



**HAL**  
open science

# Origine des variations de taux d'évolution moléculaire inter-spécifiques : apport d'un modèle génomique en milieu souterrain

Nathanaëlle Saclier

► **To cite this version:**

Nathanaëlle Saclier. Origine des variations de taux d'évolution moléculaire inter-spécifiques : apport d'un modèle génomique en milieu souterrain. Génomique, Transcriptomique et Protéomique [q-bio.GN]. Université de Lyon, 2019. Français. NNT : 2019LYSE1301 . tel-03245700

**HAL Id: tel-03245700**

**<https://theses.hal.science/tel-03245700>**

Submitted on 2 Jun 2021

**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.

N°d'ordre NNT : 2019LYSE1301



**THESE de DOCTORAT DE L'UNIVERSITE DE LYON**  
opérée au sein de  
**l'Université Claude Bernard Lyon 1**  
**Ecole Doctorale N° 341**  
**(Evolution Ecocystèmes Microbiologie Modélisation)**

**Spécialité de doctorat** : Génomique évolutive  
**Discipline** : Biologie

Soutenue publiquement le 12/12/2019, par :  
**Nathanaëlle Saclier**

---

**Origine des variations de taux  
d'évolution moléculaire inter-spécifiques :  
apport d'un modèle génomique en milieu  
souterrain**

---

Devant le jury composé de :

ACHAZ, Guillaume	Maître de conférence, MNHN Paris	Rapporteur
GLÉMIN, Sylvain	Directeur de recherche, CNRS	Rapporteur
MOUCHIROUD, Dominique	Professeure, UCB Lyon 1	Examinatrice
SAMADI, Sarah	Professeure, MNHN Paris	Examinatrice
DOUADY, Christophe	Professeur, UCB Lyon 1	Directeur de thèse
LEFÉBURE, Tristan	Maître de conférence, UCB Lyon 1	Co-directeur de thèse



---

**TITRE : Origine des variations de taux d'évolution moléculaire interspécifiques : apport d'un modèle génomique en milieu souterrain**

**Résumé** La vitesse à laquelle les séquences d'ADN évoluent varie selon les espèces. Ces différences peuvent venir de caractéristiques intrinsèques de l'espèce (taux métabolique, traits d'histoire de vie) ou de son environnement (rayonnements ionisants). L'objectif de cette thèse est de tester les hypothèses expliquant les variations de taux d'évolution moléculaire entre les espèces. Pour cela, les particularités des Asellidae souterrains ont été couplées avec des données de séquençage nouvelle génération dans le génome nucléaire et le génome mitochondrial. L'utilisation des Asellidae comme modèle biologique nous permet d'avoir, au sein du même groupe, des espèces ayant indépendamment effectué une transition vers le milieu souterrain. Cette transition étant accompagnée de nombreux changements, tant biologiques (longévité, taux métabolique, temps de génération) qu'environnementaux, elle nous permet, au sein du même groupe, de pouvoir comparer des espèces contrastées en termes de longévité, de rayonnements ionisants ou encore de productivité et de température. Ce modèle biologique nous permet ainsi de préciser et de quantifier les facteurs responsables de variations du taux d'évolution moléculaire entre les espèces. Par cette approche, un effet du temps de génération sur le taux d'évolution du génome nucléaire et une absence d'effet sur le génome mitochondrial a pu être mis en évidence. Nous avons ensuite montré un effet de la radioactivité naturelle, d'une ampleur analogue à celle du temps de génération, mais cette fois dans les deux compartiments génomiques. Enfin, l'étude des variations des taux d'évolution moléculaire à une échelle globale a révélé des biais dans les calculs des taux de substitutions qui devront être pris en compte dans les études cherchant à établir le lien entre le taux de mutations et la diversification.

**Mots-clés** Taux d'évolution moléculaire – Taux de mutations – Taux de substitutions – Asellidae – Traits d'histoire de vie – Radioactivité – Macro-écologie – Approche comparative – biodiversité souterraine.

---

**TITLE : Understanding the determinants of the rate of molecular evolution : contribution of a subterranean genomic model**

**Abstract** The rate at which DNA accumulates substitutions varies widely among species. Rate variations have been imputed to species intrinsic features (metabolic rate, life history traits) or to the environment characteristics (ionizing radiations, selection pressure). The aim of this PhD project was to investigate the main hypotheses explaining variations in the rate of molecular evolution between species. To achieve that, we combined the unique properties of subterranean isopods from the Asellidae family and high-throughput sequencing data from the nuclear and mitochondrial genome. Asellidae species have made multiple independent transitions to subterranean environments where subterranean species have repeatedly evolved a lower metabolic rate, a longer lifespan and a longer generation time. Moreover, because they are poor dispersers, they are exposed to the same environment across many generations, allowing us to compare species with long-term contrasted features in term of life history traits and environmental characteristics. We found that generation time negatively impact the rate of molecular evolution in the nuclear genome whereas the mitochondrial rate remained unchanged. We also found an increase of the mutation rate for species living in naturally highly radioactive environments. Finally, the study of the rate of molecular evolution variation at a global scale brought forward a systematic bias which needs to be taken into account in studying the link between the mutation rate and diversification.

---

## Remerciements

Si la rédaction du manuscrit de thèse est un travail assez personnel, son contenu est lui le résultat de nombreuses collaborations. Dans chaque partie de ce manuscrit, je me suis ainsi attachée à décrire la contribution scientifique de toutes les personnes qui ont participé à l'élaboration de ce travail, sans cependant pouvoir les remercier, tout comme les personnes qui ont contribué de façon plus indirecte à l'aboutissement de ce manuscrit.

Je tiens donc premièrement à remercier mes deux directeurs de thèse : Christophe Douady qui, malgré ses responsabilités, a su rester présent tout en m'accordant une grande confiance et Tristan Lefébure pour son investissement au quotidien et sa patience. Merci également à Florian Malard qui, s'il n'était pas dans la direction de cette thèse, a apporté une contribution tout aussi importante. Merci à vous trois d'avoir construit la scientifique que je suis, de m'avoir transmis votre passion, votre rigueur et votre curiosité.

Je tiens à remercier les membres de mon jury pour avoir accepté d'évaluer ce travail : Guillaume Achaz et Sylvain Glémin pour leurs retours sur le manuscrit, Dominique Mouchiroud pour avoir accepté de présider la soutenance, et Sarah Samadi pour avoir examiné ce travail. Un sincère merci à tous les quatre pour les échanges scientifiques autour de mon travail mais aussi pour votre flexibilité face aux grèves qui ont entraîné le report de la soutenance.

Je remercie également les membres de mon comité de pilotage de thèse. Merci à Nicolas Lartillot pour sa grande disponibilité et son aide technique et scientifique tout au long de ces trois ans. À Marie Sémon pour son écoute et ses conseils avisés, et pour avoir été un modèle de réussite et d'humilité en tant que femme de science. À Laurent Duret pour son regard constructif et ses remarques toujours extrêmement pertinentes sur mon travail. Merci à Nicolas Galtier pour avoir apporté son regard extérieur, pour sa pédagogie, et pour m'avoir offert l'opportunité de continuer à faire de la recherche en post-doctorat, et à Benoît Nabholz pour ses retours honnêtes et pertinents sur mon travail.

Je tiens à remercier chaleureusement Lara Konecny qui a généré la grande majorité des données moléculaires utilisées dans cette thèse, me permettant de travailler avec des données d'une rare qualité, et qui a également encadré les stagiaires pour les étapes de manipulations en laboratoire, toujours avec patience, humour et pédagogie. Merci également à Colin Issartel pour le travail de dissection fastidieux et pourtant très rigoureux qu'il a effectué.

Merci à toutes les personnes avec qui j'ai pu travailler au cours de cette thèse et qui m'ont apporté leurs compétences et leur aide au laboratoire ou sur le ter-

---

rain : David Eme, Clémentine François, Patrick Chardon, Vincent Breton, Arnaud Bellec, Emilien Luquet, Sandrine Plenet, Cyril Degletagne, Sylvie Bodino, Jean-Claude Caissard, Gilles Escarguel, Julien Clavel, Bernard Kaufmann, Marcel Meyssonier, Audric et Aymeric Berjoan, Josiane et Bernard Lips et Audrey Brechet.

Je remercie également sincèrement tous les membres de l'équipe E3S et du LEHNA qui m'ont offert un environnement de travail enrichissant. Merci en particulier à mes collègues de bureau qui ont toujours su entretenir une ambiance joviale et bienveillante : Ivan Paz-Viñas, Mailys Gauthier, Laura Grice et Héloïse Verdier. Merci à ceux qui m'ont facilité la vie administrativement : Nadjette Houriez, Dominique Martin, Pierre Marmonier, Nadine Brochet, Nathalie Lyvet et Marilynne Brennat.

Et merci à tous les collègues thésards, post-doctorants et stagiaires anciens et actuels : Natacha Foucreau, Julia Desprat, Adeline Dumet, Mathilde Pigneret, Jeremy Voisin, Valerian Estragnat, Florian Vautrin, Antoine Gosset, Marion Cordonier, Mylène Dutour, Mélanie Boël, Mathieu Mortz, Morgane Touzot, Louise Barberis, Pauline Guenser, Juliette Tariel, Hugo Basquin, Yohan Lebon, Elisa Thorat, Marie Cariou, Mohammad Wazne, Samuel Mouron, Camille Touchet, Simon Navel, Nicolas Noclin, Samuel Ginot, Jules Segrestin, Kiana Lagadou, Juliette Berger. . . et toutes les personnes que j'aurais pu oublier.

Merci à mes parents pour m'avoir appris à toujours me surpasser, à mes sœurs et mon frère pour leur solidarité et à Jonathann pour son humour, son honnêteté et son soutien sans faille au quotidien.

# Table des matières

Table des matières . . . . .	v
Table des figures . . . . .	ix
Liste des tableaux . . . . .	xi
<b>Introduction</b>	<b>1</b>
<b>1 Cadre conceptuel et objectifs</b>	<b>7</b>
1 Origine des variations du taux de substitutions inter-spécifiques. Que dit la littérature? . . . . .	8
1.1 Facteurs affectant le taux de mutations . . . . .	9
1.2 Niveau populationnel : variation de la probabilité de fixation	18
1.3 Compromis évolutifs . . . . .	21
1.4 Démêler les hypothèses . . . . .	26
2 Difficultés et lacunes dans l'étude des taux d'évolution moléculaire	31
3 Objectifs de la thèse . . . . .	34
<b>2 Tester les hypothèses relatives aux taux d'évolution moléculaire</b>	<b>37</b>
1 Estimer les taux d'évolution moléculaire . . . . .	39
2 Les différentes approches comparatives . . . . .	44
2.1 Approches en deux étapes . . . . .	45
2.2 Approche de modélisation de la co-évolution entre traits et taux de substitutions . . . . .	49
3 Approche choisie dans cette thèse . . . . .	51
<b>3 Les Asellidae souterrains : développement du modèle pour l'étude des taux d'évolution moléculaire.</b>	<b>53</b>
1 Le modèle biologique des Asellidae . . . . .	54
1.1 Impact de la transition vers le milieu souterrain chez les Asellidae . . . . .	55
1.2 Milieux de vie et méthodes d'échantillonnage . . . . .	56
1.3 Délimitation des espèces d'Asellidae . . . . .	61



1.4	Phylogénie du groupe . . . . .	66
1.5	Les Asellidae : un modèle original pour étudier les taux de substitutions . . . . .	78
2	Données moléculaires . . . . .	80
2.1	Séquençage Sanger . . . . .	80
2.2	Données transcriptomiques . . . . .	82
2.3	Génomomes mitochondriaux . . . . .	84
2.4	Gènes se non-fonctionnalisant en milieu souterrain . . . . .	85
<b>4</b>	<b>Impact des traits d’histoire de vie sur les taux d’évolution moléculaire</b>	<b>93</b>
1	Introduction . . . . .	93
2	Matériel et Méthodes . . . . .	97
3	Résultats . . . . .	101
4	Discussion . . . . .	107
<b>5</b>	<b>Impact de l’environnement mutagène sur les taux d’évolution moléculaire</b>	<b>113</b>
1	Introduction . . . . .	113
2	Matériel et Méthodes . . . . .	117
3	Résultats . . . . .	125
4	Discussion . . . . .	126
<b>6</b>	<b>Existe-t-il des patrons de variations des taux d’évolution moléculaire à l’échelle globale ?</b>	<b>131</b>
1	Introduction . . . . .	131
2	Matériel et Méthodes . . . . .	136
3	Résultats . . . . .	143
4	Discussion . . . . .	148
5	Conclusion . . . . .	153
<b>7</b>	<b>Synthèse, Discussion &amp; Perspectives</b>	<b>155</b>
1	Synthèse . . . . .	155
2	Discussion . . . . .	158
3	Perspectives . . . . .	162
3.1	Tester l’existence d’une contrainte sur le taux d’évolution moléculaire en lien avec l’investissement parental . . . . .	164
3.2	Tester l’existence d’une contrainte sur le taux de substitutions en lien avec la compatibilité mito-nucléaire . . . . .	169

3.3	Existe-t-il un lien entre les taux de substitutions et les taux de diversification? . . . . .	174
	<b>Bibliographie</b>	<b>179</b>
	<b>Annexes</b>	<b>211</b>



# Table des figures

1.1	Schéma des facteurs pouvant impacter le processus mutationnel . . . . .	9
1.2	Distribution des effets de fitness . . . . .	20
1.3	Hypothèses sur les variations de taux de substitutions . . . . .	27
2.1	Exemple de l'impact de l'inertie Phylogénétique . . . . .	38
2.2	Quatre approches de mesure des taux de substitutions dans une analyse comparative . . . . .	46
3.1	Comparaison des périodes principales du cycles vital de trois Aselloidea . . . . .	56
3.2	Echantillonnage au pinceau . . . . .	58
3.3	Echantillonnage au filet . . . . .	59
3.4	Echantillonnage au filet Cvet-Kov . . . . .	60
3.5	Echantillonnage à la pompe Bou-Rouch . . . . .	61
3.6	Pléopodes II d'Asellidae . . . . .	63
3.7	Processus d'identification . . . . .	67
3.8	Phylogénie calibrée dans le temps des Asellidae . . . . .	74
3.9	Carte de répartition du genre <i>Proasellus</i> . . . . .	75
3.10	Sensibilité de chaque point de calibration (LTT plot) . . . . .	76
3.11	Effet du temps de colonisation sur l'estimation du $d_S$ . . . . .	87
3.12	Calcul du temps de colonisation . . . . .	89
4.1	Taux de substitutions des génomes nucléaires et mitochondriaux . . . . .	102
4.2	Corrélation entre le $d_S$ et le temps de colonisation . . . . .	103
4.3	Corrélation entre le $d_S$ et le ratio ARN/protéine . . . . .	104
4.4	Effet des UVs : Zscores . . . . .	105
4.5	Impact de la différence de puissance dans les deux génomes . . . . .	106
4.6	Hypothèses de découplage de la réplication des deux génomes . . . . .	110
5.1	Carte échantillonnage radioactivité . . . . .	118
5.2	Phylogénie et carte des espèces échantillonnées . . . . .	120

TABLE DES FIGURES

---

5.3	Corrélations taux de substitutions - radioactivité . . . . .	127
5.4	Spectre mutationnel en lien avec la radioactivité . . . . .	129
6.1	Relation taux de substitutions - temps de divergence . . . . .	145
6.2	Distribution des branches courtes en fonction de la latitude . . . .	149
6.3	Points retirés de la corrélation taux de substitutions - latitude . .	149
6.4	Time-dependency : surestimation des taux de substitutions dans le temps récent . . . . .	152
7.1	Phylogénie des individus de l'espèce <i>Physsa acuta</i> . . . . .	171
7.2	Ordre des gènes dans les génomes mitochondriaux des 4 mitotypes de <i>Physsa acuta</i> . . . . .	172
7.3	Comparaison des gènes nucléaires, mitochondriaux et des gènes en collaboration chez <i>Physsa acuta</i> . . . . .	172
7.4	Taux de diversification le long de la phylogénie des Asellidae . . .	176

# Liste des tableaux

1.1	Attendus des hypothèses relatives aux taux de substitutions . . . .	28
3.1	Mix utilisé pour l'amplification par PCR . . . . .	81
3.2	Cycles du thermocycleur utilisés pour l'amplification par PCR . .	81
5.1	Résultats des tests PGLS de la corrélation entre radioactivité et taux de substitutions . . . . .	126
5.2	Résultats des tests PGLS de la corrélation entre radioactivité et spectre mutationnel . . . . .	128
6.1	Nombre d'espèces dans les jeux de données . . . . .	144
6.2	Résultats des tests PGLS entre taux de substitution et environnement	147
6.3	Résultats des tests PGLS entre taux de substitution et environne- ment avant nettoyage des données . . . . .	148



# Introduction

Dans les années soixante, l'analyse des premières séquences protéiques complètes chez plusieurs espèces a permis à plusieurs auteurs (Zuckerlandl et Pauling, 1962; Margoliash, 1963; Doolittle et Blombäck, 1964) de faire le constat que le nombre de différences observées sur une protéine entre deux espèces semblait correspondre au temps de divergence entre ces deux espèces. Après avoir montré une accumulation à vitesse constante des mutations sur les protéines de l'hémoglobine chez plusieurs espèces de mammifères, Zuckerlandl et Pauling proposent en 1965 l'idée d'utiliser les données moléculaires pour dater la divergence entre les espèces. Ils proposent ainsi que les mutations sur les protéines s'accumulent à vitesse constante au cours du temps et dans toutes les lignées, cette vitesse étant en revanche, propre à chaque protéine. Sous cette hypothèse, on peut déduire de la distance génétique entre deux espèces, le temps qui s'est écoulé depuis leur spéciation en utilisant une donnée temporelle (*e.g.* fossile) pour calibrer la vitesse d'accumulation des mutations au cours du temps. Ainsi, avec un seul point de calibration et une seule protéine, on pourrait calibrer une horloge moléculaire permettant de dater les événements de spéciation. Cette hypothèse d'horloge moléculaire a suscité un vif intérêt dans la communauté scientifique car elle offre un outil puissant permettant de dater des événements évolutifs pour lesquels les données fossiles ne sont pas disponibles.

Historiquement, l'hypothèse d'horloge moléculaire intervient en plein milieu du débat entre sélectionnistes et neutralistes. Au début des années soixante, deux visions s'opposent pour expliquer le maintien du polymorphisme dans les popu-



lations : une vision purement sélectionniste expliquant le maintien de plusieurs allèles par sélection naturelle (*e.g.* principe de la sélection balancée), et une vision neutraliste, principalement défendue par Motoo Kimura, expliquant ce maintien par un effet totalement stochastique. Malgré le fait qu'Emile Zuckerkandl était plutôt en faveur de la vision sélectionniste, l'hypothèse d'horloge moléculaire a été un argument majeur en support de la théorie neutraliste de l'évolution (Kimura, 1979). Dès 1968, Kimura utilise les résultats de Zuckerkandl et Pauling pour montrer que les mutations s'accumulent au cours du temps à un taux proche de leur taux d'apparition, suggérant ainsi que la sélection naturelle n'élimine ou ne favorise que très peu d'entre elles et que ces mutations sont donc, pour la plupart, neutres. En 1971, Kimura et Ohta proposent que les différences de taux d'évolution moléculaire observées entre les gènes codant pour des protéines proviennent du nombre de changements neutres possibles sur chaque gène. Ainsi, les gènes qui auraient un nombre de changements neutres possibles très faible comme la cytochrome c, évolueraient très lentement, contrairement à des gènes dont la fonction permettrait un plus grand nombre de changements neutres. D'autres articles confirmant la constance des taux d'évolution moléculaire entre les espèces ont ensuite appuyé l'hypothèse d'horloge moléculaire (Sarich et Wilson, 1967; McLaughlin et Dayhoff, 1970) mais des déviations importantes ont rapidement été mises en évidence (Laird et al., 1969). L'essor du séquençage d'ADN et la possibilité de mesurer des taux de substitutions synonymes a, dans un premier temps, confirmé ces exceptions à l'horloge moléculaire, en montrant notamment une accélération importante du taux d'évolution chez les rongeurs et un ralentissement chez les hominidés (Britten, 1986; Li et al., 1987). Les évidences de variations de taux d'évolution entre les lignées se sont ensuite multipliées (Li, 1993), remettant alors totalement en cause l'hypothèse d'horloge moléculaire. Des variations des taux de substitutions ont ainsi été montrées à toutes les échelles : entre positions d'un gène (Kimura, 1983), entre différentes régions chromosomiques (Wolfe et al.,

1989), entre types de génomes (Wolfe et al., 1987) et entre lignées (Ohta, 1993; Martin et Palumbi, 1993).

L'observation d'écart à l'horloge moléculaire entre les lignées a rapidement suscité l'intérêt de la communauté scientifique. En 1986, Britten propose que le ralentissement chez les hominoïdes et l'accélération chez les rongeurs seraient dû à des différences de sélection sur l'efficacité de la réplication ou de la réparation de l'ADN. Mais rapidement Li et al. (1987) proposent une explication alternative en accord avec la théorie neutraliste en réutilisant la conclusion de Laird et al. (1969) qui avaient à l'époque été critiqués sur les temps de divergence utilisés : les différences s'expliqueraient uniquement par le temps de génération, c'est-à-dire la fréquence à laquelle l'ADN est répliqué pour former la génération suivante. Il y aurait donc bien une accumulation constante des mutations, mais par génération et non par unité de temps. Par la suite, Schlotteröer et al. (1991) suggèrent que le faible taux d'évolution moléculaire des cétacés pourrait s'expliquer par une exposition plus faible aux radiations. L'année suivante, Martin et al. (1992) expliquent les différences observées entre les primates et les requins par l'effet du métabolisme qui produirait des agents mutagènes. D'autres hypothèses se sont ajoutées par la suite, faisant appel à la longévité (Nabholz et al., 2007), à la salinité (Herbert et al., 2002) ou encore à la taille de population (Lynch, 2010a). Aujourd'hui, à l'exception du temps de génération, aucune de ces hypothèses ne fait consensus. Déterminer les facteurs qui contrôlent les variations de ce taux est ainsi un véritable défi qui anime les débats scientifiques depuis la découverte des déviations à l'horloge moléculaire.

*In fine*, derrière tous ces facteurs, se pose la question de l'influence relative des trois grandes forces -mutation, sélection, dérive- dans l'évolution des espèces. Cette question qui a pris ses racines dans le débat entre sélectionnistes et neutralistes, reste une question centrale en biologie évolutive dont l'objectif est d'identifier les facteurs qui ont forgé la diversité des organismes que l'on observe aujourd'hui.

Ceci explique l'intérêt porté au taux d'évolution moléculaire et à leur lien avec les patrons de biodiversité (Gillman et Wright, 2013), la dynamique des populations (Nikolaev et al., 2007), la diversification (Duchene et Bromham, 2013) ou encore les interactions biotiques (Lutzoni et Pagel, 1997). Un aspect central de l'étude des taux d'évolution moléculaires est leur lien avec les traits d'histoire de vie (Gillooly et al., 2005; Thomas et al., 2010), qui permet d'aborder la question plus globale de la rétroaction de l'histoire de vie des espèces sur leur trajectoire évolutive. Ainsi, la compréhension des facteurs faisant varier les taux d'évolution moléculaire occupe une place centrale en biologie évolutive et permet de faire l'interface entre les processus évolutifs et l'écologie des organismes.

Le point de blocage scientifique majeur dans l'étude des taux d'évolution moléculaire consiste à différencier tous les facteurs et à quantifier leur importance relative. L'absence de consensus malgré la littérature foisonnante suggère que les taux d'évolution moléculaire ne sont pas contrôlés par un seul facteur majoritaire. D'autre part, la plupart de ses hypothèses ne sont pas mutuellement exclusives. Cependant, l'établissement de l'importance relative de ces hypothèses se heurte à plusieurs difficultés : premièrement, de nombreux facteurs covarient, rendant leur déconvolution difficile, deuxièmement il existe des difficultés méthodologiques inhérentes à l'étude des taux d'évolution moléculaire et troisièmement, les taux d'évolution moléculaire se calculant sur des temps très longs, l'expérimentation *in vitro* permet difficilement de tester les différentes hypothèses. Ainsi, il faut trouver *in naturae* les conditions permettant de tester les différentes hypothèses.

L'enjeu de cette thèse est ainsi d'identifier les facteurs à l'origine de variations des taux de substitutions et d'évaluer leur importance relative. Afin de répondre aux différentes difficultés soulevées, je ferai dans le premier chapitre, un bilan des connaissances sur les taux d'évolution moléculaire, en m'attachant à apporter des éléments permettant de faire la part des différentes hypothèses de la littérature et aux moyens de les tester. Je présenterai ensuite, dans le deuxième chapitre,

les méthodes permettant de tester les corrélations entre les taux d'évolution moléculaire et d'autres facteurs en allant de la construction du jeu de données aux méthodes statistiques, en passant par les méthodes de calcul des taux d'évolution moléculaires. Le chapitre 3 présentera ensuite le modèle biologique des Asellidae utilisé dans cette thèse et les données acquises sur ce groupe pour répondre aux différents objectifs. Je présenterai ensuite dans les chapitres suivants, les résultats obtenus lors de cette thèse : dans le chapitre 4, je présenterai les résultats obtenus sur l'influence des traits d'histoire de vie des espèces sur leur taux d'évolution moléculaire ; dans le chapitre 5, l'effet de l'environnement mutagène a été testé en évaluant la corrélation entre les taux d'évolution moléculaire et la radioactivité naturelle. Ces deux premiers chapitres de résultats ont fait l'objet de publications qui sont en annexe de cette thèse et que je présenterai dans une version synthétique en français dans le manuscrit. Dans le chapitre 6, je décrirai les résultats obtenus sur les variations des taux d'évolution moléculaire en lien avec des gradients environnementaux (température, productivité). Le chapitre 7 me permettra ensuite de présenter l'impact de ses résultats sur les hypothèses de la littérature et de proposer des perspectives à ce travail.



# Chapitre 1

## Cadre conceptuel et objectifs

Le taux d'évolution moléculaire est la vitesse à laquelle les séquences d'ADN accumulent des changements. Cette vitesse dépend à la fois du nombre de mutations qui se produit au sein des individus (le taux de mutations, noté  $\mu$ ) et de l'interaction entre la sélection et la dérive génétique qui détermine si ces mutations vont persister dans la population ou disparaître (la probabilité de fixation, notée  $p$ ) (Bromham, 2009). On entend ainsi par taux d'évolution moléculaire, le taux de mutations fixées dans la population ( $\mu * p$ ), aussi appelé taux de substitutions (qui ne doit pas être confondue avec le terme substitution en génétique qui désigne le changement d'un nucléotide en un autre - *i.e.* une mutation ponctuelle dans le cadre de cette thèse). De nombreux facteurs peuvent intervenir au cours de ce processus de substitutions. Ainsi, déterminer les facteurs responsables de variations du taux de substitutions demande dans un premier temps de comprendre à quel étape du processus substitutionnel ces facteurs agissent pour pouvoir identifier leurs effets. L'objectif étant ici d'identifier les facteurs responsables de variations de taux de substitutions entre espèces, il est ensuite nécessaire de mettre en place une analyse comparative tenant compte des relations phylogénétiques entre les espèces, permettant de tester statistiquement ces variations. Cependant, l'étude

des taux de substitutions par une approche phylogénétique se heurte à de nombreuses difficultés : premièrement un grand nombre de facteurs covariant, ce qui rend difficile l'assignation d'un changement de taux d'évolution à un facteur en particulier. Deuxièmement le calcul des taux de substitutions peut présenter des difficultés affectant la puissance de l'analyse.

Dans ce chapitre, je m'attacherai ainsi à établir un état des connaissances des facteurs agissant au cours du processus substitutionnel. La deuxième partie détaillera les difficultés et lacunes qui empêchent une meilleure compréhension de l'impact des différents facteurs sur les taux de substitutions. Je présenterai ensuite comment ces difficultés ont été approchées à travers les objectifs de cette thèse.

## **1 Origine des variations du taux de substitutions inter-spécifiques. Que dit la littérature ?**

De nouvelles mutations apparaissent à chaque génération chez tous les organismes. Lorsqu'une nouvelle mutation apparaît elle peut soit disparaître, soit se répandre dans la population jusqu'à être éventuellement trouvée chez tous les individus, on dit alors qu'elle est fixée. Ainsi, les facteurs qui modifient le taux de substitutions peuvent agir aux différentes étapes entre l'apparition d'une mutation et sa fixation dans la population. Dès le début de ce processus, des facteurs mutagènes peuvent favoriser l'apparition de nouvelles mutations. Ensuite, les facteurs impactant la balance entre sélection et dérive vont déterminer si cette mutation va se fixer. Enfin, à plus large échelle temporelle, le taux de substitutions est probablement sous l'influence de compromis évolutifs entre le fardeau mutationnel (*i.e* l'accumulation de mutations délétères) et le coût de la réparation, le point d'équilibre de ce compromis pourrait dépendre de la stratégie évolutive de l'espèce (Figure 1.1).

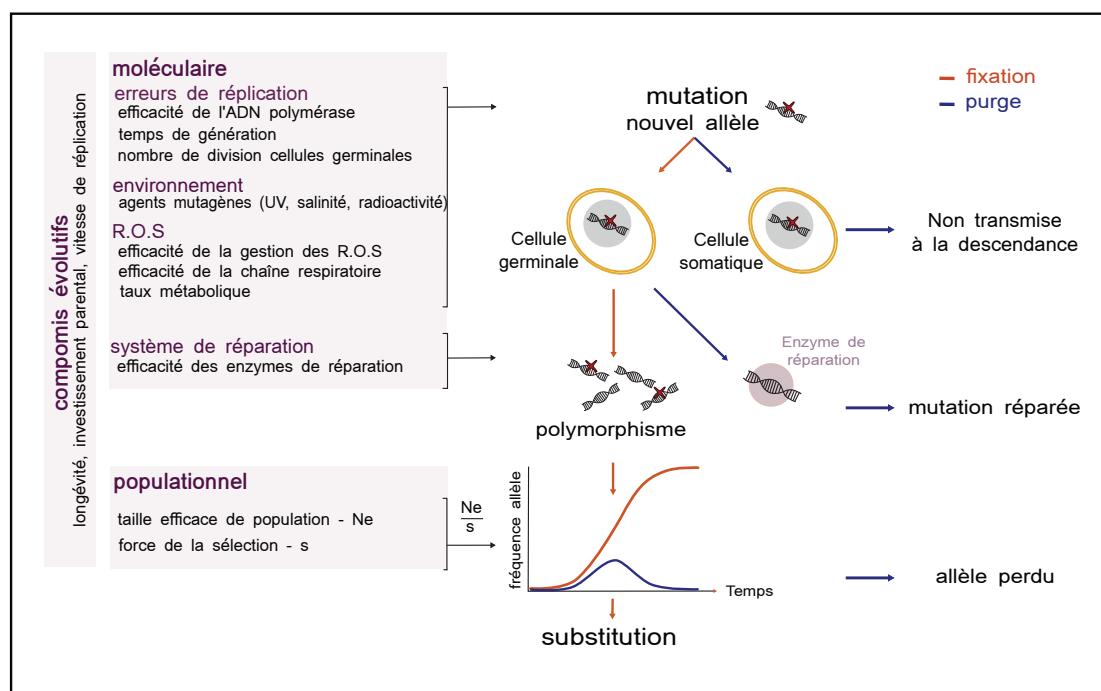


FIGURE 1.1 – Facteurs intervenant au cours du processus menant à la fixation ou à la purge d’une mutation. Les facteurs interviennent à trois étapes de ce processus : en modifiant le taux de mutations, la probabilité de fixation, ou l’équilibre du compromis entre fardeau mutationnel et coût de la réparation.

## 1.1 Facteurs affectant le taux de mutations

Une mutation génétique est un changement de l’information génétique dans les séquences nucléotidiques (ADN ou ARN). On peut distinguer plusieurs types de mutations :

- Les mutations ponctuelles changent un nucléotide en un autre (*e.g.* une base Adénine devient une Cytosine)
- Les insertions-délétions sont la perte ou le gain d’un ou plusieurs nucléotides.
- Les insertions d’éléments transposables sont l’insertion, au milieu des séquences, d’éléments mobiles autonomes : des séquences capables de se déplacer et se répliquer de façon totalement autonome dans le génome.
- Les réarrangements inter ou intra chromosomiques sont des échanges de matériel génétique entre les chromosomes ou entre les régions d’un chromosome. Ces réarrangements se produisent généralement après une mauvaise



réparation d'une cassure double-brin de l'ADN (Pfeiffer et al., 2000).

- Les duplications sont des multiplications du matériel génétique. Il peut s'agir de la duplication d'un gène, d'une portion de chromosome, voire d'un chromosome entier ou même du génome complet.

En revanche, les modifications épigénétiques ne sont pas considérées comme des mutations car, si elles peuvent changer l'activité d'un gène, elles ne modifient pas les séquences d'ADN.

Le taux de substitutions n'est estimé, dans l'immense majorité des cas, qu'à partir des mutations ponctuelles (changement d'une base en une autre base). Ainsi les autres types de mutations ne sont pas prises en compte dans l'analyse de ces taux. De plus, une substitution ne peut se produire que si la mutation est transmise à la descendance, ainsi seules les mutations germinales (*i.e.* mutations dans l'ADN des cellules germinales) ont une influence sur le taux de substitutions. Ces mutations se produisent soit en raison d'erreurs lors de la réplication de l'ADN pendant la gamétogenèse, soit en raison de dommages physico-chimiques causés à l'ADN par des agents appelés mutagènes. Des mécanismes de protection et de réparation peuvent empêcher ou réparer ces mutations, le taux de substitutions dépend donc aussi de l'efficacité de ces mécanismes (Figure 1.1).

### **Erreurs de réplication dans la lignée germinale**

Pour assurer la réplication de l'ADN lors de la méiose et de la mitose, les cellules disposent de multiples ADN polymérases (voir Hübscher et al., 2002). Ces polymérases sont des complexes enzymatiques qui utilisent un brin d'ADN comme matrice pour produire un brin d'ADN complémentaire. La fidélité de cette réplication est assurée par trois mécanismes : (i) la sélectivité de l'ADN polymérase, c'est-à-dire sa capacité à insérer la bonne base, (ii) l'activité exonucléase intrinsèque à l'ADN polymérase qui est une auto-correction permettant de faire une 'relecture' de la base précédemment insérée et de la corriger en cas d'erreur,

et enfin (iii) l'activité de réparation des mésappariements, qui est un système de correction post-réplication pour réparer les éventuelles erreurs restantes. Par exemple, chez *Escherichia coli*, le taux d'erreurs de l'ADN polymérase avant auto-correction se situe autour de  $10^{-5}$  (une erreur sur 100 000 bases Schaaper, 1993). L'activité exonucléase, permet de réduire le taux d'erreurs à  $10^{-7}$ . Après les corrections post-transcriptionnelles le taux d'erreurs passe à  $10^{-10}$ . Les erreurs de réplifications lors de la méiose (et des mitoses) sont considérées comme la source principale de mutations dans la lignée germinale. En effet, Chang et al. (1994) montrent que, chez les rongeurs, le ratio du taux de mutations mâle/femelle est équivalent au ratio du nombre de divisions méiotiques mâle/femelle. Ce résultat suggère que, chez ces organismes, la majorité des mutations se produit lors de la réplication. Cette prévalence des mutations générées lors de la réplication a ensuite été montrée chez des organismes très éloignés phylogénétiquement allant de l'être humain (Kong et al., 2012) aux champignons (Gangloff et al., 2017), suggérant qu'il s'agit d'une règle universelle. Il faut cependant noter que même si elles sont minoritaires, les mutations hors réplifications existent et pourraient avoir un effet important sur les trajectoires évolutives car elles sont différentes des mutations se produisant lors de la réplication (Gangloff et al., 2017; Achaz et al., 2019). Ainsi, si la majorité des mutations se produit lors de la réplication, le taux de substitutions devrait dépendre fortement de deux paramètres : le taux d'erreurs - non corrigées - lors de la réplication et le taux de réplifications des cellules germinales (nombre de réplifications par unité de temps).

**Taux d'erreurs lors de la réplication** Le taux d'erreurs lors de la réplication est variable selon les espèces. Nous avons vu que chez *Escherichia coli*, ce taux se situe autour de  $10^{-10}$ . Mais déjà au sein de cette espèce, on observe des variations importantes avec des phénotypes 'mutateurs' dont le taux d'erreurs montent à  $10^{-5}$  (Matic et al., 1997). Ces différences chez les bactéries proviennent généralement de mutations au niveau des gènes *mutS* et *mutL* qui sont impliqués dans

la réparation des mésappariements après la réplication (Oliver et al., 2002). A plus large échelle taxinomique, on observe de fortes variations de la fidélité de la réplication. Alors que les virus à ARN montrent des taux d'erreurs de  $10^{-4}$  (Drake, 1993), le taux de mutations chez l'humain se situe autour de  $10^{-8}$ , avec des variations en fonction de l'âge des parents (Kong et al., 2012). Ces variations de la fidélité de la réplication sont majoritairement expliquées dans la littérature comme étant le résultat d'un compromis entre le coût de la fidélité de la réplication et le fardeau mutationnel (Dawson, 1998, voir partie 1.3).

**Taux de réplication de la lignée germinale** Le taux de réplications de la lignée germinale dépend, quant à lui, de deux paramètres : (i) le nombre de réplications subies par les cellules germinales pour donner de nouvelles cellules germinales et (ii) le temps moyen qui s'écoule entre une génération de cellules germinales et la suivante. Ce temps correspond au temps de génération.

**Nombre de réplications des cellules germinales** Le nombre de réplication pour produire un gamète est extrêmement variable. Des différences s'observent même entre mâles et femelles d'une même espèce. Il a été montré chez l'Homme et certains oiseaux que le nombre de divisions pour produire un spermatozoïde est plus important que pour un ovule : les mâles génèrent donc un plus grand nombre de mutations que les femelles (Haldane, 1946; Li et al., 2002). Ces différences viennent du processus de gamétogenèse qui est différent entre les sexes et selon l'espèce. Par exemple les mammifères femelles produisent tout leur stock d'ovocytes durant la phase embryonnaire. C'est-à-dire que l'individu ne produit des ovocytes que pendant une courte période embryonnaire et n'en produira plus ensuite. En revanche, la spermatogenèse se fait de façon continue, amenant à de nombreuses divisions de mitoses avant les divisions de méiose, et donc à un plus grand nombre de divisions. Chez les protostomiens, la production semble se faire tout au long de la vie de l'animal, de la même façon que les mammifères

mâles, dans les deux sexes (Charniaux-Cotton, 1973). Chez les plantes, ce nombre de divisions est un trait extrêmement plastique qui peut varier fortement entre générations et entre espèces proches (Petit et Hampe, 2006). Les mécanismes de différenciation de la lignée germinale chez les plantes sont encore assez mal connus. Il a souvent été admis que cette différenciation ne se faisait que tardivement à partir de cellules somatiques, cependant cette hypothèse repose sur peu de données (voir review de Lanfear, 2018). Une étude récente (Wang et al., 2019) a montré que le fraisier présente des taux de mutations dans les différents compartiments de la plante compatibles avec une ségrégation précoce de la lignée germinale, avec des mutations moins nombreuses dans les parties reproductrices. Cependant, toutes les autres plantes de l'étude présentent une ségrégation tardive de leur lignée germinale, suggérant ainsi qu'il n'existe pas de règle universelle. Chez les plantes à ségrégation tardive, des mutations somatiques devraient être transmises à la descendance. Lanfear et al. (2013) ont proposé que les variations du taux de substitutions chez les plantes s'expliquent principalement par le taux de mitoses, c'est-à-dire le nombre de divisions de mitoses par unité de temps subies par les cellules somatiques qui donneront la lignée germinale.

**Temps entre deux générations de cellules germinales** Le temps de génération correspond au temps moyen qui s'écoule entre deux générations (Coale, 1972). L'hypothèse du temps de génération (Li et Tanimura, 1987) prévoit que le taux de substitutions est plus élevé chez les espèces à temps de génération court, car elles répliquent leur ADN plus de fois que les espèces à temps de génération long sur un même intervalle de temps. L'effet du temps de génération a été décrit chez les mammifères (Ohta, 1993), les oiseaux (Mooers et Harvey, 1994), les arthropodes et mollusques (Thomas et al., 2010) et même chez les plantes (Smith et Donoghue, 2008). Chez ces dernières, cet effet est encore mal compris car si la ségrégation de la lignée germinale se fait très tardivement à partir de lignée somatique, le nombre de mutations présentes dans la descendance devrait dépendre

principalement du taux de mitoses. Lanfear et al. (2013) font l'hypothèse que cette différence de taux de substitutions entre les plantes à court et à long temps de génération pourrait être liée à la covariation de ce temps de génération avec le taux de mitoses. C'est à dire que les plantes à temps de génération court répliqueraient leurs cellules somatiques plus rapidement que les plantes à temps de génération long.

### **Impact des agents mutagènes**

Les organismes sont au contact d'agents mutagènes qui peuvent être extrinsèques (*i.e.* provenant de l'environnement) ou intrinsèques (générés par le métabolisme). De nombreux agents extrinsèques dégradent l'ADN de façon physique (*e.g.* rayonnements ionisants) ou chimique (*e.g.* la salinité). Certains de ces agents ont des effets bien définis sur l'ADN. Par exemple, les rayonnements ultraviolets (UV) sont connus pour provoquer des torsions de l'ADN par photo-excitation des sites dipyrimidiques (*i.e.* composés de cytosines et/ou de thymine), empêchant ainsi la transcription (Setlow et Carrier, 1966). Afin de pouvoir transcrire cet ADN endommagé des polymérase moins fidèles sont utilisés par la machinerie cellulaire. Le fonctionnement de ces polymérase provoque un changement de bases, remplaçant les nucléotides endommagés par des thymine. Ainsi, une mutation caractéristique des UV est le passage de CC vers TT (Brash, 2015). En revanche, pour la plupart des autres agents mutagènes, aucune signature caractéristique n'a été décrite, ce qui rend leur identification difficile. L'effet *in situ* des agents mutagènes « naturels » extrinsèques sur l'ADN a toutefois rarement été mis en évidence (Palmeira et al., 2006). Il existe trois explications principales à cela : premièrement, les espèces présentent des comportements d'évitement et de protection, notamment vis-à-vis des rayonnements UV. Deuxièmement, les cellules germinales sont souvent plus difficiles à atteindre pour des agents mutagènes que les cellules somatiques. Troisièmement, certains agents mutagènes tels que les mé-

taux lourds, n'ont qu'un effet indirect sur l'ADN en générant un stress au niveau des cellules qui vont alors produire elles même des agents mutagènes, il est alors difficile d'identifier l'origine de ce stress.

La production d'agents mutagènes intrinsèques à l'organisme a lieu lors du métabolisme de l'oxygène dans les mitochondries. Les réactions chimiques qui ont lieu produisent des espèces réactives de l'oxygène (ROS) qui causent des dommages aux nucléotides (Barja, 2002). La production de ces ROS est un résultat du fonctionnement normal du métabolisme. Lorsque le taux métabolique augmente, la respiration mitochondriale s'intensifie, on s'attend donc à une augmentation de la quantité de ROS produite. L'hypothèse du taux métabolique (Martin et al., 1992) prévoit ainsi une augmentation du taux de substitutions chez les espèces ayant un taux métabolique élevé. La quantité de mutations générées par le stress oxydant étant beaucoup plus élevée dans l'ADN mitochondrial que dans l'ADN nucléaire (Richter et al., 1988), on s'attend à un impact plus important du stress oxydant sur le génome mitochondrial que sur le nucléaire. Les études sur la relation entre taux de substitutions et taux métabolique ont fourni des résultats équivoques. Pour certaines, cette relation est inexistante (Mooers et Harvey, 1994; Bromham et al., 1996), alors que pour d'autres, elle est positive (Martin et Palumbi, 1993; Bleiweiss, 1998; Gillooly et al., 2005). Toutefois, la plupart des études rapportant une relation positive utilisent des proxy du taux métabolique (*e.g.* la taille corporelle ou la température de l'environnement), ne tiennent pas compte de l'inertie phylogénétique entre taxons (voir Chapitre 2) et ne concernent souvent que les mammifères. Lanfear et al. (2007) ont repris l'ensemble des études portant sur la relation entre taux métabolique et taux d'évolution moléculaire et ne montrent pas de relation. Cependant, le nombre de gènes utilisés dans cette étude pour le calcul des taux est faible et les analyses statistiques peu puissantes, suggérant que si le taux métabolique a un effet, ce n'est pas un facteur prédominant. A cela, il y a plusieurs explications : premièrement, l'activité de la chaîne respira-

toire est plus faible dans les cellules germinales que dans les cellules somatiques (Paula et al., 2013). Deuxièmement, pour un même taux métabolique, la quantité de ROS générée varie en fonction de l'efficacité de la chaîne respiratoire. Enfin, l'impact des ROS sur l'ADN dépend de l'efficacité des mécanismes de protection et de réparation (Galtier et al., 2009).

Les séquences sont donc exposées en permanence à des agents mutagènes. Cependant, il existe des mécanismes qui permettent d'empêcher ces agents mutagènes de créer des mutations. Ainsi, les espèces riches en GC seraient moins sensibles à ces agents (Boussau et al., 2008; Naya et al., 2002). En effet, les bases A-T sont reliées par deux liaisons hydrogènes, alors que les bases G-C en comptent trois, ce qui les rend plus difficiles à endommager. Pour ce qui est des agents mutagènes intrinsèques, il existe de nombreuses défenses antioxydantes (superoxyde dismutase, catalase), qui prennent en charge les ROS afin de les empêcher d'atteindre les molécules (Hulbert et al., 2007). Ainsi, les mutations dues aux ROS ne se produisent que lorsque la balance entre leur production et les systèmes d'élimination est déséquilibrée.

### **Systèmes de réparation des mutations**

L'ADN subit en permanence des dommages qu'il est nécessaire de réparer. Nous avons vu que les erreurs lors de la réplication sont corrigées par l'activité exonucléase des ARN polymérase et par le système de réparation des mésappariements, avec une efficacité très variable entre les espèces (Drake et al., 1998). Ce système est spécifique à la réplication et a la particularité de n'intervenir que sur le brin nouvellement formé. Dans ce cas particulier, les mutations se font uniquement sur ce brin et le système peut utiliser le brin matrice pour réinsérer la bonne base. Les dommages causés par les agents mutagènes peuvent créer des lésions de différents types. A chaque type de lésion correspond un système de réparation (voir Friedberg et al., 2005, pour un descriptif complet des systèmes de

réparation) :

- La réparation par réversion des dommages : Les dommages causés par les UV ou encore l'alkylation sont un des rares types de mutations réversibles. Pour ces types de mutations, il existe des mécanismes permettant simplement de revenir en arrière sans retirer de base. Par exemple, chez les Procaryotes et certains Eucaryote, il existe une photolyase capable de directement transférer un électron aux dimères de pyrimidines formés par les UV afin de réparer la lésion (qui avait été créée par l'arrachage d'un électron par les rayonnements UV).
- La réparation par excision de base (BER) intervient lorsqu'une seule base est endommagée. La base en question est retirée par une glycosylase puis une ADN polymérase remplace la base en utilisant l'autre brin comme matrice.
- La réparation par excision de nucléotides (NER) intervient lorsque plusieurs bases sont endommagés. Une endonucléase coupe la région endommagée, puis une ADN polymérase synthétise la région en utilisant le brin complémentaire comme matrice.
- La réparation par jonction d'extrémité non homologue (NHEJ) permet de réparer les cassures double brins sans avoir besoin de matrice. Elle ne restaure pas la séquence initiale mais restaure la continuité de l'ADN. Pour cela, les bases endommagées autour de la cassure sont retirées et les deux parties sont ensuite reliées. Ce système de réparation engendre généralement une délétion au niveau de la cassure.
- La réparation par recombinaison homologue intervient également dans le cas de cassure double brin. Cependant, une étape de recherche de séquence homologue permet d'utiliser un autre brin d'ADN comme matrice pour réparer la lésion.

Tous ces systèmes permettent ainsi d'assurer la stabilité de l'ADN. Cependant,



il arrive que certaines mutations ne soient pas réparées. Lorsqu'une fourche de réplication est confrontée à une lésion non réparée, soit parce qu'elles n'ont pas été détectées, soit parce que les dommages se produisent pendant la réplication dans la phase où l'ADN est simple brin, ce qui empêche la réparation, la synthèse du brin peut être ralentie, voir bloquée. Cependant l'ADN a une certaine tolérance pour les mutations et des mécanismes permettent de contourner les lésions lors de la réplication. Il existe ainsi des polymérases translésionnelles qui permettent de répliquer les zones endommagées. Contrairement aux polymérases réplcatives, ces polymérases sont capables d'ajouter des nucléotides face aux nucléotides lésés. Ces polymérases sont toutefois beaucoup moins fidèles que les polymérases réplcatives et créent ainsi des mutations lors de la réplication (voir Baynton et Fuchs, 2000, pour une review sur la synthèse translésionnelle). Ainsi, l'efficacité des mécanismes de réparation, et en particulier l'efficacité à détecter les mutations pour déclencher ces mécanismes, mais aussi la fidélité des polymérases translésionnelles pourrait influencer sur le taux de substitutions. La difficulté à caractériser l'efficacité de ces différents mécanismes fait qu'ils ont été très peu considérés dans l'étude des taux de substitutions.

## **1.2 Niveau populationnel : variation de la probabilité de fixation**

Une mutation non réparée dans la lignée germinale se transmet à la descendance et va, soit disparaître, soit augmenter en fréquence dans la population jusqu'à éventuellement arriver à fixation. La probabilité de fixation dépend de la balance entre la partie aléatoire, c'est-à-dire la dérive génétique, et les forces directionnelles, c'est-à-dire la sélection naturelle et le biais de conversion vers GC. Si l'on ignore pour le moment ce biais de conversion vers GC (nous y reviendrons à la fin de cette partie), la fixation d'une mutation dans la population dépend donc de la balance entre dérive génétique et sélection naturelle. La force de la

dérive génétique dépend de la taille efficace de population (notée  $N_e$ ). Lorsque  $N_e$  augmente, la force de la dérive génétique diminue, car l'impact proportionnel de chaque événement aléatoire qui concerne un seul individu diminue. La force de la sélection dépend du caractère avantageux ou délétère de la mutation, qu'on appelle valeur sélective (noté  $s$ ). Ainsi pour les mutations très fortement délétères ou très fortement avantageuses ( $|s| > 1/N_e$ ), la dérive aura très peu, voir aucun effet. De la même façon, pour une mutation totalement neutre ( $s = 0$ ), la sélection n'aura pas d'effet, seul le hasard déterminera si cette mutation se répandra dans la population ou non. En revanche la fixation des mutations intermédiaires, c'est à dire légèrement délétères ou légèrement avantageuses ( $s \approx 1/N_e$ ) dépend des effets combinés de la sélection naturelle et de la dérive génétique. Ainsi quand  $N_e$  décroît, une mutation légèrement délétère (ou avantageuse) peut devenir effectivement neutre. La valeur sélective d'une mutation n'est cependant pas une valeur absolue, elle dépend de la valeur de trait qui est étudié. En effet, si le trait étudié est déjà à sa valeur optimale dans le contexte dans lequel vit l'organisme, alors une mutation sera souvent délétère. En revanche si le trait est très loin de cette valeur optimale, cette même mutation pourrait avoir un effet positif. Ainsi, on peut représenter les probabilités de fixation des mutations sous la forme d'une distribution des effets de fitness (notée DFE). Cette DFE prend en compte la valeur sélective de la mutation dans les conditions actuelles (taille de population et valeur de trait, Figure 1.2).

Ainsi, on peut voir que la balance entre la sélection et la dérive va modifier la probabilité de fixation des mutations ayant un effet sur la fitness de l'individu, en revanche la fixation des mutations totalement neutres ne dépend pas du tout de la sélection naturelle mais uniquement de la dérive génétique. Cette dérive est plus forte dans les petites populations, cependant la théorie prédit que cet effet est exactement contrebalancé par l'augmentation du polymorphisme dans les grandes populations (Kimura, 1983), amenant à un taux de fixation indépendant de  $N_e$ .

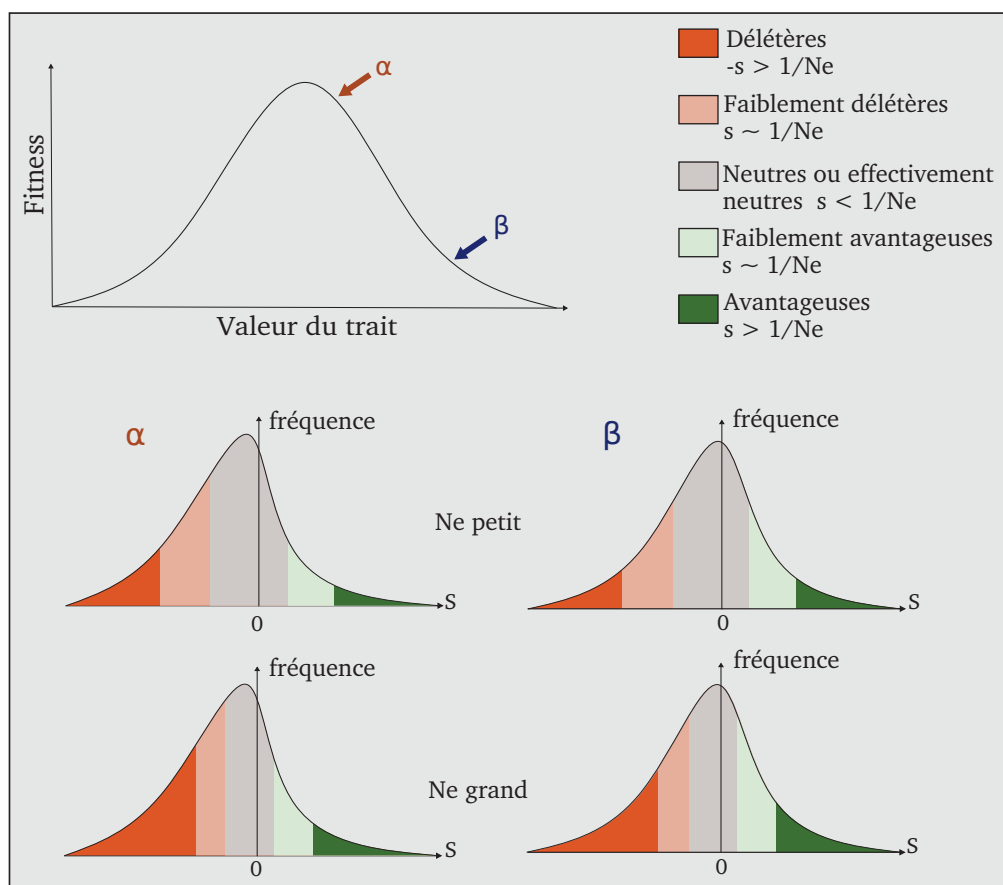


FIGURE 1.2 – Exemples fictifs de distributions des effets de fitness pour des mutations se produisant sur un trait dont la valeur est proche de l’optimum ( $\alpha$ ) ou éloignée de l’optimum ( $\beta$ ), dans une population de taille efficace ( $N_e$ ) petite ou grande. On peut voir que la valeur actuelle du trait change la distribution de la valeur sélective ( $s$ ) des mutations alors que la taille de population change l’équilibre entre la valeur sélective et la force de la dérive ( $s * N_e$ ).

Lynch (2010a) propose cependant qu’une baisse de  $N_e$  entraîne une baisse de l’efficacité de la sélection sur le contrôle des taux de mutations (efficacité de la réplication et de la réparation), augmentant le taux de substitutions neutres (voir partie 1.3).

Une autre force directionnelle qui peut influencer la probabilité de fixation est la conversion génique biaisée vers GC (gBGC). Ce biais est lié au fait que, lors de la recombinaison méiotique, un mésappariement entre une base A ou T et une base C ou G aura tendance à être converti plus souvent vers les bases G ou C. Ce

biais ne change pas la probabilité de mutations (car il ne s'agit pas d'un nouveau variant, les deux formes existaient déjà dans les cellules parentales) mais modifie la probabilité de fixation et ressemble ainsi à de la sélection directionnelle (Nagylaki, 1983). Tout comme la sélection, l'intensité de ce biais augmente avec la taille de population. Cependant, c'est un processus neutre car les bases ainsi favorisées le seront indépendamment de leur effet sur la fitness. Si une mutation vers C ou G se produit, elle aura plus de chance d'arriver à fixation qu'une mutation vers A ou T. Ainsi, le patron substitutionnel sera impacté d'une façon qui mime une sélection vers les bases GC. L'effet sur le taux de substitutions global va dépendre de la force du biais et de la composition en bases des séquences. En effet, si une séquence est riche en AT, la gBGC va faire augmenter le taux de substitutions global. A l'inverse, la gBGC fera baisser le taux de substitutions global si la séquence est riche en GC (Duret et Arndt, 2008). Les modifications de taux de substitutions seront alors accompagnées d'une modification de la composition en bases. Deux effets notables de ce biais permettent de le distinguer de la sélection naturelle : le biais est plus fort sur les régions fortement recombinantes et se fait sur toutes les positions de l'ADN, même neutres (Galtier et Duret, 2007).

### 1.3 Compromis évolutifs

Si les mutations sont 'le carburant' nécessaire pour l'évolution, ces mutations sont aussi la source de maladies, de malformations et de mortalité (Lynch, 2010b; Jackson et Loeb, 1998). Ainsi, ces mutations peuvent être à l'origine d'une réduction de la fitness, soit via la production de descendants non viables, soit via une réduction de la longévité entraînant une baisse de la fécondité. Si la majorité des mutations sont neutres ou délétères comme le prévoit la théorie quasi-neutraliste de l'évolution (Kimura, 1968; Ohta, 1973; Eyre-Walker et Keightley, 2007), les organismes devraient tous évoluer vers un taux de mutations nul. Cependant, une réduction du taux de mutations pourrait également avoir un coût. Deux prin-

cipaux coûts sont évoqués dans la littérature : le coût physiologique des mécanismes empêchant les mutations et le coût évolutif d'une évolvabilité faible. Le point d'équilibre du compromis entre ces coûts et le fardeau mutationnel pourrait dépendre des traits d'histoire de vie de l'espèce.

**Coût physiologique.** Réduire le taux de mutations demande d'augmenter l'efficacité de la machinerie de réplication et/ou de la réparation afin d'empêcher, de détecter ou de réparer plus de mutations. L'évolution d'un système parfait se heurte à deux problèmes majeurs : premièrement un système de réplication ou de relecture qui ne laisserait passer aucune erreur serait probablement très coûteux énergétiquement et/ou serait extrêmement lent, ralentissant ainsi la vitesse de réplication. Les organismes ayant besoin d'une vitesse de réplication élevée pour survivre ne pourraient donc pas évoluer vers un tel système. L'hypothèse du coût de la fidélité (Dawson, 1998) propose qu'il existe un compromis entre le coût de la réplication et le fardeau mutationnel. Selon cette hypothèse, il existe un taux de mutations évolutivement stable qui dépend des espèces. Cette hypothèse est soutenue par une étude montrant que des drosophiles prélevées dans la nature puis exposées *in vitro* à des agents mutagènes développent rapidement des adaptations pour réduire leur taux de mutations basal qui reste ensuite stable (Nöthel, 1987). Deuxièmement, il pourrait exister des effet pléiotropes sur les différents systèmes de réparation. En comparant le taux de mutations de phages ayant différents allèles 'anti-mutateurs' sur le gène codant pour l'ADN polymérase, Drake (1993) montrent que ces phages ont des taux de mutations équivalent au type 'normal', suggérant que lorsque l'efficacité de la polymérase est améliorée pour certaines mutations, elle est réduite pour les autres types de mutations.

**Coût évolutif.** Une réduction du taux de mutations implique également un plus petit nombre de changements sur lesquels l'évolution (sélection naturelle ou dérive) peut agir. Ainsi, une baisse du taux de mutations pourrait impacter la

rapidité des processus évolutifs pour les espèces. Même si d'autres mécanismes (*e.g.* l'hybridation ou les mécanismes épigénétiques) pourraient permettre une évolution sans mutations, un trop faible nombre de mutations pourrait amener à l'incapacité des populations à générer de la variabilité génétique héritable permettant de s'adapter à un environnement changeant. Plusieurs auteurs ont proposé qu'un taux de mutations élevé amènerait une diversification plus rapide (Duchene et Bromham, 2013; Lanfear et al., 2010). Cette proposition est à la base d'hypothèses controversées liant patron de répartition de la biodiversité et taux de substitutions. L'*evolutionary speed hypothesis* (Rohde, 1992) propose ainsi que les différences de température vont impacter soit directement (par une augmentation du métabolisme), soit indirectement (par des changements de temps de génération) les taux de mutations, qui vont eux même impacter la diversification, expliquant ainsi la variation latitudinale de biodiversité. L'*integrated evolutionary speed hypothesis* (Gillman et Wright, 2013) va encore plus loin en proposant que tous les patrons de répartition de la biodiversité (latitude, altitude, profondeur des océans) peuvent s'expliquer par des variations de taux de mutations ou de substitutions causées par les conditions environnementales. Ces hypothèses reposent cependant sur l'idée que la diversification serait limitée par la mutation. Alors que certains auteurs trouvent une corrélation positive entre le taux de diversification et le taux de substitutions (Barracough et Savolainen, 2001; Jobson et Albert, 2002; DugoCota et al., 2015), d'autres ne confirment pas cette corrélation (Goldie et al., 2011). Certains auteurs proposent qu'une telle corrélation pourrait s'expliquer, à l'inverse, par une augmentation du taux de substitutions suite à un événement de spéciation en raison d'une augmentation très forte de la sélection positive sur certains locus ou en raison d'une réduction de la taille de population en cas d'isolement reproducteur (Pagel et al., 2006; Venditti et Pagel, 2010). Cette hypothèse peine cependant à expliquer la relation entre taux de substitutions neutres et diversification. Lynch (2010a) propose qu'une réduction de la

taille de population pourrait amener à une augmentation du taux de substitutions synonymes par une baisse de l'efficacité de la sélection (et donc une augmentation de la dérive génétique) sur les gènes contrôlant le taux de mutations. Certaines études montrent, en effet, une relation négative entre la taille de population et le taux de mutations (Piganeau et Eyre-Walker, 2009; Lartillot et Delsuc, 2011). Cependant, Martincorena et Luscombe (2013) évaluent que les taux de mutations observés empiriquement sont à minima 2 à 3 fois supérieurs à la limite imposée par la dérive génétique seule, suggérant que l'effet de la taille de population sur les taux de substitutions neutres est probablement négligeable par rapport à la balance entre le fardeau mutationnel et le coût des mécanismes qui empêchent les mutations. Ainsi, si la présence d'un coût physiologique est bien supporté par les données, le coût évolutif d'un faible taux de mutations est encore largement discuté.

**Compromis.** Le taux de mutations serait donc au coeur d'un compromis entre le fardeau mutationnel et le coût d'un faible taux de mutations. Le point d'équilibre de ce compromis, c'est-à-dire le taux de mutations évolutivement stable, pourrait cependant ne pas être identique pour toutes les espèces. En effet, le fardeau mutationnel, tout comme le coût d'une réplication plus lente dépendent des caractéristiques des espèces. Ainsi, le taux de mutations serait un trait à part entière de la stratégie évolutive d'une espèce, dont la valeur optimale dépend de ses autres traits d'histoire de vie. Deux traits semblent prépondérants dans ce compromis : le nombre de descendants et la longévité.

La corrélation positive montrée par Welch et Waxman (2008) entre taux de substitutions synonymes nucléaires et la fécondité chez des mammifères semble indiquer que les espèces pour lesquelles les individus produisent peu de descendants transmettent moins de mutations. Ce résultat appuie l'hypothèse de Britten (1986) qui propose que pour une espèce ayant un petit nombre de descendants, la

sélection devrait favoriser l'investissement parental, c'est-à-dire la production de jeunes de 'bonne qualité'. En effet, la mort de l'un d'entre eux représentant une proportion importante des descendants, maintenir un faible taux de mutations évite ainsi de produire une quantité importante de jeunes ayant des mutations délétères. Ainsi, les espèces ayant peu de descendants devraient subir des pressions de sélection sur les gènes impliqués dans le contrôle du taux de mutations (réplication, réparation, protection).

La relation négative entre la longévité et le taux de substitutions (Martin et Palumbi, 1993; Hua et al., 2015) a longtemps servi de support à l'hypothèse du taux métabolique (voir partie 1.1). Les défauts de cette hypothèse, ainsi que les échecs de démonstration empirique (Lanfear et al., 2007) ont amené Nabholz et al. (2007) à proposer l'hypothèse de longévité. Cette hypothèse se base sur les théories qui lient le vieillissement des organismes au fonctionnement de la mitochondrie. Harman (1956) est le premier à proposer que les radicaux libres (ou espèces réactives de l'oxygène, notées ROS) produits par la respiration mitochondriale provoquent le vieillissement de l'organisme. Ces ROS produits par le métabolisme de l'oxygène vont endommager la plupart des molécules dans les cellules, aussi bien les lipides membranaires (Esterbauer, 1993) que les molécules d'ADN (Ames, 1989). Ainsi, l'accumulation des dommages au cours de la vie de l'organisme va peu à peu entraîner sa sénescence. Aujourd'hui, la théorie du vieillissement mitochondrial (voir Kujoth et al., 2007, pour une review complète) propose que les ROS vont endommager les molécules dans les cellules et notamment les molécules d'ADN qui codent pour le fonctionnement de la chaîne respiratoire mitochondriale. Cette chaîne respiratoire va ainsi perdre en efficacité et produire plus de ROS. L'organisme entre donc dans un cercle vicieux qui va l'amener graduellement vers la sénescence. La rapidité de ce processus déterminerait ainsi la longévité des organismes. Cette rapidité dépend de la quantité de ROS produits par la chaîne respiratoire mitochondriale, mais aussi de l'efficacité



des mécanismes de protection et de réparation des dégâts causés par les ROS. Ainsi, Nabholz et al. (2007) ont proposé que chez les organismes longévifs, il y a une pression de sélection forte pour maintenir des systèmes de réparation et de protection contre les ROS afin de ralentir la vitesse du vieillissement et d'éviter la production de mutations délétères avant la reproduction de l'animal. En supposant que ces systèmes dans les cellules somatiques sont les mêmes que dans les cellules germinales, les organismes longévifs devraient avoir un taux de mutations germinal également plus faible.

## 1.4 Démêler les hypothèses

Les hypothèses concernant les variations de taux de substitutions sont nombreuses et il semble difficile d'évaluer leur importance respective. En effet toutes ces hypothèses cherchent à expliquer les patrons de variations de taux de substitutions observés et, en particulier, les relations entre les traits d'histoire de vie et le taux de substitutions. Les traits d'histoire de vie étant fortement corrélés entre eux, mais aussi avec d'autres paramètres tels que le taux métabolique ou la taille de population, de nombreuses hypothèses décrivent le même patron mais avec un mécanisme sous-jacent différent. Ainsi, très peu d'entre elles sont consensuelles dans la communauté scientifique. Si on s'intéresse aux principales hypothèses (Figure 1.3), on peut voir par exemple que trois hypothèses sont fondées sur l'idée d'une balance entre les coûts des mécanismes empêchant les mutations et les bénéfices d'un taux de mutations faibles (hypothèse de l'investissement parental, hypothèse de Lynch et l'hypothèse du coût de la fidélité).

D'autre part, on peut observer que des hypothèses différentes impliquent des traits d'histoire de vie fortement corrélés entre eux, mais selon un mécanisme très différent. Ainsi, le temps de génération et la longévité sont deux facteurs très fortement corrélés (au moins chez les mammifères) et les hypothèses relatives à ces traits devraient aboutir au même patron mais par des mécanismes très différents.

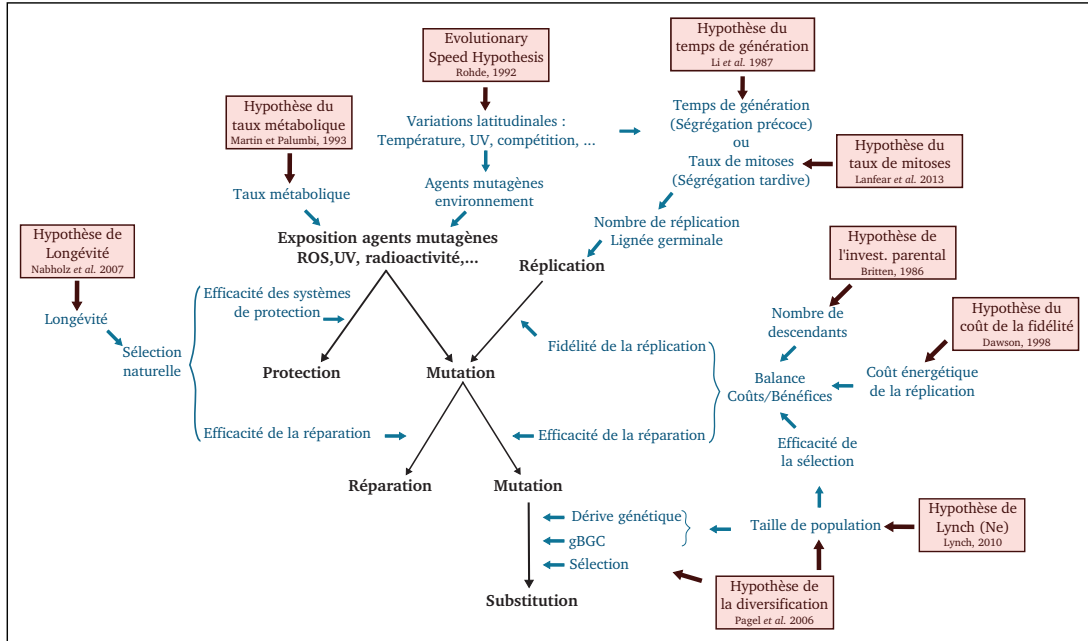


FIGURE 1.3 – Synthèse des hypothèses proposées dans la littérature expliquant les variations de taux de substitutions. Les hypothèses sont encadrées, les facteurs impliqués dans ces hypothèses sont indiqués en bleu. Ces facteurs agissent à différentes étapes du processus de substitution de l'apparition d'une mutation à la substitution.

Ainsi, il est nécessaire, pour démêler toutes ces hypothèses, d'établir des attendus théoriques plus précis. Trois moyens permettent notamment de distinguer une grande partie de ces hypothèses : (i) la comparaison entre les substitutions sur des positions neutres (substitutions synonymes) et sur des positions non neutres (substitutions non-synonymes). (ii) La comparaison entre les taux de substitutions observés dans le compartiment mitochondrial et le compartiment nucléaire. (iii) L'effet observé sur les gènes liés au taux de mutations, c'est-à-dire impliqués dans des fonctions de réparation ou de fidélité de la réplication (Table 1.1).

Le taux de substitutions synonymes (noté  $d_S$ ) correspond aux substitutions qui ne modifient pas la séquence protéique du gène. On oppose ce taux de substitutions synonymes, au taux de substitutions non-synonymes (noté  $d_N$ ). Sous l'hypothèse qu'une mutation qui ne modifie pas la séquence protéique est neutre, le taux de substitutions synonymes correspond au taux de mutations. En effet, la probabilité de fixation des mutations neutres ne dépend ni de la taille de population, ni de

TABLE 1.1 – Synthèse et attendus des hypothèses relatives aux facteurs faisant varier les taux de substitutions. Les différentes hypothèses relatives aux facteurs faisant varier les taux de substitutions peuvent se différencier en fonction de leurs prédictions sur le taux de substitutions synonymes ( $d_S$ ) ou non-synonymes ( $d_N$ ), sur l'ADN mitochondrial (mtDNA) ou nucléaire (nuDNA) et sur les gènes dont les fonctions pourraient faire varier le taux de mutations. Pour certaines hypothèses, d'autres prédictions doivent être testées.

