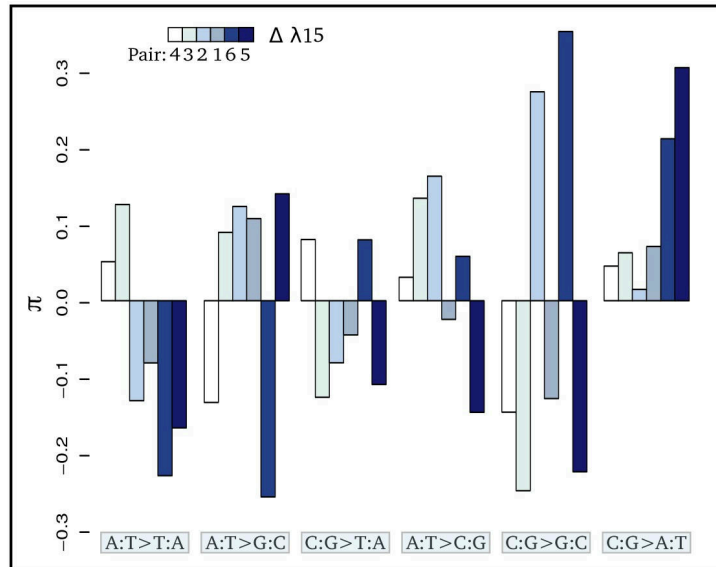


Fig. 37. – Contrastes (π) de la proportion relative de chaque type de mutation $p_{i:j \rightarrow k:l}$ dans chaque paires d'espèces : $\pi_m = \log\left(\frac{p_{i:j \rightarrow k:l}^+}{p_{i:j \rightarrow k:l}^-}\right)$ où + et - correspondent respectivement à l'espèce exposée à la plus forte et à la plus faible radioactivité dans la paire m . Les barres positives représentent donc une plus forte proportion d'une mutation donnée chez l'espèce la plus fortement exposée dans la paire alors qu'une barre négative représente une plus forte proportion de la mutation chez l'espèce la plus faiblement exposée de la paire. Les barres sont ordonnées de la paire ayant la plus faible différence de proportion de roches métamorphiques ou magmatiques dans un rayon de 15km (λ_{15}) à celle ayant la plus forte différence de λ_{15} ($\Delta_{\lambda_{15}}$). Les mutations sont ordonnées de celle étant la plus négativement corrélée à λ_{15} à celle étant la plus positivement corrélée. Les numéros sous l'échelle de couleurs indiquent le numéro de la paire d'espèces tel qu'indiqué dans la Fig. 35. Tiré de Saclier et al. (2020).

(Shibutani et al. 1991), ce qui suggère que le rayonnement tellurique naturelle influence le taux de mutation en générant un excès de radicaux libres, en particulier dans les mitochondries.

1.5. Convergence phénotypique et expression des gènes

Les phénotypes convergents sont très répandus dans la nature, mais nous n'avons qu'une compréhension fragmentaire des mécanismes généraux qui sous-tendent leur émergence. Parmi les exemples emblématiques, on peut citer la pigmentation, le métabolisme C4 des plantes, le vol, le nanisme insulaire, la bioluminescence, les résistances aux insecticides, l'écholocation ... ainsi que les nombreux traits associés aux environnements extrêmes. Une question importante qui découle de ces exemples est de savoir quels changements dans les génomes sous-tendent la convergence des phénotypes : les changements génomiques sont-ils eux-mêmes convergents ou est-ce que de multiples voies génomiques peuvent conduire aux mêmes phénotypes ? Si les changements génomiques sont eux-mêmes convergents, cela signifie qu'il n'y a que peu de chemins évolutifs disponibles pour atteindre un phénotype donné. En revanche, si les modifications génomiques sous-tendant les phénotypes convergents sont idio-

synchratiques, on peut conclure que le contexte génomique n'apporte que peu de contraintes à l'évolution des phénotypes. L'étude de l'évolution génomique convergente permet donc de comprendre en général à quel point l'évolution peut être prévisible (Gompel & Prud'homme 2009). L'étude des phénomènes de convergence phénotypique et génomique permet ainsi de mieux comprendre les contraintes évolutives imposées par la structure des génomes et la relation entre génotype et phénotype. Depuis 10 ans, beaucoup de travaux s'attaquent à cette question (Sackton & Clark 2019; Bolnick et al. 2018). Nous l'avons également abordée aux travers des nombreuses convergences observées chez les organismes souterrains, qu'elles soient régressives ou non (Chapitre 1.1.3.2). Les premiers travaux que nous avons amorcés s'intéressent à la convergence des niveaux d'expression de gènes (l'analyse de la convergence des séquences protéiques est en projet). L'expression d'un gène est déjà constitutive d'un phénotype, on n'étudie donc pas directement des phénomènes de convergence génomique, mais comment des convergences phénotypiques comme la régression oculaire ou la perte de pigmentation sont accompagnées par des phénomènes de convergence dans l'expression des gènes.

Pour cela nous avons d'abord étudié l'expression de gènes à l'échelle du transcriptome entre **populations souterraines de la même espèce** pour trois espèces différentes (*A. aquaticus*, *P. coxalis* et *P. meridianus*, Fig. 38), puis dans un second temps **entre espèces d'asellidés de surface et souterrains** (dans un contexte de paires d'espèces ou non).

Changement d'expression de gènes chez les populations souterraines.

Entre populations de surface et souterraines d'une même espèce, on retrouve entre 1% et 13% de gènes différentiellement exprimés (DE) avec à peu près autant de gènes sur- et sous-exprimés (Fig. 39). La très grande majorité des gènes DE entre populations ne sont pas les mêmes dans les 3 espèces testées. Un petit pool de gènes semble montrer un signal de sur-expression convergente en particulier entre les populations souterraines d'*A. aquaticus* et *P. coxalis*. Parmi les convergences attendues, les gènes impliqués dans la photo-réception devraient présenter un signal de sous-expression convergente dans les populations souterraines qui présentent une régression totale ou partielle des structures oculaires. Pour tester cette hypothèse, nous avons recherché les familles de gènes impliqués dans la photo-réception (les LIT pour « Light Interaction Toolkit ») telles que reconstruites par Speiser et al. (2014). Sur les 69 familles de gènes LIT retrouvées dans les transcriptomes de ces trois espèces, 21 gènes sont DE entre populations souterraines et de surface. Néanmoins, parmi ces 21 gènes, peu sont DE de façon convergente, seuls 4 gènes montrent des évidences de sous-expression convergente, parmi lesquels on retrouve la famille des opsines. A l'échelle populationnelle, on a donc un signal

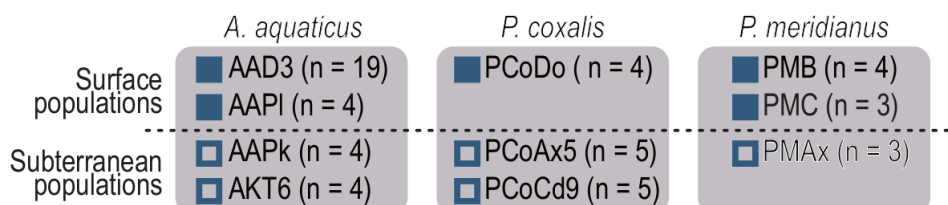


Fig. 38. – Design de l'étude des variations et des convergences d'expression de gènes entre populations souterraines et de surface pour 3 espèces d'asellidés. n: nombre d'individus séquencés indépendamment en RNA-seq.

Tiré de Grice et al in prep.

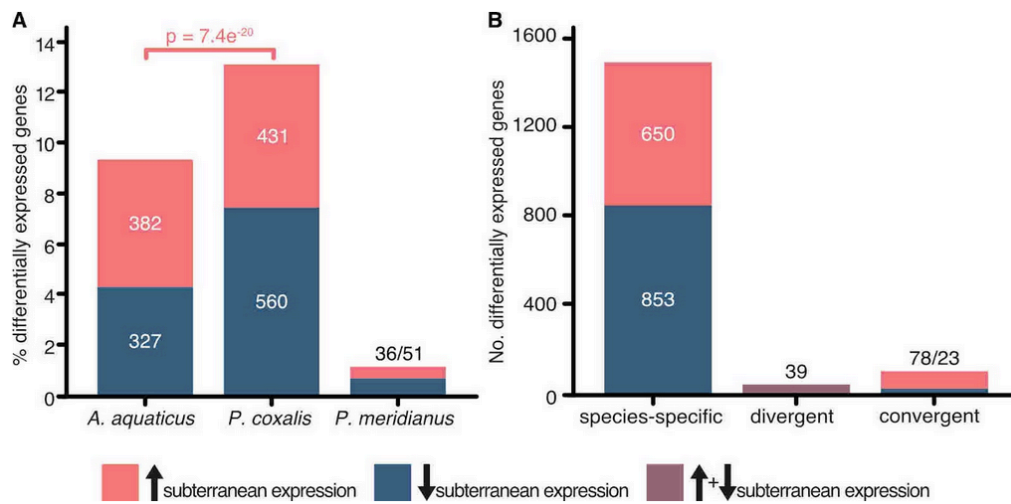


Fig. 39. – Différences d'expression entre populations de surface et souterraines, pour chaque espèce étudiée. **A** : pourcentage de gènes différentiellement exprimés (DE) entre populations de surface et souterraines (sur un set total de 7552 gènes partagés). La valeur de p indique que les gènes sur-exprimés chez les populations souterraines d'*A. aquaticus* et de *P. coxalis* se chevauchent plus qu'attendu par hasard. Toutes les autres comparaisons sont non-significatives. **B** : Nombre de gènes DE issus du panel A mais regroupés selon qu'ils sont trouvés DE dans une seule espèce (« species-specific »), DE dans au moins 2 espèces mais dans un sens opposé (« divergent »), ou DE dans au moins 2 espèces dans le même sens (« convergent »). Tiré de Grice et al in prep.

faible de convergence : beaucoup de gènes changent d'expression, et le signal de convergence est limité à quelques gènes, même pour des fonctions en cours de régression où la convergence d'expression était fortement attendue.

L'étude populationnelle est néanmoins complexe : il est difficile d'avoir beaucoup de réplicats populationnels, les temps de divergence et de colonisation du milieu souterrain sont probablement très variables et difficiles à estimer, et il est possible que cette échelle temporelle soit bruitée par le polymorphisme. En effet des études sur différentes populations d'*A. aquaticus* ont montré de fortes variations morphologiques entre populations de surface et souterraines mais également de fortes variations entre populations souterraines (Balázs et al. 2021; Konec et al. 2015). A l'inverse de la vision classique du syndrome de troglomorphie où les phénomènes de convergence dominant l'évolution des phénotypes, il est possible que dans la première phase de colonisation du milieu souterrain, les populations opèrent de nombreux changements, certains convergents mais un grand nombre divergents. Un changement brutal d'habitat pourrait d'abord conduire à une forte dérégulation des phénotypes. A plus grand échelle temporelle, beaucoup de nouvelles combinaisons de phénotypes disparaissent, voire des populations souterraines entières, les populations survivantes portant elles un signal de convergence potentiellement plus fort.

Convergence de l'expression des gènes entre espèces souterraines.

La littérature semble indiquer que la divergence influe fortement sur le taux de convergence moléculaire (Bohutínská & Peichel 2024). Plus deux taxons seraient proches, plus le nombre d'opportunités de convergence serait grand : allèles en commun, gènes en commun, architecture génomique commune... autant d'éléments sur lesquels la sélection peut agir de manière convergente. Si l'on prend l'ensemble du jeu de données Asellidae pour lequel nous disposons de données d'expression pour plusieurs individus (42 taxons au total) et que l'on regarde la

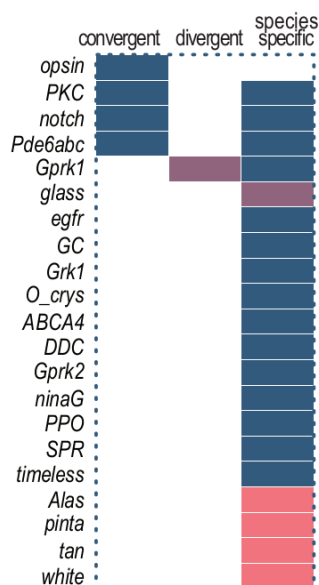


Fig. 40. – 21 familles de gènes LIT montrant des évidences d’expression différentielle entre populations de surface et souterraines (parmi 69 familles annotées). Le code couleur est le même que celui de la figure Fig. 39. Seules 4 familles montrent des évidences claires de sous-expression convergente, parmi lesquelles on retrouve les opsines. Tiré de Grice et al in prep.

relation entre divergence moléculaire (nombre de substitutions par site dans les séquences codantes) et nombre de gènes différentiellement exprimés, on trouve une très forte relation positive (Fig. 41) : plus deux espèces divergent, plus elles présentent un nombre important de gènes DE. Si maintenant on sépare les taxons en fonction de leur habitat, on observe un effet additionnel : le nombre de gènes différentiellement exprimés augmente quand l’un des taxons est souterrain. Le passage à la vie souterraine conduit donc à une accélération de l’évolution du transcriptome (du moins des niveaux d’expression). Cette accélération est d’autant plus forte que les deux taxons sont souterrains : chaque transition environnementale conduit donc à un excès de gènes différentiellement exprimés même si les transitions sont convergentes. Cette première analyse semble donc indiquer que l’évolution de l’expression des gènes est d’abord influencée par le temps de divergence puis par le nombre de changements d’habitat qu’ils soient convergents ou non. Ce double effet est compatible avec un scénario de dérive génétique supplémentée d’un processus de sélection diversifiante associée aux sauts d’habitat, mais laisse peu de place à des processus de convergence. En effet, si les processus de convergence sont dominants, alors le nombre de gènes différentiellement exprimés devrait chuter quand deux taxons souterrains sont comparés.

Nous avons ensuite explicitement recherché le nombre de gènes dont l’expression est convergente en formant 10 paires d’espèces souterraines et de surface (Fig. 42). Le jeu de données d’expression dans sa globalité est d’abord fortement structuré phylogénétiquement, mais deux axes d’une ACP sur la matrice d’expression de gènes sont corrélés au statut surface/souterrain (Fig. 42). Environ 7% des gènes testés montrent un effet significatif de l’habitat sur leur expression, avec un excès de gènes sous-exprimés (Fig. 44). Si l’on compare au nombre de gènes dont l’expression est convergente entre populations souterraines d’une même espèce (cf section précédente), nous sommes dans les mêmes proportions alors que cette fois nous comparons 20 taxons. En augmentant le nombre de taxons, et surtout en intégrant des taxons beaucoup plus divergents, vu la rapide accumulation de gènes différentiellement exprimés avec la divergence,

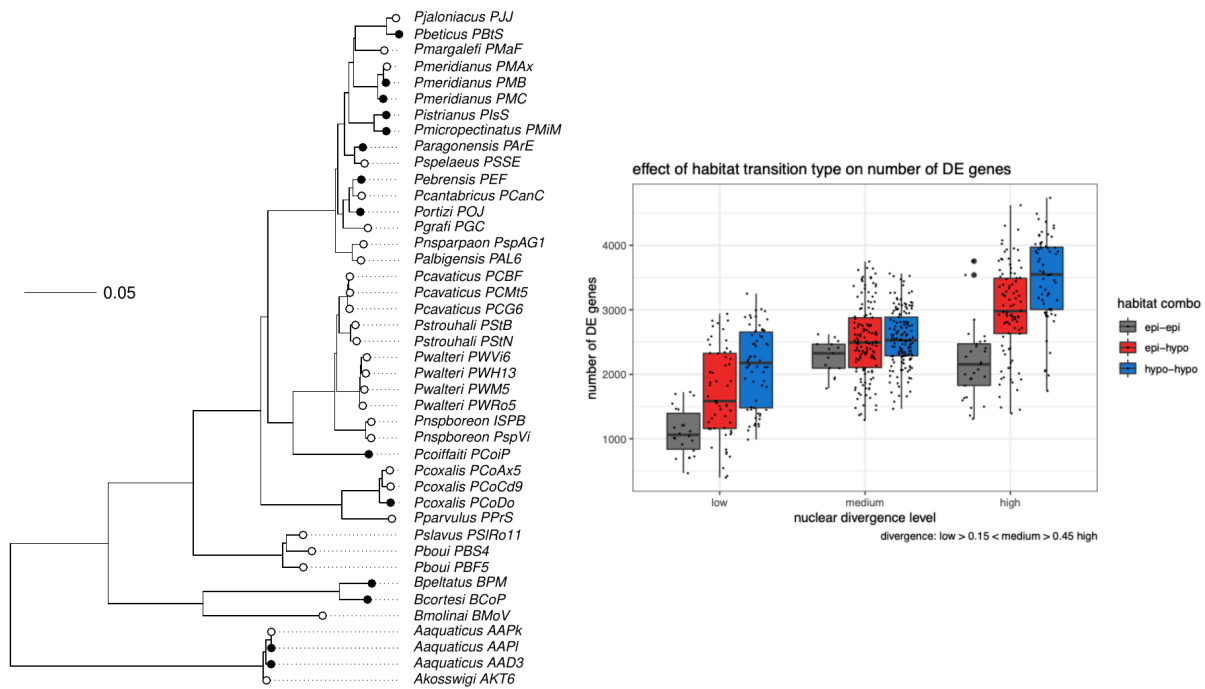


Fig. 41. – Jeu de données d'expression de gènes complet pour les asellidés (données intra et inter-espèces) représenté par un arbre phylogénétique (à gauche), et relation entre la divergence moléculaire et le nombre de gènes différentiellement exprimés (DE, à droite). Les divergences ont été regroupées en 3 groupes correspondant globalement aux relations entre espèces proches appartenant à un même sous-clade, à des espèces de sous-clades différents mais du même genre, et des espèces de genres différents. Les différences d'expression de gènes ont été testées sur 8756 familles de gènes (1000 gènes = 11% du jeu de données). Tiré de Grice et al in prep.

on augmente aussi la probabilité de trouver des gènes qui sont convergents par chance chez les espèces d'un même habitat. On peut essayer d'estimer le nombre de gènes convergents dans leur expression sous l'hypothèse que l'habitat n'a pas d'effet particulier en permutant aléatoirement les habitats mais en conservant la même structure en paires d'espèces. Les résultats de ces permutations sont assez étonnants (Fig. 43) : le nombre de gènes convergents observés est parfaitement centré au milieu de la distribution nulle empirique. Cette analyse ne soutient donc pas que la convergence soit issue de pression de sélection convergente agissant indépendamment sur un même nombre de cibles limité, mais plutôt que la convergence dans l'expression des gènes soit un simple effet secondaire de la saturation du processus, un attendu aléatoire, quels que soient les habitats des taxons étudiés.

Si l'on regarde le set de gènes convergents touchant les fonctions liées à la photoréception, on observe qu'une part conséquente des gènes du LIT (« Light Interaction Toolkit ») montre des évidences d'expression convergente (15 gènes sur les 83 testés, Fig. 44). C'est trois fois plus qu'attendu si le phénomène de convergence était distribué aléatoirement sur l'ensemble des gènes testés (18% des gènes du LIT contre 7% du transcriptome). On remarque également que la plupart de ces gènes convergents montraient déjà des variations d'expression dans au moins une espèce dans l'étude populationnelle précédente, bien que ce soit parfois dans une direction opposée. A l'inverse des précédents résultats, ces observations supportent l'hypothèse de convergence « directionnelle » (c'est à dire ciblant les mêmes gènes) et ré-ouvre donc la possibilité que la sélection naturelle modèle l'expression des gènes de façon convergente dans les populations souterraines.

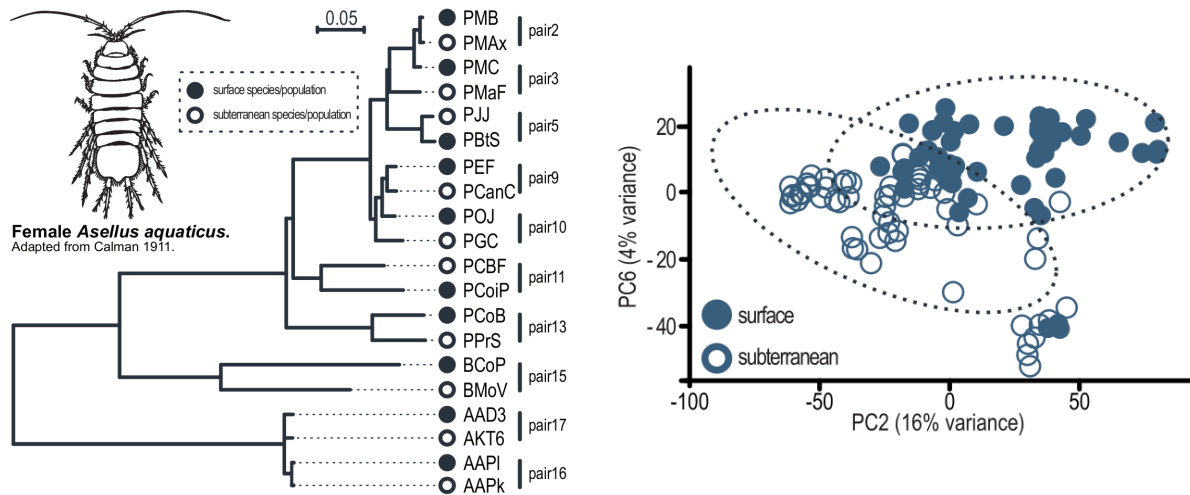


Fig. 42. – Jeu de données d'expression de gènes entre 10 paires d'espèces souterraines et de surface (phylogénie de gauche) et relation entre l'expression des gènes des différents individus telle que représentée par une ACP. Seuls les axes de l'ACP (axes 2 et 6) montrant une corrélation entre le statut souterrain et de surface sont représentés. Les axes 1, 3 et 4 montrent eux un fort signal phylogénétique (K de Blomsberg) chaque axe récapitulant un part de la structure phylogénétique (sous-clade, clade, genre). Tiré de Grice et al in prep.

Limites et perspectives.

Ces premiers résultats sont assez difficiles à interpréter et souffrent probablement de la forte sensibilité de la méthode à la divergence. Nous utilisons l'approche classique de DESeq2 qui consiste à modéliser les comptages de lectures RNA-seq pour chaque gène en utilisant un modèle linéaire généralisé avec un loi binomiale négative. Cette approche est critiquable pour plusieurs raisons :

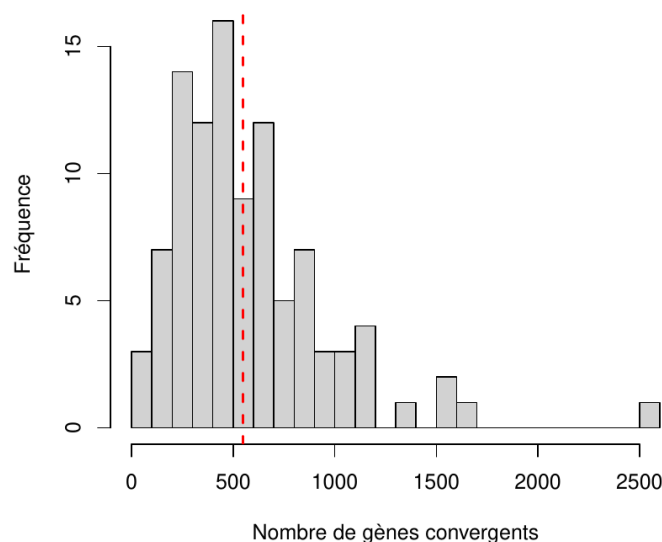


Fig. 43. – Nombre de gènes convergent dans leur expression après permutations aléatoires des statuts souterrain/surface pour le jeu de données à 20 espèces. La ligne verticale rouge indique le nombre de gènes convergents dans le jeu de données non permuté.

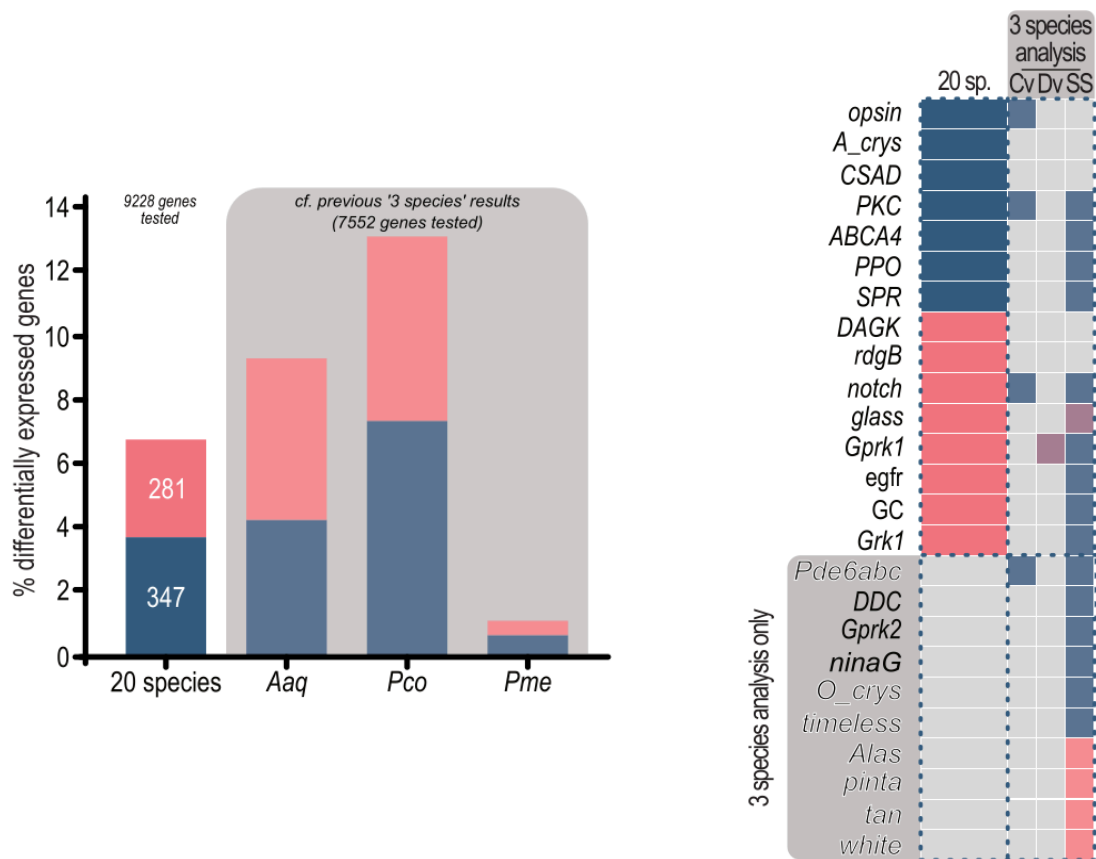


Fig. 44. – Nombre de gènes montrant des évidences d'expression convergente chez les espèces souterraines (à gauche) et gènes du LIT convergents dans l'étude à 20 espèces ou l'étude populationnelle (à droite). 9228 familles de gènes sont testées avec DESeq2 avec un modèle incorporant la structure en paires et l'habitat (`read_count ~ pair + habitat`). Tiré de Grice et al in prep.