Hypothèse	Prédiction $d_N$ VS $d_S$	Prédiction mtDNA VS nuDNA	Gènes liés au taux de mutations	Autres prédictions
Taille de population	$d_N$ uniquement	Pas de prédiction	Effet similaire au reste du génome	$d_N$ augmente dans le génome entier
Sélection naturelle	$d_N$ uniquement	Pas de prédiction	Aucun effet	$d_N$ augmente sur les gènes sous sélection
Hypothèse de Lynch (Ne)	Impact $d_N \gg d_S$	Pas de prédiction	$d_N/d_S$ augmentent sur les gènes impliqués dans la réparation et la fidélité de la réplication dans les petites populations (de la même façon que sur tous les gènes)	$d_N/d_S$ augmente sur tous les gènes
Hypothèse de longévité	Impact $d_N \leq d_S$	Effet plus important sur mtDNA (ROS)	Efficacité des gènes impliqués dans la réparation et la fidélité de la réplication corrélée à la longévité de l'espèce	
Hypothèse du taux métabolique	Impact $d_N \leq d_S$	Effet plus important sur mtDNA (ROS)	Effet similaire au reste du génome	$d_S$ corrélé à la quantité de dommages causés par les ROS dans le tissu germinale
Hypothèse du temps de génération	Impact $d_N \leq d_S$	Effet identique sur les deux compartiments	Effet similaire au reste du génome	
Hypothèse de l'investissement parental	Impact $d_N \leq d_S$	Effet identique sur les deux compartiments	Efficacité des gènes impliqués dans la réparation et la fidélité de la réplication corrélée à l'investissement parental	
Hypothèse du coût de la fidélité	Impact $d_N \leq d_S$	Effet identique sur les deux compartiments	Efficacité des gènes impliqués dans la réparation et la fidélité de la réplication corrélée au taux de croissance	
Impact d'agents mutagènes	Impact $d_N \leq d_S$	Effet identique sur les deux compartiments	Effet similaire au reste du génome	$d_S$ corrélés à la quantité d'agents mutagènes dans le milieu

la sélection (voir partie 1.2). Ainsi, les variations de taux de substitutions causées par la taille de population ou la sélection naturelle n'impactent que le  $d_N$  et pas le  $d_S$ . La sélection n'affecte généralement pas un nombre important de positions, en estimant le taux de substitutions à l'échelle du génome, c'est-à-dire sur de nombreux gènes ayant des fonctions très différentes, les taux de substitutions estimés devraient majoritairement refléter l'effet de la dérive génétique et non de la sélection. Les vecteurs de maladie constituent toutefois une exception notable : des études suggèrent que des bactéries ayant un taux de mutations global plus élevé pourrait acquérir un avantage sélectif pour échapper aux défenses de son hôte (Giraud et al., 2002). A l'inverse, une diminution radicale des pressions de sélection pourrait également entraîner une augmentation du  $d_N$  à l'échelle du génome. Ce type de changement a été montré lors de la transition vers le mutualisme, car de nombreuses fonctions, auparavant vitales, sont alors apportées par l'hôte, relâchant la sélection sur de nombreux gènes à la fois (Lutzoni et Pagel, 1997). Ainsi, il est nécessaire de s'assurer en amont que le  $d_S$  est une bonne estimation du taux de mutations. Les autres hypothèses impliquant un changement dans le taux de mutations devraient impacter le  $d_S$  et le  $d_N$ . Si la pression de sélection purifiante est forte, le  $d_N$  devrait être moins impacté. L'hypothèse de Lynch (2010a) impliquant un effet indirect de la taille de population est un peu à part car le relâchement de la sélection sur les gènes impliqués dans le taux de mutations devraient impacter le  $d_N$  et le  $d_S$  mais la réduction de taille de population associés devrait impacter uniquement le  $d_N$ . Ainsi, l'ampleur de l'effet devrait être plus importante sur le  $d_N$  que sur le  $d_S$ . Cependant, nous venons de voir que d'autres mécanismes peuvent faire augmenter le  $d_N$ . Ainsi, tester l'hypothèse de Lynch demande de vérifier qu'il y a effectivement une sélection relâchée sur les gènes impliqués dans la réparation et la fidélité de la réplication.

D'autre part, les hypothèses sur les variations de taux de mutations ne prévoient pas toutes des effets similaires sur le compartiment nucléaire et le compar-

timent mitochondrial. En effet, les ROS étant produits dans la mitochondrie, leur impact mutagène devraient être plus important sur le génome mitochondrial que nucléaire (voir partie 1.1). Ainsi, on s'attend à ce que l'effet des ROS soit plus marqué sur le taux de mutations mitochondrial que nucléaire. Les autres hypothèses prévoient un effet similaire sur les deux compartiments, cette prédiction restant toutefois à démontrer.

Une fois ces deux paramètres pris en compte, l'analyse seule des taux de substitutions ne permet plus de distinguer les hypothèses. Il est alors nécessaire d'évaluer les pressions de sélection qui s'exercent sur les gènes liés au taux de mutations (gènes de réparation, de gestion des ROS, de réplication, appelés MRg par la suite). En effet, plusieurs hypothèses sont fondées sur l'idée que la sélection naturelle pourrait agir sur l'efficacité de ces MRg (hypothèse du coût de la fidélité, de l'investissement parental, de longévité mais aussi l'hypothèse de Lynch). Un moyen d'évaluer les pressions de sélection qui s'exercent sur les MRg est de comparer le  $d_N$  et le  $d_S$ . Si le ratio  $d_N/d_S$  est proche de 0 alors le gène évolue sous forte sélection purifiante (la plupart des mutations non neutres sont éliminées), si  $d_N/d_S$  est proche de 1, le gène évolue de façon neutre, si le  $d_N/d_S$  est supérieur à 1, le gène évolue sous sélection positive (des mutations non neutres sont sélectionnées). Les hypothèses qui proposent que la sélection naturelle va entraîner une augmentation ou une baisse de l'efficacité des protéines codées par les MRg, entraînant ainsi une baisse (ou une augmentation) du  $d_S$  prévoient ainsi que le  $d_N/d_S$  de ces gènes devraient covarier avec le  $d_S$ . Cependant, l'efficacité d'un gène n'est pas forcément corrélée à son  $d_N/d_S$ , en effet on peut imaginer qu'une seule mutation puisse impacter fortement l'efficacité d'un gène. Ainsi, une meilleure façon d'évaluer les hypothèses qui sont basées sur un changement d'efficacité des MRg serait de mesurer le taux de mutations quand le gène est activé ou inactivé (CRISPR-CAS9).

Enfin, certaines hypothèses ne peuvent pas être différenciées sur la base de

traits moléculaires. En effet, différencier l'hypothèse de l'investissement parental et du coût de la fidélité demande d'estimer la corrélation entre le  $d_S$  et les traits biologiques impliqués dans ces hypothèses. Alors que l'investissement parental concerne l'énergie allouée à chaque descendant, le coût de la fidélité concerne le coût de la réplication. Selon cette hypothèse les espèces à croissance rapide devraient avoir une sélection relâchée sur le fidélité de la réplication afin de pouvoir répliquer leur ADN (et donc leur cellules) rapidement. Cependant, le taux de croissance étant généralement corrélé à l'investissement, il est nécessaire de mettre les deux facteurs dans la même analyse afin d'évaluer lequel explique la plus grande partie de la variance.

Ainsi des approches couplant physiologie, études biométriques et génétique fonctionnelle permettraient de démêler ces hypothèses. Cependant, il est nécessaire pour cela de disposer d'estimations robustes des taux de substitutions, mais également d'une approche comparative assez puissante pour distinguer ces hypothèses. Tester des relations entre taux de substitutions en relation avec des traits ou des facteurs environnementaux présente quelques particularités qu'il est nécessaire de prendre en compte.

## **2 Difficultés et lacunes dans l'étude des taux d'évolution moléculaire**

Il existe ainsi de nombreuses hypothèses quant aux facteurs affectant le taux de substitutions. A l'exception de l'hypothèse du temps de génération et de l'effet indéniable des agents mutagènes, aucune de ces hypothèses ne fait consensus dans la communauté scientifique. Et au-delà de leur validation, l'impact relatif des différents facteurs est totalement inconnu. En effet, pour la plupart des facteurs, certains articles confirment leur impact alors que d'autres non. On peut trouver plusieurs explications à ces divergences importantes : (i) la covariation

entre les traits d'histoire de vie, (ii) l'utilisation de traits actuels pour expliquer des variations de taux à long terme, (iii) l'utilisation de données moléculaires ne permettant pas une estimation précise des taux de substitutions et enfin (iv) un échantillonnage taxinomique ne permettant pas d'affirmer l'universalité des hypothèses.

**Covariation entre les traits d'histoire de vie.** Les hypothèses liées aux traits d'histoire de vie sont confrontées à la covariation importante entre les facteurs, en lien avec la stratégie évolutive de l'espèce. Ainsi, une espèce longévive a tendance à avoir un temps de génération long, un taux métabolique réduit, peu de descendants et une faible taille de population. Une corrélation entre l'un de ces facteurs et le taux de substitutions n'a donc pas valeur de démonstration. Cependant, nous avons vu qu'en testant plus formellement les attendus de chaque hypothèse sur les deux compartiments génomiques, sur les différents taux de substitutions (synonymes ou non-synonymes), ainsi que sur les gènes impliqués dans le taux de mutations, on peut faire la part de certaines de ces hypothèses.

**Différence de temporalité.** Lorsque l'on fait une corrélation entre le taux de substitutions et d'autres facteurs (traits d'histoire de vie ou environnement), un problème qui se pose est la différence temporelle entre l'estimation du taux de substitutions sur un intervalle de temps (*i.e.* une branche) et la mesure du facteur qui est immédiate. En effet, le taux de substitutions est un taux moyen calculé sur le temps que représente une branche (voir Chapitre 2, 1). Le taux de substitutions obtenu est donc le reflet de l'histoire évolutive de l'espèce. En revanche, les traits ou paramètres environnementaux que l'on corrèle à ce taux sont mesurés dans le présent. Afin de contourner ce problème, il est possible de modéliser l'évolution du trait le long de la branche et de le corrélérer à l'évolution du taux. C'est le principe du logiciel CoEvol (Lartillot et Poujol, 2011), qui permet ainsi de modéliser l'évolution de traits d'histoire de vie évoluant de façon continue. Pour les traits

phénotypiques évoluant de façon binaire, une reconstruction des changements de traits peut-être effectuée (Mayrose et Otto, 2011). Cependant, cette différence de temporalité est rarement prise en compte dans les faits et certains traits ne peuvent pas être facilement modélisés (*e.g.* changement d'habitat).

**Données moléculaires utilisées.** L'estimation des taux de substitutions présente aussi de nombreuses difficultés. Étant effectuée sur une phylogénie, elle dépend de la capacité des modèles à estimer les paramètres à partir de l'information contenue dans les données. Cette information dépend à la fois des temps de divergence entre les espèces et du nombre de marqueurs moléculaires utilisés. Le temps de divergence entre les espèces dépend de l'échantillonnage taxinomique. Lorsque les espèces sont phylogénétiquement trop éloignées, les substitutions multiples entraînent une sous-estimation du nombre réel de substitutions, par effet de saturation. Il est possible de pallier ce problème en ne comptant que les substitutions rares. Cependant, au delà des problèmes d'estimation, la pertinence biologique de comparaisons entre des espèces ayant des traits trop différents, telle qu'entre les primates et les rongeurs (Li et al., 1996) est discutable. En effet, comparer deux espèces qui diffèrent par un grand nombre de traits d'histoire de vie, ne permet pas d'assigner une différence de taux à un trait en particulier. Pour ce qui est du nombre de marqueurs moléculaires, si les études plus récentes commencent à utiliser des données génomiques, le nombre de marqueurs utilisé est généralement faible dans la littérature (entre 1 et 15). Ainsi, les estimations sont effectuées avec peu de substitutions et sont potentiellement plus liées à l'histoire des gènes utilisés qu'à celles des espèces.

**Échantillonnage taxinomique.** Enfin, à quelques exceptions près, la plupart des études ont été effectuées sur des vertébrés endothermes (mammifères et oiseaux). Ce biais taxinomique pose deux problèmes. Premièrement, il est difficile de généraliser les processus décrits à tous les animaux (et encore moins aux plantes).



Deuxièmement, certaines hypothèses sont difficiles à tester sur ces animaux. Par exemple, tester l'hypothèse selon laquelle les animaux vivant près de l'équateur auraient un taux de substitutions plus élevé sous l'effet de la température en raison du lien avec le taux métabolique (Rohde, 1992) est discutable sur des animaux endothermes.

Ainsi, il apparaît que pour faire avancer la connaissance sur les facteurs responsables des variations des taux d'évolution moléculaire, il est nécessaire (i) de mettre au point des plans expérimentaux permettant de faire la part des différents facteurs, (ii) de prendre en compte la différence de temporalité entre le temps sur lequel le taux de substitutions est calculé et la mesure des traits actuels, (iii) d'effectuer des comparaisons entre des espèces ayant des temps de divergence adéquats et n'ayant pas des phénotypes trop différents (iv) de mettre au point des jeux de données ayant un grand nombre de marqueurs moléculaires et enfin (v) de tester les hypothèses proposées dans la littérature sur des animaux ectothermes.

### 3 Objectifs de la thèse

L'objectif global de cette thèse est ainsi de tester les hypothèses principales concernant les variations de taux de substitutions entre les espèces. Pour cela, nous avons couplé les particularités des Asellidae souterrains avec des données de séquençage dans les deux compartiments génomiques. Cet objectif global est ainsi composé de trois objectifs :

**Objectif 1 : Tester le lien entre les taux de substitutions et les traits d'histoire de vie.**

La majorité des hypothèses sur les taux de substitutions s'est focalisée sur leur lien avec les traits d'histoire de vie (temps de génération, longévité, taux métabolique). Afin de déconvoluer ces différentes hypothèses, nous avons utilisé la transition vers le milieu souterrain afin de comparer des espèces ayant des traits

d’histoire de vie différents (espèces de surface versus espèces souterraines). Afin de déconvoluer les différentes hypothèses, nous avons utilisé leurs attendus différents sur les compartiments nucléaires et mitochondriaux.

**Objectif 2 : Tester le lien entre les taux de substitutions et l’environnement mutagène.**

Très tôt dans l’étude des variations de taux de substitutions, l’environnement mutagène a été proposé comme un facteur pouvant faire varier ces taux (Schlotteröer et al., 1991). Cependant, ce facteur a été très peu considéré par la suite, probablement pour des raisons principalement techniques : il est en effet très difficile de comparer des espèces soumises à des environnements mutagènes différents *in naturae*, en ayant plusieurs réplicats indépendants, et en s’assurant que les espèces vivent dans ces environnements mutagènes depuis assez de générations pour que cet environnement ait un effet mesurable sur le taux de substitutions mesuré sur des millions d’années. En utilisant des espèces souterraines, qui dispersent très peu, nous avons comparé des espèces vivant dans des habitats géologiques très variés et ayant ainsi des expositions contrastées à la radioactivité naturelle afin d’en évaluer l’impact sur les taux de substitutions. Les espèces souterraines présentent l’avantage de ne pas être exposées aux UV permettant d’éliminer l’effet de ce facteur potentiellement mutagène.

**Objectif 3 : Évaluer les hypothèses proposant un lien entre patrons de diversité à l’échelle globale et taux d’évolution moléculaire.**

En 1992, Rhode propose “L’Evolutionary Speed Hypothesis” qui explique le patron latitudinal de biodiversité par des changements de taux de mutations engendrés par les facteurs environnementaux. De nombreuses études ont cherché à tester ce lien entre les taux d’évolution moléculaire et les facteurs environnementaux variant latitudinalement, donnant des résultats parfois contradictoires mais

généralement plutôt en faveur de cette hypothèse. Cependant ces études souffrent de nombreuses lacunes, notamment le fait que très peu d'entre elles ont évalué l'effet de l'environnement sur des taux de substitutions neutres, ne permettant pas ainsi de montrer un lien avec le taux de mutations et donc de tester réellement l'hypothèse de Rohde. Le dernier objectif de cette thèse sera ainsi d'utiliser deux jeux de données : (i) la phylogénie complète des Asellidae échantillonnés à ce jour pour calculer des taux d'évolution moléculaire sur un grand nombre d'espèces (ii) 64 espèces pour lesquelles les transcriptomes ont été séquencés afin de calculer les taux d'évolution sur un grand nombre de gènes. Dans les deux cas, les espèces sont réparties sur un gradient latitudinal. Ces données seront couplées à des données environnementales (latitude, température, productivité) afin d'évaluer l'existence d'une corrélation avec les taux de substitutions (synonymes ou non synonymes).

## Chapitre 2

# Tester les hypothèses relatives aux taux d'évolution moléculaire

Lorsque l'on cherche à tester une corrélation entre deux traits chez différentes espèces, les traits sont mesurés directement sur les individus de ces espèces. La particularité des taux de substitutions est qu'ils ne peuvent pas être calculés directement sur une espèce, et sont estimés en comparant les séquences d'espèces différentes. Ainsi, estimer le taux de substitutions consiste à estimer le nombre de substitutions qui se sont produites dans une lignée sur une période de temps définie. Ce nombre de substitutions est estimé en calculant la longueur des branches le long d'une phylogénie. Tester la corrélation entre un trait et le taux de substitution consiste donc à tester la relation entre ce trait mesuré chez une espèce et le taux de substitutions calculé sur la (ou les) branche(s) menant à cette espèce. La plupart des analyses statistiques font l'hypothèse que les observations sont indépendantes entre elles. Cependant, dans le cas de comparaisons interspécifiques, le fait que deux espèces proches ont tendance à se ressembler plus que deux espèces éloignées rend les données non-indépendantes. C'est ce que l'on appelle l'inertie phylogénétique. La Figure 2.1 montre un cas extrême où la corrélation entre deux

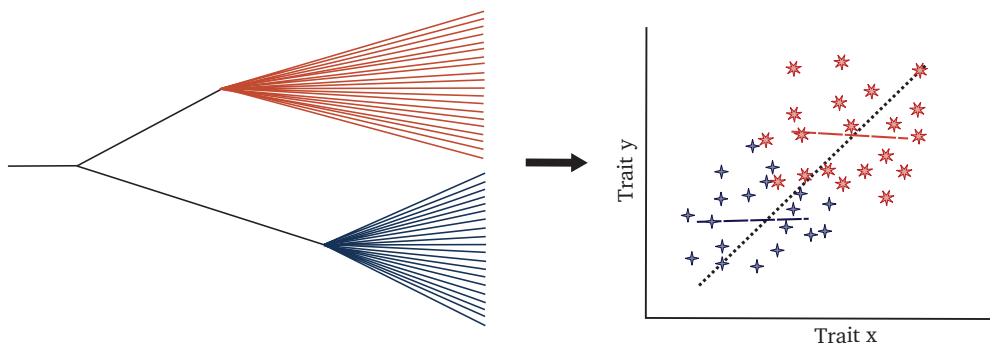


FIGURE 2.1 – Influence de l'inertie phylogénétique dans les corrélations. Une phylogénie “extrême” montrant deux clades composés chacun de 20 espèces (Gauche). Et une relation typique de deux traits évoluant de façon Brownienne le long de cette phylogénie (Droite). La ligne pointillée noire montre la relation lorsque la phylogénie n'est pas prise en compte dans l'analyse, les lignes pointillées bleue et orange montrent la relation à l'intérieur de chaque clade <sup>1</sup>.

traits provient uniquement de la structure hiérarchique des données créée par les relations phylogénétiques entre les taxons. On voit à travers cet exemple que la non prise en compte de ces relations entraîne une surestimation de la corrélation entre les traits. Les approches comparatives phylogénétiques (phylogenetic comparative methods, noté PCM) ont été inventées pour tenir compte de cette inertie dans les données. Il existe plusieurs méthodes et le choix de l'une d'entre elles dépend (i) de la question posée, (ii) des données disponibles et (iii) des connaissances sur la phylogénie des espèces étudiées. Après avoir présenté comment les taux de substitutions sont calculés, je présenterai les différentes PCM et les cas dans lesquels elles peuvent être utilisées, ainsi que les tests statistiques associés à ces méthodes.

---

1. Inspiré de “Phylogenies and the Comparative Method”, Felsenstein Joseph, volume 125 No. 1 pp. 1-15, copyright (1985); avec la permission de American Society of Naturalists, permission transmise par Copyright Clearance Center, Inc.

## 1 Estimer les taux d'évolution moléculaire

Quelque soit la méthode ou les tests utilisés, le calcul des taux de substitutions se fait toujours à partir d'un alignement de séquences et éventuellement d'une phylogénie. Il est ensuite nécessaire de choisir le modèle d'évolution et les méthodes utilisées pour estimer les taux de substitutions. Dans le cas où la datation des noeuds est nécessaire, il faut alors estimer les temps de divergence.

### Les données

Les taux d'évolution moléculaires sont estimés à partir d'un alignement de séquences. Ainsi la qualité de cet alignement est un point crucial de l'estimation. La qualité de l'alignement dépend non seulement du fait de bien aligner les positions homologues, mais également des taxons et des séquences qui constituent cet alignement. En fonction des méthodes utilisées pour estimer les taux de substitutions et des connaissances sur les relations de parenté du groupe, la phylogénie sera reconstruite de façon conjointe ou une phylogénie pourra être imposée.

Il existe de nombreux algorithmes permettant d'aligner des séquences (*e.g.* clustalw (Thompson et al., 2003), muscle (Edgar, 2004), prank (Löytynoja et Goldman, 2008), MAFFT (Katoh et Standley, 2013)). Leur performance dépend notamment du type de séquence (gène ribosomique, gène protéique) mais aussi de la divergence entre les séquences. Lorsque les séquences sont très différentes (soit parce qu'elles évoluent vite, soit parce que les taxons sont très distants), il est préférable de faire l'alignement en protéines. Il est parfois impossible d'être certain de l'homologie de deux positions, dans ce cas des outils permettent d'éliminer les sites qui montrent des variabilités trop importantes (*e.g.* Gblocks Castresana, 2000). Cependant, dans le cas de contaminations ou d'erreurs d'assemblage (chimères, décalages du cadre de lecture), les procédures automatisées ne sont généralement pas très performantes et vont soit ne pas détecter les problèmes (dans le cas de la contamination), soit retirer un grand nombre de sites en raison d'un mauvais

alignement. Idéalement, il est donc préférable de vérifier les alignements *de visu*.

La précision de l'estimation dépend, d'autre part, de l'échantillonnage taxinomique. En effet, un trop faible nombre de taxons ou un manque d'échantillonnage dans une partie de l'arbre ne permettra pas aux méthodes de reconstruction de placer les mutations sur les bonnes branches de l'arbre, impactant ainsi la qualité de l'estimation. Ainsi, il est préférable d'estimer les taux d'évolution sur le maximum de taxons possibles et de ne récupérer les taux d'évolution que pour les taxons d'intérêt plutôt que d'effectuer une estimation sur un nombre réduit de taxons. Cependant l'ajout de taxons peut amener à créer un déséquilibre du nombre de taxons entre les clades, ce qui peut créer un artefact de densité de noeuds, et peut aussi amener à insérer un grand nombre de branches très courtes dans la phylogénie. Ces courtes branches posent des difficultés pour l'estimation des taux de substitutions.

En effet, un élément important dans l'estimation du taux de substitutions est le nombre de substitutions sur lequel il est calculé. L'estimation du taux de substitution est effectuée sur un intervalle de temps (*i.e.* une longueur de branche) pendant lequel le taux de substitutions varie selon un mouvement brownien. Cette estimation est donc moins précise que la mesure directe d'un trait, car le taux de substitutions fluctue autour de la moyenne sur cet intervalle de temps. Ainsi, pour avoir une estimation précise il est nécessaire d'avoir un nombre suffisant de données. Dans le cas des taux de substitutions, la quantité de données est le nombre de substitutions sur cet intervalle de temps. Lorsque ce nombre est trop faible, soit parce que les espèces ont divergé trop récemment, soit parce que les séquences sélectionnées évoluent trop lentement ou encore parce que le nombre de positions dans l'alignement est trop faible, il y a trop peu de mutations et on a alors un problème d'incertitude autour de la mesure qui nuit à la fiabilité de l'estimation. L'estimation sur les branches très courtes est alors beaucoup plus variable que sur les branches très longues. Cette hétéroscédasticité réduit de façon importante la

puissance de l'analyse comparative. Pour éviter ce problème, Welsch et Waxman (2008) ont proposé un test permettant de définir la divergence minimale entre les séquences dans le jeu de données. Les branches terminales en dessous de ce seuil doivent être retirées de l'analyse. Afin d'éviter cette réduction du jeu de données qui peut parfois être importante, Hua et al. (2015) ont proposé l'utilisation d'une régression de Poisson plutôt qu'une régression linéaire. La régression de Poisson permet de modéliser la survenue d'événements rares (ici la mutation) dans un intervalle de temps. Inclure explicitement dans le modèle que sur un court intervalle de temps, la probabilité d'observer une mutation est faible, permet de tenir compte de la différence dans l'estimation des taux de substitutions entre les courtes et longues branches. Cependant, la comparaison d'espèces ayant divergées récemment pose aussi le problème du polymorphisme ancestral. En effet, les modèles d'évolution considèrent toutes les différences observées entre deux espèces comme étant des substitutions s'étant produites après la divergence de ces deux espèces. Si deux individus pris dans les deux espèces montrent des différences qui viennent en fait de polymorphisme dans la population ancestrale, la divergence de ces positions est en réalité plus ancienne que la spéciation. Des mutations s'étant produites avant la spéciation seront modélisées comme s'étant produites après la spéciation. Ainsi, le polymorphisme ancestral a tendance à augmenter le taux de substitutions réel (Charlesworth, 2010). Lorsque l'on compare des espèces ayant divergé depuis longtemps, ce polymorphisme représente une très faible proportion des substitutions et est donc négligeable. En revanche, lorsque les espèces comparées ont divergées récemment, ce polymorphisme peut représenter une proportion importante des substitutions. Ainsi, il est préférable d'éviter d'inclure dans les jeux de données des espèces ayant divergé récemment.

A l'inverse, un nombre trop élevé de mutations pose également des problèmes dans l'estimation des taux de substitutions. En effet, si plusieurs mutations se sont produites sur le même site de l'alignement, les méthodes de reconstruction vont



difficilement pouvoir retracer l'histoire substitutionnelle. Dans ce cas les modèles vont sous-estimer le taux de substitutions sur les longues branches de l'arbre. C'est ce que l'on appelle la saturation. Pour éviter ce problème il est nécessaire de veiller à ne pas faire des comparaisons entre espèces trop divergentes. Il est également possible de ne compter que les mutations les plus rares, en comptant par exemple uniquement les transversions qui sont plus rares que les transitions.

Enfin, la qualité de l'estimation dépendra fortement des marqueurs utilisés dans l'alignement. Premièrement parce qu'une bonne estimation des taux d'évolution demande un nombre suffisant de mutations pour faire le calcul, deuxièmement parce qu'un seul marqueur ne permet pas d'estimer le taux d'évolution moléculaire génomique. En effet, il existe de fortes variations du taux de substitutions entre les compartiments génomiques et entre les gènes d'une même espèce (*e.g.* Wolfe et al., 1987, 1989). Ainsi, pour étudier l'influence d'un facteur sur les variations de taux d'évolution, il est nécessaire de calculer ces taux sur plusieurs marqueurs, provenant idéalement de compartiments génomique différents.

## Le calcul

Les longueurs de branches sont généralement calculées avec des méthodes de maximum de vraisemblance (PhyML (Guindon et al., 2010), fastTree (Price et al., 2010), RaxML (Stamatakis, 2014), IQtree (Nguyen et al., 2015),...) ou bayésiennes (PhyloBayes (Lartillot et al., 2009), CoEvol (Lartillot et Poujol, 2011), MrBayes (Ronquist et Huelsenbeck, 2003), Beast (Drummond et Rambaut, 2007),...). La méthode du maximum de Parcimonie est à exclure des analyses de taux d'évolution car elle n'admet pas plus d'une substitution par site sur chaque branche de l'arbre, elle modélise ainsi très mal les mutations dans les séquences. Les différentes méthodes utilisent des modèles d'évolution qui font des suppositions sur le processus substitutionnel, ainsi le choix du modèle d'évolution est un paramètre important de l'estimation des taux d'évolution moléculaire. Des logiciels

permettent de déterminer le modèle correspondant le mieux aux données (*e.g.* jModelTest (Posada, 2008)). Bien que généralement le modèle le plus complexe (GTR+G+I) donne des résultats similaires au modèle le plus ajusté (Abadi et al., 2019), ce n'est pas toujours le cas (Arbiza et al., 2011). Certains logiciels ne permettent toutefois pas d'utiliser le modèle le plus ajusté.

### Calcul du temps de divergence

Selon l'approche utilisée pour tester les variations de taux d'évolution moléculaire (voir partie 2), il peut-être nécessaire de calibrer la phylogénie dans le temps, c'est-à-dire de dater les événements de spéciation. La datation d'événements de spéciation peut se faire sur la base de données moléculaires, de données fossiles, ou de données paléogéographiques (régressions marines, mouvements tectoniques).

Il existe de nombreuses méthodes de datations moléculaires (PhyloBayes (Lartillot et al., 2009), MCMCtree (Yang, 2007), Beast (Drummond et Rambaut, 2007)). Ces méthodes utilisent la vitesse d'évolution des séquences pour dater les événements de spéciations. Nous avons déjà vu que l'hypothèse d'horloge moléculaire est rarement vérifiée, les séquences n'évoluent donc pas de façon homogène le long de la phylogénie. Ainsi, ces méthodes estiment des taux différents sur chaque branche de l'arbre. Selon le modèle, ces taux peuvent être auto-corrélés (le taux estimé sur une branche dépend du taux estimé sur la branche ancestrale) ou totalement libres (non-corrélés). L'utilisation d'une horloge auto-corrélée semble plus réaliste et donne de meilleurs résultats dans la plupart des cas (Lepage et al., 2007).

Les données moléculaires seules permettent ainsi d'estimer, avec plus ou moins de précision, les taux de substitutions sur une échelle de temps relative. En revanche, elle ne permettent pas de dater sur une échelle de temps absolue (*i.e.* en millions d'années). Pour cela il est nécessaire d'inclure des données fossiles ou paléogéographiques. Les fossiles d'un âge supérieur à 1Ma n'ont pas d'ADN ex-

exploitable. Ainsi, ces deux types de données sont utilisées de façon similaire dans les phylogénies : non pas via des données moléculaires mais en positionnant sur l'arbre des points de calibration. Les données fossiles permettent d'avoir accès à des lignées disparues. L'analyse de ces fossiles permet de classer ces organismes disparus au sein de la phylogénie (avec une plus ou moins grande incertitude). Ces fossiles sont datés (*e.g.* via l'analyse de la couche sédimentaire dans laquelle ils sont retrouvés), ce qui permet de faire des hypothèses telles que : le groupe X est apparu il y a au moins N millions d'années, car un fossile du groupe X datant de N millions d'années a été retrouvé. Les données paléogéographiques s'utilisent de la même façon. Par exemple, si l'on voit qu'un clade terrestre s'est diversifié dans un lieu restreint qui était sous la mer durant une période géologique, on peut faire des hypothèses telles que : le clade X s'est diversifié au plus tôt après le retrait de la mer, il y a N millions d'années. On peut ainsi ajouter ces informations aux logiciels de datations moléculaires en indiquant des intervalles de dates pour certains noeuds dans l'arbre, permettant ainsi d'obtenir des dates absolues de divergence. On peut alors diviser les taux de substitutions obtenues sur une branche, par le temps que présente cette branche, on obtient ainsi un taux de substitutions par million d'années moyen par branche. Le logiciel CoEvol permet d'éviter l'étape de division en modélisant directement le taux par intervalle de temps. Si des points de calibrations sont intégrés dans l'analyse, la sortie est un taux par million d'années, en l'absence de point de calibration, l'âge de la racine est fixé à 1 et le taux est alors en âge relatif.

## 2 Les différentes approches comparatives

Dans le cas des taux d'évolution moléculaire, on peut distinguer deux grands types d'approches : les approches en deux étapes qui estiment les taux d'évolution sur les branches de l'arbre, ces taux sont ensuite corrélés aux traits d'intérêt ; et les approches intégratives, qui modélisent la co-évolution entre le taux de substi-

tutions et les traits d'intérêts.

## 2.1 Approches en deux étapes

### Approche par paires d'espèces

Une des approches les plus utilisées pour tester des différences de taux de substitutions est l'approche par paires d'espèces. Cette approche consiste à échantillonner deux espèces qui diffèrent pour le trait qui nous intéresse et qui partagent un ancêtre commun exclusif dans l'arbre (Figure 2.2a). Les taux de substitutions sont ensuite calculés sur les branches allant de cet ancêtre commun à chacune des deux espèces actuelles. On considère qu'à partir de la spéciation, les traits évoluent de façon indépendantes. La différence de taux d'évolution entre ces deux espèces (appelée *contraste*) est donc ensuite considérée comme une variable indépendante. On peut ainsi corrélérer le *contraste* du trait étudié (*e.g.* une différence de longévité), avec le *contraste* de taux de substitutions. Cette approche a l'avantage de permettre un traitement statistique simple avec des tests standards et ne demande pas de connaissance poussée de la phylogénie. Elle permet d'inclure dans la même analyse des paires d'espèces de groupes très éloignés. Par exemple, Thomas et al. (2010) ont utilisé cette approche qui leur a permis de tester l'influence du temps de génération sur une très grande diversité d'arthropodes sans disposer de la phylogénie complète du groupe. Cette approche a cependant le défaut de ne pas prendre en compte toute l'information contenue dans la phylogénie, ainsi l'inertie phylogénétique entre les deux membres d'une paire est prise en compte, mais pas celle entre les paires. De plus, la puissance statistique est fortement réduite car il faut deux espèces pour obtenir un point dans l'analyse. Enfin, lorsque les relations de parentés entre les taxons sont mal établies ou les données non disponibles, le calcul des taux de substitutions est effectué sur seulement 3 taxons (les deux qui forment la paire et un groupe externe), ce qui pose des problèmes d'estimation, comme nous l'avons vu dans la partie précédente. Dans cette approche, pour tes-

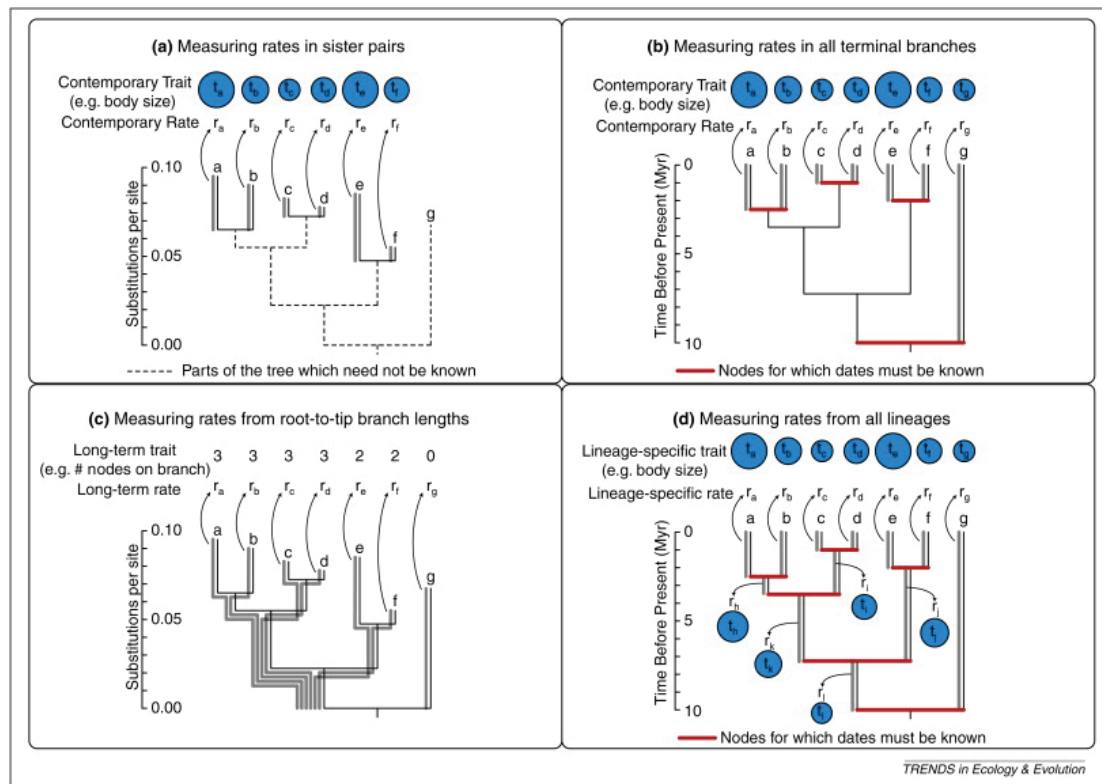


FIGURE 2.2 – Quatre type d’analyse comparative pour tester des associations entre des traits et des taux de substitutions. La phylogénie sous-jacente est identique mais la méthode diffère de par la façon dont les taux de substitutions sont estimés. Les lignes grises indiquent la partie de la phylogénie qui est utilisée pour l’estimation des taux de substitutions. L’approche par paires d’espèces (a) et par branche terminale (b) estiment les taux à partir des branches terminales de la phylogénie. L’approche par paires d’espèces (a) ne demande aucune datation, et la plus grande partie de la phylogénie n’a pas besoin d’être connue (lignes pointillées). La méthode des branches terminales (b) nécessite de connaître la phylogénie complète et de dater les noeuds terminaux (en rouge). Les méthodes ‘root-to-tip’ (c) et lignée complète (d) utilisent les taux de la phylogénie complète et nécessite que toutes les dates soient connues<sup>2</sup>.

ter la relation entre le taux de substitutions et un trait, c’est la relation entre la différence de taux d’évolution entre les deux membres des paires et la différence de traits entre ces deux membres qui est testée. Ces différences, appelées contrastes, sont considérées comme statistiquement indépendantes. Ainsi, les tests statistiques standards qui ne prennent pas en compte la phylogénie peuvent être utilisés : régression linéaire (si distribution normale), test de Wilcoxon...

2. Reprinted from Trends in Ecology and Evolution Vol 25 issue 9, Robert Lanfear, John J. Welch and Lindell Bromham, “Watching the clock : Studying variation in rates of molecular evolution between species”, Pages No. 495-503, Copyright (2010), with permission from Elsevier.

### Approche utilisant l'arbre complet

Une autre approche consiste à effectuer le calcul du taux de substitutions sur toutes les branches de l'arbre. Pour cela, la phylogénie de toutes les espèces à étudier est reconstruite. Cela nécessite donc d'être en mesure de reconstruire cette phylogénie. Ensuite, en fonction de la question posée, seules les branches terminales ou la totalité des branches peuvent être utilisées. Si l'on souhaite corrélérer les taux de substitutions aux traits actuels, alors, on utilise les branches terminales de l'arbre, qui sont alors considérées comme les taux d'évolution actuels. Cette approche nécessite que les valeurs de taux de substitutions soient comparables entre les espèces, les taux de substitutions doivent donc être exprimés par unité de temps. Les noeuds terminaux doivent donc être datés, au moins de façon relative (Figure 2.2b).

L'approche de la racine à la feuille (root-to-tip) a été proposée pour corrélérer les taux de substitutions à un processus ayant cours sur l'arbre entier (Figure 2.2c). Elle consiste à additionner les valeurs des longueurs de branches de la racine à la feuille. Elle a été utilisée dans le cadre de l'étude des taux de diversification pour tester la corrélation entre les taux de substitutions et le nombre d'événements de spéciation menant à une espèce (DugoCota et al., 2015). Cependant cette approche présente de nombreux problèmes, notamment le fait de compter de multiple fois les mêmes branches ainsi que les mêmes événements de spéciation. D'autre part cette approche est, par nature, très sensible à l'artefact de densité de noeuds. Cet artefact engendre une sous-estimation du taux de substitutions dans les zones de l'arbre où il y a moins de taxons (Fitch et Bruschi, 1987; Hugall et Lee, 2007). Il est créé par le fait que, dans les zones où il y a moins de taxons, les substitutions multiples (*i.e.* plusieurs substitutions sur le même site de l'alignement) ne peuvent pas être détectées, menant ainsi à une sous-estimation du nombre de substitutions réelles. La présence de ce biais peut, à lui seul expliquer, la corrélation entre diversification (estimé par le nombre de taxons dans un clade)

et taux de substitutions dans les études utilisant la méthode root-to-tip. Ainsi de nombreuses études utilisant cette méthode ont été invalidées ou remises en question (Hugall et Lee, 2007).

Enfin, il est possible de corrélérer les taux estimés sur chaque branche de l'arbre avec la valeur du trait estimé sur chaque branche (Figure 2.2d). Cette approche nécessite non seulement de dater tous les noeuds de l'arbre afin que toutes les mesures de taux de substitutions soient à la même échelle, mais aussi d'avoir une estimation des traits ancestraux sur chaque noeud de l'arbre. Cette approche peut-être utilisée avec des données de modélisation de traits ancestraux ou avec des données fossiles. Dans les faits, cette approche a été très peu utilisée, au profit de nouvelles méthodes de modélisation conjointe des taux et des traits.

Pour ces approches, les valeurs obtenues ne sont pas indépendantes. Il est donc nécessaire de prendre en compte l'inertie phylogénétique. Des tests ont été développés dans ce but. La première méthode qui a été proposée pour prendre en compte la phylogénie est le PIC (Phylogenetically Independent Contrasts, Felsenstein, 1985). Cette méthode calcule les contrastes sur tous les noeuds de l'arbre et permet d'obtenir des valeurs 'corrigees pour la phylogénie', indépendantes entre elles. On peut alors faire les tests statistiques standards (modèles linéaire, wilcoxon). Cette méthode suppose d'avoir une phylogénie avec des longueurs de branches connues. De plus, la méthode suppose que les traits évoluent de façon brownienne le long des branches de la phylogénie. De ce mouvement brownien, résulte une distribution normale du trait dans la phylogénie. Cependant certains mécanismes (*e.g.* comme la sélection stabilisante) rendent le modèle brownien inapproprié pour modéliser l'évolution des traits. Pour ces différentes raisons, Grafen et Hamilton (1989) puis Martins et Hansen (1997) ont proposé une généralisation de la méthode de Felsenstein permettant d'intégrer d'autres modèles d'évolution ainsi que des polytomies dans la phylogénie. Cette méthode consiste à appliquer les modèles des moindres carrés généralisés (GLS) à la phylogénie (PGLS). Une matrice car-

rée de variance-covariance est construite à partir de la phylogénie (la covariance quantifiant la ressemblance entre deux espèces et la variance, la divergence depuis leur ancêtre commun) qui est ensuite utilisée dans le modèle. Ce test s'applique toutefois uniquement à des traits évoluant de façon continue. Liang et Zeger (1986) ont introduit l'utilisation des modèles linéaires généralisés appliqués à des données corrélées (generalized estimating equations GEEs) afin de gérer des traits binaires ou catégoriels.

## 2.2 Approche de modélisation de la co-évolution entre traits et taux de substitutions

Dans les approches précédentes, les taux de substitutions sont modélisés de façon totalement indépendante des traits auxquels on cherche à les associer. Ainsi, les taux de substitutions sont estimés dans une première étape puis un test est ensuite effectué pour tester la corrélation entre ces taux et les traits étudiés. D'autres approches cherchent à reconstruire de façon conjointe l'évolution du phénotype et le taux de substitutions le long de la phylogénie. Ainsi, O'Connor et Mundy (2009) ont proposé une approche qui permet de détecter une corrélation entre un trait binaire et le taux de substitutions de certains sites. Cette méthode permet de détecter, dans les séquences, des sites associés au trait testé mais ne permet pas de tester une association entre le taux de substitutions global et le phénotype. C'est pourquoi Mayrose et Otto (2011) ont proposé une approche permettant de tester la corrélation entre le phénotype et le taux de substitutions global. Pour un caractère binaire, les changements de ce caractère d'un état à l'autre sont modélisés le long de la phylogénie grâce à des modèles d'évolution phénotypique. L'évolution des séquences est ensuite modélisée en fonction des changements d'états de caractère sous l'hypothèse que le taux de substitutions est influencé par ce caractère. Le modèle est ensuite comparé à un modèle nul dans un test de comparaison de vraisemblance. Ces approches demandent une phylogénie enracinée et datée. Pour



tester la corrélation entre les taux de substitutions et les traits, un test de rapport de vraisemblance (LRT) est effectué afin de tester la différence d'adéquation du modèle aux données entre un modèle nul (qui considère que le taux de substitutions évolue indépendamment du trait) et le modèle qui considère que le taux de substitutions évolue sous l'influence du trait.

La seule méthode permettant à ce jour d'analyser la relation entre les taux de substitutions et un trait continu a été proposée par [Lartillot et Poujol \(2011\)](#). Cette méthode permet de reconstruire les valeurs des traits et des taux de substitutions le long de la phylogénie et d'estimer conjointement la covariance entre les traits, ainsi qu'entre les traits et les paramètres du modèle d'évolution. Alors que les approches précédentes estiment les taux de substitutions comme étant la moyenne des taux de substitutions le long d'une branche, l'approche de [Lartillot et Poujol \(2011\)](#) modélise le taux de substitution instantané le long des branches de l'arbre comme un paramètre évoluant de façon continue. Cette approche modélise ainsi les traits phénotypiques et les taux de substitutions dans un même processus multivarié, permettant ainsi, non seulement, d'utiliser les données moléculaires pour obtenir une estimation des traits phénotypiques passés, mais aussi de tester la corrélation entre les différents traits phénotypiques et entre les taux de substitutions et les traits phénotypiques. Cette approche offre la possibilité d'inclure des points de calibrations permettant de dater les événements de spéciation dans l'arbre, mais en leur absence, l'arbre est arbitrairement calibré à 1, on obtient alors des valeurs de taux de substitutions comparables entre les différentes branches de l'arbre. Pour ce qui est du test statistique, la matrice de covariance est estimée grâce à des méthodes de Monte-Carlo par chaînes de Markov. La chaîne est échantillonnée pour obtenir une distribution postérieure de chaque entrée de la matrice de covariance. A partir de cette matrice, le coefficient de corrélation entre les traits avec les taux de substitutions est calculé pour chaque point de la distribution postérieure. La probabilité postérieure d'une corrélation est obtenue

simplement en comptant la proportion des points pour lesquels le coefficient de corrélation est supérieur à zéro. Dans le cas où plusieurs traits sont inclus dans l'analyse, la co-variation entre tous ces traits est également modélisé. On obtient alors des probabilités postérieures partielles, reflétant la corrélation entre les traits et les taux d'évolution, corrigées pour la covariation avec les autres traits.

### 3 Approche choisie dans cette thèse

Nous venons de voir qu'il existe une grande variété d'approches ayant toutes leurs avantages et leurs limites d'utilisation. L'approche par paire d'espèces ne demande que peu d'information sur le groupe mais nécessite de former des paires. Ainsi toutes les espèces ne permettant pas de former une paire sont exclues du jeu de données. A l'inverse les autres approches permettent d'intégrer tous les taxons, mais demandent des connaissances plus poussées sur la phylogénie et les temps de divergence. Les approches de modélisation conjointe des taux et des traits étudiées imposent, quant à elles, un modèle d'évolution aux traits étudiés qui ne peut pas s'appliquer à tous les traits. Dans le cadre de cette thèse, une modélisation des traits étudiés sur tout l'arbre pose plusieurs problèmes. Premièrement, modéliser la transition vers le milieu souterrain est particulièrement compliqué car les espèces de surface ayant donné des populations souterraines ont très probablement disparues (voir Chap. 3 partie 2.4). Deuxièmement, la profondeur de ce clade est très importante, l'ancêtre des Asellidae étant estimé à environ 250 Millions d'années (voir Chap. 3 partie 1.4). Ainsi, modéliser des variables environnementales selon un processus brownien le long de l'arbre alors que la géographie a totalement changé le long de cet arbre n'a pas vraiment de sens. Enfin, si les temps de divergence de ces espèces sont présentés dans la partie suivante, il faut noter qu'ils ont été acquis à la fin de cette thèse, ne permettant pas de les utiliser dans les analyses de taux de substitutions. Pour toutes ces raisons, dans cette thèse, nous avons utilisé le logiciel CoEvol d'une façon légèrement détournée. En effet,

le logiciel est conçu pour modéliser la coévolution des taux de substitutions par unité de temps et de traits morphologiques le long de la phylogénie. Si aucune valeur de trait n'est donnée au logiciel, les taux de substitutions par unité de temps sont modélisés seuls sous un modèle d'évolution Muse et Gaut (1994). En l'absence de points de calibrations, ces taux sont estimés en temps relatif à la racine, permettant de comparer les taux entre eux, mais pas de comparer les valeurs entre plusieurs analyses. La distribution postérieure des taux de substitutions des branches terminales a ensuite été extraite et la médiane a été utilisée dans les corrélations. Cette utilisation n'étant pas standard, dans le premier chapitre, les résultats ont été comparés avec une approche par paire qui a donné des conclusions identiques.

## Chapitre 3

# Les *Asellidae* souterrains : développement du modèle pour l'étude des taux d'évolution moléculaire.

Afin d'évaluer les hypothèses relatives aux taux d'évolution moléculaire sur des organismes ectothermes, nous avons utilisé comme modèle biologique des pancrustacés isopodes aquatiques de la famille des *Asellidae*. Ces organismes présentent des caractéristiques originales qui permettent de tester la plupart des hypothèses faites sur les taux d'évolution moléculaire. Afin de pouvoir mener à bien les différentes parties de cette thèse, un travail d'approfondissement des connaissances sur l'histoire évolutive de ce groupe a du être mené en amont. Dans ce chapitre, je présenterai dans un premier temps en détail le modèle biologique et la façon dont les échantillons sont récoltés puis traités, je détaillerai également la systématique et l'histoire évolutive du groupe. Dans une deuxième partie, je montrerai en quoi les caractéristiques exposées permettent de tester les hypothèses relatives aux

taux d'évolution moléculaire. Je détaillerai ensuite les différents types de données moléculaires qui ont été utilisés pour répondre aux objectifs de cette thèse.

## 1 Le modèle biologique des Asellidae

Les Asellidae sont des Isopodes aquatiques marcheurs. Ils se déplacent sur les sédiments au fond de l'eau et sont incapables de nager. Ils se nourrissent de la matière organique présente dans les sédiments et présentent une thigmotaxie positive, c'est-à-dire qu'ils éprouvent la nécessité de rester au contact d'un substrat solide. Ainsi, on trouve les individus la plupart du temps accrochés sous les rochers ou les morceaux de bois dans l'eau. Ils sont généralement résistants au courant, certaines espèces affectionnent les cours d'eau rapides alors que d'autres sont trouvées dans tous les milieux, y compris stagnants. Cette famille d'isopodes présente la caractéristique de renfermer à la fois des espèces vivant dans les eaux souterraines et des espèces vivant dans les eaux de surface (rivière, lac). Le passage vers le milieu souterrain est accompagné de la régression de certains traits (*e.g.* yeux) souvent irréversible, rendant le retour vers le milieu de surface très peu probable. Ainsi, en comparant une espèce souterraine à une espèce de surface, on peut supposer que la transition s'est faite du milieu de surface vers le milieu souterrain. D'autre part, au sein de cette famille on trouve, dans des clades différents, des espèces de surface et des espèces souterraines, nous indiquant que cette transition s'est produite plusieurs fois indépendamment. Ainsi, cette famille nous offre l'opportunité d'avoir plusieurs réplicats indépendants du même changement d'habitat. Cette transition écologique a de nombreuses conséquences sur les caractéristiques des espèces, notamment sur leurs traits biologiques, leur répartition et leur capacité de dispersion. Après avoir présenté les différentes caractéristiques de cette famille, je présenterai son histoire évolutive à travers les travaux effectués durant cette thèse qui ont permis de définir les plans expérimentaux utilisés dans la suite de ce travail.

## 1.1 Impact de la transition vers le milieu souterrain chez les Asellidae

Le milieu souterrain est souvent considéré comme un milieu extrême dans le sens où des organismes non adaptés à ce milieu semblent ne pas pouvoir y effectuer leur cycle de vie en entier. Ainsi, lors de la transition vers le milieu souterrain on observe de nombreux changements de traits associés à ce milieu, qui créent de remarquables convergences dans les traits d'histoire de vie, mais également des convergences morphologiques, physiologiques, et même comportementales chez des organismes très variés tels que les arthropodes (malacostracés, arachnides,...), les téléostéens (*e.g.* *Astyanax*) ou encore les amphibiens (le protée). La comparaison entre les Asellidae de surface et les Asellidae souterrains montrent qu'ils n'échappent pas à ces règles de convergence. Du point de vue des traits d'histoire de vie, les Asellidae souterrains présentent une longévité accrue. Henry (1976) décrit une longévité en moyenne 5 fois plus longue chez les espèces souterraines qu'il a mis en élevage par rapport aux espèces de surface. Cette longévité accrue est accompagnée d'un rythme de mue plus lent, d'un temps de génération plus long et d'une maturité sexuelle plus tardive (Figure 3.1). La plupart des observations sur le nombre de descendants chez les espèces souterraines semblent indiquer une réduction du nombre de descendants en faveur d'une taille plus grosse (Vandel, 1964), cependant les résultats d'une étude plus approfondie chez les Asellidae semblent montrer une plasticité importante de la taille des oeufs en fonction des conditions environnementales, donnant une image bien plus complexe de la réalité (Annexe 1). Morphologiquement, les transformations les plus flagrantes sont la régression du système oculaire et la dépigmentation. Mais on observe également un allongement du corps et des appendices, et une régression des crochets copulateurs, en lien avec le changement de comportement reproducteur. Physiologiquement, il a été montré chez quelques espèces de malacostracés souterrains une résistance à l'hypoxie bien plus grande que leurs équivalents de surface (Malard et Hervant,

1999) et une meilleure survie lors de longues périodes sans nourriture (Hervant et Renault, 2002). Au niveau du comportement, on observe une activité motrice réduite (Henry, 1976) mais également des comportements sexuels différents : chez les Asellidae de surface il y a une chevauchée nuptiale au moment de l'accouplement avec formation d'un couple durable. Ce comportement d'accouplement a totalement disparu chez les espèces souterraines (Henry, 1976).

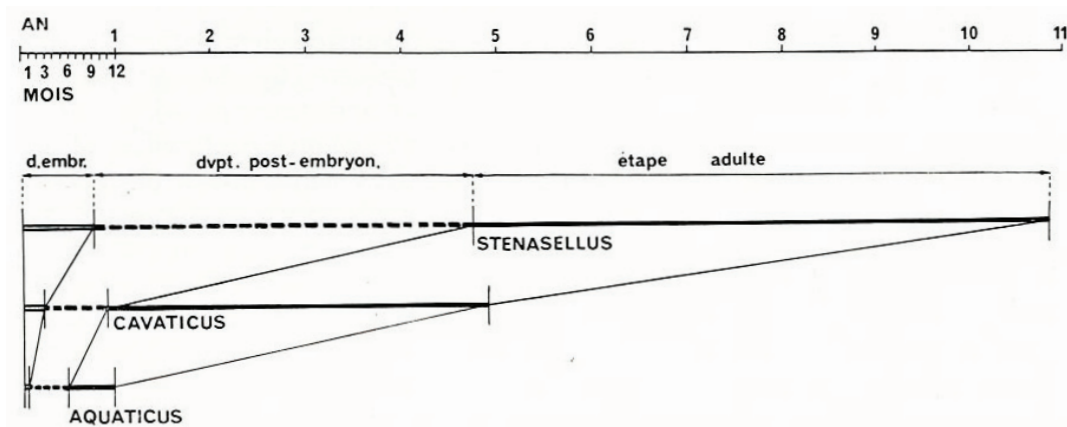


FIGURE 3.1 – Durée relative comparée des trois périodes principales du cycle vital (développement embryonnaire, développement post-embryonnaire et étape adulte) chez *Stenasellus* (Stenasellidae), *Proasellus cavaticus* et *Asellus aquaticus* (Asellidae)<sup>3</sup>. Henry faisait l'hypothèse que la différence observée entre *Stenasellus* et *Proasellus* venait d'une colonisation plus ancienne du milieu souterrain par *Stenasellus*.

## 1.2 Milieux de vie et méthodes d'échantillonnage

On trouve les Asellidae en milieu de surface ou en milieu souterrain. Il existe plusieurs types d'habitats souterrains. Ces habitats ont des caractéristiques communes : premièrement, ces habitats sont caractérisés par l'absence de lumière qui engendre une absence de photosynthèse et donc une productivité primaire très limitée. Ces habitats sont ainsi des milieux très pauvres en nutriments (Francois et al., 2016b) mais aussi potentiellement en oxygène (Malard et Hervant, 1999) dont les apports dépendent principalement du milieu extérieur. Deuxièmement,

3. Tiré de Henry (1976)

le réseau trophique est généralement simplifié en milieu souterrain, c'est-à-dire que le nombre de niveaux dans la chaîne alimentaire est réduit. Troisièmement, ces milieux montrent une grande stabilité thermique. La température du milieu souterrain varie très peu et reste toujours autour de la moyenne annuelle régionale (Freeze et Cherry, 1979). Ces caractéristiques en font un milieu dans lequel les organismes non adaptés ne peuvent pas effectuer tout leur cycle de vie (survie et reproduction). Cependant, au delà de ces caractéristiques communes, on peut différencier deux grands types d'habitats souterrains : les habitats dans les roches consolidées et l'eau qui s'écoule dans les roches non consolidées.

### **Echantillonnage des espèces de surface**

Les espèces de surface se trouvent dans les cours d'eau ou les lacs. On les trouve généralement accrochées à un substrat (racines, branches de bois, roches) ou dans le premier centimètre de sédiments. Leur échantillonnage se fait soit en soulevant les substrats et en les récoltant avec un pinceau, soit en filtrant la première couche de sédiment avec un filet surber. Ce filet est entouré d'un cadre métallique permettant de racler directement les sédiments.

### **Habitat dans les roches consolidées**

Il s'agit de l'eau qui est contenue dans les fissures de la roche. Ces habitats se trouvent principalement dans les formations karstiques. Le karst est une structure géomorphologique principalement constituée de roches calcaires érodées par les eaux souterraines : l'érosion élargie les fissures de la roche jusqu'à former des grottes. L'habitat karstique reste toutefois principalement constitué de fissures dans lesquelles des organismes peuvent vivre. Ce karst est donc en majorité inaccessible à l'homme. On peut accéder à cet habitat par les grottes ou par les sources d'eau qui jaillissent de ce karst, ce qui ne nous donne donc une image que très partielle du milieu de vie des organismes. Dans les grottes on peut trouver les organismes dans des gours et des flaques, sur des parois où l'eau s'écoule de



façon continue, ou dans des rivières souterraines. Dans ce type de milieu, soit les individus sont sous les pierres dans la rivière et sont récoltés à l'aide d'un pinceau (Figure 3.2), soit on prélève la surface des sédiments au fond de la rivière avec un filet (Figure 3.3), permettant de récupérer les organismes à la surface du sédiment. Pour les sources d'eau, le positionnement d'un filet au niveau de la source permet de récupérer les organismes qui sont projetés en dehors du milieu.



FIGURE 3.2 – Photo d'un *Proasellus strouhali* au moment de l'échantillonnage au pinceau dans la rivière d'une cavité souterraine (Mine de Breté, Bourgogne, France)<sup>4</sup>.

---

4. Droit photo : Bernard Lips



FIGURE 3.3 – Échantillonnage avec un filet qui est utilisé pour prélever la surface des sédiments et récupérer les organismes qui y vivent <sup>4</sup>.

### Habitat dans les roches non-consolidées

Il s'agit de l'eau qui s'écoule dans des alluvions : souvent proche de la surface du sol, elle peut accompagner les fleuves ou les rivières. Les Asellidae souterrains vivent alors dans les interstices entre les graviers qui sont remplis d'eau. On accède à ces habitats en échantillonnant dans les puits qui ont été creusés pour prélever de l'eau, soit en utilisant une pompe pour prélever l'eau souterraine de la zone hyporhéique (eau qui s'écoule sous la rivière).

Pour échantillonner dans les puits, on utilise un filet phréatobiologique de type CvetKov (Cvetkov, 1968, Figure 3.4) qui consiste en un filet lesté accroché à une corde permettant de prélever la faune dans le fond du puits.

La zone hyporhéique est définie comme l'ensemble des sédiments saturés en eau, situés au dessous et à côté d'une rivière, contenant une certaine proportion d'eau de surface (White, 1993; Datry et al., 2008). Cette zone est une zone

4. Droit photo : Bernard Lips

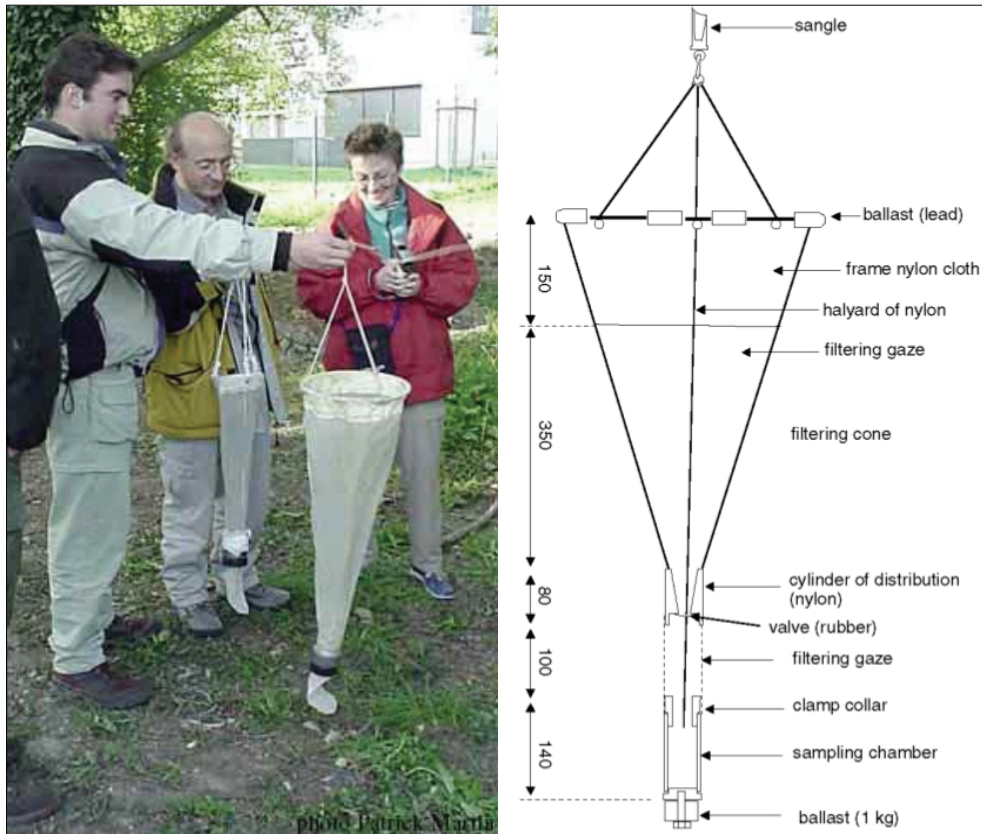


FIGURE 3.4 – Photo (gauche)<sup>5</sup> et schéma fonctionnel (droite)<sup>6</sup> d'un filet Cvetkov utilisé pour l'échantillonnage en puits.

d'échange entre les eaux souterraines et les eaux de surface. On y trouve donc à la fois de la faune de surface et de la faune souterraine dans des proportions variables en fonction de la rivière, du niveau de l'eau et de la profondeur à laquelle on prélève les organismes. Dans ce milieu, la faune est prélevée par sondage Bou-Rouch. Ce sondage consiste à enfoncer un tube d'acier crépiné sur lequel on fixe une pompe manuelle. On peut ainsi pomper l'eau interstitielle avec les sédiments fins et les organismes qui y vivent, que l'on récolte dans un filet à la sortie de la pompe (Figure 3.5).

5. Tiré de "Sampling manual for the assessment of regional groundwater biodiversity. Protocols for the Assessment and Conservation of Aquatic Life in the Subsurface (PASCALIS)" avec l'autorisation de F. Malard.

6. Tiré de Bou (1974)

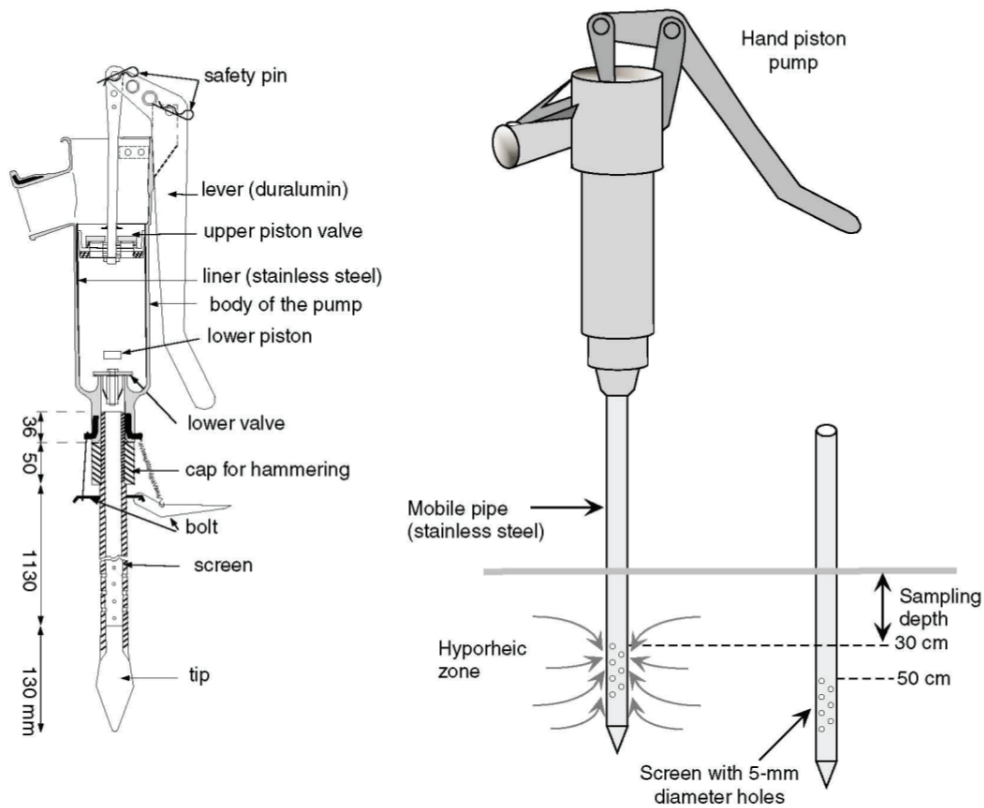


FIGURE 3.5 – Schéma d'une pompe Bou-Rouch<sup>6</sup>.

### 1.3 Délimitation des espèces d'Asellidae

Avant l'arrivée des outils moléculaires, les espèces d'Asellidae ont été établies à partir de la description de la forme de l'organe copulateur mâle (seconds pléopodes, Figure 3.6). Cet organe est le support de détermination morphologique des espèces. Les travaux de Morvan et al. (2013) ont cependant mis en évidence que cette famille contenait un grand nombre d'espèces indifférenciées sur des critères morphologiques. Sur les 114 espèces considérées par Morvan et al. (2013), les auteurs ont délimité 204 unités taxinomiques grâce aux outils moléculaires (appelées MOTU pour Molecular Operational Taxonomic Unit). Parmi les 42 espèces déterminées à partir de la morphologie, pour lesquelles les auteurs disposaient de plus d'un site d'échantillonnage, 78,5% contenaient plusieurs espèces délimitées

6. Tiré de Bou (1974)

avec des outils moléculaires. Ainsi, il est nécessaire de définir les entités évolutives à partir d'une délimitation moléculaire des espèces. Cependant, une approche moléculaire pose d'autres problèmes. Premièrement, il existe différents outils de délimitation qui donne des résultats différents. Deuxièmement, beaucoup d'outils de délimitation moléculaire sont basés sur une phylogénie, ainsi les résultats dépendent de l'échantillonnage taxinomique. De ce fait, les délimitations changent au cours du temps avec l'ajout de nouvelles données. Troisièmement, si le lien n'est pas fait avec la morphologie, les données de la littérature ne peuvent plus être mises en lien avec les nouvelles données. Afin de gérer le lien entre l'identification morphologique et les différentes identifications moléculaires, un protocole d'identification des individus a été mis en place au sein du laboratoire, ainsi qu'une base de données permettant de tracer ce protocole (Malard et al., 2019). Cette base de données permet de répertorier les données acquises dans le laboratoire mais également les données de la littérature et finalement de lier les différentes identifications avec les données géographiques.

**Protocole d'identification** Afin d'assurer l'homogénéité des données, un protocole d'identification a été mis en place dans le laboratoire. Lors de l'échantillonnage sur le terrain, les individus sont stockés en alcool 96° et conservés dans des glaciaires afin de garantir la conservation de l'ADN tout en permettant l'identification morphologique des individus. Une fois au laboratoire, les échantillons sont stockés à  $-20^{\circ}\text{C}$  en attendant leur traitement. Pour chaque station, un premier tri est effectué par un spécialiste du groupe (Florian Malard, CR CNRS) afin de détecter la présence de plusieurs espèces dans l'échantillon. Si c'est le cas, les espèces sont séparées. 2 à 3 individus mâles par espèce sont ensuite extraits de l'échantillon afin d'être identifiés morphologiquement. Cette identification se fait en disséquant les Pléopodes II, organes copulateurs mâles (Figure 3.6). Ces pléopodes sont placés entre lame et lamelle et observés au microscope afin de déterminer le taxon morphologique. Les lames reçoivent un identifiant et sont

conservées dans les collections du laboratoire. Les pattes des individus disséqués, ainsi que celles d'individus femelles sont ensuite utilisées pour effectuer une extraction d'ADN, puis une amplification par PCR des gènes 16S et COI qui sont ensuite séquencés en Sanger (protocole détaillé en partie 2.1).

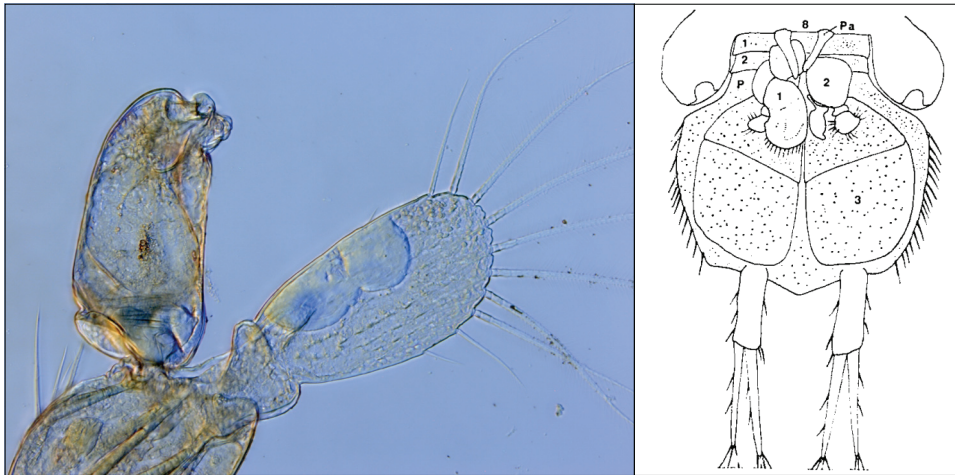


FIGURE 3.6 – Photo d'un pléopode II d'un mâle de *Proasellus beticus* utilisé pour la détermination spécifique (gauche)<sup>7</sup>. Pléotelson d'un Asellidae (droite)<sup>8</sup> montrant le positionnement des pléopodes, notamment le pléopode mâle copulateur (n° 2).