1. Pour compenser les différences de profondeur de séquençage et la nature compositionnelle de ce type de jeu de données¹⁸, la méthode s'appuie sur une sorte de régression par la moyenne (la médiane des ratios entre comptages et moyenne géométrique par gène, Anders & Huber (2010)). Dit plus simplement, la méthode part du principe que la plupart des gènes ne sont pas différentiellement exprimés et servent à « aligner » le jeu de données. Ici certaines espèces montrent plus de 30% de gènes différentiellement exprimés, et il est possible que la méthode soit en dehors de sa gamme d'usage. Cela reste néanmoins à être testé, cette méthode est par exemple fréquemment utilisée pour comparer l'expression différentielle de gènes entre organes avec des variations inter-organes qui dépassent globalement les variations entre taxons (Brawand et al. 2011).
2. Alors que le signal phylogénétique est très prégnant (les axes 1, 3 et 4 d'une ACP sur les comptages normalisés montrent un fort signal phylogénétique), il n'est pris en compte que partiellement dans l'analyse. Le modèle utilise un facteur paires d'espèces qui permet de faire varier l'expression des gènes entre paires, mais n'intègre pas la divergence, la structure en sous-clades...

¹⁸Bien que les approches RNA-seq permettent d'obtenir des comptages, ce ne sont que des estimateurs de proportions relatives. Le problème devient alors épineux : un gène qui augmente fortement en expression va par définition faire baisser la proportion observée des autres gènes alors qu'ils ne sont pas différentiellement exprimés. La méthode de `DESeq2` gère ce problème et celui des différences de profondeur de séquençage en calculant des facteurs de tailles pour chaque échantillon.

3. Il n'y a pas d'intégration de cofacteurs importants comme des différences de temps de colonisation ou des phénotypes souterrains plus ou moins marqués.

Enfin, notre jeu de données biologique est lui aussi limité : le nombre de réplicats biologiques est peut-être faible pour certaines espèces (minimum $n=3$), et les données de RNA-seq sont issues d'extraction sur organismes entiers. Nos données de RNA-seq reflètent des variations globales d'expression, il n'est alors plus possible de différencier des variations d'expression à l'échelle d'un tissu de celles générées par des compositions différentes en tissus.

Un symptôme de la sous-optimalité de notre approche statistique est sa probable faible spécificité. L'évolution rapide des niveaux d'expression et leur analyse dans un contexte non-phylogénétique conduisent probablement à générer des faux positifs qui masquent un réel signal fonctionnel que l'on observe sur les gènes du LIT.

Il existe plusieurs solutions méthodologiques que nous allons explorer avec mon collègue de laboratoire Julien Clavel :

1. La première solution est d'utiliser des approches d'évolution de caractères continus qui intègrent directement les relations phylogénétiques. On ne cherche plus des gènes dont l'expression tend à augmenter ou baisser en lien avec un facteur, mais on cherche des gènes dont l'expression évolue vers une moyenne différente selon que le taxon est souterrain ou de surface. Pour cela on peut par exemple appliquer un processus d'Ornstein-Uhlenbeck (OU) sur les données d'expression (Chen et al. 2019) en intégrant les variations d'expression entre individus d'une même population (Rohlf et al. 2014). Pour chaque gène, on peut alors comparer un modèle OU à un modèle d'évolution non-directionnelle basé sur un mouvement brownien afin d'inférer les processus pilotant l'évolution de l'expression des gènes (dérive, sélection stabilisante, sélection divergente; Price et al. (2022)). Cette approche permet d'intégrer naturellement la structure phylogénétique mais également la forte dépendance à la divergence. Stern & Crandall (2018) l'ont par exemple utilisé sur des espèces d'écrevisses souterraines et de surface et ont pu montrer que les gènes liés à la vision montraient des expressions dont l'évolution accélérerait et était moins contrainte par la sélection.
2. Au lieu de décrire l'expression de chaque gène indépendamment, une approche alternative est d'utiliser les modèles précédents dans un format multivarié. En effet, il est attendu que l'expression de beaucoup de gènes covarie. Leur modélisation conjointe impose de travailler en grandes dimensions mais augmente le signal par ensemble de gènes. C'est une approche classique en morphométrie (Clavel et al. 2019) mais plus rare en génomique (Cope et al. 2020). Dans notre cadre, il serait intéressant de savoir si l'expression des gènes liés à la vision covarie, et si oui, comment leur évolution est impactée par la colonisation du milieu souterrain.
3. Jusqu'à présent l'étude du changement d'habitat s'est faite dans un cadre non-phylogénétique (étude de paires d'espèces) ou en contrôlant nos régressions pour la phylogénie (pGLS), mais sans jamais vraiment modéliser directement le changement d'habitat. Les approches phylogénétiques permettent de travailler sur des ensembles de branches (« branch model ») qui peuvent être assignés à un état de surface ou souterrain. Néanmoins il est presque impossible que le saut d'habitat se situe sur un noeud de la phylogénie : il aura

lieu le long d'une branche qui sera pourtant entièrement assignée au statut souterrain. De même, les modèles d'évolution de traits continus peuvent intégrer des changements de paramètres (par exemple l'optimum θ dans un modèle OU), voire changer entièrement de processus en passant d'une branche à une autre, cependant ces changements induiront une évolution graduelle des traits alors que des sauts d'habitats peuvent induire des changements brusques de traits. Une solution au problème de changement le long des branches et de saltationnisme pourrait être d'utiliser un processus de Lévy (Landis et al. 2013; Duchon et al. 2017). Le processus de Lévy permet des sauts plus ou moins brusques et à une fréquence plus ou moins élevée de valeurs de traits le long des branches. L'idée serait ici d'utiliser un processus de Lévy pour modéliser l'évolution de l'expression des gènes, potentiellement en guidant le processus sur les branches où ont eu lieu les sauts, voire en informant le processus des dates des sauts quand cette information est disponible.

1.6. Bilan et devenir du modèle Asellidae

A posteriori, construire un modèle de génomique évolutive autour des Asellidae et l'utiliser pour aborder des questions de biologie qui dépassent le modèle, apparaît comme une stratégie bien réfléchie. Utiliser ce modèle de sauts d'habitat pour démontrer comment l'écologie des espèces à travers les tailles de population, les traits d'histoire de vie ou l'environnement impacte l'évolution des génomes semble correspondre à un programme de recherche non-adaptationniste élaboré en réponse à l'adaptationnisme encore très prégnant en biologie. Il n'en est rien. La construction du modèle a bien sûr été réfléchie, mais son usage s'est développé au gré des opportunités. Pour tout dire, le débat entre adaptationnisme et neutralisme m'intéressait mais je le regardais comme une question qu'il fallait dépasser, une controverse historique... Je ne pensais pas que le débat était finalement toujours d'actualité et que j'allais si souvent y être confronté.

Bien qu'à l'origine le débat neutraliste/adaptationniste se soit construit sur des questions de taux de polymorphismes (paradoxe de Lewontin, Lewontin (1974)) et de substitution (l'hypothèse d'horloge moléculaire, Kimura & Ohta (1971)), il s'est étendu à d'autres traits génomiques comme la taille des génomes, la structure des chromosomes et des gènes, la composition en bases, la régulation et l'expression des gènes... (Galtier 2024). La théorie de la barrière de la dérive est applicable à tous les mécanismes moléculaires qui subissent une pression mutationnelle : la fidélité de la réplication, les mécanismes de réparation, de transcription, d'épissage, ... Face à un mécanisme ou une structure moléculaire complexe, on retrouve souvent deux scénarios alternatifs : la complexité de ces structures est issue d'un long processus d'optimisation, ou bien cette complexité est en partie issue de l'accumulation d'évènements faiblement délétères. La fixation d'évènements faiblement délétères n'implique pas forcément une baisse continue de la fitness des organismes, des évènements adaptatifs sont toujours possibles du moment qu'ils sont assez fortement adaptatif ($N_e \cdot s > 1$). Les génomes peuvent alors être vus comme un grand bazar construit par bricolage (Jacob 1977; Koonin

2009), un bricolage qui vient parfois compenser ou se construire sur des événements délétères, induisant une complexité croissante des génomes.

Une controverse récente a éclaté à la suite de la publication des travaux du projet ENCODE. Après une analyse de grande ampleur de l'expression et de l'activité du génome humain, les auteurs remettent en question l'idée selon laquelle une grande partie du génome humain serait non-codant et sans fonction (Dunham et al. 2012). La réponse ne se fait pas attendre : une activité n'implique pas à une fonction (Graur et al. 2013; Eddy 2013; Doolittle & Brunet 2017). Cette vieille controverse neutraliste contre adaptationniste a stimulé et continue à alimenter le champ disciplinaire de l'évolution moléculaire (Galtier 2024). Les termes du débat se sont élargis : on ne confronte plus les forces, mais on étudie leur équilibre et impact relatif. Modéliser l'impact conjoint des forces évolutives demeure cependant extrêmement complexe. Une solution est de construire un modèle nul sur les parties non-fonctionnelles des génomes pour rendre compte des effets démographiques et mutationnels, puis d'incorporer dans un second temps les effets sélectifs dans les régions fonctionnelles (Johri et al. 2020). Le débat est aujourd'hui de savoir si un tel modèle nul neutraliste et son extension quasi-neutre (nearly neutral model), continuent à avoir du sens et être informatifs (Kern & Hahn 2018; Jensen et al. 2019). En effet, même des sites neutres, à cause du déséquilibre de liaison avec des sites sous sélection, peuvent avoir des dynamiques qui seront mal modélisées par les modèles neutres ou quasi-neutres (Pouyet & Gilbert 2021).

Au lieu d'apporter des nouvelles approches de modélisation pour aborder ce débat complexe, le modèle Asellidae essaie de résoudre cette difficulté en se plaçant dans un cadre phylogénétique et écologique tellement contraint et fort qu'une analyse plus rudimentaire est suffisante. L'idée n'est pas nouvelle, et s'apparente à celle d'une expérience dans un « laboratoire souterrain de l'évolution » (Poulson & White 1969; Mammola 2019). Les conditions de l'expérimentation sont choisies pour que l'analyse des résultats soit simple. Je pense que nous avons apporté des observations importantes et contribué à une meilleure compréhension de l'évolution de la taille des génomes, de la composition des protéines et l'influence de l'environnement sur le taux de mutation. Le modèle Asellidae est toujours en développement et nous avons produit bien plus de données que nous ne sommes en capacité d'analyser. Pour la suite, nous avons trois projets d'applications du modèle : étudier (1) la relation entre complexité génomique et N_e au delà des tailles de génomes, (2) l'adaptation lors d'un saut d'habitat, et (3) le lien entre perte de phénotype et perte de gène.

1.6.1. Mais où est l'adaptation en milieu souterrain ?

Grâce aux multiples colonisations et réductions de taille de population, le modèle Asellidae est excellent pour l'étude de la dérive génétique. Pour être tout à fait franc, ce n'est pas pour cela que j'ai commencé à travailler dessus. Le projet de recherche que j'ai proposé en 2010 pour intégrer mon poste actuel s'intitulait « Génomique évolutive en milieu souterrain : un nouveau modèle biologique, environnemental et analytique pour comprendre le vieillissement ». Je voulais étudier les mécanismes moléculaires permettant aux organismes souterrains de multiplier par 5 à 10 leur longévité, et plus globalement comment des organismes peuvent s'adapter à des conditions environnementales si différentes. Pourquoi avoir changé de cap ? Une fois les premières données génomiques assemblées, l'impact des forces non-adaptatives saute aux yeux : gros génomes, expansion des éléments répétés, augmentation du d_N/d_S , baisse du taux

de substitution... Cet ensemble de phénomènes génomiques est également observé chez des téléostéens souterrains (Zhao et al. 2022), on pourrait alors parler de syndrome génomique souterrain.

La théorie quasi-neutraliste implique que les mutations faiblement délétères et avantageuses se comportent comme des mutations neutres. Néanmoins, les mutations à fort effet délétère ou avantageux ne devraient pas être impactées par une baisse de la taille efficace. L'adaptation a forcément impacté l'évolution des génomes d'Asellidae souterrains. Nous ne l'avons pas observée car nous ne l'avons pas cherchée. Avec les données de génomes complets qui s'accumulent en ce moment même (l'assemblage du 20^{ème} génomes d'asellidés est en cours au laboratoire) et l'ajout de données de polymorphisme (séquençage Illumina pour 10 individus par espèce), nous allons pouvoir aborder la question de l'adaptation en milieu souterrain plus sereinement. Plusieurs chantiers vont pouvoir démarrer :

1. Recherche d'évidence de sélection directionnelle sur l'expression de gènes chez les organismes souterrains (ou de modules de gènes) tel que décrit dans la section 1.5.0.3
2. Recherche de convergence dans les séquences des protéines des espèces souterraines (Duchemin et al. 2023; Rey et al. 2019).
3. Recherche de locus évoluant sous sélection positive par des approches classiques de scan de sélection (test MK, recherche de balayage sélectif, ...). Il faudra néanmoins prendre en compte les variations de taille efficace de population dans l'application de ces tests classiques.

1.6.2. Origine de la complexité génomique

Le rôle central de la sélection naturelle dans l'émergence de traits phénotypiques complexes ne fait généralement pas débat. Un exemple classique est celui des structures oculaires : la vision gradualiste darwienne peut être jugée trop simpliste au profit de modules de gènes pouvant interagir au cours du développement d'un organe et générer des sauts dans l'évolution des traits (Oakley & Speiser 2015), mais on ne trouvera néanmoins pas grand monde pour mettre en avant des processus non adaptatifs dans l'origine de la complexité des structures oculaires¹⁹. Pour les traits génomiques, la situation est bien différente. L'origine de presque tous les traits génomiques complexes, en particulier ceux portés par les eucaryotes, peut être reformulée dans le contexte du débat neutralisme/adaptationnisme. C'est le cas par exemple de la structure des gènes en intron/exon (Lynch 2002; 2006) ou bien de l'épissage alternatif (Saudemont et al. 2017; Xu et al. 2019). Ces phénomènes complexes peuvent avoir une origine et un rôle adaptatif, ou bien ce sont des mutations neutres ou faiblement délétères qui ont réussi à se fixer chez les eucaryotes en raison de leur faible taille de population.

Ces questions peuvent (et seront) abordées avec le modèle Asellidae, mais un risque est de se placer dans un cadre phylogénétique trop court : certains de ces traits génomiques évoluent lentement, et le groupe des Asellidae n'est potentiellement pas assez vieux pour les étudier. De même en restant au sein des Asellidae, on perd en universalité. Au sein du projet NeGA, nous testons l'impact d'une réduction de N_e dans plusieurs groupes biologiques distincts mais aussi

¹⁹il y a débat sur le rôle de la sélection naturelle dans la régression des structures oculaires chez les organismes souterrains, voir la section 1.6.3

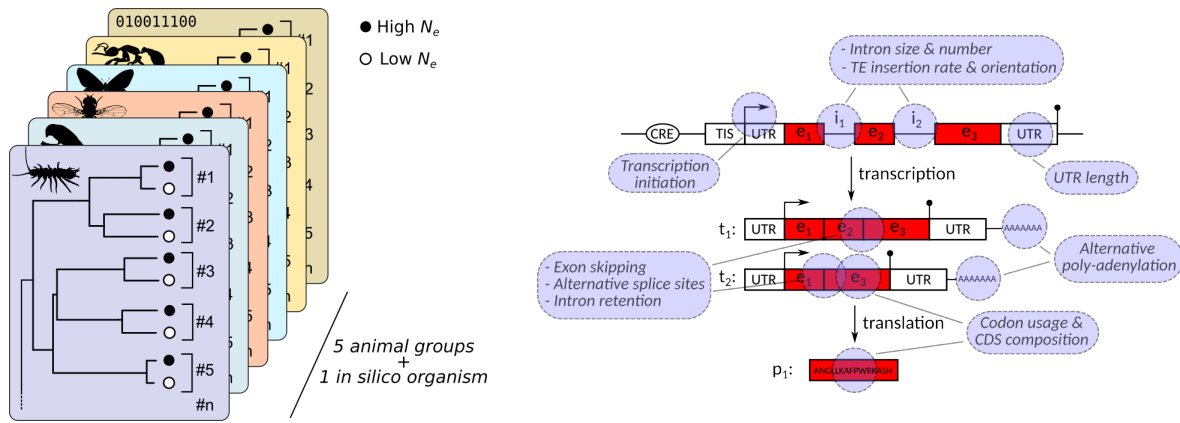


Fig. 45. – Groupes d’organismes (à gauche) pour lesquels l’influence de N_e est testée sur plusieurs traits génomiques dans le cadre du projet NeGA, dont la taille de génome, le contenu en éléments transposables, et la complexité de la structure, la composition et l’expression des gènes (à droite).

à l’échelle des métazoaires (Fig. 45). Dans sa thèse, Florian Bénétière a abordé la question de l’épissage et de la composition en base des gènes à l’échelle des métazoaires (Florian Bénétière, Duret, et al. 2024). Il a d’abord montré que le taux d’épissage des gènes est corrélé à N_e , et que donc la plupart des variants d’épissage sont probablement des erreurs de la machinerie transcriptionnelle (Florian Bénétière, Necsulea, et al. 2024). Nous avons également étudié le lien entre l’intensité de la sélection traductionnelle et N_e . La sélection traductionnelle conduit à l’usage des codons synonymes s’associant au tRNA les plus abondants. On s’attend donc à un usage préférentiel de ces codons dans les gènes les plus exprimés là où l’effet sur la fitness est le plus fort. Nous avons estimé le coefficient de sélection traductionnelle ($S = 4N_e s$) en comparant l’usage des codons synonymes optimaux entre les gènes les plus et les moins exprimés (dos Reis & Wernisch 2009) :

$$S^{hx} = \ln \left(\frac{FOP^{hx}}{1 - FOP^{hx}} \right) - \ln \left(\frac{FOP^{lx}}{1 - FOP^{lx}} \right)$$

Étonnement, alors que la théorie de la barrière de la dérive prédit que le processus d’optimisation sera d’autant plus efficace chez les espèces à grand N_e , les covariations entre S^{hx} et différents proxys de N_e sont très faibles (Fig. 46, Bénétière et al. (2024)). Dit autrement, N_e est un mauvais prédicteur de l’intensité de la sélection traductionnelle. Certaines espèces de métazoaire à fort N_e , comme la plupart des ordres d’insectes, montrent très peu d’évidence de sélection traductionnelle. In fine, mis à part les diptères et les nématodes, peu de métazoaires montrent de fortes évidences de sélection traductionnelle. Ce qui contrôle l’intensité de la sélection traductionnelle demeure donc un mystère : si N_e joue, il y a probablement d’autres facteurs qui agissent et masquent en grande partie son effet. Si l’effet de N_e n’est pas clair, celui des autres processus non-adaptatifs sur les biais d’usage des codons est lui flagrant. Chez les animaux, l’usage des codons synonymes (et donc la composition des gènes en troisième position) est avant tout déterminé par le biais mutationnel ou les biais de fixation dans la région génomique des gènes concernés (Fig. 47). Les espèces montrant de fortes variations génomiques de leur contenu en GC, comme l’espèce humaine, montrent de concert de très fortes variations du biais d’usage du code.

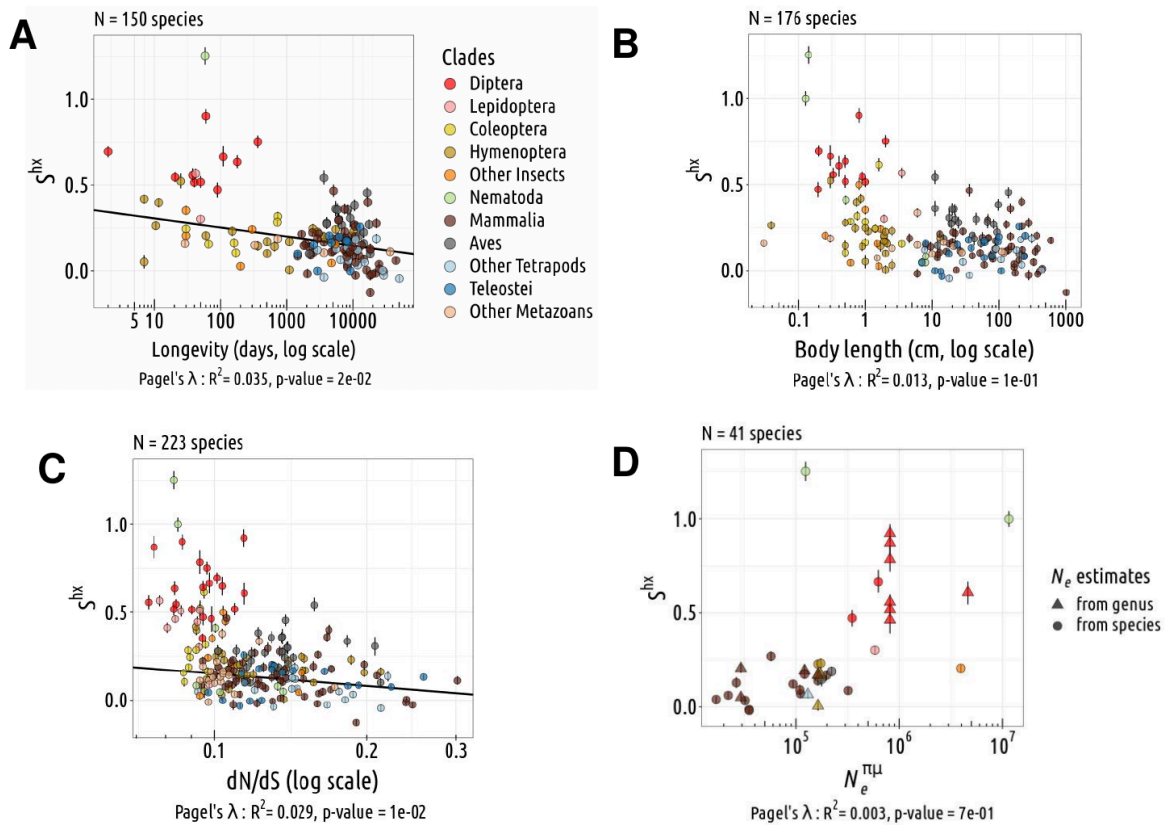


Fig. 46. – Relation entre N_e et l'intensité de la sélection traductionnelle (S^{hx}) et la longévité (panel A), la longueur du corps (panel B), le d_N/d_S (panel C), et des estimations de la taille efficace à partir de données de polymorphisme et de taux de mutation ($N_e^{\pi\mu}$) ($n_{\max} = 223$ espèces). L'intensité de la sélection traductionnelle S^{hx} est mesurée sur les 2% de gènes les plus fortement exprimés. Le modèle lambda de Pagel est utilisé pour prendre en compte la structure phylogénétique des données dans un modèle de régression (la droite de régression est affichée lorsque la corrélation est significative à 5%). Tiré de Bénitière et al. (2024).

La relation entre N_e et taille de génome n'est pas non plus observée à l'échelle des métazoaires (Marino et al. 2024). La notion d'échelle temporelle pour l'étude des effets de N_e est peut-être essentielle : les traits génomiques observés (taille de génomes, teneur en éléments transposables, taux d'épissage alternatif, usage des codons synonymes, ...) évoluent à des vitesses spécifiques et en interaction avec des forces qui varient elles-aussi dans des gammes temporelles particulières. L'observation de leurs interactions à partir de données empiriques macroévolutives pourrait alors dépendre directement de l'échelle évolutive de ces données. Un mauvais « focus » phylogénétique et la relation entre l'évolution des génomes et les forces évolutives devient floue (voire artificielle pour certaines observations à grande échelle comme Lynch & Conery (2003)). De même, les études à grande échelle évolutive, comme celle à l'échelle des métazoaires, apportent en principe plus de puissance statistique et de capacité de généralisation. Ces approches ont néanmoins une puissance statistique difficilement définissable (quelle gamme de variation et combien de transitions observe-t-on ?). Enfin, plus les distances phylogénétiques augmentent plus on compare l'incomparable : on étudie la covariation de quelques traits dans un contexte réel beaucoup plus complexe où une multitude de traits évoluent et co-évoluent. La probabilité d'avoir des covariants cachés augmente probablement avec la profondeur phylogénétique d'une étude.