Les chromatogrammes sont analysés et nettoyés avec FinchTV (Geospiza, Inc., version 1.3.1, <https://digitalworldbiology.com/FinchTV>). Les séquences sont validées une par une en vérifiant l'absence de codon stop, d'insertion ou délétion modifiant le cadre de lecture. Si des individus provenant de la même station et du même morpho-taxon présentent des différences importantes dans les séquences de COI, le gène 16S est utilisé pour vérifier que cette divergence est réelle. Si la divergence n'est pas confirmée par le 16S, la séquence est retirée et une nouvelle amplification du COI est effectuée avec des amorces différentes. Les séquences du gène COI sont ensuite utilisées pour effectuer l'identification moléculaire (Scripts et

---

7. Droits photo : Marie-José Dole-Olivier.

8. Tiré de Henry et Magniez (1983)

tutoriel disponibles sur github [https://github.com/TristanLefebure/MOTU\\_delimitation\\_pipeline](https://github.com/TristanLefebure/MOTU_delimitation_pipeline)). Les séquences de COI de tous les individus disponibles sont alignés avec Muscle (Edgar, 2004) puis l'alignement est vérifié *de visu* (qualité de l'alignement, bon cadre de lecture). Les séquences identiques sont éliminées pour réduire l'alignement aux haplotypes, avec le script *collapse\_to\_uniq\_seq* ([https://github.com/TristanLefebure/collapse\\_to\\_uniq\\_seq](https://github.com/TristanLefebure/collapse_to_uniq_seq)). Si des séquences sont identiques sur la partie alignée mais diffèrent par leur longueur, la séquence la plus longue est conservée. Si les séquences ne diffèrent qu'en raison d'ambiguïtés (positions pour lesquelles la base est indéterminée), la séquence contenant le moins de code ambigu est conservée. Un arbre phylogénétique est ensuite construit avec PhyML (Guindon et al., 2010) sous un modèle GTR + G + I. L'arbre obtenu est enraciné au niveau de la dichotomie entre Asellidae et Stenasellidae. L'arbre de COI est ensuite utilisé pour effectuer la délimitation des MOTU. trois méthodes sont utilisées :

- La méthode de seuil (Lefebure et al., 2006) : sur la base d'une comparaison des divergences morphologiques et moléculaires d'espèces de "crustacés", Lefebure et al. (2006) ont proposé un seuil de divergence moléculaire de 16% sur le gène du COI au delà duquel deux individus sont considérés comme étant issus de deux espèces différentes.
- Le Poisson Tree Process (Zhang et al., 2013, PTP). Cette méthode modélise le taux de substitutions (représenté par les longueurs de branches dans l'arbre) sous un modèle de poisson, pour définir un seuil en deçà duquel les longueurs de branches correspondent à un processus de coalescence intrapopulationnel. Pour simplifier, les longues branches correspondent à un processus de spéciation, c'est-à-dire à des divergences entre espèces, alors que les courtes branches correspondent à un processus de coalescence, c'est-à-dire des divergences entre individus de la même espèce. Le modèle permet de définir le seuil entre ce qui est considéré comme une longue et une courte

branche.

- Le Bayesian Poisson Tree Process (BPTP). Le modèle est identique à la méthode précédente, seule l'implémentation change : il s'agit d'une implémentation Bayésienne alors que le PTP standard est implémenté en maximum de vraisemblance.

Alors que la méthode de seuil est plus conservative, le PTP (ou BPTP) délimite un plus grand nombre de MOTUs.

Une fois toutes ces méthodes utilisées, pour chaque individu séquencé, on dispose d'une identification morphologique et de 3 identifications moléculaires. Toutes les informations à propos de l'échantillonnage, du stockage et des identifications sont ensuite répertoriées dans une base de données. La figure 3.7 résume ce processus d'identification mis en place dans le laboratoire.

**Gestion des données** Au cours du processus d'acquisition des échantillons et d'identification des espèces, un grand nombre de données est généré. En effet, lors de l'échantillonnage des données relatives à la station de collecte, telles que le nombre d'individus collectés, la date, les coordonnées, la description du lieu, la température, la conductivité, les collecteurs sont relevées. Ensuite lorsque le taxon est identifié morphologiquement, il reçoit un nom de taxon. Les individus utilisés pour l'identification reçoivent un numéro de collection. Les pattes des individus envoyés au séquençage sont stockés sous forme d'extraction d'ADN avec un identifiant et les produits de PCR également. Les résultats du séquençage sont également conservés avec un identifiant. Les individus reçoivent ensuite des numéro de MOTU en fonction de la méthode et du jeu de données utilisés. Enfin, les séquences déposées dans les bases de données reçoivent un accession number. Toutes ces données nécessitent d'être soigneusement conservées afin d'assurer la reproductibilité des analyses effectuées, mais également afin de satisfaire aux normes de plus en plus nombreuses relatives à la collecte de matériel vivant (*e.g.* protocole de Nagoya). Dans un but plus fonctionnel, ces données ont besoin d'être



interconnectées afin de faciliter leur gestion et leur utilisation. Pour ces raisons, la base de données GOTIT (<https://github.com/GOTIT-DEV/GOTIT>) dont la structure est adaptée au processus d'identification précédemment exposé a été créée (Figure 3.7, Malard et al., 2019, Annexe 2). Cette base de données permet à la fois de gérer toutes les données acquises par le laboratoire mais également d'analyser la répartition spatiale de ces données. Ainsi, on peut voir la répartition de chaque taxon définie sur la base de la morphologie ou sur une base moléculaire.

## 1.4 Phylogénie du groupe

Les Asellidae sont une famille d'isopodes aquatiques comprenant 439 taxa décrits dans la littérature sur la base de la morphologie. La monophylie des Asellidae a été remise en doute par un article (Lins et al., 2012) basé principalement sur le gène 18S. Les phylogénies construites au cours de cette thèse, basées sur un grand nombre de gènes et sur les deux compartiments génomiques, les définit bien comme un clade, groupe frère des Stenasellidae (voir Chapitre 6). Afin de mettre en place les différents protocoles expérimentaux pour cette thèse, nous avons reconstruit une phylogénie, calibrée dans le temps, de la super famille des des Aselloidea (Asellidae et Stenasellidae).

### Matériel et Méthodes

**Données moléculaires** 1731 séquences de COI appartenant à 233 taxa, décrits sur la base de la morphologie, ont été assemblées (liste des séquences et accession number en Annexe 3), soit environ 53% des taxa décrits dans la littérature. Sur ces 233 taxa, 151 font partie du genre *Proasellus* qui compte 185 taxa décrits, couvrant ainsi 81% des taxa du genre. Sur ces 1731 séquences, 74 proviennent de la base de données NCBI, les autres ont été générées par le laboratoire au cours des 15 dernières années lors de grandes campagnes d'échantillonnage dans toute l'Europe et aux États-Unis (en collaboration avec un bu-

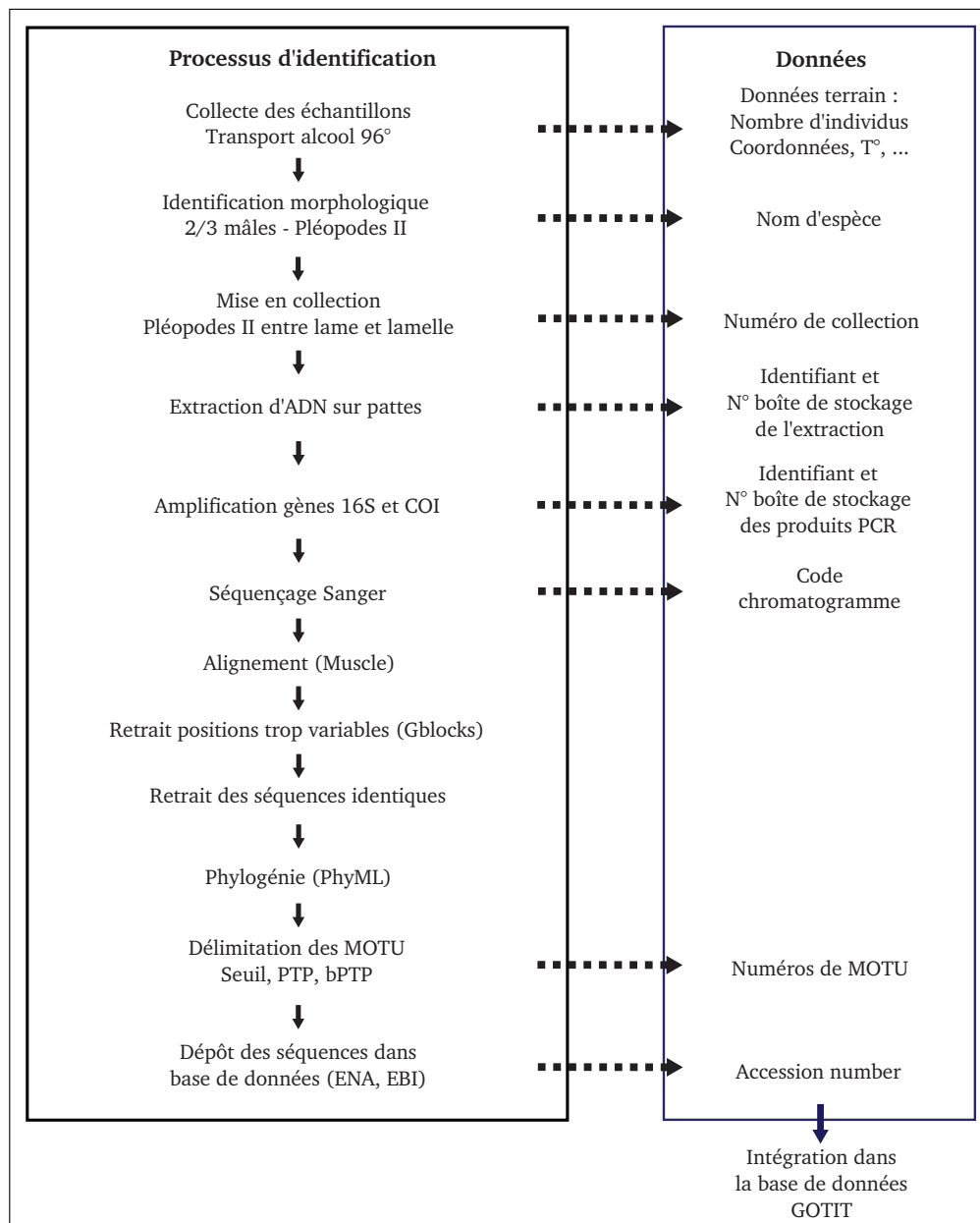


FIGURE 3.7 – Processus d'identification des espèces dans le laboratoire et données générées au cours de ce processus qui sont ensuite intégrées dans la base de données du laboratoire.

reau d'étude). Ces 1731 séquences ont été utilisées pour délimiter les MOTU en suivant le pipeline présenté en partie 1.3. En utilisant la méthode du PTP, on obtient 409 MOTUs. Pour tous les MOTUs pour lesquels le matériel biologique était disponible, les gènes 16S, 28S et FAST-2 (FAST kinase-like protein, subdomain

2) ont été séquencés en Sanger (protocole partie 2.1) pour au moins 2 individus par MOTU, provenant si possible de deux stations différentes. La meilleure séquence de chaque gène a été sélectionnée avec le script `chimera_assembler` ([https://github.com/TristanLefebure/chimera\\_assembler.pl](https://github.com/TristanLefebure/chimera_assembler.pl)), qui conserve la séquence la plus longue ayant le moins de code ambigu, nous permettant ainsi d'obtenir 4 gènes pour 367 MOTUs parmi les 409 (Liste des MOTUs et accession number en Annexe 4).

**Reconstruction phylogénétique** Les gènes 16S et 28S des 367 MOTU ont été alignés avec MAFFT (Kato et Standley, 2013) et les gènes COI et FAST-2 ont été alignés avec Prank (Löytynoja et Goldman, 2008). Les positions trop variables pour être alignées sans ambiguïté ont ensuite été retirées avec Gblocks (Castresana, 2000). Les 4 gènes ont ensuite été concaténés pour construire une phylogénie avec le logiciel bayésien PhyloBayes (Lartillot et al., 2009) sous un modèle CAT-GTR. Ce logiciel permet d'obtenir un arbre consensus majoritaire, à partir des arbres de la chaîne de Markov. Afin de ne pas avoir de polytomies dans l'arbre, un seuil de 10% a été utilisé pour définir les bipartitions majoritaires. L'arbre a été enraciné en utilisant les Stenasellidae comme outgroup.

**Estimation des temps de divergence** La topologie donnée par PhyloBayes a ensuite été utilisée pour reconstruire les temps de divergence, en y ajoutant des points de calibrations. N'ayant qu'un seul fossile exploitable dans notre contexte, des évènements climatiques et tectoniques ont été utilisés pour calibrer les noeuds dans l'arbre :

- Point de calibration n° 1 et 3 : Phillips et al. (2013), en utilisant le plus vieux fossile d'isopode connu (*Hesslerella*), daté à -300 Ma, datent la divergence entre les Ligiidae et les Oniscidae (deux groupes d'isopodes) comme étant plus récente que ce fossile. La divergence entre ces deux isopodes étant plus ancienne que la divergence entre les Stenasellidae et les Asellidae (Lins

et al., 2012), on peut en conclure que cette dernière divergence est plus récente que 300 Ma.

- Point de calibration n° 2 : La répartition des Stenasellidae suggère une origine du groupe à partir du paléo-océan Neo-Tethys (Magniez (2008), dans (Malard et al., 2014)), l'ouverture de ce paléo-océan est datée à environ -250 Millions d'années (Stampfli, 2000)). Ainsi, la racine des Stenasellidae sera considérée comme ne pouvant pas être plus ancienne que -250 Ma.
- Point de calibration n° 3 : Le genre *Gallasellus*, trouvé exclusivement en France, est le groupe frère du genre *Caecidotea*, trouvé exclusivement aux Etats-Unis. Ainsi, leur divergence ne peut pas être plus récente que le dernier contact entre les deux continents, daté à -54 Ma (Skogseid et al., 2000).
- Points de calibration n° 2, 4 et 5 : La divergence entre les espèces Sardo-Corses et les espèces pyrénéennes est nécessairement antérieure ou concomitante à la séparation de la Corse et du continent, il y a 29 Millions d'années (Orsini et al., 1980). On trouve des espèces proches continentales et sur la Corse (et sur la Sardaigne) dans deux groupes du genre *Stenasellus* et dans un clade ibero-aquitain du genre *Proasellus*.
- Point de calibration n° 6 : Les espèces bétiques (*Proasellus aff. escolai*, *Proasellus aff. lagari*, *Proasellus bellesi*, *Proasellus beticus*, *Proasellus comasi*, *Proasellus escolai*, *Proasellus espanoli*, *Proasellus goubaultae*, *Proasellus granadensis*, *Proasellus jaloniacus*, *Proasellus lagari*, *Proasellus meijersae*, et *Proasellus solanasi*) ont un ancêtre commun exclusif et sont réparties sur une zone géographique restreinte qui était marine jusqu'à la régression de la mer il y a 19 Ma (Meulenkamp et Sissingh, 2003). Considérant que la migration complète d'un clade entier est peu probable, on peut supposer que la diversification de ce clade s'est faite dans cette zone et ne peut donc pas être plus ancienne que 19 Ma.

- Point de calibration n° 7 : De la même façon les espèces de la vallée d'Ason (clade de *Proasellus cantabricus*, *Proasellus chappuisi*, *Proasellus grafi*, *Proasellus sp.* (Cuvera2), *Proasellus stocki*, *Proasellus ebrensis*, *Proasellus ortizi*) ont un ancêtre commun exclusif et sont toutes contenues dans une zone restreinte qui était marine jusqu'il y à 54 Ma (Payros et al., 2006; Orue-Etxebarria et al., 2001; Baceta et al., 2004)
- Point de calibration n° 8 : Idem pour les espèces Basques (clade de *Proasellus aquaecalidae*, *Proasellus aragonensis*, *Proasellus guipuzcoensis*, *Proasellus n. sp.* (Aquerreta), *Proasellus n. sp.* (Etcheberrigaray), *Proasellus spelaeus* et *Proasellus vandeli*), dont la zone de répartition était sous la mer jusqu'il y à 49 Ma (Plaziat, 1981; Pujalte et al., 2000; Baceta et al., 2004).
- Point de calibration n° 9 : Les *Bragasellus lagari* sont également dans une zone restreinte qui était marine jusqu'il y à 75 Ma (Tyson et Funnell, 1987; Dercourt et al., 2000).
- Points de calibration n° 10 à 14 : 4 lignées du clade Alpin du genre *Proasellus*, ainsi que le clade des *Proasellus anophthalmus* à l'exception de *Proasellus karamani* étaient marines jusqu'il y à 70 Ma (Tyson et Funnell, 1987; Kędzierski et Leszczyński, 2013).
- Point de calibration n° 15 : L'île de Crête est sortie de l'eau il y a maximum 17 Ma, ainsi les espèces proches de *Proasellus coxalis* qui sont sur l'île, se sont très probablement séparées après cette date.
- Point de calibration n° 16 : Les espèces *Proasellus arnautovici*, *Proasellus gjorgjevici* et *Proasellus remyi* sont endémiques du lac Ohrid (Frontière Albanie, Macédoine du Nord). Ces espèces se répartissent non pas en trois mais en deux lignées génétiques dont l'une est trouvées dans les milieux profonds du lac et l'autre dans les milieux de surface, laissant supposer une spéciation intralacustrine (Wysocka et al., 2008). Ainsi la divergence entre

ces espèces ne peut pas être plus ancienne que l'apparition du lac, datée maximum à 2 Ma (1.3 à 1.9 dans Wagner et al. (2017), >1.2 dans Wagner et al. (2014), env. 2 dans Stelbrink et al. (2016) et Lindhorst et al. (2015)).

- Points de calibration n° 17 et 18 : De la même façon, la divergence entre les espèces proches endémiques du lac Baikal (Russie) ne peuvent pas être plus ancienne que l'apparition du lac, datée à environ 27 Ma (Grabowski et al., 2017; Petit et Déverchère, 2006). Ainsi, la divergence entre *Limnoasellus* et *Mesoasellus*, et la divergence entre les deux espèces de *Baikalasellus* ne peuvent pas être plus ancienne que -27 Ma.

Tous ces points ont été utilisés dans PhyloBayes (Lartillot et al., 2009) comme des soft-bounds (les dates maximales et minimales imposées peuvent être dépassées avec une probabilité plus faible), sous un modèle CAT-GTR + G+ I avec un processus birth-death sur les temps de divergence et une auto-corrélation log-normale des taux de substitutions entre les branches (le taux de substitutions d'une branche est tirée d'une distribution log-normale centrée autour du taux de substitutions de la branche parente). Afin de vérifier l'impact de chacun des points de calibration sur l'estimation, des chronogrammes ont été réalisés en ôtant à chaque fois un des points de calibration.

## Résultats

**Délimitation des MOTUs.** La délimitation moléculaire utilisant le PTP a permis de définir 409 MOTUs (355 avec la méthode du seuil) au sein des 233 taxa identifiés morphologiquement (appelés morpho-taxa par la suite). Dans la plupart des cas la délimitation moléculaire découpe les morpho-taxa en plusieurs unités sauf pour 14 MOTUs (3,7%) qui contiennent plusieurs morpho-taxa. Dans ces 14 MOTUs, 5 contiennent des sous-espèces de *Stenasellus virei*, 3 contiennent des sous-espèces de *Proasellus coxalis*, 1 contient 2 sous-espèces de *Proasellus meridianus* et 1 contient deux sous-espèces de *Proasellus intermedius*. Les 3 espèces du lac Ohrid : *Proasellus arnautovici*, *remyi* et *gjorgjevici* sont réparties dans 2

MOTUs de façon incongruente avec leur identification morphologique. *Proasellus slavus* et *Proasellus nolli* sont dans le même MOTU. Et enfin l'individu dont les séquences ont été récupérées dans NCBI, identifié comme *Asellus aquaticus* (acc.num. AY531762) est dans le MOTU des *Asellus kosswigi*. Hormis pour ce dernier, les critères ayant été utilisés pour délimiter des sous-espèces ne semblent pas être confirmés par les outils moléculaires.

**Phylogénie et répartition.** La phylogénie obtenue montrent globalement des supports de branches élevés (Figure 3.8). Globalement, la monophylie des différents genres de la famille des Asellidae présents dans cette phylogénie est bien soutenue avec des probabilités postérieures maximales (*i.e.* égale à 1). Deux genres ne sont pas monophylétiques : le genre *Caecidotea* qui est réparti dans 2 clades distincts, et le genre *Proasellus* dans lequel on trouve l'espèce *Chthonasellus bodoni*. On peut raisonnablement supposer que le genre *Chthonasellus* n'a pas de réel support, la morphologie ne justifiant pas la distinction de cette espèce en un genre distinct du genre *Proasellus* (communication personnelle Florian Malard). En considérant cette espèce comme étant en réalité une espèce du genre *Proasellus*, on peut considérer que ce genre est monophylétique. Le genre *Asellus* se positionne à la base des Asellidae, le genre *Gallasellus* dont toutes les espèces connues sont françaises, se positionne au milieu du clade contenant les espèces américaines (genre *Lirceus* et *Caecidotea*). Les deux espèces du lac Baikal (Russie) se positionnent à la base de ce clade. Le genre *Proasellus*, qui contient le plus d'espèces, se divise en 4 grands clades assez bien délimités géographiquement (Figure 3.9). Le clade slavus est à la base de la phylogénie, ce clade ne comporte que des espèces souterraines qui sont réparties entre le sud de l'Allemagne et la Croatie (avec des exceptions dans le sud de la France). Les relations de parentés entre les 3 autres clades sont moins robustes, la monophylie coxalis-alpin n'étant soutenue qu'avec une probabilité postérieure de 0,45. Le clade des alpins est réparti, comme son nom l'indique, sur la chaîne des Alpes. La seule espèce de surface connue dans ce clade

est *Proasellus coiffaiti*. Le clade coxalis contient autant d'espèces souterraines que d'espèces de surface. Si on regarde la répartition des espèces souterraines, l'aire est limitée dans les pays autour de la mer adriatique. Si on ajoute les espèces de surface, l'aire de répartition du clade est beaucoup plus étendue et va jusqu'au nord de l'Afrique et au sud de la France, en passant par la Corse et la Sardaigne (Figure 3.9). Le clade des Ibero-aquitains qui contient un grand nombre d'espèces, majoritairement souterraines mais également de surface, a une répartition très large sur l'Europe, y compris pour les espèces souterraines. On voit qu'à l'intérieur de ces clades on trouve des espèces de surface et des espèces souterraines en proportions variables. Globalement, la majorité des morpho-taxa sont monophylétiques (87%), avec toutefois quelques exceptions. Notamment *Proasellus coxalis* et *Proasellus anophthalmus* dont les identifications sont incongruentes avec la délimitation et la phylogénie, suggérant que les critères d'identification de ces groupes devraient être reconsidérés. D'autres morpho-taxa ne sont pas monophylétiques mais les probabilités postérieures faibles ne permettent pas de remettre en cause avec certitude leur monophylie (p.p.<0.7).

Les espèces souterraines ont des aires de répartition globalement plus restreintes que les espèces de surface (Figure 3.9), avec des exceptions notables. En effet si l'on s'intéresse aux espèces souterraines, 74,4% d'entre elles ne sont trouvées que sur une seule station, contre 54,3% pour les espèces de surface. Si on regarde la distance linéaire entre les deux stations les plus éloignées de l'aire de distribution d'une espèce (Maximum linear extend, noté MLE), on obtient une moyenne de 21 km pour les espèces souterraines contre 163 km pour les espèces de surface. On observe cependant 3 exceptions notables : *Proasellus cavaticus* (1314 km), *Proasellus strouhali* (1253km) et *Proasellus slavus* (664 km).

**Temps de divergence** La phylogénie calibrée dans le temps (Figure 3.8) donne une divergence entre Stenasellidae et Asellidae datée à 262.3Ma (IC 95% 298,665-202,36). L'origine des Asellidae est datée à -234Ma (IC 95% 288,588-173,111).



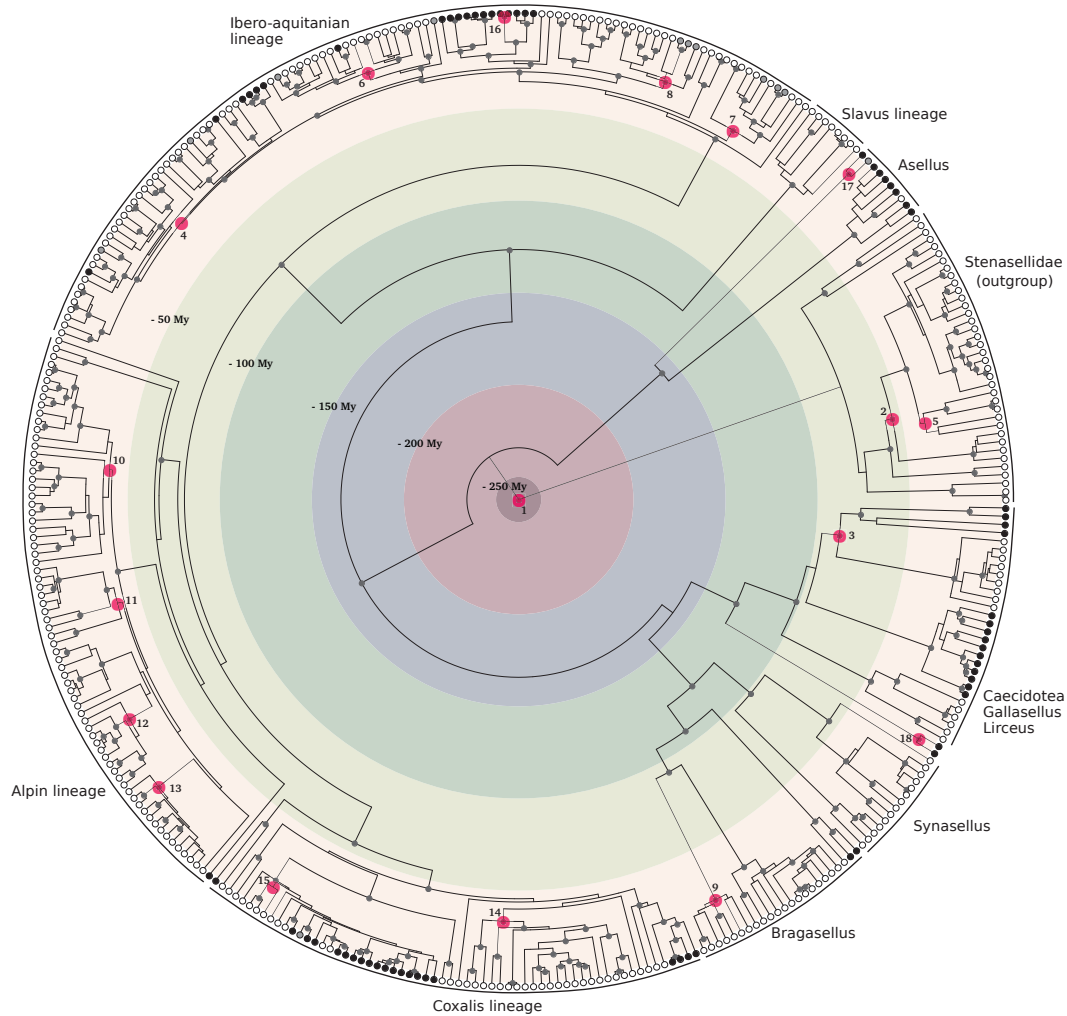


FIGURE 3.8 – Phylogénie calibrée dans le temps des Asellidae. L'arbre est enraciné en utilisant les Stenasellidae comme outgroup. Le temps est indiqué par des cercles de couleur. Les points de calibration sont indiqués par des ronds roses, les numéros correspondent à ceux indiqués dans le matériel et méthodes. Les supports de branches supérieurs à 0.90 sont indiqués par des ronds gris sur les noeuds fils. Les ronds des feuilles indiquent les espèces pigmentées et oculées (noir), totalement dépigmentées et anophthalmes (blanc) et partiellement dépigmentées et microphthalmes (gris).

L'apparition du genre *Proasellus*, genre le mieux échantillonné, est datée à -127,6Ma (IC 95% 159,25-99,50). Le retrait d'un point de calibration, quel qu'il soit, a peu d'impact sur l'estimation des temps de divergence (Figure 3.10). Les deux calibrations qui ont le plus d'impact sont l'ouverture de l'atlantique nord (point 2) et la diversification du clade contenant les espèces bétiques (point 6) avec des estimations qui restent dans les intervalles de crédibilité du chronogramme

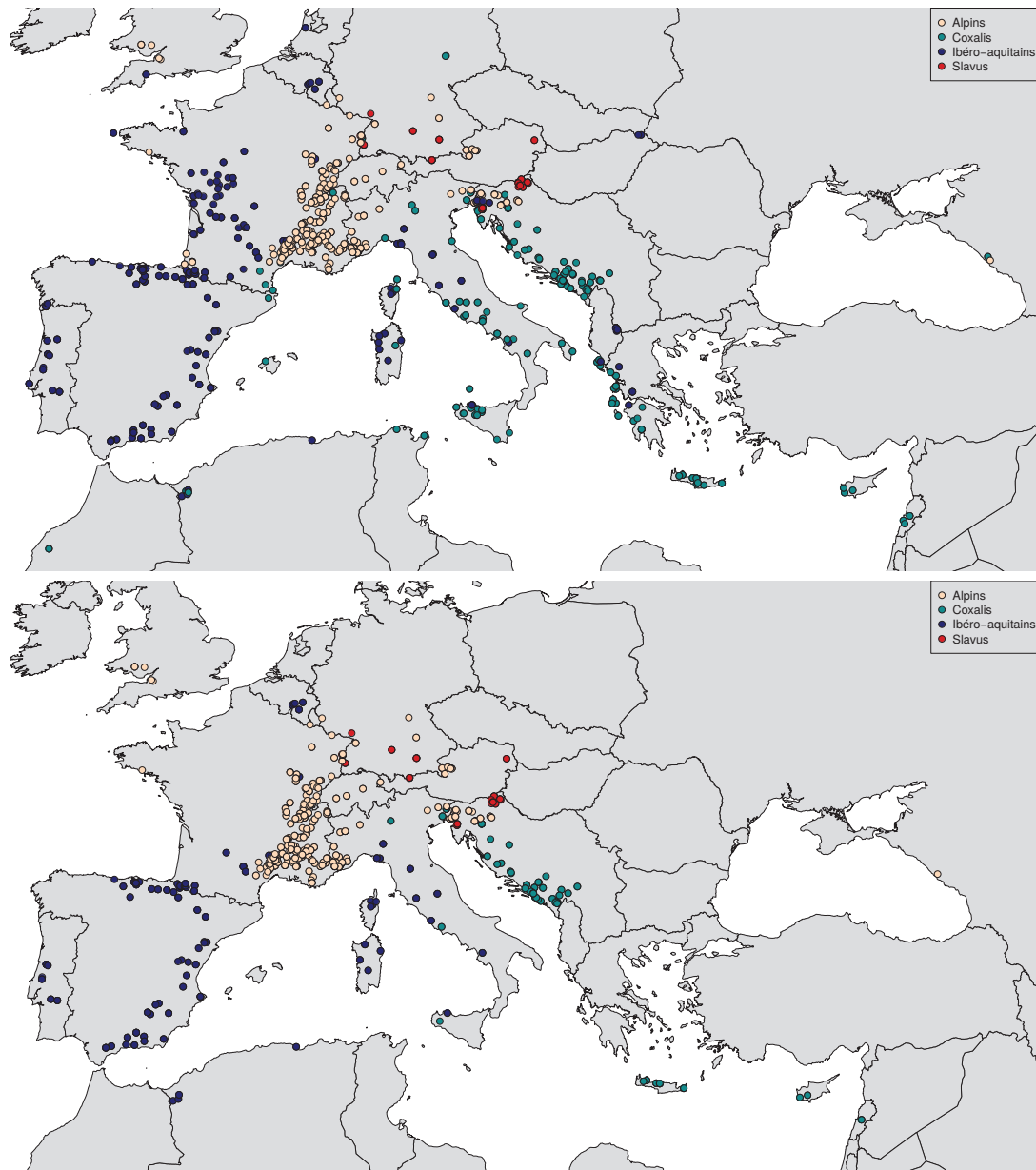


FIGURE 3.9 – Carte de répartition de tous les individus (haut) et des individus vivant en milieu souterrain (bas) appartenant au genre *Proasellus* pour lesquels une séquence de COI est disponible. Les 4 grands clades du genre sont représentés par des couleurs différentes.

construit avec tous les points. Les trois principaux clades de *Proasellus* se sont tous diversifiés entre -50 et -40 Ma.

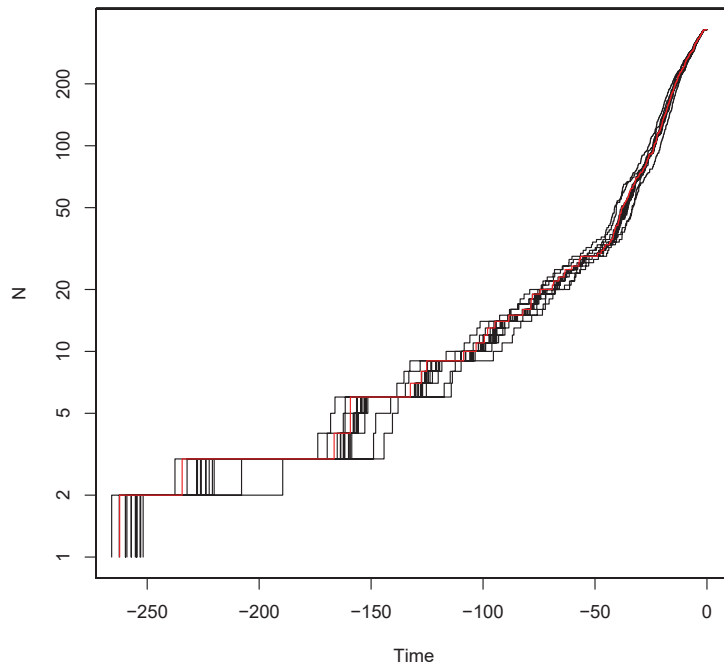


FIGURE 3.10 – Logarithme du nombre de lignées dans la phylogénie en fonction du temps (LTT plot). Chaque courbe correspond à chacune des phylogénies calibrées dans le temps avec un point de calibration en moins. Le temps est indiqué en millions d'années. La ligne rouge correspond à la phylogénie calibrée avec tous les points de calibrations.

## Discussion

**Délimitation des MOTUs** Globalement la délimitation moléculaire effectuée scinde en unités plus petites les taxa délimités morphologiquement. Cette délimitation met en évidence que les critères d'identification morphologique permettent dans la plupart des cas de délimiter des unités évolutives avec cependant des lacunes pour certains groupes. En effet, cette délimitation suggère que les critères utilisés pour délimiter les sous-espèces de *Stenasellus virei*, *Proasellus anophthalmus*, *Proasellus coxalis* et les espèces du lac Ohrid ne permettent pas de délimiter des unités évolutives. Dans le but de conserver le lien entre la morphologie et la délimitation moléculaire, un retour aux descriptions et aux identifications morphologiques pour ces individus serait nécessaire afin de vérifier la validité des critères permettant de définir ces espèces. Pour les différentes espèces et sous-

espèces identifiées dans le lac Ohrid, les 7 unités morphologiques sont regroupées en deux MOTUs qui correspondent aux deux lignées décrites par Wysocka et al. (2008). Ainsi, les espèces décrites par les auteurs s'apparenteraient plus en réalité à des populations et les lignées à des espèces. En revanche cette délimitation confirme la conclusion des auteurs : un des deux MOTU regroupe des individus collectés dans les milieux profonds du lac alors que l'autre regroupe des individus collectés dans les milieux de surface, confortant l'idée d'une spéciation écologique intra-lacustrine.

**Phylogénie** Cette phylogénie calibrée dans le temps des Aselloidea confirme que ce groupe fournit un modèle biologique pertinent pour étudier de nombreuses questions d'évolution moléculaire. Premièrement, espèces de surface et espèces souterraines apparaissent ensemble dans plusieurs clades, indiquant que la transition vers le milieu souterrain s'est effectuée de nombreuses fois indépendamment au sein des Asellidae. Deuxièmement, l'aire géographique très restreinte des grands clades suggère que les espèces souterraines dispersent très peu. A l'exception de trois espèces : *Proasellus cavaticus*, *Proasellus valdensis* et *Proasellus strouhali*. Ces trois espèces ont déjà été décrites comme des "exceptions" (Eme et al., 2013) et leur aires de répartition s'expliquent par des vagues de colonisation à partir de refuges, à la suite d'épisodes glaciaires. On peut donc bien supposer que les espèces souterraines ont des capacités de dispersion très faibles et faire ainsi l'hypothèse que les coordonnées géographiques actuelles sont proches des coordonnées passées, permettant d'estimer les paramètres environnementaux qu'elles expérimentent depuis de nombreuses générations. Troisièmement, ces espèces sont présentes partout en Europe dans des environnements très variés en terme de température, de productivité ou encore d'impact historique des glaciations.

**Contributions et Perspectives** L. Konecny-Dupré a effectué le séquençage des gènes. F. Malard a effectué la détermination morphologique des espèces et a

rassemblé les données géographiques permettant le calcul des MLE. N. Saclier a effectué la délimitation des MOTUs, les alignements et la construction de la phylogénie calibrée dans le temps, ainsi que les calculs des MLE. C. Douady a défini les points de calibrations et a supervisé la totalité du projet. Cette phylogénie fera l'objet d'un article sur la dispersion et la diversification des Asellidae. Elle sera également utilisée dans un article portant sur la diversification et la dispersion de l'espèce polytypique *Proasellus coxalis* autour de la Méditerranée, en lien avec la crise Messinienne. A plus long terme, cette phylogénie datée permettra d'étudier l'effet de transgression marine au niveau du Fleuve Rhône sur la diversification au sein de l'espèce *Proasellus walteri*. L'ajout de séquences en provenance des Etats-Unis permettra également de comprendre l'histoire évolutive du groupe sur le continent Américain et la relation des espèces Américaines avec les espèces Européennes.

## 1.5 Les Asellidae : un modèle original pour étudier les taux de substitutions

Les caractéristiques des Asellidae souterrains que je viens de présenter ont déjà fait leurs preuves pour répondre à des questions théoriques de génomique évolutive (Lefébure et al., 2017), d'écologie fonctionnelle (Francois et al., 2016b) ou encore de macro-écologie (Eme et al., 2018). Dans cette thèse, nous utiliserons les caractéristiques de ce modèle dans trois volets principaux :

1- Nous avons vu dans la partie 1.1 que la transition vers le milieu souterrain est accompagnée de nombreux changements de traits, notamment une longévité accrue, un temps de génération plus long, ou encore un taux métabolique réduit. Le fait que cette transition ait été effectuée indépendamment par de nombreuses espèces du groupe nous permet d'étudier les corrélations entre ces changements phénotypiques et les taux de substitutions avec de nombreux réplicats. Ainsi 13 paires d'espèces représentant chacune une transition du milieu de surface vers

le milieu souterrain ont été utilisées pour répondre au premier objectif de cette thèse. En comparant les taux de substitutions nucléaires et mitochondriaux, nous pourrions démêler l'impact des différents traits d'histoire de vie sur les taux de substitutions.

2- Nous avons vu dans la partie 1.4 que les isopodes souterrains ont également la particularité d'avoir une dispersion très faible. Le milieu souterrain étant très compartimenté, la dispersion au delà d'un aquifère est difficile pour ces espèces qui ne peuvent pas passer par les eaux de surface. Ainsi, l'aire de distribution moyenne d'une espèce se situe en moyenne dans un rayon de 15 km (Eme et al., 2018). Cette caractéristique permet de supposer que les lieux où l'on trouve les espèces actuellement sont les lieux où les espèces vivent depuis de nombreuses générations. Ainsi, on peut estimer qu'en caractérisant l'environnement actuel, on a une bonne estimation de leur environnement passé (dans la limite des changements paléogéographiques, tels que les fluctuations climatiques, les transgressions marines ou la tectonique des plaques). Contrairement aux animaux très mobiles (la plupart des mammifères, oiseaux, de nombreux insectes) qui sont capables de parcourir de grandes distances en peu de temps, ce modèle biologique nous permet donc de caractériser l'environnement qui a pu engendrer des changements de taux de substitutions. Nous pouvons ainsi caractériser la radioactivité naturelle à laquelle sont exposées ces espèces et tester l'impact de cette radioactivité sur les taux de substitutions sur 14 espèces du groupe.

3- Nous avons enfin vu dans la partie 1.4 que ces isopodes souterrains sont répartis sur tout l'hémisphère nord avec un échantillonnage très important en Europe, on les trouve ainsi dans des environnements très variables (latitude, température, productivité,...) nous permettant ainsi d'estimer des taux de substitutions le long de gradients environnementaux. Ainsi les 367 MOTUs du groupe pour lesquelles nous disposons de données moléculaires seront utilisés pour tester la corrélation entre les variables environnementales et taux de substitutions.

## 2 Données moléculaires

Les différentes parties de cette thèse se sont basées sur différents types de données moléculaires. En effet, la délimitation des espèces, la phylogénie calibrée dans le temps ainsi qu'une partie du chapitre 6 utilisent des données de séquençage Sanger. Les chapitres 4, 5 et 6 utilisent des données transcriptomiques mais également des génomes mitochondriaux assemblés à partir de différents types de données. Enfin, afin de prendre en compte explicitement le temps depuis lequel une espèce est devenue souterraine, nous avons cherché des gènes se non-fonctionnalisant chez les espèces lors de la transition du milieu de surface au milieu souterrain.

### 2.1 Séquençage Sanger

Dans le but de délimiter les espèces et de construire la phylogénie du groupe les gènes mitochondriaux COI et 16S ont été séquencés. Pour chaque morphotaxon de chaque site, un mâle disséqué et deux individus non identifiés ayant des tailles différentes (pour maximiser les chances de trouver des espèces cryptiques) ont été séquencés lorsque cela était possible (quelques stations n'ont pas permis la collecte de 3 individus ou de mâles). Pour chaque MOTU délimité, les gènes 28S et FAST-2 ont été séquencés pour au minimum deux individus, provenant au minimum de deux stations différentes si le MOTU était trouvé sur plusieurs stations.

Les individus collectés sur le terrain sont immédiatement placés dans de l'alcool à 96° et placés dans des glaciaires jusqu'au retour au laboratoire où ils sont placés à -20°C jusqu'aux analyses moléculaires. L'ADN est extrait en utilisant le protocole Chelex développé par Casquet et al. (2012) : une partie (pattes ou autre fragment) de chaque individu est incubé pendant 2h à 56°C dans 10 $\mu$ L de Protéinase K (15mg/mL) et 150 $\mu$ L de solution Chelex à 7% (Bio-Rad, Richmond, CA, USA). L'ADN est ensuite amplifié en utilisant le mix PCR indiqué dans la Table 3.1 et les amorces décrites dans Morvan et al. (2013). Le mix est ensuite

CHAPITRE 3. LES ASELLIDAE SOUTERRAINS : DÉVELOPPEMENT DU  
MODÈLE POUR L'ÉTUDE DES TAUX D'ÉVOLUTION MOLÉCULAIRE.

TABLE 3.1 – Mix utilisé pour l'amplification par PCR

	Volume par échantillon ( $\mu\text{L}$ )	Concentration finale
H <sub>2</sub> O	18,3	
Tampon 10X avec 30mM MgCl <sub>2</sub>	2,5	1X
10 $\mu\text{L}$ primer sens	0,6	0,24 $\mu\text{M}$
10 $\mu\text{L}$ primer anti-sens	0,6	0,24 $\mu\text{M}$
BSA 100X (10mg/mL Biolabs B9000S lot 0061506)	0,25	0,1mg/mL
dNTP 5mM chacun (Promega)	0,5	0,4mM
EUROBIOTAQ DNA polymerase 5U/ $\mu\text{L}$	0,25	0,05U/ $\mu\text{L}$
ADN	2	

TABLE 3.2 – Cycles du thermocycleur utilisés pour l'amplification par PCR

Température	Temps	Nombre cycles
95°C	5min	1
95°C	20 à 30 sec	35 à 40
51°C(COI), 53°C(16S et FAST2), 62°C(28S)	30 à 45 sec	
72°C	30 à 45 sec	
72°C	5min	1

placé dans un thermocycleur avec un cycle présenté en Table 3.2 . Une migration de 5 $\mu\text{L}$  du produit de PCR est ensuite effectuée sur gel d'Agarose (1.3% Agarose, Tampon TAE 1X, 0,6  $\mu\text{L}$  de Ladder 100pb JENA Bioscience Gel Red, 1 $\mu\text{L}$  de Bleu GelRed) afin de vérifier la réussite de l'amplification. Pour les gènes mitochondriaux, deux ensembles d'amorces différents sont toujours utilisés afin de mettre en évidence la présence de NUMTs (séquence mitochondriale dupliquée dans le génome nucléaire). En cas d'échec de l'amplification, une PCR semi-nichée est effectuée. A partir du produit PCR, on effectue une deuxième amplification en utilisant une des deux amorces ayant servi à la première amplification et une autre amorce plus proche que celle de l'amplification précédente. Une fois l'amplification réalisée avec succès, le séquençage Sanger est effectué par BIOFIDAL (Vaulx-en-Velin, France).

Le résultat du séquençage sous forme de chromatogramme est analysé avec le logiciel FinchTV version 1.3.1 (Geospiza Inc. ; <http://www.geospiza.com>). Les séquences sont ensuite alignées dans Seaview (Version 4.0) puis une vérification visuelle est effectuée afin de valider les séquences. Il existe deux raisons principales à la non validation de séquences : la présence de codons stop ou de frameshift,



ou la présence de nombreuses différences entre individus d'une même station (>5 différences). Si une séquence n'est pas validée, une nouvelle amplification avec des amorces différentes puis un nouveau séquençage sont effectués.

## 2.2 Données transcriptomiques

Afin de calculer les taux d'évolution moléculaire sur un grand nombre de gènes, les différentes parties de cette thèse ont toutes utilisé des données transcriptomiques. Pour cela, les transcriptomes des espèces utilisées dans chaque partie ont été séquencés puis assemblés *de novo*. Les familles de gènes orthologues ont ensuite été définies afin de pouvoir obtenir un ensemble de gènes orthologues utilisé pour le calcul des taux de substitutions. Au total, les transcriptomes de 67 espèces ont été séquencés. Pour les espèces sélectionnées pour le séquençage de transcriptome, les individus ont été récoltés sur le terrain et immédiatement congelés vivant dans de l'azote liquide. De retour au laboratoire, ils ont été placés à  $-80^{\circ}\text{C}$  jusqu'à l'extraction d'ARN.

**Séquençage** L'ARN total a été extrait en utilisant du TRI Reagent (Molecular Research Center). La qualité des extractions a été vérifiée sur une puce à ARN BioAnalyser (Agilent Technologies) et la concentration en ARN a été estimée en utilisant un fluorimètre Qubit (Life Technologies). Avant chaque séquençage, l'identification de l'individu a été vérifiée par le séquençage du gène 16S. Pour les transcriptomes séquencés avant 2013 (Liste des espèces avec date de séquençage et individus poolés en annexe 5), un pool équimolaire d'au minimum 5 individus a été effectué pour obtenir un minimum de  $10\mu\text{g}$  d'ARN, puis les volumes ont été réduits en utilisant un Concentrator-Plus (Eppendorf) pour obtenir approximativement  $10\mu\text{L}$ . Des ADNc double brins ont ensuite été produits en utilisant le kit Mint2 (Evrogen) en utilisant le protocole standard, excepté pour la synthèse du premier brin pour laquelle l'adaptateur CDS-1 a été utilisé avec le plugOligo-Adapter du kit Mint1 ( $5'_{-}\text{AAGCAGTGGTATCAACGCAGAGTACGGGGG}_{-}P_{3}$ ). Après

sonication avec le Bioruptor Nextgen UCD300 (Diagenode) et purification avec MinElute (Qiagen), les bibliothèques Illumina ont été préparées en utilisant le kit NEB-Next (New England BioLabs) et amplifiées en utilisant une amorce indexée unique pour chaque espèce. Après purification avec MinElute, les fragments de 400-500 paires de bases ont été sélectionnés sur un gel d'Agarose. Les bibliothèques ont été séquencées en paired-end sur un séquenceur Illumina HiSeq2000 en utilisant 100 cycles. Le séquençage a été effectué par le High-throughput DNA Sequencing Center (Copenhague, Danemark). Environ 25 millions de fragments ont été séquencés en paired-end par espèce.

Pour les transcriptomes séquencés à partir de 2013, une bibliothèque par individu a été préparée en utilisant le kit TruSeq™ RNA Sample Prep Kit v2 (Illumina). Pour chaque espèce, une bibliothèque a été séquencée en paired-end en utilisant 100 cycles sur un séquenceur Illumina HiSeq2500. Pour chaque espèce utilisée dans le chapitre 4, en plus de la bibliothèque paired-end, 7 bibliothèques ont été séquencées en single-end en utilisant 50 cycles sur un séquenceur Illumina HiSeq2500 également. Le séquençage a été effectué par la plateforme GenomEast de l'IGBMC (Illkirch, France). Environ 50 millions de fragments ont été séquencés pour les bibliothèques paired-end et 30 millions pour les bibliothèques single-end.

Toutes les données de séquençage haut-débit sur le modèle Asellidae produites par l'équipe E3S sont regroupées et déposées dans ENA sous le numéro de projet PRJEB14193 (<https://www.ebi.ac.uk/ena/data/view/PRJEB14193>).

**Assemblage** Les lectures ont été nettoyées et les transcriptomes séquencés en paired-end ont été assemblés *de novo* avec Trinity (Grabherr et al., 2011). Le pipeline d'assemblage étant différent dans chaque partie de la thèse, il sera décrit au sein des différentes parties.

## 2.3 Génomes mitochondriaux

Afin de calculer les taux de substitutions sur des gènes mitochondriaux, j'ai assemblé les génomes mitochondriaux de toutes les espèces dont le transcriptome a été séquencé. Pour assembler ces génomes mitochondriaux, 3 types de données ont été utilisés :

- Du séquençage de génome faible couverture pour 22 espèces issues de Lefébure et al. (2006).
- Les transcriptomes des 55 espèces n'ayant pas de génome séquencé.
- Des données de séquençage Sanger

**Assemblage à partir de séquençage de génomes faible couverture** Le génome mitochondrial étant présent en de multiples copies par cellules, un séquençage faible couverture permet d'assembler les génomes mitochondriaux complets. Les génomes de 22 espèces (faisant partie des 67 espèces dont le transcriptome a été séquencé) publiés dans Lefébure et al. (2017) été utilisés. Les lectures ont été utilisées pour assembler les génomes mitochondriaux avec MITObim (Hahn et al., 2013) en utilisant le gène du COI (voir partie 2.1) comme point de départ pour reconstruire le génome mitochondrial complet.

**Assemblage à partir des transcriptomes** Pour chacune des 55 espèces pour lesquelles nous n'avions pas de séquençage de génome, les gènes mitochondriaux de l'espèce la plus proche parmi les 22 assemblés précédemment ont été utilisés comme point de départ dans MITObim. Dans tous les cas sauf un (*Jaera prae-hirsuta* pour lequel le génome mitochondrial complet a été directement assemblé par Trinity), MITObim est parvenu à assembler de grands contigs mitochondriaux mais pas la totalité du génome mitochondrial. Les longs contigs mitochondriaux ainsi assemblés sont ensuite alignés sur un génome mitochondrial complet (la synténie des gènes mitochondriaux étant bien conservée chez les Asellidae). Lorsqu'un recouvrement permettait d'assembler deux contigs, les deux contigs ont été

fusionnés. Lorsqu'aucun recouvrement ne permettait de lier les différents contigs, la séquence a été complétée par du séquençage Sanger en utilisant les contigs autour de la séquence pour créer les amorces.

Les génomes mitochondriaux ont ensuite été annotés en utilisant le serveur web MITOS (Bernt et al., 2013). Nous avons ainsi pu obtenir les 13 gènes mitochondriaux codant pour des protéines, nous permettant de calculer les taux de substitutions mitochondriaux sur le maximum de gènes possible.

## 2.4 Gènes se non-fonctionnalisant en milieu souterrain

Nous avons vu que la transition du milieu de surface vers le milieu souterrain était accompagnée de la régression de nombreux traits (Partie 1.1). Nous avons également vu qu'une difficulté de l'étude des taux de substitutions est la différence de temporalité entre le taux de substitutions estimé sur une branche et un trait mesuré sur les organismes vivant actuellement (Chap. 1, partie 2). La branche terminale représente l'intervalle de temps entre la dernière spéciation connue et aujourd'hui (temps de spéciation). Le taux de substitutions calculé sur cette branche reflète donc le taux de substitutions moyen depuis la spéciation à aujourd'hui. Hors, l'impact des changements de traits liés à la transition vers le milieu souterrain sur les taux de substitutions ne commence qu'à partir du moment où l'espèce colonise ce nouveau milieu. Si l'espèce a colonisé le milieu récemment, relativement au temps de spéciation, la proportion de la branche pour laquelle l'espèce est souterraine est faible et le taux moyen sera plus proche de celui d'une espèce de surface alors que l'espèce est catégorisée comme souterraine (Figure 3.11). La non prise en compte de ce temps de colonisation du milieu souterrain pourrait donc avoir pour effet de 'noyer' le signal et de donner un résultat faussement négatif. Ce problème n'est pas spécifique au milieu souterrain, comme nous l'avons vu dans la partie 2 du Chapitre 1, il s'agit du problème général des corrélations entre des traits actuels et des taux de substitutions calculés sur un

intervalle de temps. La prise en compte de ce problème permettrait de gagner fortement en puissance dans les analyses. Pourtant, il est très rarement pris en compte dans les études sur les taux de substitutions. Dans le cas des espèces souterraines, une estimation du temps de colonisation permettrait de tenir compte de cette différence de temporalité.

**Box 1 : Modèles de diversification des espèces souterraines**

La diversification des espèces souterraines peut se faire de deux façons : soit par la diversification en milieu souterrain, soit à partir d'espèces de surface qui colonisent le milieu souterrain. La distribution restreinte des espèces souterraines, ainsi que leur spécialisation poussée ont amenées à des hypothèses relatives à la diversification en milieu souterrain majoritairement basées sur la colonisation du milieu par les espèces de surface (Howarth, 1987; Barr et Holsinger, 1985), développant l'idée que la spéciation en milieu souterrain est rare. Cependant, les travaux testant la diversification en milieu souterrain sont assez rares et il existe donc assez peu de support empirique accréditant l'hypothèse que le milieu souterrain serait un 'cul-de-sac' évolutif. Récemment, Stern et al. (2017) suggèrent même que ce n'est pas le cas. Théoriquement, la diversification en milieu souterrain devrait se faire par les processus identiques à ceux en surface (vicariance ou adaptation à de nouvelle niche). En revanche, le processus par lequel des espèces de surface sont à l'origine d'espèces souterraines est plus spécifique à l'écologie des milieux souterrains. Deux principaux modèles ont été proposés pour expliquer l'apparition des espèces souterraines : les modèles de relictés climatiques et le modèle de saut adaptatif.

Le modèle de relictés climatiques (Peck et Finston, 1993) est basée sur l'observation d'un grand nombre d'espèces dites relictés en milieu souterrain, à savoir des taxa qui apparaissent complètement isolés des espèces de surface dans la phylogénie. Imaginons une rivière en contact avec les eaux souterraines. Des individus d'une espèce stygophile (qui peut séjourner en milieu souterrain) vivant dans la rivière occupent à la fois les eaux de surface et les eaux souterraines. L'apport régulier d'individus provenant de la rivière conserve un brassage génétique entre les deux milieux, il n'y a donc pas formation de nouvelles espèces. Un changement climatique provoque l'arrêt de l'écoulement de la rivière, il y a alors une déconnexion des milieux souterrains et un arrêt des échanges génétiques avec la surface menant à la formation de nombreuses populations souterraines déconnectées pouvant amener une spéciation allopatrique par un processus d'isolement géographique. Ainsi, à partir d'une espèce de surface on obtient plusieurs espèces indépendantes malgré des pressions de sélection identiques. Ce modèle prévoit que toute la biodiversité souterraine est issue de la biodiversité de surface. Il existe d'autres modèles de relictés climatiques (Boutin et Coineau, 1990; Notenboom, 1988; Holsinger, 1994). Tous ces modèles font appel à une cause extrinsèque (glaciation, assèchement, régression,...) interrompant les flux géniques entre populations épigées et hypogées. Ces modèles proposent une spéciation allopatrique concomitante à la colonisation du milieu. On s'attend dans les modèles précédents à ce qu'il y ait très peu d'espèces de surface et un grand nombre d'espèces souterraines. Cependant, on trouve parfois au même endroit, des espèces de surface et des espèces souterraines. Le modèle de saut adaptatif proposé par Howarth (1987) explique ce phénomène par un processus de sélection différentielle entre le milieu de surface et le milieu souterrain, amenant à une sélection disruptive. Ce modèle prévoit que si le taux de migration vers le milieu souterrain est faible alors l'homogénéisation n'est pas suffisante, permettant une spéciation. Pour des organismes qui ont des potentialités à survivre dans le milieu souterrain ce processus peut se produire un grand nombre de fois.

Les approches standards d'inférence des changements de milieux ou de traits par modélisation le long de la phylogénie ne sont pas adaptées au modèle souterrain. En effet, les principales hypothèses expliquant la colonisation du milieu

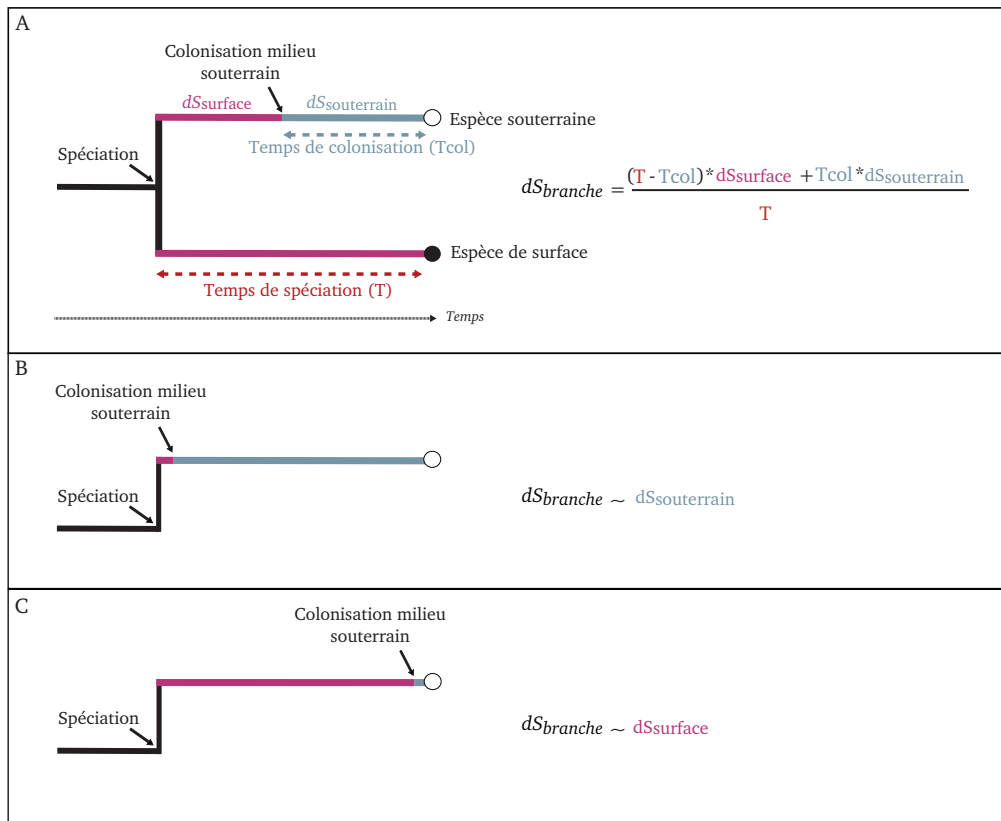


FIGURE 3.11 – Effet du temps de colonisation ( $T_{col}$ ) du milieu souterrain sur l'estimation du  $d_S$  calculée sur la branche de l'arbre. La longueur de la branche correspond au temps de spéciation ( $T$ ). Le  $d_S$  de la branche est une moyenne du  $d_S$  d'une espèce de surface et du  $d_S$  d'une espèce souterraine, pondérée par le temps passé dans chaque habitat (A). Si une espèce a colonisé le milieu peu de temps après la spéciation, le  $d_S$  calculé sur la branche sera proche de celui d'une espèce souterraine (B). Si la colonisation est très récente par rapport au temps de spéciation, le  $d_S$  calculé sur la branche sera proche de celui d'une espèce de surface (C).

souterrain impliquent une extinction de l'espèce de surface ancestrale (Box 1). Ainsi, dans la phylogénie, on trouve de grands groupes ne contenant que des espèces souterraines probablement en raison de la disparition des espèces de surface, rendant une approche par parcimonie erronée. En effet, une telle approche supporterait un scénario dans lequel une seule transition très ancestrale a eu lieu.

Un autre moyen d'estimer des changements majeurs (transitions écologiques, changement de régime alimentaire,...) est l'utilisation de la pseudogénération. En

effet, après des transitions importantes, il arrive que certaines fonctions n'aient plus aucune utilité. On peut citer par exemple les bactéries endosymbiotiques qui perdent un grand nombre de fonctions redondantes avec leur hôte (Gil et al., 2003; Ham et al., 2003), la perte des dents chez les mysticètes (DavitBéal et al., 2009) ou encore la perte des récepteurs de la saveur umami chez les carnivores (Jiang et al., 2012). En lien avec la régression de ces traits, les gènes codant pour ces traits ne sont plus sous sélection et passent d'un régime de sélection purifiante à un régime de sélection neutre. En comparant l'intensité de la sélection qui s'exerce sur des fonctions devenues inutiles en milieu souterrain entre des espèces de surface et des espèces souterraines, on peut estimer le temps de colonisation de ces espèces souterraines. L'intensité de la sélection naturelle peut être estimée en calculant le ratio du taux de substitutions non-synonymes ( $d_N$ ) sur le taux de substitutions synonymes ( $d_S$ ). Dans le cas d'une sélection purifiante les mutations modifiant la séquence protéique (mutations non-synonymes) sont en grande partie éliminées, ainsi le ratio  $d_N/d_S$  est proche de zéro. Dans le cas d'une sélection positive, les mutations non-synonymes sont favorisées par la sélection,  $d_N/d_S$  est alors supérieur à 1. Dans le cas d'une absence de sélection, le taux de substitutions non-synonymes dépend simplement du taux de mutations, tout comme le taux de substitutions synonymes,  $d_N/d_S$  vaut donc 1. Ainsi, en considérant que le gène est sous sélection purifiante chez les espèces de surface actuelles, on peut calculer un  $d_N/d_S$  du gène avant la colonisation (*i.e.* on fait l'hypothèse qu'avant de coloniser le milieu l'espèce avait un régime de sélection identique aux espèces de surface actuelles). On considère qu'après la colonisation, le gène évolue de façon neutre, le  $d_N/d_S$  de ce gène vaut donc 1. Le  $d_N/d_S$  de ce gène estimé pour l'espèce souterraine sur la branche terminale correspond à un  $d_N/d_S$  moyen depuis la spéciation. Ce  $d_N/d_S$  moyen est donc une moyenne du  $d_N/d_S$  de surface et du  $d_N/d_S$  souterrain, pondéré par le temps passé dans chaque milieu (Figure 3.12) :

$$d_N/d_{S_{\text{Branche}}} = \frac{d_N/d_{S_{\text{Surface}}} * T_{\text{Surface}}}{\text{Temps}} + \frac{d_N/d_{S_{\text{Souterrain}}} * T_{\text{Souterrain}}}{\text{Temps}}$$

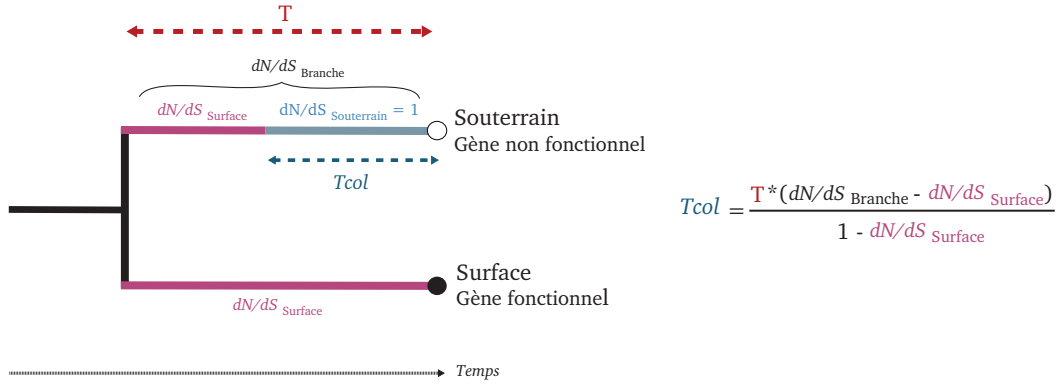


FIGURE 3.12 – Principe du calcul du temps de colonisation en utilisant les valeurs de  $d_N/d_S$  calculées sur un gène devenant non-fonctionnel en milieu souterrain.

Le  $d_N/d_{S_{\text{Souterrain}}}$  vaut 1, le temps passé en souterrain correspond au temps de colonisation ( $T_{\text{col}}$ ), le temps passé en surface sur la branche correspond au temps depuis la spéciation ( $T$ ) moins le temps de colonisation.

$$d_N/d_{S_{\text{Branche}}} = \frac{d_N/d_{S_{\text{Surface}}} * (T - T_{\text{col}}) + T_{\text{col}}}{T}$$

$d_N/d_{S_{\text{Branche}}}$ ,  $d_N/d_{S_{\text{Surface}}}$ , et  $T$  peuvent être estimés à partir d'un alignement et d'une phylogénie calibrée dans le temps. Nous pouvons ainsi calculer le temps de colonisation :

$$T_{\text{col}} = \frac{T * (d_N/d_{S_{\text{Branche}}} - d_N/d_{S_{\text{Surface}}})}{1 - d_N/d_{S_{\text{Surface}}}}$$

Ce qui nous intéresse n'est pas le temps absolu depuis lequel l'espèce est souterraine mais la proportion de la branche pour laquelle l'espèce est souterraine ( $T_{\text{col}}/T$ ). Ainsi, le calcul d'un temps absolu n'est pas nécessaire, la longueur de branche (nombre de substitutions par site) suffit.

La plupart des gènes candidats pour étudier l'évolution régressive sont des gènes codant pour des photorécepteurs liés à la vision. Alors que certains au-



teurs ont montré des évidences de non-fonctionnalisation de ses gènes chez les espèces souterraines (Leys Remko et al., 2005; Niemiller et al., 2012; Jacobs, 2013), d'autres études n'ont montré aucun signe de non-fonctionnalisation (Crandall et Hillis, 1997; Friedrich et al., 2011). La plupart des études ont cependant été faites avec très peu d'espèces (souvent une espèce de surface et une souterraine) et parfois sur des espèces ayant divergées très récemment, ne laissant pas assez de temps pour que des mutations non-fonctionnalisantes se produisent et se répandent dans la population. Afin de rechercher des gènes candidats pouvant être utilisés dans le calcul du temps de colonisation, des gènes liés à des fonctions qui régressent en milieu souterrain ont été sélectionnés. Ainsi, des gènes liés à la vision ainsi qu'au rythme circadien ont été sélectionnés dans la littérature. Une étape d'amplification et de séquençage de ces gènes a ensuite été effectuée afin de tester si ces gènes montrent des signes de régression en milieu souterrain : soit en cherchant des signes de non-fonctionnalisation (insertions délétions provoquant un décalage du cadre de lecture ou codons stop dans la séquence codante), soit en testant une augmentation du  $d_N/d_S$  en milieu souterrain. A ce jour, les seuls gènes montrant une non-fonctionnalisation évidente chez les espèces souterraines sont les gènes des opsines (photorécepteurs).

Nous avons ainsi estimé le temps depuis lequel une espèce a colonisé le milieu souterrain en utilisant la non-fonctionnalisation d'un des gènes de l'opsine. En effet, de façon concomitante avec la régression du système oculaire, certaines espèces souterraines présentent des mutations non-fonctionnalisantes sur le gène de l'opsine (Niemiller et al., 2012) qui indiquent une perte de contrainte fonctionnelle. Pour 89 espèces, nous avons pu reconstruire la séquence d'un gène de l'opsine (deux paralogues de l'opsine sont connus chez les Asellidae). Les séquences ont été déterminées en utilisant des séquences issues (i) des transcriptomes, (ii) de séquençage Sanger en utilisant les primers PCR (LWF1a et Scylla) définis dans Taylor et al. (2005) et (iii) des lectures Illumina issues des génomes. Ces

lectures Illumina ont été alignées sur la séquence de l'opsine de l'espèce la plus proche en utilisant BWA (algorithme d'alignement) (Li et Durbin, 2009) et une séquence consensus a été créée en utilisant la suite SAMtools (Li et al., 2009). Les séquences ont été alignées de visu en raison de la présence d'un intron et d'une grande insertion chez les espèces du groupe *Proasellus cavaticus*. Une séquence codant pour une protéine fonctionnelle a été utilisée comme référence afin de déterminer le cadre de lecture. Dans le cas où plusieurs espèces montraient des signes identiques de non fonctionnalisation (même mutation non-sens, délétion ou insertion), une seule séquence a été conservée pour le calcul du temps de colonisation. En effet, le calcul du temps de colonisation ne fonctionne que si la colonisation a eu lieu le long de la branche terminale. Si la diversification s'est faite après la colonisation, il est donc nécessaire de ne conserver qu'une seule branche. L'intron, les insertions et les frameshift ont été supprimés avec le script `addCodonStructureToAln.pl` et `struct2codonAln.pl` du module *BioPseudogene* (<https://github.com/TristanLefebure/BioPseudogene>) afin de ne conserver que la séquence codante du gène. Les  $d_N/d_S$  ont ensuite été calculés avec la suite *Bio++* (Dutheil et Boussau, 2008). Un modèle non-homogène (modèle NY68) a été premièrement appliqué à l'alignement avec *BppML* puis le programme *MapNH* (Version 1.1.1) du package *TestNH* (Guéguen et Duret, 2018) a été utilisé pour reconstruire les taux de substitutions synonymes et non-synonymes sur chaque branche. La phylogénie calibrée dans le temps a ensuite été utilisée pour calculer le temps de colonisation. Enfin ce temps de colonisation a été divisé par le temps de spéciation afin d'obtenir la proportion de la branche qui correspond à une espèce souterraine. En effet, c'est bien cette proportion qui nous indique si le taux de substitutions calculé sur cette branche correspond à un état majoritairement souterrain ou de surface.

**Contributions** De nombreuses données moléculaires utilisées dans cette thèse sont issues de nombreuses années de travail effectué par l'équipe de recherche

### CHAPITRE 3. LES ASELLIDAE SOUTERRAINS : DÉVELOPPEMENT DU MODÈLE POUR L'ÉTUDE DES TAUX D'ÉVOLUTION MOLÉCULAIRE.

---

Ecologie et Evolution des Écosystèmes Souterrains du LEHNA. Ainsi, une grande partie des données a été générée en amont de cette thèse. Les individus utilisés pour le séquençage ont été en grande partie échantillonnés par les membres du laboratoire depuis 2008 au cours de grandes campagnes d'échantillonnage auxquelles j'ai participé ces 4 dernières années, mais sont également issus de nombreux envois de la part de la communauté scientifique, en raison de l'expertise de l'équipe sur ce groupe. Le travail de laboratoire a, en très grande partie, été effectuée par Lara Konecny-Dupré, ma participation sur cette partie a été anecdotique. L'assemblage des transcriptomes a été effectuée par Tristan Lefébure. J'ai effectué la délimitation des gènes orthologues pour le chapitre 5 mais celle du chapitre 4 provient du travail de thèse de Clémentine François et celle du Chapitre 6 a été effectuée par Laura Grice. J'ai effectué l'assemblage de tous les génomes mitochondriaux. Le séquençage Sanger des gènes de l'opsine a été effectué en grande partie avant mon arrivée au laboratoire. J'ai encadré plusieurs stages qui ont consisté à chercher d'autres gènes se non fonctionnalisant, mais toutes les tentatives se sont soldés par des échecs.

# Chapitre 4

## Impact des traits d’histoire de vie sur les taux d’évolution moléculaire

Ce chapitre a fait l’objet d’une publication dans *Molecular Biology and Evolution*, Volume 35, Issue 12, Décembre 2018, Pages 2900–2912, présentée en Annexe 6.