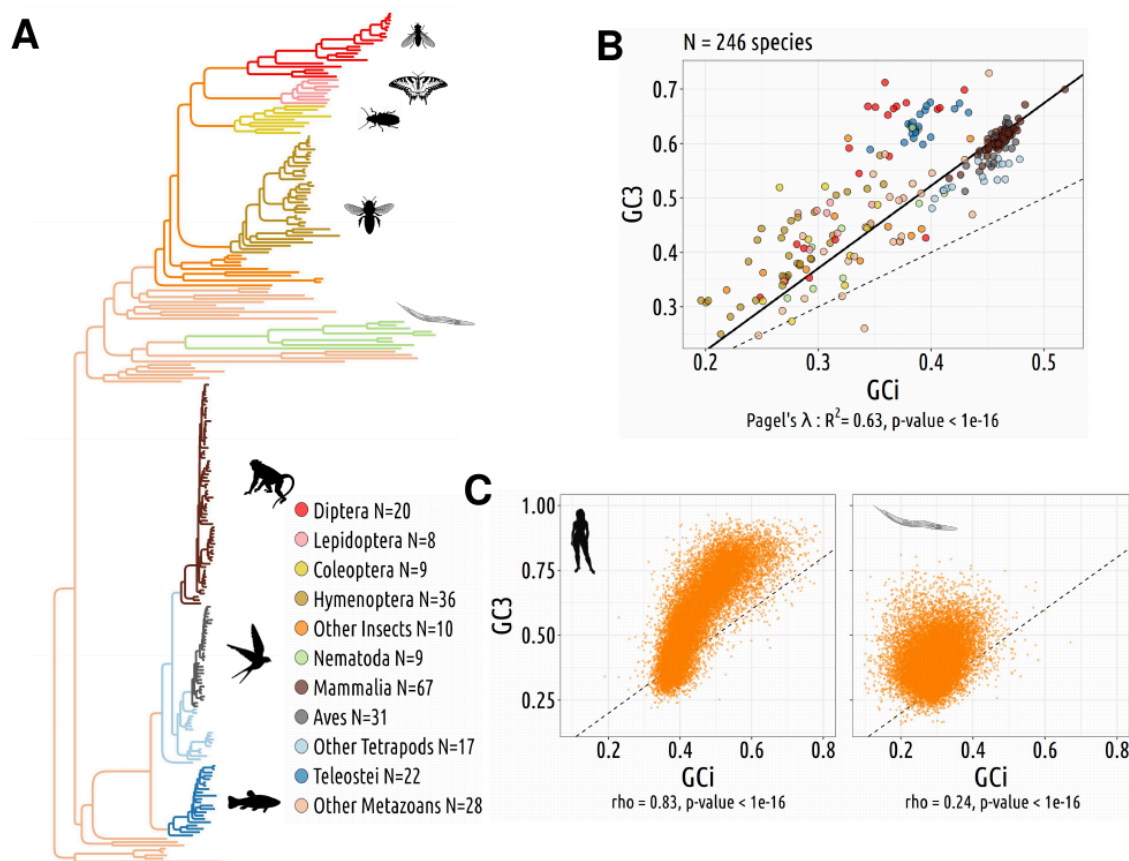


Fig. 47. – Les variations de l’utilisation des codons sont largement dues à des processus non-adaptatifs. A : Arbre phylogénétique des 257 espèces étudiées. B : Contenu GC moyen des gènes en troisième position des codons (GC_3) et GC moyen des gènes dans les introns (GC_i) pour chaque espèce. C : Corrélation entre GC_3 et GC_i chez *Homo sapiens* (à gauche) et chez *Caenorhabditis elegans* (à droite), chaque point correspondant à un gène. Le rho de Spearman et les valeurs de p correspondantes sont affichés sous le graphique. Les lignes pointillées correspondent à $x = y$. Tiré de Bénitiere et al. (2024).

Jusqu’à présent, on étudie l’effet de la barrière de la dérive, et donc le produit $N_e s$, sur un trait génomique en observant l’impact de variations de N_e . On sous-entend donc que s ne varie pas (ou plutôt que la distribution des effets sur la fitness des mutations, le DFE, est stable), alors qu’il est possible que l’évolution de traits cachés (le métabolisme, la thermorégulation, l’apparition d’un système de répression des éléments transposables...) impacte s et masque ainsi la relation N_e et traits génomiques. Malgré l’explosion de données génomiques, travailler sur un groupe moins profond et plus homogène dans sa biologie, avec un nombre mieux défini de transitions de N_e , comme les Asellidae, me semble une stratégie complémentaire qui garde sa raison d’être.

1.6.3. Perte de fonction et pseudogénéisation

Alors que les structures vestigiales, correspondant à des structures non-fonctionnelles ayant régressé à partir d’un état fonctionnel, ont participé à la construction de la théorie darwinienne de l’évolution, elles demeurent une des grandes questions ouvertes en biologie évolutive. En effet, les mécanismes par lesquels ces modifications régressives opèrent et l’importance quantitative de ces régressions dans l’adaptation à de nouveaux environnements restent largement débattus. Pour Darwin l’existence de telles structures était difficilement explicable par un mécanisme de sélection naturelle et, concernant la régression du système oculaire chez

les téléostéens cavernicoles, il finit par recourir à une notion de non-utilisation (« disuse ») (Darwin et al. 1859). Plus récemment deux grandes familles d'hypothèses ont été formulées. D'une part une hypothèse neutraliste où l'émergence de structures vestigiales est liée à la disparition de sélection purifiante et à la fixation par dérive de mutations non-fonctionnalisantes. A l'opposé les théories adaptatives (énergétique ou pléiotropique) reposent sur l'hypothèse que les structures fonctionnelles et non utilisées ont un impact sur la fitness des organismes les portant (Fong et al. 1995).

Les organismes souterrains sont depuis longtemps considérés comme des modèles pour comprendre les processus évolutifs à l'origine de traits régressifs tels que la perte de la pigmentation et des yeux (Culver & Pipan 2009). Des recherches dans les années 90 sur l'amphipode *Gammarus minus* et le téléostéen cavernicole *Astyanax mexicanus*, deux espèces avec des populations de surface et souterraines, ont révélé des résultats contradictoires. Des comparaisons de fitness ont montré que les caractères régressifs étaient soumis à la sélection naturelle dans les populations souterraines de *G. minus* (Jones et al. 1992). Au contraire, la formation d'hybrides avec des yeux fonctionnels restaurés et une importante variabilité inter-individuelle dans les caractères régressifs ont conduit Wilkens (1988) à soutenir un modèle purement neutre. Des recherches plus récentes ont affiné cette vision dualiste de l'évolution dans au moins deux directions.

Premièrement, les caractères régressifs semblent être contrôlés par des modifications développementales et des mutations de perte de fonction (« loss of function mutations ») similaires au sein de plusieurs populations et espèces. Les différentes populations cavernicoles d'*Astyanax* ont développé une expansion commune de leur signalisation Sonic Hedgehog (*Shh*) le long de la ligne médiane antérieure de l'embryon (Yamamoto et al. 2004). La surexpression de *Shh* induit l'apoptose du cristallin, ce qui déclenche la régression de l'œil (Yamamoto & Jeffery 2000). L'analyse de croisements hybrides a révélé que des mutations dans le même gène (*oca2*) sont responsables de l'albinisme dans différentes populations cavernicoles d'*Astyanax* (Protas et al. 2006). Il est frappant de constater qu'un défaut analogue dans les premières étapes de la voie de production de la mélanine est responsable de l'albinisme chez deux espèces de sauteuses souterraines (Bilandžija et al. 2012). De même, les gènes d'opsine montrent des signes de relâchement des pressions de sélection et d'accumulation de mutations non-fonctionnalisantes chez des espèces aussi différentes que des coleoptères Dytiscidae souterrains (Leys et al. 2005) et les poissons cavernicoles Amblyopsidae (Niemiller et al. 2013) (sans oublier les Asellidae souterrains).

Deuxièmement, il existe aujourd'hui de nombreuses preuves que des gènes pléiotropes peuvent en fait contrôler la régression et l'apparition synchrones de caractères au cours du développement d'un organisme (Protas & Jeffery 2012). Outre le contrôle de la régression des yeux, la signalisation *Shh* est également associée au nombre de papilles gustatives et à la taille de la bouche chez *Astyanax* (Yamamoto et al. 2009). Tout en perdant leurs yeux, ces téléostéens développent un comportement d'attraction vibratoire (VAB) lié à la présence de récepteurs sensoriels neuromastaux superficiels (SN) qui les aident à trouver de la nourriture dans l'obscurité. En croisant des organismes de populations de surface et souterraines, Yoshizawa et al. (2012) ont trouvé deux loci de caractères quantitatifs (QTL) qui contrôlent à la fois la taille des yeux, le nombre de SN et le VAB, ce qui suggère que les mêmes gènes

pléiotropes (ou des gènes voisins) sont impliqués dans l'évolution des caractères régressifs et constructifs. Il a également été constaté chez *Astyanax* que la signalisation Shh, en modifiant le développement du cerveau, contrôlait le compromis entre la recherche de nourriture et les comportements agressifs et d'agrégation en banc (Elipot et al. 2013). Enfin les mutations non-fonctionnalisantes dans *oca2* ont un effet sur la pigmentation et la perte de sommeil (O'Gorman et al. 2021).

Alors que les mécanismes génomiques associés à l'évolution régressive lors la colonisation des eaux souterraines sont de mieux en mieux catalogués pour quelques groupes modèles (voir par exemple Protas et al. (2011)), et que le débat neutraliste contre adaptationniste semble en partie dépassé (Culver et al. 2023), nous sommes aujourd'hui confrontés à deux questions majeures : La relative convergence des mécanismes d'évolution régressive entre les taxons souterrains indique-t-elle un nombre d'opportunités mutationnelles restreintes ? Si la sélection peut agir par mutation sur des gènes pléiotropes contrôlant les compromis entre les caractères régressifs et constructifs, quels sont ces gènes ?

Sur ces questions, les téléostéens souterrains concentrent la plupart des recherches, alors qu'ils montent paradoxalement peu d'évidences de perte de gènes comparé aux mammifères fouisseurs, possiblement car la plupart des téléostéens souterrains sont souterrains depuis peu de temps (Policarpo et al. 2021). Une future application du modèle Asellidae est d'aborder cette problématique avec cette fois une plus grande profondeur phylogénétique et un bien plus grand nombre de contrastes souterrain/surface. Plus spécifiquement, nous pourrons :

1. Quantifier le processus de régression moléculaire et tester l'effet d'un saut d'habitat sur la fréquence de ce processus.
2. Rechercher des familles de gènes qui s'étendent ou se contractent parallèlement chez les espèces souterraines (Balart-García et al. 2023)
3. Tester si au sein des pathways liés à la vision et la pigmentation, ce sont les mêmes gènes qui régressent.
4. Tester si les données de polymorphismes indiquent des évènements de balayages sélectifs autour des gènes qui régressent.

Les premiers résultats dont nous disposons indiquent que les populations souterraines tendent à perdre les gènes liés à la photoréception mais que les évènements de convergences strictes sont sans doute extrêmement rares (le meilleur exemple demeure les opsines, Fig. 48). A partir des génomes de quatre paires d'espèces souterrain/surface, si l'on compte le nombre de familles de gènes montrant une contraction du nombre d'homologues chez les espèces souterraines (i.e. des candidats à la régression convergente, Fig. 49), on trouve 344 familles candidates à la régression (sur 8924 familles, soit environ 4%). On trouve néanmoins à peu près le même nombre de gènes montrant le patron inverse (337 familles). Ces résultats préliminaires semblent indiquer que les sauts d'habitat et la régression de nombreux traits n'augmentent pas le taux de perte de gènes chez les espèces souterraines. Ils tendent plutôt à soutenir un modèle où l'épistasie et la pléiotropie conduisent au maintien de gènes même si la fonction principale qu'ils supportent a disparu (Forni et al. 2024).

Cette première analyse est néanmoins limitée par deux biais importants. Les espèces souterraines ayant des génomes plus gros et plus riches en éléments transposables, il est plus difficile d'annoter leurs gènes. Enfin un simple comptage de famille de gènes ne permet pas de

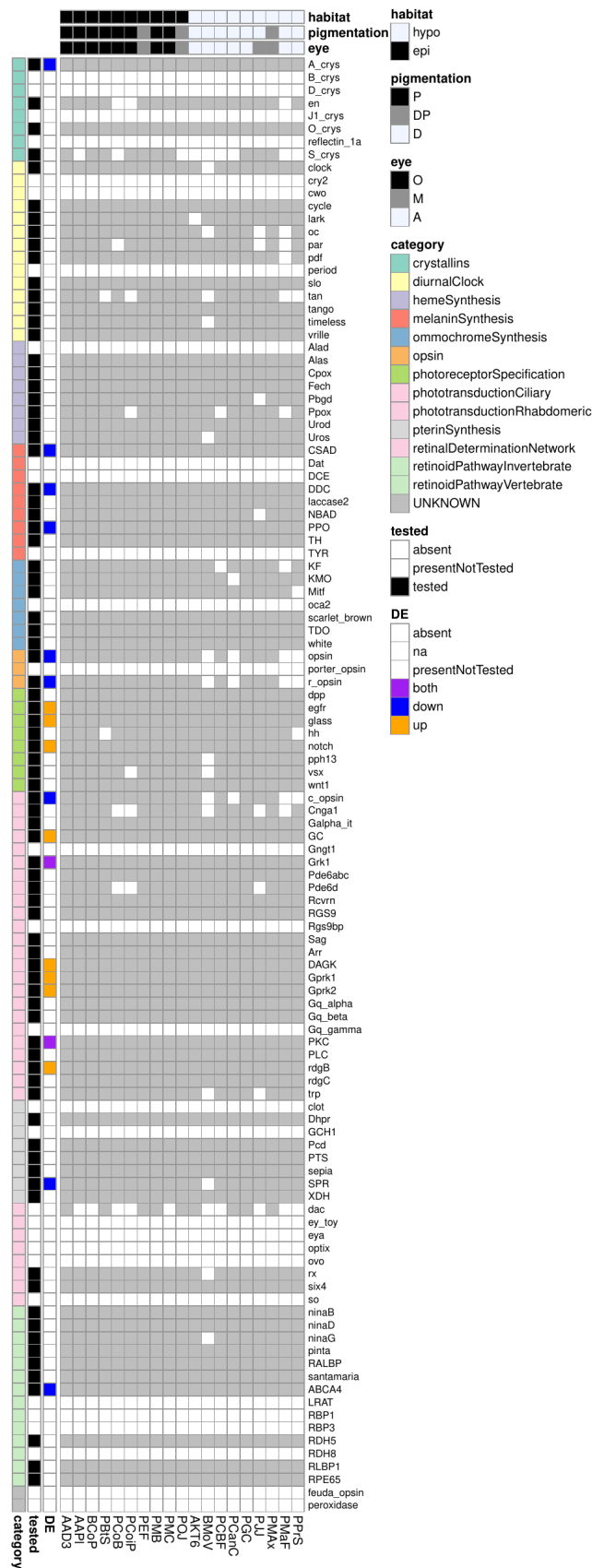


Fig. 48. – Recherche de gènes dont la fonction est liée à l'interaction avec la lumière (Light Interaction Toolkit, LIT; (Speiser et al. 2014)) dans 19 transcriptomes d'espèces et de populations d'asellidae de surface et souterrains. 55 familles du LIT sont présentes dans les 19 transcriptomes quel que soit l'habitat. 29 autres familles ont un ou plusieurs représentants dans au moins neuf populations. Aucun cas de famille de LIT spécifiques aux populations de surface n'a été détecté mais des familles comme les opsines montrent une tendance à être sous-exprimées ou absentes chez les populations souterraines. Tiré de Grice et al in prep.

Couple	1		2		3		4			
Espèce	●	○	●	○	●	○	●	○	Indice	Patron
Orthogroupe 1	X > Y	X > Y	X > Y	X > Y	X > Y	X > Y	X > Y	X > Y	4*(X>Y)	Candidat
Orthogroupe 2	X > Y	X > Y	X > Y	X > Y	X > Y	X > Y	X = Y	X = Y	3*(X>Y)	Candidat
Orthogroupe 3	X > Y	X > Y	X > Y	X > Y	X = Y	X = Y	X = Y	X = Y		NA
Orthogroupe 4	X > Y	X > Y	X > Y	X > Y	X < Y	X < Y	X < Y	X < Y		NA
Orthogroupe 5	X < Y	X < Y	X < Y	X < Y	X < Y	X < Y	X < Y	X < Y	4*(X<Y)	Inverse
Orthogroupe 6	X < Y	X < Y	X < Y	X < Y	X < Y	X < Y	X > Y	X > Y	3*(X<Y)	Inverse

Fig. 49. – Méthode de recherche de régression de famille de gènes chez 4 espèces souterraines. Les ronds noirs représentent les espèces de surface, les ronds blancs les espèces souterraines. Pour chaque familles de gènes, le nombre d'homologues est compté et comparé au sein de chaque paire d'espèce. Quand 3 ou 4 paires d'espèces montrent un nombre d'homologues plus important chez les espèces de surface, la famille est considérée comme candidate à la régression. Dans le cas inverse, on parle de famille « inverse ». Tiré du M2 de F. Blanchard (2024).

différencier un scénario où une famille tend à perdre des homologues dans un habitat, d'une famille qui tend à en gagner par duplication dans l'autre habitat. Une meilleure annotation des éléments transposables, l'utilisation d'approche phylogénétique où les taux de perte et de duplication sont estimés indépendamment, et bien sûr le passage à 10 paires d'espèces, devraient augmenter la sensibilité et la spécificité de ce test.

2. Projets en cours d'émergence : traces ADN et gestion de la ressource en eau

Dans cette section, j'aborde des programmes de recherche qui ont récemment débuté (projets Aliquot en 2024 et ADNe-STEU en 2025). Ils s'articulent toujours autour de thématiques d'écologie génomique mais abordent des problématiques très différentes de celles présentées précédemment, à savoir le lien entre biodiversité, ADN environnemental et dynamique des écosystèmes aquatiques. Ces recherches sont aussi beaucoup plus appliquées et répondent à un fort enjeu sociétal : une gestion durable de la ressource en eau.

Les écosystèmes d'eaux douces et les zones humides soutiennent de nombreux et essentiels services écosystémiques (Abigail J. Lynch et al. 2023). Ces milieux ont néanmoins subi de très fortes dégradations au cours du 20^{ème} siècle, et ils se dégradent plus vite que d'autres écosystèmes (Millennium Ecosystem Assessment 2005; Carpenter et al. 2011). La gouvernance de l'eau et les pratiques actuelles conduisent à un compromis entre l'accès à la ressource en eau et le maintien des autres services écosystémiques, conduisant souvent à la mise en compétition des différentes parties prenantes (Dudgeon et al. 2006).

Alors qu'une partie conséquente des services reposent sur la biodiversité aquatique, cette biodiversité est confrontée à de nombreuses menaces : surexploitation, pollution de l'eau, modification du débit, destruction ou dégradation de l'habitat, et invasion par des espèces exotiques (Dudgeon et al. 2006). A ce cocktail déjà compliqué s'est rajouté le changement climatique, les maladies infectieuses, la prolifération d'algues nuisibles, l'expansion de l'énergie hydroélectrique, les contaminants émergents (microplastiques, PFAS, ...), la pollution lumineuse, la salinisation de l'eau douce, le déclin du calcium ... (Reid et al. 2019).

En conséquence, à l'échelle globale, on observe de fortes baisses des abondances et des aires de distribution de la biodiversité aquatique des eaux douces (baisse de 86% du Living Planet Index depuis 1970, WWF (2024)). Les actions de conservation sont difficiles : ces écosystèmes sont des milieux receveurs des activités terrestres (eg. agricoles) et leur biodiversité est riche en espèces endémiques ce qui complique les plans de gestion (Dudgeon et al. 2006). Une étude sur plusieurs groupes taxonomiques montrent qu'un quart des espèces aquatiques d'eaux douces sont en risque d'extinction, en particulier à cause de la pollution, des barrages et de l'extraction d'eau, de l'agriculture et des espèces envahissantes (Sayer et al. 2025). La situation est considérée comme extrêmement urgente : l'empreinte humaine est bien trop forte sur les écosystèmes aquatiques d'eaux douces (Albert et al. 2021).

En Europe, probablement grâce à la Directive Cadre sur l'Eau (DCE) qui impose une surveillance de la qualité des eaux, une amélioration de la biodiversité des cours d'eau a été

observée au cours des décennies 1990 et 2000, mais cette amélioration s'est arrêtée depuis 2010 (Haase et al. 2023; Sinclair et al. 2024). Le tableau est sombre, même avec des actions de conservation urgentes et d'ampleur, avec l'intensification du changement global, les stress multiples et le probable dépassement de plusieurs points de bascule, il est peut-être déjà trop tard pour enrayer la perte irréversible de la biodiversité aquatique (Fig. 50, Dudgeon & Strayer (2025)). C'est dans ce contexte que je développe des outils moléculaires pour mieux évaluer la biodiversité des hydrosystèmes (en incorporant la diversité inconnue) et mieux suivre les impacts de certaines activités.

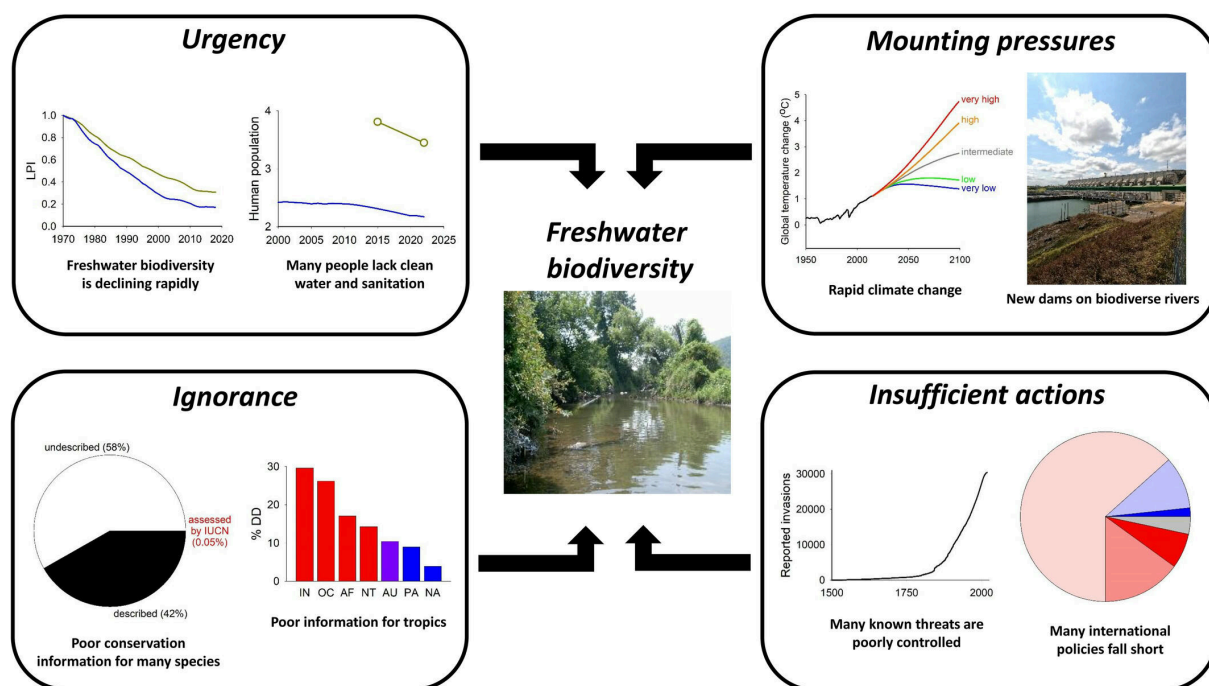


Fig. 50. – Principaux obstacles à l'infléchissement de la courbe de la perte de biodiversité aquatique continentale. **Urgence** : la biodiversité des eaux douces diminue rapidement, comme le montrent les tendances de l'Indice Planète Vivante (ligne bleue : animaux d'eaux douces, verte : animaux terrestres), tandis qu'une grande partie de la population humaine mondiale (en milliards de personnes) manque d'eau gérée en toute sécurité (ligne bleue) et d'assainissement (ligne verte). **Ignorance** : Les connaissances sur de nombreuses espèces d'eau douce sont insuffisantes, en particulier sous les tropiques. Le panneau de gauche montre la proportion d'éphéméroptères, de plécoptères et de trichoptères qui ont été évalués par l'UICN (0,05% des espèces), ainsi que ceux qui ont été décrits (noir) et qui restent à décrire (blanc). Le panneau de droite montre le pourcentage d'espèces de poissons d'eau douce jugées par l'UICN (2024) comme ayant des données insuffisantes (data deficient, DD) dans les domaines principalement tropicaux (rouge ; IN = Indomalayen, OC = Océanien, AF = Afrotropical, NT = Néotropical), les domaines principalement tempérés (bleu ; PA = Paléarctique, NA = Néarctique), et un domaine à climat mixte (violet ; AU = Australasien). **Des pressions croissantes** : le climat change rapidement et continuera de changer plus ou moins fortement en fonction de notre efficacité à réduire nos émissions de gaz à effet de serre, et de nouveaux grands barrages continuent d'être planifiés et construits sur de grands fleuves riches en biodiversité. **Actions insuffisantes** : le nombre d'invasions biologiques continue de s'accroître (panneau de gauche), montrant que les contrôles existants sont inadéquats, et les politiques internationales de conservation n'atteignent souvent pas leurs objectifs, le panneau de droite montrant les progrès accomplis dans la réalisation des 60 éléments contribuant aux objectifs d'Aichi définis en 2010 à la conférence de Nagoya sur la biodiversité : bleu foncé = critères dépassés ; bleu clair = critères atteints ; rose clair = critères non atteints mais quelques progrès réalisés ; rose moyen = pas de changement ; rouge = tendance à s'éloigner de l'objectif ; gris = pas d'évaluation possible. Tiré de Dudgeon & Strayer (2025).

2.1. Les acides-nucléiques environnementaux

Les 20 dernières années ont vu le développement exponentiel de l'écologie moléculaire et des approches dites « omiques » basées sur les technologies de séquençage à haut débit (Shokralla et al. 2012). Ces méthodes révolutionnent notre capacité à caractériser, inventorier et quantifier la biodiversité et les processus écosystémiques, y compris pour les groupes taxonomiques et les environnements qui étaient auparavant largement sous-étudiés. En particulier, les acides nucléiques environnementaux (ADNe et ARNe, appelés par la suite ANe) présentent un grand potentiel pour résoudre les problèmes de gestion des ressources aquatiques (Pawlowski et al. 2018; Cordier et al. 2021). On entend par ANe tout acide-nucléique qui se trouve dans un échantillon de l'environnement, peu importe son origine taxonomique ou son état de dégradation. C'est une définition très large, qui repose seulement sur la méthode d'échantillonnage. Certains collègues préfèrent une définition plus étroite, par exemple excluant les micro-organismes (voir par exemple la discussion initiée par Pawlowski et al. (2020)). Il existe plein de façon d'échantillonner des ANe aquatiques : filtration d'eau, collecte de sédiments, biofilms, capteurs passifs (Fig. 51)... Il existe aussi des versions intermédiaires : par exemple des échantillons issus de filets Surber contenant des organismes mais aussi des matrices environnementales (sédiments, litières) peuvent être broyé dans leur globalité avant extraction d'ADN. On parlera d'ADN de communauté ou ADN de bulk non trié (Fig. 52), qui peuvent rentrer dans la définition d'un ANe si on utilise une définition très large d'un échantillon d'ANe.