### 1 Introduction

Nous avons vu, dans le premier chapitre, que la plupart des hypothèses cherchant à expliquer les variations de taux de substitutions observées entre les espèces étaient basées sur les traits d’histoire de vie des organismes. L’hypothèse du temps de génération (Li et al., 1987) explique les différences de taux de substitutions entre les mammifères par le fait que, toutes choses étant égales par ailleurs, un organisme dont le temps de génération est plus long copiera son ADN moins de fois sur un même intervalle de temps qu’une espèce dont le temps de généra-

tion est court. Ainsi, les erreurs liées à la réplication se produiront moins souvent, engendrant un taux de mutations par millions d'années, plus faible. Cet effet a été montré chez les mammifères (Ohta, 1993; Bromham et al., 1996), les oiseaux (Mooers et Harvey, 1994), les protostomiens (Thomas et al., 2010) et les plantes (Smith et Donoghue, 2008). L'hypothèse du taux métabolique (Martin et al., 1992) suggère que les espèces de faible poids génèrent une plus grande quantité d'espèces réactives de l'oxygène (ROS) que les espèces ayant un poids élevé, en lien avec leur taux métabolique par unité de masse plus élevé. Ces ROS ayant des effets mutagènes (Barja, 2002), une augmentation de leur production entraînerait une augmentation du nombre de mutations. L'hypothèse de longévité propose une interprétation évolutive du lien entre la production de ROS et le taux de substitutions (Nabholz et al., 2007). Cette hypothèse formalise, à une échelle évolutive, les prédictions de la théorie du vieillissement mitochondrial (Kujoth et al., 2007) qui propose que les mutations dans les cellules somatiques qui se produisent dans la mitochondrie affectent l'efficacité de la chaîne respiratoire mitochondriale. Cette baisse d'efficacité augmente la production de ROS, déclenchant un cercle vicieux qui amène à terme au vieillissement des organismes. A une échelle évolutive, les pressions de sélection qui s'exercent sur les espèces longévives devraient réduire le taux de mutations mitochondriales en diminuant l'impact des ROS, en augmentant la fidélité de la réplication ou en améliorant l'efficacité de la réparation. On voit donc apparaître un compromis entre le coût des mécanismes permettant de limiter les mutations et le fardeau mutationnel, dont l'équilibre dépend de la longévité de l'espèce.

Ces trois hypothèses montrent cependant des lacunes. L'hypothèse du temps de génération semble incapable d'expliquer la différence de taux d'évolution entre requins et primates, malgré des temps de génération similaires (Martin et al., 1992). Alors que certains travaux montrent un effet du taux métabolique (Mooers et Harvey, 1994; Bromham et al., 1996), Lanfear et al. (2007), en incluant une

large variété de taxons dans leur analyse, ne montrent pas de relation entre le taux métabolique et le taux de substitutions. Ce résultat ne remet pas en cause l'effet mutagène des ROS mais montre que la relation entre ROS, taux métabolique et taux de substitutions est loin d'être claire. En revanche, aucun élément ne permet, aujourd'hui, de remettre en cause l'hypothèse de longévité, si ce n'est qu'elle n'a jamais été testée en dehors des mammifères.

Afin de tester ces différentes hypothèses nous avons utilisé la transition vers le milieu souterrain effectuée indépendamment par différentes espèces d'Asellidae. Nous avons vu dans le premier chapitre que cette transition écologique induit l'apparition, chez les espèces souterraines, d'une modification importante des traits d'histoire de vie avec un allongement de la longévité (de 5 à 15X), un retardement de la maturité sexuelle et une baisse du nombre de descendants (Henry, 1976). Cette famille nous permet, de plus, de disposer d'une phylogénie bien résolue (Morvan et al., 2013), d'un nombre suffisant de comparaisons, ainsi que d'un temps de divergence qui permet des estimations robustes des taux d'évolution moléculaire. Cependant, la transition d'habitat peut varier fortement entre les espèces. Comme expliqué dans la partie 2.4 du Chapitre 3, ces différences de temps de colonisation peuvent impacter la corrélation entre le taux de substitutions et les traits de ces espèces, il est donc important d'estimer la date de colonisation et d'en tenir compte dans les analyses.

L'hypothèse de longévité et l'hypothèse du taux métabolique impliquent que les taux d'évolution devraient être plus faibles chez les espèces longévives, avec un ralentissement plus marqué dans le compartiment mitochondrial que nucléaire : le vieillissement étant lié aux mutations mitochondriales, le ralentissement devrait être plus important dans ce compartiment. En revanche l'hypothèse du temps de génération prévoit un effet similaire dans les deux compartiments. Quelle que soit l'hypothèse, l'effet devrait être d'autant plus marqué que l'espèce a colonisé le milieu depuis longtemps (voir Chap. 3 partie 2.4).

Cette transition est toutefois associée à des facteurs confondants et à des biais méthodologiques pouvant brouiller l’empreinte des traits d’histoire de vie sur les taux de substitutions. Premièrement, les rayonnements ultraviolets (UV) étant reconnus pour leur effet mutagène, ils pourraient augmenter le taux de mutations des espèces de surface. Il est cependant possible de tester leur effet grâce à leur signature mutationnelle. Deuxièmement, pour avoir une estimation du taux de mutations, le taux de substitutions synonymes ( $d_S$ ) est utilisé, celui-ci étant considéré comme neutre (Kimura, 1968). Toutefois le  $d_S$  n’est effectivement neutre qu’en l’absence de sélection traductionnelle. Cette sélection engendre un choix préférentiel entre les codons qui codent pour un même acide aminé, en favorisant ceux qui correspondent aux ARNs de transfert les plus abondants, favorisant ainsi la rapidité de la traduction (Ikemura, 1982). Enfin, le génome mitochondrial évolue beaucoup plus rapidement que le génome nucléaire et possède beaucoup moins de gènes. Il est ainsi plus sensible aux problèmes de saturation (plus d’une mutation sur un site de l’alignement) et il a moins de puissance que le génome nucléaire pour détecter des différences de taux d’évolution. La puissance et la saturation sont donc à prendre en compte dans l’analyse.

Nous avons donc cherché à comparer les taux d’évolution nucléaires et mitochondriaux de 13 couples d’espèces phylogénétiquement proches, mais vivant en environnements contrastés, afin de comprendre l’effet d’une augmentation de la longévité sur les taux de substitutions. En utilisant des gènes se non-fonctionnalisant en milieu souterrain, nous avons pu dater les évènements de colonisation de ce milieu, et ainsi, prendre en compte ce paramètre dans notre analyse. L’utilisation d’un proxy du taux métabolique nous a permis de tester la relation entre le métabolisme et le taux de substitutions. Enfin, nous avons également estimé l’influence possible de facteurs confondants (UV et sélection traductionnelle) ou de biais analytiques (saturation et puissance) sur nos résultats.

## 2 Matériel et Méthodes

**Données moléculaires et calcul des taux de substitutions** En utilisant une large phylogénie des Asellidea (Morvan et al., 2013), 13 paires d'espèces, contenant chacune une espèce épigée et une espèce hypogée, ont été délimitées, représentant ainsi 13 transitions indépendantes vers le milieu souterrain. Pour chacune des 26 espèces, les individus ont été récoltés dans leur environnement et congelés vivants dans les minutes suivant leur collecte. Pour chaque espèce, les données transcriptomiques de Francois et al. (2016a) ont été utilisées et le génome mitochondrial a été assemblé (voir Chap. 3, partie 2.3), nous permettant d'obtenir 382 gènes nucléaires et 12 gènes mitochondriaux. Les gènes ont été alignés avec PRANK (Löytynoja et Goldman, 2008), et les sites potentiellement mal alignés ont été masqués avec GBlocks (Castresana, 2000). Les gènes ont ensuite été concaténés indépendamment pour le compartiment nucléaire et mitochondrial. Un arbre phylogénétique a été construit avec PhyML v3.0 (Guindon et al., 2010) sous un modèle GTR+G+I. Les taux de substitutions synonymes par unité de temps ont été estimés en utilisant le logiciel CoEvol (Lartillot et Poujol, 2011). Ce programme implémente un modèle de codon Muse et Gaut (Muse et Gaut, 1994), avec une variation Brownienne du  $d_S$  et du  $d_N/d_S$  le long des branches de l'arbre.

**Temps de colonisation** Nous avons estimé le temps depuis lequel l'espèce a colonisé le milieu souterrain en utilisant la non-fonctionnalisation du gène de l'opsine (voir Chap. 3, partie 2.4). Pour 23 espèces, nous avons pu reconstruire la séquence du gène de l'opsine. Les  $d_N/d_S$  des opsines ont été estimés en utilisant le programme CodeML du package PAML (Yang, 2007) et ont ensuite été utilisés pour calculer la proportion de la branche qui correspond à une espèce souterraine (RCT).

**Taux métabolique** Sous l'hypothèse que l'activité de la chaîne respiratoire influence le taux de mutations, le taux métabolique devrait être corrélé au  $d_S$ . Pour



tester cette hypothèse, le ratio RNA/protein a été utilisé comme proxy du taux métabolique. Ce ratio est un bon proxy du nombre de ribosomes actifs par protéine synthétisée (Cox, 2003; Karpinets et al., 2006). Lorsqu'il est mesuré à l'échelle de l'organisme, ce ratio reflète la vitesse à laquelle les protéines sont renouvelées. Une corrélation positive entre ce ratio et le taux de croissance a été observée chez des organismes variés (Neidhardt et Magasanik, 1960; Leick, 1968; Cox, 2003; Karpinets et al., 2006; Wojewodzic et al., 2011; Houlihan et al., 1990; Pierce et al., 1999; Mathers et al., 1992; Peragón et al., 2001; Wagner et al., 2001; McKEE et Knowles, 1987; Xing et al., 2016). Ce ratio devrait donc être positivement corrélé avec le taux métabolique, même si cette relation devrait être validée chez les isopodes. Pour 16 espèces, nous avons récupéré le ratio RNA/protein mesuré dans Lefébure et al. (2017) : pour 10 individus par espèce, l'ARN total et les protéines totales ont été extraits avec un protocole TRI-Reagent (Molecular Research Center). L'ARN total a été quantifié par fluorimétrie (Qubit ; Life technologies) et les protéines totales ont été obtenues en utilisant l'acide Bicinchoninique (Smith et al., 1985). Le ratio ARN/protein a été estimé par l'ARN total normalisé par la biomasse de protéine.

**Rayonnements Ultraviolets** Une différence environnementale majeure entre espèces de surface et souterraines, directement susceptible d'affecter les taux de substitutions est la présence de rayonnement ultra-violet (UV). L'effet de ces rayonnements a été testé en utilisant le patron de substitutions spécifique des UV. Ces rayonnements provoquent une photo-excitation des sites dipyrimidiques (*i.e.* composés de cytosines et/ou de thymines) qui forme une liaison covalente à la place de la liaison hydrogène entraînant une torsion de la molécule d'ADN et empêchant la transcription. Les polymérases capables de transcrire ces molécules endommagées remplacent les deux pyrimidines par des T (Brash, 2015). Ainsi, une différence d'effet des UVs devrait se traduire par une différence de composition en dinucléotides CpT, TpC CpC ou TpT. Cette différence peut être dans les deux

sens : l'espèce la plus exposée pourrait présenter un plus grand nombre de sites TT par effet mutagène ou pourrait montrer un déficit de sites dipyrimidiques, la sélection naturelle favorisant les sites non sensibles aux UV. Dans le but de détecter un biais de composition en dinucléotides CpT, TpC CpC ou TpT, la statistique *Zscore*, proposée par (Palmeira et al., 2006) dans le package sequinR (Charif et Lobry, 2007), a été calculée pour chaque espèce sur la concaténation des 382 gènes sous un modèle 'syncodon' avec 1000 permutations. Cette méthode permet de tenir compte de la fréquence de chaque nucléotide de la séquence ainsi que de la structure en codons, ce qui permet d'obtenir des résultats similaires à ceux obtenus sur des régions non codantes. Une différence de *Zscore* entre espèces de surface et souterraines a été testée par un test des rangs signés de Wilcoxon pour données appariées, permettant de considérer la structure en paire d'espèces. Les *Zscores* n'ont été calculés que sur les gènes nucléaires car le système de réparation du génome mitochondrial ne répare pas les mutations liées aux UV, les mitochondries endommagées étant éliminées (Clayton et al., 1974).

**Sélection traductionnelle** Pour un même acide aminé, certains codons peuvent être préférentiellement choisis par rapport à d'autres pour des raisons d'efficacité de la traduction (rapidité et fidélité). Cette sélection traductionnelle a pour effet de rendre certaines substitutions synonymes non neutres et potentiellement d'invalider les conclusions sur les taux d'évolution moléculaire calculés sur l'ensemble des positions synonymes. Afin de vérifier l'absence de biais dus à cette sélection, l'usage relatif des codons synonymes (RSCU) a été calculé pour les 10% de gènes les plus exprimés et pour les 10% de gènes les moins exprimés. En cas de sélection traductionnelle, les gènes les plus exprimés présentent la plus forte fréquence de codons optimaux. Le RSCU permet de comparer la fréquence attendue d'un codon, en fonction de la composition en base du génome, et la fréquence réellement observée. La différence de RSCU entre gènes fortement et faiblement exprimés a ensuite été calculée, permettant ainsi de détecter les codons favorisés ou évités.

Pour les gènes mitochondriaux, n'ayant que 13 gènes qui sont tous transcrits en un seul bloc (et n'ont donc pas ou peu de différence d'expression), le calcul d'un RSCU est impossible. Le nombre effectif de codons (ENC) a été calculé en utilisant le package `seqinR` du logiciel R. L'ENC varie entre 20 et 61, 20 indiquant qu'un codon unique est utilisé pour chaque acide aminé et 61 que tous les codons sont utilisés équiprobablement pour un acide aminé. L'ENC ne permet pas de détecter la sélection traductionnelle. Néanmoins, il permet de détecter un biais d'usage du code qui peut être lié à la sélection traductionnelle.

**Saturation** Les gènes mitochondriaux évoluant beaucoup plus rapidement que les gènes nucléaires, l'estimation des taux de substitutions peut être mauvaise du fait de problèmes de saturation (plusieurs substitutions sur le même site). Afin d'évaluer l'impact de cette saturation, des graphiques de la divergence observée (nombres de différences observées entre deux séquences) en fonction de la divergence estimée sous un modèle d'évolution GTR + G + I (distance patristique) ont été tracés en différenciant les taux de transversions et de transitions et les différentes positions des codons. Les transversions (*i.e.* substitution d'une purine -A ou G- vers une pyrimidine -C ou T- ou inversement) étant plus rares que les transitions (*i.e.* substitution d'une purine vers une autre purine ou d'une pyrimidine vers une autre pyrimidine), elles ont tendance à être moins saturées. Sur la base de ces graphiques, et afin de limiter l'effet de la saturation dans les gènes mitochondriaux, seules les transversions synonymes ont été utilisées pour estimer les taux d'évolution dans le génome mitochondrial. Pour cela, la suite Bio++ (Dutheil et Boussau, 2008) a été utilisée : les paramètres du modèle d'évolution ont été optimisés avec le programme BppML avec un modèle non-homogène (Modèle NY68), puis le programme MapNH du package TestNH (version 1.1.1) a permis de reconstruire les états ancestraux et d'estimer le nombre de transversions synonymes. Les transversions synonymes ainsi obtenues ont été divisées par les longueurs de branches du chronogramme afin d'obtenir un taux comparable entre

toutes les espèces.

**Puissance** Le nombre de gènes mitochondriaux étant largement inférieur au nombre de gènes nucléaires (12 contre 382), des différences de résultats pourraient venir d'une différence de puissance entre les deux compartiments. Afin de tester cette hypothèse, un ré-échantillonnage de codons a été effectué dans l'alignement des gènes nucléaires. La vitesse d'évolution des deux compartiments génomiques étant différente, le signal porté par un codon n'est pas le même dans les deux compartiments. On ne peut donc pas simplement ré-échantillonner le compartiment nucléaire de façon à ce qu'il y ait le même nombre de sites que dans l'alignement mitochondrial, car il y aurait alors beaucoup plus de mutations dans la mitochondrie. Il faut donc échantillonner des codons nucléaires de façon à obtenir le même nombre moyen de substitutions que dans le génome mitochondrial. Ainsi, 1000 alignements de 26566 codons ont été échantillonnés dans l'alignement nucléaire. Sur chaque échantillon, le nombre de transversions synonymes a été calculé avec MapNH puis divisé par le chronogramme. La moyenne du contraste du nombre de transversions synonymes [ $S = \frac{1}{13} \sum_{i=1}^{13} \log\left(\frac{TvS_{surface_i}}{TvS_{subterranean_i}}\right)$ ] a été calculée pour chaque échantillon et sur la concaténation des gènes mitochondriaux afin d'estimer la probabilité que la concaténation de gènes mitochondriaux soit équivalente à un sous-échantillon des gènes nucléaires.

### 3 Résultats

**Taux de substitutions** Sous l'hypothèse de longévité, on s'attend à observer une baisse du taux de substitutions synonymes ( $d_s$ ) chez les espèces souterraines dans les deux compartiments génomiques, avec une baisse plus fortement marquée dans le génome mitochondrial. Dans le compartiment nucléaire, on observe effectivement un taux de substitutions synonymes plus faible chez les espèces souterraines (PGLS p.value=0.019). En revanche, de façon surprenant, ce patron n'est

pas retrouvé dans le génome mitochondrial (PGLS,  $p.value=0.562$ , Figure 4.1).

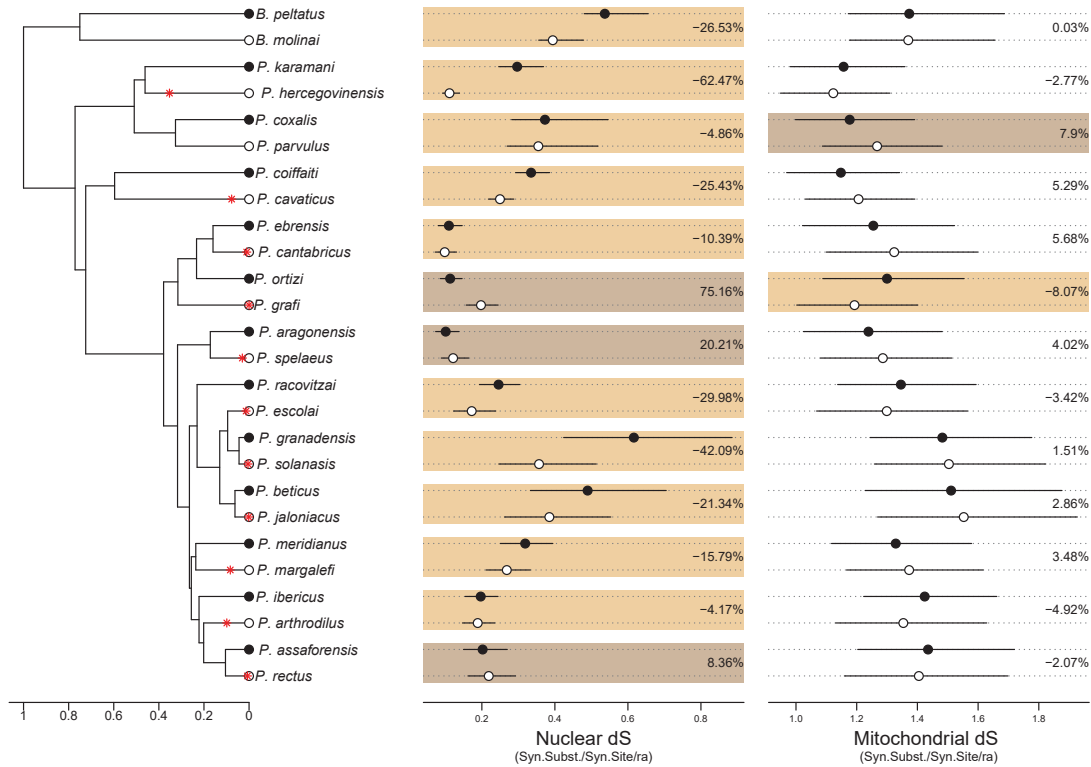


FIGURE 4.1 – Taux de substitutions des génomes nucléaires (centre) et mitochondriaux (droite) des 13 paires espèces d'Aselloidea (gauche) composées chacune d'une espèce de surface (cercle noir) et d'une espèce souterraine (cercle blanc). L'échelle de temps du chronogramme (gauche) est arbitrairement fixée de façon à ce que l'âge à la racine soit 1. Lorsqu'un gène de l'opsine a pu être obtenu pour l'espèce, une étoile rouge sur le chronogramme indique le temps de colonisation estimé avec la non-fonctionnalisation du gène de l'opsine. Les taux sont exprimés en taux de substitutions synonymes par site synonyme ( $d_S$ ) relativement à l'âge de la racine (ra) avec l'intervalle de crédibilité à 95% tel qu'estimé par CoEvol. Pour chaque paire, le pourcentage indique la variation chez l'espèce souterraine, relativement à l'espèce de surface. Lorsque cette variation a une probabilité postérieure d'être différente de zéro supérieure à 0.9, la paire d'espèces est surlignée avec une boîte colorée foncée (augmentation) ou claire (diminution)

**Temps de colonisation** Sous l'hypothèse d'une baisse du taux de substitutions en milieu souterrain, plus la proportion de la branche où l'espèce est souterraine est grande, plus le taux de substitutions doit être réduit. On observe en effet, dans

le génome nucléaire, qu'il y a une corrélation négative entre le taux de substitutions synonymes ( $d_S$ ) et le temps de colonisation relatif (RCT, proportion de la branche pour laquelle l'espèce souterraine, Figure 4.2, PGLS,  $p.value = 3.10^{-4}$ ). Dans le génome mitochondrial, aucune corrélation entre le taux de transversions synonymes et le RCT n'est retrouvée (PGLS,  $p.value = 0.562$ ), confirmant les résultats obtenus en utilisant l'habitat.

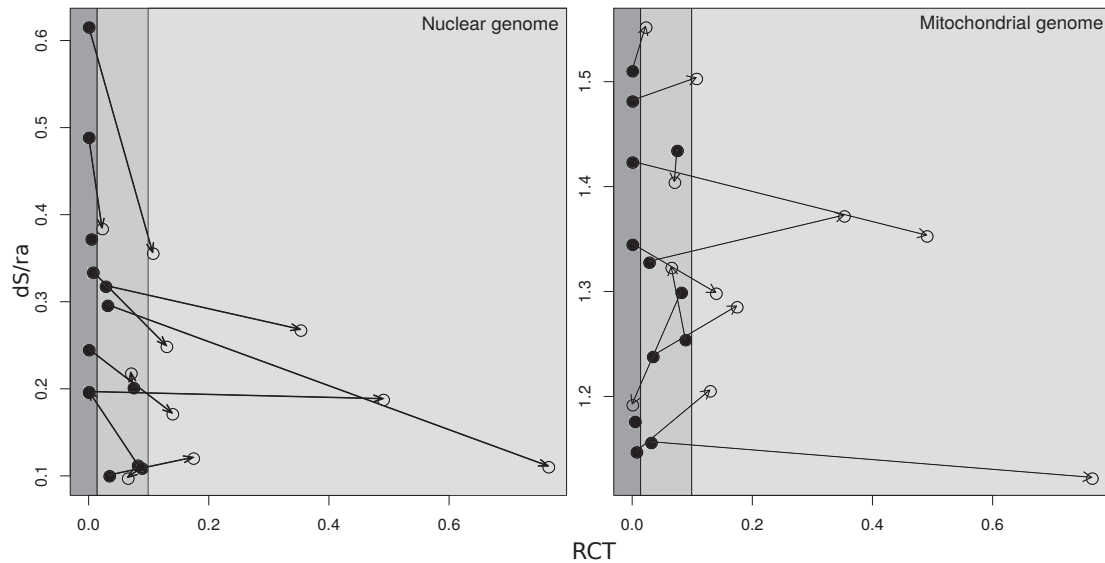


FIGURE 4.2 – Corrélation entre le taux de substitutions synonymes relatif à l'âge de la racine ( $d_S/ra$ ) et la proportion de la branche terminale qui correspond à une espèce souterraine (RCT) pour le génome nucléaire (gauche) et mitochondrial (droite). Les espèces souterraines sont représentées avec des cercles blancs et les espèces de surface avec des cercles noirs. Les deux espèces d'une paire sont reliées avec une ligne noire. Les différents niveaux de gris indiquent les espèces pour lesquels le mode de vie souterrain représente  $<1\%$  de la branche terminale (gauche) ou  $>10\%$  de la branche terminale (droite).

**Taux métabolique** Sous l'hypothèse que l'activité de la chaîne respiratoire influence le taux de mutations, le taux métabolique devrait être corrélé au  $d_S$ . Nous trouvons ici une corrélation positive entre le ratio ARN/protéine et le  $d_S$  nucléaire (Figure 4.3, PGLS,  $p.value = 7.10^{-4}$ ). En revanche aucune corrélation n'est trouvée avec le  $d_S$  mitochondrial (PGLS,  $p.value = 0.380$ ).

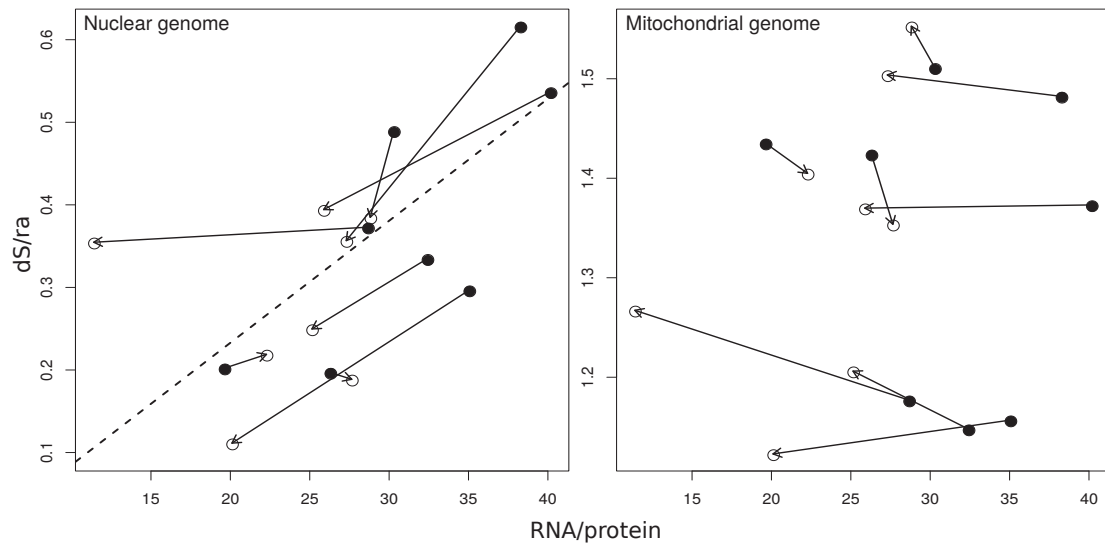


FIGURE 4.3 – Corrélation entre le taux de substitutions synonymes relatif à l'âge de la racine ( $d_S/ra$ ) et le ratio ARN/protéin pour le génome nucléaire (gauche) et mitochondrial (droite). Les espèces souterraines sont représentées avec des cercles blancs et les espèces de surface avec des cercles noirs. Les deux espèces d'une paire sont reliées avec une ligne noire. La ligne pointillée indique la régression linéaire estimée par un PGLS.

**Rayonnements ultra-violets** Aucun des quatre dinucléotides dipyrimidiques supposés être impactés par les UV (TpT, CpT, TpC ou CpC) n'est significativement sur ou sous-représenté (Figure 4.4, Wilcoxon, p.value > 0.05). Cette observation va à l'encontre de l'hypothèse selon laquelle les espèces soumises aux UV présenteraient plus de mutations de CC vers TT, ainsi que l'hypothèse selon laquelle la sélection amènerait à un évitement des sites dipyrimidiques.

**Sélection traductionnelle** La sélection traductionnelle engendre un choix préférentiel de certains codons synonymes pour des raisons de rapidité de la traduction. Pour estimer les taux d'évolution moléculaire on utilise le  $d_S$ , car on le considère comme une bonne estimation du taux de mutations, ce qui n'est vrai que si le  $d_S$  évolue de façon neutre. Lorsqu'il y a de la sélection traductionnelle, le  $d_S$  n'est plus neutre et ne peut donc plus être utilisé pour la mesure des taux d'évolution moléculaire. Au niveau des gènes nucléaires, on trouve une tendance non-significative vers une augmentation du nombre de codons optimaux chez les

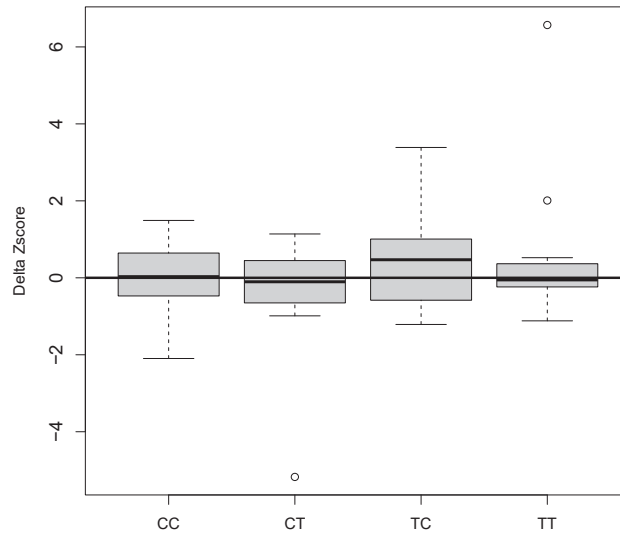


FIGURE 4.4 –  $\Delta Zscore$  ( $Zscore$  surface –  $Zscore$  souterrain) pour les quatre dinucléotides sensibles aux rayonnements UV. Le  $Zscore$  (Palmeira et al., 2006) calcule l’excès de chaque dinucléotide sur les positions synonymes en comparaison à un modèle neutre basé sur la fréquence en nucléotides. Si l’absence d’exposition aux UVs avait un effet direct sur le taux de mutations, une corrélation positive avec le  $\Delta Zscore$  devrait être trouvée.

espèces de surface (PGLS, p.value=0.18). Ces résultats indiquent que la sélection traductionnelle est très faible chez ces espèces. La sélection traductionnelle ne peut pas expliquer les différences de taux d’évolution moléculaire observées dans le génome nucléaire, car une augmentation de cette sélection chez les espèces de surface devrait réduire le  $d_S$  et non l’augmenter. Les valeurs d’ENC calculées sur les gènes mitochondriaux vont de 43 à 55, excepté pour *Bragasellus peltatus* qui a une valeur à 38. Les valeurs sont élevées (Wright, 1990), indiquant un faible biais d’usage. De plus, l’ENC n’est pas corrélé avec l’habitat (PGLS, p.value=0.38). En résumé, il n’y a aucune évidence que la sélection sur l’usage du code pourrait biaiser les  $d_S$  estimés, que ce soit dans le compartiment nucléaire ou mitochondrial.

**Saturation** La saturation entraîne une sous-estimation du nombre réel de substitutions qui peut gommer un signal existant. Le génome mitochondrial évoluant plus rapidement que le génome nucléaire, il est plus sujet à la saturation. Les ana-



lyses de saturations montrent que le génome mitochondrial présente une saturation importante au niveau des transitions synonymes. Cependant les transversions synonymes (ici estimés en utilisant la troisième position des codons) évoluent de façon monotone avec la distance des espèces, et ne sont donc pas saturées. Tout comme le  $d_S$ , le nombre de transversions synonymes ne va pas dans le sens d'un ralentissement du taux de substitutions en milieu souterrain tel qu'attendu sous l'hypothèse de longévité (PGLS, p.value=0,847) et il n'y a de corrélation significative ni avec le temps de colonisation (PGLS, p.value=0.881), ni avec le ratio ARN/protéine (PGLS, p.value=0.559).

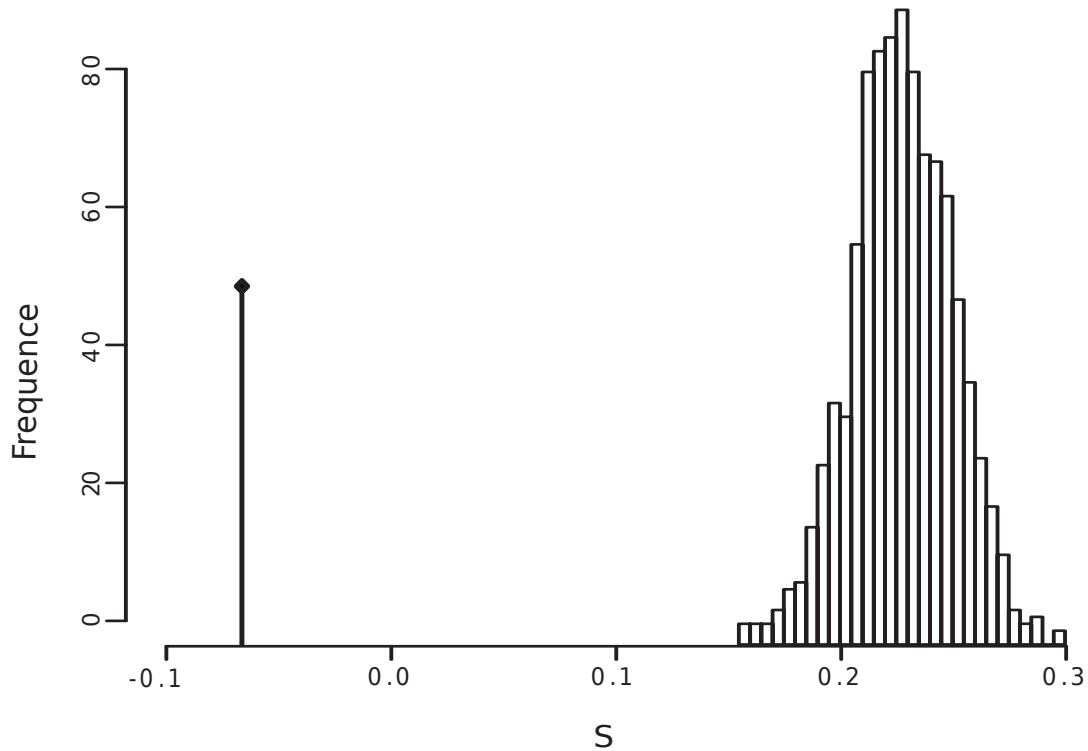


FIGURE 4.5 – Distribution de la statistique  $S = \frac{1}{13} \sum_{i=1}^{13} \log\left(\frac{TvS_{surface_i}}{TvS_{subterranean_i}}\right)$  permettant de tester si le découplage observé entre le taux de substitutions nucléaire et le taux de substitutions mitochondrial est dû à la différence de taille des jeux de données. Cette statistique a été calculée sur 1000 rééchantillonnages de 26566 codons dans l'alignement nucléaire (en blanc) et sur l'alignement mitochondrial (ligne noire).  $TvS_{surface_i}$  est le nombre de transversions synonymes estimé sur la branche terminale menant à l'espèce de surface de la paire  $i$  et  $TvS_{subterranean_i}$  est l'estimation analogue pour l'espèce souterraine de la paire.

**Puissance statistique** Le ré-échantillonnage du génome nucléaire montre que, malgré la réduction du nombre de sites pour le calcul, les espèces de surface conservent un taux d'évolution moléculaire plus élevé que les espèces souterraines (Figure 4.5 de l'article). En effet, pour tous les échantillons nucléaires, la moyenne des contrastes du nombre de transversions synonymes est comprise entre 0.18 et 0.35, alors que dans le génome mitochondrial, on observe une valeur proche de zéro (-0.0409). Ainsi la valeur de contraste de transversions synonymes, calculée sur la concaténation de gènes mitochondriaux, est très éloignée de la distribution des contrastes calculée sur les gènes nucléaires, malgré un même nombre de substitutions. L'absence de différence de taux de substitutions observée sur les gènes mitochondriaux ne provient donc pas d'un problème de puissance.

## 4 Discussion

**Reconsidération des principales hypothèses** Les Asellidae souterrains ont un taux de substitutions synonymes plus faibles que les Asellidae de surface dans le compartiment nucléaire, mais pas dans le compartiment mitochondrial. Cette observation ne correspond aux attendus d'aucunes des hypothèses présentes dans la littérature. Ni la saturation, ni le manque de puissance ou les facteurs confondants (UV, sélection traductionnelle) ne permettent d'expliquer ce découplage entre les deux compartiments génomiques. De plus aucune relation n'est trouvée entre le ratio ARN/protéine et le  $d_S$ , suggérant une absence de lien avec le taux métabolique. Ce résultat est en accord avec une étude précédente (Lanfear et al., 2007) qui n'avait trouvé aucune relation entre le taux de substitutions mitochondriale et le taux métabolique. Un impact du taux métabolique sur le compartiment nucléaire mais pas sur le compartiment mitochondrial est très peu probable car les agents mutagènes, les ROS, sont produits dans la mitochondrie. Le taux métabolique étant souvent fortement corrélé au temps de génération (Martin et al., 1992), la corrélation entre le  $d_S$  nucléaire et le ratio ARN/protéine suggère que le

temps de génération, et non le taux métabolique, impact le taux de substitutions, mais uniquement dans le compartiment nucléaire.

**Lien entre longévité, vieillissement et taux de mutations** On peut trouver deux explications à l'absence de relation entre la longévité et le taux de substitutions. Premièrement, une augmentation de la longévité peut ne pas être associée à une meilleure gestion des ROS. En effet, Jobson et al. (2010) ne trouvent pas de changements de  $d_N/d_S$  sur les gènes liés au stress oxydant chez les mammifères longévifs. En revanche, les auteurs notent une baisse du  $d_N/d_S$  sur les gènes qui codent pour les protéines membranaires. La composition de ces protéines modifie la résistance des tissus aux dégâts oxydatifs. Ainsi, une stratégie des espèces longévives consisterait, non pas à réduire la production des ROS, mais à mieux protéger les tissus, ce qui n'aurait pas d'impact sur les taux d'évolution moléculaire. Deuxièmement, une différence d'impact des ROS au niveau des cellules somatiques ne signifie pas que cette différence se retrouve dans les cellules germinales. L'hypothèse de longévité est fondée sur la théorie du vieillissement mitochondrial. Cette théorie prévoit que les ROS produits tout au long de la vie d'un individu engendrent un phénotype sénescence. Par analogie, et sous l'hypothèse que les mécanismes sont ubiquistes, l'hypothèse de longévité propose que cet effet individuel entraîne une pression de sélection sur les gènes, de façon à baisser le taux de mutations dans les cellules somatiques, engendrant une baisse du taux de mutations germinales. Or, l'activité de la chaîne respiratoire étant plus faible dans les cellules germinales que dans les cellules somatiques (Paula et al., 2013), la production de ROS y est plus faible. Une baisse du taux de mutations somatiques pourrait donc être totalement découplée d'une baisse du taux de mutations germinales. Ainsi les ROS n'ont potentiellement que peu d'impact sur le taux de substitutions. Le rejet de l'hypothèse de longévité ne remet donc pas en cause la théorie du vieillissement mitochondrial qui se situe à une échelle individuelle. En revanche, il remet en cause le lien entre cette théorie et les taux d'évolution

moléculaire.

**Découplage des deux compartiments génomiques** Contrairement aux hypothèses proposées jusqu'ici, le  $d_S$  mitochondrial ne semble pas évoluer sous l'influence des mêmes facteurs que le  $d_S$  nucléaire. Un tel découplage entre les deux compartiments génomiques n'est proposé par aucune hypothèse de la littérature. En effet, l'hypothèse de longévité et l'hypothèse du taux métabolique prévoient des variations des deux compartiments avec des amplitudes différentes, mais avec un sens commun, ce qui n'est pas observé ici. L'absence d'hypothèse prévoyant un découplage entre les deux compartiments peut venir du fait qu'elles ont été formulées à partir d'observations faites chez les mammifères ou les oiseaux qui présentent des particularités au niveau de la gamétogenèse. En effet, chez ces organismes, les ovocytes sont formés au stade embryonnaire, ils entament une méiose qui se bloque en prophase I, puis ils restent en quiescence jusqu'à la fécondation. Pendant ce blocage, les mitochondries ne se divisent pas ou peu (Paula et al., 2013; John et al., 2010), imposant au génome mitochondrial un arrêt des divisions correspondant au temps de génération. Chez les arthropodes, la gamétogenèse est continue, les cellules germinales primordiales (cellules souches gamétiques) font des mitoses durant toute la vie de l'animal, puis une partie entre périodiquement en méiose. Comme le génome mitochondrial se réplique indépendamment du génome nucléaire, il est possible que le nombre de réplifications des deux compartiments soit complètement découplé. L'absence de découplage observé chez les mammifères et les oiseaux pourrait venir du temps d'action de ce découplage. En effet chez les mammifères, ce temps d'action pourrait être très court (car les mitochondries sont à l'arrêt la majorité du temps), alors que chez les arthropodes, ce temps d'action pourrait être permanent. Ainsi, le découplage serait général mais pas détectable chez tous les organismes. Notons que les spermatogonies chez les mammifères montrent un processus de spermatogenèse similaire à celui des autres animaux. Cependant, la mitochondrie ayant une transmission maternelle stricte,

le nombre de divisions du génome mitochondrial ne dépend pas du processus de spermatogenèse.

Dans le cas des Asellidae, on observe un allongement du temps de génération en milieu souterrain. Les données disponibles sur ces espèces (Henry, 1976) semblent indiquer qu'il y a un allongement de tout le cycle vital de l'espèce en milieu souterrain (allongement de la durée de stade embryonnaire, retardement de la puberté), ainsi qu'une baisse du taux de croissance (estimé par le ratio ARN/protéine). Ainsi, on peut supposer qu'un allongement du temps de génération correspond à une diminution de la vitesse de réplication des cellules, qui entraîne une baisse du taux de substitutions nucléaire. Si cette baisse de vitesse de réplication nucléaire n'est pas liée à une baisse de réplication mitochondriale, alors elle explique l'absence de ralentissement observée dans le génome mitochondrial chez ces espèces (Figure 4.6).

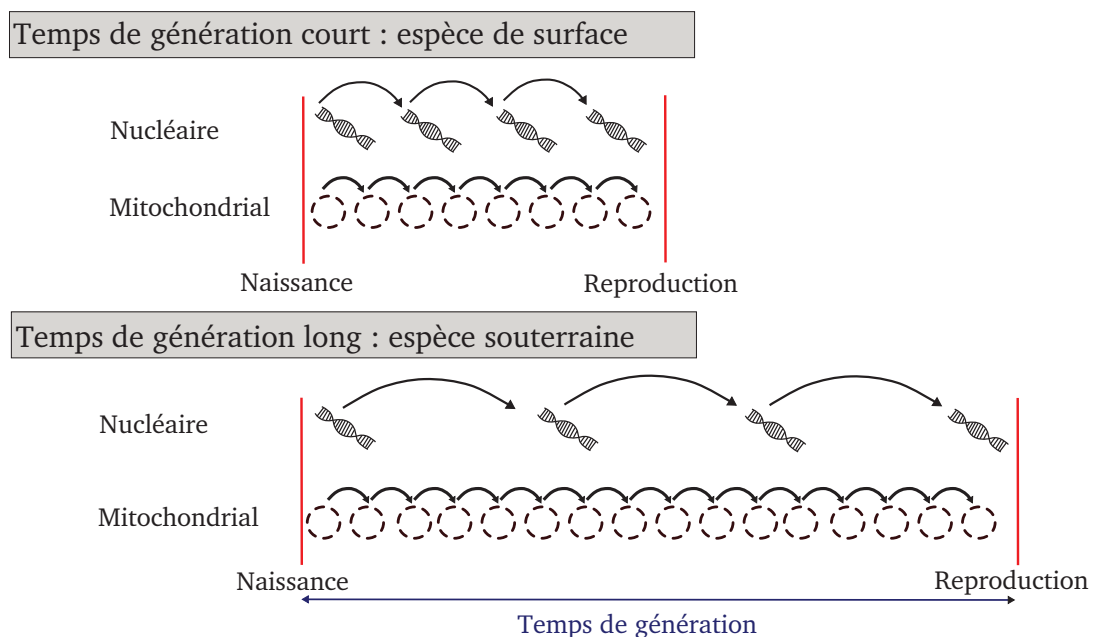


FIGURE 4.6 – Schéma de l'hypothèse d'indépendance des dynamiques de réplication du génome mitochondrial et du génome nucléaire chez les Asellidae. L'allongement du temps de génération serait accompagné d'un ralentissement des divisions cellulaires, entraînant un ralentissement des réplifications du génome nucléaire expliquant la baisse du  $d_S$  nucléaire. Les réplifications du génome mitochondrial seraient indépendantes de celle du génome nucléaire.

En considérant, le nombre moyen de réplifications subies par le génome jusqu'à la fécondation et le temps de génération, on peut donc formuler de façon synthétique, une expression globale du taux de substitutions synonymes par million d'années :

$$d_S/Ma = \frac{\text{Nombre moyen de réplifications} \times \text{Erreurs lors de la réplification}}{\text{Temps de génération}}$$

Le nombre de réplifications du génome jusqu'à la fécondation dépend de l'espèce et du compartiment génomique (mitochondrial ou nucléaire). Les erreurs lors de la réplification vont dépendre de l'efficacité des polymérase ou du système de réparation. Selon l'espèce et selon qu'on s'intéresse au compartiment nucléaire ou mitochondrial, l'importance de chaque élément sera différente en fonction des processus auxquels sont soumises les séquences ou selon la biologie et les traits d'histoire de vie des espèces. Ainsi, dans leur comparaison entre requins et primates, Martin et al. (1992) ont rejeté l'hypothèse du temps de génération en négligeant le nombre de réplifications et la quantité d'erreurs à chaque réplification.



# Chapitre 5

## Impact de l'environnement mutagène sur les taux d'évolution moléculaire

Ce chapitre a fait l'objet d'une publication soumise pour publication présentée en Annexe 7

### 1 Introduction

Alors que l'une des premières hypothèses expliquant les variations de taux d'évolution moléculaire était celle d'un effet de la radioactivité (Schlotteröer et al., 1991), l'effet des traits d'histoire de vie a occupé le champ thématique, laissant finalement peu de place à l'étude de l'environnement mutagène. C'est dans le champ d'étude de la carcinogenèse que les recherches ont mis en évidence des agents dans l'environnement capables de faire muter l'ADN tels que l'Arsenic (Tkeshelashvili et al., 1980), le Cadmium (Jin et al., 2003) ou encore le Radon (Shanahan et al., 1996). Ces études sur l'impact d'agents mutagènes sur les mu-



tations somatiques ne permettent cependant pas d'étendre leurs effets aux taux de substitutions qui dépendent des mutations germinales. Les évidences que les cellules germinales subissent également des dégâts liés à ses agents sont faibles. Alors que certaines études ont montré une augmentation des mutations germinales après une exposition à des agents mutagènes (Somers et al., 2002; Dubrova et al., 1996), d'autres études ont trouvé le résultat contraire (Satoh C et al., 1996; Czeizel et al., 1991). Ces résultats sont donc non seulement contradictoires mais ont aussi été obtenus suite à des expositions artificielles, soit accidentelles, soit lors de tests *in vitro*, les rendant difficilement généralisables à l'effet d'une exposition naturelle. Si cette littérature permet d'affirmer qu'il existe des agents mutagènes dans l'environnement, elle ne nous permet pas d'évaluer l'impact d'une exposition dans l'environnement sur les taux de substitutions.

Des corrélations encore mal comprises entre les taux de substitutions et certains facteurs environnementaux soulèvent la question de l'impact de l'environnement mutagène sur les taux de substitutions. Ainsi, Hebert et al. (2002) propose que la salinité pourrait impacter le taux de substitutions par son effet mutagène. De façon indirecte, la corrélation entre les taux de substitutions et la latitude (Wright et al., 2010) ou l'altitude (Lin et al., 2019) pourrait s'expliquer par l'effet mutagène des rayonnements UV. D'autre part, les environnements riches en agents mutagènes sont souvent des lieux où la biodiversité est très riche et endémique (*e.g.* zones équatoriales, sources hydrothermales, la forêt de Kerala en Inde). Ces exemples questionnent ainsi l'impact de l'environnement mutagène sur l'évolution de cette biodiversité. Cependant, ces environnements présentent d'autres particularités (température, productivité) qui pourraient expliquer cette richesse. Il existe en effet souvent des covariants importants dans l'environnement empêchant de montrer de façon évidente des relations de cause à effet, ce qui explique pourquoi la plupart des études ont utilisé une approche expérimentale, ne permettant pas d'étudier ces questions à une échelle évolutive.

La principale source d'exposition à des agents mutagènes exogènes pour les organismes (il existe des agents mutagènes produits par le métabolisme, voir Chap. 1, partie 1.1) provient de l'exposition aux radiations ionisantes. Les sources de radiations ionisantes sont : les rayonnements cosmiques, les rayonnements telluriques (radioactivité des roches), l'exposition au radon dans l'air, et la radioactivité naturelle présente dans tous les organismes vivants, notamment à cause du potassium 40 (United Nations Scientific Committee on the Effects Of Atomic Radiation, 2000). Ces radiations sont des rayonnements ayant une énergie assez importante pour transformer un atome en ion, c'est-à-dire que l'atome perd ou gagne un ou plusieurs électrons, devenant ainsi instable. De façon directe, les rayonnements peuvent affecter l'ADN en séparant le squelette sucre-phosphate, créant ainsi des cassure double brin (Hoeijmakers, 2001; Gent et al., 2001). Ces cassures sont difficiles à réparer car les deux brins étant rompu, il n'y a plus de brin matrice pour reconstituer l'information génétique. Le système de réparation est alors capable de restaurer la stabilité génomique mais au prix de nombreuses mutations (voir Chap. 1, partie 1.1). Les rayonnements peuvent également affecter les liaisons hydrogènes entre les bases d'ADN, rendant ainsi l'ADN instable (Collyns et al., 1965). De façon indirecte, les rayonnements ionisants provoquent la radiolyse des molécules d'eau présentes dans les cellules (Roots et Okada, 1975). Cette radiolyse correspond à la décomposition des molécules d'eau qui aboutit à la formation d'espèces chimiques sous forme de radicaux hydrogène ( $H\bullet$ ) et hydroxyle ( $HO\bullet$ ), mais aussi sous forme de peroxyde d'hydrogène ( $H_2O_2$ , qui donnera ensuite des radicaux  $HO\bullet$  par la réaction de Fenton), d'électrons libres, de dihydrogène ( $H_2$ ) et d'ions hydronium ( $H_3O^+$ ). Les radicaux libres formés ( $H\bullet$  et  $HO\bullet$ ) sont instables (ils ont un ou plusieurs électrons manquants) et interagissent avec l'ADN, provoquant des cassures à l'origine de mutations. L'effet mutagène de radiations à fortes doses est aujourd'hui bien établie (Dubrova et al., 1996; Ziegler et al., 1993) mais les effets multi-générationnels d'une exposition naturelle ont été très

peu étudiés.

L'étude des effets des radiations ionisantes naturelles sur l'évolution posent de nombreux problèmes expérimentaux. En effet, une étude *in vitro* ne permet pas de tester l'effet d'une exposition à une échelle évolutive. Sur des organismes multicellulaires, une telle étude prendrait plusieurs dizaines à plusieurs centaines d'années. Ainsi, il est nécessaire d'utiliser une approche de phylogénie comparative *in naturae* en comparant des espèces vivant avec des expositions différentes. Dans ce cadre, la seule source de radiations qui semble pouvoir être étudiée est celle provenant de l'exposition tellurique. En effet, l'exposition cosmique est extrêmement aléatoire et dépend d'évènements tels que les supernova. Il semble ainsi impossible de définir des espèces étant plus ou moins exposées à ces rayonnements. L'exposition par le radon dans l'air se produit majoritairement dans des endroits confinés et concerne principalement les humains. Enfin, la radioactivité présente dans la nourriture est impossible à contrôler *in naturae*. En revanche la radioactivité dans les roches est très variable en fonction des régions, permettant de caractériser des régions avec de fortes ou de faibles expositions. Cette radioactivité, même si elle décroît depuis 2 Milliards d'années, n'est pas aléatoire, elle dépend de la constitution des roches. En revanche, il faut s'assurer que les conditions telluriques dans lesquelles on trouve l'espèce actuellement correspondent aux conditions expérimentées par l'espèce depuis de nombreuses générations, et plus précisément sur le temps que représente la branche de l'arbre sur laquelle sera calculée le taux d'évolution moléculaire (voir Chap. 2, partie 1). En effet, si l'espèce a passé un long moment en zone non radioactive et qu'elle est arrivée très récemment dans la zone radioactive, le taux de substitutions mesuré le long de la branche reflètera davantage des conditions non radioactives.

En étudiant l'effet de la radioactivité tellurique sur des espèces d'Asellidae souterrains, on peut lever en grande partie ces difficultés. En effet, les Asellidae souterrains ont des capacités de dispersion faibles (voir Chap. 3, partie 1.4) ce

qui nous permet de supposer que les populations actuelles sont le résultat d'une évolution *in situ* depuis plusieurs milliers de générations. Ils sont répartis largement et sont trouvés dans des zones avec des niveaux de radioactivité naturelle variables. Leur phylogénie bien établie permet ainsi de mettre au point un jeu de données moléculaires avec des espèces phylogénétiquement proches mais vivant dans des environnements contrastés en terme de radioactivité.

Afin de tester l'effet de la radioactivité tellurique sur les taux de substitutions, nous avons sélectionné 14 espèces d'Asellidae vivant dans des environnements avec des niveaux de radioactivité contrastées. Un transcriptome par espèce a été séquencé et assemblé *de novo*. Le génome mitochondrial complet a également été assemblé, nous permettant de calculer les taux de substitutions dans chaque compartiment génomique. Afin d'évaluer plus finement l'impact de la radioactivité sur l'évolution des génomes, le patron mutationnel (la proportion de chaque type de mutation) a également été calculé à partir des transcriptomes de 8 individus par espèces. Pour chaque espèce, l'environnement mutagène a été caractérisé en mesurant tous les radionucléides présents dans les sédiments dans lesquelles les individus ont été récoltés. En partant du constat que ces espèces dispersent très peu, un proxy de la radioactivité régionale a également été utilisé afin d'évaluer la radioactivité à laquelle l'espèce a été exposée depuis des milliers de générations.

## 2 Matériel et Méthodes

Les protocoles, machines et normes utilisés pour les mesures de radioactivité et le séquençage sont détaillés dans l'article en Annexe 7

### Echantillonnage

En croisant la carte de la répartition de l'Uranium en France (Ielsch et al., 2017) et les données d'occurrence d'Asellidae (Données disponibles dans la base de données GOTIT, (Malard et al., 2019)), nous avons défini des zones d'échan-

tillonnage en France ayant à la fois une radioactivité naturelle élevée et des occurrences d'Asellidae. Peu d'espèces étant connues dans les environnements métamorphiques (riches en radioactivité naturelle), une étape importante de prospection s'est d'abord faite en particulier dans la région des Cévennes (Figure 5.1). D'autres campagnes d'échantillonnage ont ensuite été effectuées (Figure 5.1). Pour chacun des 58 sites où la présence d'Asellidae a été détectée, des sédiments ont été prélevés afin d'évaluer la radioactivité du site. Pour chacune de ces stations, les individus ont été identifiés morphologiquement puis moléculairement en séquençant le gène 16S qui a ensuite été comparé aux séquences connues d'Asellidae. La radioactivité de chaque site a été estimée en mesurant la radioactivité alpha globale dans les sédiments.

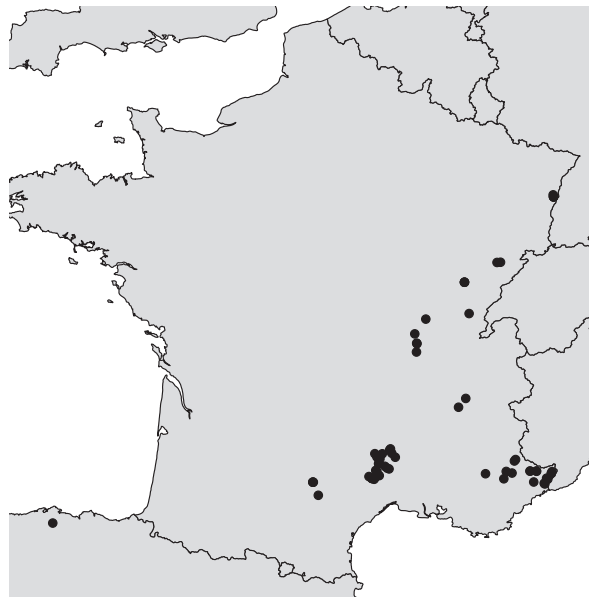


FIGURE 5.1 – Carte des 58 sites où des Asellidae ont été trouvés et où la radioactivité alpha globale a été mesurée.

Sur la base de l'identification, de la mesure de radioactivité, et de la phylogénie connue des Asellidae (Morvan et al., 2013). Des espèces ont été sélectionnées afin de former des paires d'espèces indépendantes, chaque paire contenant une

espèce vivant en milieu avec une radioactivité naturelle forte ( $>0.9$  Bq/g, voir Box 1 pour les unités de la radioactivité) et une espèce vivant dans un milieu avec une radioactivité naturelle faible ( $<0.7$  Bq/g), en imposant une différence de radioactivité de 3X entre les deux membres d'une paire. Pour les 14 espèces ainsi sélectionnées (Figure 5.2), une deuxième campagne d'échantillonnage a été effectuée. Pour chaque espèce, les individus ont été récoltés et congelés vivants en azote liquide, le plus rapidement possible après leur récolte. Pour chaque site, des sédiments fins ont été récoltés afin de mesurer tous les radionucléides permettant d'évaluer la dose reçue par les organismes.

**Box 1 : Les unités de mesure de la radioactivité**

Pour mesurer la radioactivité, trois unités de mesure sont utilisées :

- Le Becquerel (Bq) représente l'activité d'une matière radioactive, c'est à dire le nombre de désintégration radioactive qui se produit par seconde dans cette matière. 1 Bq correspond ainsi à une désintégration par seconde. On parle ainsi en Becquerel par quantité de matière (Bq/g ou Bq/Kg). Henry Becquerel (1852-1908) a découvert la radioactivité en 1896. Cette découverte lui a valu un prix Nobel avec Marie et Pierre Curie en 1903.
- Le Gray (Gy), mesure la quantité de rayonnements absorbés par un corps exposé à la radioactivité. C'est-à-dire que cette unité mesure la quantité d'énergie communiquée à un corps. Ainsi 1 Gy correspond à 1 Joule par kilogramme de matière irradiée. On utilise le Gray pour parler de la dose reçue par unité de temps (appelé débit-dose), on utilise ainsi généralement le Gray par heure Gy/h. 1 Gy représente une dose importante de radioactivité, dans la pratique, on utilise le milliGray (mGy) ou le microGray ( $\mu$ Gy). Louis Harold Gray (1905-1965) est considéré comme l'un des pères de la radiobiologie et de la médecine nucléaire. Il a défini cette unité de mesure permettant d'évaluer la dose de rayonnements absorbée par un corps à laquelle on donnera plus tard son nom.
- Le Sievert (Sv) permet de mesurer l'effet biologique d'une exposition humaine à des rayonnements. Cette mesure prend en compte l'effet du type de rayonnement (alpha, bêta ou gamma) et le tissu qui est exposé. On parle de dose équivalente biologique lorsque l'on mesure l'effet sur un organe et de dose Efficace lorsque l'on mesure l'effet sur un organisme entier. Le Sievert correspond à la même unité que le Gray (1 Joule/Kg), mais on applique des coefficients qui corrigent la dose reçue par l'effet biologique du rayonnement. Rolf Sievert (1896-1966) a contribué à la connaissance des effets des radiations sur la santé, mais a également mis au point le calcul de l'exposition en un point permettant de traiter les tumeurs par radiothérapie.

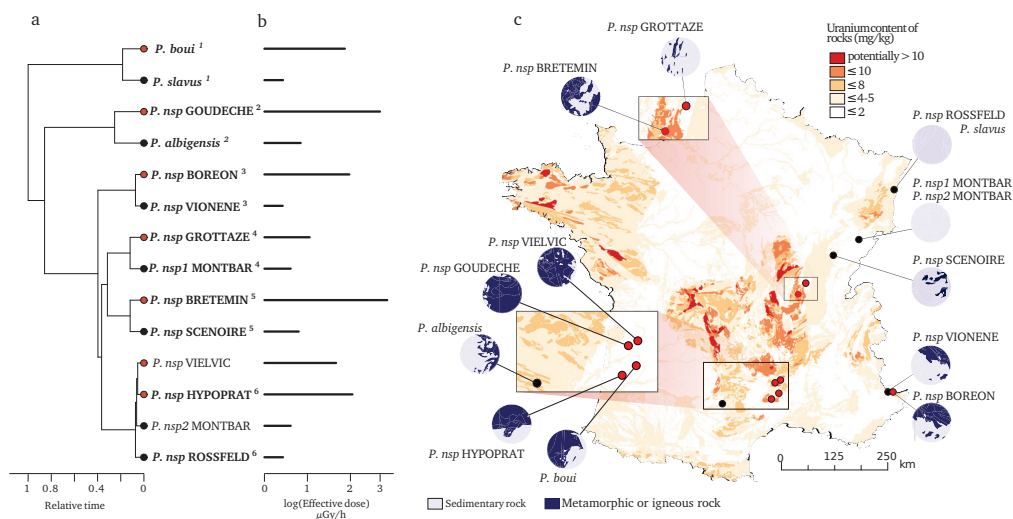


FIGURE 5.2 – Espèces et localités sélectionnées pour étudier l’impact de la radioactivité naturelle sur le taux et le spectre de mutations. 14 espèces avec des expositions contrastées à la radioactivité naturelle ont été sélectionnées (points noirs : faible exposition, points rouge : forte exposition). 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 (b). 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 (c).

## Mesures de la radioactivité

**Alpha radioactivity** Pour chacun des 58 sites échantillonnés, 3 échantillons d’environ 50g de sédiments ont été récoltés dans des bouteilles en plastique (polyéthylène). La mesure de la radioactivité alpha globale a été effectuée par le service LABRADOR de L’Institut de Physique Nucléaire de Lyon (France).

**Dose Reçue** Dans le but d’estimer la dose reçue par les organismes, 3 échantillons de 100g de sédiments fins ( $<100\text{m}$ ) représentant les sédiments dans lesquels les organismes se nourrissent ont été récoltés. Les radionucléides ont été mesurés au Centre d’Etude Nucléaire de Bordeaux. Ces mesures détaillées ont été entrées

dans l'outil ERICA V1.2.1, (Brown et al., 2016) qui extrapole les connaissances radiobiologiques connues sur une espèce modèle pour évaluer la dose reçue par les organismes. Ici, le modèle 'crustacé', basé sur des études faites sur la crevette, a été utilisé. La dose reçue dépendant du type d'exposition, nous avons assumé que les organismes passaient 10% de leur temps à la surface des sédiments et 90% à l'intérieur des sédiments, avec une exposition interne liée à l'ingestion de ces sédiments.

Ces mesures de radionucléides ont révélé une rupture dans l'équilibre séculaire de deux stations. A l'état naturel, tous les éléments des chaînes de désintégrations radioactives descendent les uns des autres et sont donc trouvés en équilibre dans le milieu, c'est-à-dire que la quantité de chaque élément dépend de la quantité de l'élément parent, ainsi que du temps de désintégration de cet élément parent. Cependant, lors d'activités minières, les éléments d'intérêt (uranium, plomb argentifère,...) sont extraits des roches et le reste (appelé stériles) est laissé sur place. Malgré leur nom ces stériles sont radioactifs, le minerai extrait ne représentant souvent qu'une petite partie de la radioactivité de la roche. Alors que la roche initiale étaient non seulement plus profonde mais aussi en un bloc, les stériles sont sous formes de résidus, laissant ainsi s'échapper plus facilement le radon qui se trouve sous forme de gaz. Ainsi, à proximité des mines d'extraction, même des années après l'exploitation, la radioactivité mesurée y est plus forte que la radioactivité naturellement présente et les éléments radioactifs ne sont pas dans les proportions naturelles. Ainsi, les activités mesurées dans ces deux stations ne correspondent pas à la radioactivité à laquelle sont exposées les espèces depuis des milliers de générations, l'exploitation minière ne datant que des années 1950. Le taux de substitutions étant mesuré sur la branche de l'arbre, donc sur des millions d'années, l'impact de cette radioactivité artificielle sur ce taux sera probablement nul. Les tests utilisant la radioactivité mesurée sur site ont donc été effectués sans ces points. La mesure de la radioactivité régionale permet cependant d'intégrer



ces points dans l'analyse.

### **Proportion de roches métamorphiques ou magmatiques dans un rayon**

**de 15km** Une mesure de la radioactivité au niveau du point d'échantillonnage pose deux problèmes : Premièrement, cette radioactivité représente la radioactivité dans laquelle vivent les individus échantillonnés et non toute l'espèce, alors que le dS estimé sur la branche de l'arbre représente celui de l'espèce. Deuxièmement, même si ces espèces dispersent peu, sur les millions d'années que représentent les branches terminales, elles ont pu bouger de quelques kilomètres. Ainsi, une estimation régionale de la radioactivité permettrait de mieux prendre en compte l'exposition réelle de l'espèce sur ce temps. A partir de la carte géologique de France (échelle : 1/1 000 000, ©BRGM), la proportion de roches magmatiques ou métamorphiques dans un rayon de 15km autour du point d'échantillonnage a été calculé (notée  $\lambda_{15}$ ), un cercle de 30km représentant l'aire de répartition moyenne d'une espèce souterraine (Eme et al., 2018).

### **Taux de substitutions**

Afin d'évaluer l'effet de la radioactivité naturelle sur les taux d'évolution moléculaire, ces taux ont été calculés indépendamment sur des gènes nucléaires et sur des gènes mitochondriaux. Pour le taux de substitutions nucléaires, un transcriptome par espèce a été séquencé en paired-end puis assemblé avec Trinity (voir Chap. 3, partie 2.2). Les familles de gènes orthologues ont été délimitées avec Blastp et Silix. 2490 gènes orthologues en copie unique partagés par les 14 espèces ont ainsi été identifiés. Ces 2490 gènes ont été alignés avec PRANK (Löytynoja et Goldman, 2008) avec un model de codons et les sites dont l'alignement était ambigu ont été retirés avec Glocks (Castresana, 2000). Les gènes ont ensuite été concaténés pour construire une phylogénie avec PhyML v3.0 (Guindon et al., 2010) sous un model GTR+G+I avec 100 réplicats de bootstrap. Les deux espèces de la lignée Slavus (*Proasellus bowi* et *Proasellus slavus*) ont été utilisées comme groupe

externe (voir Chap.3, partie 1.4). Un arbre par gène a également été construit avec les mêmes paramètres. La phylogénie de chaque gène a ensuite été comparée avec celle de la concaténation en utilisant ProfileNJ (Noutahi et al., 2016) avec un seuil de bootstrap à 90% (seuls les noeuds ayant un support de bootstrap supérieur à 90% peuvent être considérés comme incongruents entre les deux phylogénies), permettant de mettre en évidence que certains gènes avaient une phylogénie différente de celle de la concaténation (que nous considérons comme étant la 'vraie' phylogénie). Ces incongruences peuvent poser des problèmes dans l'estimation du  $d_S$ . En effet, le  $d_S$  est calculé sur une topologie définie. Si l'on impose un calcul de  $d_S$  sur des séquences en donnant une phylogénie qui ne correspond pas à l'histoire évolutive de ces séquences, cela revient à positionner les mutations sur les branches de l'arbre de façon à ce qu'elles correspondent à une histoire imposée qui n'est pas la leur. Cette incongruence aura tendance à allonger les branches terminales car les mutations qui ne peuvent pas être conciliées avec la topologie imposée seront placées sur ces branches, seul endroit où elles ne peuvent pas être en contradiction avec une mauvaise topologie. Afin d'éviter cela, tous les gènes soutenant des phylogénies contradictoires à la phylogénie obtenue avec la concaténation ont été retirés du calcul de  $d_S$ . Les gènes ayant des séquences trop courtes pour certaines espèces (<3/4 de l'alignement) ont également été retirés car un trop faible nombre de mutations biaise l'estimation du  $d_S$  (voir Chap. 2, partie 1). Le calcul de  $d_S$  nucléaire a ainsi été fait sur 769 gènes.

Pour les taux de substitutions mitochondriaux, les génomes mitochondriaux complets ont été assemblés et annotés (voir Chap. 3, partie 2.3), les 13 gènes codants pour des protéines ont ainsi été utilisés. Ces gènes ont été alignés et une phylogénie a été construite de la même façon que pour les gènes nucléaires. Les phylogénies mitochondriale et nucléaire obtenues présentent des topologies identiques.

Ces deux concaténations de gènes et les deux phylogénies correspondantes

ont été utilisées indépendamment pour calculer le  $d_S$  dans CoEvol (Lartillot et Poujol, 2011) avec un modèle de codons (Muse et Gaut, 1994), et une variation Brownienne du  $d_S$  et du  $d_N/d_S$  le long de l'arbre. Deux chaînes ont été lancées pour chaque concaténation, les convergences ont été vérifiées visuellement et en utilisant le programme tracecomp. Aucun point de calibration n'ayant été donné, la racine est fixée à 1 et les valeurs obtenues sont exprimées en  $d_S$  par unité de temps relative à la racine (noté  $d_S/ra$ ). Pour chaque branche terminale, la médiane des valeurs de  $d_S/ra$  après la convergence a été utilisée pour tester les corrélations. Les tests ont été effectués avec des PGLS pour prendre en compte l'inertie phylogénétique

### Spectre mutationnel

Afin d'évaluer l'effet de la radioactivité naturelle sur le spectre mutationnel, les transcriptomes de 7 individus par espèce ont été séquencés en single-end, permettant de calculer le spectre mutationnel sur 8 individus par espèces (1 transcriptome paired-end et 7 transcriptomes single-end). Pour calculer le spectre mutationnel de chaque population, nous avons utilisé les 2490 gènes orthologues en copie unique trouvés chez les 14 espèces étudiées. Nous avons premièrement cherché à reconstruire la séquence ancestrale de ces gènes pour chaque espèce. Puis les différences entre cette séquence ancestrale et les séquences des 8 individus ont été utilisées pour reconstruire le spectre mutationnel. Pour reconstruire la séquence ancestrale, nous avons utilisé une approche par paires d'espèces. Au sein des 14 espèces, nous avons pu former 6 paires d'espèces phylogénétiquement indépendantes. En effet, alors que la phylogénie basée sur le COI indiquait une paire *Proasellus nsp.* HYPOP-RAT et *Proasellus nsp.* MONTBAR contrastée en terme de radioactivité, et une autre paire *Proasellus nsp.* VIELVIC et *Proasellus nsp.* ROSSFELD, la phylogénie finale a révélé que les espèces *Proasellus nsp.* VIELVIC et *Proasellus nsp.* HYPOP-RAT, toutes deux prélevées en environnement radioactif, étaient en réalité monophylétiques, ne permettant donc de faire qu'une paire au lieu de deux.

Nous avons ainsi conservé la paire *Proasellus nsp.* HYPOPRAT-*Proasellus nsp.* ROSSFELD pour tester le spectre mutationnel. Sur ces 6 paires, nous avons utilisé une approche par parcimonie pour reconstruire les séquences ancestrales. Pour les individus des deux espèces formant une paire, si une base unique était partagée par les deux espèces, cette base a été considérée comme la base ancestrale. Par exemple, si pour une espèce donnée à une position donnée, on détecte les bases A et T au sein des 8 individus, et que dans l'espèce soeur on détecte les bases A et C, A est considéré comme ancestrale (si on avait détecté A et T également dans la deuxième espèce, la base ancestrale aurait été considérée comme inconnue). Pour chaque espèce, nous avons ensuite calculé la fréquence de chaque type de mutations en fonction de la fréquence en base ancestrale. Cette fréquence dépendant du taux de mutations, la proportion de chaque type de mutation a été calculée. Les mutations complémentaires ont ensuite été additionnées afin d'augmenter la puissance statistique, le nombre de mutations étant assez faible. La relation entre la proportion de chaque type de mutation et la radioactivité a ensuite été testée avec des PGLS afin de prendre en compte l'inertie phylogénétique.