Quand on échantillonne des ANe, on échantillonne un mélange très hétérogène d'acides nucléiques dans différents états (Fig. 53), de différentes classes de taille, et plus ou moins dégradés (Power et al. 2023). L'ADNe peut être libéré dans l'environnement sous différentes formes, changer de forme, être transporté, stocké, adsorbé à des particules minérales, associé

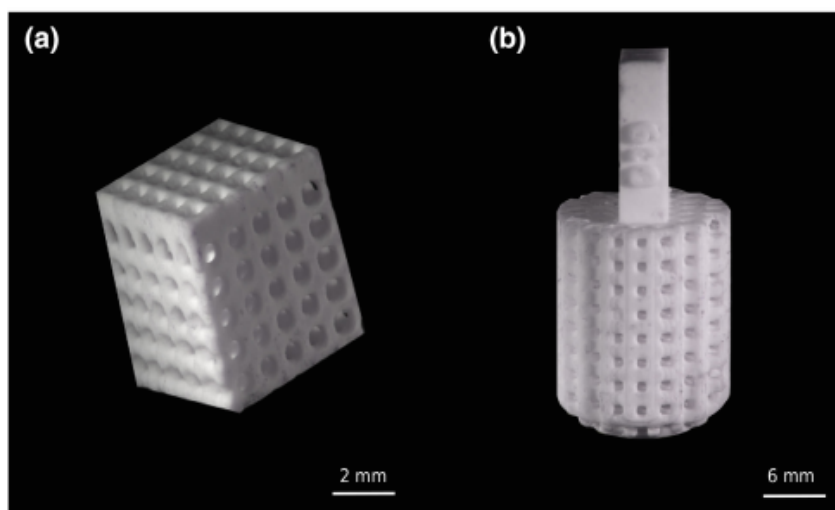


Fig. 51. – Capteurs passifs à ADN en hydroxyapatite obtenus par impression 3D. Nous avons développé et testé ces capteurs en laboratoire et sur le terrain. Ils captent les ADN mais montrent pour le moment une trop forte variabilité de rendement pour être utilisés en routine. Tiré de Verdier et al. (2022).

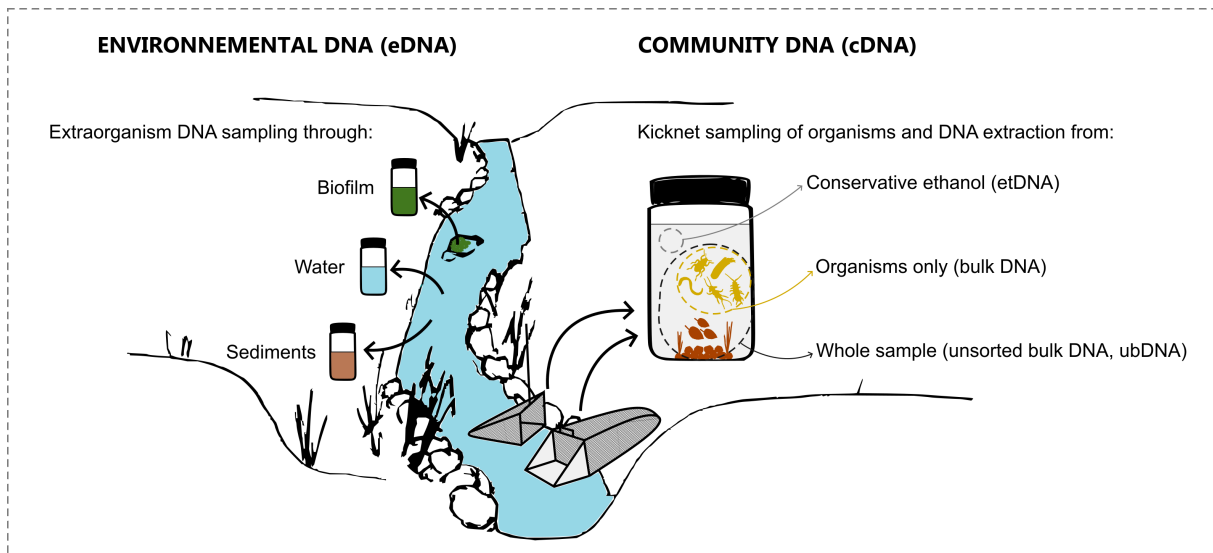


Fig. 52. – Exemple d'échantillons d'ADNe et d'ADN de bulk dans une rivière. Tiré de Gauthier et al in prep.

à des biofilms, remobilisé... Le cycles de vie d'un ADNe est donc potentiellement extrêmement complexe, ce qui a conduit certains auteurs à parler d'« écologie des ADNe » (Barnes & Turner 2016).

L'idée d'utiliser les ANe pour caractériser la biodiversité d'un écosystème date de 2008 (Ficetola et al. 2008). Dix sept ans plus tard, les ANe sont à la pointe du biomonitoring, de la biosurveillance et de la bioindication (Cordier et al. 2021; Bowers et al. 2021; Zhang et al. 2022). Malgré des progrès scientifiques considérables dans l'utilisation des ANe, le transfert opérationnel reste cependant limité, probablement en raison (i) d'un manque de connaissances sur l'« écologie des ANe » (production, transfert, dégradation) le long du continuum hydrologique (Barnes et al. 2021), (ii) du retard dans l'application de ces approches sur certaines masses d'eau (par exemple les eaux souterraines), et (iii) du manque d'outils stables et standardisés destinés aux parties prenantes.

Les travaux sur la variabilité spatiale et temporelle des ANe sont assez rares et se concentrent souvent sur quelques groupes modèles (souvent des téléostéens) (Pont et al. 2018; Searcy et al. 2022). Il est essentiel de caractériser et de comprendre cette variabilité à différentes échelles pour rendre ces approches opérationnelles dans un cadre réglementaire. Par exemple, on

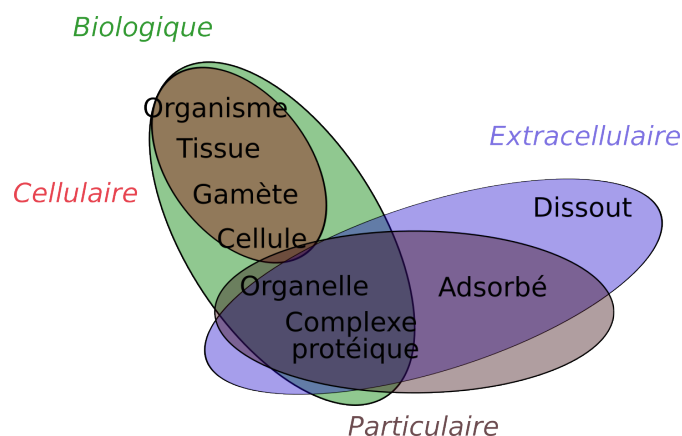


Fig. 53. – Les différents états de l'ADN environnemental.

considère souvent que les ANe fournissent des informations fiables sur l'occurrence mais des informations dégradées sur l'abondance des organismes (Sassoubre et al. 2016), ce qui rend difficile l'intercalibration avec des méthodes non moléculaires. Néanmoins, la dégradation de ce signal quantitatif peut simplement être le résultat de nos difficultés à comprendre la dynamique spatio-temporelle des ANe dans les hydrosystèmes. A l'exception de quelques taxons, on connaît mal les distances de dispersion des ANe et l'influence des paramètres hydrauliques et sédimentaires sur cette dispersion, les taux de dégradation/stockage ou les événements potentiels de remobilisation des ANe.

Même le taux de production des ANe est mal connu : par exemple, on ne sait pas si l'influence de certains traits biologiques sur la quantité et les caractéristiques des ANe émis comme la biomasse (Yates et al. 2021), le métabolisme, la présence d'un exosquelette (Verdier et al. 2024) observés chez quelques taxons est généralisable (Thalinger et al. 2021). De même, le rôle de la phénologie (eg. reproduction, mue) dans l'émission des ANe reste inexploré pour la plupart des organismes. Enfin, la dynamique temporelle des hydrosystèmes est susceptible d'influencer le devenir des ANe le long des réseaux hydrographiques : quelle est l'influence des crues et des périodes d'assec, des phases de déstratification des lacs, des fluctuations de la température, du pH ou de l'exposition aux UV ? Toutes ces connaissances de base sur « l'écologie des ANe » font encore défaut et entravent considérablement la définition du cadre spatio-temporel des études basées sur les ANe.

Enfin, beaucoup d'usages des ANe, comme les inventaires de biodiversité par métabarcoding, reposent sur des bases de données taxonomiques moléculaires dites de référence dans lesquelles des séquences de gènes (16S, ITS, COI, RBCL...) sont attribuées à des espèces. Des groupes taxonomiques entiers, voire des écosystèmes (par exemple, la plupart des masses

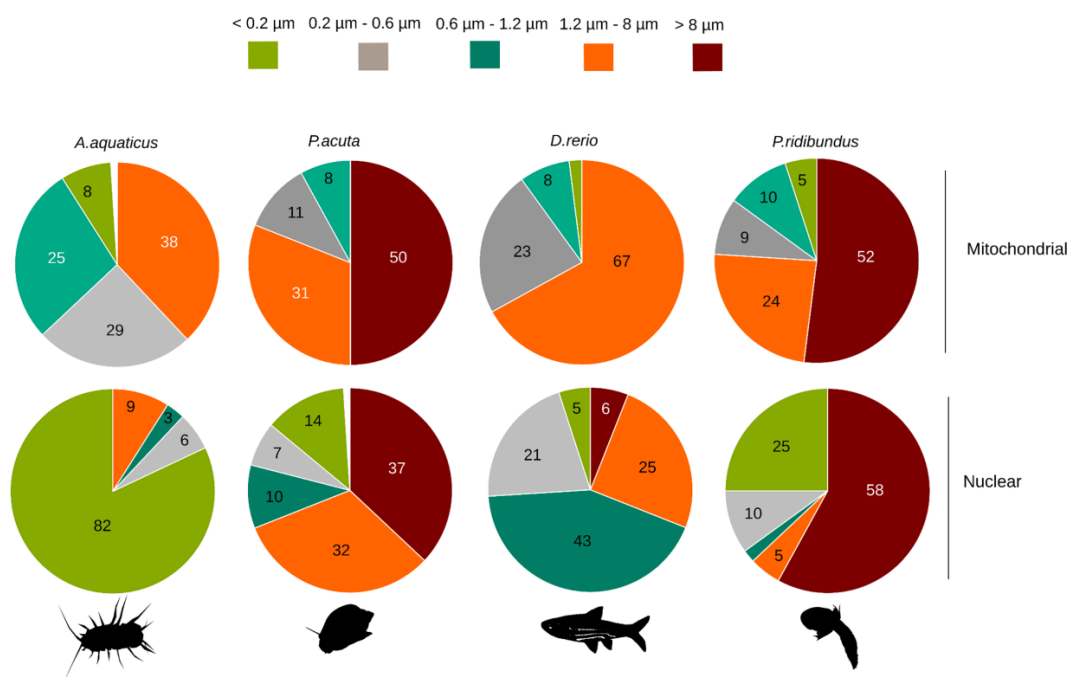


Fig. 54. – Taille de particules émises par quatre métazoaires très contrastés (isopode, gastéropode, téléostéen, amphibien). Les différentes classes de particules sont obtenues après filtration en série d'échantillons d'eau issus de mésocosmes, puis quantification par qPCR. En plus de variation forte de la taille des particules émises, les isopodes émettent de 100 à 1000 fois moins d'ADNe que les autres taxons. Tiré de Verdier et al. (2024).

d'eau tropicales, les « crustacés » souterrains), sont sous-représentés dans ces bases de données, ce qui limite l'utilisation des ANe pour la biosurveillance de ces groupes ou écosystèmes (Sinniger et al. 2016; Keck et al. 2023).

Le projet Aliquot du PEPR OneWater que je coordonne et qui regroupe une dizaine de laboratoires partenaires va tenter de relever certains de ces défis (et d'autres). Par la suite, je décrirai deux projets de recherche associés à l'utilisation des ANe : (1) Utiliser les ADNe pour mesurer les impacts d'une pollution organique dans un contexte d'intermittence, (2) mieux comprendre la dynamique spatio-temporelle des ADNe par la modélisation hydraulique.

2.2. Mesurer des impacts sans échantillonner les organismes

En Europe, la directive-cadre sur l'eau (DCE, directive 2000/60/CE) impose la surveillance des masses d'eau et l'évaluation des impacts anthropiques à l'aide de méthodes normalisées combinant des évaluations chimiques et biologiques (Grizzetti et al. 2017). Alors que les évaluations chimiques fournissent des instantanés à court terme de la qualité de l'eau, les éléments de qualité biologique (EQB) - tels que les communautés d'invertébrés benthiques, diatomées, macrophytes et téléostéens - apportent une vision intégrative des perturbations (Marchant et al. 2006). Parmi ces éléments, les macroinvertébrés aquatiques²⁰ sont largement utilisés en raison de leur sensibilité à divers facteurs de stress environnementaux. Les méthodes traditionnelles de biosurveillance utilisant des macroinvertébrés (ci-après dénommées méthodes morpho-taxonomiques) impliquent la collecte, le tri et l'identification morphologique des organismes, suivis du calcul d'indices de qualité basés sur l'abondance et l'occurrence des espèces, ainsi que sur des données relatives à leurs traits biologiques (par exemple, polluosensibilité, polyvoltinisme). Bien que cette approche ait été normalisée pour les rivières pérennes, son application reste très difficile et souvent inadéquate pour des écosystèmes plus complexes, tels que les rivières intermittentes.

Les cours d'eau non pérennes, dont le débit s'arrête à un moment de l'année, sont extrêmement fréquents et répandus sur tous les continents (Messenger et al. 2021). Ces écosystèmes uniques, alternant des phases humides et sèches, créent des habitats diversifiés, tant aquatiques (lotiques et lentiques) que terrestres (canaux secs) (Datry et al. 2018). Ce dynamisme façonne la distribution et la diversité des organismes aquatiques, ce qui se traduit par des communautés très distinctes des rivières pérennes (Leigh & Datry 2017). Cette forte variabilité temporelle et spatiale rend difficile l'utilisation d'outils développés pour surveiller l'état écologique des ri-

²⁰J'ai longtemps lutté contre l'usage du terme « invertébré » : un bel exemple de groupe paraphylétique, formé sur l'absence d'un trait (pas sa perte, sa non-existence), anthropocentré, aux contours flous... On pourrait le remplacer par protostomien, ou par un ensemble de taxons, principalement les mollusques, annélides et arthropodes (les MAA ?)... mais c'est une cause perdue, tous les écologues utilisent et comprennent le terme invertébré (sans pouvoir le définir). Je me résigne donc à l'utiliser...



Fig. 55. – Exemple d’une rivière intermittente, l’Albarine (Ain, France), avec un tronçon de rivière où l’écoulement s’est arrêté mais où demeurent des zones d’eau stagnante (à gauche). Sur le même cours d’eau, exemple de rejet de station de traitement des eaux usées (à droite).

vières pérennes, conduisant les rivières intermittentes à être souvent oubliées des programmes de biosurveillance (Stubbington et al. 2018).

La biosurveillance utilisant la méthode morfo-taxonomique est confrontée à deux difficultés lorsqu’il s’agit d’évaluer l’état écologique d’une rivière intermittente. Cette méthode ne prend pas en compte l’effet de l’assèchement (effet qui s’observe longtemps après la remise en eau) et qui vient masquer les perturbations anthropiques. Enfin beaucoup d’organismes des rivières intermittentes sont sous forme de stades dormants qui sont difficilement identifiables, on a alors souvent une description de ces communautés à un très gros grain taxonomique.

Nous avons testé si une approche d’écologie moléculaire pouvait être utile à la bioindication des rivières intermittentes, en particulier :

1. Est-ce qu’une approche de métabarcoding en travaillant à un grain taxonomique beaucoup plus fin permet de mieux détecter un impact ?
2. Est-ce qu’une approche utilisant une matrice ADNe permet d’observer un impact local en rivière ? Le transport des ADNe par le courant rend potentiellement difficile la détection d’impacts locaux.

Pour cela, nous avons sélectionné trois rejets de stations de traitement des eaux usées (STEU) situés sur un gradient d’intermittence sur la rivière Albarine (Fig. 56). Un suivi classique (norme I2M2, Mondy et al. (2012)), des ADN de bulks non triés (Fig. 52), et des ADNe d’eau filtrée ont été échantillonnés en amont et aval de chaque rejet. Après extraction d’ADN, un fragment du COI a été amplifié (Leese et al. 2021), et séquencé en masse.

Les résultats montrent que seules les matrices ADNe permettent d’observer des différences significatives de communautés en amont et en aval de rejets dans tous les contextes d’intermittence (Permanova, Fig. 57). De même, seuls les échantillons d’ADNe permettent d’observer une baisse de diversité en aval des rejets dans tous les contextes d’intermittence. Enfin le signal de bioindication est en grande partie associé au patron de présence/absence d’ADNe assignés à une cinquantaine de genre de Chironomidae différents.

Ces résultats sont très encourageants mais soulèvent d’autres interrogations :

- pourquoi détecte-t-on mieux les effets locaux avec de l’ADNe ? Notre hypothèse est que le taux de détection des taxons biodiindicateurs est plus fort avec l’ADNe

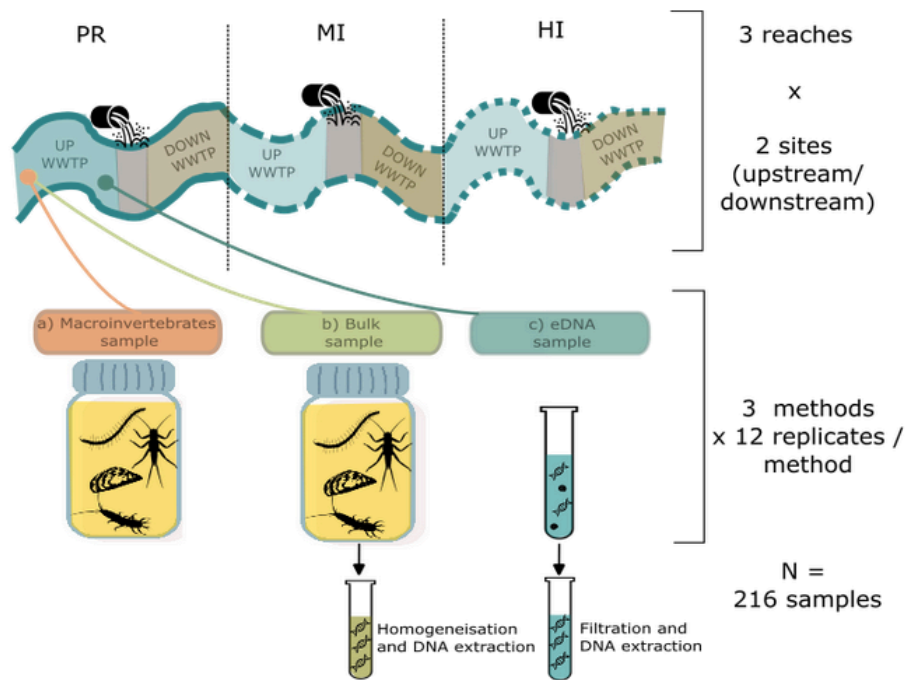


Fig. 56. – Plan d'échantillonnage pour tester la capacité d'ADNe de rivières intermittentes à détecter une pollution organique (rejet de STEU). PR: tronçon de rivière pérenne, MI: intermittence moyenne, HI: forte intermittence. Tiré de Verdier et al in prep.

- pourquoi la diversité des ADNe augmente le long de l'Albarine, alors qu'elle baisse avec les approches morpho-taxonomiques et ADN-bulk ? Notre hypothèse est que les ADNe dispersent des amonts de l'Albarine vers les aval, ce qui enrichit les échantillons pris en aval.
- Comment concilier les deux hypothèses précédentes : comment est-il possible qu'une méthode aie la capacité de détecter des effets locaux tout en accumulant les ADNe des amonts ?

Nous allons tester la généralité de ces résultats en échantillonnant les amonts et aval de 50 stations d'épuration du bassin du Rhône le long d'un gradient d'intermittence (projet ADNe-STEU financé par l'OFB). La question du signal local versus global sera testée à partir d'une approche de modélisation hydraulique (partie suivante).

2.3. Dynamique spatio-temporelle des ANe

L'utilisation des ANe est limitée par notre capacité à appréhender la complexité de son cycle de vie dans un hydrosystème. Une façon classique de tester la faisabilité d'une approche est tout simplement de la tester et de l'adapter pour maximiser sa sensibilité et spécificité. On peut par exemple développer des amorces de ddPCR spécifique à un taxon (par exemple une

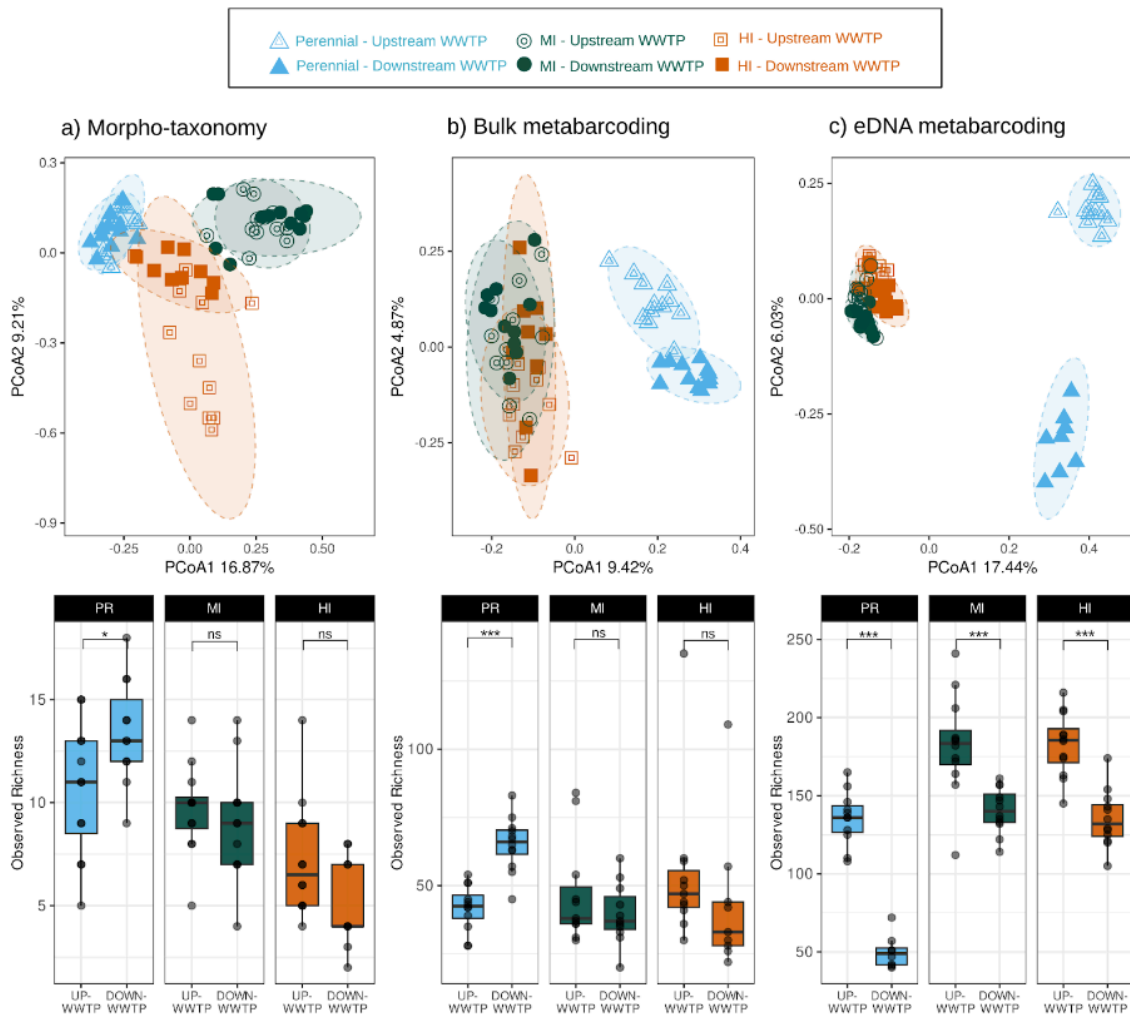


Fig. 57. – Les panneaux supérieurs montrent l’analyse en composantes principales (PCoA) basée sur les indices de dissimilarité de Jaccard de la composition de la communauté de macroinvertébrés aquatiques pour (a) la morpho-taxonomie (identification grossière), (b) le métabarcoding COI des ADN bulk (OTUs) et (c) le métabarcoding de l’ADNe d’eau filtrée (OTUs). Chaque couleur correspond à la permanence de l’écoulement du site d’échantillonnage (PR = Pérenne, MI = Intermittence moyenne et HI = Intermittence élevée). Les points vides représentent les échantillons prélevés en amont des stations d’épuration, tandis que les points pleins représentent les échantillons prélevés en aval des stations d’épuration. Les panneaux inférieurs représentent la richesse observée (richesse en taxons et en OTUs) des macroinvertébrés obtenue avec les trois méthodes d’échantillonnage, en amont et en aval des stations d’épuration (WWTP) et en fonction de la permanence du débit. Tiré de Verdier et al in prep.

espèce invasive ou patrimoniale, un pathogène...), et mesurer la capacité de cette méthode à détecter ce taxon sur des échantillons d’ANE dans des contextes où l’on connaît l’occurrence du taxon. En fonction des résultats, on pourra valider la méthode ou la rejeter, mais dans les deux cas, on n’aura pas appris grand chose de la dynamique des ANe de ce taxon. Cette approche est très fréquente depuis les premiers travaux de 2008 sur les ADNe, elle a permis le développement de nombreux outils, mais elle est limitée dans son application : si l’on change de taxon, d’environnement, il faut alors tout recommencer. Une autre approche de la question complexe de la dynamique spatio-temporelle des ANe est celle de la modélisation.

Plusieurs équipes ont déjà commencé à modéliser le cycle de vie des ADNe dans un hydrosystème (Wilcox et al. 2016; Shogren et al. 2017; Carraro et al. 2023; Pont 2024). Ces modèles utilisent plusieurs processus:

- un processus d'émission des ADNe : les ADNe sont émis par les organismes en fonction de paramètres comme la densité d'organismes et suivant une distribution de taille de particules.
- un processus de dégradation : les ADNe vont se dégrader avec le temps et à une vitesse qui dépend de paramètres comme la température.
- un processus de déplacement/sédimentation et remobilisation : les ADNe vont se déplacer ou se déposer selon des paramètres hydrauliques, en particulier la vitesse du courant.

A partir d'un modèle de ce type, sachant tous les paramètres contrôlant les processus décrits plus haut, on peut alors simuler la dynamique d'ADNe dans un hydrosystème et prédire par exemple la distance de dispersion moyenne ou bien la probabilité de détection d'un ADNe dans un hydrosystème. Un élément essentiel est la modélisation des vitesses de courant, qui nécessite en amont un modèle hydrologique pour la bassin étudié (prédire les débits) et un modèle hydraulique (prédire les vitesses de courant).