### 3 Résultats

#### Taux de substitutions

Sous l'hypothèse qu'une exposition chronique à la radioactivité naturelle peut engendrer une augmentation du taux de mutations germinales, on s'attend à trouver une augmentation du taux de substitutions synonymes chez les espèces vivant en milieu plus fortement radioactif. Cette augmentation n'a pas de raison d'être plus marquée dans l'un ou l'autre des compartiments génomiques. On trouve bien une corrélation positive entre le taux de substitutions synonymes et la radioactivité, quel que soit le compartiment génomique et quelle que soit la mesure de radioactivité utilisée (radioactivité  $\alpha$ , dose reçue ou  $\lambda_{15}$ ), avec des statistiques

similaires pour les deux compartiments, à l'exception de la pente qui est toujours plus forte avec le  $d_S$  mitochondrial (Table 5.1, Figure 5.3). Les modèles prédisent une augmentation du  $d_S$  nucléaire d'environ 30% chez les espèces vivant dans des environnements plus fortement radioactifs en comparaison à des espèces vivant dans des environnements très faiblement radioactifs. Cette augmentation est de 60% dans le compartiment mitochondrial.

TABLE 5.1 – Résultats des tests des moindres carrés généralisés phylogénétiques (PGLS) entre le taux de substitutions synonymes ( $d_S$ ) nucléaire et mitochondriaux et la radioactivité  $\alpha$  mesurée dans les sédiments (en Bq/g), la dose reçue (RD, en  $\mu\text{Gy/h}$ ) modélisée avec l'outil ERICA, ou la proportion de roches métamorphiques ou magmatiques dans un rayon de 15 km ( $\lambda_{15}$ ). La radioactivité  $\alpha$  et la dose reçue ont été transformées en log afin de correspondre aux hypothèses des modèles linéaire (distribution normale). Chaque ligne correspond à un test LRT entre un modèle nul et un modèle prenant en compte la variable. Les  $R^2$  sont des pseudo  $R^2$  de Cox-Snell.

	Nuclear dS				Mitochondrial dS				N taxa
	Slope	L. Ratio	p. value	$R^2$	Slope	L. Ratio	p. value	$R^2$	
log( $\alpha$ radio.)	0.034	5.995	<b>0.014</b>	0.393	0.506	7.895	<b>0.005</b>	0.482	12
log(RD)	0.038	6.51	<b>0.011</b>	0.419	0.491	5.981	<b>0.015</b>	0.392	12
$\lambda_{15}$	0.076	9.039	<b>0.003</b>	0.476	1.097	11.680	<b>0.001</b>	0.566	14

## Patron mutationnel

On observe une différence de spectre mutationnel, corrélée au niveau de radioactivité. En effet, on observe une corrélation positive forte entre les mutations C:G→A:T et le niveau de radioactivité quelle que soit la mesure de radioactivité utilisée (Table 5.2 et Figure 5.4). En revanche, deux types de mutations baissent avec la radioactivité : les mutations A:T→T:A et A:T→G:C.

## 4 Discussion

### Taux de substitutions

On observe une augmentation du taux de substitutions dans les deux compartiments génomiques dans les environnements naturellement plus radioactifs avec

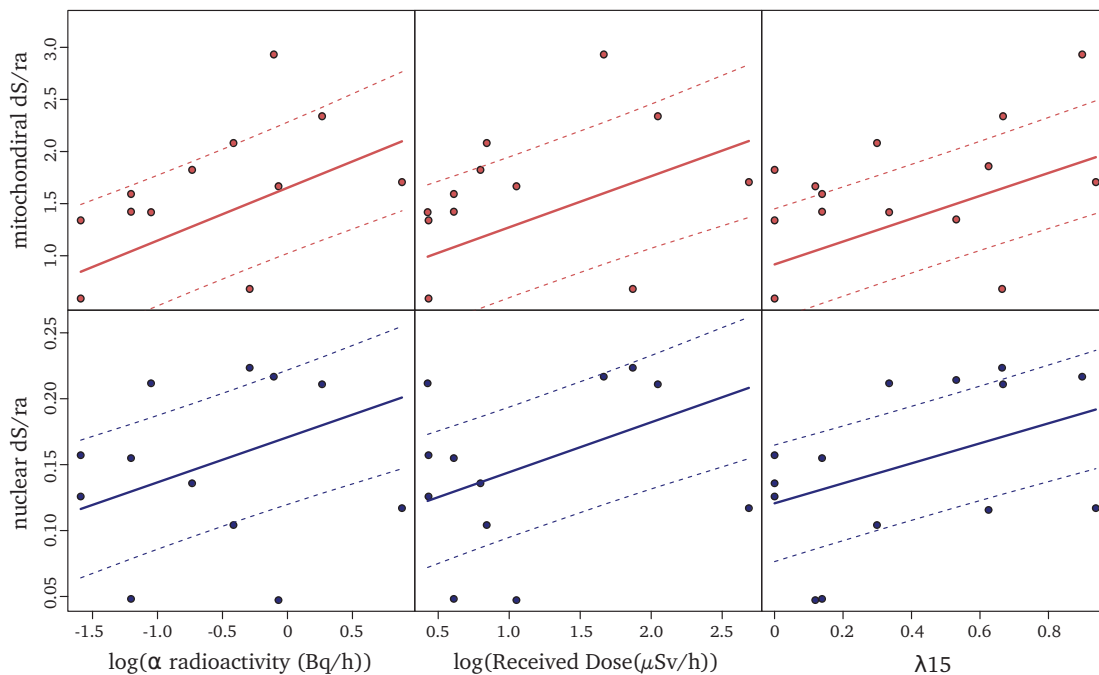


FIGURE 5.3 – Corrélations entre les taux de substitutions synonymes relatifs à l'âge de la racine ( $d_S/ra$ ) et la radioactivité mesurée, soit par la radioactivité  $\alpha$  dans les sédiments (gauche), soit par la dose reçue (milieu) ou par la proportion des roches magmatiques et ignées dans un rayon de 15 km autour du point d'échantillonnage,  $\lambda_{15}$  (droite). Chaque point représente une espèce. Les  $d_S$  mesurés sur les gènes mitochondriaux sont représentés par des points rouges et les  $d_S$  mesurés sur des gènes nucléaires par des points bleus. Les prédictions du modèle (PGLS) sont indiqués avec une ligne pleine et l'intervalle de confiance de ce modèle par des lignes pointillées.

un effet plus marqué dans le compartiment mitochondrial, suggérant un effet du stress oxydant. La différence entre les deux compartiments peut s'expliquer si les environnements radioactifs augmentent l'activité de la chaîne respiratoire mitochondriale ou réduisent son efficacité, générant ainsi plus de stress oxydant dans ce compartiment. En revanche, la radiolyse de l'eau, tout comme les dommages directs des rayonnements sur l'ADN ne sont pas ciblés dans un compartiment. Le génome mitochondrial ayant des capacités de réparation plus faible que le génome nucléaire (Mandavilli et al., 2002), ce dernier pourrait être plus sensible à un impact mutagène. Ainsi, ces résultats suggèrent que l'effet mutagène de la radioactivité serait lié au stress oxydant qu'elle génère. Des tests de comparaison de l'activité et de l'efficacité de la chaîne respiratoire mitochondriale en présence

TABLE 5.2 – Résultats des tests des moindres carrés généralisés phylogénétiques (PGLS) entre la proportion relative de chaque type de mutation et la radioactivité  $\alpha$  mesurée dans les sédiments (en Bq/g), la dose reçue (RD, en  $\mu\text{Gy/h}$ ) modélisée avec l'outil ERICA, ou la proportion de roches métamorphiques ou magmatiques dans un rayon de 15 km ( $\lambda_{15}$ ). La radioactivité  $\alpha$  et la dose reçue ont été transformées en log afin de correspondre aux hypothèses des modèles linéaire (distribution normale). Chaque ligne correspond à un test LRT entre un modèle nul et un modèle prenant en compte la variable. Les  $R^2$  sont des pseudo  $R^2$  de Cox-Snell.

Variable dépendante	Variable explicative	Pente	L.Ratio	P.value	$R^2$	N
P(A:T→T:A)	log( $\alpha$ radio)	-0.009	6.8187	0.009	0.433	12
	log(RD)	-0.011	9.3479	0.002	0.541	14
	$\lambda_{15}$	-0.025	6.994	0.008	0.442	14
P(A:T→C:G)	log( $\alpha$ radio)	0.001	0.121	0.728	0.010	12
	log(RD)	0.000	0.015	0.904	0.001	14
	$\lambda_{15}$	0.001	0.005	0.945	0.000	14
P(A:T→G:C)	log( $\alpha$ radio)	-0.022	7.778	0.005	0.477	12
	log(RD)	-0.019	4.004	0.045	0.284	14
	$\lambda_{15}$	-0.014	0.394	0.530	0.032	14
P(C:G→G:C)	log( $\alpha$ radio)	0.005	2.137	0.144	0.163	12
	log(RD)	0.006	2.403	0.121	0.181	14
	$\lambda_{15}$	0.006	0.360	0.549	0.030	14
P(C:G→A:T)	log( $\alpha$ radio)	0.013	13.010	0.000	0.662	12
	log(RD)	0.014	12.079	0.001	0.635	14
	$\lambda_{15}$	0.042	13.791	0.000	0.683	14
P(C:G→T:A)	log( $\alpha$ radio)	0.012	3.323	0.068	0.242	12
	log(RD)	0.010	1.684	0.194	0.131	14
	$\lambda_{15}$	-0.009	0.183	0.669	0.015	14

ou en absence de faibles doses de radioactivité permettraient de tester cette hypothèse et d'évaluer plus finement l'impact de ces doses sur le fonctionnement cellulaire.

### Spectre mutationnel

L'analyse du spectre mutationnel montre une corrélation entre la proportion de certains types de mutations et le niveau de radioactivité. On note une augmentation des mutations C:G→A:T chez les espèces vivant en milieu plus fortement radioactif. Ces mutations sont connues pour être le résultat d'un stress oxydant (Shibutani et al., 1991). La relation entre une exposition à la radioactivité et une augmentation du stress oxydant a déjà été montrée lors d'exposition *in vitro* (Loseva et al., 2014). Nos résultats suggèrent que cette augmentation du stress oxydant a lieu également lors d'une exposition chronique à de faibles doses. Les mutations liées au stress oxydant sont généralement évitées ou réparées via l'ac-

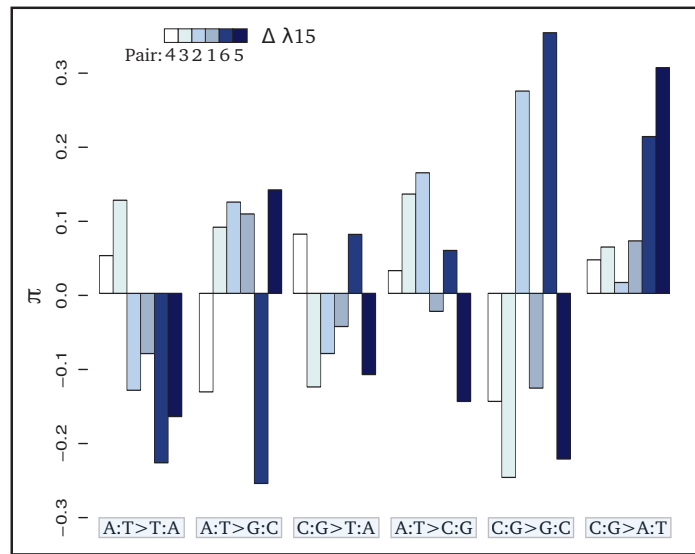


FIGURE 5.4 – Contrastes ( $\pi$ ) de la proportion relative de chaque type de mutation  $[p(i:j \rightarrow k:l)]$  dans chaque paires d'espèces soeurs :  $\pi_m = \log \frac{p(i:j \rightarrow k:l)^+}{p(i:j \rightarrow k:l)^-}$  où + et - correspondent respectivement à l'espèce exposée à la plus forte et à la plus faible radioactivité dans la paire  $m$ . Les barres positives représentent donc une plus forte proportion d'une mutation données 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 ( $\lambda_{15}$ ) à celle ayant la plus forte différence de  $\lambda_{15}$  ( $\Delta$ ). Les mutations sont ordonnées de celle étant la plus négativement corrélée à  $\lambda_{15}$  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 Figure 5.2.

tivation de systèmes cellulaires (enzymes antioxydantes, systèmes de réparation). Cependant, plusieurs études ont montré que ces systèmes ne sont pas activés lorsque le niveau des dommages est très faible (Inoue et al., 2004; Grudzenski et al., 2010; Kim et al., 2012). Il est ainsi possible que le niveau de stress oxydant augmente en milieu naturellement radioactif, augmentant le nombre de mutations, mais sans activation des systèmes de réparation.

### Impact relatif de l'environnement mutagène sur le taux de substitutions

Nous montrons ici qu'une différence d'exposition à la radioactivité d'un facteur 5 environ est capable de générer une augmentation du taux de substitutions



d'environ 30% dans le génome nucléaire et de 60% dans le génome mitochondrial. Pour le génome nucléaire cette augmentation est légèrement supérieure à ce qui est observé lors de la transition du milieu de surface vers le milieu souterrain (Chap. 4). Pour le génome mitochondrial, les traits d'histoire de vie n'étant pas corrélés au taux de substitutions de ce compartiment chez ces espèces, on trouve donc ici un impact très important. Ainsi, il semble que l'effet de l'environnement mutagène ait été largement sous estimé dans l'étude des taux d'évolution moléculaire. Cependant, il faut noter que les Asellidae se nourrissent en ingérant le sédiment. Ainsi, l'exposition à la radioactivité est interne. Une exposition interne étant beaucoup plus mutagène qu'une exposition externe, il serait important de tester l'effet de la radioactivité sur des organismes ayant un régime alimentaire différent.

# Chapitre 6

## Existe-t-il des patrons de variations des taux d'évolution moléculaire à l'échelle globale ?

### 1 Introduction

La mutation est la source de la variabilité génétique sur laquelle les processus évolutifs agissent. Lors d'un événement de spéciation, si les échanges génétiques entre les deux populations deviennent très rares ou nuls, les mutations vont alors s'accumuler aléatoirement dans les deux populations. La sélection devrait éliminer les allèles délétères au cours du temps, mais si une mutation a un effet neutre en cas de reproduction au sein de la même population mais délétère en cas de reproduction avec l'autre population, cette mutation ne sera pas contre-sélectionnée car la reproduction entre les deux populations se produit rarement. L'accumulation de ces incompatibilités hybrides devraient à terme empêcher toute reproduction entre les deux populations. Le modèle de Dobzhansky-Muller (Dobzhansky, 1936; Muller, 1942) prévoit ainsi que, par ce mécanisme, deux populations vont finir par donner deux espèces. Si l'on considère que chaque mutation ponctuelle a une probabilité non nulle de générer une incompatibilité, il n'y a alors qu'un pas à

faire pour lier le taux de mutations d'une espèce à sa diversification. Ainsi, un raisonnement intuitif permet de supposer qu'un groupe ayant un taux de mutations élevé contiendra plus d'espèces, soit parce qu'un isolement géographique amènera plus rapidement à une accumulation des incompatibilités hybrides, soit parce que la probabilité de générer un variant permettant de survivre dans l'environnement sera plus grande, réduisant ainsi la probabilité d'extinction. Il semble ainsi logique de pouvoir lier taux de diversification (taux de spéciation-taux d'extinction) à taux d'évolution moléculaire. C'est sur ce raisonnement que Rohde formule, en 1992, l'Evolutionary Speed Hypothesis (ESH) qui propose que le gradient latitudinal de diversité (LDG, Box 1) s'explique par des variations latitudinales de taux de mutations, elles mêmes engendrées par des facteurs environnementaux variant latitudinalement.

Mais si ce raisonnement semble intuitif, il omet les nombreux autres facteurs impliqués dans la spéciation. En effet, cette hypothèse implique premièrement que les mutations ponctuelles jouent un rôle prépondérant dans l'isolement génétique des populations. Hors, des réarrangements génomiques (translocations, duplications) ont un impact majeur dans le processus de spéciation (Brennan et al., 2019; Fuller et al., 2018). D'autre part, le modèle de Dobzhansky-Muller permet de faire l'hypothèse qu'une espèce ayant un taux de mutations élevé donnera plus rapidement des populations isolées génétiquement en cas d'isolement géographique, mais cela implique une première rupture géographique qui ne dépend pas du taux de mutations. Enfin, les hypothèses expliquant les patrons de diversité sont extrêmement nombreuses et font appel à des concepts de quantités de niches, de ressources disponibles ou encore d'interactions biotiques. Toutes ces hypothèses et ces mécanismes étant mutuellement non exclusifs, on peut s'interroger sur le rôle relatif du taux de mutations, s'il existe, parmi tous ces facteurs.

L'ESH est restée longtemps dans la longue liste des hypothèses expliquant le LDG sans être testée, du fait de la difficulté d'obtenir des données moléculaires sur

un gradient latitudinal suffisant. Dès que ces données ont été disponibles, plusieurs auteurs se sont attachés à tester l'ESH, montrant ainsi que les taux de substitutions sont plus élevés dans les tropiques pour une grande diversité d'organismes allant des mammifères (Gillman et al., 2009; Weir et Schluter, 2011) aux plantes (Davies et al., 2004; Wright et al., 2006; Gillman et al., 2010; Lancaster, 2010) en passant par les foraminifères (Allen et al., 2006). Ces résultats ont poussé Gillman et Wright (2013) à proposer que tous les patrons de répartition de la biodiversité étaient régis par les taux d'évolution moléculaire, expliquant ainsi non seulement le LDG mais aussi les patrons de diversité altitudinaux ou de profondeur dans les océans. Cependant, la relation entre taux de substitutions et latitude semble ne pas être universelle. En effet, des méta-analyses à large échelle taxinomique (Webster et al., 2003; Orton et al., 2019) semblent indiquer que ce patron existe dans seulement 30 à 50% des clades. En revanche, le grand nombre de démonstrations chez les plantes semble indiquer que cette corrélation est robuste pour ce groupe.

S'il apparaît ainsi que les taux de substitutions sont plus élevés dans les tropiques pour un grand nombre de taxa, les études ayant testé l'ESH souffrent de plusieurs lacunes. Premièrement, les études sont, pour la plupart, basées sur une approche en paires d'espèces (voir Chap. 2, partie 2) entre des espèces tropicales et des espèces à latitudes plus élevées. La plupart du temps, il n'y a pas de test pour évaluer la corrélation entre la valeur du contraste de taux de substitutions et la valeur du contraste de latitude. Ainsi, Weir et Schluter (2011) ont montré que si les espèces tropicales ont bien un taux de substitutions plus élevé que les espèces à des latitudes plus élevées, la différence de taux de substitutions n'est pas pour autant corrélée à la différence de latitude, suggérant un effet non linéaire. Deuxièmement, l'ESH a été principalement testée en comparant les taux de substitutions d'espèces vivant à des latitudes différentes. Or, la latitude n'est qu'un proxy de facteurs environnementaux et ne peut pas influencer les taux de

substitutions. Ainsi, les relations gagneraient sans doute en puissance en utilisant les valeurs de température ou de rayonnements UV dans les tests plutôt que la latitude. Troisièmement, la plupart des études ont mesuré les taux d'évolution de façon globale, sans distinguer les taux synonymes et non-synonymes. L'ESH propose que l'environnement impacte les taux de mutations. Ainsi, pour tester cette hypothèse il est nécessaire de regarder les substitutions synonymes qui sont, en l'absence d'autres biais, un bon estimateur du taux de mutations. Un taux global ne permet pas de différencier l'impact de l'environnement de celui de la taille de population ou de la sélection naturelle. Enfin, les taux de substitutions sont calculés, la plupart du temps sur quelques gènes, souvent mitochondriaux. Or on sait que les taux de substitutions sont impactés par des facteurs différents dans les deux compartiments génomiques et un calcul de taux de substitutions sur quelques gènes ne reflète pas le taux de substitutions du génome. Ainsi, ces lacunes poussent à se demander s'il existe réellement une corrélation entre les taux de mutations et les facteurs environnementaux corrélés aux patrons de diversité.

**Box 1 : Le gradient latitudinal de diversité** Le gradient latitudinal de diversité (LDG) décrit la diminution de la richesse spécifique en allant de l'équateur vers les pôles. Le LDG est l'un des patrons les plus frappants qui soit, notamment parce qu'il est retrouvé chez toutes les espèces qu'elles soient ectothermes ou endothermes, marines ou terrestres. Cependant, si ce patron est très bien décrit, il n'en reste pas moins mal expliqué. De façon ultime le nombre d'espèces en un lieu dépend de 3 forces : la spéciation, l'extinction et la dispersion. Une centaine d'hypothèses a été proposée pour expliquer cette répartition de la biodiversité sur terre. Toutes ces hypothèses se différencient par le poids qu'elles donnent à ces trois forces et par les facteurs qui vont les impacter. Mittelbach et al. (2007) classent ces hypothèses en 3 catégories : (i) les hypothèses écologiques qui expliquent la richesse par les facteurs environnementaux permettant la coexistence ou le maintien des espèces dans le milieu, (ii) les hypothèses évolutives qui expliquent ce patron par des différences latitudinales de diversification et (iii) les hypothèses historiques qui font appel à des différences latitudinales de stabilité climatique à long terme. Ainsi, parmi les nombreux facteurs proposés comme étant à l'origine de ce gradient on peut citer : la température, la productivité, la variabilité saisonnière, la quantité de niches, les interactions biotiques, la dispersion ou encore les tailles de populations. L'Evolutionary Speed Hypothesis (Rohde, 1992), qui propose que les variations de diversification proviennent de variations de taux de mutations, elles mêmes engendrées par des facteurs environnementaux variant latitudinalement, est donc une hypothèse parmi beaucoup d'autres. Cependant c'est une hypothèse un peu particulière car c'est la seule à proposer un mécanisme moléculaire à l'origine du LDG.

Afin d'évaluer la relation entre les variables environnementales variant latitudinalement et les taux de substitutions, nous avons utilisé les propriétés particulières des Asellidae souterrains. Cette famille d'isopodes aquatiques présente, en effet, plusieurs particularités qui permettent de lever certaines difficultés inhérentes à l'étude des gradients environnementaux. Premièrement, ces organismes ne sont pas soumis aux UV, permettant d'éliminer un facteur dans l'analyse. Deuxièmement, la température en milieu souterrain est très stable et correspond à la température moyenne annuelle en surface, ainsi cette température moyenne annuelle est une valeur qui représente bien les conditions de vie des organismes et permet de retirer l'effet de la variabilité saisonnière qui varie avec la latitude (Eme et al., 2014). Troisièmement, ces organismes ayant des capacités de dispersion très faibles (Chap.3, partie 1.4, Eme et al., 2018), on peut utiliser les coordonnées géographiques actuelles pour reconstruire leur environnement passé. Pour certains paramètres tels que la latitude, on peut même supposer que les conditions environnementales actuelles représentent relativement bien les conditions auxquelles les organismes sont soumis depuis de nombreuses générations.

Afin d'évaluer la relation entre les facteurs environnementaux variants latitudinalement et les taux de mutations, deux jeux de données ont été constitués : un set de 367 espèces d'Aselloidae pour lesquelles les taux de substitutions ont été calculés sur deux gènes nucléaires (28S et FAST2) et deux gènes mitochondriaux (16S et COI) et un set de 64 espèces pour lesquelles les transcriptomes ont été séquencés et les génomes mitochondriaux assemblés, nous permettant de calculer les taux de substitutions sur 346 gènes nucléaires et 13 gènes mitochondriaux. Pour chaque espèce, les facteurs pouvant varier latitudinalement (température, productivité, aire de distribution, impact des glaciations) ont été extraits afin d'évaluer leur corrélation avec les taux de substitutions.

Le temps depuis lequel l'espèce est devenue souterraine a une influence sur l'estimation des taux de substitutions (voir Chapitre 3). Afin d'intégrer cet effet, la

non-fonctionnalisation du gène de l'opsine a été utilisée pour dater les événements de colonisation pour 89 espèces.

## 2 Matériel et Méthodes

### Données moléculaires

Afin de tester l'impact de facteurs environnementaux variants latitudinalement, deux jeux de données ont été construits. Le premier jeu de données, que nous appellerons `data367sp` par la suite, reprend les données utilisées pour construire la phylogénie calibrée dans le temps (Chap. 3, partie 1.4), nous permettant d'évaluer la relation sur une concaténation de deux gènes mitochondriaux (`data367sp_mito`) et sur une concaténation de deux gènes nucléaires (`data367sp_nucl`) pour 367 taxa répartis sur l'Europe et, dans une moindre mesure, les Etats-Unis. Chaque concaténation contenant un gène ribosomal, le calcul du  $d_S$  est impossible, ainsi le  $d_S$  a été calculé sur le gène COI (`data367_COI`) et sur le gène FAST2 (`data367sp_FAST2`).

Parmi ces 367 espèces, 64 espèces réparties sur un gradient de latitude ont été sélectionnées pour former le jeu de données `data64sp`. Pour ces 64 espèces, les transcriptomes ont été séquencés et les génomes mitochondriaux assemblés (voir Chap. 3, partie 2.3, la délimitation des familles de gènes orthologues a permis d'obtenir 346 gènes nucléaires (`data64sp_nucl`) pour le calcul des taux de substitutions synonymes. Pour le génome nucléaire, les 13 gènes mitochondriaux codant pour des protéines ont été utilisés pour le calcul du taux de substitutions synonymes (`data64sp_mito`).

### Délimitation des familles de gènes orthologues du `data64sp`

Afin d'obtenir les familles de gènes orthologues en copie unique (c'est-à-dire sans paralogues) présents chez les 64 espèces d'Aselloidea, ainsi que chez l'espèce *Jaera praehirsuta* (isopode de la famille des Janiroidea, groupe frère des Aselloi-

dea) qui est utilisée comme outgroup afin de vérifier la monophylie des Asellidea remise en cause par Lins et al. (2012), un pipeline bio-informatique optimisé a été mis en place. Les transcriptomes ont été séquencés (voir Chap. 3, partie 2.2), les lectures ont été nettoyées avec Trimmomatic (Bolger et al., 2014), normalisée (les lectures identiques observées plus de 50 fois sont retirées) puis assemblés avec Trinity (Grabherr et al., 2011). Lorsque la qualité de l'assemblage était trop faible, des données de séquençage single-end ont été ajoutées. Les cadres ouverts de lectures (ORF, Open Reading Frame) corrigées pour les décalages de cadre de lecture (frameshift) ont été prédits en utilisant un programme personnalisé appelé assembly2orf (<https://github.com/ellefeg/assembly2orf>). A chaque contig a été assigné la séquence ayant le meilleur hit parmi une base de données de séquences protéiques de métazoaires (comprenant les séquences d'Asellidae du laboratoire et la base de données Ensembl Metazoa) avec Diamond BLASTx v0.9.22 (Buchfink et al., 2015) : les contigs sans hit de blast ont été conservés. Les frameshift dans les contigs par rapport à la séquence de référence assignée, ont été retirés avec Exonerate v2.2.0 (Slater et Birney, 2005), nous permettant d'obtenir des séquences sans frameshift et limitées à la partie alignée avec la référence. Les ORFs ont ensuite été prédits en utilisant TransDecoder (<http://transdecoder.sourceforge.net>, Haas et al., 2013). Pendant le processus de TransDecoder, l'homologie entre les ORFs candidats et la base de données mentionnée précédemment, a été testée avec un BLASTx. L'homologie des domaines protéiques a été testée avec HMMER 3.1b2 (<http://hmmerr.org/> Potter et al., 2018). Les séquences annotées ont été retenues par TransDecoder en fonction de leur score de prédiction de région codante.

Les familles de gènes orthologues ont été définies en utilisant OrthoFinder2 (Emms et Kelly, 2018, , paramètres : -S diamond -M msa). La redondance dans les isoformes a ensuite été retirée en conservant l'isoforme le plus exprimé. La contamination a été retirée en utilisant Diamond BLASTx sur une base de données contenant 937 taxons (100 Eucaryotes et 837 Procaryotes, pers. comm. de



C.M. Francois). Les séquences dont l'ensemble des meilleurs hits correspondaient à des séquences non-arthropodes ont été retirées. Les familles de gènes orthologues contenant une séquence pour chaque espèce ont ensuite été extraites, nous permettant d'obtenir 348 gènes orthologues.

Les 348 gènes orthologues ont été alignés en protéines avec PRANK (Löytynoja et Goldman, 2008), repassés en nucléotides, puis vérifiés et corrigés si nécessaire *de visu*. Les positions trop variables ont été retirées avec Gblocks (Castresana, 2000). Les séquences ont été concaténées puis une phylogénie concaténée a été construite avec PhyML (GTR+G+I). En parallèle, une phylogénie a été construite pour chaque gène avec les mêmes paramètres et 100 répliquats de bootstrap. La congruence entre la phylogénie concaténée et la phylogénie donnée pour chaque gène a été vérifiée avec ProfileNJ (Noutahi et al., 2016) avec un seuil de bootstrap de 90% (les supports de branches inférieurs à cette valeur ne sont pas considérés comme incongruents avec la phylogénie concaténée). Cette étape permet de retirer des gènes dont l'histoire évolutive ne correspondrait pas à l'histoire des espèces, mais aussi de détecter des cas de contaminations par d'autres espèces ou entre les espèces échantillonnées. Sept gènes montraient des incongruences notables, deux gènes montrant des évidences de paralogie ont été retirés. Pour les 5 autres gènes, à chaque fois une seule séquence montrait des anomalies. Dans deux cas, une des séquences était une probable chimère, la séquence a alors été tronquée pour conserver la partie s'alignant avec les autres séquences. Dans les 3 autres cas, une espèce se branchait à la base de l'arbre et montrait une branche extrêmement longue, suggérant une contamination, la séquence a donc été supprimée de l'alignement.

### **Enracinement de la phylogénie**

Une nouvelle concaténation a été effectuée avec les 346 familles de gènes orthologues nettoyées. Une nouvelle phylogénie a été reconstruite avec PhyML avec un modèle GTR+G+I avec 100 répliquats de bootstrap. En parallèle les gènes mitochondriaux ont été alignés avec Prank (Löytynoja et Goldman, 2008) et vérifiés

*de visu*, les positions trop variables ont été retirées avec Gblocks (Castresana, 2000). Les gènes ont ensuite été concaténés et une phylogénie a été construite de la même façon que pour le nucléaire. Ces deux arbres ont ensuite été enracinés en utilisant *Jaera praeheirsuta* (isopode de la super famille des Janiroidea, groupe frère des Aselloidea) comme outgroup. Ces deux arbres confirment que les Asellidae forment un groupe monophylétique, groupe frère de la famille des Stenasellidae (Score de bootstrap à 100% pour les deux compartiments génomiques). L'espèce *Jaera praeheirsuta* montrant des divergences génétiques assez fortes avec les autres espèces, cette espèce a été retirée pour les calculs des taux de substitutions. Les deux espèces de Stenasellidae ont été utilisées par la suite comme groupe externe.

### Calcul des taux de substitutions

**data367sp** La phylogénie calibrée dans le temps présentée en partie 1.4 du chapitre 3 a été utilisée dans CoEvol (Lartillot et Poujol, 2011). Les taux de substitutions ont été calculés sur 4 sets de gènes différents : (i) une concaténation des deux gènes mitochondriaux, (ii) une concaténation des deux gènes nucléaires, (iii) le gène COI seul, et (iv) le gène FAST2 seul. Pour les deux premiers sets, le taux global varie de façon brownienne, pour les deux derniers, un modèle de codon (Muse et Gaut, 1994) est utilisé dans lequel le  $d_N$  et le  $d_S$  varie de façon Brownienne. Pour chaque set, deux chaînes ont été lancées, les convergences ont été vérifiées visuellement et en utilisant le programme tracecomp implémenté dans CoEvol. Pour le gène COI, un  $d_S$  sur les transversions a été calculé avec mapNH, de la même façon que pour la concaténation de gènes mitochondriaux.

**data64sp** L'espèce *Jaera praeheirsuta* a été retirée des deux concaténations (mitochondriale et nucléaire). Un arbre nucléaire et un arbre mitochondrial des 64 espèces ont ensuite été construits avec PhyML (GTR+G+I). Ces phylogénies et les deux concaténations ont ensuite été utilisées pour le calcul des taux de substitutions dans CoEvol (Lartillot et Poujol, 2011) avec un modèle de codon (Muse

et Gaut, 1994), et une variation Brownienne du  $d_S$  et du  $d_N/d_S$  le long de l'arbre. Deux chaînes ont été lancées pour chaque concaténation, les convergences ont été vérifiées visuellement et en utilisant le programme tracecomp. Aucun point de calibration n'ayant été donné, la racine est fixée à 1 et les valeurs obtenues sont exprimées en  $d_S$  par unité de temps relative à la racine (noté  $d_S/ra$ ).

**Fiabilité des estimations** Comme expliqué dans le chapitre 2, l'estimation des taux de substitutions peut s'avérer peu fiable lorsque le nombre d'événements de substitutions sur lequel le calcul est effectué est trop faible. Ce nombre d'événements dépend du nombre de positions sur lequel les taux de substitutions sont estimés, du temps de divergence entre les espèces, et de la vitesse d'évolution des gènes utilisés. Ainsi, les estimations effectuées sur les courtes branches doivent être évitées car elles sont généralement peu fiables et très variables. En effet, la valeur minimale étant bornée à zéro alors qu'il n'y a aucune borne supérieure, si la distribution de probabilité est estimée à partir de trop peu de données, elle est alors biaisée vers les valeurs fortes. Ainsi, Debruyne et Poinar (2009) montrent que le "time-dependency" des taux de substitutions, c'est-à-dire le fait que les taux de substitutions sont plus rapides dans les branches courtes, pourrait être créé par cet artefact. Le test diagnostique proposé par Welch et Waxman (2008) pour les contrastes, permet de vérifier cet artefact. Cette procédure consiste à tester une corrélation négative entre le taux mesuré et le temps que représente la branche, puis à définir une longueur de branche minimale au delà de laquelle cette corrélation disparaît. Les branches plus courtes que cette valeur seuil sont alors retirées de l'analyse. Welch et Waxman (2008) montrent par des simulations que cette procédure, même en retirant la moitié des points, permet de gagner en puissance.

Si les branches courtes sont un problème, les branches longues le sont également, car elles présentent le problème de la saturation. Dans ce cas on devrait observer un taux de substitutions plus faible dans les longues branches. La même

approche que pour les branches courtes, qui consiste à évaluer la relation entre le taux de substitutions et le temps a été utilisée pour retirer les branches saturées.

### Calcul des variables environnementales

Afin d'évaluer la corrélation entre les taux de substitutions et les variables environnementales corrélées à la latitude, 3 variables ont été estimées pour chaque localisation échantillonnée (1731 individus) : (i) l'évapotranspiration actuelle (AET) qui est une estimation de la quantité d'eau (en mm) qui repart dans l'atmosphère. Plus la couverture végétale est forte, plus cette valeur est élevée, ce qui fait de l'AET un bon estimateur de la productivité du milieu, (ii) la température, (iii) la différence entre le température au dernier maximum glaciaire (LGM) et la température actuelle. En effet, [Eme et al. \(2014\)](#) ont montré que la répartition de la biodiversité du groupe était principalement expliquée par ces épisodes glaciaires qui ont eu pour effet des extinctions massives des régions nordiques qui ont ensuite été recolonisées par les rares espèces souterraines ayant des capacités de dispersion importantes. Ainsi, dans le cadre de l'Evolutionary Speed Hypothesis, il semble important de tester l'effet d'un facteur reconnu comme un acteur majeur de la distribution de la diversité.

Pour chaque station dans laquelle un individu utilisé dans la délimitation moléculaire des espèces (Chap. 3, partie 1.3) était présent, les données d'AET ont été extraites à partir d'une carte mondiale ayant une résolution de 30 secondes d'arc publiée par [Trabucco et Zomer \(2019\)](#). Les valeurs de température ont été extraites de la base de données WorldClim ([http : //www.worldclim.org](http://www.worldclim.org)) avec une résolution spatiale de 2.5 minutes d'arc. Les valeurs de température au LGM ont été extraites également de la base de données WorldClim en utilisant les 3 modèles climatiques disponibles avec une résolution spatiale de 30 secondes d'arc : CCSM4, MIROC-ESM et MPI-ESM. La moyenne de ces trois modèles a ensuite été utilisée.

Pour chaque espèce, la moyenne des stations contenant au moins un individu

appartenant à cette espèce a été calculée pour chaque variable. Certains points étant très proches de la mer, la résolution ne permet pas d'avoir une mesure sur une zone ne contenant pas de zone marine. Dans ce cas, le calcul de l'AET est impossible. Pour une espèce, la donnée est manquante pour la totalité des stations, il n'y a donc pas de mesure d'AET pour cette espèce (*Proasellus aff. hypogeus* (ABDELKAD), Species\_278).

### **Temps de colonisation**

Le temps de colonisation pouvant impacter les taux de substitutions nucléaires, nous avons utilisé le gène de l'opsine pour estimer la proportion de la branche terminale qui représente une espèce souterraine. Le temps de colonisation a ainsi été estimé pour 89 espèces (voir Chap. 3, partie 2.4) et a été intégré dans les tests.

### **Tests statistiques**

Les espèces de surface ont été retirées du jeu de données du fait de leur larges aires de dispersion. En effet, pour ces espèces, l'environnement actuel représente probablement mal leur environnement passé, ce qui ne permet pas de les corrélérer au taux de substitutions moyens sur une longue période de temps avec les données environnementales mesurées aujourd'hui. Ainsi, on passe de 367 à 299 espèces pour le data367sp, et de 64 espèces à 49 espèces pour le data64sp.

La fiabilité des points a été vérifiée en testant la corrélation entre le taux estimé sur la branche (en dS/ra) et le temps que représente la branche, les branches les plus courtes ont été retirées jusqu'à ne plus trouver de corrélation. La présence de branches saturées a été vérifiée en corrélant le dS et le temps que représente la branche. Les points formant un plateau (*i.e.* le dS est beaucoup plus faible qu'attendu étant donné le temps qui s'est écoulé) ont également été retirés de l'analyse. La corrélation entre les taux de substitutions et les variables environnementales a ensuite été testée par un PGLS (Phylogenetic Generalized Least Square), en utilisant la phylogénie calibrée dans le temps pour calculer la matrice

de variance-covariance sous un modèle Brownien.

### 3 Résultats

#### Fiabilité des estimations de taux de substitutions

Afin d'éliminer les branches dont l'estimation des taux de substitutions n'est pas fiable, soit en raison de branches trop courtes, soit du fait de la saturation, une corrélation négative entre le  $d_S/ra$  et le temps que représente la branche a été testée.

**Gènes mitochondriaux** On trouve effectivement une corrélation négative entre le  $d_S/ra$  et le temps de divergence dans le compartiment mitochondrial (PGLS,  $p.value < 0.05$ ) quel que soit le jeu de données (Figure 6.1). Pour le `data367sp_mito` et `data367_COI`, cette corrélation disparaît lorsque les branches inférieures à 13 Ma sont retirées. Pour le `data64sp_mito`, la corrélation disparaît lorsque l'on enlève les branches inférieures à 6 Ma, les deux espèces dont la divergence est supérieur à 100 Ma semblent saturées et ont donc été retirées également.

**Gènes nucléaires** De façon surprenante, on trouve une corrélation positive dans les `data367sp` (deux gènes nucléaires ou FAST2 seul) entre le  $d_S/ra$  et le temps de divergences (PGLS,  $p.value < 0.05$ ). Cette corrélation positive est surprenante car rien n'explique une augmentation du taux de substitutions dans les longues branches. Une absence de biais dans les courtes branches dans le nucléaire est peu probable. En effet les gènes mitochondriaux, qui ont un taux de mutations supérieur et sont donc moins sujets aux problèmes d'estimation *a priori*, montrent un effet. L'explication la plus probable est que la variance est très élevée dans toutes les branches étant donné le faible nombre de positions et le taux d'évolution très faible de ces deux gènes nucléaires. Ainsi, pour ces deux gènes, la totalité des points a été conservée mais la puissance statistique est probablement

très faible, une absence de résultat sera donc difficilement interprétable. Pour le data64sp\_nucl, on ne voit pas de corrélation entre le taux de substitution et le temps que représente les branches (PGLS, p.value=0.63). Cependant, on observe visuellement une relation positive sur les plus longues branches et une variance importante sur les courtes branches (Figure 6.1). En retirant ces très courtes branches (<6Ma), on retrouve la corrélation positive (PGLS, p.value=0.014) observée dans le data367sp, qui reste toutefois inexplicée.

Ainsi, après retrait des points non fiables, les data367sp\_mito et data367\_COI sont substantiellement réduits. Après avoir retiré les espèces de surface et les points non fiables, on garde les 299 espèces pour les gènes nucléaires, mais seulement 118 espèces pour les gènes mitochondriaux. Pour le data64sp, on garde 41 espèces pour le compartiment mitochondrial et 43 pour le compartiment nucléaire (Table 6.1). Les données de temps de colonisation n'étant disponibles que pour 89 espèces, l'intégration de ce paramètre dans les modèles réduits encore considérablement le jeu de données : data367sp\_nucl et data367\_FAST2 passe ainsi à 71 espèces, data367\_mito et data367\_COI passe à 32 espèces, data64sp\_nucl passe à 26 espèces et data64sp\_mito à 25 (Table 6.1).

TABLE 6.1 – Jeux de données utilisés pour le calcul des taux de substitutions globaux ou synonymes (Substi.) avec le nombre de gènes (N gènes) et le nombre d'espèces sur lequel le calcul a été effectué. Les arbres ont été construits avec toutes les espèces (N sp tot), puis seules les espèces souterraines (N sp sout) ont été utilisées dans les corrélations, les branches trop courtes et saturées ont été retirées (N sp clean). Enfin, l'impact du temps de colonisation a été testé sur les espèces pour lesquelles le gène de l'opsine était disponible (N sp tcol).

Jeu de données	N gènes	N sp tot	N sp sout	N sp clean	N sp tcol	Substi.
data367sp_nucl	2	367	299	299	71	Global
data367sp_mito	2	367	299	118	32	Global
data367sp_FAST2	1	367	299	299	71	$d_S$
data367sp_COI	1	367	299	118	32	$d_S$
data64sp_nucl	346	64	49	43	26	$d_S$
data64sp_mito	13	64	49	41	25	$d_S$

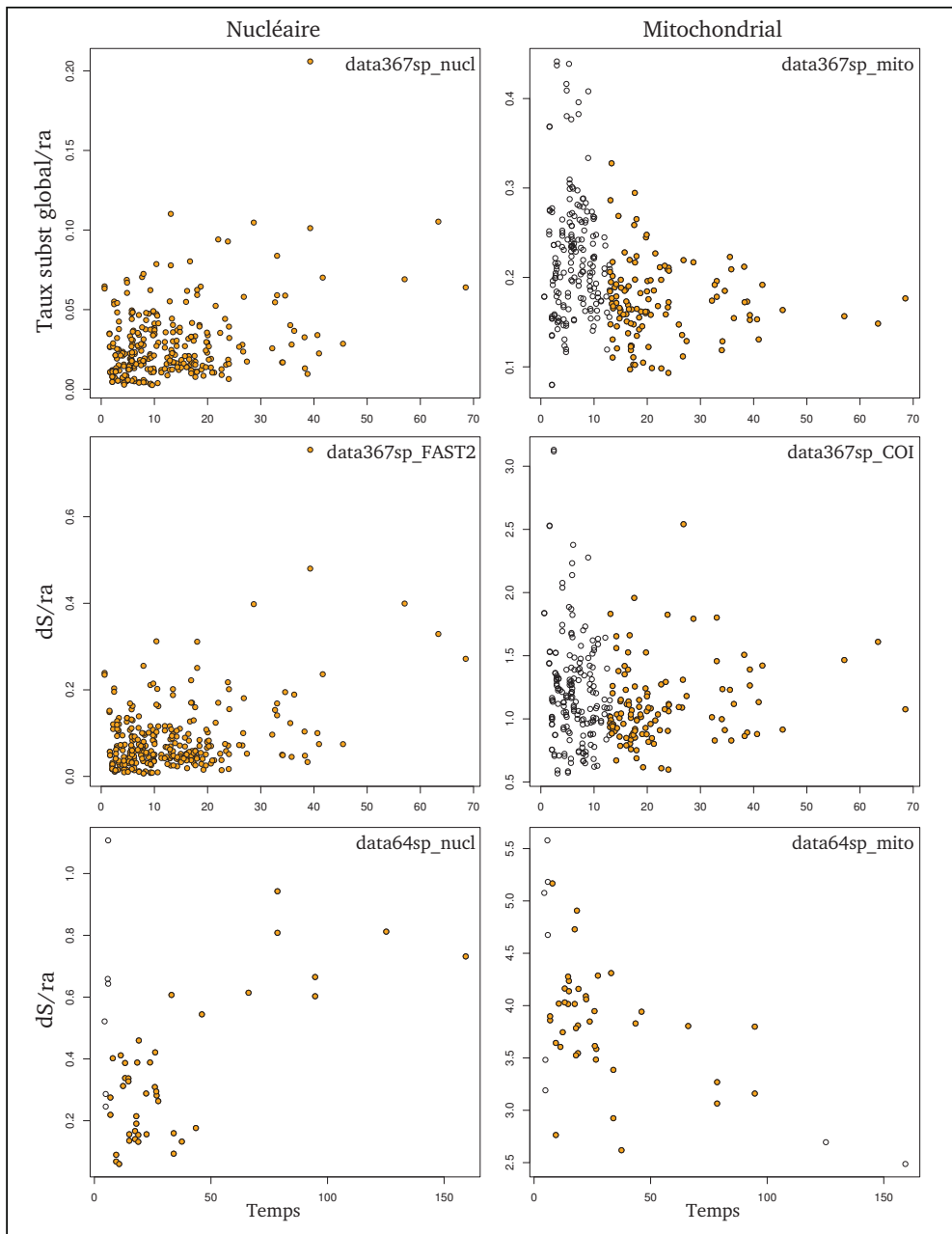


FIGURE 6.1 – Relation entre le taux de substitutions relatif à l'âge de la racine (global ou synonyme) et le temps que représente la branche sur laquelle il est calculé, dans les différents jeux de données. Les points orange sont les branches qui ont été conservées dans l'analyse, les points blancs indiquent les branches qui ont été retirées.

### Corrélations avec les variables environnementales

**Temps de colonisation.** Quel que soit le jeu de données, on ne retrouve aucune corrélation entre le taux de substitutions mitochondrial et le temps de colonisation



(PGLS,  $p.\text{values} > 0.1$ ), ce qui est conforme aux résultats du chapitre 3. Une tendance négative mais non significative est trouvée dans le compartiment nucléaire pour `data367sp_FAST2` (PGLS,  $p.\text{value} = 0.091$ ) et `data367sp_nucl` (PGLS,  $p.\text{value} = 0.097$ ). On retrouve bien la corrélation négative avec le taux de substitution nucléaire dans `data64sp_nucl` (PGLS,  $p.\text{value} = 0.0025$ ). L'intégration du temps de colonisation dans les relations a amené à réduire fortement le nombre de taxons. L'absence de significativité dans les gènes nucléaires des `data367sp` laisse ainsi supposer un problème de puissance, ne permettant pas réellement d'interpréter les résultats de ces tests. Les résultats sur la fiabilité des estimations laissent déjà présager un problème de puissance pour ce jeu de données. Ainsi, il n'est pas surprenant de ne pas être capable de mettre en évidence une corrélation significative. Le temps de colonisation n'étant pas corrélé au taux de substitutions mitochondrial, son intégration dans les modèles est inutile. Pour le taux de substitutions nucléaire, les tests seront effectués avec et sans le temps de colonisation.

**Variables environnementales** Quel que soit le jeu de données et le compartiment, aucune relation n'est trouvée entre le taux de substitutions et les variables environnementales (Table 6.2). Ni la latitude, ni la productivité (estimée par l'évapotranspiration), ni la température ou encore la différence de température depuis le dernier maximum glaciaire, ne permettent d'expliquer les variations de taux de substitutions. Une tendance positive est trouvée entre le taux de substitutions nucléaire et la productivité pour le `data64sp_nucl`, cette tendance disparaît lorsque l'on teste toutes les variables dans le même modèle (Annexe 8). Excepté ce dernier cas, le fait de tester les variables indépendamment ou de les inclure dans un même modèle, tout comme le fait d'ajouter le temps de colonisation dans la corrélation ne change pas les résultats (Annexe 8).

**Effet des courtes branches** Le retrait des courtes branches pose deux problèmes : la réduction drastique du jeu de données et le retrait du signal porté

CHAPITRE 6. EXISTE-T-IL DES PATRONS DE VARIATIONS DES TAUX  
D'ÉVOLUTION MOLÉCULAIRE À L'ÉCHELLE GLOBALE ?

TABLE 6.2 – Résultats des tests PGLS de la corrélation entre les taux de substitutions et les variables environnementales : Latitude, Température, évapotranspiration actuelle (AET) et différence de température entre aujourd'hui et le dernier maximum glaciaire (Delta LGM). Le nombre de taxa utilisés est réduit de 1 pour l'AET, un des points étant trop proche de la mer. Les branches trop courtes et saturées ont été retirées.

		data367sp nucl	data367sp mito	data367sp FAST2	data367sp COI	data64sp nucl	data64sp mito
	N	299/298	118/117	299/298	118/117	43	41
	Taux	global	global	$d_S$	$d_S$	$d_S$	$d_S$
Latitude	Pente	-0.01550	-0.00976	-0.01275	-0.00294	-0.03646	-0.00809
	P.value	0.41420	0.20570	0.53790	0.68810	0.09610	0.28740
	Cox-Snell $R^2$	0.00223	0.01348	0.00127	0.00137	0.06235	0.02722
Température	Pente	0.01907	0.00063	0.01247	0.00706	0.05082	0.01668
	P.value	0.11920	0.93270	0.35020	0.31370	0.10320	0.12060
	Cox-Snell $R^2$	0.00809	0.00006	0.00291	0.00856	0.06271	0.05706
AET	Pente	0.00006	0.00003	-0.00006	0.00001	0.00050	0.00009
	P.value	0.74620	0.75110	0.75710	0.92590	0.05420	0.30780
	Cox-Snell $R^2$	0.00035	0.00086	0.00044	0.00007	0.08260	0.02505
$\Delta$ LGM	Pente	0.01588	-0.01237	0.01701	0.02062	-0.04392	-0.02946
	P.value	0.71150	0.44340	0.71620	0.17610	0.41620	0.10890
	Cox-Snell $R^2$	0.00046	0.00497	0.00044	0.01539	0.01526	0.06076

par ces courtes branches. En effet, on pourrait imaginer que le fait de retirer les courtes branches biaise le jeu de données en tronquant une partie du gradient latitudinal par exemple. Les tests ont été effectués avec tous les points pour voir si la réduction des jeux de données avait un impact sur les résultats des tests. Lorsque tous les points sont inclus, on trouve une corrélation négative entre la latitude et le taux de substitutions mitochondrial et une corrélation positive avec l'AET dans le data367sp\_mito (Table 6.3). La corrélation avec la productivité est également trouvée dans le data367sp\_COI mais pas celle avec la latitude. De plus, lorsque les deux facteurs (AET et latitude) sont inclus dans le même modèle, la corrélation avec la latitude devient non significative (PGLS, p.value=0.9354), suggérant que c'est la corrélation entre AET et latitude qui génère cette relation. On note cependant que les  $R^2$  sont extrêmement faibles. On trouve également une corrélation positive entre la température et le  $d_S$  mitochondrial dans le data64sp\_mito, ainsi qu'une corrélation positive entre la productivité et le  $d_S$  nucléaire dans le data64sp\_nucl.

La répartition des points retirés dans le data367sp\_mito et le data367sp\_COI,

TABLE 6.3 – Résultats des tests PGLS de la corrélation entre les taux de substitutions et les variables environnementales lorsque les valeurs mal estimées ne sont pas retirées : Latitude, température, évapo-transpiration actuelle (AET) et différence de température entre aujourd'hui et le dernier maximum glaciaire ( $\Delta$  LGM). Le nombre de taxa utilisés est réduit de 1 pour l'AET, un des points étant trop proche de la mer.

		data_367sp	data367sp	data64sp	data64sp
		mito	COI	mito	nucl
	N	299/298	299/298	49	49
	Taux	global	$d_S$	$d_S$	$d_S$
Latitude	Pente	-0.01595	-0.00761	-0.00890	-0.04218
	P.value	0.03250	0.25070	0.27440	0.07800
	Cox-Snell $R^2$	0.01517	0.00440	0.02409	0.06140
Température	Pente	0.00261	0.00723	0.02523	0.05094
	P.value	0.58940	0.09070	0.02200	0.12460
	Cox-Snell $R^2$	0.00097	0.00953	0.10152	0.04700
AET	Pente	0.00022	0.00015	0.00011	0.00057
	P.value	0.00340	0.02830	0.28190	0.04770
	Cox-Snell $R^2$	0.02836	0.01600	0.02335	0.07691
Delta LGM	Pente	-0.00915	0.01687	-0.02896	-0.04784
	P.value	0.58910	0.26040	0.15750	0.43460
	Cox-Snell $R^2$	0.00098	0.00423	0.03996	0.01238

révèle une concentration de branches courtes entre de 43° et 44° (Figure 6.2). Si l'on regarde la répartition des branches considérées comme trop courtes dans la relation entre l'AET et le taux de substitutions mitochondrial (Figure 6.3), on peut voir que les valeurs de taux de substitutions les plus élevées ont toutes été retirées. Ces résultats suggèrent que la surestimation des taux de substitutions dans les courtes branches pourrait être à l'origine de la corrélation observée entre le taux de substitutions et la productivité dans le jeu de données non nettoyé.

## 4 Discussion

**Puissance des jeux de données** Malgré le grand nombre de taxa et de données transcriptomiques utilisées dans cette analyse, l'élimination des branches courtes qui ne permettent pas une estimation fiable des taux de substitutions a amené à une réduction très importante de la taille du data367sp sur le compartiment mitochondrial. En effet, ce retrait a engendré la perte de plus de la moitié des points. Cependant, il n'est pas nécessaire de retirer des analyses toutes les

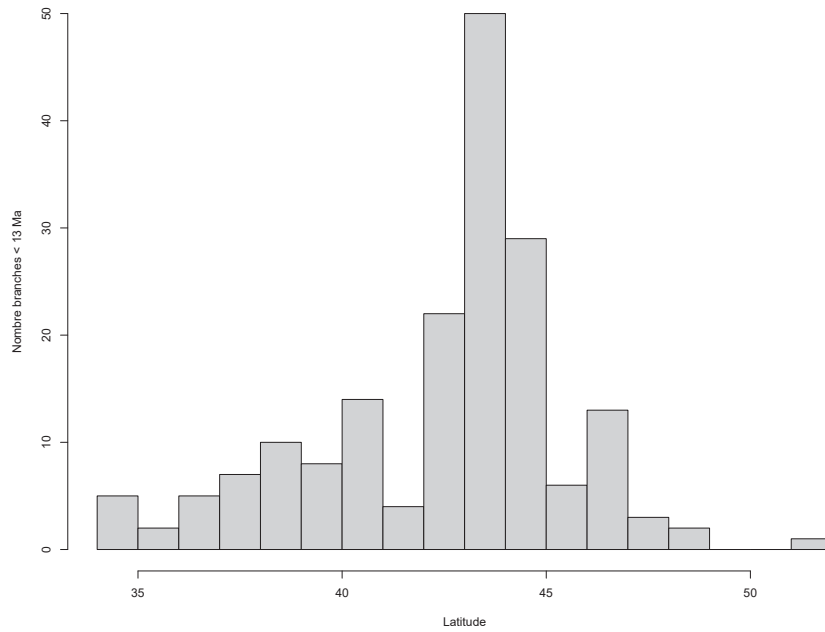


FIGURE 6.2 – Histogramme du nombre de branches retirées dans le data367sp (<13 Ma) en fonction de la latitude.

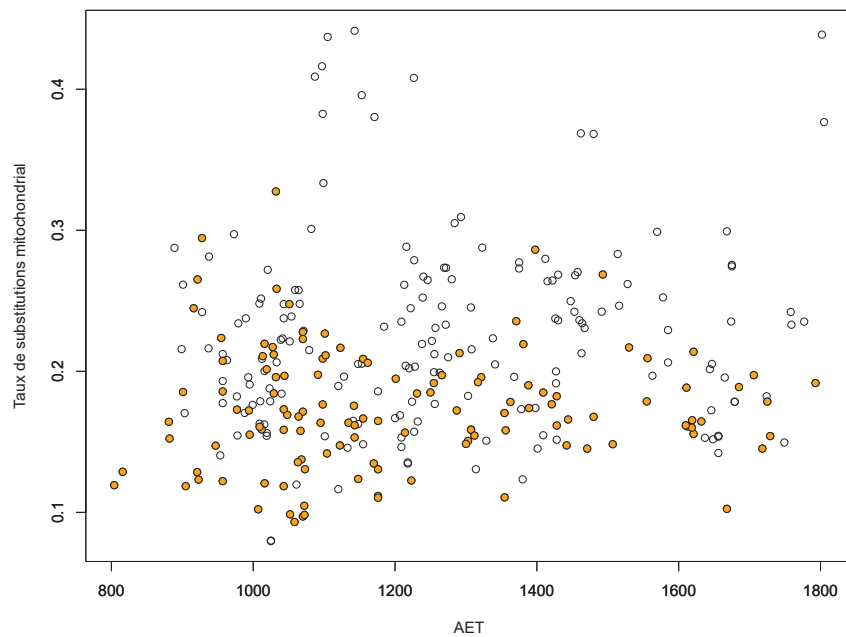


FIGURE 6.3 – Taux de substitutions global mitochondrial (data367sp\_mito) en fonction de l'évapotranspiration actuelle (AET) Les branches considérées comme bien estimées sont indiquées par des cercles orange, les branches retirées de la corrélation sont indiquées par des cercles blancs.

branches inférieures à 13 Ma. En effet, lorsque l'on retire une branche courte d'un arbre, la branche qui formait une paire avec la branche retirée sera alors allongée et son retrait ne sera pas nécessaire. Ainsi, un recalcul des taux de substitutions en retirant seulement une partie des branches trop courtes (c'est à dire en retirant une partie des espèces) permettrait de faire les calculs sur des branches plus longues pour un plus grand nombre de branches, et augmenter ainsi la puissance de l'analyse. Cependant, la puissance ne dépend pas uniquement du nombre d'espèces mais aussi du nombre de gènes utilisés dans l'analyse. Pour le jeu de données avec 367 espèces, le  $d_S$  n'a pu être calculé que sur un seul gène mitochondrial et un gène nucléaire. L'absence d'augmentation de la variance dans les courtes branches pour les gènes nucléaires suggère que la variance est très élevée même dans les branches les plus longues, et que la puissance de ces données pour détecter des changements de taux de substitutions est très faible. Ainsi, augmenter le nombre de gènes codant pour des protéines (permettant le calcul d'un  $d_S$ ) augmenterait la puissance de l'analyse. Toutefois, cette augmentation du nombre de gènes ne devrait pas se faire aux dépens du nombre d'espèces, l'échantillonnage taxinomique étant important dans la reconstruction phylogénétique (Philippe et al., 2011).

**Corrélation avec les variables environnementales** Les résultats présentés ici après retrait des branches considérées comme mal estimées, suggèrent qu'il n'existe pas de relation entre les variables environnementales variant latitudinalement et les taux de substitutions. Cependant, on peut noter deux particularités des Asellidae qui pourraient en faire un cas particulier. Premièrement, ils ne sont pas exposés aux UV. La quantité de rayonnements UV augmente lorsque la latitude baisse. Cependant, la quantité d'UV augmente avec l'altitude alors que la richesse spécifique montre le patron inverse (*e.g.* Touaylia et al., 2011; Lin et al., 2019), rendant peu probable un effet des UV sur la diversification via un changement du taux de mutations. Deuxièmement, il est probable que les traits d'histoire de vie des espèces souterraines ne varient pas latitudinalement. Ainsi,

une des hypothèses de Rohde (1992) était que le temps de génération étant corrélé à la latitude, les taux de substitutions devraient varier latitudinalement. Les résultats du Chapitre 4 indiquent que des changements de traits d'histoire de vie affectent les taux de substitutions nucléaires. Ici, nous ne trouvons pas de corrélation entre la latitude et le taux de substitutions nucléaire, suggérant que le temps de génération n'est donc pas corrélé à la latitude. Ainsi, si ces résultats indiquent une absence d'effet direct des variables environnementales sur les taux de substitutions, ils ne permettent pas de rejeter l'hypothèse d'une corrélation avec les traits d'histoire de vie pour les autres groupes (*e.g.* mammifères). Cependant, la diversité des Asellidae étant corrélée à la productivité du milieu et à la différence de température avec le dernier maximum glaciaire (Eme et al., 2014), l'absence de corrélation entre les taux de substitutions et ces deux variables environnementales suggère que la relation entre l'environnement et les taux de substitutions, même indirectement, ne permet pas d'expliquer la répartition de la biodiversité. Ces résultats vont à l'encontre de la première partie de l'Evolutionary Speed Hypothesis (Rohde, 1992) qui propose que les variables environnementales vont impacter le taux de mutations qui impactera à son tour le taux de diversification.

**Impact des branches courtes** Lorsque les branches trop courtes, pour lesquelles les taux de substitutions ont tendance à être surestimés, sont intégrées dans le jeu de données, une corrélation négative entre les taux de substitutions et la productivité est trouvée dans le compartiment mitochondrial pour le jeu de données data367sp. La variance expliquée par le modèle est cependant très faible ( $\approx 2\%$ ) et la corrélation n'est pas retrouvée dans le jeu de données data65sp. Dans le data367sp on trouve des corrélations avec la température pour le  $d_S$  mitochondrial et avec l'évapotranspiration pour le  $d_S$  nucléaire. Ces résultats sont difficiles à interpréter et suggèrent que ces corrélations sont des artefacts créés par la surestimation des taux de substitutions dans les courtes branches. En plus de la mauvaise estimation des taux de substitutions sur les branches trop courtes, il existe un dé-

bat sur l'existence d'un “time-dependency” sur les taux d'évolution moléculaire (Bandelt, 2008; Emerson et Hickerson, 2015; Ho et al., 2005, 2007, 2011). Ce time-dependency propose qu'il existe une décroissance exponentielle des taux de substitutions qui est fonction de l'intervalle de temps depuis le présent sur lesquels ils sont calculés (Figure 6.4).

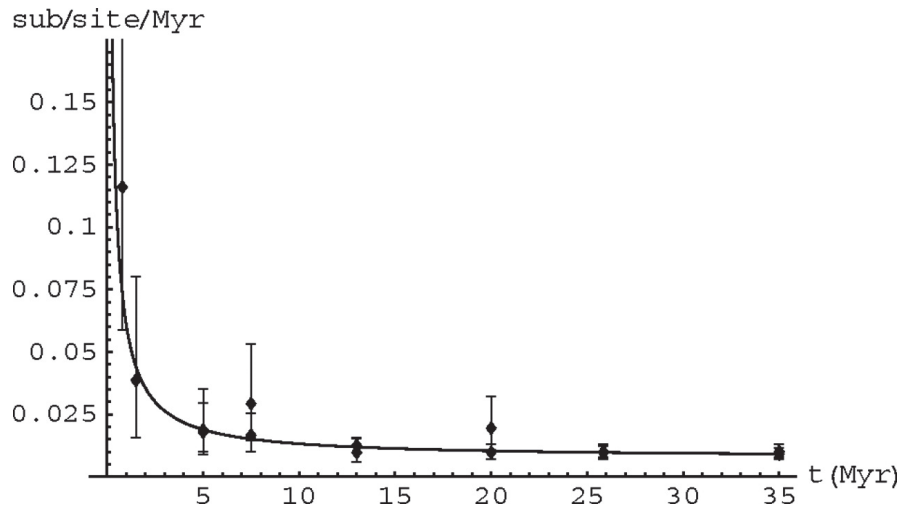


FIGURE 6.4 – Exemple de “time-dependency” des taux de substitutions en fonction de l'intervalle de temps (depuis le présent) sur lequel ils sont calculés. On voit une décroissance exponentielle de l'estimation lorsque l'intervalle de temps augmente<sup>10</sup>.

Si la raison de cette relation est fortement débattue entre effet du polymorphisme qui sera éliminé par la sélection naturelle, de la saturation, des erreurs de séquençage ou des erreurs de calibrations, il n'en reste pas moins que cette relation est souvent observée dans les jeux de données. Ces biais entraînant une surestimation des taux de substitutions dans les branches terminales trop courtes sont un problème particulièrement important dans l'étude de la relation avec des facteurs variant avec la diversité. En effet, les clades dont le taux de diversification est le plus élevé devraient avoir tendance à avoir un plus grand nombre de branches terminales très courtes. De plus, les clades les moins diversifiés devraient, quant à eux, avoir des longueurs de branches très importantes pouvant amener à une saturation. Ces deux biais pourraient ainsi amener à des branches courtes dont le taux de substitutions est surestimé pour les clades avec un taux de diversification

élevé et des branches longues pour lequel le taux de substitutions est sous-estimé dans les clades diversifiant peu.

Cet effet est supporté dans nos données par la concentration de branches très courtes entre 43° et 44°. Chez les Asellidae, il a été montré qu'il y a une grande concentration d'espèces à cette latitude (Eme et al., 2014). Ainsi, la relation entre variables environnementales, taux de substitutions et diversification pourrait n'être que le résultat de problèmes d'estimation des taux de substitutions liés à la structure de l'arbre. Si l'on revient sur les analyses ayant testées la relation entre taux de substitutions et latitude, très peu d'entre elles prennent en compte l'impact des courtes branches. Les rares études qui l'ont explicitement pris en compte montrent que la relation entre les taux de substitutions et les variables environnementales est faible (Orton et al., 2019; Bromham et al., 2015). Ces résultats suggèrent ainsi que les différences de taux de diversification entre les taxa pourraient générer une structure de l'arbre phylogénétique. Cette structure avec des branches courtes dans les clades diversifiés et des branches longues dans les clades peu diversifiés, pourrait générer une corrélation entre les taux de substitutions et la diversification. La diversification étant corrélée à des variables environnementales, cet artefact pourrait amener à une corrélation entre les variables environnementales et les taux de substitutions.

## 5 Conclusion

Les résultats de cette analyse suggèrent que les difficultés liés à l'estimation des taux de substitutions réduisent considérablement la puissance des jeux de données à détecter une corrélation entre les variables environnementales corrélées à la diversité et le taux de substitutions. D'autre part, la différence de résultats lorsque les branches courtes sont incluses ou non dans l'analyse suggère que les

---

10. Reprinted from *Molecular Biology and Evolution* Volume 23, Issue 12, Woodhams Michael, "Can Deleterious Mutations Explain the Time Dependency of Molecular Rate Estimates?", Pages 2271–2273, Copyright (2006), with permission from Oxford University Press.



corrélations trouvées dans la littérature pourraient être le résultat de biais dans l'estimation des taux de substitutions. Ainsi, pour pouvoir tester avec une puissance suffisante l'Evolutionary Speed Hypothesis, une meilleure compréhension du 'time-dependency' des taux de substitutions serait nécessaire afin de pouvoir mieux prendre en compte ce biais dans les analyses sans avoir à retirer une part substantielle des données.

# Chapitre 7

## Synthèse, Discussion & Perspectives

### 1 Synthèse

L'objectif de ce travail de thèse était de tester les hypothèses relatives aux variations des taux d'évolution moléculaire inter-spécifiques et d'évaluer leur impact relatif. Pour cela nous avons développé les données et connaissances autour des isopodes souterrains de la famille des Asellidae dont les caractéristiques sont particulièrement pertinentes dans le cadre de l'étude des taux d'évolution moléculaire. Afin d'évaluer les différentes hypothèses (temps de génération, taux métabolique, longévité, environnement), les taux de substitutions ont été calculés sur le compartiment nucléaire et sur le compartiment mitochondrial.

Ce travail de thèse a permis de délimiter les espèces du groupe et d'établir leurs relations de parenté et les temps de divergence. Ainsi, une phylogénie des 367 espèces d'Asellidae pour lesquelles nous disposons des données moléculaires suffisantes a été construite. Cette phylogénie a ensuite été calibrée dans le temps grâce à des événements bio-géographiques. Cette phylogénie a été utilisée ici pour calculer les taux de substitutions du chapitre 6 mais son usage potentiel est bien plus large. En effet, elle pourra, par exemple, permettre d'évaluer les variations

des taux de diversification au sein du groupe et ainsi d'identifier les facteurs influençant la diversification et la répartition des espèces. Elle permettra également de définir d'autres paires d'espèces souterrain-surface pour étudier l'effet de cette transition sur l'expression des gènes, la non-fonctionnalisation, la taille des populations, ...

Dans la première partie de ce travail de thèse (Chapitre 4), nous avons utilisé la transition vers le milieu souterrain pour établir des paires d'espèces proches, l'une vivant en milieu de surface et l'autre en milieu souterrain, permettant d'avoir des répliquats de la même transition. La transition vers le milieu souterrain étant accompagnée d'un allongement de la durée de vie et du temps de génération ainsi que d'une baisse du taux métabolique, la comparaison des espèces vivant dans les deux habitats nous a permis de tester les 3 principales hypothèses liant traits d'histoire de vie et taux de substitutions : l'hypothèse du temps de génération (Li et al., 1987), l'hypothèse du taux métabolique (Martin et al., 1992), et l'hypothèse de longévité (Nabholz et al., 2007). Alors que pour la première hypothèse, on s'attendait à un impact similaire dans les deux compartiments génomiques, les deux autres hypothèses prévoyaient un impact plus important dans le compartiment mitochondrial. La comparaison des taux de substitutions dans les deux compartiments nous a permis de mettre en évidence que les espèces souterraines avaient un taux de substitutions plus faible que leurs homologues de surface dans le compartiment nucléaire mais pas dans le compartiment mitochondrial. Ces résultats vont ainsi à l'encontre des trois hypothèses testées. Nous avons donc fait l'hypothèse que le temps de génération n'impacte pas le taux de substitutions mitochondrial en raison du fait que la réplication de ce génome est indépendante de la division cellulaire et donc du temps de génération, contrairement au génome nucléaire. Cette hypothèse permet d'expliquer qu'un effet est observé chez les mammifères car ces derniers ont une période de pause dans la gamétogenèse qui est égale au temps de génération et impacte les deux compartiments génomiques.

Dans la deuxième partie (Chapitre 5), nous avons utilisé la caractéristique des espèces souterraines d'être de faibles disperseurs afin de tester l'impact de l'environnement mutagène. En effet, cette très faible dispersion nous permet de faire l'hypothèse que l'environnement actuel est une bonne représentation de l'environnement dans lequel l'espèce persiste depuis de nombreuses générations. Ainsi nous avons mesuré la radioactivité présente dans les sédiments qui sont ingérés par les espèces pour 14 espèces souterraines d'Asellidae. Nous avons pu mettre en évidence dans cette partie un effet de l'environnement mutagène analogue à l'effet des traits d'histoire de vie montré dans la première partie, mais cette fois dans les deux compartiments génomiques. Nous avons également pu mettre en évidence un effet de la radioactivité naturelle sur le spectre mutationnel. En effet, la proportion de mutations C:G→A:T augmente avec la quantité de radioactivité. Ces mutations étant connues pour être générées par le stress oxydant (Shibutani et al., 1991), ces résultats suggèrent que la radioactivité génère un stress oxydant provoquant des mutations. Ce niveau de stress semble assez faible pour ne pas être détecté par les mécanismes de réparation de l'ADN.

Dans la dernière partie de ce travail (Chapitre 6), nous avons utilisé toutes les données moléculaires disponibles pour le groupe (Sanger, transcriptomes, génomes mitochondriaux, opsine) afin de tester le lien entre les taux de substitutions et les gradients environnementaux corrélés à la latitude sur des espèces souterraines. En effet, l'hypothèse de Rohde (1992) propose que les variables environnementales impactent les taux de mutations qui vont à leur tour impacter le taux de diversification, expliquant ainsi le gradient latitudinal de biodiversité. Les résultats suggèrent une absence de lien entre les variables environnementales variant latitudinalement et les taux de substitutions. Cependant, le retrait de certains points dont l'estimation du taux de substitutions était peu fiable a amené à une réduction très importante de la taille du jeu de données faisant douter de la puissance statistique de ces données pour détecter un effet. Enfin, la corrélation trouvée entre

les taux de substitutions et les variables environnementales lorsque les mauvaises estimations ne sont pas retirées laisse supposer que la mauvaise estimation des taux de substitutions dans les courtes branches pourrait générer une corrélation artefactuelle qui pourrait être courante dans la littérature.

Globalement, le couplage de données génomiques à la fois nucléaires et mitochondriales, sur un grand nombre de marqueurs avec des approches interdisciplinaires, appliqué au modèle biologique des Asellidae souterrains a permis (i) de mettre en évidence l'effet du temps de génération sur les taux de substitutions des protostomiens (ii) de montrer un effet de l'environnement mutagène et (iii) de mettre en évidence un artefact possible dans les tests liant les taux de diversification et les taux de substitutions.

## 2 Discussion

**Tester les différentes hypothèses** A travers de ce travail de thèse, nous avons pu voir que les hypothèses expliquant les variations de taux d'évolution moléculaires sont nombreuses et leur déconvolution difficile. En effet, à partir du constat que les taux d'évolution moléculaire sont corrélés au continuum de traits d'histoire de vie, de nombreux mécanismes sous-jacents ont été proposés pour expliquer un même patron (temps de génération, longévité, taux métabolique, hypothèse de Lynch, ...). Nous avons pu voir qu'établir une définition plus précise des attendus théoriques de chaque hypothèse permettait, en partie, de les différencier, mettant cependant en évidence la nécessité de différenciation des taux de substitutions synonymes et non-synonymes, mitochondriaux et nucléaires, ainsi que le besoin d'approches interdisciplinaires. Ainsi, le chapitre 4 nous a permis de montrer qu'en différenciant les taux de substitutions synonymes sur les deux compartiments génomiques on pouvait mettre en évidence un effet du temps de génération sur les taux de substitutions nucléaires chez les Asellidae. L'absence d'effet sur le compartiment mitochondrial nous a permis de rejeter l'hypothèse de longévité et du

taux métabolique et de remettre en cause l'universalité des attendus théoriques des différentes hypothèses. Cette approche purement génomique montre cependant rapidement ses limites pour différencier les autres hypothèses. Le chapitre 5 a ainsi mis en évidence qu'une approche interdisciplinaire (*i.e.* la génomique et la radiobiologie) permettait de montrer un effet de l'environnement mutagène sur les taux d'évolution moléculaire. Le chapitre 6 a finalement montré que le lien entre génomique et macro-écologie pouvait permettre de tester des hypothèses faisant le lien entre l'environnement et la diversification des espèces, ce champ d'étude demandant cependant à être développé. Ainsi, pour tester d'autres hypothèses, telles que les compromis évolutifs liés à la qualité des descendants ou au coût énergétique de la fidélité, des approches de biométrie (taille des descendants), de physiologie (réserves énergétiques) en encore de génomique fonctionnelle (impact des différents gènes sur le taux de mutations) semblent nécessaires.

**Effet des traits d'histoire de vie** Nous avons pu mettre en évidence l'effet bien connu du temps de génération sur les taux de substitutions. Cependant, nous avons pu montrer que cet effet n'existait que dans le compartiment nucléaire chez les Asellidae. Pour expliquer ce résultat surprenant, nous avons proposé une hypothèse de découplage de la réplication entre les deux compartiments, ce découplage se faisant sur des durées différentes chez les mammifères et les protostomiens, expliquant les différences de résultats. Ainsi, si l'on additionne ces résultats avec ceux de la littérature, il semble que seul le taux de réplifications des génomes dans les lignées germinales soit unanimement confirmé. En effet, l'impact du temps de génération pour les animaux, les plantes ou les bactéries semblent être confirmé par toutes les études qui l'ont testé. Ainsi, les erreurs lors de la réplication semblent être un facteur intrinsèque universel expliquant une part importante des variations des taux de substitutions. Cependant, certains articles commencent à remettre en cause la prédominance des mutations d'origine réplcatives (*e.g.* Gao et al., 2019; Sasani et al., 2019). De plus, le temps de génération ne permet pas d'expliquer

la totalité des variations observées. Dans le cas des Asellidae, le pseudo  $R^2$  de Cox-Snell pour le modèle expliquant le taux de substitutions nucléaire par le changement d'habitat est de 0.4. Cette valeur suggère qu'une assez grande partie des variations de taux de substitutions nucléaires reste inexpliquée. Pour le taux de substitutions mitochondrial, aucune relation avec les traits d'histoire de vie n'a été trouvée. Pourtant, le taux de substitutions mitochondrial est corrélé aux traits d'histoire de vie chez d'autres organismes. Par exemple, des variations de taux de substitutions mitochondriaux en lien avec les traits d'histoire de vie ont été mis en évidence chez des Sébastes (téléostéens, Hua et al. 2015), alors qu'aucune différence n'est observée dans le compartiment nucléaire, relançant l'hypothèse du taux métabolique. Ces résultats contradictoires suggèrent que même chez des groupes largement étudiés, il est nécessaire de réévaluer plus précisément la relation entre les traits d'histoire de vie, le taux métabolique et les taux d'évolution moléculaire avec des mesures précises des temps de génération, de la longévité, du nombre de descendants et des dégâts causés par le stress oxydant. De plus, Lanfear et al. (2007) ayant montrés que l'utilisation de proxy pouvait amener à des conclusions erronées sur la relation entre taux de substitutions et taux métabolique, une approche mêlant physiologie et génomique permettrait de tester de façon plus précise cette hypothèse.

**Universalité des hypothèses** Les attendus théoriques des différentes hypothèses ont été établies principalement à partir des connaissances chez les mammifères. Ces organismes ectothermes, ayant des tailles de populations plutôt réduites (en comparaison aux insectes, par exemple), des soins parentaux particulièrement importants et dont la gamétogenèse présente des particularités non négligeables, sont assez peu représentatifs de la biodiversité. Ainsi, le chapitre 4 a permis de mettre en évidence que des analyses chez d'autres organismes révèlent des résultats allant à l'encontre des hypothèses pré-existantes. Allio et al. (2017) ont également montré que le ratio de taux de substitutions mitochondrial et nucléaire

était différent entre les vertébrés et les protostomiens, confirmant que les facteurs impactant les taux d'évolution moléculaire n'ont pas la même influence chez tous les organismes. Il est aujourd'hui nécessaire de développer une stratégie pour démêler les facteurs qui ont un impact universel, de ceux qui agissent spécifiquement dans un groupe. Ce type d'approche à très large échelle taxinomique n'est cependant pas facile à conduire, ne serait-ce que dans la construction du jeu de données moléculaire. En effet, les approches maximisant le nombre d'espèces amènent à ne travailler que sur quelques gènes orthologues et ne sont pas les meilleures approches dans le cas des taux de substitutions car cela conduirait à une mauvaise estimation de ces taux.

**Impact de l'environnement mutagène** Le chapitre 5 a permis de mettre en évidence un impact de l'environnement mutagène sur les taux de substitutions, d'une ampleur analogue à l'effet du temps de génération. Ainsi, alors que la littérature s'est principalement focalisée sur l'effet des traits d'histoire de vie, nous montrons que l'environnement pourrait avoir un impact aussi important sur le taux d'évolution moléculaire. Un effet de la salinité sur les taux de substitutions avait déjà été montré chez des branchiopodes (Hebert et al., 2002), mais ce type d'analyse reste rare, suggérant que l'impact de l'environnement mutagène sur les taux d'évolution moléculaire a peut-être été négligé. Nous concluons, dans ce chapitre, que ces résultats questionnent sur le lien entre cet effet mutagène et les points chauds de diversité situés dans des zones de forte radioactivité, tels que les sources hydrothermales. Ce lien pose la question plus générale du lien entre le taux de mutations et la diversification qui est abordée dans le chapitre 6 à travers le lien entre des paramètres environnementaux corrélés à la distribution de la diversité et les taux de substitutions. Les résultats de ce dernier chapitre suggèrent qu'une telle relation n'existe pas, cependant les relations ne sont pas aisées à tester en raison des biais d'estimation dans les courtes et longues branches et demanderaient certains développements méthodologiques.



**Problèmes d'estimation des taux de substitutions** A travers le chapitre 6, nous avons pu aborder les problèmes d'estimation des taux d'évolution moléculaire dans les courtes branches et le débat autour du "time-dependency" des taux d'évolution moléculaire. Cette tendance à surestimer les taux de substitutions lorsqu'ils sont calculés sur des branches courtes et à les sous-estimer lorsque les branches sont trop longues pourrait être à l'origine de corrélations avec les facteurs liés à la diversification. La plupart du temps, ces problèmes d'estimation ne font que réduire la puissance des tests en raison de la trop grande variance dans l'estimation des taux de substitutions (Welch et Waxman, 2008). Dans le cas de la diversification, ces biais sont plus problématiques car ils pourraient créer une corrélation inexistante. En effet, les clades diversifiant plus fortement devraient avoir tendance à avoir des branches terminales plus courtes que les clades diversifiant faiblement. Ainsi, les taux de substitutions pourraient être surestimés dans les clades diversifiant fortement (time-dependency ou surestimation des branches trop courtes). A l'inverse, la saturation pourrait amener à une sous-estimation des taux de substitutions dans les clades qui diversifient moins et qui ont ainsi tendance à avoir des longues branches. Ainsi, il est nécessaire d'évaluer si ces problèmes d'estimation sont réellement capables de créer des corrélations auquel cas une réévaluation des corrélations trouvées entre la latitude et les taux d'évolution moléculaire serait nécessaire.

### 3 Perspectives

A travers ce travail de thèse, j'ai pu aborder les facteurs qui font varier les taux d'évolution moléculaire. Les traits d'histoire de vie, ainsi que l'environnement mutagène se sont révélés être des facteurs pouvant faire varier ces taux dans des proportions semblables. Grâce à ses caractéristiques, le modèle des *Asellidae* a permis d'aborder toutes ces questions fondamentales, et de tester certaines hypothèses sous un nouvel angle. Cependant, malgré leurs nombreux avantages,

ces organismes ne permettent pas d'aborder certaines questions. Ainsi, certaines perspectives de ce travail propose l'utilisation d'autres modèles pour aborder des questions qui n'ont pas pu l'être lors de cette thèse en raison des limitations du modèle biologique. La question qui vient naturellement en perspective de ce travail s'inclut dans la thématique plus globale de l'interaction entre les changements génomiques et les variations phénotypiques : quel est l'impact de ces variations de taux de substitutions sur les trajectoires évolutives des organismes ? Cette question très vaste peut être abordée sous de nombreux angles. Premièrement, on peut s'interroger sur l'impact des variations de taux de substitutions sur la survie ou la reproduction des organismes. En effet, si la plupart des mutations sont neutres ou délétères (Kimura, 1968; Ohta, 1987; Eyre-Walker et Keightley, 2007), le taux de mutations devrait être réduit par le processus de sélection naturelle. Cependant, cette réduction devrait se faire dans la limite du coût des mécanismes assurant la fidélité de la réplication. Le compromis entre le coût des mutations (fardeau mutationnel) et le coût de la fidélité devrait dépendre de l'ampleur de l'effet délétère des mutations. Cette ampleur pourrait dépendre des traits d'histoire de vie des espèces et notamment de l'investissement dans chaque descendant (Britten, 1986). Cet effet délétère des mutations sur les organismes peut-être 'simple' : la mutation par elle-même engendre un phénotype moins performant, ou plus complexe : la mutation engendre une incompatibilité du gène muté avec d'autres gènes, endommageant ainsi la fonction qu'ils assurent de concert. Par exemple, les complexes protéiques sont codées par plusieurs gènes qui 'coopèrent' entre eux pour assurer la formation d'une enzyme fonctionnelle. Cette coopération pourrait être une pression de sélection sur les mutations intervenant sur ces gènes. Deuxièmement, un taux élevé de mutations pourrait engendrer plus rapidement des incompatibilités hybrides, empêchant la reproduction entre les organismes portant une mutation avec ceux qui ne la porte pas. Ces incompatibilités pourraient entraîner des variations de la probabilité d'une espèce à donner de nouvelles espèces, en lien avec leur

taux de substitutions. Cette question a été abordée dans le chapitre 6 sans toutefois être testée directement. Ainsi, je présenterai dans ces perspectives les moyens qui pourraient être mis en oeuvre pour mieux comprendre l'impact de variations des taux d'évolution moléculaire sur l'évolution des organismes en m'intéressant premièrement à l'impact de l'investissement parental sur le compromis entre le fardeau mutationnel et le coût de la fidélité, deuxièmement à l'impact de la sélection pour assurer la coopération entre les gènes sur ce même compromis et enfin, à l'impact du taux de substitutions sur la probabilité de générer de nouvelles espèces.

### **3.1 Tester l'existence d'une contrainte sur le taux d'évolution moléculaire en lien avec l'investissement parental**

Dès 1986, Britten avait proposé que le taux de substitutions étaient sous l'influence d'un compromis évolutif entre le nombre de mutations transmises à la descendance et l'investissement parental. Cette hypothèse suppose ainsi que plus les individus investissent d'énergie dans chaque descendant, plus l'impact d'une mutation délétère amenant à la mort (ou la non-reproduction) d'un de ces descendants est fort. Ainsi, il devrait exister un compromis entre le fardeau mutationnel et le coût des mécanismes empêchant les mutations qui devrait dépendre de l'investissement parental dans chaque descendant. Tester l'hypothèse que le taux de substitutions est au coeur d'un compromis entre l'investissement parental et le coût des mécanismes empêchant les mutations demande premièrement d'estimer le taux de substitutions et l'investissement parental. Cependant, l'investissement parental est généralement corrélé à d'autres traits d'histoire de vie dont nous avons montré que certains peuvent faire varier les taux de substitutions : la longévité, le temps de génération et le taux métabolique. Ainsi, afin de prendre en compte ces différents traits, il est nécessaire de les inclure dans le modèle. Cependant, pour déterminer les parts de variance expliquée par ces différents facteurs, il est

nécessaire de disposer de valeurs précises de ces facteurs et non de proxy. Ainsi, les données disponibles dans la littérature ou dans les bases de données ne nous permettent pas d'atteindre le niveau de précision suffisant. Il faut donc développer un modèle biologique et un protocole expérimental nous permettant de faire des mesures de ces différents traits d'histoire de vie. Les données génomiques nécessaires au calcul des taux d'évolution moléculaire devront ensuite être acquises afin de tester la corrélation entre ces taux et les traits d'histoire de vie mesurés. A cette approche corrélative, on peut ajouter une approche plus 'mécanistique' en étudiant le fonctionnement des gènes liés au taux de mutations, c'est-à-dire des gènes assurant les fonctions de fidélité de la réplication, de réparation ou encore de protection contre les dégâts oxydatifs.

### **Estimation des traits d'histoire de vie**

L'investissement dans la descendance peut-être évalué en mesurant les réserves énergétiques des descendants à la naissance (avant qu'ils ne se nourrissent). Il semble que le plus aisé soit d'utiliser des organismes ovipares afin de mesurer la vitellogénine dans les oeufs. Cette vitellogénine est la lipoprotéine dominante dans la constitution du vitellus. Elle est produite dans le foie puis excrétée dans le sang, elle migre ensuite dans les ovocytes où elle est stockée. La vitellogénine peut être dosée en chromatographie en phase liquide couplée à la spectrométrie de masse en tandem (LC-MS/MS). Les protéines extraites des oeufs sont digérées par une enzyme, les peptides ainsi créés sont ensuite séparés par chromatographie en fonction de leur hydrophobicité puis ionisés afin d'être identifiés et quantifiés par spectrométrie de masse.

Cette mesure de l'investissement parental ne suffit cependant pas, il faut également mesurer les autres traits d'histoire de vie qui pourraient impacter les taux de substitutions : la longévité, le temps de génération et le taux métabolique.

Le paramètre à mesurer pour estimer la longévité est la longévité moyenne des individus. Ainsi, il est nécessaire de travailler sur des espèces dont la longévité en

captivité est mesurable, c'est-à-dire des organismes qui ne sont pas trop longévifs et qui supportent bien les conditions d'élevage. La nécessité d'étudier un gradient de longévité élimine tous les groupes ne contenant que des espèces annuelles.

Le temps de génération correspond au temps moyen qui s'écoule entre un individu et ses descendants. Il ne s'agit donc pas de l'âge de la maturité sexuelle ou de l'âge à la première reproduction. Pour estimer ce temps de génération, il faut relever l'âge des individus à chaque reproduction et faire une moyenne pour chaque individu de toutes les reproductions au cours de sa vie.

Le taux métabolique est la quantité d'énergie dépensée par l'organisme par unité de temps. On mesure généralement le métabolisme de base qui reflète la quantité d'énergie minimum dépensée par l'organisme, au repos, pour assurer les fonctions vitales. Cette mesure est cependant discutée car elle ne reflète pas le métabolisme réel de l'organisme. De plus, elle est difficile à obtenir pour les petits animaux. D'autre part, la température à laquelle l'organisme est considéré comme au repos pour les ectothermes varie en fonction des espèces et elle est difficile à déterminer. L'hypothèse du taux métabolique (Martin et Palumbi, 1993) est basée sur l'impact mutagène des espèces réactives de l'oxygène (ROS) formées par la chaîne respiratoire. Ainsi, mesurer directement la quantité de dommages créés par les ROS apporterait une plus grande précision. Les ROS ont un effet bien connu sur l'ADN : ils créent une oxydation des guanines qui sont alors transformées en 8-oxo-2'-désoxyguanosine (8-OH-dG). Ces 8-OH-dG sont dosables après une extraction d'ADN et une digestion enzymatique afin de séparer les bases puis de doser les résidus de guanine oxydés. Le dosage se fait à l'aide de kit (*e.g.* DN A/RNA Oxydative Damage EIA kit, Cayman 589320). Cependant, comme nous l'avons conclu dans le chapitre 4, le taux métabolique de l'organisme n'est probablement pas représentatif du taux métabolique des cellules germinales. Or, seules les mutations dans les cellules germinales sont transmises à la descendance et ont un impact sur les taux d'évolution moléculaire. Ainsi, ce dosage devra être

fait sur des cellules germinales, nécessitant des organismes pouvant être disséqués pour prélever ces cellules.

### **Choix du modèle biologique et protocole expérimental**

La mesure de ces différents paramètres demande un modèle qui comporte plusieurs espèces, ovipares, élevable en laboratoire, avec des longévités différentes mais pas trop longues et qu'il est possible de disséquer pour obtenir des cellules germinales sur lesquelles mesurer les dégâts oxydatifs. Les Asellidae utilisés dans cette thèse présentent le gros désavantage d'être difficilement mises en élevage. Leur longévité importante rendrait de plus toute expérimentation extrêmement longue ( $\approx 8$  ans). De plus, les expériences tentées sur les oeufs (Annexe 1) nous ont permis de voir qu'il était difficile d'en obtenir un grand nombre et que leur petite taille ne permet pas de mesurer les ressources qu'ils contiennent. Ainsi différentes espèces de 'poissons' ou de gastéropodes sembleraient permettre de répondre à tous ces objectifs.

Plusieurs espèces devront être sélectionnées sur un gradient de traits d'histoire de vie. Pour pouvoir mesurer l'âge à chaque reproduction, afin de ne pas contraindre la reproduction dans les élevages, le plus simple est de mettre en élevage des individus de même âge. Les élevages devront être visités chaque jour afin de ramasser les pontes pour mesurer la quantité de vitellogénine dans les oeufs. La date de chaque ponte devra être relevée pour estimer le temps de génération. La date de décès des individus devra également être relevée pour estimer la longévité de chaque espèce. Quelques individus devront être utilisés pour chaque espèce (une dizaine de chaque sexe) afin d'effectuer une dissection pour prélever les cellules germinales, sur lesquelles un dosage des 8-OH-dG sera effectué (le séquençage d'ADN peut-être effectué sur les mêmes individus).

### **Estimation des taux de substitutions et corrélation avec les traits d'histoire de vie**

Afin d'estimer le taux de substitutions de façon robuste pour toutes les espèces, il est nécessaire d'avoir un nombre important de gènes. On peut séquencer le génome, le transcriptome ou faire de la capture de gènes (cette dernière méthode demandant cependant un travail préalable de design de sondes pour sélectionner les gènes d'intérêt). Un individu par espèce devra être utilisé pour le séquençage. Les individus utilisés pour les mesures de dégâts oxydatifs peuvent être utilisés en prenant toutefois garde à la conservation des échantillons pour le séquençage de transcriptome. A partir de ces données et des connaissances sur les taxa sélectionnés, une phylogénie des espèces devra être construite.

Les séquences obtenues ainsi que la phylogénie et les traits mesurés pour chaque espèce pourront être intégrés directement dans le logiciel CoEvol, conçu pour estimer la co-variation des taux de substitutions et des traits d'histoire de vie, en tenant compte des corrélations entre les différents traits. Pour les mesures de 8-OH-dG, les mesures effectuées sur les deux sexes devront être intégrées dans la corrélation avec les taux de substitutions nucléaires. En revanche, seules les mesures effectuées sur les femelles devront être intégrées dans la corrélation avec le taux de substitutions mitochondrial. Pour les autres traits, aucune différence entre les deux sexes n'est théoriquement prévue. Cependant il serait intéressant de tester si les traits d'un des deux sexes expliquent mieux les variations de taux de substitutions.

### **Efficacité des gènes liés au taux de mutations**

Afin de tester plus précisément l'implication des gènes liés au taux de mutations, l'efficacité de ces gènes pourrait être testée chez les différentes espèces en inactivant ces gènes avec la technique CRISPR-CAS9. Cependant, il est probable que l'inactivation de ces gènes entraîne des effets très délétères, voire la mort des

individus. Une approche plus standard de calcul du  $d_N/d_S$  de ces gènes pourrait également être effectuée. Un résultat positif pourra être interprété comme un effet mais un résultat négatif ne pourra être interprété. En effet, une seule mutation pourrait être à l'origine d'une baisse d'efficacité et ne ferait pas varier le  $d_N/d_S$ .

Le dispositif expérimental présenté devrait ainsi permettre de tester l'existence d'une relation entre l'investissement dans chaque descendant et le taux de substitutions, tout en testant la relation avec le taux métabolique, la longévité et le temps de génération.

### **3.2 Tester l'existence d'une contrainte sur le taux de substitutions en lien avec la compatibilité mito-nucléaire**

Certains gènes assurent des fonctions en collaboration avec d'autres gènes. Par exemple, la cytochrome c oxydase est une enzyme composée de 14 sous-unités protéiques, chacune codée par un gène. Même si une mutation sur l'un de ces gènes n'a pas d'influence directe sur la sous-unité protéique produite, elle peut impacter l'interaction de cette sous-unité avec les autres sous-unités et *in fine* le phénotype. Ainsi, cette coopération entre différents gènes pourrait agir comme une force pour réduire le taux de mutations.

Un des exemples les plus parlant de collaboration génétique est la collaboration qui existe entre l'ADN mitochondrial et l'ADN nucléaire pour assurer certaines fonctions essentielles telles que le métabolisme énergétique (Wolff et al., 2014). Cette collaboration est d'autant plus complexe que les forces évolutives qui s'exercent sur ces deux génomes ne sont pas identiques. Chez la plupart des organismes, le génome mitochondrial a un taux de mutations beaucoup plus élevé que le génome nucléaire, il est non-recombinant et sa taille de population efficace est plus faible. Toutes ces caractéristiques font qu'il est plus sensible à l'accumulation de mutations par dérive génétique et par hitch-hiking. Chou et Leu (2015) décrivent la coopération comme un processus de 'Reine-rouge' (Van Valen, 1973)



dans lequel le génome nucléaire est soumis à une pression de sélection forte pour compenser les mutations mitochondriales afin d'assurer la collaboration entre les deux génomes. Cette course perpétuelle devrait amener soit à une évolution rapide des loci nucléaires impliqués dans la collaboration soit à une réduction du taux de substitutions sur les gènes mitochondriaux impliqués (Schmidt et al., 2001). Ainsi, Dowling et al. (2008) considère que la coopération entre les deux génomes est un trait à part entière qui conditionne l'évolution des deux génomes. Ce trait pourrait ainsi exercer une pression de sélection pour réduire le taux de substitutions mitochondrial afin d'éviter l'émergence de mutations non-synonymes pouvant rompre cette coopération.

Cependant, un cas extrême de taux de substitutions mitochondrial découvert récemment à Lyon sur l'espèce *Physa acuta*, un escargot d'eau douce, remet en cause toute cette théorie et révèle les lacunes dans nos connaissances sur l'évolution de la coopération entre les deux génomes. J'ai pu montrer au cours de ma thèse qu'au sein de cette population, il existe 4 génomes mitochondriaux différents (appelés mitotypes par la suite), dont l'un présente un taux de substitutions extrêmement élevé (Figure 7.1).

Le mitotype divergent est associé à une inversion importante des gènes dans le génome mitochondrial (Figure 7.2). Les autres mitotypes trouvés dans la même population montrent d'autres particularités. Les mitotypes A et B ont une zone de taille inconnue qui n'a pas pu être assemblée, le mitotype C présente une inversion entre les gènes *nad4l* et *cox2*. Les inversions chez le mitotype divergent et le mitotype C se trouvent dans la même région que la partie qui n'a pas pu être assemblée chez les deux autres mitotypes. Lors de l'assemblage des génomes mitochondriaux, j'ai pu observer que les mitotypes divergents montraient très peu de polymorphisme alors que pour les autres mitotypes, on observe un polymorphisme important.

En revanche, de façon surprenante, le génome nucléaire ne montre aucune



FIGURE 7.1 – Phylogénie des individus de l'espèce *Physa acuta*. On détecte 4 mitotypes différents dont un est extrêmement divergent.

divergence entre le type 'divergent' et les types 'normaux' (Figure 7.3). Si l'on s'intéresse aux gènes nucléaires qui codent pour des fonctions en coopération avec les gènes mitochondriaux, ils montrent également très peu de variations. De plus, les individus ayant un génome mitochondrial divergent ne montrent pas de différence dans leur métabolisme énergétique qui est assuré par la chaîne respiratoire mitochondriale, dont le fonctionnement dépend de l'interaction avec le génome nucléaire (communication personnelle Cyril Degletagne).

Ainsi, on observe un taux de substitutions mitochondrial extrêmement élevé et une inversion dans la synténie des gènes mitochondriaux sans qu'aucun effet ne soit observé ni dans le génome nucléaire, y compris sur les gènes impliqués dans la collaboration mito-nucléaire, ni sur le phénotype des individus, remettant totalement en cause nos *a priori* sur l'impact du taux de mutations dans l'évolution de la coopération entre les deux génomes. Approfondir nos connaissances sur ce

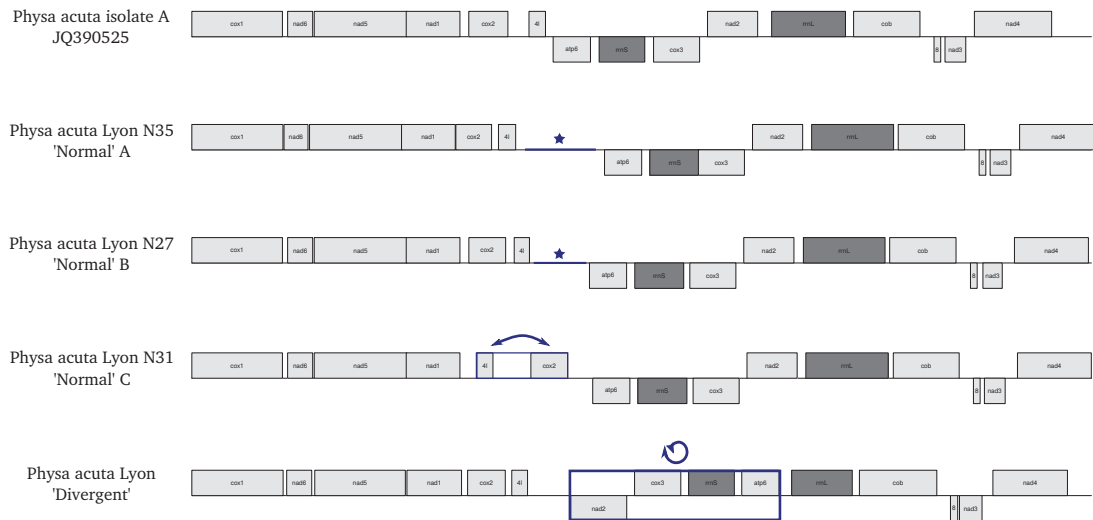


FIGURE 7.2 – Ordre des gènes dans les génomes mitochondriaux des 4 mitotypes de *Physa acuta* trouvés dans la population de Lyon, en comparaison avec le génome ‘standard’ extrait de la base de données NCBI. Les gènes codés sur le brin + sont schématisés au dessus de la ligne et les gènes codés sur le brin - sont en dessous de la ligne. Le gène nad4l est noté 4l, le gène ATP6 est noté 6 et le gène ATP8 est noté 8. Les inversions sont indiqués par des flèches. Le génome ‘Divergent’ montre une grande inversion d’un fragment par rapport aux autres mitotypes. Le mitotype C montre une inversion des gènes nad4l et nad4. Les mitotypes de Lyon A et B ont une zone de taille inconnue qui n’a pas pu être assemblée (indiquée par une étoile).

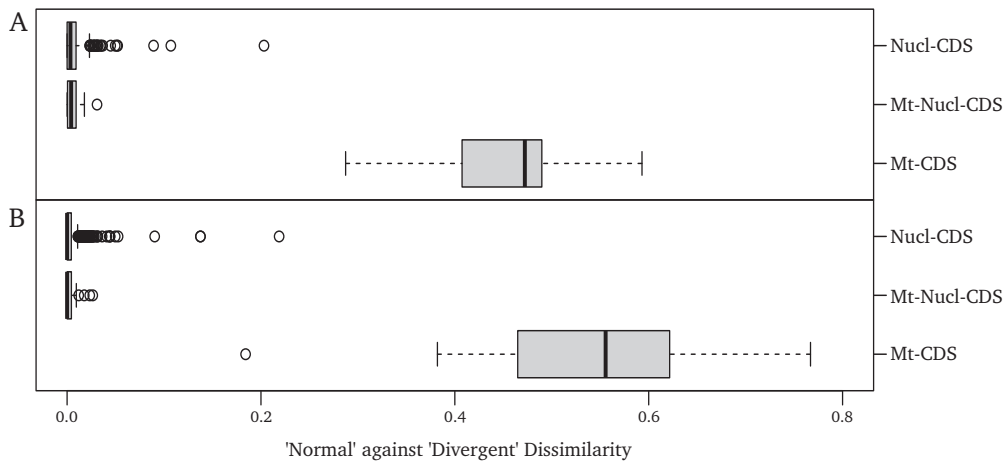


FIGURE 7.3 – Scores de dissimilarité (p-distances) en nucléotides (A) ou en protéines (B) entre les gènes du mitotype C et les gènes du mitotype ‘Divergent’ sur les gènes mitochondriaux (Mt-CDS), nucléaires (Nucl-CDS) et nucléaires codant pour des fonctions en interactions avec des gènes mitochondriaux (Mt-Nucl-CDS).

modèle pourrait nous permettre de comprendre comment un taux de substitutions aussi élevé peut évoluer sans impacter la coopération mito-nucléaire.

La première interrogation lorsque l'on voit la forte similarité entre les individus du mitotype divergent est de savoir si le taux de mutations est toujours élevé dans le génome mitochondrial ou si cette augmentation a été passagère et que le taux de mutations est aujourd'hui revenu à la normale. Une approche expérimentale par lignée d'accumulation (mère + descendants) nous permettrait d'estimer le taux de mutations actuel. Cette espèce se reproduisant très vite, il est possible de séquencer le génome mitochondrial de plusieurs générations de la même lignée. Un séquençage de génome faible couverture est suffisant pour assembler les mitochondries complètes, celle-ci étant présente en de nombreux exemplaires dans chaque cellule.

Afin de mieux comprendre cette augmentation du taux de substitutions, il est ensuite nécessaire de comprendre l'origine des remaniements observés dans le génome mitochondrial et leur lien avec le taux de mutations. Chez les deux mitotypes ne montrant pas d'inversion, une zone proche du site d'inversion s'est avérée impossible à assembler. Même les PCR sur cette zone ont échouées, laissant supposer que cette zone est trop grande et/ou trop répétée. Il pourrait s'agir de la région contrôle, région d'où démarre la transcription du génome mitochondrial et qui n'a pas été identifiée chez ces espèces. Cette région contrôle est généralement une zone difficile à identifier tant sa structure est variable en fonction des espèces. Par exemple, chez les insectes cette région peut faire de 350 à 4600 pb selon les espèces (voir Zhang et Hewitt, 1997). Certaines caractéristiques permettent cependant de l'identifier : c'est une région généralement riche en AT avec des répétitions en tandem qui évoluent de façon concertée. Un événement au niveau de la région contrôle, tel qu'une insertion d'élément transposable, pourrait expliquer les inversions observées. Cette région étant très répétée et sa taille étant inconnue, un assemblage à partir de lectures génomiques courtes a de grandes chances

d'échouer. En utilisant la technologie de séquençage par nanopore (MinION) on pourrait séquencer tout l'ADN mitochondrial en un seul fragment, ce qui devrait permettre de décrire cette région sans étape d'assemblage afin d'analyser sa composition.

Ces différentes analyses nous permettront de mieux comprendre les observations faites sur cette population et à terme d'évaluer les mécanismes qui ont permis une telle augmentation du taux de substitutions sans impacter la coopération entre les deux génomes.

### **3.3 Existe-t-il un lien entre les taux de substitutions et les taux de diversification ?**

Nous avons vu, dans le chapitre 6, qu'il existe des hypothèses qui lient le taux de substitutions et le taux de diversification. Ce lien pourrait se faire de deux façons : premièrement, un taux de substitutions élevé pourrait amener à augmenter la vitesse d'apparition des incompatibilités hybrides lorsque deux populations sont isolées. Deuxièmement, un taux de substitutions élevé pourrait favoriser l'apparition de variants permettant soit la survie dans un environnement changeant, soit la colonisation de nouvelles niches. Nous nous intéresserons ici au lien entre taux de diversification et taux de substitutions.

L'étude du lien entre taux de diversification et taux de substitutions pose de nombreux problèmes méthodologiques. Premièrement, la façon dont les taux de diversification sont intégrés dans la corrélation est très variable. Les différentes études qui ont testé ce lien ont souvent utilisé la taille du clade (diversité en espèce du clade). Le problème de cette approche est que l'on obtient une seule mesure de diversification par clade qui doit être confrontée à une seule mesure de taux de substitutions par clade. Les auteurs sélectionnent alors généralement une espèce du clade pour faire le calcul des taux de substitutions. Une très grande quantité d'information est donc perdue dans cette procédure. Une autre méthode consiste

à calculer le taux de substitutions de la racine à la feuille pour chaque feuille puis de corrélérer cette mesure avec le nombre de noeuds entre la racine et chaque feuille. Cette méthode pose deux problèmes majeurs : premièrement, les mêmes branches internes sont utilisées un grand nombre de fois, deuxièmement le nombre de noeuds est souvent corrélé à la longueur cumulée des branches en raison d'un artefact de densité de noeuds (Hugall et Lee, 2007). Cet artefact engendre une augmentation des taux de substitutions lorsque le nombre de noeuds sur le chemin entre la racine à la feuille, augmente. Cela vient du fait que lorsqu'il y a un grand nombre de noeuds (et donc un grand nombre d'espèces), les modèles de reconstruction phylogénétique prennent mieux en compte la saturation et reconstruisent une plus grande proportion des mutations qui se sont produites le long des branches.

Pour éviter l'artefact du nombre de noeuds, la meilleure solution est souvent de travailler sur les branches terminales. Cependant dans les cas des taux de diversification, cela implique d'être capable d'estimer le taux de diversification sur les branches terminales. Si cela semble contre-intuitif, cette estimation est néanmoins possible grâce à l'approche récemment publiée par Maliet et al. (2019). Cette approche permet de modéliser l'évolution du taux de diversification le long des branches de l'arbre. Le taux de diversification de la branche fille étant issu d'une distribution normale centrée sur le taux de diversification de la branche parente. L'utilisation de ce modèle sur la phylogénie des Asellidae révèle de grandes variations du taux de diversification le long de l'arbre (Figure 7.4).

Cependant, les résultats du chapitre 6 nous ont amené à conclure que certains biais dans le calcul des taux de substitutions pourraient amener à des corrélations artéfactuelles entre les taux de substitutions et les taux de diversification, même en les calculant sur des branches terminales (time-dependency et mauvaise estimation des courtes branches). Ainsi, pour mieux comprendre la relation entre ces deux taux, la première étape serait d'évaluer l'impact de ces biais. Une approche passant par des simulations permettrait de déterminer si la structure de l'arbre peut créer

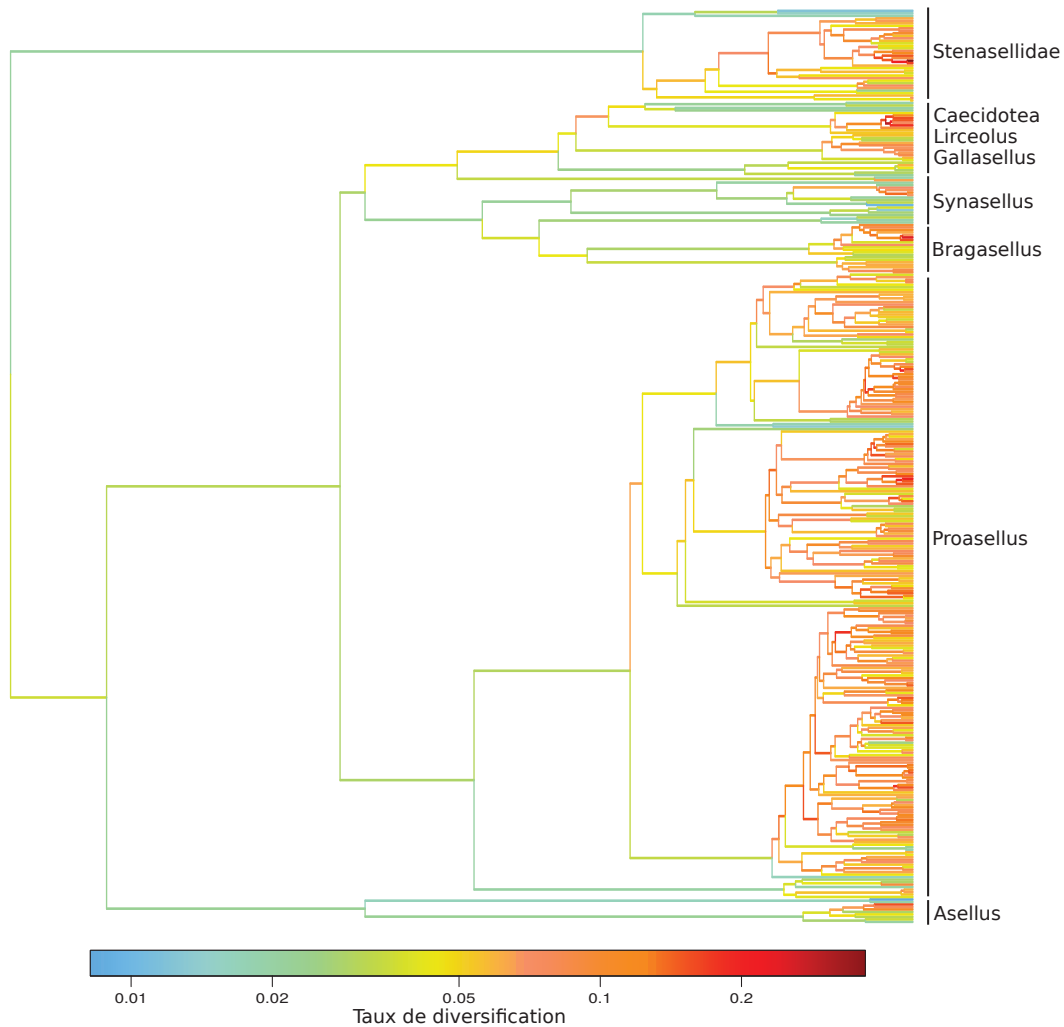


FIGURE 7.4 – Taux de diversification le long de la phylogénie des Asellidae, calculés avec le logiciel ClaDS (Maliot et al., 2019) sous un modèle ClaDS2 (taux de spéciation et d’extinction variables).

artificiellement une corrélation.

En utilisant la phylogénie des Asellidae, on peut simuler des alignements de séquences sous l’hypothèse qu’il n’y a aucune corrélation entre la diversification et le taux de substitutions. A partir de ces alignements, on peut calculer les taux de substitutions le long des branches. On peut ensuite tester la corrélation entre les taux de substitutions ainsi calculés et les taux de diversification connus. Si une corrélation est trouvée entre les taux de substitutions et les taux de diversification, on pourra conclure que ces biais sont capables de générer la relation.

Un moyen simple d’éviter un tel biais serait alors de supprimer les branches

trop courtes et trop longues de l'arbre. Cependant, le retrait de points est toujours problématique car l'information principale pourrait se trouver dans les espèces les plus récentes (notamment parce qu'elles reflètent mieux l'environnement actuel). Ainsi, la meilleure approche serait de prendre en compte ce biais dans les analyses. Pour cela, une première approche serait d'inclure une augmentation de la variance dans les courtes branches (hétéroscédasticité) et la décroissance exponentielle de la moyenne à l'intérieur du test. Une deuxième approche serait le développement d'un logiciel permettant d'estimer la co-évolution entre les taux de diversification et les taux de substitutions le long de l'arbre, en intégrant directement dans le modèle ces biais d'estimation. Une telle approche permettrait, de plus, de tenir compte de l'effet de la variance dans l'estimation des dates de divergence sur l'estimation des taux de diversification.

L'utilisation du modèle *Asellidae* pour tester cette relation diversification-taux de substitutions demanderait également l'acquisition d'un plus grand nombre de gènes pour toutes les espèces. En restreignant l'étude au genre *Proasellus*, une approche par capture de gènes permettrait d'obtenir les 346 gènes orthologues en copie unique, trouvés dans les transcriptomes, aux 252 espèces de ce genre. Cette méthode consiste à créer des sondes d'ARN biotinylées correspondant aux gènes que l'on souhaite obtenir. Ces sondes vont s'hybrider sur les séquences d'intérêt qui seront 'capturées' (les sondes biotinylées viennent d'accrocher sur des billes) puis séquencées. Cette technique permettrait d'obtenir un nombre suffisant de gènes sans passer par un séquençage du transcriptome complet qui serait difficilement réalisable pour un nombre aussi élevé d'espèces.





# Bibliographie

- Abadi S, Azouri D, Pupko T, Mayrose I. 2019. Model selection may not be a mandatory step for phylogeny reconstruction. *Nature Communications*. 10 :934.
- Achaz G, Gangloff S, Arcangioli B. 2019. The quiescent X, the replicative Y and the Autosomes. *bioRxiv*. .
- Allen AP, Gillooly JF, Savage VM, Brown JH. 2006. Kinetic effects of temperature on rates of genetic divergence and speciation. *Proceedings of the National Academy of Sciences*. 103 :9130–9135.
- Allio R, Donega S, Galtier N, Nabholz B. 2017. Large Variation in the Ratio of Mitochondrial to Nuclear Mutation Rate across Animals : Implications for Genetic Diversity and the Use of Mitochondrial DNA as a Molecular Marker. *Molecular Biology and Evolution*. 34 :2762–2772.
- Ames BN. 1989. Endogenous DNA damage as related to cancer and aging. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 214 :41–46.
- Arbiza L, Patricio M, Dopazo H, Posada D. 2011. Genome-Wide Heterogeneity of Nucleotide Substitution Model Fit. *Genome Biology and Evolution*. 3 :896–908.
- Baceta JJ, Pujalte V, Serra-Kiel J, Robador A, Orue-Etxebarria X. 2004. El Maastrichtiense final, Paleoceno e Ilerdiense inferior de la Cordillera Pirenaica. In : Geología de España, pp. 308–313.

- Bandelt HJ. 2008. Clock debate : when times are a-changin' : Time dependency of molecular rate estimates : tempest in a teacup. *Heredity*. 100 :1–2.
- Barja G. 2002. Endogenous oxidative stress : relationship to aging, longevity and caloric restriction. *Ageing Research Reviews*. 1 :397–411.
- Barr TC, Holsinger JR. 1985. Speciation in Cave Faunas. *Annual Review of Ecology and Systematics*. 16 :313–337.
- Barraclough TG, Savolainen V. 2001. Evolutionary Rates and Species Diversity in Flowering Plants. *Evolution*. 55 :677–683.
- Baynton K, Fuchs RP. 2000. Lesions in DNA : hurdles for polymerases. *Trends in Biochemical Sciences*. 25 :74–79.
- Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritsch G, Pütz J, Middendorf M, Stadler PF. 2013. MITOS : Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*. 69 :313–319.
- Bleiweiss R. 1998. Slow rate of molecular evolution in high-elevation hummingbirds. *Proceedings of the National Academy of Sciences*. 95 :612–616.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic : a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30 :2114–2120.
- Bou C. 1974. Recherches sur les eaux souterraines. Méthodes de récolte dans les eaux souterraines interstitielles. *Ann. SpCICol*. 29 :611419.
- Boussau B, Blanquart S, Necșulea A, Lartillot N, Gouy M. 2008. Parallel adaptations to high temperatures in the Archaean eon. *Nature*. 456 :942–945.
- Boutin CL, Coineau N. 1990. Regression Model', 'Modèle Biphase'd'volution et origine des Micro-organismes stygobies interstitiels continentaux. *Rev. Micropaléont.* 33 :302–322.

- Brash DE. 2015. UV Signature Mutations. *Photochemistry and Photobiology*. 91 :15–26.
- Brennan AC, Hiscock SJ, Abbott RJ. 2019. Completing the hybridization triangle : the inheritance of genetic incompatibilities during homoploid hybrid speciation in ragworts (Senecio).  *AoB PLANTS*. 11.
- Britten RJ. 1986. Rates of DNA sequence evolution differ between taxonomic groups. *Science*. 231 :1393–1398.
- Bromham L. 2009. Why do species vary in their rate of molecular evolution? *Biology Letters*. p. rsbl.2009.0136.
- Bromham L, Hua X, Lanfear R, Cowman PF. 2015. Exploring the Relationships between Mutation Rates, Life History, Genome Size, Environment, and Species Richness in Flowering Plants. *The American Naturalist*. 185 :507–524.
- Bromham L, Rambaut A, Harvey PH. 1996. Determinants of rate variation in mammalian DNA sequence evolution. *Journal of Molecular Evolution*. 43 :610–621.
- Brown JE, Alfonso B, Avila R, Beresford NA, Copplestone D, Hosseini A. 2016. A new version of the ERICA tool to facilitate impact assessments of radioactivity on wild plants and animals. *Journal of Environmental Radioactivity*. 153 :141–148.
- Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*. 12 :59–60.
- Casquet J, Thebaud C, Gillespie RG. 2012. Chelex without boiling, a rapid and easy technique to obtain stable amplifiable DNA from small amounts of ethanol-stored spiders. *Molecular Ecology Resources*. 12 :136–141.

- Castresana J. 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution*. 17 :540–552.
- Chang BH, Shimmin LC, Shyue SK, Hewett-Emmett D, Li WH. 1994. Weak male-driven molecular evolution in rodents. *Proceedings of the National Academy of Sciences*. 91 :827–831.
- Charif D, Lobry JR. 2007. SeqinR 1.0-2 : A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. Springer Berlin Heidelberg.
- Charlesworth D. 2010. Don't forget the ancestral polymorphisms. *Heredity*. 105 :509–510.
- Charniaux-Cotton. 1973. Description et contrôle de l'ovogenèse chez les Crustacés supérieurs. *Annales de Biologie Animale Biochimie Biophysique*. 13 :21–37.
- Chou JY, Leu JY. 2015. The Red Queen in mitochondria : cyto-nuclear co-evolution, hybrid breakdown and human disease. *Frontiers in Genetics*. 6.
- Clayton DA, Doda JN, Friedberg EC. 1974. The Absence of a Pyrimidine Dimer Repair Mechanism in Mammalian Mitochondria. *Proceedings of the National Academy of Sciences*. 71 :2777–2781.
- Coale AJ. 1972. Growth and Structure of Human Populations : A Mathematical Investigation. Princeton University Press.
- Collins B, Okada S, Scholes G, Weiss JJ, Wheeler CM. 1965. Chain Scission and Hydrogen Bond Breakage on Irradiation of DNA. *Radiation Research*. 25 :526–536.
- Cox RA. 2003. Correlation of the rate of protein synthesis and the third power

- of the RNA :protein ratio in *Escherichia coli* and *Mycobacterium tuberculosis*. *Microbiology*. 149 :729–737.
- Crandall KA, Hillis DM. 1997. Rhodopsin evolution in the dark. *Nature*. 387 :667–668.
- Cvetkov L. 1968. Un filet phréatobiologique. *Bulletin de l'Institut de Zoologie et Musée, Sofia*. 27 :215–218.
- Czeizel AE, Elek C, Susánszky E. 1991. The evaluation of the germinal mutagenic impact of Chernobyl radiological contamination in Hungary. *Mutagenesis*. 6 :285–288.
- Datry T, Dole-Olivier MJ, Marmonier P, Claret C, Perrin JF, Lafont M, Breil P. 2008. *La zone hyporhéique, une composante à ne pas négliger dans l'état des lieux et la restauration des cours d'eau*. p. 17.
- Davies TJ, Savolainen V, Chase MW, Moat J, Barraclough TG. 2004. Environmental energy and evolutionary rates in flowering plants. *Proceedings of the Royal Society of London. Series B : Biological Sciences*. .
- DavitBéal T, Tucker AS, Sire JY. 2009. Loss of teeth and enamel in tetrapods : fossil record, genetic data and morphological adaptations. *Journal of Anatomy*. 214 :477–501.
- Dawson KJ. 1998. Evolutionarily Stable Mutation Rates. *Journal of Theoretical Biology*. 194 :143–157.
- Debruyne R, Poinar HN. 2009. Time Dependency of Molecular Rates in Ancient DNA Data Sets, A Sampling Artifact ? *Systematic Biology*. 58 :348–360.
- Dercourt J, Gaetani M, Vrielynck B, de la carte géologique du monde C. 2000. Atlas Peri-Tethys : Paleogeographical maps.

- Dobzhansky T. 1936. Studies on Hybrid Sterility. II. Localization of Sterility Factors in *Drosophila Pseudoobscura* Hybrids. *Genetics*. 21 :113–135.
- Doolittle RF, Blombäck B. 1964. Amino-Acid Sequence Investigations of Fibrinopeptides from Various Mammals : Evolutionary Implications. *Nature*. 202 :147–152.
- Dowling DK, Friberg U, Lindell J. 2008. Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends in Ecology & Evolution*. 23 :546–554.
- Drake JW. 1993. General Antimutators are Improbable. *Journal of Molecular Biology*. 229 :8–13.
- Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. Rates of Spontaneous Mutation. *Genetics*. 148 :1667–1686.
- Drummond AJ, Rambaut A. 2007. BEAST : Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*. 7 :214.
- Dubrova YE, Nesterov VN, Krouchinsky NG, Ostapenko VA, Neumann R, Neil DL, Jeffreys AJ. 1996. Human minisatellite mutation rate after the Chernobyl accident. *Nature*. 380 :683–686.
- Duchene D, Bromham L. 2013. Rates of molecular evolution and diversification in plants : chloroplast substitution rates correlate with species-richness in the Proteaceae. *BMC Evolutionary Biology*. 13 :1.
- DugoCota CastroviejoFisher S, Vilà C, GonzalezVoyer A. 2015. A test of the integrated evolutionary speed hypothesis in a Neotropical amphibian radiation. *Global Ecology and Biogeography*. 24 :804–813.
- Duret L, Arndt PF. 2008. The Impact of Recombination on Nucleotide Substitutions in the Human Genome. *PLOS Genetics*. 4 :e1000071.

- Dutheil J, Boussau B. 2008. Non-homogeneous models of sequence evolution in the Bio++ suite of libraries and programs. *BMC Evolutionary Biology*. 8 :1.
- Edgar RC. 2004. MUSCLE : multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. 32 :1792–1797.
- Eme D, Malard F, Colson-Proch C, Jean P, Calvignac S, Konecny-Dupré L, Hervant F, Douady CJ. 2014. Integrating phylogeography, physiology and habitat modelling to explore species range determinants. *Journal of Biogeography*. 41 :687–699.
- 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 :5685–5699.
- Eme D, Zgamaister M, Delić T, et al. (11 co-authors). 2018. Do cryptic species matter in macroecology? Sequencing European groundwater crustaceans yields smaller ranges but does not challenge biodiversity determinants. *Ecography*. 41 :424–436.
- Emerson BC, Hickerson MJ. 2015. Lack of support for the time-dependent molecular evolution hypothesis. *Molecular Ecology*. 24 :702–709.
- Emms D, Kelly S. 2018. OrthoFinder2 : fast and accurate phylogenomic orthology analysis from gene sequences. *bioRxiv*. p. 466201.
- Esterbauer H. 1993. Cytotoxicity and genotoxicity of lipid-oxidation products. *The American Journal of Clinical Nutrition*. 57 :779S–786S.
- Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. *Nature Reviews Genetics*. 8 :610–618.
- Felsenstein J. 1985. Phylogenies and the Comparative Method. *The American Naturalist*. 125 :1–15.



- Fitch WM, Bruschi M. 1987. The evolution of prokaryotic ferredoxins—with a general method correcting for unobserved substitutions in less branched lineages. *Molecular Biology and Evolution*. 4 :381–394.
- Fišer C, Zagmajster M, Zakšek V. 2013. Coevolution of life history traits and morphology in female subterranean amphipods. *Oikos*. 122 :770–778.
- Foucreau N. 2013. *Étude des réponses écophysiologicals et fonctionnelles de populations de l'organisme clé Gammarus pulex (crustacea, Amphipoda) dans un contexte de changement climatique, au sein de la vallée du Rhône.* .
- Francois CM, Duret L, Simon L, Mermillod-Blondin F, Malard F, Konecny-Dupré L, Planel R, Penel S, Douady CJ, Lefébure T. 2016a. No Evidence That Nitrogen Limitation Influences the Elemental Composition of Isopod Transcriptomes and Proteomes. *Molecular Biology and Evolution*. 33 :2605–2620.
- Francois CM, Mermillod-Blondin F, Malard F, Fourel F, Lécuyer C, Douady CJ, Simon L. 2016b. Trophic ecology of groundwater species reveals specialization in a low-productivity environment. *Functional Ecology*. 30 :262–273.
- Freeze RA, Cherry JA. 1979. Groundwater. Englewood Cliffs, N.J : Prentice-Hall.
- Friedberg EC, Walker GC, Siede W, Wood RD. 2005. DNA Repair and Mutagenesis. American Society for Microbiology Press. Google-Books-ID : VAKs-BAAAQBAJ.
- Friedrich M, Chen R, Daines B, Bao R, Caravas J, Rai PK, Zagmajster M, Peck SB. 2011. Phototransduction and clock gene expression in the troglobiont beetle Ptomaphagus hirtus of Mammoth cave. *Journal of Experimental Biology*. 214 :3532–3541.
- Fuller N, Ford AT, Nagorskaya LL, Gudkov DI, Smith JT. 2018. Reproduction in the freshwater crustacean Asellus aquaticus along a gradient of radionuclide

- contamination at Chernobyl. *Science of The Total Environment*. 628-629 :11–17.
- Galtier N, Duret L. 2007. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. *Trends in Genetics*. 23 :273–277.
- Galtier N, Jobson RW, Nabholz B, Glémin S, Blier PU. 2009. Mitochondrial whims : metabolic rate, longevity and the rate of molecular evolution. *Biology Letters*. p. rsbl.2008.0662.
- Gangloff S, Achaz G, Francesconi S, Villain A, Miled S, Denis C, Arcangioli B. 2017. Quiescence unveils a novel mutational force in fission yeast. *eLife*. p. 19.
- Gao Z, Moorjani P, Sasani TA, Pedersen BS, Quinlan AR, Jorde LB, Amster G, Przeworski M. 2019. Overlooked roles of DNA damage and maternal age in generating human germline mutations. *Proceedings of the National Academy of Sciences*. 116 :9491–9500.
- Gent DCv, Hoeijmakers JHJ, Kanaar R. 2001. Chromosomal stability and the DNA double-stranded break connection. *Nature Reviews Genetics*. 2 :196–206.
- Gil R, Silva FJ, Zientz E, et al. (13 co-authors). 2003. The genome sequence of *Blochmannia floridanus* : Comparative analysis of reduced genomes. *Proceedings of the National Academy of Sciences*. 100 :9388–9393.
- Gillman LN, Keeling DJ, Gardner RC, Wright SD. 2010. Faster evolution of highly conserved DNA in tropical plants. *Journal of Evolutionary Biology*. 23 :1327–1330.
- Gillman LN, Keeling DJ, Ross HA, Wright SD. 2009. Latitude, elevation and the tempo of molecular evolution in mammals. *Proceedings of the Royal Society B : Biological Sciences*. 276 :3353–3359.

- Gillman LN, Wright SD. 2013. Patterns of Evolutionary Speed : In Search of a Causal Mechanism. *Diversity*. 5 :811–823.
- Gillooly JF, Allen AP, West GB, Brown JH. 2005. The rate of DNA evolution : Effects of body size and temperature on the molecular clock. *Proceedings of the National Academy of Sciences*. 102 :140–145.
- Giraud A, Matic I, Radman M, Fons M, Taddei F. 2002. Mutator Bacteria as a Risk Factor in Treatment of Infectious Diseases. *Antimicrobial Agents and Chemotherapy*. 46 :863–865.
- Goldie X, Lanfear R, Bromham L. 2011. Diversification and the rate of molecular evolution : no evidence of a link in mammals. *BMC Evolutionary Biology*. 11 :1.
- Grabherr MG, Haas BJ, Yassour M, et al. (21 co-authors). 2011. Trinity : reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature biotechnology*. 29 :644.
- Grabowski M, Mamos T, Baćela-Spychalska K, Rewicz T, Wattier RA. 2017. Neogene paleogeography provides context for understanding the origin and spatial distribution of cryptic diversity in a widespread Balkan freshwater amphipod. *PeerJ*. 5 :e3016.
- Grafen A, Hamilton WD. 1989. The phylogenetic regression. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*. 326 :119–157.
- Grudzenski S, Raths A, Conrad S, Rube CE, Löbrich M. 2010. Inducible response required for repair of low-dose radiation damage in human fibroblasts. *Proceedings of the National Academy of Sciences*. 107 :14205–14210.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies : Assessing the Performance of PhyML 3.0. *Systematic Biology*. 59 :307–321.

- Guéguen L, Duret L. 2018. Unbiased Estimate of Synonymous and Nonsynonymous Substitution Rates with Nonstationary Base Composition. *Molecular Biology and Evolution*. 35 :734–742.
- Haas BJ, Papanicolaou A, Yassour M, et al. (23 co-authors). 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*. 8 :1494–1512.
- Hahn C, Bachmann L, Chevreur B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research*. 41 :e129–e129.
- Haldane JBS. 1946. The Mutation Rate of the Gene for Haemophilia, and Its Segregation Ratios in Males and Females. *Annals of Eugenics*. 13 :262–271.
- Ham RCHJv, Kamerbeek J, Palacios C, et al. (16 co-authors). 2003. Reductive genome evolution in *Buchnera aphidicola*. *Proceedings of the National Academy of Sciences*. 100 :581–586.
- Harman D. 1956. Aging : A Theory Based on Free Radical and Radiation Chemistry. *Journal of Gerontology*. 11 :298–300.
- Hebert PDN, Remigio EA, Colbourne JK, Taylor DJ, Wilson CC. 2002. Accelerated Molecular Evolution in Halophilic Crustaceans. *Evolution*. 56 :909–926.
- Henry JP, Magniez G. 1983. Introduction pratique à la systématique des organismes des eaux continentales françaises - 4. Crustacés Isopodes (principalement Asellotes). *Publications de la Société Linnéenne de Lyon*. 52 :319–357.
- Henry JPA. 1976. Recherches sur les Asellidae hypogés de la lignée cavaticus : Crustacea, Isopoda, Asellota. [s.n.]. [S.l.].
- Hervant F, Renault D. 2002. *Fasting and refeeding in crustaceans*. p. 9.

- Ho SYW, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigo AG, Cooper A. 2011. Time-dependent rates of molecular evolution. *Molecular Ecology*. 20 :3087–3101.
- Ho SYW, Phillips MJ, Cooper A, Drummond AJ. 2005. Time Dependency of Molecular Rate Estimates and Systematic Overestimation of Recent Divergence Times. *Molecular Biology and Evolution*. 22 :1561–1568.
- Ho SYW, Shapiro B, Phillips MJ, Cooper A, Drummond AJ. 2007. Evidence for Time Dependency of Molecular Rate Estimates. *Systematic Biology*. 56 :515–522.
- Hoeijmakers JHJ. 2001. Genome maintenance mechanisms for preventing cancer. *Nature*. 411 :366–374.
- Holsinger JR. 1994. Amphipoda. In : *Encyclopaedia Biospeologica*. (Eds C. Juberthie and V. Decou.), pp. 147–163.
- Houlihan DF, McMillan DN, Agnisola C, Genoino IT, Foti L. 1990. Protein synthesis and growth in *Octopus vulgaris*. *Marine Biology*. 106 :251–259.
- Howarth F. 1987. The evolution of non-relictual tropical troglobites. *International Journal of Speleology*. 16.
- Hua X, Cowman P, Warren D, Bromham L. 2015. Longevity Is Linked to Mitochondrial Mutation Rates in Rockfish : A Test Using Poisson Regression. *Molecular Biology and Evolution*. p. msv137.
- Hugall AF, Lee MSY. 2007. The Likelihood Node Density Effect and Consequences for Evolutionary Studies of Molecular Rates. *Evolution*. 61 :2293–2307.
- Hulbert AJ, Pamplona R, Buffenstein R, Buttemer WA. 2007. Life and Death : Metabolic Rate, Membrane Composition, and Life Span of Animals. *Physiological Reviews*. 87 :1175–1213.

- Hübscher U, Maga G, Spadari S. 2002. Eukaryotic DNA Polymerases. *Annual Review of Biochemistry*. 71 :133–163.
- Ielsch G, Cuney M, Buscail F, Rossi F, Leon A, Cushing ME. 2017. Estimation and mapping of uranium content of geological units in France. *Journal of Environmental Radioactivity*. 166 :210–219.
- Ikemura T. 1982. Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. *Journal of Molecular Biology*. 158 :573–597.
- Inoue M, Shen GP, Chaudhry MA, Galick H, Blaisdell JO, Wallace SS. 2004. Expression of the Oxidative Base Excision Repair Enzymes is not Induced in TK6 Human Lymphoblastoid Cells after Low Doses of Ionizing Radiation. *Radiation Research*. 161 :409–417.
- Jackson AL, Loeb LA. 1998. The Mutation Rate and Cancer. *Genetics*. 148 :1483–1490.
- Jacobs GH. 2013. Losses of functional opsin genes, short-wavelength cone photopigments, and color vision—A significant trend in the evolution of mammalian vision. *Visual Neuroscience*. 30 :39–53.
- Jiang P, Josue J, Li X, Glaser D, Li W, Brand JG, Margolskee RF, Reed DR, Beauchamp GK. 2012. Major taste loss in carnivorous mammals. *Proceedings of the National Academy of Sciences*. 109 :4956–4961.
- Jin YH, Clark AB, Slebos RJC, Al-Refai H, Taylor JA, Kunkel TA, Resnick MA, Gordenin DA. 2003. Cadmium is a mutagen that acts by inhibiting mismatch repair. *Nature Genetics*. 34 :326–329.
- Jobson RW, Albert VA. 2002. Molecular Rates Parallel Diversification Contrasts between Carnivorous Plant Sister Lineages<sup>1</sup>. *Cladistics*. 18 :127–136.

- Jobson RW, Nabholz B, Galtier N. 2010. An Evolutionary Genome Scan for Longevity-Related Natural Selection in Mammals. *Molecular Biology and Evolution*. 27 :840–847.
- John JCS, Facucho-Oliveira J, Jiang Y, Kelly R, Salah R. 2010. Mitochondrial DNA transmission, replication and inheritance : a journey from the gamete through the embryo and into offspring and embryonic stem cells. *Human Reproduction Update*. 16 :488–509.
- Karpinets TV, Greenwood DJ, Sams CE, Ammons JT. 2006. RNA :protein ratio of the unicellular organism as a characteristic of phosphorous and nitrogen stoichiometry and of the cellular requirement of ribosomes for protein synthesis. *BMC Biology*. 4 :30.
- Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7 : Improvements in Performance and Usability. *Molecular Biology and Evolution*. 30 :772–780.
- Kim BM, Rhee JS, Seo JS, Kim IC, Lee YM, Lee JS. 2012. 8-Oxoguanine DNA glycosylase 1 (OGG1) from the copepod *Tigriopus japonicus* : Molecular characterization and its expression in response to UV-B and heavy metals. *Comparative Biochemistry and Physiology Part C : Toxicology & Pharmacology*. 155 :290–299.
- Kimura M. 1968. Evolutionary rate at the molecular level. *Nature*. 217 :624–626.
- Kimura M. 1979. The Neutral Theory of Molecular Evolution. *Scientific American*. 241 :98–129.
- Kimura M. 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press.
- Kimura M, Ohta T. 1971. On the rate of molecular evolution. *Journal of Molecular Evolution*. 1 :1–17.

- Kong A, Frigge ML, Masson G, et al. (21 co-authors). 2012. Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature*. 488 :471–475.
- Kujoth GC, Bradshaw PC, Haroon S, Prolla TA. 2007. The Role of Mitochondrial DNA Mutations in Mammalian Aging. *PLoS Genetics*. 4 :161–173.
- Kędzierski M, Leszczyński S. 2013. A paleoceanographic model for the Late Campanian–Early Maastrichtian sedimentation in the Polish Carpathian Flysch basin based on nannofossils. *Marine Micropaleontology*. 102 :34–50.
- Laird CD, McCONAUGHY BL, McCARTHY BJ. 1969. Rate of Fixation of Nucleotide Substitutions in Evolution. *Nature*. 224 :149.
- Lancaster LT. 2010. Molecular evolutionary rates predict both extinction and speciation in temperate angiosperm lineages. *BMC Evolutionary Biology*. 10 :1–10.
- Lanfear R. 2018. Do plants have a segregated germline? *PLOS Biology*. 16 :e2005439.
- Lanfear R, Ho SYW, Jonathan Davies T, Moles AT, Aarssen L, Swenson NG, Warman L, Zanne AE, Allen AP. 2013. Taller plants have lower rates of molecular evolution. *Nature Communications*. 4 :1879.
- Lanfear R, Ho SYW, Love D, Bromham L. 2010. Mutation rate is linked to diversification in birds. *Proceedings of the National Academy of Sciences*. 107 :20423–20428.
- 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.



- Lartillot N, Delsuc F. 2011. Joint Reconstruction of Divergence Times and Life-History Evolution in Placental Mammals Using a Phylogenetic Covariance Model. *Evolution*. 66 :1773–1787.
- Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3 : a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*. 25 :2286–2288.
- Lartillot N, Poujol R. 2011. A Phylogenetic Model for Investigating Correlated Evolution of Substitution Rates and Continuous Phenotypic Characters. *Molecular Biology and Evolution*. 28 :729–744.
- 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.
- Lefébure T, Morvan C, Malard F, et al. (17 co-authors). 2017. Less effective selection leads to larger genomes. *Genome Research*. 27 :1016–1028.
- Leick V. 1968. Ratios between Contents of DNA, RNA and Protein in Different Micro-organisms as a Function of Maximal Growth Rate. *Nature*. 217 :1153–1155.
- Lepage T, Bryant D, Philippe H, Lartillot N. 2007. A General Comparison of Relaxed Molecular Clock Models. *Molecular Biology and Evolution*. 24 :2669–2680.
- Leys Remko, Cooper Steven JB, Strecker Ulrike, Wilkens Horst. 2005. Regressive evolution of an eye pigment gene in independently evolved eyeless subterranean diving beetles. *Biology Letters*. 1 :496–499.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 25 :1754–1760.

- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25 :2078–2079.
- Li WH. 1993. So, what about the molecular clock hypothesis? *Current Opinion in Genetics & Development*. 3 :896–901.
- Li WH, Ellsworth DL, Krushkal J, Chang BHJ, Hewett-Emmett D. 1996. Rates of Nucleotide Substitution in Primates and Rodents and the Generation–Time Effect Hypothesis. *Molecular Phylogenetics and Evolution*. 5 :182–187.
- Li WH, Tanimura M. 1987. The molecular clock runs more slowly in man than in apes and monkeys. *Nature*. 326 :93–96.
- Li WH, Tanimura M, Sharp PM. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *Journal of Molecular Evolution*. 25 :330–342.
- Li WH, Yi S, Makova K. 2002. Male-driven evolution. *Current Opinion in Genetics & Development*. 12 :650–656.
- Liang KY, Zeger SL. 1986. Longitudinal data analysis using generalized linear models. *Biometrika*. 73 :13–22.
- Lin G, Huang Z, Wang L, Chen Z, Zhang T, Gillman LN, Zhao F. 2019. Evolutionary Rates of Bumblebee Genomes Are Faster at Lower Elevations. *Molecular Biology and Evolution*. 36 :1215–1219.
- Lindhorst K, Krastel S, Reicherter K, Stipp M, Wagner B, Schwenk T. 2015. Sedimentary and tectonic evolution of Lake Ohrid (Macedonia/Albania). *Basin Research*. 27 :84–101.
- Lins LSF, Ho SYW, Wilson GDF, Lo N. 2012. Evidence for Permo-Triassic colonization of the deep sea by isopods. *Biology Letters*. 8 :979–982.

- Loseva O, Shubbar E, Haghdoost S, Evers B, Helleday T, Harms-Ringdahl M. 2014. Chronic Low Dose Rate Ionizing Radiation Exposure Induces Premature Senescence in Human Fibroblasts that Correlates with Up Regulation of Proteins Involved in Protection against Oxidative Stress. *Proteomes*. 2 :341–362.
- Lutzoni F, Pagel M. 1997. Accelerated evolution as a consequence of transitions to mutualism. *Proceedings of the National Academy of Sciences*. 94 :11422–11427.
- Lynch M. 2010a. Evolution of the mutation rate. *Trends in Genetics*. 26 :345–352.
- Lynch M. 2010b. Rate, molecular spectrum, and consequences of human mutation. *Proceedings of the National Academy of Sciences*. 107 :961–968.
- Löytynoja A, Goldman N. 2008. Phylogeny-Aware Gap Placement Prevents Errors in Sequence Alignment and Evolutionary Analysis. *Science*. 320 :1632–1635.
- Magniez G. 2008. From 1896 to 2008, a sight of the stygobitic family Stenasellidae (Crustacea, Isopoda, Asellota).
- Malard F, Grison P, Duchemin L, KonecnyDupré L, Lefébure T, Saclier N, Eme D, Martin C, Callou C, Douady CJ. 2019. GOTIT : a laboratory application software for optimizing multi-criteria species based research. *Methods in Ecology and Evolution*. 0.
- Malard F, Henry JP, Douady CJ. 2014. The scientific contribution of Guy Magniez (1935–2014). *Subterranean Biology*. 13 :55–64.
- Malard F, Hervant F. 1999. Oxygen supply and the adaptations of animals in groundwater. *Freshwater Biology*. 41 :1–30.
- Maliet O, Hartig F, Morlon H. 2019. A model with many small shifts for estimating species-specific diversification rates. *Nature Ecology & Evolution*. 3 :1086–1092.

- Mandavilli BS, Santos JH, Van Houten B. 2002. Mitochondrial DNA repair and aging. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 509 :127–151.
- Margoliash E. 1963. PRIMARY STRUCTURE AND EVOLUTION OF CYTOCHROME C. *Proceedings of the National Academy of Sciences of the United States of America*. 50 :672–679.
- Martin AP, Naylor GJP, Palumbi SR. 1992. Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. *Nature*. 357 :153.
- Martin AP, Palumbi SR. 1993. Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences*. 90 :4087–4091.
- Martincorena I, Luscombe NM. 2013. Non-random mutation : The evolution of targeted hypermutation and hypomutation. *BioEssays*. 35 :123–130.
- Martins EP, Hansen TF. 1997. Phylogenies and the Comparative Method : A General Approach to Incorporating Phylogenetic Information into the Analysis of Interspecific Data. *The American Naturalist*. 149 :646–667.
- Mathers EM, Houlihan DF, Cunningham MJ. 1992. Nucleic acid concentrations and enzyme activities as correlates of growth rate of the saithe *Pollachius virens* : growth-rate estimates of open-sea fish. *Marine Biology*. 112 :363–369.
- Matic I, Radman M, Taddei F, Picard B, Doit C, Bingen E, Denamur E, Elion J. 1997. Highly Variable Mutation Rates in Commensal and Pathogenic *Escherichia coli*. *Science*. 277 :1833–1834.
- Mayrose I, Otto SP. 2011. A Likelihood Method for Detecting Trait-Dependent Shifts in the Rate of Molecular Evolution. *Molecular Biology and Evolution*. 28 :759–770.