Au sein du projet Aliquot, dans le cadre d'une thèse, nous allons utiliser ce type de modélisation pour :

1. Avant même de confronter le modèle à des données empiriques, tester l'influence de différents paramètres avec une analyse de sensibilité sur un bassin comme l'Albarine.
2. Optimiser les paramètres du modèle en injectant des ANe et en mesurant en aval leurs abondances. L'idée est d'injecter différents types d'ADNe comme des ADN synthétiques ou des ADNe issus de matières organiques ou d'organismes exogènes (par exemple en injectant de l'eau issue de mésocosmes où est maintenu un élevage de taxons absents de la région).

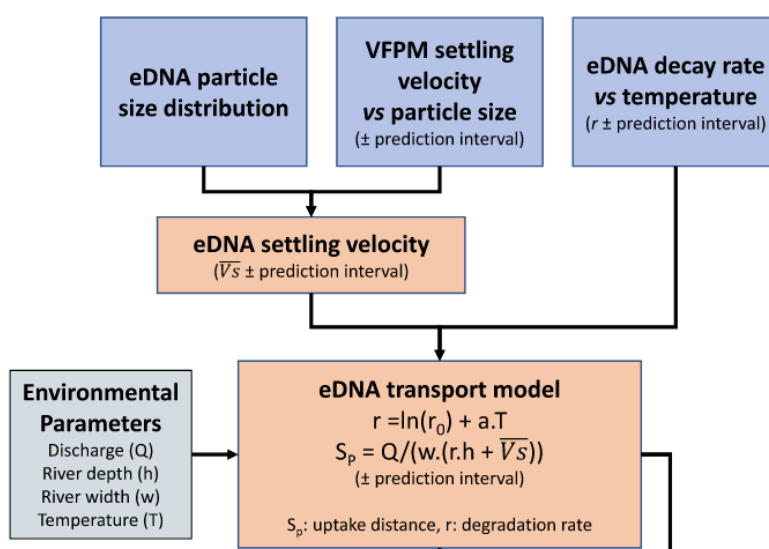


Fig. 58. – Modèle de transport de l'ADNe de Pont (2024). Le modèle contient trois processus : émission (quantité et distribution de taille des ADNe), dispersion (vitesse de sédimentation) et dégradation (en fonction de la température). Les trois processus sont combinés dans un modèle de transport qui permet de décrire la distance moyenne pour que 63% des particules aient disparu (uptake distance, S_p). Tiré de Pont (2024).

3. Pour un bassin, comme l'Albarine, dont on connaît déjà bien la distribution de diversité en « macroinvertébrés » benthiques (par des approches traditionnelles ou d'ADN-bulk) et pour lequel on possède déjà des modèles hydrologiques (Mimeau et al. 2024) et hydrauliques (Morel et al. 2020), confronter les sorties du modèle à un échantillonnage d'ADNe à l'échelle du bassin. La même analyse pourra être faite à différentes saisons afin de tester l'influence des variations de débit. Nous pourrions aborder la question de l'influence des amonts sur la diversité des ANe des avals, en utilisant des paramètres globaux comme la part relative de processus de « turn-over » et de « nestedness » (Baselga 2010) dans les communautés décrites par des ADNe.

3. La direction de recherche et l'encadrement de doctorant·e

« L'habilitation à diriger des recherches sanctionne la reconnaissance du haut niveau scientifique du candidat, du caractère original de sa démarche dans un domaine de la science, de son aptitude à maîtriser une stratégie de recherche dans un domaine scientifique ou technologique suffisamment large et de sa capacité à encadrer de jeunes chercheurs. »

– [Arrêté du 23 novembre 1988 relatif à l'habilitation à diriger des recherches, article 1](#)

Il y aurait sans doute une étude intéressante à faire sur l'histoire et l'évolution de l'HDR depuis sa mise en place en 1984 par la loi Savari pour remplacer la thèse d'état, jusque sa fonction et son usage actuel par nos administrations académiques²¹. L'HDR autorise avant tout l'accès au statut officiel de directeur de thèse. Jusqu'à peu mon école doctorale autorisait la co-direction de thèse sans HDR, c'est aujourd'hui fini, et je me retrouve donc dans l'impossibilité de démarrer des projets de recherche pourtant déjà financés. J'ai bien du mal à trouver des aspects positifs à cet exercice : une tâche supplémentaire dans la liste toujours plus longue des tâches d'un enseignant-chercheur, un exercice abscons (prouver qu'on est en capacité de faire quelque chose en l'ayant déjà fait pour avoir le droit de le (re)faire), une relique d'une structure académique encore empreinte de mandarinat, une énième étape de reconnaissance par les pairs de l'excellence de notre travail, un frein supplémentaire à l'égalité homme-femme dans la recherche... C'est aussi un exercice qui souvent rate sa cible : diriger une thèse nécessite plus de compétences sociales, psychologiques, d'empathie et de patience que d'excellence scientifique. Je reconnais qu'écrire cela dans mon HDR est un peu hypocrite. Il y a néanmoins un point positif à avoir retardé cet exercice le plus tard possible : j'ai déjà une petite expérience de direction et de suivi de thèses en tant que co-directeur ou tuteur. Toutes les thèses sont différentes mais j'en tire peut-être quand même quelques enseignements. Je voudrais tout d'abord dire ce qui me semble une évidence : la thèse est un dispositif d'apprentissage par la recherche autour d'un projet et au sein d'une équipe qui est efficace et essentiel. Cela dit, il me semble souffrir de plusieurs défauts.

La thèse est souvent une expérience éprouvante et difficile. C'est souvent la première expérience professionnelle dans un environnement où les codes du travail sont parfois flous. C'est un travail long et exigeant, multi-tâches, dont les résultats se font attendre et avec des temps de publication (surtout pour les sciences expérimentales) peu compatibles avec une thèse en 3 ans. Une étude de 2018 sur plus de 2000 doctorants montre que 41% des étudiant·e·s présentent des signes d'anxiété modérée à aiguë, les doctorants ont six fois plus de risques d'éprouver

²¹Le nombre d'HDR intervient dans le montant des financements récurrents, le nombre de bourses de thèse par école doctorale, la possibilité de promotion aux fonctions de professeur...

des troubles anxigènes ou dépressifs (Evans et al. 2018). L'augmentation de ces troubles lors des dernières décennies est débattue, le phénomène étant compliqué à démêler du meilleur diagnostic et de la baisse probable de l'auto-censure de la part des étudiants (Eleftheriades et al. 2020). Sur 300 doctorants australiens, les troubles dépressifs, anxigènes et les risques suicidaires sont fréquemment associés au syndrome de l'imposteur, à de l'auto-dénigrement et de la solitude (Mills et al. 2024). Le problème touche tous les pays dont la France : sur 136 doctorants de l'UCBL, entre 42% et 56% montrent des niveaux anormaux de stress, dépression et anxiété (Marais et al. 2018). Néanmoins dans une enquête de 2021 par le réseau national des collèges doctoraux portant sur 11545 doctorants en France, la majorité des doctorants français expriment un haut niveau de satisfaction de leur expérience doctorale (Pommier et al. 2022). On peut néanmoins regretter que cette étude organisée par nos écoles doctorales ne questionne pas les doctorants sur leur vécu psychologique pendant la thèse. La seule information sur le ressenti des doctorants est sur leur avenir professionnel : un tiers est inquiet vis-à-vis de leur avenir professionnel. Nos administrations seraient-elle en train d'éviter la question de la santé mentale de nos doctorants ?

Un incubateur de la culture du « publish or perish ». Le fonctionnement de la recherche, et donc le système dans lequel évoluent les doctorants, est probablement en grande partie responsable des difficultés décrites au dessus. En quelques décennies la recherche s'est accélérée pour pouvoir être qualifiée de « fast-science » comme on parle de fast-food (Alleva 2006). Il est difficile de trouver quelques avantages à la fast-science (Frith 2020) : une quantité exponentielle d'articles qu'on ne pourra pas lire, un fonctionnement proche de celui d'un marché financier avec des thématiques bulles et des crash, la baisse de la santé mentale des chercheurs, de fréquents arrêts de carrière, la baisse de la diversité des profils des chercheurs, la crise de la reproductibilité... Certains auteurs prédisent que le modèle actuel, s'il n'est pas rapidement



*"Surely you were aware when you accepted the position,
Professor, that it was publish or perish."*

Fig. 59. – Première utilisation de l'expression "Publish or perish" par Coolidge, professeur d'histoire à Harvard, en 1932. Tiré de Rawat & Meena (2014).

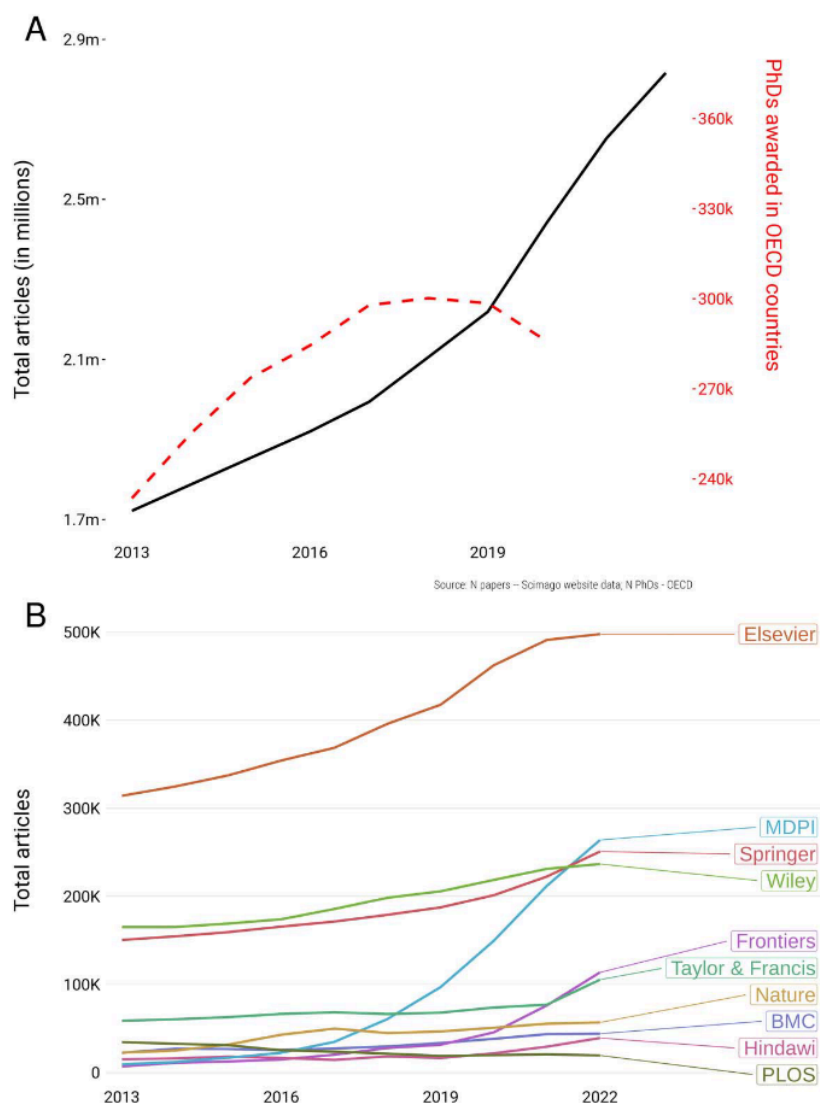


Fig. 60. – Augmentation du nombre d’articles publiés entre 2013 et 2022 (ligne pleine) alors que le nombre de thèses se stabilise (ligne pointillée, panel A). L’augmentation du nombre d’articles touche la plupart des éditeurs, mais des nouveaux éditeurs comme MDPI et Frontiers montrent les plus fortes dynamiques (panel B), en particulier en augmentant considérablement le nombre d’articles publiés dans des « special issues ». Tiré de Hanson et al. (2024).

transformé, conduira à l’augmentation de comportements non-éthiques, à la corruption du système et la perte de confiance du public (Edwards & Roy 2017).

La reconnaissance de la qualité de la carrière des chercheurs, et donc leur capacité de promotion et d’obtention de financements d’excellence, par le nombre de publications ou de citations, le « publish or perish » (Fig. 59), a fondamentalement bouleversé notre système. Notre système de publication scientifique est sous pression avec toujours plus d’articles publiés sans augmentation du nombre de chercheurs (47% d’augmentation en 2022 comparé à 2016), avec une part conséquente de nouveaux éditeurs avec des pratiques questionnables (Fig. 60, Hanson et al. (2024)). Cette accélération profite surtout à une petite « élite » de chercheurs : le 1% des chercheurs les plus cités concentrent 21% des citations, une concentration qui ne fait que de s’accroître et qui touche particulièrement la recherche européenne (Nielsen & Andersen 2021). Alors que le nombre de publications augmente, leur qualité n’augmente pas

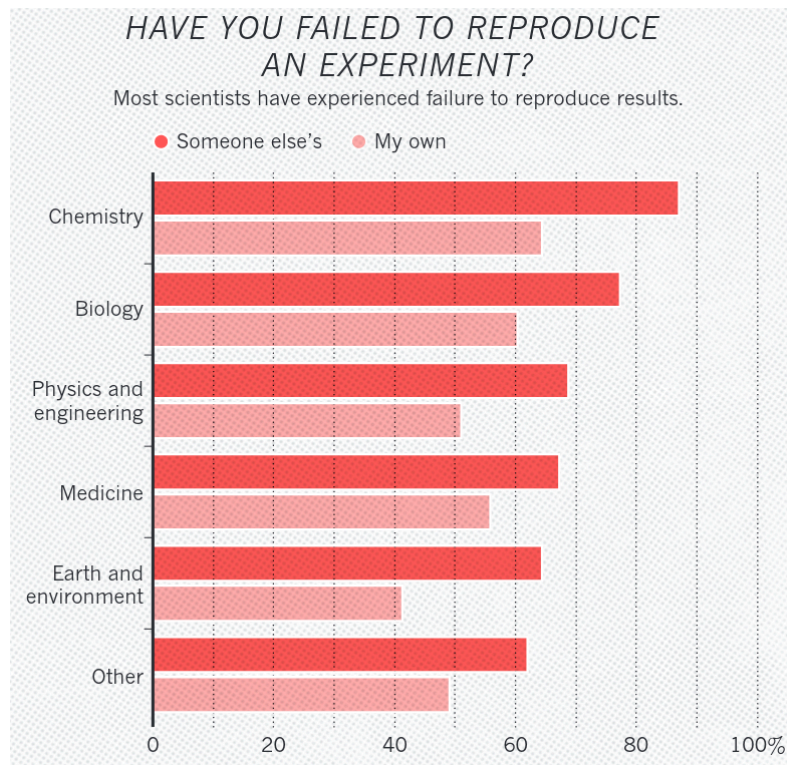


Fig. 61. – Résultat d’une enquête sur la reproductibilité des résultats scientifiques auprès de 1576 chercheurs.
Tiré de Baker (2016).

et souffre d’une crise de reproductibilité. Sur 354 articles en écologie, 63% exagèrent les effets observés (Kimmel et al. 2023). Sur 807 chercheurs en écologie et évolution, entre 40% et 64% reportent des pratiques de recherche questionnables (choix des résultats présentés en fonction de leur significativité, choix des statistiques, définition des hypothèses a posteriori... Fraser et al. (2018)). Plus de 70% des chercheurs d’une enquête reconnaissent avoir essayé sans succès de reproduire les résultats d’une étude (Fig. 61, Baker (2016)).

Un exercice de modélisation montre que l’avancement de carrière basé principalement sur le nombre de publications conduit à l’augmentation de faux positifs. Le problème est structurel et émerge de lui-même, car le système sélectionne les pratiques de fast-science (Smaldino & McElreath 2016). Enfin, notre système de financements peu nombreux et compétitifs est contre-productif : il conduit à la baisse de rendement (mesuré par le nombre de citations par exemple), la baisse de l’impact scientifique des recherches (plus forte probabilité d’échec, plus faible diversité des thématiques scientifiques abordées, perte de créativité, système moins résilient et moins évolvable), la perte d’un tissu scientifique diversifié, l’augmentation de la compétition entre chercheurs et thématiques, et la diminution de l’évolvabilité des structures de recherche (Aagaard et al. 2020). Nos responsables d’institution de la recherche française ne semblent néanmoins pas à remettre en cause ce modèle de fast-science (Fauvelle 2019)...

Nous avons probablement transféré ce modèle de fast-science sur les doctorants, ce qui peut générer une vision désabusée de la recherche scientifique, une perte de sens et un cadre de travail parfois toxique (compétition, stress, dévalorisation, surmenage...). Alors que la science se définit comme la recherche de la vérité dans un grand effort collectif, un peu de recul montre surtout que nous sommes souvent poussés à la recherche de reconnaissance et de financements.

Comment adapter son encadrement. Beaucoup de ces problèmes sont structureaux ou conjecturaux, un encadrant de thèse peut sembler bien désarmé. Il me semble que nous avons néanmoins quelques marges de manœuvre :

- Il est mieux d'encadrer à plusieurs : meilleure écoute, plus de disponibilité (surtout pour les enseignants-chercheurs), plus de professionnalisme. Attention néanmoins à ne pas diluer les responsabilités et se retrouver dans une thèse sans direction claire.
- Aller vers un modèle qui reste exigeant mais qui est bienveillant et protège la santé mentale des doctorants. Il n'y a pas de solution miracle : les attendus de chacun doivent être clairement énoncés en début de thèse, il faut parfois en faire moins qu'espéré (et arrêter la course aux publications), et adapter le contenu de la thèse en fonction des difficultés et avancées. Il faut aussi peut-être encadrer peu d'étudiants.
- Pour les thèses avec beaucoup de développements ou d'expérimentations, trouver une solution de financement pour rajouter 6 à 12 mois à la thèse. On finit souvent par devoir bricoler une solution en fin de thèse, cela est stressant pour tout le monde. Nous devrions pouvoir inclure 6 à 12 mois de salaire supplémentaire dans nos budgets de projet sans que cela pose problème.
- Il ne faut pas appliquer le modèle que l'on a subi : la course à la découverte, à la publication dans des magazines scientifiques (Nature, Science, ...), aux rendements, mais replacer les valeurs académiques (vérité, équité, probité, connaissances, travail collaboratif...) au centre du travail de thèse.
- La thèse ne forme pas qu'à la recherche, elle reste une expérience enrichissante pour plein d'autres métiers, arrêtons d'en faire une expérience traumatisante en projetant nos fantasmes de réussite scientifique. En sciences du vivant, les doctorants sont de moins en moins embauchés dans l'académique (59% en 2016, 50% en 2018). La part du secteur public hors recherche et la part du privé hors recherche augmente : après leurs thèses les docteurs font de moins en moins de recherche - la thèse est de facto devenue une formation généraliste (Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation 2021).
- Il doit devenir possible d'arrêter une thèse sans que cela soit un échec. Si un doctorant n'est pas en capacité de conduire des recherches ou de rédiger sa thèse, ce n'est pas aux encadrants de le faire à sa place. Ce sujet est aujourd'hui tabou, mais il ne devrait pas l'être. Des mécanismes d'accompagnement et de réorientation (passage d'un statut doctorant à ingénieur par exemple) devraient sans doute être réfléchis et proposés pendant la thèse sans que la situation ne se transforme en drame pour toutes les parties impliquées (doctorant, encadrant, financeur, école doctorale).
- Il n'est pas cohérent de demander à la fois des publications acceptées à la soumission du manuscrit de thèse et de ne pas autoriser à dépasser 3 ans de thèse. Autrement dit, l'objectif de la thèse doit redevenir le manuscrit de thèse, pas les publications.

Je réalise que ce récit est un peu sombre et très critique de nos pratiques de travail. Le travail avec des doctorants reste pour moi une expérience très enrichissante et très positive. Imaginer à plusieurs une expérience, la construire, l'analyser et la coucher sur le papier, reste une rare expérience de co-construction artisanale et intellectuelle que je trouve particulièrement

émouvante. Accompagner un ou une jeune chercheur·se dans cette démarche est sans doute la meilleure des façons de partager et apprendre notre métier.

4. L'enseignement par la conduite de projets

Depuis 2010, j'essaie de développer à l'UCBL des enseignements et des filières où l'apprentissage est moins encyclopédique et désincarné. Beaucoup d'étudiants ne sont pas réceptifs à l'enseignement pyramidal : perte de motivation, savoirs éphémères, peu de gain de compétences... Cela reste néanmoins l'enseignement dominant à l'université (en grande partie car c'est le moins cher). Il y a plusieurs approches pour casser cette dynamique (cours inversés, débats, ...), pour ma part j'essaie de développer des approches par projets où l'étudiant devient acteur de son apprentissage. Je vais développer quelques enseignements par la suite, mais vous pourrez trouver un ensemble de fascicules de TP ici : <https://forge.univ-lyon1.fr/tristan.lefebure/cours>.

TP - Outils Moléculaires pour l'Ecologie et l'Evolution de L3.

Peu de temps après mon arrivée à l'université, je monte des TP longs d'écologie moléculaire (30h de TP, UE Outils moléculaires pour l'écologie et l'évolution) qui consistent à identifier moléculairement des arthropodes et génotyper une population de tritons alpestres du campus. Les deux projets sont menés de front par les étudiants et consistent à :

1. échantillonner et décrire les organismes
2. procéder aux extractions d'ADN et aux amplifications, et envoyer les amplicons pour analyses (Fig. 62).
3. analyser les chromatogrammes, effectuer l'identification moléculaire, génotyper les organismes et calculer les fréquences alléliques et les indices de fixation
4. tenir un cahier de laboratoire pour l'ensemble des analyses



Fig. 62. – TP de barcoding et de génotypage des étudiants de L3 parcours Sciences de la Biodiversité. Les étudiants en binôme gèrent leur TP sur plusieurs jours et tiennent un cahier de laboratoire.

TP - Technologie d'Acquisition de Données de M1.

Suite aux bouleversements technologiques des dernières décennies et vu l'inculture des étudiants de biologie de Master 1 face à ces techniques, je monte en 2017²² une nouvelle UE méthodologique pour trois masters de l'UCBL (BEE, Biologie végétale et Microbiologie). L'UE se compose principalement de TP d'une quinzaine d'heures. Le choix des thématiques est forcément un peu arbitraire, mais ce sera :

1. Nouvelles techniques de séquençage et application en métabarcoding: analyse de la communauté de rhizosphère de plantes (échantillonnage, extraction d'ADN, amplification, et analyse de données).
2. Analyse de niche trophique par isotopes stables et marquage ADN-SIP
3. Quantification de biomasse microbienne par PCR quantitative
4. Etude du paysage par Système d'Information Géographique

Parcours de master en Génomique Environnementale.

En master, j'ai participé à la construction d'un nouveau parcours de Génomique Environnementale afin de répondre à une triple demande :

- le manque de personnes compétentes dans cette thématique qui croise écologie, biologie moléculaire et bioinformatique.
- une forte demande du secteur public et privé
- l'absence de formation de ce type en France

La master est construit autour d'un projet de recherche mené par les étudiants et en collaboration avec le CREA Mont-Blanc qui est une ONG scientifique dont la mission est d'explorer l'impact du changement climatique sur la biodiversité et de partager ces connaissances avec les décideurs et les citoyens. Globalement il s'agit de savoir si des échantillons d'ADNe prélevés dans un cours d'eau permettent d'inventorier la biodiversité terrestre d'une vallée alpine. En parallèle, les étudiants apprennent à séquencer un génome, à l'assembler et l'analyser (la promotion 2024-2026 étudie la cardamine des Alpes). Le détail des questions abordées par les étudiants est défini pendant le master. Le projet inclut (Fig. 64) :

1. en tout début de M1, l'échantillonnage des ADNe d'un torrent de montagne à Vallorcine accompagné du CREA et de scientifiques spécialistes d'écologie alpine (2 jours de terrain, Fig. 63). Des organismes d'intérêts sont aussi prélevés pour le séquençage de leur génome.
2. l'extraction des ADNe, la préparation de librairies de métabarcoding Illumina, et la préparation d'une librairie d'ADN HMW pour séquençage Nanopore (UE Biologie moléculaire pour la génomique environnementale, 30h)
3. l'analyse bioinformatique des lectures de séquençage et les analyses statistiques (UE Atelier d'analyse en génomique environnementale, 60h)
4. l'écriture d'un article et sa présentation

La première promotion du parcours GE a donné lieu à la présentation de poster à deux conférences scientifiques (International Environmental and Agronomical Genomics symposium 2024, SFE2 2024). Le dépôt d'un premier manuscrit est en cours de préparation.

²²Je ne suis plus responsable de l'UE depuis la rentrée 2023



Fig. 63. – Echantillonnage à Vallorcine d’ADNe pour la promotion 2022-2024.