- McKee MJ, Knowles CO. 1987. Levels of protein, RNA, DNA, glycogen and lipid during growth and development of *Daphnia magna* Straus (Crustacea : Cladocera). *Freshwater Biology*. 18 :341–351.
- McLaughlin PJ, Dayhoff MO. 1970. Eukaryotes versus Prokaryotes : An Estimate of Evolutionary Distance. *Science*. 168 :1469–1471.
- Meulenkamp JE, Sissingh W. 2003. Tertiary palaeogeography and tectonostratigraphic evolution of the Northern and Southern Peri-Tethys platforms and the intermediate domains of the African–Eurasian convergent plate boundary zone. *Palaeogeography, Palaeoclimatology, Palaeoecology*. 196 :209–228.
- Mittelbach GG, Schemske DW, Cornell HV, et al. (22 co-authors). 2007. Evolution and the latitudinal diversity gradient : speciation, extinction and biogeography. *Ecology Letters*. 10 :315–331.
- Mooers AO, Harvey PH. 1994. Metabolic Rate, Generation Time, and the Rate of Molecular Evolution in Birds. *Molecular Phylogenetics and Evolution*. 3 :344–350.
- 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 :512–522.
- Muller HJ. 1942. Isolating mechanisms, evolution and temperature.
- Muse SV, Gaut BS. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Molecular Biology and Evolution*. 11 :715–724.
- Nabholz B, Glémin S, Galtier N. 2007. Strong Variations of Mitochondrial Mutation Rate across Mammals—the Longevity Hypothesis. *Molecular Biology and Evolution*. 25 :120–130.

- Nagylaki T. 1983. Evolution of a finite population under gene conversion. *Proceedings of the National Academy of Sciences*. 80 :6278–6281.
- Naya H, Romero H, Zavala A, Alvarez B, Musto H. 2002. Aerobiosis Increases the Genomic Guanine Plus Cytosine Content (GC%) in Prokaryotes. *Journal of Molecular Evolution*. 55 :260–264.
- Neidhardt FC, Magasanik B. 1960. Studies on the role of ribonucleic acid in the growth of bacteria. *Biochimica et Biophysica Acta*. 42 :99–116.
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE : A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*. 32 :268–274.
- Niemiller ML, Near TJ, Fitzpatrick BM. 2012. Delimiting Species Using Multilocus Data : Diagnosing Cryptic Diversity in the Southern Cavefish, *Typhlichthys Subterraneus* (teleostei : Amblyopsidae). *Evolution*. 66 :846–866.
- Nikolaev SI, Montoya-Burgos JI, Popadin K, Parand L, Margulies EH, Program NIOHISCCS, Antonarakis SE. 2007. Life-history traits drive the evolutionary rates of mammalian coding and noncoding genomic elements. *Proceedings of the National Academy of Sciences*. 104 :20443–20448.
- Notenboom J. 1988. Phylogenetic Relationships and Biogeography of the Groundwater-Dwelling Amphipod Genus *Pseudoniphargus* (Crustacea), with Emphasis on the Iberian Species. *Bijdragen tot de Dierkunde*. 58 :159–204.
- Noutahi E, Semeria M, Lafond M, Seguin J, Boussau B, Guéguen L, El-Mabrouk N, Tannier E. 2016. Efficient Gene Tree Correction Guided by Genome Evolution. *PLOS ONE*. 11 :e0159559.
- Nöthel H. 1987. Adaptation of *Drosophila melanogaster* populations to high mutation pressure : evolutionary adjustment of mutation rates. *Proceedings of the National Academy of Sciences*. 84 :1045–1049.

- O'Connor TD, Mundy NI. 2009. Genotype–phenotype associations : substitution models to detect evolutionary associations between phenotypic variables and genotypic evolutionary rate. *Bioinformatics*. 25 :i94–i100.
- Ohta T. 1973. Slightly Deleterious Mutant Substitutions in Evolution. *Nature*. 246 :96–98.
- Ohta T. 1987. Very slightly deleterious mutations and the molecular clock. *Journal of Molecular Evolution*. 26 :1–6.
- Ohta T. 1993. An examination of the generation-time effect on molecular evolution. *Proceedings of the National Academy of Sciences*. 90 :10676–10680.
- Oliver A, Baquero F, Blázquez J. 2002. The mismatch repair system (mutS, mutL and uvrD genes) in *Pseudomonas aeruginosa* : molecular characterization of naturally occurring mutants. *Molecular Microbiology*. 43 :1641–1650.
- on the Effects of Atomic Radiation UNSC. 2000. Sources and Effects of Ionizing Radiation : Sources. United Nations Publications. Google-Books-ID : 57YmhD4ZBpQC.
- Orsini JB, COULON C, COCOZZA T. 1980. DERIVE (LA) CENOZOIQUE DE LA CORSE ET DE LA SARDAIGNE.-. *Géologie alpine*. pp. PP., 169–202, FIG., CARTES, TABL., BIBLIOGR.
- Orton MG, May JA, Ly W, Lee DJ, Adamowicz SJ. 2019. Is molecular evolution faster in the tropics? *Heredity*. 122 :513–524.
- Orue-Etxebarria X, Pujalte V, Bernaola G, Apellaniz E, Baceta JI, Payros A, Nuñez-Betelu K, Serra-Kiel J, Tosquella J. 2001. Did the Late Paleocene thermal maximum affect the evolution of larger foraminifers? Evidence from calcareous plankton of the Campo Section (Pyrenees, Spain). *Marine Micropaleontology*. 41 :45–71.

- Pagel M, Venditti C, Meade A. 2006. Large Punctuational Contribution of Speciation to Evolutionary Divergence at the Molecular Level. *Science*. 314 :119–121.
- Palmeira L, Guéguen L, Lobry JR. 2006. UV-Targeted Dinucleotides Are Not Depleted in Light-Exposed Prokaryotic Genomes. *Molecular Biology and Evolution*. 23 :2214–2219.
- Paula WBMD, Lucas CH, Agip ANA, Vizcay-Barrena G, Allen JF. 2013. Energy, ageing, fidelity and sex : oocyte mitochondrial DNA as a protected genetic template. *Phil. Trans. R. Soc. B*. 368 :20120263.
- Payros A, Orue-Etxebarria X, Pujalte V. 2006. Covarying sedimentary and biotic fluctuations in Lower–Middle Eocene Pyrenean deep-sea deposits : Palaeoenvironmental implications. *Palaeogeography, Palaeoclimatology, Palaeoecology*. 234 :258–276.
- Peck SB, Finston TL. 1993. Galapagos islands troglobites : the questions of tropical troglobites, parapatric distributions with eyed-sister-species, and their origin by parapatric speciation. *Mémoires de Biospéléologie*. pp. 19–37.
- Peragón J, Barroso JB, Garcia-Salguero L, de la Higuera M, Lupiáñez JA. 2001. Growth, protein-turnover rates and nucleic-acid concentrations in the white muscle of rainbow trout during development. *The International Journal of Biochemistry & Cell Biology*. 33 :1227–1238.
- Petit C, Déverchère J. 2006. Structure and evolution of the Baikal rift : A synthesis. *Geochemistry, Geophysics, Geosystems*. 7.
- Petit RJ, Hampe A. 2006. Some Evolutionary Consequences of Being a Tree. *Annual Review of Ecology, Evolution, and Systematics*. 37 :187–214.
- Pfeiffer P, Goedecke W, Obe G. 2000. Mechanisms of DNA double-strand break repair and their potential to induce chromosomal aberrations. *Mutagenesis*. 15 :289–302.



- Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Wörheide G, Baurain D. 2011. Resolving Difficult Phylogenetic Questions : Why More Sequences Are Not Enough. *PLOS Biology*. 9 :e1000602.
- Phillips MJ, Page TJ, de Bruyn M, Huey JA, Humphreys WF, Hughes JM, Santos SR, Schmidt DJ, Waters JM. 2013. The linking of plate tectonics and evolutionary divergence. *Current Biology*. 23 :R603–R605.
- Pierce GJ, Key LN, Boyle PR, Siegert KJ, Gonçalves JM, Porteiro FM, Martins HR. 1999. RNA concentration and the RNA to protein ratio in cephalopod tissues : sources of variation and relationship with growth rate. *Journal of Experimental Marine Biology and Ecology*. 237 :185–201.
- Piganeau G, Eyre-Walker A. 2009. Evidence for Variation in the Effective Population Size of Animal Mitochondrial DNA. *PLOS ONE*. 4 :e4396.
- Plaziat JC. 1981. Late cretaceous to late eocene palaeogeographic evolution of southwest Europe. *Palaeogeography, Palaeoclimatology, Palaeoecology*. 36 :263–320.
- Posada D. 2008. jModelTest : Phylogenetic Model Averaging. *Molecular Biology and Evolution*. 25 :1253–1256.
- Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER web server : 2018 update. *Nucleic Acids Research*. 46 :W200–W204.
- Poulson TL. 1963. Cave Adaptation in Amblyopsid Fishes. *The American Midland Naturalist*. 70 :257–290.
- Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE*. 5 :e9490.
- Pujalte V, Robles S, Orue-Etxebarria X, Baceta JI, Payros A, Larruzea I. 2000. Uppermost Cretaceous–middle Eocene strata of the Basque-Cantabrian region

- and western Pyrenees : a sequence stratigraphic perspective. *Revista de la Sociedad Geológica de España*. 13(2) :191–211.
- Pöckl M. 1993. Reproductive potential and lifetime potential fecundity of the freshwater amphipods *Gammarus fossarum* and *G. roeseli* in Austrian streams and rivers. *Freshwater Biology*. 30 :73–91.
- Richter C, Park JW, Ames BN. 1988. Normal oxidative damage to mitochondrial and nuclear DNA is extensive. *Proceedings of the National Academy of Sciences*. 85 :6465–6467.
- Rohde K. 1992. Latitudinal Gradients in Species Diversity : The Search for the Primary Cause. *Oikos*. 65 :514–527.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3 : Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 19 :1572–1574.
- Roots R, Okada S. 1975. Estimation of Life Times and Diffusion Distances of Radicals Involved in X-Ray-Induced DNA Strand Breaks or Killing of Mammalian Cells. *Radiation Research*. 64 :306–320.
- Saclier N, François CM, 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 :2900–2912.
- Sarich VM, Wilson AC. 1967. Immunological Time Scale for Hominid Evolution. *Science*. 158 :1200–1203.
- Sasani TA, Pedersen BS, Gao Z, Baird L, Przeworski M, Jorde LB, Quinlan AR. 2019. Large, three-generation human families reveal post-zygotic mosaicism and variability in germline mutation accumulation. *eLife*. 8 :e46922.

- Satoh C, Takahashi N, Asakawa J, Kodaira M, Kuick R, Hanash S M, Neel J V. 1996. Genetic analysis of children of atomic bomb survivors. *Environmental Health Perspectives*. 104 :511–519.
- Schaaper RM. 1993. Base selection, proofreading, and mismatch repair during DNA replication in *Escherichia coli*. *Journal of Biological Chemistry*. 268 :23762–23765.
- Schlotteröer C, Amos B, Tautz D. 1991. Conservation of polymorphic simple sequence loci in cetacean species. *Nature*. 354 :63–65.
- Schmidt TR, Wu W, Goodman M, Grossman LI. 2001. Evolution of Nuclear- and Mitochondrial-Encoded Subunit Interaction in Cytochrome c Oxidase. *Molecular Biology and Evolution*. 18 :563–569.
- Setlow RB, Carrier WL. 1966. Pyrimidine dimers in ultraviolet-irradiated DNA's. *Journal of Molecular Biology*. 17 :237–254.
- Shanahan EM, Peterson D, Roxby D, Quintana J, Morely AA, Woodward A. 1996. Mutation rates at the glycophorin A and HPRT loci in uranium miners exposed to radon progeny. *Occupational and Environmental Medicine*. 53 :439–444.
- 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.
- Skogseid J, Planke S, Faleide JI, Pedersen T, Eldholm O, Neverdal F. 2000. NE Atlantic continental rifting and volcanic margin formation. *Geological Society, London, Special Publications*. 167 :295–326.
- Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*. 6 :1–11.

- Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD, Fujimoto EK, Goeke NM, Olson BJ, Klenk DC. 1985. Measurement of protein using bicinchoninic acid. *Analytical Biochemistry*. 150 :76–85.
- Smith SA, Donoghue MJ. 2008. Rates of Molecular Evolution Are Linked to Life History in Flowering Plants. *Science*. 322 :86–89.
- Somers CM, Yauk CL, White PA, Parfett CLJ, Quinn JS. 2002. Air pollution induces heritable DNA mutations. *Proceedings of the National Academy of Sciences*. 99 :15904–15907.
- Stamatakis A. 2014. RAxML version 8 : a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30 :1312–1313.
- Stampfli GM. 2000. Tethyan oceans. *Geological Society, London, Special Publications*. 173 :1–23.
- Stelbrink B, Shirokaya AA, Föller K, Wilke T, Albrecht C. 2016. Origin and diversification of Lake Ohrid's endemic acroloxid limpets : the role of geography and ecology. *BMC Evolutionary Biology*. 16 :1–13.
- Stern DB, Breinholt J, PedrazaLara C, LópezMejía M, Owen CL, BrackenGrissom H, Fetzner JW, Crandall KA. 2017. Phylogenetic evidence from freshwater crayfishes that cave adaptation is not an evolutionary dead-end. *Evolution*. 71 :2522–2532.
- Taylor SD, Cruz KDdl, Porter ML, Whiting MF. 2005. Characterization of the Long-Wavelength Opsin from Mecoptera and Siphonaptera : Does a Flea See ? *Molecular Biology and Evolution*. 22 :1165–1174.
- Thomas JA, Welch JJ, Lanfear R, Bromham L. 2010. A generation time effect on the rate of molecular evolution in invertebrates. *Molecular Biology and Evolution*. p. msq009.

- Thompson JD, Gibson TJ, Higgins DG. 2003. Multiple Sequence Alignment Using ClustalW and ClustalX. *Current Protocols in Bioinformatics*. 00 :2.3.1–2.3.22.
- Tkeshelashvili LK, Shearman CW, Zakour RA, Koplitz RM, Loeb LA. 1980. Effects of Arsenic, Selenium, and Chromium on the Fidelity of DNA Synthesis. *Cancer Research*. 40 :2455–2460.
- Touaylia S, Garrido J, Bejaoui M, Boumaiza M. 2011. Altitudinal Distribution of Aquatic Beetles (Coleoptera) in Northern Tunisia : Relationship between Species Richness and Altitude. *The Coleopterists Bulletin*. 65 :53–62.
- Trabucco a, Zomer RJ. 2019. Global High-Resolution Soil-Water Balance.
- Tyson RV, Funnell BM. 1987. European cretaceous shorelines, stage by stage. *Palaeogeography, Palaeoclimatology, Palaeoecology*. 59 :69–91.
- Van Valen L. 1973. A new evolutionary law. *Evol Theory*. 1 :1–30.
- Vandel AA. 1964. Biospéologie : la biologie des animaux cavernicoles. Gauthier-Villars éditeur. Paris.
- Venditti C, Pagel M. 2010. Speciation as an active force in promoting genetic evolution. *Trends in Ecology & Evolution*. 25 :14–20.
- Wagner B, Wilke T, Francke A, Albrecht C, Baumgarten H, Bertini A, Combourieu-Nebout N, Cvetkoska A, D'Addabbo M, Donders TH. 2017. *The environmental and evolutionary history of Lake Ohrid (FYROM/Albania) : interim results from the SCOPSCO deep drilling project*. p. 24.
- Wagner B, Wilke T, Krastel S, et al. (19 co-authors). 2014. The SCOPSCO drilling project recovers more than 1.2 million years of history from Lake Ohrid. *Scientific Drilling*. 17 :19–29.

- Wagner MM, Campbell RG, Boudreau CA, Durbin EG. 2001. Nucleic acids and growth of *Calanus finmarchicus* in the laboratory under different food and temperature conditions. *Marine Ecology Progress Series*. 221 :185–197.
- Wang L, Ji Y, Hu Y, et al. (16 co-authors). 2019. The architecture of intra-organism mutation rate variation in plants. *PLOS Biology*. 17 :e3000191.
- Webster AJ, Payne RJH, Pagel M. 2003. Molecular Phylogenies Link Rates of Evolution and Speciation. *Science*. 301 :478–478.
- Weir JT, Schluter D. 2011. Are rates of molecular evolution in mammals substantially accelerated in warmer environments? *Proceedings of the Royal Society B : Biological Sciences*. 278 :1291–1293.
- Welch JJ, Waxman D. 2008. Calculating independent contrasts for the comparative study of substitution rates. *Journal of Theoretical Biology*. 251 :667–678.
- White DS. 1993. Perspectives on Defining and Delineating Hyporheic Zones. *Journal of the North American Benthological Society*. 12 :61–69.
- Wojewodzic MW, Rachamim T, Andersen T, Leinaas HP, Hessen DO. 2011. Effect of temperature and dietary elemental composition on RNA/protein ratio in a rotifer. *Functional Ecology*. 25 :1154–1160.
- Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences*. 84 :9054–9058.
- Wolfe KH, Sharp PM, Li WH. 1989. Mutation rates differ among regions of the mammalian genome. *Nature*. 337 :283.
- Wolff JN, Ladoukakis ED, Enríquez JA, Dowling DK. 2014. Mitonuclear interactions : evolutionary consequences over multiple biological scales. *Philosophical Transactions of the Royal Society B : Biological Sciences*. 369 :20130443.

- Wright F. 1990. The 'effective number of codons' used in a gene. *Gene*. 87 :23–29.
- Wright S, Keeling J, Gillman L. 2006. The road from Santa Rosalia : A faster tempo of evolution in tropical climates. *Proceedings of the National Academy of Sciences*. 103 :7718–7722.
- Wright SD, Gillman LN, Ross HA, Keeling DJ. 2010. Energy and the tempo of evolution in amphibians. *Global Ecology and Biogeography*. 19 :733–740.
- Wysocka A, Kostoski G, Kilikowska A, Wróbel B, Sell J. 2008. The Proasellus (Crustacea, Isopoda) species group, endemic to the Balkan Lake Ohrid : a case of ecological diversification? *Fundamental and Applied Limnology / Archiv für Hydrobiologie*. 172 :301–313.
- Xing W, Shi Q, Liu H, Liu G. 2016. Growth rate, protein :RNA ratio and stoichiometric homeostasis of submerged macrophytes under eutrophication stress. *Knowledge and Management of Aquatic Ecosystems*. p. 25.
- Yang Z. 2007. PAML 4 : Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*. 24 :1586–1591.
- Zhang DX, Hewitt GM. 1997. Insect mitochondrial control region : A review of its structure, evolution and usefulness in evolutionary studies. *Biochemical Systematics and Ecology*. 25 :99–120.
- Zhang J, Kapli P, Pavlidis P, Stamatakis A. 2013. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*. 29 :2869–2876.
- Ziegler A, Leffell DJ, Kunala S, Sharma HW, Gailani M, Simon JA, Halperin AJ, Baden HP, Shapiro PE, Bale AE. 1993. Mutation hotspots due to sunlight in the p53 gene of nonmelanoma skin cancers. *Proceedings of the National Academy of Sciences*. 90 :4216–4220.

Zigler KS, Cooper GM. 2011. *Brood size of the stygobiotic asellid isopod Caecidotea bicrenata bicrenata from Franklin County, Tennessee, USA*. p. 3.

Zuckerkindl E, Pauling L. 1962. Molecular disease, evolution, and genetic heterogeneity. *Horizons in biochemistry*. pp. 189–225.

Zuckerkindl E, Pauling L. 1965. Evolutionary Divergence and Convergence in Proteins. In : *Evolving Genes and Proteins*, Elsevier, pp. 97–166.





# Annexes

## Annexe 1 :

### Influence de la transition vers le milieu souterrain sur la stratégie reproductive chez les Asellidae

Nathanaëlle Saclier, Colin Issartel, Florian Malard

#### Introduction

A partir de l'observation de que les organismes souterrains produisent des œufs plus gros mais en moins grand nombre, Vandel (1964) définit une « loi » selon laquelle les organismes souterrains auraient le même investissement reproducteur que les espèces de surface mais avec un changement dans la stratégie reproductive : la réduction du nombre d'œufs serait ainsi parfaitement corrélée à l'augmentation de leur taille.

Des observations de changements de la taille des œufs ont en effet été observée chez de nombreux animaux souterrains tels que des téléostéens (Poulson, 1963), ou des malacostracés (Henry, 1976; Zigler et Cooper, 2011; Fišer et al., 2013). Cependant, ces observations se réduisent souvent à une comparaison visuelle, sans mesure de la taille des œufs ou de la taille des femelles. Les observations plus poussées de ce changement de taille et de nombre d'œufs (*e.g.* Henry, 1976) n'ont, quant à elle, jamais été faites dans un cadre comparatif, ne permettant ainsi pas de prendre en compte l'inertie phylogénétique.

Afin de tester cette hypothèse de changements de stratégie de reproduction en lien avec une transition vers le milieu souterrain, nous proposons d'utiliser les caractéristiques originales de la famille des Asellidae (Rafinesque, 1815), qui présente de nombreux avantages. En effet, elle comprend des espèces épigées, des espèces exclusivement hypogées, ainsi que des espèces « intermédiaires », vivant en général en milieu souterrain mais qui ne présentent pas encore tous les traits troglomorphiques caractéristiques (dépigmentation, anophtalmie...). Il a été mon-

tré qu'au sein de cette famille, plusieurs espèces ont colonisé indépendamment le milieu souterrain (Francois et al., 2016a; Lefébure et al., 2017; Saclier et al., 2018), nous permettant d'avoir des répliqués de la même transition. De plus, chez les espèces de cette famille, les femelles portent les œufs à l'intérieur d'une poche incubatrice, ce qui permet de faire le lien entre les œufs et les caractéristiques de la femelle qui les a produits, et en particulier sa taille.

Le nombre et de la taille des œufs ont été comparés entre des espèces de surface (oculées et pigmentées), des espèces « intermédiaires » (microphthalmes et partiellement dépigmentées), et des espèces strictement souterraines (anophthalmes et dépigmentées)

Cependant, les espèces de surface subissant des variations saisonnières qui sont quasiment absentes en milieu souterrain, une variation saisonnière de la taille et/ou du nombre d'œufs chez les espèces de surface pourrait entraîner un biais potentiel. Il a été montré chez plusieurs espèces de Gammaridae (Latreille, 1802) une variation saisonnière de la taille des œufs ainsi que de leur nombre (Pöckl, 1993; Foucreau, 2013), c'est pourquoi l'effet de la saisonnalité sur ces paramètres sera étudié au sein d'une population de surface.

## Matériel et méthodes

**Échantillonnage** Les animaux prélevés sont stockés dans l'éthanol à 96% puis ramenés au laboratoire où ils sont triés. Les femelles ovigères d'Asellidae sont mises à l'écart et stockées dans des tubes remplis d'éthanol à 96%. Chaque tube de femelles ovigères se voit attribuer un code mentionnant le nom de code de la station de prélèvement, la date, le nom de l'espèce et le nombre de femelles ovigères.

L'étude de la variabilité saisonnière du nombre d'œufs et de leur taille est réalisée sur une population d'*Asellus aquaticus* (Linnaeus, 1758) de surface, localisée sur l'espace expérimental «Les étangs» situé sur le campus de la Doua, à Villeurbanne. Deux séries de prélèvements ont été réalisées : une première en mars 2018, une seconde en octobre 2018.

Enfin, la variabilité interspécifique a été étudiée sur un total de 17 espèces dont tous les individus ont été prélevés en Europe (Table 1)

**Dissection** Les pattes de la femelle ovigère sont retirées à l'aide de pinces. La femelle est déposée sur une lame contenant une petite goutte d'eau. Deux photographies de la femelle ont été prises : une de l'animal entier, une autre focalisée sur le céphalon. Pour quelques individus, des photographies de la poche contenant les œufs ont aussi été prises avant d'avoir été ouvertes. La femelle est ensuite placée sur une lame sur laquelle ont été déposées quelques gouttes de glycérine. Les œufs sont extraits de la poche incubatrice sous un stéréomicroscope Olympus SZX16, à l'aide d'aiguilles très fines et d'une aiguille dont l'extrémité a été recourbée. Les œufs extraits et la femelle sont photographiés sous le stéréomicroscope à l'aide

TABLE 1 – Liste des espèces utilisées pour étudier l’effet de la transition écologique vers le milieu souterrain au niveau inter-spécifique, et leur milieu de vie.

Espèce	Écologie
<i>Asellus aquaticus</i>	Surface
<i>Asellus aquaticus</i>	Surface
<i>Proasellus cavaticus</i>	Souterrain
<i>Proasellus cavaticus</i>	Souterrain
<i>Proasellus beticus</i>	Surface
<i>Proasellus coiffaiti</i>	Surface
<i>Proasellus albigensis</i>	Souterrain
<i>Proasellus coxalis</i>	Surface
<i>Proasellus aragonensis</i>	Intermédiaire
<i>Proasellus ibericus</i>	Surface
<i>Proasellus escolai</i>	Souterrain
<i>Proasellus granadensis</i>	Intermédiaire
<i>Proasellus valdensis</i>	Souterrain
<i>Proasellus hercegovinensis</i>	Souterrain
<i>Proasellus meridianus</i>	Surface
<i>Proasellus ortizi</i>	Intermédiaire
<i>Proasellus jaloniacus</i>	Souterrain

d’une caméra (Olympus, modèle DP25) et du logiciel CellD (version 5.1). Après avoir été extraits, les œufs contenus dans la glycérine sont placés entre lame et lamelle. Des petits morceaux de pâte à modeler ont préalablement été déposés aux quatre angles de la lamelle, dans le but de la surélever et de ne pas écraser les œufs lorsque la lamelle est posée sur la lame. Les œufs sont photographiés directement après. Chez certains individus, les œufs contenus dans la poche étaient complètement agglomérés entre eux, rendant difficile leur séparation sans les abîmer. Pour remédier à ce problème, le paquet d’œufs a été extrait de la poche puis et placé dans un tube rempli d’eau, qui a ensuite été agité pendant quelques minutes. Les lames et les femelles disséquées sont conservés dans les collections de zoologie du LEHNA. Au total, 321 individus ont été disséqués. La dissection d’une femelle ovigère et la prise des photographies prend en moyenne 40 minutes pour un manipulateur expérimenté, cette durée pouvant varier fortement selon le nombre d’œufs produits (une grande quantité d’œufs nécessitera un grand nombre de photographies d’œufs). Toutes les dissections ont été effectuées par Colin Issartel dans le cadre de son contrat d’alternance.

### Mesures morphométriques, mesures du nombre et de la taille des œufs

Toutes les mesures ont été réalisées avec le logiciel ImageJ (version 1.51j), à partir des photographies prises sous le stéréomicroscope. Deux mesures ont été effectuées : le nombre d’œufs produits par chaque femelle, et la surface visible (en deux dimensions) de chaque œuf. Deux paramètres ont été mesurés pour évaluer la taille de la femelle : la longueur du corps et la largeur de la capsule céphalique (Figure 5).

Pour mesurer la taille des oeufs, chaque photographie est convertie en image en niveaux de gris, elle-même convertie ensuite en image binaire (uniquement en

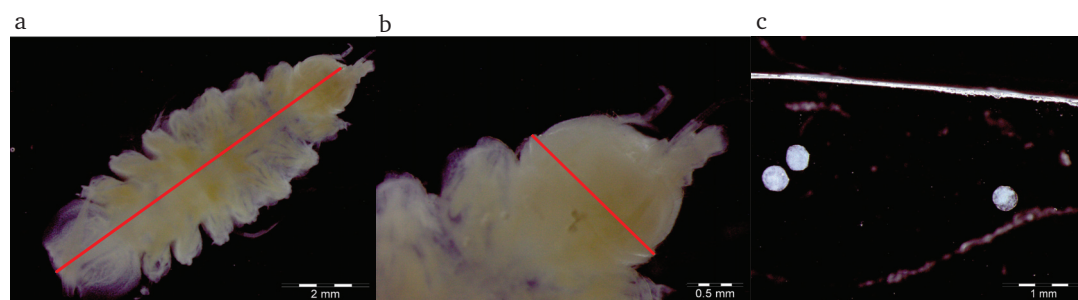


FIGURE 5 – Mesures effectuées pour chaque individu : longueur du corps (a), largeur du céphalon (b) et superficie de chaque œufs (c). *Asellus aquaticus* souterrain.

noir et blanc), ceci afin de «nettoyer» l'image (c'est-à-dire enlever les éléments qui ne seront pas mesurés) et de mettre un seuil à l'image, ajusté de sorte à détecter les œufs dont la surface doit être mesurée.

### Tests statistiques

**Étude de l'effet de la taille du corps** Afin de déterminer l'effet de la taille du corps sur le nombre et la taille des œufs, des modèles linéaires ont été utilisés entre la taille et le nombre d'œufs avec les différentes mesures de la femelle. Une corrélation entre la taille de la femelle et la taille ou le nombre d'œufs peut générer des corrélations entre l'habitat (souterrain-surface) et la taille ou le nombre d'œufs si la taille des femelles change avec le type d'habitat. Cette relation a été testée par un test de Kruskal-Wallis, les données n'étant pas distribuée normalement.

**Étude de la variabilité saisonnière** Afin de déterminer un effet de la saison sur la taille et le nombre d'œufs, une comparaison de ces paramètres entre mars et octobre 2018 a été effectuée. Les nombres moyens d'œufs et la taille des œufs entre mars et octobre ont été comparés statistiquement à l'aide d'un test de Student.

**Étude de la variabilité inter-spécifique** Des différences dans le nombre ou la taille des œufs entre les espèces pigmentées, partiellement dépigmentées et dépigmentées, ont été testées par un PGLS en utilisant des packages ape et nlme du logiciel R.

### Résultats

**Étude de l'effet de la taille du corps** On voit une corrélation positive entre le nombre d'œufs et la longueur du corps ( $\text{lm}$ ,  $p < 2,26\text{e-}16$ ,  $R^2 = 0,724$ ), la largeur du céphalon ( $\text{lm}$ ,  $p < 2,26\text{e-}16$ ,  $R^2 = 0,7373$ ), ainsi la surface du corps ( $\text{lm}$ ,  $p < 2,26\text{e-}16$ ,  $R^2 = 0,7853$ , Figure 6). Il existe également une corrélation positive

entre la taille des œufs et la longueur du corps ( $lm$ ,  $p = 1,251e-8, R^2=0,161$ ), la largeur du céphalon ( $lm$ ,  $p = 4,477e-7, R^2=0,1782$ ) et la surface de l'individu ( $lm$ ,  $p = 2,754e-7, R^2=0,86$ ) (Figure 7). On voit donc que le nombre d'œufs est en très grande partie expliqué par la surface de la femelle ( $R^2=0,86$ ), La taille des œufs dépend également de la taille de la femelle mais dans une moindre mesure. Il est donc nécessaire de prendre en compte la taille de la femelle pour comprendre l'évolution de la taille des œufs. Le paramètre qui explique le mieux la taille et le nombre d'œufs étant la surface de la femelle (longueur \* largeur du céphalon), c'est cette mesure qui sera utilisée par la suite.

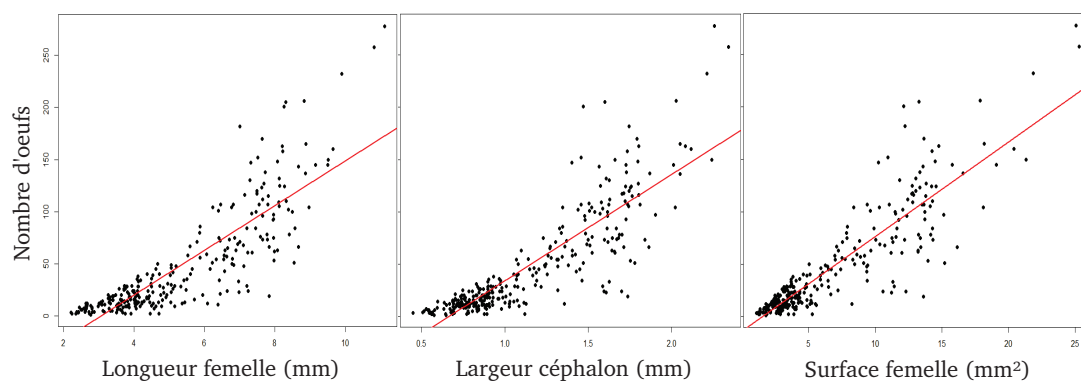


FIGURE 6 – Nombre d'œufs en fonction de longueur du corps (gauche), largeur du céphalon (centre) et superficie de la femelle (droite).

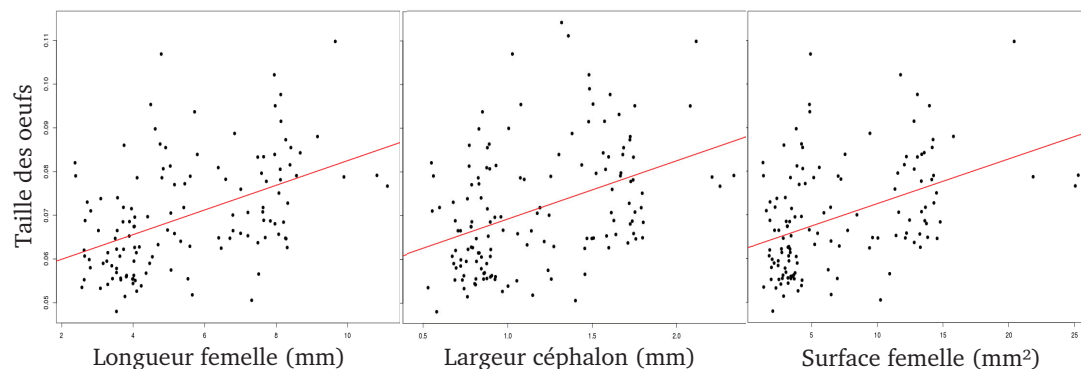


FIGURE 7 – Taille des œufs en fonction de longueur du corps (gauche), largeur du céphalon (centre) et superficie de la femelle (droite).

Afin d'évaluer si la corrélation avec la taille de la femelle peut créer la relation entre la taille et le nombre d'œufs et l'habitat, la relation entre la taille de la femelle et l'habitat (catégorisé en fonction du phénotype de la femelle : pigmentés, partiellement dépigmentés et dépigmentés) a été testée. La surface du corps de la femelle change bien en fonction du phénotype de la femelle (Kruskal-Wallis,  $p = 1,95e-9$ ) : les individus pigmentés une surface du corps moyenne significativement

supérieure à celle des individus partiellement dépigmentés et des individus dépigmentés, en revanche ces deux dernière catégories ne sont pas différentes entre elles (Figure 8).

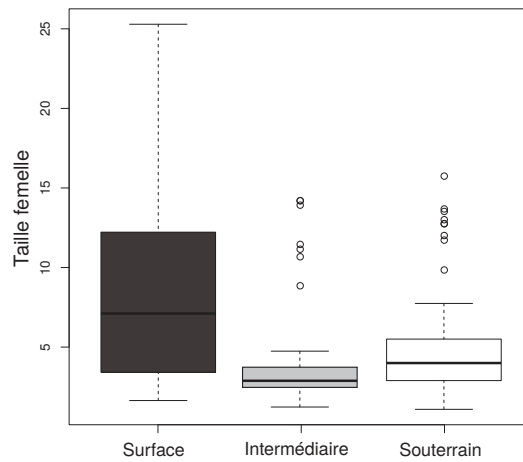


FIGURE 8 – Taille des femelles (surface en  $\text{mm}^2$  en fonction de l'habitat).

**Étude de la variabilité saisonnière** Le nombre d'œufs produits par les femelles en mars n'est pas significativement différent du nombre d'œufs produits par les femelles en octobre (test de Student,  $p = 0,6516$ ). En revanche, la taille des œufs produits en octobre est significativement supérieure à la taille des œufs produits en mars (test t de Welch,  $p = 0,002397$ ) (Figure 9).

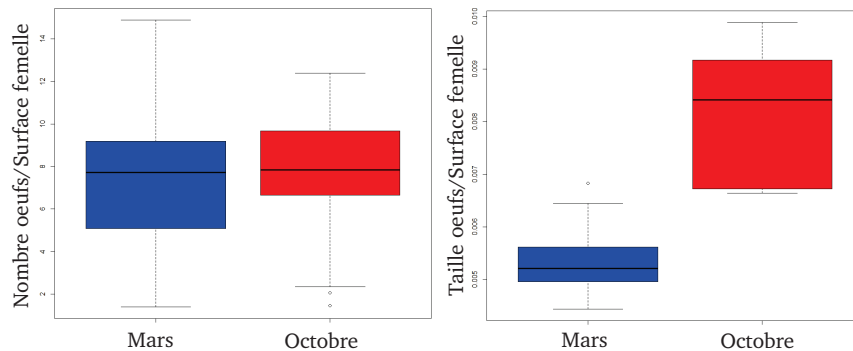


FIGURE 9 – Nombre (gauche) et taille (droite) des œufs en fonction de la saison (mars ou octobre) chez *Asellus aquaticus*

**Étude de la variabilité inter-spécifique** Le nombre moyen d'œufs est significativement plus élevé chez les espèces pigmentées que chez les espèces dépigmentées (PGLS,  $p.\text{value} = 0,0076$ ) ; en revanche, il n'y a pas de différence significative entre le nombre moyen d'œufs produit par les espèces dépigmentées et le nombre moyen d'œufs produit par les espèces partiellement dépigmentées (PGLS,  $p = 0,5746$ ) (Figure 10).

En ce qui concerne la taille des œufs, les résultats n'indiquent pas de différence significative entre la taille moyenne des œufs produits par les espèces pigmentées et les espèces dépigmentées (PGLS,  $p$ .value = 0,6332); la taille moyenne des œufs produits par les espèces partiellement dépigmentées n'est pas non plus significativement différente de la taille moyenne des œufs produits par les espèces dépigmentées (PGLS,  $p$ .value = 0,0735).

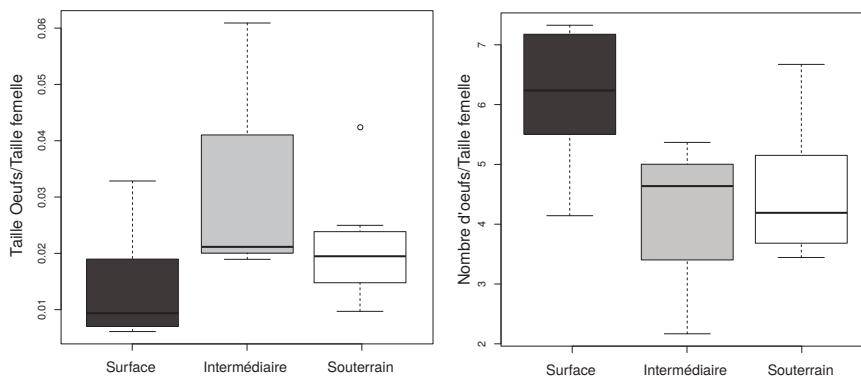


FIGURE 10 – Nombre (gauche) et taille (droite) des œufs en fonction de la saison (mars ou octobre) chez *Asellus aquaticus*

## Discussion

Cette étude nous indique premièrement que le premier facteur qui influence le nombre des œufs, et dans une moindre mesure la taille, est la taille de la femelle qui les porte. D'autre part la taille de cette femelle est significativement différente entre l'habitat souterrain et l'habitat de surface. En effet les femelles souterraines sont plus petite que les espèces de surface. D'autre part, la mesure de la largeur et celle de la longueur n'explique pas exactement la même part de variance. Ceci est du au fait que les femelle souterraines sont plus allongées que les espèces de surface. Le nombre et la taille des œufs étant mieux expliquée par la surface que par la longueur, on peut en déduire que la forme de la femelle impacte les œufs qu'elle porte. Ainsi le paramètre qui pourrait limiter la taille et le nombre d'œufs pourrait être la place disponible dans la poche incubatrice. Ainsi, une corrélation entre la taille ou le nombre d'œufs pourrait provenir uniquement des changement dans la taille et la forme de la femelle qui change en milieu souterrain. Ces paramètres n'ayant pas été pris en compte dans les précédentes observations, cela remettrait potentiellement en cause les conclusions.

L'étude la variabilité saisonnière nous indiquent que la taille des œufs varie avec la saison. Les femelles d'octobre, produisent des œufs plus gros qu'en mars. On peut supposer que les femelles de mars sont des femelles qui ont passées l'hiver dans les sédiments (leur grande taille permet de supposer qu'elle ne sont pas née à la sortie de l'hiver). En revanche les femelles d'octobre sont nées au printemps et ont connu une période très faste pendant l'été. Ainsi, on peut supposer que les






femelles d'octobre ont plus de réserves énergétique à allouer à la production. Elles semblent allouer cette réserve à la taille de œufs et non au nombre, peut-être parce que le nombre de cellules germinales est fixé ou parce que ce nombre ne peut pas être régulé par la quantité de ressources.

La comparaison surface souterrain nous a permis de montrer que les femelles souterraines produisent moins d'œufs. En revanche, ils ne sont pas plus gros. L'habitat souterrain étant un habitat pauvre en énergie, on ce serait attendu, étant donné les résultats précédents à voir un patron sur la taille et non sur le nombre d'œufs. Cependant il y a une différence temporelle importante entre les deux comparaisons. Dans le cas des espèces souterraines elles sont constamment dans un milieu pauvre en énergie, ainsi si le nombre d'œufs est fixé génétiquement, les pressions de sélection ont pu amener à des individus qui produisent moins d'œufs, un habitat pauvre en énergie ne permettant pas d'accumuler assez d'énergie pour produire un grand nombre d'œufs. Ainsi dans cet habitat, les femelles ayant moins de descendants auraient été favorisées car les portées auraient survécues alors que les femelles ayant un grand nombre d'œufs n'auraient pas pu les nourrir. La faible variabilité de la taille des œufs au sein d'une portée suggère que la même énergie est allouée à chaque œufs. La femelle ne serait ainsi pas en capacité de favorisé une partie des œufs au dépend des autres. Ainsi, un manque de nourriture pourrait amener à l'avortement de toute la portée. Ces hypothèses restant toutefois à tester.

**Annexe 2 :**

**Article : GOTIT : a laboratory application software for optimizing multicriteria species based research**

# GOTIT: A laboratory application software for optimizing multi-criteria species-based research

Florian Malard<sup>1</sup>  | Philippe Grison<sup>2</sup> | Louis Duchemin<sup>1</sup> | Lara Konecny-Dupré<sup>1</sup> |  
Tristan Lefébure<sup>1</sup> | Nathanaëlle Saclier<sup>1</sup> | David Eme<sup>3</sup>  | Chloé Martin<sup>2</sup> |  
Cécile Callou<sup>2</sup> | Christophe J. Douady<sup>1,4</sup> 

<sup>1</sup>UMR5023 Ecologie des Hydrosystèmes Naturels et Anthropisés, Univ. Lyon 1, ENTPE, CNRS, Univ. de Lyon, Villeurbanne, France

<sup>2</sup>BBEES, Unité Bases de données sur la Biodiversité, Ecologie, Environnement et Sociétés, Muséum National d'Histoire Naturelle, CNRS, Paris, France

<sup>3</sup>New Zealand Institute for Advanced Studies, School of Natural and Computational Sciences, Massey University, Auckland, New Zealand

<sup>4</sup>Institut Universitaire de France, Paris, France

## Correspondence

Florian Malard  
Email: florian.malard@univ-lyon1.fr

## Funding information

Centre National de la Recherche Scientifique, Grant/Award Number: STYGOMICS (CNRS Défi Enviromics); Royal Society of New Zealand, Grant/Award Number: Marsden Grant 15-MAU-132; European Cooperation in Science and Technology, Grant/Award Number: COST Action DNAqua-Net (CA15219); Agence Nationale de la Recherche, Grant/Award Number: CONVERGENOMICS (ANR-15-CE32-0005), DEEP (ANR 08JJC012001) and EUR H2O'Lyon (ANR-17-EURE-0018)

Handling Editor: Michael Matschiner

## Abstract

1. Storing, managing and sharing the ever-increasing amount of species occurrence data and biological vouchers is a cornerstone of biodiversity science. The convergence of morphology- and DNA-based methods for species delimitation has accentuated the need for optimizing laboratory workflow by generating new analytical processes, diversifying taxonomic reference systems and identification protocols and multiplying vouchers. Yet, laboratory information management systems dedicated to multi-criteria species-based research are rare or they have not been made openly available to biodiversity professionals.
2. We developed GOTIT (Gene Occurrence and Taxa in Integrative Taxonomy), a database structure and web application for optimizing productivity at various steps of a laboratory's workflow including research project design, sampling gap analysis, sampling planning, species identification, DNA sequencing, information sharing and contributing data to global biodiversity facilities. GOTIT manages the different steps of a species-occurrence data production process from sampling to sequencing, the storage of specimen lots, slides and DNA extracts and their assignment to morphology-based and DNA-based species hypotheses. The application also accommodates species occurrence data and DNA sequence metadata from external sources and the bibliographic referencing of information.
3. GOTIT is designed to optimize research activities on the diversity and evolution of species which have a strong geographic component, such as biogeography and phylogeography. A key feature of GOTIT is the ability to assign multiple species hypotheses based on morphology and molecular delimitation methods to the same set of specimens. This integrative taxonomic approach allows to explore how different aspects of speciation, such as morphological distinguishability and genetic isolation, vary in strength relative to each other across taxa and geographic regions. Another key feature is traceability which promotes repeatability and minimizes duplicate work among multiple users. Any species data refer to a set of localities, sampling methods, biological vouchers, analytical protocols and delimitation methods that can be traced.

4. The application is free and open-source with extensive documentation and demo version. Developers can reuse the source code or develop new add-ons to fulfill a laboratory's specific requirements although the application can be used as such to manage species occurrence data for a wide range of taxa.

#### KEYWORDS

biodiversity data management, biogeography, integrative taxonomy, laboratory information management system, species delimitation, species distribution, workflow management software

## 1 | INTRODUCTION

Storing, managing and sharing the ever-increasing amount of species occurrence data and their associated vouchers is a cornerstone of biodiversity science (Fišer, 2019; Hobern et al., 2012). The advent of molecular methods for species delimitation have urged the need for novel species data management tools by introducing new taxonomic references, increasing the diversity of identification protocols and vouchers and generating an ever-increasing flow of data (Prakash et al., 2017). A number of centralized biodiversity facilities enable end-users to deposit, annotate and retrieve species occurrence and molecular data for a wide range of organisms (Flemons, Guralnick, Krieger, Ranipeta, & Neufeld, 2007; Maitner et al., 2018; Prakash et al., 2017; Ratnasingham & Hebert, 2013). However, the web services provided by these facilities are not meant to assist laboratories in efficiently managing and updating their biodiversity data and materials. Rather, they are either meant to ease transferring laboratories' data into or retrieving data from these centralized biodiversity facilities.

Software-based solutions to support a laboratory's workflow in life sciences are often referred to as laboratory information management systems (LIMS). Despite sharing a base set of functionalities, existing LIMS always have specific strengths because they are developed to meet the requirements of different research activities (Table 1). For example, the LabKey platform for collaborative research in the clinical domain provides an impressive suite of tools for data integration, analysis and sharing (Nelson et al., 2011). Nonetheless, other less elaborated LIMS, such as MolabIS for molecular genetics (Truong, Groeneveld, Morgenstern, & Groeneveld, 2011) and Collec-Science for field studies (Plumejeaud-Perreau et al., 2019) provide expert tools for converting DNA sequences to multiple formats and for the tracking of field samples (Table 1). To our knowledge, no existing free and open-source LIMS have been specifically designed to provide research laboratories with an efficient tool for jointly managing, visualizing and querying their morphology-based and DNA-based species occurrence data.

Here we present GOTIT (Gene Occurrence and Taxa in Integrative Taxonomy), a ready-to-use application for managing multi-criteria species occurrence data. The application is particularly suitable for biodiversity projects that employs an integrative taxonomic approach combining morphology-based and DNA-based

species delimitation methods to explain species distribution patterns (Padial, Miralles, De la Riva, & Vences, 2010). First, we present the structure of the database designed to manage species occurrence data and their assignment to morphology- and DNA-based species hypotheses (SHs). Then, we describe the architecture and functions of the web interface used to easily access, map and manage data stored within the database. Finally, we demonstrate using benchmark data how GOTIT facilitates research aiming to elucidate biodiversity patterns.

## 2 | MATERIALS AND METHODS

### 2.1 | GOTIT relational database

The database has five distinct pathways containing 'tables' (Figure 1, Appendixes S1 and S2). The first pathway portrays a species-occurrence data production process as classically followed by many research laboratories. As part of this internal process, a 'Site' is visited and sampled ('Sampling'), thereby providing specimen lots ('Internal biological material'). Then, specimens are isolated from a lot ('Specimen') for morphological and/or molecular analysis. Isolated specimens are used for mounting on microscope slides ('Slide') and/or producing DNA sequences of different genes ('DNA', 'PCR', 'Chromatogram' and 'Internal sequence'). The second pathway traces the 'Storage' of all vouchers produced during laboratory workflow. The third pathway accommodates external data that are not produced by the hosting laboratory. This enables any user working on a particular set of taxa to collect all data concerning those taxa into a single database. External data are incorporated into two separate tables – 'External biological material' and 'External sequence' – which are linked via the table 'Sampling' to the table 'Site' containing geographic coordinates. This enables mapping simultaneously external and internal species occurrence data. The fourth pathway links biological materials and sequences to the 'Source' table, which provides the origin of data (e.g. journal articles, database and personal data).

The fifth pathway assigns morphology-based and DNA-based SHs to occurrence data by using two taxonomic datasets. The first dataset in the 'Taxon' table holds a list of binomial species names that correspond, mostly, to species described using morphological

**TABLE 1** A comparison of features among four free and open-source laboratory information management systems

Features	Laboratory information management systems			
	LabKey platform <sup>a</sup>	MolabIS	Collec-Science	GOTIT
Type of laboratory	Clinical	Molecular genetics	Field laboratory	Biogeography, phylogeography
Types of data/metadata	Sample, clinical assay, flow cytometry, microarray experiments, proteomics, multiplexed genotyping, NGS experiments	Sample, spatial coordinates, species, DNA, PCR, electrophoresis, Sanger sequencing, microsatellite	Sample, spatial coordinates, various metadata depending on sample type	Sample, spatial coordinates, species, MOTU, DNA, PCR, Sanger sequencing
Strength	Provide collaboration tools to exchange and mine data	Manage diverse sequence data formats	Allow sample labelling with QR code in the field	Tools to manage and map multi-criteria species data
Web-based sample request	+	-	-	-
User interface for scripting in R	+	-	-	-
Link to external meta-data and data	++ <sup>b</sup>	-	-	+
Upload of data in batches/ data export	++	+	+/-	+
Complex queries on data	++ <sup>c</sup>	+ <sup>d</sup>	-	+ <sup>d</sup>
Advanced web interface features	++ <sup>e</sup>	+	-	++ <sup>f</sup>
Sample and storage traceability	+	+	++	+
Management of many-to-many relationships	+	+/-	-	++
Management/ visualization of geographic data	-	+	+	++

<sup>a</sup>Only the 'community edition' is free.

<sup>b</sup>Dynamic links.

<sup>c</sup>Application Programming Interface.

<sup>d</sup>Using SQL queries.

<sup>e</sup>Customizable interface.

<sup>f</sup>Embedded web forms, interactive maps.

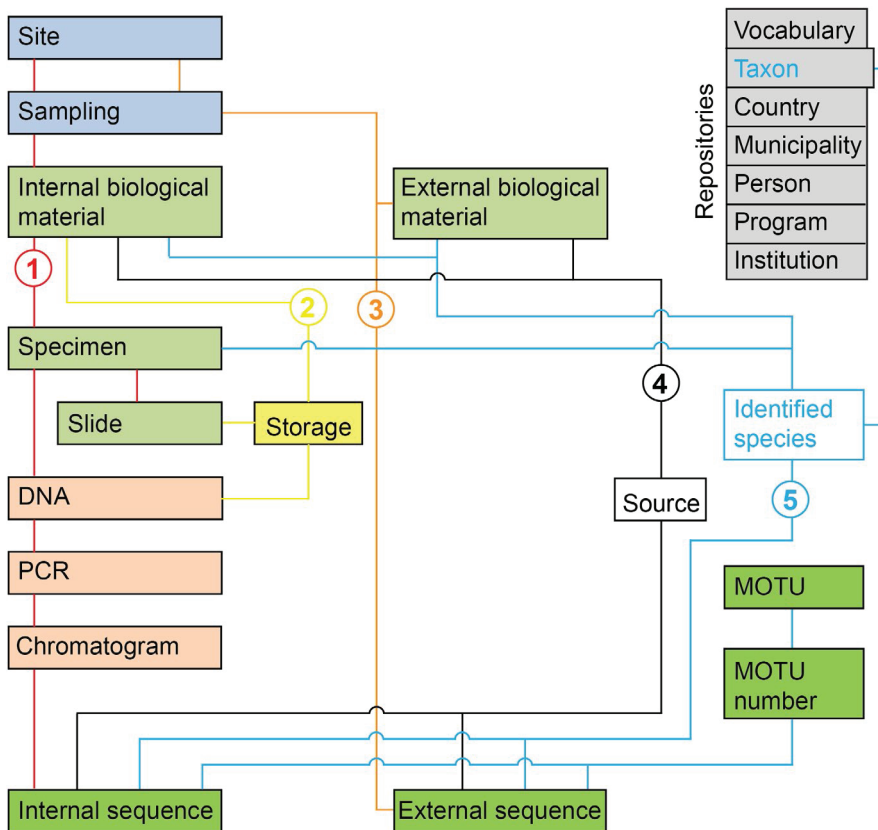
criteria. Specimen lots, individual specimens and sequences, are linked to that dataset via the table 'Identified species', which specifies the identification criterion used for species assignment. The criteria list is set up by the user; this can include morphological, bioacoustic or molecular criteria. Species assignment is highly flexible. A user can attribute a DNA sequence, and then the sequence-bearing specimen, to a morphology-based SH using the similarity between that sequence and a reference sequence assigned to that SH (i.e. barcoding; Ratnasingham & Hebert, 2013). The database also maintains the history of species assignment since the same specimen can be assigned to different species over time.

The second taxonomic dataset holds a list of 'MOTU numbers' representing molecular operational taxonomic units (Figure 1). The database does not generate MOTUs because molecular species delimitation methods are too many and continuously evolving (Fišer, Robinson, & Malard, 2018). Instead, MOTUs are generated externally and imported into the table 'MOTU number'. Sequences

are linked to 'MOTU numbers' and the details of the external analytical procedure used to delimit MOTUs and reference to the corresponding article are stored in the table 'MOTU'. Apart from molecular species delimitation, the 'MOTU number' table can accommodate any kind of OTUs delimited based on whatever criteria defined by the user. Therefore, that table provides taxonomic flexibility while keeping the link to binomial species names that are used as currencies for communication among scientists.

## 2.2 | GOTIT web interface

The web interface menu has nine headings, each containing a number of subheadings (Figure 2a,b, see detailed functionalities in Appendix S3). The first four headings successively entitled 'Field data', 'Morphological analysis', 'Molecular analysis' and 'DNA sequence', follow the different steps of the data production process. Each subheading gives access to a content form which features



**FIGURE 1** Simplified schematic model of GOTIT database. Numbered and colored lines correspond to distinct pathways. (1) red: internal data production; (2) yellow: storage; (3) orange: external data; (4) black: data sources; (5) blue: species assignment. Colors of boxes as in Figure 2b to show the links between the database and web interface

buttons to delete records, upload multiple records and access record forms for creating, reading and updating individual records (Figure 2c). Other buttons direct the user to the next step of the data production process. Record forms provide the fields to be completed for each record: data entry is most often by means of menu scrolling down the items entered in the repositories (Figure 2d). A number of specialized tools are included in the report forms such as the cartographic visualization in the site creation form of already existing sites near a candidate site. This tool prevents the user from creating several site codes for the same sampling site.

The other headings provide access to content pages for managing the 'Storage' of vouchers, the data sources linked to specimen lots and DNA sequences, and all items contained in each of the seven 'Repositories'. User privileges are granted or revoked by the administrator in the 'User' heading by selecting one of the following four roles in decreasing order of privileges: administrator, project member, collaboration member and invited member. GOTIT comprises 29 tools for uploading data in batches. Prior to uploading data using formatted csv file templates, the software checks and reports errors, such as identifier duplicates and incorrect data format.

### 3 | CASE EXAMPLES

GOTIT was developed primarily to manage species occurrence data on Aselloidea (Isopoda) produced in the lead author's laboratory

(Eme et al., 2018). In the following section, we use these data to illustrate the functionality of ready-to-use queries implemented in the 'Species search' add-on of the interface.

#### 3.1 | Examining molecular species boundaries within morphospecies

GOTIT assesses the exhaustiveness of molecular sampling within a morphospecies by comparing its geographic range as drawn using external and internal specimen lots to that of its COI sequences (Figure 3). The search returns the number of localities, latitudinal midpoints and maximum linear extent of specimen lots and COI sequences and maps of external and internal specimen lots and COI sequences. Mapping of MOTUs can be performed for any MOTU dataset and molecular delimitation methods entered by the user in the database. In Figure 4, the Poisson tree process (PTP) delimitation method reveals six MOTUs out of 62 COI sequences of *Proasellus cavaticus*. One of these MOTUs is widely distributed, whereas the other five have narrow geographic distribution ranges.

#### 3.2 | Comparing sets of SHs

Using morphology and molecular methods to delimit species boundaries often provides different sets of SHs. The comparison between two sets of SHs is implemented in GOTIT by placing each SH into one of four intuitive categories: match, split, lump and re-shuffling (Zhang, Hao, Yang, & Shi, 2016). Up to 361 molecular

(a) **GOTIT** *Gene, Occurrence and Taxa in Integrative Taxonomy* ADMIN

**MENU**

- Dashboard
- Field data
- Morphological analysis
- Molecular analysis
- DNA sequence
- Storage
- Source
- Repositories
- User
- Species search

**Field data**

Site	5558
Sample	6566

**Molecular data**

PCR	13192
Chromatogram	9996
DNA sequence	7431

**Literature data**

Lot	4933
DNA sequence	678
Reference	395

**Biological material**

Lot	1683
Specimen	5153
Slide	1681
DNA	4474

**Species hypotheses**

Taxa	509
Internal sequence assigned to MOTU	1178
External sequence assigned to MOTU	325

**Last updates**

**Tools**

- Upload site
- Upload sampling
- Upload biological material
- Upload specimen
- Upload specimen slide
- Upload ext. biological material
- Upload DNA
- Upload PCR
- Upload chromatogram
- Upload int. sequence
- Upload ext. sequence
- Upload MOTU
- Upload source
- Store biological material
- Store slide

**ADMIN**

- User profile
- Switch English / French
- Log Out

(b) **Sub-headings**

**Field data**

- Site
- Sampling

**Morphological analysis**

- Internal biological material
- Specimen
- Specimen slide
- External biological material

**Molecular analysis**

- DNA
- PCR
- Chromatogram

**DNA sequence**

- Internal sequence
- External sequence
- MOTU

**Storage**

- Biological material
- Specimen slide
- DNA

**Source**

- Source

**Repositories**

- Vocabulary
- Taxon
- Country
- Municipality
- Person
- Program
- Institution

**User**

- User

**Species search**

- COI data set
- MOTU
- Species hypotheses
- Distribution data
- Species names

(c) **Specimen content page**

**Specimen** ?

Show all records Create new specimens Import new set of specimens and related data

Site code	Specimen molecular code	Tube code	Specimen molecular nb	Last taxon name	Date Up	Show	Edit	Del	DNA	Slide
ROSSFELD	Pwalteri ROSSFELD_201409_11am	CRYO	11am	PROASELLUS_WALTERI	2018-12-17 13:25:40					
SHOPVICA	CAEsp ROSSFELD_201409_11am	A1	1ID	CAECIDOTEA_SP	2018-11-30 07:18:29					

(d) **Specimen update form**

**Specimen** Back to Biological material ?

Internal biological material: PROASELLUS\_WALTERI|ROSSFELD\_201409

Tube code: CRYO

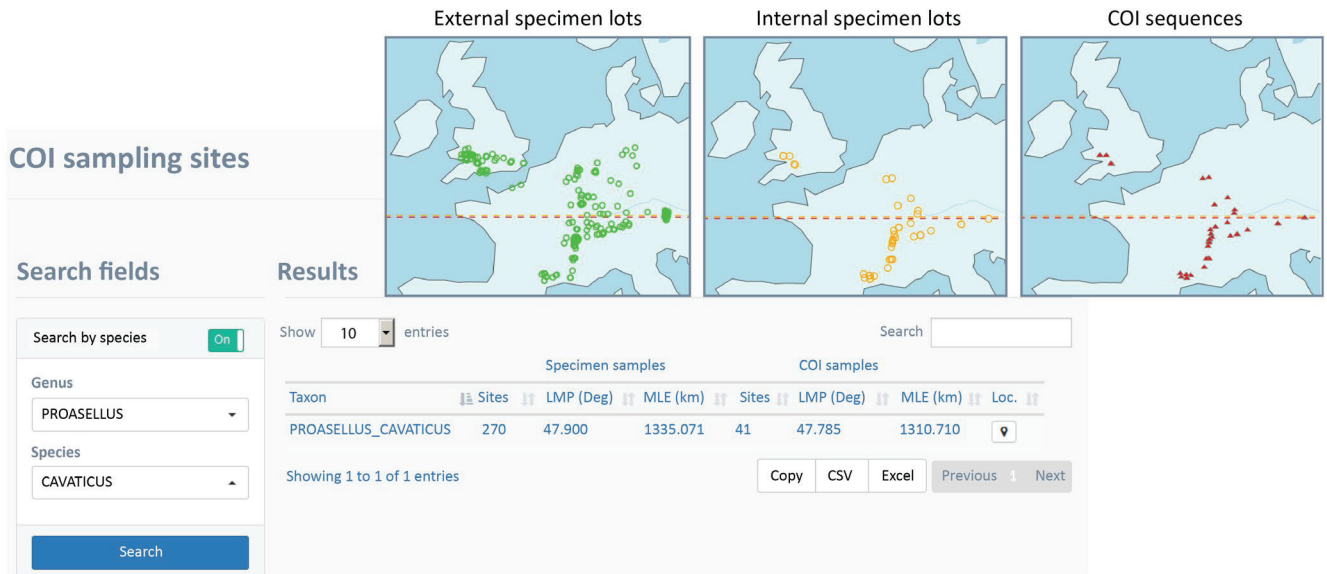
Specimen morphological code: PROASELLUS\_WALTERI|ROSSFELD\_201409[CRYO]amPWRo11

Specimen type: CRYOCONSERVED SPECIMEN

Specimen molecular number: 11am

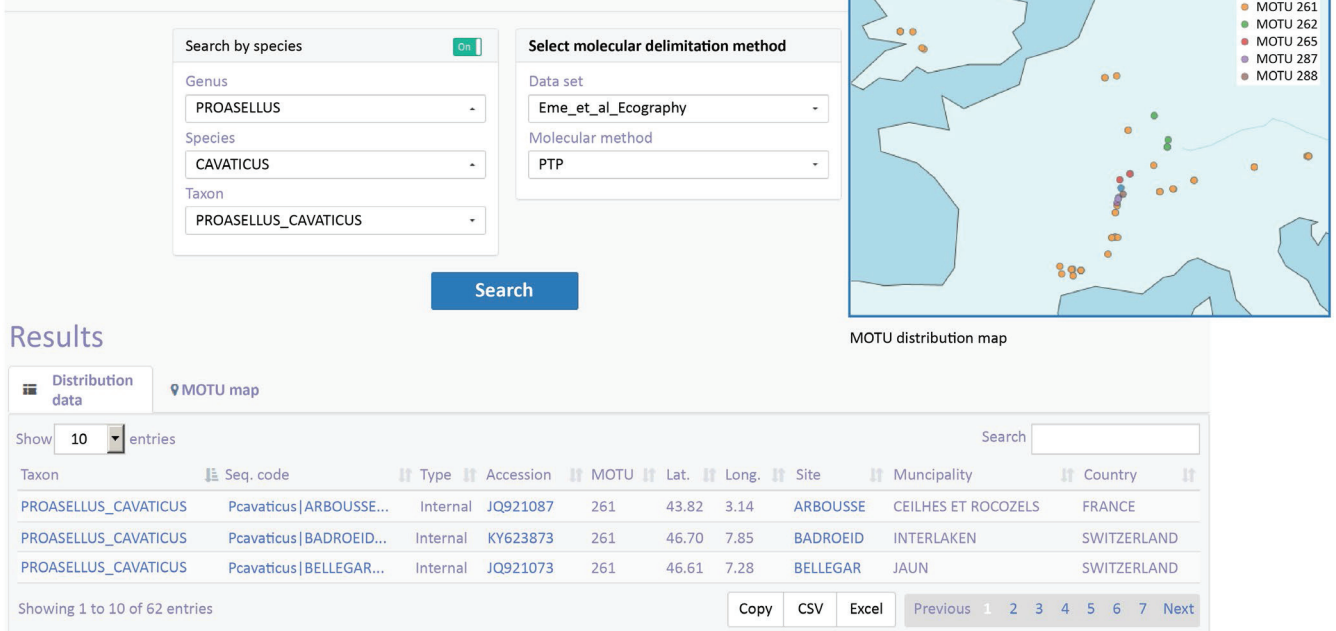
Specimen molecular code: Pwalteri|ROSSFELD\_201409\_11am

**FIGURE 2** Overview of GOTIT web interface showing extracts of the dashboard (a), sub-headings (b), a content page (c) and a record form (d)



**FIGURE 3** Assessing the exhaustiveness of molecular intra-taxon sampling. The table and maps depicts the distribution of external and internal specimen lots and COI sequences for the selected species. LMP: latitudinal midpoint of the species range. MLE: maximum linear extent of the species range

## Search fields



**FIGURE 4** Counting and mapping MOTUs within morphospecies

SHs were recovered by Eme et al. (2018) from a dataset containing 1,501 COI sequences from 152 morphological SHs of Aselloidea (Figure 5). Results show that molecular methods essentially split morphospecies into smaller clusters of individuals. The query also allows to reverse the comparison (from MOTU to morphospecies) and compare two molecular sets of SHs.

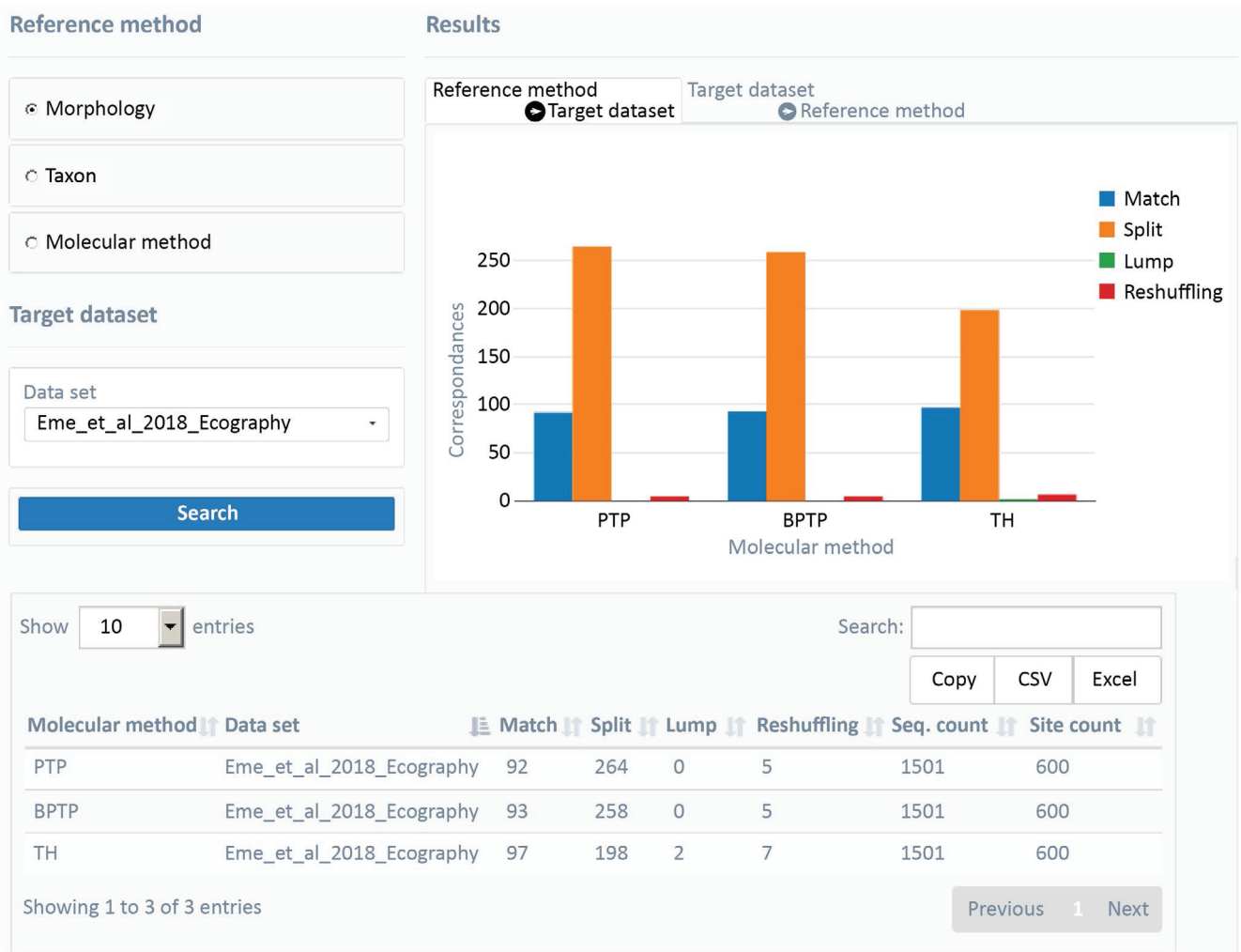
Multiple-criteria species datasets in which individual specimens are attributed to multiple SHs can be retrieved from the database using SQL query to unveil geographic variation in species

richness among different sets of SHs (Figure 6, Appendix S4). Using the dataset from Figure 5 reveals that the species richness peak at latitudes ranging from 42° to 46°N becomes increasingly pronounced using molecular delimitation methods (Figure 6).