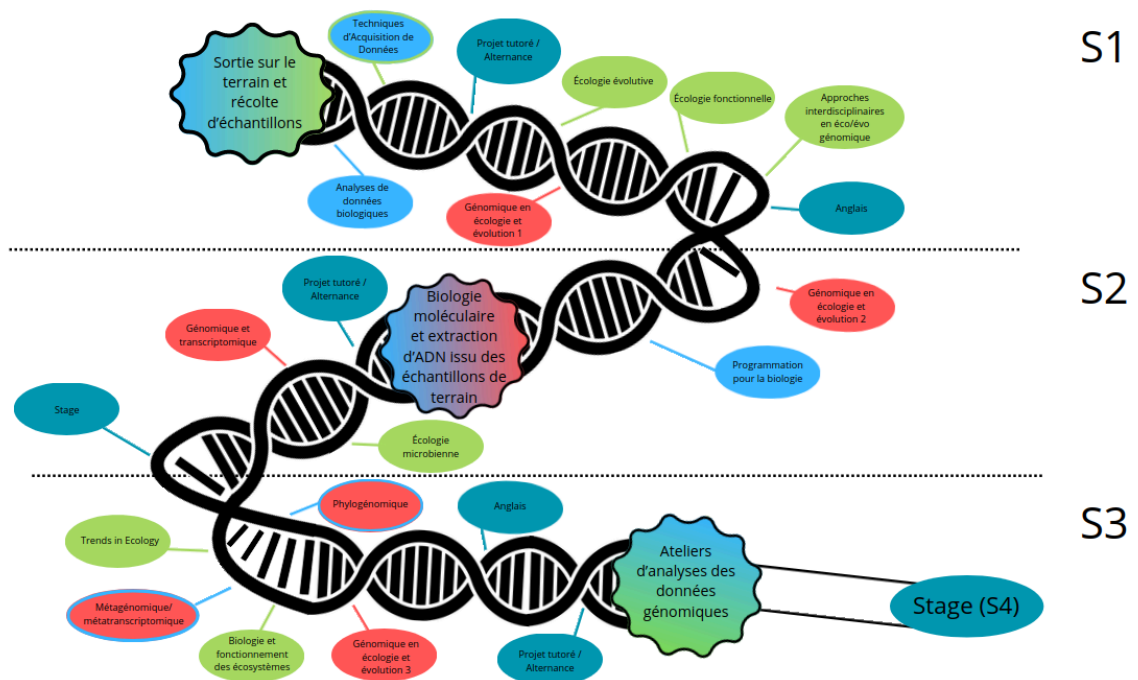


Fig. 64. – Maquette du parcours de master en Génomique Environnementale. <https://master-bee.univ-lyon1.fr/gelyon/>

5. CV détaillé

- **Fonction** : Maître de conférence à l'Université Claude Bernard Lyon 1
- **Laboratoire** : Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés (LEHNA, UMR 5023), Université Claude Bernard Lyon 1, CNRS, ENTPE
- **Equipe de recherche** : Ecologie, Evolution, Ecosystèmes souterrains (E3S)
- **HAL** : [IDHal tristan-lefebure](https://hal.archives-ouvertes.fr/tristan-lefebure)
- **ORCID** : [0000-0003-3923-8166](https://orcid.org/0000-0003-3923-8166)

Responsabilités administratives

- co-responsable du parcours de **Licence Sciences de la Biodiversité**, UCBL, depuis 2021
- co-responsable du parcours de **Master en Génomique Environnementale**, UCBL, depuis 2022
- co-directeur de la plateforme du **DTAMB** (Développement de Techniques et Analyse Moléculaire de la Biodiversité), FR BioEEnViS, UCBL, depuis 2021

Parcours académique

- **Maître de conférence** (2010 – présent), Université Lyon 1, France
- **Chercheur associé** (2009 – 2010), Cornell University, USA
- **Chercheur en post-doctorat** (2006 – 2008), Cornell University, USA
- **Doctorant** allocataire de recherche (2001 – 2005), Université Lyon 1, France

Formation universitaire

- **Doctorat**, écologie évolutive, Université Lyon 1, 2005
 - directeurs de recherche : Janine Gibert & Christophe Douady
 - titre : « Origine, évolution et mesure de la biodiversité des eaux souterraines : analyse moléculaire du genre *Niphargus* (Crustacea) »
 - soutenance : 7 octobre 2005 à Villeurbanne (Fr), n°151-2005

- **DEA**, analyse et modélisation des systèmes biologiques, Université Lyon 1, 2001
 - directeur de recherche : Janine Gibert
 - titre: « Utilisation de l’outil moléculaire pour estimer la biodiversité souterraine cachée : premiers résultats chez le Crustacé Amphipode *Niphargus virei* »
- **Maîtrise**, biologie des populations et des écosystèmes. Université Lyon 1, 2000.
- **Licence**, biologie des organismes et des populations. Université Lyon 1, 1999.
- **DEUG**, biologie. Université Lyon 1, 1998.

Projets de recherche

Coordinateur

- **Stoichiogenomics** : Déterminisme trophique de l’évolution des génomes, 15k€ (2012, n°70632, Appel à projets en génomique environnementale, CNRS)
- **ExoSout** : Modèle souterrain pour l’étude du vieillissement en milieu naturel, 17k€ (2014, PEPS ExoMod, CNRS)
- **CAPTAS** : Captation d’ADN souterrain, 80k€ (2017, X-LIFE, CNRS).
- **AmphibGeneFam** : Développement d’un pipeline de reconstruction de familles de gènes homologues à partir de transcriptomes de novo d’amphibiens, 5k€ (2017, FR BioEnviS).
- **Capteur à ADN** : Biomonitoring des habitats aquatiques complexes par capteurs à ADN environnementaux. Contrat de collaboration accompagnement CIFRE (2019–2022, budget de fonctionnement 79k€, Eurofins Hydrobiologie en sus de la subvention ANRT).
- **MISTRAL**: Metabarcoding pour le Suivi des opérations de Restauration de l’Albarine, 40k€ (2020-2022, ZABR / Agence de l’eau).
- **NeGA** : Influence of effective population size (Ne) on animal Genome Architecture, 575k€ (2021-2025, ANR AAPG2020).
- **ALIQUOT** : Acides nucléiques, molécules sentinelles des écosystèmes aquatiques continentaux, 1.5M€ (2024–2028, PEPR OneWater, France2030).
- **ADNe-STEU** : Apport de l’ADNe dans la détection des effets des rejets de station de traitement des eaux usées (STEU) en cours d’eau, 333k€ (2025–2027, Contrat R&D OFB, n°OFB-24-1128).

Porteur de tâche

- **Stygomics** : Evolution génomique et transcriptomique sous environnement ionisant (2014, PEPS Interdisciplinaires, CNRS)
- **Convergenomix**: Integrative study of convergent genomic evolution across three animal groups (2016–2019, ANR, ANR-15-CE32-0005)

Partenaire

- projet européen **PASCALIS** : Protocols for the ASsessment and Conservation of Aquatic Life In the Subsurface (2001–2004, n°EESD-2000-2.2.3)
- NIH contract Zoonoses Research Unit **Campylobacter** : Food & Waterborne Diseases Integrated Research Network (2006–2010, n°N01-AI-30054 (ZC003-05))

- NIH grant **Streptococcus** : Evolutionary genomics and population genetics of pathogenic streptococci (2008–2013, n°AI073368-01A2)
- Projet de coopération scientifique et universitaire franco-danoise **Next Generation Biological Models** (NGBM) : Promoting subterranean organisms as models for environmental and health Sciences (2011)
- **MINIGAN** : Mito-Nuclear Interactions and Gynodioecy in ANimals (2019–2023, ANR, AAPG2019)
- **ARTEMIS** : Les ADN environnementaux pour la détection et le suivi des populations de moustiques tigres en milieu urbain (2022, FR BioEEnViS).
- **Capteur d'ADNe** : mesurer facilement la biodiversité dans les milieux renaturés (2021–2025, Institut Carnot Eau & Environnement)
- **FUNCEE** : Des gènes fonctionnels au fonctionnement des écosystèmes : vers une nouvelle vision intégrative des liens biodiversité-fonction dans les hydrosystèmes (2023–2024, CNRS EC2CO)

Encadrements

- **Doctorat**
 - *Claire Morvan* (co-direction, E2M2, 2010–2013): Approche phylogénétique pour l'étude des patrons macroévolutifs et processus sous-jacents dans les eaux souterraines
 - *Clémentine Francois* (co-direction, E2M2, 2011–2014): Le déterminisme trophique de l'évolution des génomes
 - *Nathanaëlle Saclier* (co-direction, E2M2, 2016–2019): Un modèle génomique en milieu souterrain pour comprendre les déterminants des variations de vitesse d'évolution moléculaire.
 - *Mailys Gauthier* (co-encadrant, E2M2, 2017–2020) : Dynamique des métacommunautés dans les têtes de bassin versant fragmentées : une perspective moléculaire.
 - *Héloïse Verdier* (co-direction, E2M2, 2019–2022): Biomonitoring des habitats aquatiques complexes par capteurs à ADN environnementaux.
 - *Florian Bénitère* (co-direction, E2M2, 2021–2024): Impact de la dérive génétique sur l'évolution de l'architecture des gènes et de la complexité du transcriptome.
- **Master** : Pierre Charrier (2013), Benjamin Benti (2014), Damien Maury (2014), Nathanaëlle Saclier (2016), Juliette Berger (2018), Florian Petrilli (2023), Uliana Podgourskaia (2023), Florian Blanchard (2024), Ibtissam Dich (2025).
- Participation à des **comités de pilotage de thèse**: David Eme, Florent Lassale, Marie Cariou, Carine Rey, Laëtitia Villiers, Morgane Touzot, Rémi Allio, Chloé Haberkorn, Mathilde Barthe, Eliette Reboud.
- **Tuteur de thèse** : Eugénie Pessia, Yann Lesèque, Magalie Semeria, Michel Lecocq, Marlène Roy, Samuel Barreto, Wandrille Duchemin, Adrián Davín, Monique Aouad, Cecilia Banho, Maud Gautier, Camille Sessegolo, Djivan Prentout, Théo Tricou, Vincent Lhuillier, Théo-

phile Franzino, Hugo Menet, Benjamin Guinet, Alexia Nguyen Trung, Alice Genestier, Julien Joseph.

Enseignements

- **Responsable d'UE :**
 - Outils moléculaires pour l'écologie et l'évolution, L3, 70 étudiants, depuis 2012
 - Ateliers d'analyses en génomique environnementale, M2, 15 étudiants, depuis 2023
- **Enseignants en :**
 - Génomique évolutive (CM et TP), UCBL, Master, 2016–present
 - Génomique environnementale (CM et TP), UCBL, 2016–present
 - Bioinformatique (TP), UCBL, Master, 2016–present
 - Evolution (TD, TP), UCBL, L3, 2010–présent
 - Ecologie moléculaire (TD, TP), UCBL, L3, 2012–présent
 - Zoologie (TP), UCBL, L2 biologie, 2011–présent
- Fascicule de TP : <https://forge.univ-lyon1.fr/tristan.lefebure/cours>

Activité d'expertise

- **Comités de sélection :**
 - Maître de conférence ENS, IGFL, 2012
 - Maître de conférence Univ. Lyon 1, LBBE, 2014
 - Maître de conférence Univ. Lyon 1, LBBE, 2017
 - Maître de conférence Univ. Lyon 1, LEHNA, 2017
 - Maître de conférence Univ. Paris Sud, EGCE, 2018
 - Maître de conférence Univ. Grenoble Alpes, LECA, 2018
 - Ingénieur d'étude, ENS de Lyon, 2021
 - Maître de conférence Univ. Lyon 1, LEHNA, 2022
- **Référent** pour les journaux : Biology Letters, Briefings in Bioinformatics, BMC Evolutionary Biology, Genome Biology and Evolution, Genome Research, Journal of Zoological Systematics and Evolutionary Research, Knowledge and Management of Aquatic Ecosystems, Molecular Biology and Evolution, Molecular Ecology, Molecular Phylogenetics and Evolution, PLoS Genetics, PLoS One, ZooKeys, Zoologischer Anzeiger
- **Jury :**
 - thèse de Vincent Mérel, 2021, Univ Lyon 1
 - thèse de Juliette Berger, 2024, MNHN

- HDR de Benoit Nabholz, 2019, Univ Montpellier
- **Expertise de projet :**
 - AAP Chercheur(se)s d'avenir 2013, Région Languedoc-Roussillon.
 - AAP Labex OSUG 2010
 - AAP 2019 du Swiss National Science Foundation
 - AAP de l'Institut polaire français Paul-Emile Victor (IPEV) 2020

Publications

48. Allio R, Teullet S, Lutgen D, Magdeleine A, Koual R, Tilak MK, de Thoisy B, Emerling C, Lefébure T, Delsuc F. **2025**. Transcriptomic data reveal divergent paths of chitinase evolution underlying dietary convergence in anteaters and pangolins. *Genome Biology and Evolution*, [10.1093/gbe/evaf002](https://doi.org/10.1093/gbe/evaf002).
47. Verdier H, Datry T, Logez M, Konecny-Duprè L, Gauthier M, Lefébure T. **2024**. Environmental DNA particle size distribution and quantity differ across taxa and organelles. *Environmental DNA*, [6:e598](https://doi.org/10.1093/envdna/6/e598).
46. Laugier F, Saclier N, Béthune K, Braun A, Konecny L, Lefébure T, Luquet E, Plénet S, Romiguier J, David P. **2024**. Both nuclear and cytoplasmic polymorphisms are involved in genetic conflicts over male fertility in the gynodioecious snail *Physa acuta*. *Evolution*, [10.1093/evolut/qpae053](https://doi.org/10.1093/evolut/qpae053).
45. Saclier N, Duchemin L, Konecny-Duprè L, Grison P, Eme D, Martin C, Callou C, Lefébure T, François C, Issartel C, et al. **2024**. A collaborative backbone resource for comparative studies of subterranean evolution: The World Asellidae database. *Molecular Ecology Resources*, [24:e13882](https://doi.org/10.1111/1365-3113.12382).
44. Casane D, Saclier N, Policarpo M, François C, Lefébure T. **2023**. Evolutionary genomics and transcriptomics in groundwater animals. In: Malard F, Griebler C, Rétaux S, editors. *Groundwater Ecology and Evolution (Second Edition)*. San Diego: Academic Press. [393–412](https://doi.org/10.1016/B978-0-12-820333-3.00041-2).
43. Malard F, Verdier H, Konecny-Duprè L, Eme D, Douady CJ, Lefébure T. **2022**. Subterranean species inventory in the molecular era. *Karstologia*, 79: 73–78.
42. David P, Deglétagne C, Saclier N, Jennan A, Jarne P, Plénet S, Konecny L, François C, Guéguen L, Garcia N, Lefébure T, Luquet E. **2022**. Extreme mitochondrial DNA divergence underlies genetic conflict over sex-determination. *Current Biology*, [32: 2325–2333](https://doi.org/10.1016/j.cub.2022.05.033).
41. Verdier H, Konecny-Duprè L, Marquette C, Reveron H, Tadier S, Grémillard L, Barthès A, Datry T, Bouchez A, Lefébure T. **2022**. Passive sampling of environmental DNA in aquatic environments using 3D-printed hydroxyapatite samplers. *Molecular Ecology Resources*, [22\(6\): 2158–2170](https://doi.org/10.1111/1365-3113.12158).

40. Touzot M, [Lefébure T](#), Lengagne T, Secondi J, Dumet A, Konecny-Dupré L, Verber P, Navratil V, Duchamp C, Mondy N. **2021**. Transcriptome-wide deregulation of gene expression by artificial light at night in tadpoles of common toads. *Science of The Total Environment*, [818](#), [151734](#).
39. Blackman R, Altermatt F, Foulquier A, [Lefébure T](#), Gauthier M, Bouchez A, Stubbington R, Weigand AM, Leese F and Datry T. **2021** Unlocking our understanding of intermittent rivers and ephemeral streams with genomic tools. *Frontiers in Ecology and the Environment*, [19\(10\)](#): [574–583](#).
38. Saclier N, Chardon P, Malard F, Konecny-Dupré L, Eme D, Bellec A, Breton V, Duret L, [Lefébure T](#), Douady CJ. **2020** Bedrock radioactivity influences the rate and spectrum of mutation. *eLife*, [9](#):[e56830](#).
37. Francois C, Laurent S, Malard F, [Lefébure T](#), Douady C, Mermillod-Blondin F. **2020**. Trophic selectivity in aquatic isopods increases with the availability of resources. *Functional Ecology*, [34](#): [1078--1090](#).
36. Malard F, Grison P, Duchemin L, Konecny-Dupré L, [Lefébure T](#), Saclier N, Eme D, Martin C, Callou C, Douady CJ. **2020**. GOTIT: a laboratory application software for optimizing multi-criteria species based research. *Methods in Ecology and Evolution*, [11](#): [159–167](#).
35. Gauthier M, Konecny-Dupré L, Nguyen A, Elbrecht V, Datry T, Douady CJ, [Lefébure T](#). **2020**. Enhancing DNA metabarcoding performance and applicability with bait capture enrichment and DNA from conservative ethanol. *Molecular Ecology Resources*, [20](#): [79–96](#).
34. Saclier N, François CF, Konecny-Dupré L, Lartillot N, Guéguen L, Duret L, Malard F, Douady CJ, [Lefébure T](#). **2018** Life history traits impact the nuclear rate of substitution but not the mitochondrial rate in isopods. *Molecular Biology and Evolution*, [35\(12\)](#): [2900–2912](#).
33. [Lefébure T](#), Morvan C, Malard M, François C, Konecny-Dupré L, Guéguen L, Weiss-Gayet M, Seguin-Orlando A, Ermini L, Der Sarkissian C, Charrier NP, Eme D, Mermillod-Blondin F, Duret L, Vieira C, Orlando L, Douady CJ. **2017**. Less effective selection leads to larger genomes. *Genome Research*, [27](#): [1016–1028](#).
32. Francois CF, Duret L, Simon L, Mermillod-Blondin F, Malard F, Konecny-Dupré L, Planel R, Penel S, Douady CJ, [Lefébure T](#). **2016**. No evidence that nitrogen limitation influences the elemental composition of isopod transcriptomes and proteomes. *Molecular Biology and Evolution*, [33\(10\)](#): [2605–2620](#).
31. Huguet-Tapia JC, [Lefébure T](#), Badger JH, Guan D, Pettis GS, Stanhope MJ, Loria R. **2016**. Genome content and phylogenomics reveal both ancestral and lateral evolutionary pathways in plant pathogenic *Streptomyces* species. *Applied Environmental Microbiology*, [82\(7\)](#): [2146–2155](#).
30. Richards VP, [Lefébure T](#), Pavinski Bitar P, Dogan B, Simpson KW, Schukken YH, Stanhope MJ. **2015**. Genome based phylogeny and comparative genomic analysis of intra-mammary pathogenic *Escherichia coli*. *PLoS ONE*, [10\(3\)](#): [e0119799](#).
29. Dogan B, Suzuki H, Herlekar D, Sartor B, Campbell BJ, Roberts CL, Stewart K, Scherl EJ, Araz Y, Pavinski Bitar P, [Lefébure T](#), Chandler B, Schukken YH, Stanhope MJ, Simpson

- KW. **2014**. Inflammation-associated adherent-invasive *E. coli* are enriched in pathways for use of propanediol and iron, and M cell translocation. *Inflammatory Bowel Diseases*, [20\(11\): 1919–1932](#).
28. Griebler C, Malard F, Lefébure T. **2014**. Current developments in groundwater ecology - from biodiversity to ecosystem functions and services. *Current Opinion in Biotechnology*, [27: 159–167](#).
27. Eme D, Malard F, Konecny-Dupré L, Lefébure T, Douady CJ. **2013**. Bayesian phylogeographic inferences reveal contrasting colonization dynamics among European groundwater isopods. *Molecular Ecology*, [22\(22\): 5685–5699](#).
26. Capderrey C, Kaufmann B, Jean P, Malard F, Konecny-Dupré L, Lefébure T, Douady CJ. **2013**. Microsatellite development and first population size estimates for the groundwater isopod *Proasellus walteri*. *PLoS ONE*, 8(9): [e76213](#).
25. Palmer SR, Miller JH, Abranches J, Zeng L, Lefébure T, Richards VP, Jose A Lemos JA, Stanhope MJ, Robert A. Burne RA. **2013**. Phenotypic heterogeneity of genomically-diverse isolates of *Streptococcus mutans*. *PLoS ONE*, [8\(4\): e61358](#).
24. Morvan C, Malard F, Paradis E, Lefébure T, Konecny-Dupré L, Douady CJ. **2013**. Timetree of Aselloidea reveals species diversification dynamics in groundwater. *Systematic Biology*, [62\(4\): 512–522](#).
23. Cornejo O, Lefébure T, Pavinski Bitar PD, Lang P, Richards PD, Eilertson K, Do T, Beighton D, Zeng L, Ahn SJ, Burne R, Siepel A, Bustamante CD, Stanhope MJ. **2013**. Evolutionary and population genomics of the cavity causing bacteria *Streptococcus mutans*. *Molecular Biology and Evolution*, [30\(4\): 881–893](#).
22. Lin C, den Bakker H, Suzuki H, Lefébure T, Ponnala L, Sun Q, Stanhope MJ, Wiedmann M, Duhamel G. **2013**. Complete genome sequence of the porcine type strain *Brachyspira pilosicoli* P43/6/78T. *Genome Announcements*, [1: e00215-12](#).
21. Richards VP, Lefébure T, Pavinski Bitar PD, Stanhope MJS. **2013**. Comparative characterization of the virulence gene clusters (lipooligosaccharide [LOS] and capsular polysaccharide [CPS]) for *Campylobacter coli*, *Campylobacter jejuni* subsp. *jejuni* and related *Campylobacter* species. *Infection, Genetics and Evolution*, [14: 200–213](#).
20. Bell, JA, Jerome JP, Plovanich-Jones AE, Smith EJ, Gettings JR, Kim HY, Landgraff JR, Lefébure T, Kopper JJ, Rathinam VA, St. Charles JL, Buffa BA, Brooks AP, Poe SA, Eaton KA, Stanhope MJS, Mansfield LS. **2013**. Outcome of infection of C57BL/6 IL-10^{-/-} mice with *Campylobacter jejuni* strains is correlated with genome content of open reading frames up-and down-regulated in vivo. *Microbial Pathogenesis*, [54: 1–19](#).
19. Richards VP, Zadoks RN, Pavinski Bitar PD, Lefébure T, Lang P, Werner B, Tikofsky L, Moroni P, Stanhope MJS. **2012**. Genome characterization and population genetic structure of the zoonotic pathogen, *Streptococcus canis*. *BMC Microbiology*, [12: 293](#).
18. Burne RA, Zeng L, Ahn SJ, Palmer S, Liu Y, Lefébure T, Stanhope MJ, Nascimento M. **2012**. Progress dissecting the oral microbiome in caries and health. *Advances in Dental Research*, [24\(2\): 77–80](#).

17. Lefébure T, Richards PV, Lang P, Pavinski Bitar PD, Stanhope MJ. **2012**. Gene repertoire evolution of *Streptococcus pyogenes* inferred from phylogenomic analysis with *Streptococcus canis* and *Streptococcus dysgalactiae*. *PLoS ONE*, [7\(5\): e37607](#).
16. Suzuki H, Lefébure T, Pavinski Bitar PD, Stanhope MJ. **2012**. Comparative genomic analysis of the genus *Staphylococcus* including *Staphylococcus aureus* and its newly described sister group *Staphylococcus simiae*. *BMC Genomics*, [13:38](#).
15. Richards VP, Lang P, Pavinski Bitar PD, Lefébure T, Schukken YH, Zadoks RN, Stanhope MJ. **2011**. Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted *Streptococcus agalactiae*. *Infection, Genetics and Evolution*, [11: 1263–1275](#).
14. Suzuki H, Lefébure T, Hubis MJ, Pavinski Bitar P, Lang P, Siepel A, Stanhope MJ. **2011**. Comparative genomic analysis of the *Streptococcus dysgalactiae* species group : gene content, molecular adaptation, and promoter evolution. *Genome Biology and Evolution*, [3: 168–185](#).
13. Lefébure T, Pavinski Bitar PD, Suzuki H, Stanhope MJ. **2010**. Evolutionary dynamics of complete *Campylobacter* pan-genomes and the bacterial species concept. *Genome Biology and Evolution*, [2: 646–655](#).
12. Lang P, Lefébure T, Wang W, Pavinski Bitar P, Meinersmann R, Kaya K, Stanhope MJ. **2010**. Expanded MLST genotyping and comparative genomic hybridization evidence for host preferred groups in *Campylobacter coli*. *Applied Environmental Microbiology*, [76\(6\): 1913–1925](#).
11. Lefébure T, Stanhope MJ. **2009**. Pervasive, genome wide positive selection, leading to functional divergence in the bacterial genus *Campylobacter*. *Genome Research*, [19: 1224–1232](#).
10. Dole-Olivier MJ, Malard F, Martin D, Lefébure T, Gibert J. **2009**. Relationships between environmental variables and groundwater biodiversity at the regional scale. *Freshwater Biology*, [54\(4\): 797–813](#).
9. Trontelj P, Douady CJ, Fišer C, Gibert J, Gorički S, Lefébure T, Sket B, Zakšek V. **2009**. A molecular test for cryptic diversity in groundwater: how large are the ranges of macrostygobionts? *Freshwater Biology*, [54\(4\): 727–744](#).
8. Lang P, Lefébure T, Wang W, Zadoks RN, Schukken Y, Stanhope MJ. **2009**. Gene content differences across strains of *Streptococcus uberis* identified using oligonucleotide microarray comparative genomic hybridization. *Infection, Genetics and Evolution*, [9: 179–188](#).
7. Stanhope MJ, Lefébure T, Walsh ST, Becker JA, Lang P, Pavinski-Bitar PD, Miller LA, Italia MJ, Amrine-Madsen H. **2008**. Positive selection in penicillin binding proteins 1a, 2b, and 2x from *Streptococcus pneumoniae* and its correlation with amoxicillin resistance development. *Infection, Genetics and Evolution*, [8\(3\): 331–339](#).
6. Foulquier A, Malard F, Lefébure T, Douady CJ, Gibert J. **2008**. The imprint of Quaternary glaciers on the present-day distribution of the obligate groundwater amphipod *Niphargus virei*. *Journal of Biogeography*, [35\(3\): 552–564](#).

5. Lefébure T, Stanhope MJ. **2007**. Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome Biology*, [8: R71](#).
4. Stanhope MJ, Walsh SL, Becker JA, Miller LA, Lefébure T, Lang P, Pavinsky Bitar PD, Amrine-Madsen H. **2007**. The relative frequency of intraspecific lateral gene transfer of penicillin binding proteins 1a, 2b, and 2x, in amoxicillin resistant *Streptococcus pneumoniae*. *Infection, Genetics and Evolution*, [7\(4\): 520–34](#).
3. Lefébure T, Douady CJ, Malard F, Gibert J. **2007**. Testing dispersal and cryptic diversity in a widely distributed groundwater amphipod (*Niphargus rhenorhodanensis*). *Molecular Phylogenetics and Evolution*, [42\(3\): 676–686](#).
2. Lefébure T, Douady CJ, Gouy M, Gibert J. **2006**. Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation. *Molecular Phylogenetics and Evolution*, [40\(2\): 435–447](#).
1. Lefébure T, Douady CJ, Gouy M, Trontelj P, Briolay J, Gibert J. **2006**. Phylogeography of a subterranean amphipod reveals cryptic diversity and dynamic evolution in extreme environments, *Molecular Ecology*, [15\(7\): 1797–806](#).

6. Remerciements

Je n'ai évidemment pas fait tout ce travail tout seul.

Le **modèle Asellidae** est porté par le duo infernal et inépuisable que forment Florian Malard et Christophe Douady (terrain, échantillonnage, taxonomie, phylogénie, morphométrie...). Toute la biologie moléculaire repose sur une personne incroyable : Lara Konecny. J'ai eu la chance de participer aux thèses de Claire Morvan, Clémentine François et Nathanaëlle Saclier. Chacune de ces thèses a permis de pousser le modèle un peu plus loin et dans des directions nouvelles. Laurent Duret a toujours trouvé comment y voir plus clair dans ces données complexes. Une fois devenue ma collègue de bureau, Clémentine François a corrigé les 1000 fautes de cette HDR et a porté le projet fou, a posteriori, de séquencer 20 génomes d'Asellidae qui nous donneront assez de données pour les 10 ans à venir... Laura Grice, Laurent Gueguen, Gautier Debaecker et Marie Sémon ont également participé au développement de ce modèle.

Les projets d'**écologie moléculaire** ont démarré avec les thèses de Maïlys Gauthier puis Héloïse Verdier. J'y suis rentré pour filer un coup de main, et je suis vraiment content du travail que l'on a effectué et où cela nous a mené. De nouveau, il faut ici encore reconnaître le rôle prépondérant de Lara Konecny. Christophe Marquette et Thibault Datry ont également participé à cet axe de recherche.

Tout ce travail a été effectué au LEHNA sur ses plateaux techniques et grâce à ses personnels d'appui à la recherche. Il a été financé de multiple fois par le CNRS (appel à projet et délégation), l'ANR, l'OFB et le PEPR OneWater.

Le portage du parcours de Licence en Sciences de la Biodiversité, du parcours de Master en Génomique Environnementale, et de la plateforme DTAMB se font avec Isabelle Amat, Annabelle Haudry, Mylène Hugoni et Pascale Chevret.

Merci à toutes et tous.

Bibliographie

- Aagaard K, Kladakis A, Nielsen MW. 2020. Concentration or Dispersal of Research Funding?. *Quantitative Science Studies*. 1:117-149. doi: [10.1162/qss_a_00002](https://doi.org/10.1162/qss_a_00002).
- Acquisti C, Elser JJ, Kumar S. 2009. Ecological Nitrogen Limitation Shapes the DNA Composition of Plant Genomes. *Molecular Biology and Evolution*. 26:953-956. doi: [10.1093/molbev/msp038](https://doi.org/10.1093/molbev/msp038).
- Albert JS et al. 2021. Scientists' Warning to Humanity on the Freshwater Biodiversity Crisis. *Ambio*. 50:85-94. doi: [10.1007/s13280-020-01318-8](https://doi.org/10.1007/s13280-020-01318-8).
- Alleva L. 2006. Taking Time to Savour the Rewards of Slow Science. *Nature*. 443:271. doi: [10.1038/443271e](https://doi.org/10.1038/443271e).
- Anders S, Huber W. 2010. Differential Expression Analysis for Sequence Count Data. *Nature Precedings*. 1. doi: [10.1038/npre.2010.4282.1](https://doi.org/10.1038/npre.2010.4282.1).
- Baker M. 2016. 1,500 Scientists Lift the Lid on Reproducibility. *Nature*. 533:452-454. doi: [10.1038/533452a](https://doi.org/10.1038/533452a).
- Balart-García P et al. 2023. Parallel and Convergent Genomic Changes Underlie Independent Subterranean Colonization across Beetles. *Nature Communications*. 14:3842. doi: [10.1038/s41467-023-39603-1](https://doi.org/10.1038/s41467-023-39603-1).
- Balázs G, Biró A, Fiser, Fiser C, Herczeg G. 2021. Parallel Morphological Evolution and Habitat-Dependent Sexual Dimorphism in Cave- vs. Surface Populations of the *Asellus Aquaticus* (Crustacea: Isopoda: Asellidae) Species Complex. *Ecology and Evolution*. 11:15389-15403. doi: [10.1002/ece3.8233](https://doi.org/10.1002/ece3.8233).
- Barnes MA et al. 2021. Environmental Conditions Influence eDNA Particle Size Distribution in Aquatic Systems. *Environmental DNA*. 3:643-653. doi: [10.1002/edn3.160](https://doi.org/10.1002/edn3.160).
- Barnes MA, Turner CR. 2016. The Ecology of Environmental DNA and Implications for Conservation Genetics. *Conservation Genetics*. 17:1-17. doi: [10.1007/s10592-015-0775-4](https://doi.org/10.1007/s10592-015-0775-4).
- Baselga A. 2010. Partitioning the Turnover and Nestedness Components of Beta Diversity. *Global Ecology and Biogeography*. 19:134-143. doi: [10.1111/j.1466-8238.2009.00490.x](https://doi.org/10.1111/j.1466-8238.2009.00490.x).
- Baudouin-Cornu P. 2008. Stœchiométrie, mon cher Watson !. *médecine/sciences*. 24:483-490. doi: [10.1051/medsci/2008245483](https://doi.org/10.1051/medsci/2008245483).
- Baudouin-Cornu P, Surdin-Kerjan Y, Marlière P, Thomas D. 2001. Molecular Evolution of Protein Atomic Composition. *Science*. 293:297-300. doi: [10.1126/science.1061052](https://doi.org/10.1126/science.1061052).
- Bergeron LA et al. 2023. Evolution of the Germline Mutation Rate across Vertebrates. *Nature*. 615:285-291. doi: [10.1038/s41586-023-05752-y](https://doi.org/10.1038/s41586-023-05752-y).
- Bilandžija H, Četković H, Jeffery WR. 2012. Evolution of Albinism in Cave Planthropers by a Convergent Defect in the First Step of Melanin Biosynthesis. *Evolution & development*. 14:196-203. doi: [10.1111/j.1525-142X.2012.00535.x](https://doi.org/10.1111/j.1525-142X.2012.00535.x).
- Blommaert J. 2020. Genome Size Evolution: Towards New Model Systems for Old Questions. *Proceedings of the Royal Society B: Biological Sciences*. 287:20201441. doi: [10.1098/rspb.2020.1441](https://doi.org/10.1098/rspb.2020.1441).

- Bohutínská M, Peichel CL. 2024. Divergence Time Shapes Gene Reuse during Repeated Adaptation. *Trends in Ecology & Evolution*. 39:396-407. doi: [10.1016/j.tree.2023.11.007](https://doi.org/10.1016/j.tree.2023.11.007).
- Bolnick DI, Barrett RDH, Oke KB, Rennison DJ, Stuart YE. 2018. (Non)Parallel Evolution. *Annual Review of Ecology, Evolution, and Systematics*. 49:303-330. doi: [10.1146/annurev-ecolsys-110617-062240](https://doi.org/10.1146/annurev-ecolsys-110617-062240).
- Bourgeois Y, Boissinot S. 2019. On the Population Dynamics of Junk: A Review on the Population Genomics of Transposable Elements. *Genes*. 10:419. doi: [10.3390/genes10060419](https://doi.org/10.3390/genes10060419).
- Bowers HA et al. 2021. Towards the Optimization of eDNA/eRNA Sampling Technologies for Marine Biosecurity Surveillance. *Water*. 13:1113. doi: [10.3390/w13081113](https://doi.org/10.3390/w13081113).
- Bragg JG, Wagner A. 2009. Protein Material Costs: Single Atoms Can Make an Evolutionary Difference. *Trends in Genetics*. 25:5-8. doi: [10.1016/j.tig.2008.10.007](https://doi.org/10.1016/j.tig.2008.10.007).
- Brash DE. 2015. UV Signature Mutations. *Photochemistry and Photobiology*. 91:15-26. doi: [10.1111/php.12377](https://doi.org/10.1111/php.12377).
- Brawand D et al. 2011. The Evolution of Gene Expression Levels in Mammalian Organs. *Nature*. 478:343-348. doi: [10.1038/nature10532](https://doi.org/10.1038/nature10532).
- Britten RJ. 1986. Rates of DNA Sequence Evolution Differ Between Taxonomic Groups. *Science*. 231:1393-1398. doi: [10.1126/science.3082006](https://doi.org/10.1126/science.3082006).
- Buhay JE, Crandall KA. 2005. Subterranean Phylogeography of Freshwater Crayfishes Shows Extensive Gene Flow and Surprisingly Large Population Sizes. *Molecular Ecology*. 14:4259-4273. doi: [10.1111/j.1365-294X.2005.02755.x](https://doi.org/10.1111/j.1365-294X.2005.02755.x).
- Bénitière Florian, Duret, et al. 2024. GT-Drift: A Resource for Exploring the Interplay between Genetic Drift, Genomic and Transcriptomic Characteristics in Eukaryotes. *NAR Genomics and Bioinformatics*. 6:lqae64. doi: [10.1093/nargab/lqae064](https://doi.org/10.1093/nargab/lqae064).
- Bénitière Florian, Necsulea, et al. 2024. Random Genetic Drift Sets an Upper Limit on mRNA Splicing Accuracy in Metazoans. *Elife*. 13:RP93629.
- Bénitière F, Lefébure T, Duret L. 2024. Variation in the Fitness Impact of Translationally Optimal Codons among Animals. doi: [10.1101/2024.07.22.604600](https://doi.org/10.1101/2024.07.22.604600).
- Capderrey C et al. 2013. Microsatellite Development and First Population Size Estimates for the Groundwater Isopod *Proasellus Walteri*. *PLOS ONE*. 8:e76213. doi: [10.1371/journal.pone.0076213](https://doi.org/10.1371/journal.pone.0076213).
- Carpenter SR, Stanley EH, Zanden MJV. 2011. State of the World's Freshwater Ecosystems: Physical, Chemical, and Biological Changes. *Annual Review of Environment and Resources*. 36:75-99. doi: [10.1146/annurev-environ-021810-094524](https://doi.org/10.1146/annurev-environ-021810-094524).
- Carraro L, Blackman RC, Altermatt F. 2023. Modelling Environmental DNA Transport in Rivers Reveals Highly Resolved Spatio-Temporal Biodiversity Patterns. *Scientific Reports*. 13:8854. doi: [10.1038/s41598-023-35614-6](https://doi.org/10.1038/s41598-023-35614-6).
- Chen J et al. 2019. A Quantitative Framework for Characterizing the Evolutionary History of Mammalian Gene Expression. *Genome Research*. 29:53-63. doi: [10.1101/gr.237636.118](https://doi.org/10.1101/gr.237636.118).
- Chen Y et al. 2009. Life without Light: Microbial Diversity and Evidence of Sulfur- and Ammonium-Based Chemolithotrophy in Movile Cave. *The ISME Journal*. 3:1093-1104. doi: [10.1038/ismej.2009.57](https://doi.org/10.1038/ismej.2009.57).

- Clavel J, Aristide L, Morlon H. 2019. A Penalized Likelihood Framework for High-Dimensional Phylogenetic Comparative Methods and an Application to New-World Monkeys Brain Evolution. *Systematic Biology*. 68:93-116. doi: [10.1093/sysbio/syy045](https://doi.org/10.1093/sysbio/syy045).
- Cope AL, O'Meara BC, Gilchrist MA. 2020. Gene Expression of Functionally-Related Genes Coevolves across Fungal Species: Detecting Coevolution of Gene Expression Using Phylogenetic Comparative Methods. *BMC Genomics*. 21:370. doi: [10.1186/s12864-020-6761-3](https://doi.org/10.1186/s12864-020-6761-3).
- Cordier T et al. 2021. Ecosystems Monitoring Powered by Environmental Genomics: A Review of Current Strategies with an Implementation Roadmap. *Molecular Ecology*. 30:2937-2958. doi: [10.1111/mec.15472](https://doi.org/10.1111/mec.15472).
- Culver DC, Pipan T. 2009. *The Biology of Caves and Other Subterranean Habitats*. Oxford University Press: New York.
- Culver DC, Kowalko JE, Pipan T. 2023. Natural Selection versus Neutral Mutation in the Evolution of Subterranean Life: A False Dichotomy?. *Frontiers in Ecology and Evolution*. 11. doi: [10.3389/fevo.2023.1080503](https://doi.org/10.3389/fevo.2023.1080503).
- Darwin C, Darwin C, Murray J, William Clowes and Sons, Evans B&. 1859. *On the Origin of Species by Means of Natural Selection, or, The Preservation of Favoured Races in the Struggle for Life*. John Murray, Albemarle Street: London doi: [10.5962/bhl.title.82303](https://doi.org/10.5962/bhl.title.82303).
- Datry T et al. 2018. Flow Intermittence and Ecosystem Services in Rivers of the Anthropocene. *Journal of Applied Ecology*. 55:353-364. doi: [10.1111/1365-2664.12941](https://doi.org/10.1111/1365-2664.12941).
- Doolittle WF, Brunet TDP. 2017. On Causal Roles and Selected Effects: Our Genome Is Mostly Junk. *BMC Biology*. 15:116. doi: [10.1186/s12915-017-0460-9](https://doi.org/10.1186/s12915-017-0460-9).
- Dubrova YE et al. 1996. Human Minisatellite Mutation Rate after the Chernobyl Accident. *Nature*. 380:683-686. doi: [10.1038/380683a0](https://doi.org/10.1038/380683a0).
- Duchemin L, Lanore V, Veber P, Boussau B. 2023. Evaluation of Methods to Detect Shifts in Directional Selection at the Genome Scale. *Molecular Biology and Evolution*. 40:msac247. doi: [10.1093/molbev/msac247](https://doi.org/10.1093/molbev/msac247).
- Duchen P et al. 2017. Inference of Evolutionary Jumps in Large Phylogenies Using Lévy Processes. *Systematic Biology*. 66:950-963. doi: [10.1093/sysbio/syx028](https://doi.org/10.1093/sysbio/syx028).
- Dudgeon D et al. 2006. Freshwater Biodiversity: Importance, Threats, Status and Conservation Challenges. *Biological Reviews*. 81:163-182. doi: [10.1017/S1464793105006950](https://doi.org/10.1017/S1464793105006950).
- Dudgeon D, Strayer DL. 2025. Bending the Curve of Global Freshwater Biodiversity Loss: What Are the Prospects?. *Biological Reviews*. 100:205-226. doi: [10.1111/brv.13137](https://doi.org/10.1111/brv.13137).
- Dunham I et al. 2012. An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature*. 489:57-74. doi: [10.1038/nature11247](https://doi.org/10.1038/nature11247).
- Duret L, Galtier N. 2009. Biased Gene Conversion and the Evolution of Mammalian Genomic Landscapes. *Annual Review of Genomics and Human Genetics*. 10:285-311. doi: [10.1146/annurev-genom-082908-150001](https://doi.org/10.1146/annurev-genom-082908-150001).
- Dutheil JY et al. 2012. Efficient Selection of Branch-Specific Models of Sequence Evolution. *Molecular Biology and Evolution*.

- tion. 29:1861-1874. doi: [10.1093/molbev/mss059](https://doi.org/10.1093/molbev/mss059).
- Eddy SR. 2013. The ENCODE Project: Missteps Overshadowing a Success. *Current Biology*. 23:R259-R261. doi: [10.1016/j.cub.2013.03.023](https://doi.org/10.1016/j.cub.2013.03.023).
- Edwards MA, Roy S. 2017. Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition. *Environmental Engineering Science*. 34:51-61. doi: [10.1089/ees.2016.0223](https://doi.org/10.1089/ees.2016.0223).
- Eleftheriades R, Fiala C, Pasic MD. 2020. The Challenges and Mental Health Issues of Academic Trainees. doi: [10.12688/f1000research.21066.1](https://doi.org/10.12688/f1000research.21066.1).
- Elipot Y, Hinaux H, Callebert J, Rétaux S. 2013. Evolutionary Shift from Fighting to Foraging in Blind Cavefish through Changes in the Serotonin Network. *Current Biology*. 23:1-10. doi: [10.1016/j.cub.2012.10.044](https://doi.org/10.1016/j.cub.2012.10.044).
- Elser JJ, Acquisti C, Kumar S. 2011. Stoichiogenomics: The Evolutionary Ecology of Macromolecular Elemental Composition. *Trends in Ecology & Evolution*. 26:38-44. doi: [10.1016/j.tree.2010.10.006](https://doi.org/10.1016/j.tree.2010.10.006).
- Eme D et al. 2018. Do Cryptic Species Matter in Macroecology? Sequencing European Groundwater Crustaceans Yields Smaller Ranges but Does Not Challenge Biodiversity Determinants. *Ecography*. 41:424-436. doi: [10.1111/ecog.02683](https://doi.org/10.1111/ecog.02683).
- Evans TM, Bira L, Gastelum JB, Weiss LT, Vanderford NL. 2018. Evidence for a Mental Health Crisis in Graduate Education. *Nature Biotechnology*. 36:282-284. doi: [10.1038/nbt.4089](https://doi.org/10.1038/nbt.4089).
- Fauvelle F-X. 2019. Monsieur le PDG du CNRS, savez-vous ce que votre darwinisme fait à la recherche ?.
- Felsenstein J. 2004. *Inferring Phylogenies*. Sinauer: Sunderland.
- Ficetola GF, Miaud C, Pompanon F, Taberlet P. 2008. Species Detection Using Environmental DNA from Water Samples. *Biology Letters*. 4:423-425. doi: [10.1098/rsbl.2008.0118](https://doi.org/10.1098/rsbl.2008.0118).
- Fiser ˇ, Whitehorn H, Furness T, Trontelj P, Protas M. 2024. Genetic Bias in Repeated Evolution of Pigment Loss in Cave Populations of the *Asellus Aquaticus* Species Complex. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*. jez.b.23256. doi: [10.1002/jez.b.23256](https://doi.org/10.1002/jez.b.23256).
- Fong DW, Kane TC, Culver DC. 1995. Vestigialization and Loss of Non-functional Characters. *Annual Review of Ecology, Evolution, and Systematics*. 26:249-268. doi: [10.1146/annurev.es.26.110195.001341](https://doi.org/10.1146/annurev.es.26.110195.001341).
- Forni G, Mantovani B, Mikheyev AS, Luchetti A. 2024. Parthenogenetic Stick Insects Exhibit Signatures of Preservation in the Molecular Architecture of Male Reproduction. *Genome Biology and Evolution*. 16:evae73. doi: [10.1093/gbe/evae073](https://doi.org/10.1093/gbe/evae073).
- Foulquier A et al. 2011. Surface Water Linkages Regulate Trophic Interactions in a Groundwater Food Web. *Ecosystems*. 14:1339-1353. doi: [10.1007/s10021-011-9484-0](https://doi.org/10.1007/s10021-011-9484-0).
- Francois CM et al. 2016. No Evidence That Nitrogen Limitation Influences the Elemental Composition of Isopod Transcriptomes and Proteomes. *Molecular Biology and Evolution*. 33:2605-2620. doi: [10.1093/molbev/msw131](https://doi.org/10.1093/molbev/msw131).

- Fraser H, Parker T, Nakagawa S, Barnett A, Fidler F. 2018. Questionable Research Practices in Ecology and Evolution. *PLOS ONE*. 13:e200303. doi: [10.1371/journal.pone.0200303](https://doi.org/10.1371/journal.pone.0200303).
- Frith U. 2020. Fast Lane to Slow Science. *Trends in Cognitive Sciences*. 24:1-2. doi: [10.1016/j.tics.2019.10.007](https://doi.org/10.1016/j.tics.2019.10.007).
- Galtier N. 2024. Half a Century of Controversy: The Neutralist/Selectionist Debate in Molecular Evolution. *Genome Biology and Evolution*. 16:evae3. doi: [10.1093/gbe/evae003](https://doi.org/10.1093/gbe/evae003).
- Galtier N, Jobson RW, Nabholz B, Glémin S, Blier PU. 2009. Mitochondrial Whims: Metabolic Rate, Longevity and the Rate of Molecular Evolution. *Biology Letters*. 5:413-416. doi: [10.1098/rsbl.2008.0662](https://doi.org/10.1098/rsbl.2008.0662).
- Gompel N, Prud'homme B. 2009. The Causes of Repeated Genetic Evolution. *Developmental Biology*. 332:36-47. doi: [10.1016/j.ydbio.2009.04.040](https://doi.org/10.1016/j.ydbio.2009.04.040).
- Graur D et al. 2013. On the Immortality of Television Sets: "function" in the Human Genome According to the Evolution-Free Gospel of ENCODE. *Genome Biology and Evolution*. 5:578-590. doi: [10.1093/gbe/evt028](https://doi.org/10.1093/gbe/evt028).
- Gregory TR. 2005. Genome Size Evolution in Animals. *The Evolution of the Genome*. 3-87.
- Grizzetti B et al. 2017. Human Pressures and Ecological Status of European Rivers. *Scientific Reports*. 7:205. doi: [10.1038/s41598-017-00324-3](https://doi.org/10.1038/s41598-017-00324-3).
- Günther T, Lampei C, Schmid KJ. 2013. Mutational Bias and Gene Conversion Affect the Intraspecific Nitrogen Stoichiometry of the Arabidopsis Thaliana Transcriptome. *Molecular Biology and Evolution*. 30:561-568. doi: [10.1093/molbev/mss249](https://doi.org/10.1093/molbev/mss249).
- Haase P et al. 2023. The Recovery of European Freshwater Biodiversity Has Come to a Halt. *Nature*. 620:582-588. doi: [10.1038/s41586-023-06400-1](https://doi.org/10.1038/s41586-023-06400-1).
- Hanson MA, Barreiro PG, Crosetto P, Brockington D. 2024. The Strain on Scientific Publishing. *Quantitative Science Studies*. 5:823-843. doi: [10.1162/qss_a_00327](https://doi.org/10.1162/qss_a_00327).
- Henry J. 1976. Recherches Sur Les Asellidae Hypogés de La Lignée Cavaticus (Crustacea, Isopoda, Asellota).
- Jacob F. 1977. Evolution and Tinkering. *Science*. 196:1161-1166. doi: [10.1126/science.860134](https://doi.org/10.1126/science.860134).
- Jensen JD et al. 2019. The Importance of the Neutral Theory in 1968 and 50 Years on: A Response to Kern and Hahn 2018. *Evolution*. 73:111-114. doi: [10.1111/evo.13650](https://doi.org/10.1111/evo.13650).
- Johri P, Charlesworth B, Jensen JD. 2020. Toward an Evolutionarily Appropriate Null Model: Jointly Inferring Demography and Purifying Selection. *Genetics*. 215:173-192. doi: [10.1534/genetics.119.303002](https://doi.org/10.1534/genetics.119.303002).
- Jones R, Culver D, Kane T. 1992. Are Parallel Morphologies of Cave Organisms the Result of Similar Selection Pressure ?. *Evolution*. 46:353-365.
- Jugovic J et al. 2015. Estimating Population Size of the Cave Shrimp *Troglocaris Anophthalmus* (Crustacea, Decapoda, Caridea) Using Mark-Release-Capture Data. *Animal Biodiversity and Conservation*. 38:77-86. doi: [10.32800/abc.2015.38.0077](https://doi.org/10.32800/abc.2015.38.0077).
- Keck F, Couton M, Altermatt F. 2023. Navigating the Seven Challenges of Taxonomic Reference Databases in Metabarcoding Analyses. *Molecular Ecology Resources*. 23:742-755. doi: [10.1111/1755-0998.13746](https://doi.org/10.1111/1755-0998.13746).

- Kern AD, Hahn MW. 2018. The Neutral Theory in Light of Natural Selection. *Molecular Biology and Evolution*. 35:1366-1371. doi: [10.1093/molbev/msy092](https://doi.org/10.1093/molbev/msy092).
- Kimmel K, Avolio ML, Ferraro PJ. 2023. Empirical Evidence of Widespread Exaggeration Bias and Selective Reporting in Ecology. *Nature Ecology & Evolution*. 7:1525-1536. doi: [10.1038/s41559-023-02144-3](https://doi.org/10.1038/s41559-023-02144-3).
- Kimura M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press: Cambridge doi: [10.1017/CBO9780511623486](https://doi.org/10.1017/CBO9780511623486).
- Kimura M, Ohta T. 1971. On the Rate of Molecular Evolution. *Journal of Molecular Evolution*. 1:1-17. doi: [10.1007/BF01659390](https://doi.org/10.1007/BF01659390).
- Konec M, Prevořčnik S, Sarbu SM, Verovnik R, Trontelj P. 2015. Parallels between Two Geographically and Ecologically Disparate Cave Invasions by the Same Species, *Asellus Aquaticus* (Isopoda, Crustacea). *Journal of Evolutionary Biology*. 28:864-875. doi: [10.1111/jeb.12610](https://doi.org/10.1111/jeb.12610).
- Koonin EV. 2009. Darwinian Evolution in the Light of Genomics. *Nucleic Acids Research*. 37:1011-1034. doi: [10.1093/nar/gkp089](https://doi.org/10.1093/nar/gkp089).
- Lafuente E et al. 2021. Building on 150 Years of Knowledge: The Freshwater Isopod *Asellus Aquaticus* as an Integrative Eco-Evolutionary Model System. *Frontiers in Ecology and Evolution*. 9. doi: [10.3389/fevo.2021.748212](https://doi.org/10.3389/fevo.2021.748212).
- Landis MJ, Schraiber JG, Liang M. 2013. Phylogenetic Analysis Using Lévy Processes: Finding Jumps in the Evolution of Continuous Traits. *Systematic Biology*. 62:193-204. doi: [10.1093/sysbio/sys086](https://doi.org/10.1093/sysbio/sys086).
- Lanfear R, Kokko H, Eyre-Walker A. 2014. Population Size and the Rate of Evolution. *Trends in Ecology & Evolution*. 29:33-41. doi: [10.1016/j.tree.2013.09.009](https://doi.org/10.1016/j.tree.2013.09.009).
- Lanfear R, Thomas JA, Welch JJ, Brey T, Bromham L. 2007. Metabolic Rate Does Not Calibrate the Molecular Clock. *Proceedings of the National Academy of Sciences*. 104:15388-15393. doi: [10.1073/pnas.0703359104](https://doi.org/10.1073/pnas.0703359104).
- Lanfear R, Welch JJ, Bromham L. 2010. Watching the Clock: Studying Variation in Rates of Molecular Evolution between Species. *Trends in Ecology & Evolution*. 25:495-503. doi: [10.1016/j.tree.2010.06.007](https://doi.org/10.1016/j.tree.2010.06.007).
- Leese F et al. 2021. Improved Freshwater Macroinvertebrate Detection from Environmental DNA through Minimized Nontarget Amplification. *Environmental DNA*. 3:261-276. doi: [10.1002/edn3.177](https://doi.org/10.1002/edn3.177).
- Lefébure T, Douady CJ, Gouy M, Gibert J. 2006. Relationship between Morphological Taxonomy and Molecular Divergence within Crustacea: Proposal of a Molecular Threshold to Help Species Delimitation. *Molecular Phylogenetics and Evolution*. 40:435-447. doi: [10.1016/j.ympev.2006.03.014](https://doi.org/10.1016/j.ympev.2006.03.014).
- Lefébure T et al. 2017. Less Effective Selection Leads to Larger Genomes. *Genome Research*. 27:1016-1028. doi: [10.1101/gr.212589.116](https://doi.org/10.1101/gr.212589.116).
- Leigh C, Datry T. 2017. Drying as a Primary Hydrological Determinant of Biodiversity in River Systems: A Broad-Scale Analysis. *Ecography*. 40:487-499. doi: [10.1111/ecog.02230](https://doi.org/10.1111/ecog.02230).
- Lewontin RC. 1974. *The Genetic Basis of Evolutionary Change*. Columbia University Press.

- Leys R, Cooper SJ, Strecker U, Wilkens H. 2005. Regressive Evolution of an Eye Pigment Gene in Independently Evolved Eyeless Subterranean Diving Beetles. *Biology letters*. 1:496-499. doi: [10.1098/rsbl.2005.0358](https://doi.org/10.1098/rsbl.2005.0358).
- Li W-H, Tanimura M. 1987. The Molecular Clock Runs More Slowly in Man than in Apes and Monkeys. *Nature*. 326:93-96. doi: [10.1038/326093a0](https://doi.org/10.1038/326093a0).
- Luo H, Thompson LR, Stiglitz U, Hughes AL. 2015. Selection Maintains Low Genomic GC Content in Marine SAR11 Lineages. *Molecular Biology and Evolution*. 32:2738-2748. doi: [10.1093/molbev/msv149](https://doi.org/10.1093/molbev/msv149).
- Lynch Abigail J. et al. 2023. People Need Freshwater Biodiversity. *WIREs Water*. 10:e1633. doi: [10.1002/wat2.1633](https://doi.org/10.1002/wat2.1633).
- Lynch M. 2002. Intron Evolution as a Population-Genetic Process. *Proceedings of the National Academy of Sciences*. 99:6118-6123. doi: [10.1073/pnas.092595699](https://doi.org/10.1073/pnas.092595699).
- Lynch M. 2007a. The Frailty of Adaptive Hypotheses for the Origins of Organismal Complexity. *Proceedings of the National Academy of Sciences of the United States of America*. 104:8597-8604. doi: [10.1073/pnas.0702207104](https://doi.org/10.1073/pnas.0702207104).
- Lynch M. 2006. The Origins of Eukaryotic Gene Structure. *Molecular Biology and Evolution*. 23:450-468. doi: [10.1093/molbev/msj050](https://doi.org/10.1093/molbev/msj050).
- Lynch M. 2007b. *The Origins of Genome Architecture*. Sinauer: Sunderland.
- Lynch M et al. 2016. Genetic Drift, Selection and the Evolution of the Mutation Rate. *Nature Reviews Genetics*. 17:704-714. doi: [10.1038/nrg.2016.104](https://doi.org/10.1038/nrg.2016.104).
- Lynch Michael et al. 2023. The Divergence of Mutation Rates and Spectra across the Tree of Life. *EMBO reports*. 24:e57561. doi: [10.15252/embr.202357561](https://doi.org/10.15252/embr.202357561).
- Lynch M, Conery JS. 2003. The Origins of Genome Complexity. *Science (New York, N.Y.)*. 302:1401-1404. doi: [10.1126/science.1089370](https://doi.org/10.1126/science.1089370).
- Malard F, Galassi D, Lafont M, Dolédec S, Ward J. 2003. Longitudinal Patterns of Invertebrates in the Hyporheic Zone of a Glacial River. *Freshwater Biology*. 48:1709-1725. doi: [10.1046/j.1365-2427.2003.01118.x](https://doi.org/10.1046/j.1365-2427.2003.01118.x).
- Malard F, Mangin A, Uehlinger U, Ward JV. 2001. Thermal Heterogeneity in the Hyporheic Zone of a Glacial Floodplain. *Canadian Journal of Fisheries and Aquatic Sciences*. 58:1319-1335. doi: [10.1139/f01-079](https://doi.org/10.1139/f01-079).
- Mammola S. 2019. Finding Answers in the Dark: Caves as Models in Ecology Fifty Years after Poulson and White. *Ecography*. 42:1331-1351. doi: [10.1111/ecog.03905](https://doi.org/10.1111/ecog.03905).
- Mammola S et al. 2021. Collecting Eco-Evolutionary Data in the Dark: Impediments to Subterranean Research and How to Overcome Them. *Ecology and Evolution*. 11:5911-5926. doi: [10.1002/ece3.7556](https://doi.org/10.1002/ece3.7556).
- Marais GA, Shankland R, Haag P, Fiault R, Juniper B. 2018. A Survey and a Positive Psychology Intervention on French PhD Student Well-being. *International Journal of Doctoral Studies*. 13:109-138. doi: [10.28945/3948](https://doi.org/10.28945/3948).
- Marchant R, Norris RH, Milligan A. 2006. Evaluation and Application of Methods for Biological Assessment of Streams: Summary of Papers. *Hydrobiologia*. 572:1-7. doi: [10.1007/s10750-006-0382-y](https://doi.org/10.1007/s10750-006-0382-y).

- Marino A, Debaecker G, Fiston-Lavier A-S, Haudry A, Nabholz B. 2024. Effective Population Size Does Not Explain Long-Term Variation in Genome Size and Transposable Element Content in Animals. doi: [10.1101/2024.02.26.582137](https://doi.org/10.1101/2024.02.26.582137).
- Martin AP, Naylor GJP, Palumbi SR. 1992. Rates of Mitochondrial DNA Evolution in Sharks Are Slow Compared with Mammals. *Nature*. 357:153-155. doi: [10.1038/357153a0](https://doi.org/10.1038/357153a0).
- Messenger ML et al. 2021. Global Prevalence of Non-Perennial Rivers and Streams. *Nature*. 594:391-397. doi: [10.1038/s41586-021-03565-5](https://doi.org/10.1038/s41586-021-03565-5).
- Millennium Ecosystem Assessment. 2005. *Ecosystems and Human Well-Being: Wetlands and Water Synthesis*.
- Mills L, Read GJM, Bragg JE, Hutchinson BT, Cox JA. 2024. A Study into the Mental Health of PhD Students in Australia: Investigating the Determinants of Depression, Anxiety, and Suicidality. *Scientific Reports*. 14:22636. doi: [10.1038/s41598-024-72661-z](https://doi.org/10.1038/s41598-024-72661-z).
- Mimeau L et al. 2024. Projections of Streamflow Intermittence under Climate Change in European Drying River Networks. *Hydrology and Earth System Sciences Discussions*. 1-34. doi: [10.5194/hess-2024-272](https://doi.org/10.5194/hess-2024-272).
- Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation. 2021. *Le Devenir Des Docteurs Trois Ans Après l'obtention de Leur Thèse - État de l'Enseignement Supérieur, de La Recherche et de l'Innovation En France N°17*.
- Mondy CP, Villeneuve B, Archaimbault V, Usseglio-Polatera P. 2012. A New Macroinvertebrate-Based Multimetric Index (I2M2) to Evaluate Ecological Quality of French Wadeable Streams Fulfilling the WFD Demands: A Taxonomical and Trait Approach. *Ecological Indicators*. 18:452-467. doi: [10.1016/j.ecolind.2011.12.013](https://doi.org/10.1016/j.ecolind.2011.12.013).
- Morel M, Booker DJ, Gob F, Lamouroux N. 2020. Intercontinental Predictions of River Hydraulic Geometry from Catchment Physical Characteristics. *Journal of Hydrology*. 582:124292. doi: [10.1016/j.jhydrol.2019.124292](https://doi.org/10.1016/j.jhydrol.2019.124292).
- Mugal CF, Wolf JBW, von Grünberg HH, Ellegren H. 2010. Conservation of Neutral Substitution Rate and Substitutional Asymmetries in Mammalian Genes. *Genome Biology and Evolution*. 2:19-28. doi: [10.1093/gbe/evp056](https://doi.org/10.1093/gbe/evp056).
- Nabholz B, Glémin S, Galtier N. 2008. Strong Variations of Mitochondrial Mutation Rate across Mammals—the Longevity Hypothesis. *Molecular Biology and Evolution*. 25:120-130. doi: [10.1093/molbev/msm248](https://doi.org/10.1093/molbev/msm248).
- Nielsen MW, Andersen JP. 2021. Global Citation Inequality Is on the Rise. *Proceedings of the National Academy of Sciences*. 118:e2012208118. doi: [10.1073/pnas.2012208118](https://doi.org/10.1073/pnas.2012208118).
- Nielsen R, Yang Z. 2003. Estimating the Distribution of Selection Coefficients from Phylogenetic Data with Applications to Mitochondrial and Viral DNA. *Molecular Biology and Evolution*. 20:1231-1239. doi: [10.1093/molbev/msg147](https://doi.org/10.1093/molbev/msg147).
- Niemiller ML, Fitzpatrick BM, Shah P, Schmitz L, Near TJ. 2013. Evidence for Repeated Loss of Selective Constraint in Rhodopsin of Amblyopsid Cavefishes (Teleostei: Amblyopsidae). *Evolution*. 67:732-748. doi: [10.1111/j.1558-5646.2012.01822.x](https://doi.org/10.1111/j.1558-5646.2012.01822.x).

- O'Gorman M et al. 2021. Pleiotropic Function of the *Oca2* Gene Underlies the Evolution of Sleep Loss and Albinism in Cavefish. *Current Biology*. 31:3694-3701. doi: [10.1016/j.cub.2021.06.077](https://doi.org/10.1016/j.cub.2021.06.077).
- Oakley TH, Speiser DI. 2015. How Complexity Originates: The Evolution of Animal Eyes. *Annual Review of Ecology, Evolution, and Systematics*. 46:237-260. doi: [10.1146/annurev-ecolsys-110512-135907](https://doi.org/10.1146/annurev-ecolsys-110512-135907).
- Ohta T. 1973. Slightly Deleterious Mutant Substitutions in Evolution. *Nature*. 246:96-98. doi: [10.1038/246096a0](https://doi.org/10.1038/246096a0).
- Ohta T. 1992. The Nearly Neutral Theory of Molecular Evolution. *Annual Review of Ecology and Systematics*. 23:263-286. doi: [10.1146/annurev.es.23.110192.001403](https://doi.org/10.1146/annurev.es.23.110192.001403).
- Pawlowski J, Apothéoz-Perret-Gentil L, Altermatt F. 2020. Environmental DNA: What's behind the Term? Clarifying the Terminology and Recommendations for Its Future Use in Biomonitoring. *Molecular Ecology*. 29:4258-4264. doi: [10.1111/mec.15643](https://doi.org/10.1111/mec.15643).
- Pawlowski J et al. 2018. The Future of Biotic Indices in the Ecogenomic Era: Integrating (e)DNA Metabarcoding in Biological Assessment of Aquatic Ecosystems. *Science of The Total Environment*. 1295-1310. doi: [10.1016/j.scitotenv.2018.05.002](https://doi.org/10.1016/j.scitotenv.2018.05.002).
- Petrov DA. 2002. Mutational Equilibrium Model of Genome Size Evolution. *Theoretical Population Biology*. 61:531-544. doi: [10.1006/tpbi.2002.1605](https://doi.org/10.1006/tpbi.2002.1605).
- Pipán T, Culver DC. 2012. Convergence and Divergence in the Subterranean Realm: A Reassessment. *Biological Journal of the Linnean Society*. 107:1-14. doi: [10.1111/j.1095-8312.2012.01964.x](https://doi.org/10.1111/j.1095-8312.2012.01964.x).
- Policarpo M et al. 2021. Contrasting Gene Decay in Subterranean Vertebrates: Insights from Cavefishes and Fossorial Mammals. *Molecular Biology and Evolution*. 38:589-605. doi: [10.1093/molbev/msaa249](https://doi.org/10.1093/molbev/msaa249).
- Pommier S et al. 2022. *Le Doctorat En France- Regards Croisés Sur La Formation Doctorale*.
- Pont D. 2024. Predicting Downstream Transport Distance of Fish eDNA in Lotic Environments. *Molecular Ecology Resources*. 24:e13934. doi: [10.1111/1755-0998.13934](https://doi.org/10.1111/1755-0998.13934).
- Pont D et al. 2018. Environmental DNA Reveals Quantitative Patterns of Fish Biodiversity in Large Rivers despite Its Downstream Transportation. *Scientific Reports*. 8:10361. doi: [10.1038/s41598-018-28424-8](https://doi.org/10.1038/s41598-018-28424-8).
- Porter ML, Cronin TW, McClellan DA, Crandall KA. 2007. Molecular Characterization of Crustacean Visual Pigments and the Evolution of Pancrustacean Opsins. *Molecular Biology and Evolution*. 24:253-268. doi: [10.1093/molbev/msl152](https://doi.org/10.1093/molbev/msl152).
- Poulson T, White W. 1969. The Cave Environment. *Science*. 165:971-981. doi: [10.1126/science.165.3897.971](https://doi.org/10.1126/science.165.3897.971).
- Pouyet F, Gilbert KJ. 2021. Towards an improved understanding of molecular evolution: the relative roles of selection, drift, and everything in between. *Peer Community Journal*. 1. doi: [10.24072/pcjournal.16](https://doi.org/10.24072/pcjournal.16).
- Power H, Takahashi M, Jarman S, Berry O. 2023. What Is Environmental DNA?. *Environmental DNA*. 5:1743-1758. doi: [10.1002/edn3.497](https://doi.org/10.1002/edn3.497).
- Price PD et al. 2022. Detecting Signatures of Selection on Gene Expression. *Nature Ecology & Evolution*. 6:1035-1045. doi: [10.1038/s41559-022-01761-8](https://doi.org/10.1038/s41559-022-01761-8).

- Protas ME et al. 2006. Genetic Analysis of Cavefish Reveals Molecular Convergence in the Evolution of Albinism. *Nature Genetics*. 38:107-111. doi: [10.1038/ng1700](https://doi.org/10.1038/ng1700).
- Protas ME, Trontelj P, Patel NH. 2011. Genetic Basis of Eye and Pigment Loss in the Cave Crustacean, *Asellus Aquaticus*. *Proceedings of the National Academy of Sciences of the United States of America*. 108:5702-5707. doi: [10.1073/pnas.1013850108](https://doi.org/10.1073/pnas.1013850108).
- Protas M, Conrad M, Gross JB, Tabin C, Borowsky R. 2007. Regressive Evolution in the Mexican Cave Tetra, *Astyanax Mexicanus*. *Current Biology*. 17:452-454. doi: [10.1016/j.cub.2007.01.051](https://doi.org/10.1016/j.cub.2007.01.051).
- Protas M, Jeffery WR. 2012. Evolution and Development in Cave Animals: From Fish to Crustaceans. *Wiley Interdisciplinary Reviews: Developmental Biology*. 1:823-845. doi: [10.1002/wdev.61](https://doi.org/10.1002/wdev.61).
- Protas M, Trontelj P, Prevořnik S, Fiser V. 2023. The *Asellus Aquaticus* Species Complex Malard, F, Griebler, C, & Rétaux, S, éditeurs. *Groundwater Ecology and Evolution* (Second Edition). 329-350. doi: [10.1016/B978-0-12-819119-4.00016-0](https://doi.org/10.1016/B978-0-12-819119-4.00016-0).
- Rawat S, Meena S. 2014. Publish or Perish: Where Are We Heading?. *Journal of Research in Medical Sciences : The Official Journal of Isfahan University of Medical Sciences*. 19:87-89.
- Reid AJ et al. 2019. Emerging Threats and Persistent Conservation Challenges for Freshwater Biodiversity. *Biological Reviews*. 94:849-873. doi: [10.1111/brv.12480](https://doi.org/10.1111/brv.12480).
- dos Reis M, Wernisch L. 2009. Estimating Translational Selection in Eukaryotic Genomes. *Molecular Biology and Evolution*. 26:451-461. doi: [10.1093/molbev/msn272](https://doi.org/10.1093/molbev/msn272).
- Rey C et al. 2019. Detecting Adaptive Convergent Amino Acid Evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 374:20180234. doi: [10.1098/rstb.2018.0234](https://doi.org/10.1098/rstb.2018.0234).
- Rohlfsv RV, Harrigan P, Nielsen R. 2014. Modeling Gene Expression Evolution with an Extended Ornstein–Uhlenbeck Process Accounting for Within-Species Variation. *Molecular Biology and Evolution*. 31:201-211. doi: [10.1093/molbev/mst190](https://doi.org/10.1093/molbev/mst190).
- Sackton TB, Clark N. 2019. Convergent Evolution in the Genomics Era: New Insights and Directions. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 374:20190102. doi: [10.1098/rstb.2019.0102](https://doi.org/10.1098/rstb.2019.0102).
- Saclier N et al. 2024. A Collaborative Backbone Resource for Comparative Studies of Subterranean Evolution: The World Asellidae Database. *Molecular Ecology Resources*. 24:e13882. doi: [10.1111/1755-0998.13882](https://doi.org/10.1111/1755-0998.13882).
- Saclier N et al. 2020. Bedrock Radioactivity Influences the Rate and Spectrum of Mutation Przeworski, M, Weigel, D, & Sunyaev, SR, éditeurs. *eLife*. 9:e56830. doi: [10.7554/eLife.56830](https://doi.org/10.7554/eLife.56830).
- Saclier N et al. 2018. Life History Traits Impact the Nuclear Rate of Substitution but Not the Mitochondrial Rate in Isopods Crandall, K, éditeur. *Molecular Biology and Evolution*. 35:2900-2912. doi: [10.1093/molbev/msy184](https://doi.org/10.1093/molbev/msy184).
- Sassoubre LM, Yamahara KM, Gardner LD, Block BA, Boehm AB. 2016. Quantification of Environmental DNA (eDNA) Shedding and Decay Rates for Three Marine Fish. *Environmental Science & Technology*. 50:10456-10464. doi: [10.1021/acs.est.6b03114](https://doi.org/10.1021/acs.est.6b03114).

- Saudemont B et al. 2017. The Fitness Cost of Mis-Splicing Is the Main Determinant of Alternative Splicing Patterns. *Genome Biology*. 18:208. doi: [10.1186/s13059-017-1344-6](https://doi.org/10.1186/s13059-017-1344-6).
- Sayer CA et al. 2025. One-Quarter of Freshwater Fauna Threatened with Extinction. *Nature*. 1-8. doi: [10.1038/s41586-024-08375-z](https://doi.org/10.1038/s41586-024-08375-z).
- Searcy RT et al. 2022. High-frequency and Long-term Observations of eDNA from Imperiled Salmonids in a Coastal Stream: Temporal Dynamics, Relationships with Environmental Factors, and Comparisons with Conventional Observations. *Environmental DNA*. 4:776-789. doi: [10.1002/edn3.293](https://doi.org/10.1002/edn3.293).
- Shendure J, Ji H. 2008. Next-Generation DNA Sequencing. *Nature Biotechnology*. 26:1135-1145. doi: [10.1038/nbt1486](https://doi.org/10.1038/nbt1486).
- Shibutani S, Takeshita M, Grollman AP. 1991. Insertion of Specific Bases during DNA Synthesis Past the Oxidation-Damaged Base 8-oxodG. *Nature*. 349:431-434. doi: [10.1038/349431a0](https://doi.org/10.1038/349431a0).
- Shogren AJ et al. 2017. Controls on eDNA Movement in Streams: Transport, Retention, and Resuspension. *Scientific Reports*. 7:5065. doi: [10.1038/s41598-017-05223-1](https://doi.org/10.1038/s41598-017-05223-1).
- Shokralla S, Spall JL, Gibson JF, Hajibabaei M. 2012. Next-Generation Sequencing Technologies for Environmental DNA Research. *Molecular Ecology*. 21:1794-1805. doi: [10.1111/j.1365-294X.2012.05538.x](https://doi.org/10.1111/j.1365-294X.2012.05538.x).
- Sinclair JS et al. 2024. Multi-Decadal Improvements in the Ecological Quality of European Rivers Are Not Consistently Reflected in Biodiversity Metrics. *Nature Ecology & Evolution*. 8:430-441. doi: [10.1038/s41559-023-02305-4](https://doi.org/10.1038/s41559-023-02305-4).
- Sinniger F et al. 2016. Worldwide Analysis of Sedimentary DNA Reveals Major Gaps in Taxonomic Knowledge of Deep-Sea Benthos. *Frontiers in Marine Science*. 3.
- Smaldino PE, McElreath R. 2016. The Natural Selection of Bad Science. *Royal Society Open Science*. 3:160384. doi: [10.1098/rsos.160384](https://doi.org/10.1098/rsos.160384).
- Speiser DI et al. 2014. Using Phylogenetically-Informed Annotation (PIA) to Search for Light-Interacting Genes in Transcriptomes from Non-Model Organisms. *BMC Bioinformatics*. 15:350. doi: [10.1186/s12859-014-0350-x](https://doi.org/10.1186/s12859-014-0350-x).
- Stern DB, Crandall KA. 2018. The Evolution of Gene Expression Underlying Vision Loss in Cave Animals. *Molecular Biology and Evolution*. 35:2005-2014. doi: [10.1093/molbev/msy106](https://doi.org/10.1093/molbev/msy106).
- Sterner RW, Elser JJ. 2017. *Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere*. Princeton University Press doi: [10.1515/9781400885695](https://doi.org/10.1515/9781400885695).
- Straalen NM van, Roelofs D. 2011. *An Introduction to Ecological Genomics*. OUP Oxford.
- Stubbington R et al. 2018. Biomonitoring of Intermittent Rivers and Ephemeral Streams in Europe: Current Practice and Priorities to Enhance Ecological Status Assessments. *Science of The Total Environment*. 618:1096-1113. doi: [10.1016/j.scitotenv.2017.09.137](https://doi.org/10.1016/j.scitotenv.2017.09.137).
- Thalinger B et al. 2021. The Effect of Activity, Energy Use, and Species Identity on Environmental DNA Shedding of Freshwater Fish. *Frontiers in Ecology and Evolution*. 9. doi: [10.3389/fevo.2021.623718](https://doi.org/10.3389/fevo.2021.623718).
- Trajano E, Bichuette ME, Kapoor BG, éd. 2010. *Biology of Subterranean Fishes*.

- CRC Press: Boca Raton doi: [10.1201/EBK1578086702](https://doi.org/10.1201/EBK1578086702).
- Venarsky MP, Huntsman BM, Huryn AD, Benstead JP, Kuhajda BR. 2014. Quantitative Food Web Analysis Supports the Energy-Limitation Hypothesis in Cave Stream Ecosystems. *Oecologia*. 176:859-869. doi: [10.1007/s00442-014-3042-3](https://doi.org/10.1007/s00442-014-3042-3).
- Verdier H et al. 2024. Environmental DNA Particle Size Distribution and Quantity Differ Across Taxa and Organelles. *Environmental DNA*. 6:e598. doi: [10.1002/edn3.598](https://doi.org/10.1002/edn3.598).
- Verdier H et al. 2022. Passive Sampling of Environmental DNA in Aquatic Environments Using 3D-printed Hydroxyapatite Samplers. *Molecular Ecology Resources*. 22:2158-2170. doi: [10.1111/1755-0998.13604](https://doi.org/10.1111/1755-0998.13604).
- Vörös J, Ursenbacher S, Jelić D. 2019. Population Genetic Analyses Using 10 New Polymorphic Microsatellite Loci Confirms Genetic Subdivision within the Olm, *Proteus Anguinus*. *Journal of Heredity*. 110:211-218. doi: [10.1093/jhered/esy067](https://doi.org/10.1093/jhered/esy067).
- Wang Y, Obbard DJ. 2023. Experimental Estimates of Germline Mutation Rate in Eukaryotes: A Phylogenetic Meta-Analysis. *Evolution Letters*. 7:216-226. doi: [10.1093/evlett/qrada027](https://doi.org/10.1093/evlett/qrada027).
- Wilcox TM et al. 2016. Understanding Environmental DNA Detection Probabilities: A Case Study Using a Stream-Dwelling Char *Salvelinus fontinalis*. *Biological Conservation*. 194:209-216. doi: [10.1016/j.biocon.2015.12.023](https://doi.org/10.1016/j.biocon.2015.12.023).
- Wilkens H. 1988. Evolution and Genetics of Epigean and Cave *Astyanax fasciatus* (Characidae, Pisces): Support for the Neutral Mutation Theory. *Evolutionary biology*. 23:271-367. doi: [10.1007/978-1-4613-1043-3_8](https://doi.org/10.1007/978-1-4613-1043-3_8).
- WWF. 2024. *Living Planet Report 2024 – A System in Peril*.
- Xu C, Park J-K, Zhang J. 2019. Evidence That Alternative Transcriptional Initiation Is Largely Nonadaptive. *PLOS Biology*. 17:e3000197. doi: [10.1371/journal.pbio.3000197](https://doi.org/10.1371/journal.pbio.3000197).
- Yamamoto Y, Byerly MS, Jackman WR, Jeffery WR. 2009. Pleiotropic Functions of Embryonic Sonic Hedgehog Expression Link Jaw and Taste Bud Amplification with Eye Loss during Cavefish Evolution. *Developmental biology*. 330:200-211.
- Yamamoto Y, Jeffery WR. 2000. Central Role for the Lens in Cave Fish Eye Degeneration. *Science*. 289:631-633. doi: [10.1126/science.289.5479.631](https://doi.org/10.1126/science.289.5479.631).
- Yamamoto Y, Stock DW, Jeffery WR. 2004. Hedgehog Signalling Controls Eye Degeneration in Blind Cavefish. *Nature*. 431:844-847.
- Yates MC, Cristescu ME, Derry AM. 2021. Integrating Physiology and Environmental Dynamics to Operationalize Environmental DNA (eDNA) as a Means to Monitor Freshwater Macro-Organism Abundance. *Molecular Ecology*. 30:6531-6550. doi: [10.1111/mec.16202](https://doi.org/10.1111/mec.16202).
- Yoshizawa M, Yamamoto Y, O'Quin KE, Jeffery WR. 2012. Evolution of an Adaptive Behavior and Its Sensory Receptors Promotes Eye Regression in Blind Cavefish. *BMC biology*. 10:108. doi: [10.1186/1741-7007-10-108](https://doi.org/10.1186/1741-7007-10-108).
- Zhang Z et al. 2022. Residual Chlorine Disrupts the Microbial Communities and Spreads Antibiotic Resistance in Freshwater. *Journal of Hazardous Materials*.

dous Materials. 423:127152. doi: [10.1016/j.jhazmat.2021.127152](https://doi.org/10.1016/j.jhazmat.2021.127152).

Zhao Q, Shao F, Li Y, Yi SV, Peng Z. 2022. Novel Genome Sequence of Chinese Cavefish (*Triplophysa Rosa*) Reveals Pervasive Relaxation of Natural Selection in Cavefish Genomes. *Molecular Ecology*. 31:5831-5845. doi: [10.1111/mec.16700](https://doi.org/10.1111/mec.16700).