### 3.3 | Accessing other functionalities using SQL queries

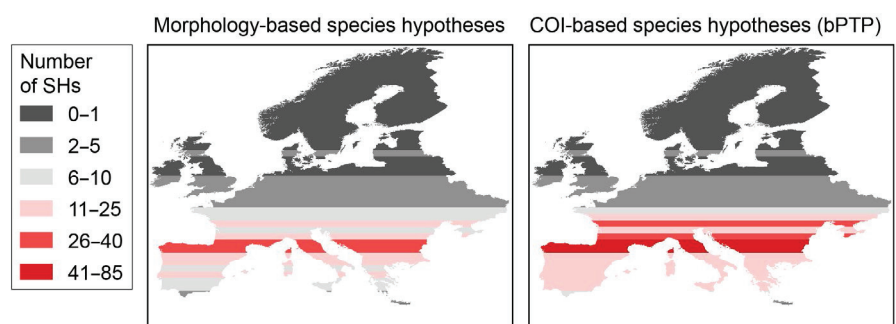
Apart from the web interface, users can access a variety of datasets using SQL queries in the open-source administration tool pgAdmin





**FIGURE 5** Comparing two sets of species hypotheses. Correspondence between a set of 197 morphospecies and three sets of MOTUs delimited using a genetic distance threshold (TH), the Poisson tree process method (PTP) and the Bayesian implementation of PTP (BPTP). Seq. count: number of sequences

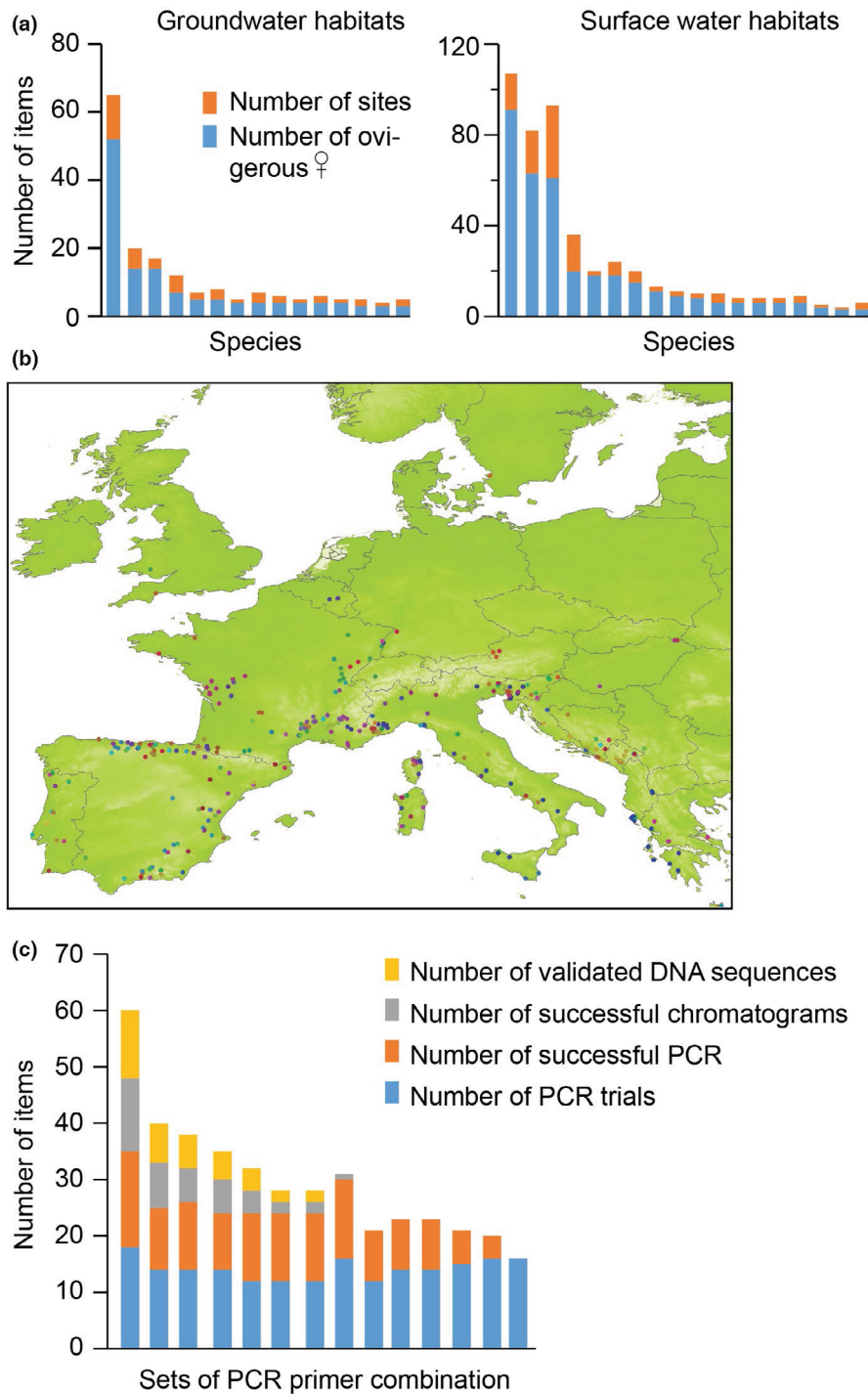
**FIGURE 6** Latitudinal patterns of Aselloidea species richness using species hypotheses delimited with morphology (left) and COI-based species hypotheses delimited with the Bayesian Poisson tree process method (bPTP) (right). Numbers are counts of species hypotheses per 0.9° latitudinal bands



(<https://www.pgadmin.org/>). Figure 7 shows a snapshot of functions of the database for optimizing research within a laboratory (Appendix S4). The first query illustrates the traceability of vouchers. The second query returns the sequence identifier codes and their associated data for inferring a phylogenetic tree. The last query provides a laboratory worker with key indicators for selecting the most suitable PCR primers to obtain DNA sequences.

#### 4 | CONCLUSIONS

GOTIT is a flexible and expandable tool for the joint management of morphology-based and DNA-based species occurrence data. These data are typically produced by research teams world-wide and collectively contribute to feed global online biodiversity databases. As experienced by the lead author's research team, the tool enhances



**FIGURE 7** A snapshot of datasets that can be retrieved from the database using SQL queries. (a) Groundwater and surface water species of Aselloidea for which at least three ovigerous females are available in storage. (b) Species occurrence data for which specimens with four targeted genes are available for inferring a phylogenetic tree of Aselloidea (data shown only for Europe). (c) Key indicators for selecting among 14 sets of PCR primer combinations to obtain COI sequences from *Proasellus walteri*. Sets are ranked in decreasing order of efficiency using the ratio of validated sequences to the number of PCR trials

productivity at various steps of a laboratory's workflow, including research project design, sampling implementation and gap analysing, species identification, DNA sequencing, information sharing and contributing data to global biodiversity facilities. A key feature of GOTIT is traceability, which promotes scientific repeatability. Any species occurrence data refer to a set of localities, sampling methods, voucher materials, analytical protocols and species delimitation methods that can be traced. Role-based access to metadata can be granted to multiple users long before sequences are publicly released, thereby promoting collaborative projects among

laboratories and reducing duplicate work. Another key feature of GOTIT is its ability to assign multiple SHs based on morphology and molecular delimitation methods to the same set of specimens. Changes in species richness distribution patterns between any two sets of SHs can be further scrutinized, thereby paving the way towards understanding how different aspects of speciation (e.g. morphological distinguishability vs. genetic isolation) vary in strength relative to each other across taxa and geographic regions (Eme et al., 2018). In addition to making the diversity of molecular SHs visible to stakeholders, keeping the history of DNA-based SHs can

lead ultimately to establishing a stable nomenclature of MOTUs, once a comprehensive taxonomic and geographic coverage of the interest group has been achieved (Morard et al., 2016). Another important feature of GOTIT is the free availability of the source code and its modular architecture. Although GOTIT probably can be used to manage species occurrence data for a wide range of taxa, programmers can implement modifications of the source code and/or new add-ons to address a particular laboratory need.

## ACKNOWLEDGEMENTS

We thank G. Lapetoule and A. Bréchet for preparing benchmark data and E. Beaumann and M.-A. Guéry for developing SQL queries. GOTIT is hosted by the CNRS/IN2P3 Computing Center (Villeurbanne, France). We thank C. Robinson for correcting the English writing and N. Puillandre and three anonymous referees for reviewing the manuscript.

## AUTHOR'S CONTRIBUTIONS

F.M. and P.G. led GOTIT development. F.M. wrote the manuscript and the database and user helps. P.G. wrote the core code and installation help. L.D. wrote the code of the 'Species search' add-on. L.K.-D. and T.L. entered benchmark data. C.M. contributed to the database structure building. D.E. and N.S. provided case examples. C.C. and C.J.D. conceived the project. All authors revised the manuscript and approved the publication.

## DATA AVAILABILITY STATEMENT

GOTIT was developed using the free and open-source database system PostgreSQL (<https://www.postgresql.org/>) and PHP Symfony framework (<https://symfony.com/>). A demo version of GOTIT is available at <https://gotit.cnrs.fr> with benchmark data. The whole project is distributed with full documentation at <https://github.com/GOTIT-DEV/GOTIT/releases/tag/v1.1.1.1> (<https://doi.org/10.5281/zenodo.3345739>) (Grison et al., 2009) under the terms of GNU General Public License.

## ORCID

Florian Malard  <https://orcid.org/0000-0001-8037-4464>

David Eme  <https://orcid.org/0000-0001-8790-0412>

Christophe J. Douady  <https://orcid.org/0000-0002-4503-8040>

## REFERENCES

- Eme, D., Zagmajster, M., Delić, T., Fišer, C., Flot, J.-F., Konecny-Dupré, L., ... Malard, F. (2018). Do cryptic species matter in macroecology? Sequencing European groundwater crustaceans yields smaller ranges but does not challenge biodiversity determinants. *Ecography*, 41, 424–436. <https://doi.org/10.1111/ecog.02683>
- Fišer, C. (2019). Collaborative databasing should be encouraged. *Trends in Ecology and Evolution*, 34, 184–185. <https://doi.org/10.1016/j.tree.2018.12.001>

- Fišer, C., Robinson, C. T., & Malard, F. (2018). Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology*, 27, 613–635. <https://doi.org/10.1111/mec.14486>
- Flemons, P., Guralnick, R., Krieger, J., Ranipeta, A., & Neufeld, D. (2007). A web-based GIS tool for exploring the world's biodiversity: The Global Biodiversity Information Facility Mapping and Analysis Portal Application (GBIF-MAPA). *Ecological Informatics*, 2, 49–60. <https://doi.org/10.1016/j.ecoinf.2007.03.004>
- Grison, P., Duchemin, L., Konecny-Dupré, L., Lefébure, T., Saclier, N., Eme, D., ... Malard, F. (2009). GOT IT -DEV/GOT IT : GOT IT v1.1.1 release.
- Hoborn, D., Apostolico, A., Arnaud, E., Bello, J. C., Canhos, D., Dubois, G., ... Willoughby, S. (2012). *Global biodiversity informatics outlook: Delivering biodiversity knowledge in the information age*. Copenhagen: Global Biodiversity Information Facility. <https://doi.org/10.15468/6jxa-yb44>
- Maitner, B. S., Boyle, B., Casler, N., Condit, R., Donoghue, J., Durán, S. M., ... Enquist, B. J. (2018). The BIEN R package: A tool to access the Botanical Information and Ecology Network (BIEN) database. *Methods in Ecology and Evolution*, 9, 373–379. <https://doi.org/10.1111/2041-210X.12861>
- Morard, R., Escarguel, G., Weiner, A. K. M., André, A., Douady, C. J., Wade, C. M., ... Kucera, M. (2016). Nomenclature for the nameless: A proposal for an integrative molecular taxonomy of cryptic diversity exemplified by planktonic foraminifera. *Systematic Biology*, 65, 925–940. <https://doi.org/10.1093/sysbio/syw031>
- Nelson, E. K., Piehler, B., Eckels, J., Rauch, A., Bellew, M., ... Igra, M. (2011). LabKey Server: An open source platform for scientific data integration, analysis and collaboration. *BMC Bioinformatics*, 12, 71. <https://doi.org/10.1186/1471-2105-12-71>
- Padial, J. M., Miralles, A., De la Riva, I., & Vences, M. (2010). The integrative future of taxonomy. *Frontiers in Zoology*, 7, 16. <https://doi.org/10.1186/1742-9994-7-16>
- Plumejeaud-Perreau, C., Quinton, E., Pignol, C., Linyer, H., Ancelin, J., Cipièrre, S., ... Bretagnolle, V. (2019). Towards better traceability of field sampling data. *Computers & Geosciences*, 129, 82–91. <https://doi.org/10.1016/j.cageo.2019.04.009>
- Prakash, P. Y., Irinyi, L., Halliday, C., Chen, S., Robert, V., & Meyer, W. (2017). Online databases for taxonomy and identification of pathogenic fungi and proposal for a cloud-based dynamic data network platform. *Journal of Clinical Microbiology*, 55, 1011–1024. <https://doi.org/10.1128/JCM.02084-16>
- Ratnasingham, S., & Hebert, P. D. N. (2013). A DNA-based registry for all animal species: The Barcode Index Number (BIN) system. *PLoS ONE*, 8, e66213. <https://doi.org/10.1371/journal.pone.0066213>
- Truong, C. V., Groeneveld, L. F., Morgenstern, B., & Groeneveld, E. (2011). MolabS – An integrated information system for storing and managing molecular genetics data. *BMC Bioinformatics*, 12, 425. <https://doi.org/10.1186/1471-2105-12-425>
- Zhang, A.-B., Hao, M.-D., Yang, C.-Q., & Shi, Z.-Y. (2016). BARCODINGR: An integrated R package for species identification using DNA barcodes. *Methods in Ecology and Evolution*, 8, 627–634. <https://doi.org/10.1111/2041-210X.12682>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Malard F, Grison P, Duchemin L, et al. GOTIT: A laboratory application software for optimizing multi-criteria species-based research. *Methods Ecol Evol*. 2019;00:1–9. <https://doi.org/10.1111/2041-210X.13307>

### **Annexe 3 :**

**Individus et séquences utilisés pour délimiter les  
MOTUs pour la phylogénie calibrée dans le temps**

Morphological identification	Sequenced individuals	Sampling Site	latitude	longitude	PTP MOTU	Accession number	
<i>Asellus (Asellus) aquaticus</i> (Linnaeus, 1758)	Aaquaticus KHARKIV_000000_0_DQ144824 NCBI	KHARKIV	50,039	36,162	Species_139	DQ144824	
	Aaquaticus KHARKIV_000000_0_DQ144826 NCBI	KHARKIV	50,039	36,162	Species_139	DQ144826	
	Aaquaticus DRASIWEL_201610_1ID_YAF590	DRASIWEL	35,231	25,602	Species_140	NA	
	Aaquaticus ZOODHOS_000000_0_DQ144828 NCBI	ZOODHOS	36,372	25,475	Species_140	DQ144828	
	Aaquaticus ASTRAKRIV_201610_1ID_YAF573	ASTRAKRIV	35,241	25,221	Species_141	NA	
	Aaquaticus DILAKKOS_201610_1ID_YAF577	DILAKKOS	35,320	25,668	Species_141	NA	
	Aaquaticus DORIESPD_201610_1ID_YAF616	DORIESPD	35,287	25,677	Species_141	NA	
	Aaquaticus KATOVRYIS_201610_1ID_YAF618	KATOVRYIS	35,191	25,222	Species_141	NA	
	Aaquaticus BASKE_000000_0_DQ144788 NCBI	BASKE	44,527	15,181	Species_142	DQ144788	
	Aaquaticus ALBORG_000000_0_DQ144816 NCBI	ALBORG	57,047	9,919	Species_143	DQ144816	
	Aaquaticus CEZEBAGN_201605_1ex_YAF010	CEZEBAGN	44,164	4,634	Species_143	NA	
	Aaquaticus CEZEBAGN_201605_2ex_YAF01	CEZEBAGN	44,164	4,634	Species_143	NA	
	Aaquaticus CRAPONNE_201605_1ex_YAF012	CRAPONNE	45,738	4,732	Species_143	NA	
	Aaquaticus DOMMEDOR_201608_1ID_YAF579	DOMMEDOR	44,811	1,219	Species_143	NA	
	Aaquaticus DONRIVE_000000_0_DQ144886	DONRIVE	53,494	-1,248	Species_143	DQ144886	
	Aaquaticus ETGDUA_201111_1ID_YAB589	ETGDUA	45,780	4,867	Species_143	KY623724	
	Aaquaticus HCOLLIAS_201605_1ex_YAF014	HCOLLIAS	43,954	4,483	Species_143	NA	
	Aaquaticus HOMPS_000000_0_AY531784 NCBI	HOMPS	43,267	2,719	Species_143	AY531784	
	Aaquaticus KAMENEC_201105_1_YAF015	KAMENEC	48,368	21,825	Species_143	NA	
	Aaquaticus KILLARNE_201204_1_YAF016	KILLARNE	52,058	-9,512	Species_143	NA	
	Aaquaticus KRCAVA_201105_1_YAF017	KRCAVA	48,356	21,834	Species_143	NA	
	Aaquaticus OLSZTYN_000000_0_DQ144821 NCBI	OLSZTYN	53,811	20,437	Species_143	DQ144821	
	Aaquaticus TANNERIE_200905_1_YAC700	TANNERIE	44,778	-0,588	Species_143	KY623728	
	Aaquaticus THRISO_201610_1ID_YAF592	THRISO	35,477	23,930	Species_143	NA	
	Aaquaticus VELKY_201105_1_YAF121	VELKY	48,345	21,835	Species_143	NA	
	Aaquaticus VRYSSSES_201610_1ID_YAF614	VRYSSSES	35,376	24,201	Species_143	NA	
	Aaquaticus WAGENIN_000000_0_DQ144844 NCBI	WAGENIN	51,968	5,663	Species_143	DQ144844	
	Aaquaticus DRAVARIV_000000_0_DQ144757 NCBI	DRAVARIV	46,385	15,936	Species_144	DQ144757	
	Aaquaticus MATELIC_000000_0_DQ144887 NCBI	MATELIC	43,255	13,012	Species_145	DQ144887	
	Aaquaticus CALCINAT_201608_1ID_YAF575	CALCINAT	45,450	10,385	Species_147	NA	
	Aaquaticus FBORMIDA_201408_1ID_YAF013	FBORMIDA	44,326	8,200	Species_147	NA	
	Aaquaticus FONTTOSI_201504_1ID_YAE408	FONTTOSI	45,562	8,555	Species_147	NA	
	Aaquaticus FONTTOSI_201504_3_YAE425	FONTTOSI	45,562	8,555	Species_147	NA	
	Aaquaticus PRATIRO_201408_1ID_YAD043	PRATIRO	45,834	9,018	Species_147	KY623725	
	Aaquaticus PRATIRO_201408_2_YAD044	PRATIRO	45,834	9,018	Species_147	KY623726	
	Aaquaticus PRATIRO_201408_3_YAD523	PRATIRO	45,834	9,018	Species_147	NA	
	Aaquaticus SURVAGO_201504_1ID_YAE409	SURVAGO	45,394	11,155	Species_147	NA	
	Aaquaticus SURVAGO_201504_2_YAE427	SURVAGO	45,394	11,155	Species_147	NA	
	Aaquaticus CERKNICA_000000_0_AY531811 NCBI	CERKNICE	45,728	14,404	Species_148	AY531811	
	Aaquaticus PIVKACAV_201504_6ID_YAD965	PIVKACAV	45,820	14,246	Species_148	NA	
	Aaquaticus PIVKACAV_201504_8_YAE392	PIVKACAV	45,820	14,246	Species_148	NA	
	Aaquaticus PLANINA_201504_6ID_YAD964	PLANINA	45,868	14,243	Species_148	NA	
	Aaquaticus PLANINA_201504_3arr_TRINITYDN50490c0g11	PLANINA	45,868	14,243	Species_148	NA	
	Aaquaticus PLANINA_201504_7_YAE426	PLANINA	45,868	14,243	Species_148	NA	
	Aaquaticus TREBICI1_000000_0_AY531762 NCBI	TREBICI1	45,673	13,825	Species_149	AY531762	
	Aaquaticus EVROS_000000_0_DQ144885 NCBI	EVROS	40,851	26,178	Species_150	DQ144885	
	Aaquaticus ASL4_201409_1IDex_YAE876	ASL4	42,484	19,154	Species_151	NA	
	Aaquaticus ASL4_201409_3ex_YAE886	ASL4	42,484	19,154	Species_151	NA	
	Aaquaticus SPUZANEX_201107_1IDex_YAC699	SPUZANEX	42,486	19,174	Species_151	KY623727	
	<i>Asellus (Asellus) aquaticus arthrobranchialis</i> Dudich, 1925	Aaquaticus arthrobranchialis FSARNOII_201503_1ID_YAE158	FSARNOII	40,813	14,616	Species_146	NA
		Aaquaticus arthrobranchialis FSARNOII_201503_2_YAE237	FSARNOII	40,813	14,616	Species_146	NA
		Ahilgendorfi AHILTOYA_000000_1_YAE658	AHILTOYA	36,696	137,214	Species_127	NA
<i>Asellus (Asellus) hilgendorfi</i> Bovallius, 1886	Ahilgendorfi AHILTOYA_000000_2_YAE727	AHILTOYA	36,696	137,214	Species_127	NA	
	Ahilgendorfi TOYAMA_000000_0_AY531829 NCBI	TOYAMA	36,696	137,214	Species_127	AY531829	
<i>Asellus (Asellus) kosswigi</i> Verovnik, Prevornik & Jugovic, 2009	Akosswig TREBICGR_200001_1ID_YAE732	TREBICGR	45,685	13,834	Species_149	NA	
	Akosswig TREBICGR_200001_2_YAE704	TREBICGR	45,685	13,834	Species_149	NA	
<i>Baicalasellus korotnevi</i> (Semenkevich, 1924)	BAikortnevi BAIKAL5B_201409_1ID_YCM937	BAIKAL5B	51,903	105,077	Species_160	KY623731	
	BAikortnevi BAIKAL5B_201409_2ID_YCM927	BAIKAL5B	51,903	105,077	Species_160	KY623732	
	BAikortnevi BAIKAL5B_201409_6IDex_YCM926	BAIKAL5B	51,903	105,077	Species_160	KY623733	
	BAislp BAIKAL5_201409_2ID_YCM909	BAIKAL5	51,903	105,077	Species_161	KY623730	
	BAislp BAIKAL5_201409_1ID_YCM909	BAIKAL5	51,903	105,077	Species_161	KY623729	
<i>Baicalasellus baicalensis</i> (Grube, 1872)	BAislp BAIKALL_000000_0_DQ144778 NCBI	BAIKALL	53,174	107,662	Species_161	DQ144778	
	<i>Balkanostenasellus skopljensis</i> (Karaman, 1937)	Balkanostenasellus skopljensis OTOKBRAC_201509_1ID_YAE719	OTOKBRAC	43,360	16,610	Species_2	NA
		Balkanostenasellus skopljensis OTOKBRAC_201509_2_YAE738	OTOKBRAC	43,360	16,610	Species_2	NA
		Balkanostenasellus skopljensis VESANOHY_201608_1arr	VESANOHY	46,218	14,699	Species_1	NA
	<i>Bragasellus lagari</i> Henry & Magniez, 1973	Blagari CF ARGONGE_200910_1ex_YCD206	ARGONGE	39,103	-1,041	Species_86	JQ921525
Blagari CF ARGONGE_200910_2ex_YCD233		ARGONGE	39,103	-1,041	Species_86	JQ921526	
Bcomasi COBALLE2_200906_2_YCD283		COBALLE2	43,185	-5,388	Species_74	JQ921529	
Bcomasi PENDONES_200906_1ID_YAS979		PENDONES	43,132	-5,247	Species_76	JQ921530	
<i>Bragasellus comasi</i> Henry & Magniez, 1976	Bcomasi FRIES_200906_1ID_YLK133	FRIES	43,423	-5,088	Species_77	JQ921532	
	Bcomasi FRIES_200906_2_YLK326	FRIES	43,423	-5,088	Species_77	JQ921533	
	Bcomasi FRIES_200906_3_YLK246	FRIES	43,423	-5,088	Species_77	JQ921534	
	Bcomasi CONCHA_201010_1IDex_YAB436	CONCHA	43,560	-6,190	Species_78	KY623734	
	Bcomasi CONCHA_201010_3cytox_YAB018	CONCHA	43,560	-6,190	Species_78	KY623735	
	Bcomasi COBALLE2_200906_1_YCD317	COBALLE2	43,185	-5,388	Species_79	JQ921528	
	Bcomasi COBALLE1_200906_1_YLK134	COBALLE1	43,186	-5,387	Species_80	JQ921527	
	Bcomasi LAFOZ_200906_1_YLK512	LAFOZ	43,149	-5,275	Species_80	JQ921535	
	Bcomasi PENDONES_200906_2ID_YCD318	PENDONES	43,132	-5,247	Species_80	JQ921531	
	Bcortesij POUCA_201010_1ID_YCD870	POUCA	41,491	-7,649	Species_60	JQ921536	
<i>Bragasellus cortesii</i> Afonso, 1989	Bcortesij POUCA_201010_2_YCD856	POUCA	41,491	-7,649	Species_60	JQ921537	
	Bcortesij POUCA_201010_3_YCD914	POUCA	41,491	-7,649	Species_60	JQ921538	
	Bcortesij POUCAFON_201504_1ID_YAE396	POUCAFON	41,490	-7,649	Species_60	NA	
	Bescolai YUSOREIN_200906_1ID_YLK175	YUSOREIN	43,022	-4,866	Species_67	JQ921539	
	Bescolai YUSOREIN_200906_4ID_YCD93	YUSOREIN	43,022	-4,866	Species_67	JQ921540	
<i>Bragasellus frontellum</i> (Braga, 1964)	Bfrontellum ABOIM_201010_1ID_YCD785	ABOIM	41,917	-8,447	Species_61	JQ921541	
	Bfrontellum ABOIM_201010_2_YCD871	ABOIM	41,917	-8,447	Species_61	JQ921542	
	Bfrontellum ABOIM_201010_2ex_YCD853	ABOIM	41,917	-8,447	Species_61	JQ921543	
<i>Bragasellus lagari</i> Henry & Magniez, 1973	Blagari SUMIDORS_200910_1ID_YCD34	SUMIDORS	38,880	-0,689	Species_83	JQ921555	
	Blagari SUMIDORS_200910_2_YCD32	SUMIDORS	38,880	-0,689	Species_83	JQ921556	
	Blagari CABRIEL_200910_1ID_YCD33	CABRIEL	39,250	-1,094	Species_84	JQ921552	
	Blagari CABRIEL_200910_2_YCD74	CABRIEL	39,250	-1,094	Species_84	JQ921553	
	Blagari CAUTAB_200910_2ID_YLK965	CAUTAB	39,144	-1,047	Species_84	JQ921549	
	Blagari CAUTAB_200910_3ID_YCD11	CAUTAB	39,144	-1,047	Species_84	JQ921550	
	Blagari CAUTAB_200910_5_YCD90	CAUTAB	39,144	-1,047	Species_84	JQ921551	
	Blagari TRILLO_200910_1ID_YAC799	TRILLO	40,691	-2,583	Species_85	KY623736	
	Blagari TRILLO_200910_2_YCD79	TRILLO	40,691	-2,583	Species_85	JQ921554	
	Blagari ALBENTOS_200910_1ID_YCD98	ALBENTOS	40,122	-0,764	Species_87	JQ921544	
	Blagari MEREMIL_200910_1ID_YCD25	MEREMIL	39,598	-0,849	Species_88	JQ921547	
	Blagari MEREMIL_200910_2_YCD36	MEREMIL	39,598	-0,849	Species_88	JQ921548	
	<i>Bragasellus lagari</i> Henry & Magniez, 1973	Blagari CASCAN2_200910_1ID_YCD24	CASCAN2	40,204	-1,130	Species_89	JQ921545
		Blagari CASCAN2_200910_2_YCD39	CASCAN2	40,204	-1,130	Species_89	JQ921546
<i>Bragasellus lagarioides</i> Henry & Magniez, 1996	Blagarioides MARFIL_200906_1_YAS876	MARFIL	42,861	-3,593	Species_69	JQ921558	
	Blagarioides MARFIL_200906_2_YLK101	MARFIL	42,861	-3,593	Species_69	JQ921559	

	Blagarioides MARFIL_200906_4ID_YLK374	MARFIL	42,861	-3,593	Species_69	JQ921560
	Blagarioides RUDRON_201010_1ID_YCD782	RUDRON	42,702	-3,918	Species_69	JQ921563
	Blagarioides RUDRON_201010_2_YCD791	RUDRON	42,702	-3,918	Species_69	JQ921564
	Blagarioides CREUS_200906_1_YLK181	CREUS	41,349	1,362	Species_71	JQ921557
	Blagarioides PERPETUA_200906_1ID_YAS954	PERPETUA	41,463	1,391	Species_71	JQ921561
	Blagarioides PERPETUA_200906_2_YLK187	PERPETUA	41,463	1,391	Species_71	JQ921562
	Blagarioides LAALFRAN_201300_2_YAD323	LAALFRAN	41,592	-0,767	Species_72	KY623737
	Blagarioides LAALFRAN_201300_3_YAF731	LAALFRAN	41,592	-0,767	Species_73	NA
Bragasellus molinae	Henry & Magniez, 1988	VALPORQU	42,907	-5,563	Species_70	JQ921565
	Bmolinae VALPORQU_201010_1ID_YCD861	VALPORQU	42,907	-5,563	Species_70	JQ921566
	Bmolinae VALPORQU_201010_2_YCD850	VALPORQU	42,907	-5,563	Species_70	JQ921567
	Bmolinae VALPORQU_201010_3_YCD909	VALPORQU	42,907	-5,563	Species_70	JQ921567
	Bmolinae VALPORQU_201010_37f_YCM374	VALPORQU	42,907	-5,563	Species_70	KY623739
	Bmolinae VALPORQU_201010_34m_YAC987	VALPORQU	42,907	-5,563	Species_70	KY623738
Bragasellus peltatus	(Braga, 1944)	MEZIO	41,282	-8,308	Species_62	JQ921568
	Bpeltatus MEZIO_201010_2_YCD790	MEZIO	41,282	-8,308	Species_62	JQ921569
	Bpeltatus MEZIO_201010_10m_YCM372	MEZIO	41,282	-8,308	Species_62	KY623740
	Bpeltatus MEZIO_201010_11m_YCM373	MEZIO	41,282	-8,308	Species_62	KY623741
	Bpeltatus SOUTO_201010_1ID_YCD869	SOUTO	41,221	-8,249	Species_62	JQ921570
Bragasellus rouchi	Henry & Magniez, 1988	DEVAPANE	43,311	-4,610	Species_75	JQ921571
Bragasellus stocki	Henry & Magniez, 1988	CONCHA	43,560	-6,190	Species_68	JQ921524
	Bstocki CONCHA_201010_2cpto_YAB019	CONCHA	43,560	-6,190	Species_68	JQ921573
	Bstocki CONCHA_201010_3_YCD928	CONCHA	43,560	-6,190	Species_68	JQ921573
Caecidotea brevicauda	(Forbes, 1876)	SPNEMCCA	37,587	-89,440	Species_100	NA
	CAEbrevicauda SPNEMCCA_201606_1_YAF497	SPNEMCCA	37,587	-89,440	Species_101	NA
	CAEbrevicauda TEAKETTL_201609_1ID_YAF682	TEAKETTL	38,774	-83,434	Species_101	NA
	CAEbrevicauda TEAKETTL_201609_2_YAF503	TEAKETTL	38,774	-83,434	Species_101	NA
	CAEbrevicauda TEAKETTL_201609_3_YAF665	TEAKETTL	38,774	-83,434	Species_101	NA
Caecidotea antricola	Creaser, 1931	ROCKSPRI	37,760	-90,480	Species_166	NA
	CAEustus ROCKSPRI_201606_1_YAF502	ROCKSPRI	37,760	-90,480	Species_166	NA
	CAEustus ROCKSPRI_201606_2_YAF537	ROCKSPRI	37,760	-90,480	Species_166	NA
	CAEustus ROCKSPRI_201606_3_YAF667	ROCKSPRI	37,760	-90,480	Species_166	NA
Caecidotea kenki	(Bowman, 1967)	PIMMIT1	38,929	-77,118	Species_90	JQ921574
	CAEkenki PIMMIT1_200902_1ID_YAS803	PIMMIT1	38,929	-77,118	Species_90	JQ921575
	CAEkenki PIMMIT1_200902_2ID_YAS704	PIMMIT1	38,929	-77,118	Species_90	JQ921575
	CAEkenki PIMMIT2_201109_1_YAC053	PIMMIT2	38,929	-77,119	Species_90	KY623742
Caecidotea reddelli	(Steeves, 1968)	ROCKCREE	38,985	-77,062	Species_90	AF255778
Caecidotea rotunda	Bowman & Lewis, 1984	BEARSPR	31,207	-97,513	Species_94	AY566522
	CAErotunda APPALACH_201608_2_YAF498	APPALACH	38,690	-83,441	Species_91	NA
	CAErotunda APPALACH_201608_1ID_YAF516	APPALACH	38,690	-83,441	Species_91	NA
	CAErotunda FROSTCAV_201511_1ID_YAF504	FROSTCAV	39,094	-83,338	Species_92	NA
	CAErotunda FROSTCAV_201511_2_YAF534	FROSTCAV	39,094	-83,338	Species_92	NA
Caecidota sp. 1		CREEKPAR	39,146	-76,491	Species_54	AF260834
Caecidota sp. 2		UNKNOWN	39,816	-105,655	Species_111	DQ889085
Caecidotea racovitzai	(Williams, 1970)	SIDECUTS	41,557	-83,672	Species_336	NA
	CAEsp SIDECUTS_201609_1ID_YAF513	SIDECUTS	41,557	-83,672	Species_336	NA
	CAEsp SIDECUTS_201609_2_YAF492	SIDECUTS	41,557	-83,672	Species_336	NA
Caecidota sp. aff. rotunda		BERNICEC	39,008	-85,388	Species_93	NA
	CAEnsprotunda AFF BERNICEC_201611_1ID_YAF499	BERNICEC	39,008	-85,388	Species_93	NA
	CAEnsprotunda AFF BERNICEC_201611_2_YAF517	BERNICEC	39,008	-85,388	Species_93	NA
Caecidotea stygia	Packard, 1871	LIMEKILN	38,166	-86,280	Species_167	NA
	CAEstygia LIMEKILN_201611_1ID_YAF520	LIMEKILN	38,166	-86,280	Species_167	NA
	CAEstygia LIMEKILN_201611_2_YAF571	LIMEKILN	38,166	-86,280	Species_168	NA
	CAEstygia CURLYSPC_201608_2_YAF518	CURLYSPC	38,817	-83,433	Species_168	NA
	CAEstygia CURLYSPC_201608_1ID_YAF519	CURLYSPC	38,817	-83,433	Species_168	NA
Chthonasellus bodoni	Argano & Messana, 1991	ANNA	44,313	7,679	Species_340	JQ921579
	Cbodoni ANNA_200904_1ID_YAS727	ANNA	44,313	7,679	Species_340	JQ921580
	Cbodoni ANNA_200904_2ID_YAS782	ANNA	44,313	7,679	Species_340	JQ921581
	Cbodoni ANNA_200904_3_YAS799	ANNA	44,313	7,679	Species_340	JQ921581
	Cbodoni CHIUPESI_200908_1ID_YLK451	CHIUPESI	44,216	7,660	Species_349	JQ921576
	Cbodoni CHIUPESI_200908_2ID_YLK495	CHIUPESI	44,216	7,660	Species_349	JQ921577
	Cbodoni CHIUPESI_200908_3_YCD94	CHIUPESI	44,216	7,660	Species_349	JQ921578
Gallasellus helyi	(Legrand, 1956)	FONTROI	46,035	-0,464	Species_130	JQ921587
	Ghelyi FONTROI_200812_1ID_YAS703	FONTROI	46,035	-0,464	Species_130	JQ921588
	Ghelyi FONTROI_200812_2_YAS685	FONTROI	46,020	-0,449	Species_130	JQ921590
	Ghelyi FONTVEUV_200812_2ID_YAS617	FONTVEUV	46,020	-0,449	Species_130	JQ921590
	Ghelyi FONTVEUV_200812_3_YAS729	FONTVEUV	46,020	-0,449	Species_130	JQ921591
	Ghelyi FPINQUET_201506_3ex_YAE805	FPINQUET	45,245	0,435	Species_130	NA
	Ghelyi FONPIPOU_201506_1_YAE690	FONPIPOU	45,308	-0,326	Species_131	NA
	Ghelyi PAMPINGR_201505_1ID_YAE693	PAMPINGR	45,753	-0,779	Species_131	NA
	Ghelyi PAMPINGR_201505_2_YAE717	PAMPINGR	45,753	-0,779	Species_131	NA
	Ghelyi VAUZELLE_201501_1ID_YAE400	VAUZELLE	45,826	-0,786	Species_131	NA
	Ghelyi SCBOUILL_201504_1ID_YAE691	SCBOUILL	46,392	-0,699	Species_132	NA
	Ghelyi SCBOUILL_201504_2_YAE667	SCBOUILL	46,392	-0,699	Species_132	NA
	Ghelyi SCBOUILL_201504_3_YAE661	SCBOUILL	46,392	-0,699	Species_132	NA
	Ghelyi ASSISESP_201505_1ID_YAE692	ASSISESP	46,305	-0,995	Species_133	NA
	Ghelyi FONTVEUV_200812_1_YAB870	FONTVEUV	46,020	-0,449	Species_133	KM249025
	Ghelyi PTSUREAU_201403_2ex_YCM631	PTSUREAU	46,121	0,105	Species_133	KY623762
	Ghelyi PTSUREAU_201403_3ex_YCM629	PTSUREAU	46,121	0,105	Species_133	KY623763
	Ghelyi PUITD_200912_1ID_YLK726	PUITD	46,085	-0,925	Species_133	JQ921582
	Ghelyi PUITD_200912_2_YLK833	PUITD	46,085	-0,925	Species_133	JQ921583
	Ghelyi PUITD_200912_3_YLK830	PUITD	46,085	-0,925	Species_133	JQ921584
	Ghelyi SCTUSSON_201502_1ID_YAE399	SCTUSSON	45,931	0,068	Species_133	NA
	Ghelyi SCTUSSON_201502_2_YAE679	SCTUSSON	45,931	0,068	Species_133	NA
	Ghelyi SOMPT_200812_1ID_YLK549	SOMPT	46,156	-0,102	Species_133	JQ921589
	Ghelyi ANDILLE1_201209_1m_YAE839	ANDILLE1	46,481	0,334	Species_134	NA
	Ghelyi ANDILLE2_201209_1ID_YAC805	ANDILLE2	46,481	0,333	Species_134	KY623743
	Ghelyi CHCURZAY_201311_1IDex_YCM748	CHCURZAY	46,485	0,053	Species_134	KY623746
	Ghelyi CHCURZAY_201311_2ex_YCM627	CHCURZAY	46,485	0,053	Species_134	KY623747
	Ghelyi CLOSROCH_201405_3ex_YCM641	CLOSROCH	46,483	0,370	Species_134	KY623750
	Ghelyi CLOSROCH_201405_1IDex_YCM628	CLOSROCH	46,483	0,370	Species_134	KY623748
	Ghelyi CLOSROCH_201405_2ex_YCM636	CLOSROCH	46,483	0,370	Species_134	KY623749
	Ghelyi FONTMART_201403_2ex_YCM749	FONTMART	46,455	0,313	Species_134	KY623752
	Ghelyi FONTMART_201403_1IDex_YCM626	FONTMART	46,455	0,313	Species_134	KY623751
	Ghelyi FTAIRVAU_201408_1_YAE418	FTAIRVAU	46,827	-0,138	Species_134	NA
	Ghelyi FTAIRVAU_201408_2_YAE436	FTAIRVAU	46,827	-0,138	Species_134	NA
	Ghelyi FTAQUARI_201403_2ex_YCM630	FTAQUARI	46,483	0,373	Species_134	KY623754
	Ghelyi FTAQUARI_201403_1IDex_YCM750	FTAQUARI	46,483	0,373	Species_134	KY623753
	Ghelyi FTHILAIR_201409_1ID_YAE397	FTHILAIR	46,912	0,699	Species_134	NA
	Ghelyi FTHILAIR_201409_2_YAE678	FTHILAIR	46,912	0,699	Species_134	NA
	Ghelyi GABOURET_201403_2ex_YCM689	GABOURET	46,457	0,165	Species_134	KY623757
	Ghelyi GABOURET_201403_3ex_YCM646	GABOURET	46,457	0,165	Species_134	KY623758
	Ghelyi GABOURET_201403_1ex_YCM645	GABOURET	46,457	0,165	Species_134	KY623756
Gallasellus helyi	(Legrand, 1956)	MERIGNY	46,641	0,935	Species_134	JQ921585
	Ghelyi MÉRIGNY_200912_1ex_YLK552	MÉRIGNY	46,641	0,935	Species_134	JQ921586
	Ghelyi MÉRIGNY_200912_2ex_YLK752	MÉRIGNY	46,641	0,935	Species_134	JQ921586
	Ghelyi PTSSOULE_201404_1IDex_YCM625	PTSSOULE	46,638	0,164	Species_134	KY623759
	Ghelyi PTSSOULE_201404_2ex_YCM635	PTSSOULE	46,638	0,164	Species_134	KY623760
	Ghelyi BATAILLE_201305_2_YAC787	BATAILLE	46,168	-0,035	Species_135	KY623745
	Ghelyi BATAILLE_201305_1_YAC797	BATAILLE	46,168	-0,035	Species_135	KY623744
	Ghelyi FTAQUARI_201403_3ex_YCM650	FTAQUARI	46,483	0,373	Species_135	KY623755
	Ghelyi PTSUREAU_201403_1IDex_YCM637	PTSUREAU	46,121	0,105	Species_135	KY623761
	Ghelyi SPTOUVRE_201612_1ex_YAF574	SPTOUVRE	45,663	0,255	Species_136	NA
	Ghelyi SPTOUVRE_201612_2ex_YAF645	SPTOUVRE	45,663	0,255	Species_136	NA
	Ghelyi GROSBOTG_201506_1ID_YAE689	GROSBOTG	46,008	0,406	Species_137	NA

	GheliyiFTMICHAU_201502_1_YAE398	FTMICHAU	45,684	0,265	Species_138	NA
	GnspRabier PUYRABIE_201404_1ID_YAD112	PUYRABIE	46,342	0,382	Species_137	KY623764
	LIMPoberezhnii BAIKAL5_201409_1ID_YCM906	BAIKAL5	51,903	105,077	Species_129	KY623765
<i>Limnoasellus poberezhnii</i> Niatyaganova & Hidding, in prep. (nomen nudum in Hidding et al. 2003)	LIMPoberezhnii BAIKAL5B_201409_2_YCM910	BAIKAL5B	51,903	105,077	Species_129	KY623766
<i>Lirceolus bisetus</i> (Steeves, 1968)	LIRCEObisetus GORMAN_000000_GM74_AY566533 NCBI	GORMAN	31,018	-98,505	Species_95	AY566533
	LIRCEObisetus GORMAN_000000_GM73_AY566532 NCBI	GORMAN	31,018	-98,505	Species_95	AY566532
	LIRCEObisetus GORMAN_000000_GM1_AY566531 NCBI	GORMAN	31,018	-98,505	Species_95	AY566531
<i>Lirceolus coccytus</i> Lewis, 2001	LIRCEOcoccytus PHANTOM_000000_PH83_AY566544 NCBI	PHANTOM	30,753	-103,599	Species_102	AY566544
	LIRCEOcoccytus PHANTOM_000000_PH82_AY566543 NCBI	PHANTOM	30,753	-103,599	Species_102	AY566543
	LIRCEOcoccytus PHANTOM_000000_PH81_AY566542 NCBI	PHANTOM	30,753	-103,599	Species_102	AY566542
	LIRCEOcoccytus SOTANO_000000_AM69_AY566528 NCBI	SOTANO	29,403	-101,334	Species_103	AY566528
	LIRCEOcoccytus SOTANO_000000_AM68_AY566527 NCBI	SOTANO	29,403	-101,334	Species_103	AY566527
	LIRCEOcoccytus SLAUGHTE_000000_SB110_AY566480 NCBI	SLAUGHTE	29,668	-101,053	Species_104	AY566480
	LIRCEOcoccytus SLAUGHTE_000000_SB111_AY566479 NCBI	SLAUGHTE	29,668	-101,053	Species_104	AY566479
<i>Lirceolus hardeni</i> Lewis & Bowman, 1996	LIRCEOhardeni PERDERNA_000000_PF10_AY566541 NCBI	PERDERNA	30,344	-98,222	Species_96	AY566541
	LIRCEOhardeni PERDERNA_000000_PF9_AY566540 NCBI	PERDERNA	30,344	-98,222	Species_96	AY566540
	LIRCEOhardeni BARTONSP_000000_SG93_AY566552 NCBI	BARTONSP	30,268	-97,766	Species_97	AY566552
	LIRCEOhardeni BARTONSP_000000_SG4_AY566551 NCBI	BARTONSP	30,268	-97,766	Species_97	AY566551
	LIRCEOhardeni BARTONSP_000000_SG3_AY566550 NCBI	BARTONSP	30,268	-97,766	Species_97	AY566550
	LIRCEOhardeni BARTONSP_000000_SG2_AY566549 NCBI	BARTONSP	30,268	-97,766	Species_97	AY566549
	LIRCEOhardeni KNEEDEEP_000000_KN100_AY566537 NCBI	KNEEDEEP	29,909	-98,618	Species_98	AY566537
	LIRCEOhardeni KNEEDEEP_000000_KN13_AY566536 NCBI	KNEEDEEP	29,909	-98,618	Species_98	AY566536
	LIRCEOhardeni JACOBWEL_000000_JW98_AY566535 NCBI	JACOBWEL	30,118	-98,130	Species_99	AY566535
<i>Lirceolus hardeni</i> Lewis & Bowman, 1996	LIRCEOhardeni JACOBWEL_000000_JW97_AY566534 NCBI	JACOBWEL	30,118	-98,130	Species_99	AY566534
	LIRCEOhardeni RATTLESN_000000_RS72_AY566548 NCBI	RATTLESN	29,941	-97,923	Species_99	AY566548
	LIRCEOhardeni RATTLESN_000000_RS19_AY566547 NCBI	RATTLESN	29,941	-97,923	Species_99	AY566547
<i>Lirceolus pilus</i> (Steeves, 1968)	LIRCEOpilus LOSTMAPL_000000_LM80_AY566539 NCBI	LOSTMAPL	29,773	-99,567	Species_105	AY566539
	LIRCEOpilus LOSTMAPL_000000_LM79_AY566538 NCBI	LOSTMAPL	29,773	-99,567	Species_105	AY566538
	LIRCEOpilus VALDINA_000000_VF59_AY566545 NCBI	VALDINA	29,524	-99,405	Species_107	AY566545
	LIRCEOpilus VALDINA_000000_VF50_AY566553 NCBI	VALDINA	29,524	-99,405	Species_107	AY566553
<i>Lirceolus n. sp. (DANDRIDG)</i>	LIRCEOsp DANDRIDG_000000_DA78_AY566530 NCBI	DANDRIDG	29,783	-101,038	Species_108	AY566530
	LIRCEOsp DANDRIDG_000000_DA77_AY566529 NCBI	DANDRIDG	29,783	-101,038	Species_108	AY566529
<i>Lirceolus n. sp. (PRESERVIC)</i>	LIRCEOsp PRESERVIC_000000_PR57_AY566546 NCBI	PRESERVIC	29,879	-98,462	Species_109	AY566546
	LIRCEOsp PRESERVIC_000000_PR49_AY566545 NCBI	PRESERVIC	29,879	-98,462	Species_109	AY566545
<i>Lirceus bicuspoidatus</i> Hubricht & Mackin, 1949	LIRbicuspidatus COLLECHA_200902_2_YAF550	COLLECHA	35,478	-93,482	Species_110	NA
<i>Lirceus fontinalis</i> Rafinesque, 1820	LIRfontinalis BRYANSTA_201607_3_YAF547	BRYANSTA	38,077	-84,415	Species_110	NA
	LIRfontinalis BRYANSTA_201607_1ID_YAF494	BRYANSTA	38,077	-84,415	Species_117	NA
	LIRfontinalis BRYANSTA_201607_2_YAF532	BRYANSTA	38,077	-84,415	Species_117	NA
<i>Lirceus garmani</i> Hubricht & Mackin, 1949	LIRgarmani RDSIDEPA_201606_1ID_YAF483	RDSIDEPA	36,198	-95,150	Species_121	NA
	LIRgarmani RDSIDEPA_201606_2_YAF489	RDSIDEPA	36,198	-95,150	Species_121	NA
<i>Lirceus hargerii</i> Hubricht & Mackin, 1949	LIRhargerii JOHNSOSP_201606_1ID_YAF055	JOHNSOSP	36,708	-82,699	Species_82	NA
	LIRhargerii JOHNSOSP_201606_2ID_YAE997	JOHNSOSP	36,708	-82,699	Species_82	NA
<i>Lirceus hoppinae</i> (Faxon, 1889)	LIRhoppinae GREATHSP_201606_1ID_YAE996	GREATHSP	36,136	-94,203	Species_113	NA
	LIRhoppinae GREATHSP_201606_2_YAF054	GREATHSP	36,136	-94,203	Species_113	NA
<i>Lirceus lineatus</i> (Say, 1818)	LIRlineatus ELKHARTR_201609_1_YAF523	ELKHARTR	41,459	-85,523	Species_114	NA
	LIRlineatus ELKHARTR_201609_2_YAF526	ELKHARTR	41,459	-85,523	Species_114	NA
<i>Lirceus louisianae</i> (Mackin & Hubricht, 1938)	LIRlouisianae MAMMOTHS_201606_1ID_YAF493	MAMMOTHS	36,496	-91,535	Species_112	NA
	LIRlouisianae MAMMOTHS_201606_2_YAF487	MAMMOTHS	36,496	-91,535	Species_112	NA
<i>Lirceus megapodus</i> Hubricht & Mackin, 1949	LIRmegapodus ROCKSPRI_201606_1ID_YAF490	ROCKSPRI	37,760	-90,480	Species_118	NA
	LIRmegapodus ROCKSPRI_201606_2_YAF548	ROCKSPRI	37,760	-90,480	Species_119	NA
	LIRmegapodus ROCKSPRI_201606_3_YAF620	ROCKSPRI	37,760	-90,480	Species_119	NA
<i>Lirceus richardsonae</i> Hubricht & Mackin, 1949	LIRrichardsonae SIDE CUTS_201609_2_YAF528	SIDE CUTS	41,557	-83,672	Species_115	NA
	LIRrichardsonae SIDE CUTS_201609_1ID_YAF524	SIDE CUTS	41,557	-83,672	Species_115	NA
<i>Lirceus setosus n. sp. Lewis, in prep.</i>	LIRhargerii CROCKETT_201607_1ID_YAF511	CROCKETT	36,695	-83,108	Species_81	NA
	LIRhargerii CROCKETT_201607_2_YAF533	CROCKETT	36,695	-83,108	Species_81	NA
<i>Magniezia gardei</i> Magniez, 1978	MAGardei ERRACHID_201004_1ID_YLK985	ERRACHID	31,953	-4,452	Species_3	JQ921605
	MAGardei ERRACHID_201004_2_YCD285	ERRACHID	31,953	-4,452	Species_3	JQ921606
<i>Mesoasellus dybowskii</i> (Semenkevich, 1924)	MESodybowskii BAIKAL5_201409_1ID_YCM911	BAIKAL5	51,903	105,077	Species_128	KY623767
	MESodybowskii BAIKAL5B_201409_3_YCM939	BAIKAL5B	51,903	105,077	Species_128	KY623768
<i>Metastenasellus camerounensis</i>	METAScamerounensis EMANA_201107_1ID_YAC080	EMANA	3,919	11,520	Species_48	KY623769
<i>Zebaze Togouet, Boulanour, Njine &amp; Boutin, 2013</i>	METAScamerounensis EMANA_201107_2ID_YAC082	EMANA	3,919	11,520	Species_49	KY623770
	METAScamerounensis EMANA_201201_3ID_YAC242	EMANA	3,919	11,520	Species_49	KY623771
	METAScamerounensis DRAGARGE_201201_1_YAF739	DRAGARGE	3,898	11,518	Species_52	NA
	METAScamerounensis NGOUL1_201010_2ID_YAF714	NGOUL1	3,543	11,339	Species_52	NA
	METAScamerounensis NGOUL1_201010_1ID_YAF713	NGOUL1	3,543	11,339	Species_52	NA
	METAScamerounensis OYOMA_201107_2IDex_YAF769	OYOMA	3,906	11,395	Species_52	NA
	METAScamerounensis OYOMA_201107_1IDex_YAF740	OYOMA	3,906	11,395	Species_52	NA
	METASsp EBOG01_201107_2_YAC081	EBOG01	3,568	11,354	Species_50	KY623772
	METASsp NGOUL1_201107_1_YAF716	NGOUL1	3,543	11,339	Species_51	NA
<i>Metastenasellus n. sp. (ARBENIN)</i>	METASsp ARBENIN_201110_1_YAB952	ARBENIN	9,060	2,280	Species_45	KY623773
	METASsp ARBENIN_201110_2_YAC002	ARBENIN	9,060	2,280	Species_45	KY623774
<i>Mexistenasellus coahuila</i> Cole & Minckley, 1972	MEXIScoahuila ANTONIO_000000_0_AY566478 NCBI	ANTONIO	29,423	-98,488	Species_44	AY566478
<i>Parastenasellus n. sp. (LOUM1)</i>	PARAnsp LOUM1 LOUM1_201201_1ID_YAC240	LOUM1	4,431	9,433	Species_47	KY623775
<i>Parastenasellus n. sp. (MAKEPE2)</i>	PARAnsp MAKEPE2 MAKEPE2_201201_5ID_YAC241	MAKEPE2	4,068	9,721	Species_46	KY623776
<i>Proasellus acutianus</i> Argano & Henry, 1972	Pacutianus FIORARIV_201411_1ex_YAD073	FIORARIV	42,657	11,619	Species_274	KY623777
	Pacutianus FIORARIV_201411_2ex_YAD132	FIORARIV	42,657	11,619	Species_274	KY623778
	Pacutianus FIORARIV_201411_3ex_YAD075	FIORARIV	42,657	11,619	Species_274	KY623779
<i>Proasellus (coxalis) africanus</i> Monod, 1924	Pafricanus MGAIEZ_200805_1ID_YAB198	MGAIEZ	36,935	10,865	Species_282	KC610501
	Pafricanus MGAIEZ_200805_2_YAB317	MGAIEZ	36,935	10,865	Species_282	KY623949
	Pafricanus DADES_201004_1_YCD10	DADES	31,367	-5,975	Species_290	JQ921137
	Pafricanus DADES_201004_2ID_YCD918	DADES	31,367	-5,975	Species_290	KY623948
	Pafricanus NZALA_201004_1ID_YCD71	NZALA	32,627	-8,410	Species_290	JQ921138
	Pafricanus NZALA_201004_2_YCD76	NZALA	32,627	-8,410	Species_290	JQ921139
	Pafricanus NZALA_201004_3_YCD77	NZALA	32,627	-8,410	Species_290	JQ921140
<i>Proasellus albigensis</i> (Magniez, 1965)	Palbigensis LAUZIE_201409_10arn_YAF595	LAUZIE	43,910	2,234	Species_201	NA
	Palbigensis FOUSTER_200804_1_YAS227	FOUSTER	43,910	2,223	Species_202	JQ921028
	Palbigensis LAUZIE_200804_2ID_YAS280	LAUZIE	43,910	2,234	Species_202	JQ921029
	Palbigensis RAYSSAGU_197700_2ID_YAS330	RAYSSAGU	43,910	2,223	Species_202	JQ921030
<i>Proasellus albigensis</i> (Magniez, 1965)	Palbigensis RAYSSAGU_200806_4_YAC821	RAYSSAGU	43,910	2,223	Species_202	KY623781
<i>Proasellus amiterminus</i> Argano & Pesce, 1979	Pamiterminus MALPASSO_201212_1IDex_YAD130	MALPASSO	41,766	12,431	Species_276	KY623796
	Pamiterminus MALPASSO_201212_2ex_YAD288	MALPASSO	41,766	12,431	Species_276	KY623797
	Pamiterminus MALPASSO_201212_3ex_YAD156	MALPASSO	41,766	12,431	Species_276	KY623798
<i>Proasellus anophthalmus</i> (Karaman, 1934)	Panophthalmus SMOKOVAC_201107_2ID_YAC008	SMOKOVAC	42,477	19,306	Species_311	KY623825
	Panophthalmus SMOKOVAC_201107_1ID_YAB918	SMOKOVAC	42,477	19,306	Species_311	KY623824
	Pnsp Dlajam SPUZBOGI_201107_1ID_YAB819	SPUZBOGI	42,481	19,181	Species_311	KY623842
	Pnsp Dlajam SPUZBOGI_201107_2ID_YAB964	SPUZBOGI	42,481	19,181	Species_311	KY623843
	Psp Ribnice RIBNICE_200609_2_YAB595	RIBNICE	42,443	19,269	Species_311	KY623846
	Panophthalmus LJUTASPR_201107_1ID_YAB821	LJUTASPR	42,540	18,380	Species_312	KY623813
	Panophthalmus LJUTASPR_201107_2ID_YAB915	LJUTASPR	42,540	18,380	Species_312	KY623814
	Panophthalmus MANITPEC_201509_1ID_YAE736	MANITPEC	44,320	15,480	Species_314	NA
	Panophthalmus MANITPEC_201509_3_YAE841	MANITPEC	44,320	15,480	Species_314	NA
	Panophthalmus DOBRO_201107_1ID_YAB922	DOBRO	42,630	19,033	Species_315	KY623806
	Panophthalmus MALENIC2_201107_1ID_YAB921	MALENIC2	42,633	19,012	Species_315	KY623815
	Panophthalmus MALENIC2_201107_2ID_YAB818	MALENIC2	42,633	19,012	Species_315	KY623816

	Panophtalmus RIBNICE_200609_1FIDex_YAB154	RIBNICE	42,443	19,269	Species_316	KY623823
	Panophtalmus SPUZSPRI_201107_1ID_YAB844	SPUZSPRI	42,511	19,199	Species_317	KY623826
	Panophtalmus SPUZSPRI_201107_2ID_YAB917	SPUZSPRI	42,511	19,199	Species_317	KY623827
	Panophtalmus MORAICA_201107_1ID_YAC387	MORAICA	42,767	19,391	Species_318	KY623819
	Panophtalmus MORAICA_201107_2ID_YAC432	MORAICA	42,767	19,391	Species_318	KY623820
	Panophtalmus BERANE2_201107_1ID_YAB873	BERANE2	42,853	19,862	Species_319	KY623799
	Panophtalmus BERANE2_201107_2ID_YAC086	BERANE2	42,853	19,862	Species_319	KY623800
	Panophtalmus GOJAKOVI_201107_1ID_YAB923	GOJAKOVI	42,988	19,496	Species_320	KY623807
	Panophtalmus GOJAKOVI_201107_2ID_YAC478	GOJAKOVI	42,988	19,496	Species_320	KY623808
	Panophtalmus KRKASPRI_201107_1ID_YAB846	KRKASPRI	44,041	16,233	Species_324	KY623811
	Panophtalmus KRKASPRI_201107_2ID_YAC057	KRKASPRI	44,041	16,233	Species_324	KY623812
	Panophtalmus MILJACKA_201107_1ID_YAB843	MILJACKA	44,002	16,018	Species_324	KY623817
	Panophtalmus MILJACKA_201107_2ID_YAC003	MILJACKA	44,002	16,018	Species_324	KY623818
	Panophtalmus GRABASCE_201312_1ex_YAF648	GRABASCE	43,640	16,771	Species_325	NA
	Panophtalmus KRENICA_201107_1ID_YAB919	KRENICA	43,712	16,728	Species_325	KY623809
	Panophtalmus KRENICA_201107_2_YAC016	KRENICA	43,712	16,728	Species_325	KY623810
	Panophtalmus DATLO_201408_2ID_YAD149	DATLO	43,062	18,499	Species_328	KY623804
<i>Proasellus anophtalmus</i> (Karaman, 1934)	Panophtalmus DATLO_201408_1_YAD045	DATLO	43,062	18,499	Species_328	KY623803
	Panophtalmus DATLO_201408_3_YAD047	DATLO	43,062	18,499	Species_328	KY623805
	Panophtalmus BREGAVE_200609_1ID_YAF601	BREGAVE	43,072	18,037	Species_329	NA
	Panophtalmus BREGAVE_201009_1ID_YAB664	BREGAVE	43,072	18,037	Species_329	KY623801
	Panophtalmus BREGAVE_201107_1ID_YAB875	BREGAVE	43,072	18,037	Species_329	KC610502
	Panophtalmus BREGAVE_201009_2_YAB662	BREGAVE	43,072	18,037	Species_329	KY623802
	Panophtalmus ORLOVAC_201107_1ID_YAC413	ORLOVAC	43,672	16,774	Species_331	KY623821
	Panophtalmus STUDENCI_201404_1ID_YCM655	STUDENCI	43,180	17,608	Species_332	KY623828
	Panophtalmus STUDENCI_201404_2ID_YCM772	STUDENCI	43,180	17,608	Species_332	KY623829
	Panophtalmus STUDENCI_201404_3ID_YCM773	STUDENCI	43,180	17,608	Species_332	KY623830
	Panophtalmus PLATSPRI_201107_2ID_YAB910	PLATSPRI	42,600	18,231	Species_334	KY623822
	PspModrooko MODROOKO_200704_1FID_YAB869	MODROOKO	43,042	17,533	Species_327	KY623844
	PspModrooko MODROOKO_200704_2_YAB357	MODROOKO	43,042	17,533	Species_327	KY623845
<i>Proasellus anophtalmus dalmatinus</i> (Karaman, 1955)	Panophtalmusdalmatinus MOCILJSK_201107_1ID_YAB618	MOCILJSK	42,689	18,072	Species_335	KC610503
	Panophtalmusdalmatinus MOCILJSK_201107_2ID_YAB714	MOCILJSK	42,689	18,072	Species_335	KY623831
<i>Proasellus anophtalmus longisetus</i> (Remy, 1941)	Panophtalmuslongisetus PROVALID_201107_1ID_YAB617	PROVALID	43,286	18,267	Species_330	KY623834
	Panophtalmuslongisetus PROVALID_201107_2ID_YAB639	PROVALID	43,286	18,267	Species_330	KY623835
	Panophtalmuslongisetus PROVALID_201107_3ID_YAB822	PROVALID	43,286	18,267	Species_330	KY623836
	Panophtalmuslongisetus SNJETICA_201107_1ID_YAB611	SNJETICA	43,291	18,251	Species_330	KY623837
<i>Proasellus anophtalmus rascicus</i> Pljajic, 1972	Panophtalmusrascicus SOPOCANI_201107_1ID_YAB906	SOPOCANI	43,116	20,370	Species_321	KY623838
	Panophtalmusrascicus SOPOCANI_201107_2ID_YAC429	SOPOCANI	43,116	20,370	Species_322	KY623839
<i>Proasellus anophtalmus rhausinus</i> (Remy, 1941)	Panophtalmusrhausinus ROZAT_201107_1ID_YAB628	ROZAT	42,676	18,137	Species_333	KY623840
	Panophtalmusrhausinus ROZAT_201107_2ID_YAB632	ROZAT	42,676	18,137	Species_333	KY623841
	Panophtalmusdalmatinus AFFZIRA_201107_1ID_YAB615	ZIRA	42,922	17,848	Species_333	KY623832
	Panophtalmusdalmatinus AFFZIRA_201107_2ID_YAB631	ZIRA	42,922	17,848	Species_333	KY623833
<i>Proasellus aquaecalidae</i> (Racovitza, 1922)	Paquaealidae GEAUXCH_200811_1m_YAS564	GEAUXCH	42,934	-0.441	Species_220	JQ921031
	Paquaealidae PEAUXCH_200811_1_YAS541	PEAUXCH	42,939	-0.440	Species_220	JQ921035
	Paquaealidae PEAUXCH_200811_2_YAS614	PEAUXCH	42,939	-0.440	Species_220	JQ921036
	Paquaealidae OUEILNEE_200811_1ID_YAS562	OUEILNEE	43,150	-0.405	Species_221	JQ921032
	Paquaealidae OUEILNEE_200811_2_YAS575	OUEILNEE	43,150	-0.405	Species_221	JQ921033
	Paquaealidae OUEILNEE_200811_3_YAS647	OUEILNEE	43,150	-0.405	Species_221	JQ921034
<i>Proasellus aragonensis</i> Henry & Magniez, 1992	Paragonensis BURGUI_200906_2_YLK543	BURGUI	42,704	-1.016	Species_227	JQ921038
	Paragonensis BURGUI_200906_1ID_YAS873	BURGUI	42,704	-1.016	Species_228	JQ921037
	Paragonensis BURGUI_200906_3_YLK245	BURGUI	42,704	-1.016	Species_228	JQ921039
	Paragonensis BURGUI_200906_4_YLK550	BURGUI	42,704	-1.016	Species_228	KY623847
	Paragonensis BURGUI_200906_7m_YAC985	BURGUI	42,704	-1.016	Species_228	KY623848
	Paragonensis BURGUI_200906_8m_YAC986	BURGUI	42,704	-1.016	Species_228	KY623849
	Paragonensis BURGUI_201405_15am_YAF612	BURGUI	42,704	-1.016	Species_228	NA
<i>Proasellus arnavutovici</i> (Remy, 1932)	Parnautovicij OHRIDMAC_000000_aar1_DQ305138 NCBI	OHRIDMAC	41,037	20,719	Species_230	DQ305138
	Parnautovicij OHRIDMAC_000000_aar2_DQ305139 NCBI	OHRIDMAC	41,037	20,719	Species_230	DQ305139
	Parnautovicij LIVADIST_200909_1ID_YLK724	LIVADIST	41,007	20,756	Species_230	JQ921041
	Parnautovicij TRPEJCA_200909_1ID_YLK835	TRPEJCA	40,961	20,783	Species_230	JQ921042
	Parnautovicij VELIDAB_200909_1ID_YLK818	VELIDAB	41,007	20,756	Species_230	JQ921043
	Parnautovicij OSOJ1_200909_1ID_YLK907	OSOJ1	40,951	20,776	Species_231	JQ921040
<i>Proasellus arnavutovici elongatus</i> (Karaman, 1953)	Parnautovicij elongatus OHRIDMAC_000000_ael1_DQ305140 NCBI	OHRIDMAC	41,037	20,719	Species_231	DQ305140
	Parnautovicij elongatus OHRIDMAC_000000_ael2_DQ305141 NCBI	OHRIDMAC	41,037	20,719	Species_231	DQ305141
<i>Proasellus arthrotilus</i> (Braga, 1945)	Parthrodilus LEGACAO_201010_1ID_YCD651	LEGACAO	40,032	-8.470	Species_174	JQ921044
	Parthrodilus LEGACAO_201010_2_YCD541	LEGACAO	40,032	-8.470	Species_174	JQ921045
	Parthrodilus LEGACAO_201010_3_YCD588	LEGACAO	40,032	-8.470	Species_174	JQ921046
	Parthrodilus LEGACAO_201010_13_YAC602	LEGACAO	40,032	-8.470	Species_174	KY623851
	Parthrodilus LEGACAO_201010_14_YAC603	LEGACAO	40,032	-8.470	Species_174	KY623852
	Parthrodilus LEGACAO_201010_11_YAC596	LEGACAO	40,032	-8.470	Species_174	KY623850
	Parthrodilus LEGACAO_201110_12_YAC597	LEGACAO	40,032	-8.470	Species_174	KY623853
<i>Proasellus assaforensis</i> Afonso, 1988	Passaforensis ASSAFORA_201010_1ID_YCD846	ASSAFORA	38,909	-9.422	Species_206	JQ921047
	Passaforensis ASSAFORA_201010_2_YCD858	ASSAFORA	38,909	-9.422	Species_206	JQ921048
	Passaforensis ASSAFORA_201010_3_YCD924	ASSAFORA	38,909	-9.422	Species_206	JQ921049
	Passaforensis ASSAFORA_201010_15m_YAC975	ASSAFORA	38,909	-9.422	Species_206	KY623854
	Passaforensis ASSAFORA_201010_17m_YAC976	ASSAFORA	38,909	-9.422	Species_206	KY623855
<i>Proasellus (coxalis) banyulensis italicus</i> (Dudich, 1925)	Pbanyulensisitalicus VERDESCA_200905_1ID_YAB241	VERDESCA	41,468	12,904	Species_290	KY623950
	Pbanyulensisitalicus VERDESCA_200905_2_YAB283	VERDESCA	41,468	12,904	Species_290	KM249027
<i>Proasellus (coxalis) banyulensis tergestinus</i> Stoch, 1985	Pbanyulensis_tergestinus VOPICINA_201504_1_YAD795	VOPICINA	45,702	13,794	Species_290	NA
	Pbanyulensis_tergestinus VOPICINA_201504_2_YAD796	VOPICINA	45,702	13,794	Species_290	NA
	Pbanyulensis_tergestinus VOPICINA_201504_3_YAD786	VOPICINA	45,702	13,794	Species_290	NA
<i>Proasellus (coxalis) banyulensis polychaetus</i> (Dudich, 1925)	Pbanyulensis_polychaetus FIUSARNOI_201503_1ID_YAD952	FIUSARNO	40,810	14,606	Species_292	NA
	Pbanyulensis_polychaetus FIUSARNOI_201503_2_YAD794	FIUSARNO	40,810	14,606	Species_292	NA
	Psp_gr_coxalis FISELESP_201602_1_YAE844	FISELESP	40,600	15,120	Species_292	NA
	Psp_gr_coxalis FISELESP_201602_3_YAE809	FISELESP	40,600	15,120	Species_292	NA
	Pcoxalis FSARNOI_201503_3ex_YAE742	FSARNOI	40,813	14,616	Species_292	NA
<i>Proasellus bardaunii</i> Alouf, Henry & Magniez, 1982	Pbardaunii QATTINE_201010_1ID_YCD590	QATTINE	33,876	35,797	Species_309	JQ921050
	Pbardaunii QATTINE_201010_2_YCD742	QATTINE	33,876	35,797	Species_309	JQ921051
	Pbardaunii QATTINE_201010_3_YAB359	QATTINE	33,876	35,797	Species_309	KY623856
<i>Proasellus bellesii</i> Henry & Magniez, 1982	Pbellesij TURON_200910_1ID_YCD145	TURON	36,790	-4.941	Species_190	JQ921052
	Pbellesij TURON_200910_2_YCD215	TURON	36,790	-4.941	Species_190	JQ921053
	Pbellesij TURON_201010_3cyto_YAB017	TURON	36,790	-4.941	Species_190	JQ921054
<i>Proasellus beroni</i> Henry & Magniez, 1968	Pberonij BERNACCI_200810_3_YCD78	BERNACCI	42,546	9,137	Species_261	JQ921055
	Pberonij ORTAMIN_200810_1ex_YCD321	ORTAMIN	42,319	9,151	Species_262	JQ921059
	Pberonij ORTAMIN_200810_2ex_YAF646	ORTAMIN	42,319	9,151	Species_263	NA
	Pberonij CASALUNA_200810_2_YCD170	CASALUNA	42,388	9,259	Species_264	JQ921057
	Pberonij CASALUNA_200810_3_YAS626	CASALUNA	42,388	9,259	Species_264	JQ921058
	Pberonij ASINAIA_200810_1ID_YCD323	ASINAIA	42,493	9,295	Species_265	JQ921060
	Pberonij CASAMOZA_200810_1ID_YAS649	CASAMOZA	42,519	9,442	Species_265	JQ921056
<i>Proasellus beticus</i> Henry & Magniez, 1992	Pbeticus SUMIDORS_200910_1ID_YCD35	SUMIDORS	38,880	-0.689	Species_238	JQ921061
	Pbeticus SUMIDORS_200910_2_YCD31	SUMIDORS	38,880	-0.689	Species_238	JQ921062
	Pbeticus SUMIDORS_200910_3_YCD193	SUMIDORS	38,880	-0.689	Species_238	JQ921063
	Pbeticus SUMIDORS_201102_16m_YAC993	SUMIDORS	38,880	-0.689	Species_238	KY623857
	Pbeticus SUMIDORS_201102_17m_YAC994	SUMIDORS	38,880	-0.689	Species_238	KY623858
<i>Proasellus boui</i> Henry & Magniez, 1969	Pbouij SAUVEGAR_000000_1ID_YAS147	SAUVEGAR	43,941	3,950	Species_152	JQ921064



	PboujSAUVEGAR_000000_2ex_YAS37	SAUVEGAR	43,941	3,950	Species_152	JQ921065
	PboujSAUVEGAR_000000_3ex_YAS47	SAUVEGAR	43,941	3,950	Species_152	JQ921066
	PboujLAVABREI_201609_1ID_YAF722	LAVABREI	44,168	4,053	Species_153	NA
	PboujLAVABREI_201609_3_YAF727	LAVABREI	44,168	4,053	Species_153	NA
	PboujSTPCOSTE_201705_2ID_YAG045	STPCOSTE	44,164	3,974	Species_153	NA
	PboujFALGIERE_201407_F5ex_YCM640	FALGIERE	44,128	3,892	Species_154	KY623859
	PboujHBUGUEIR_201609_1ID_YAF580	HBUGUEIR	44,077	4,108	Species_154	NA
	PboujHCOLLIAS_201605_2_YAF039	HCOLLIAS	43,954	4,483	Species_154	NA
	PboujMOULINDU_201407_C5ID_YCM649	MOULINDU	44,197	3,702	Species_154	KY623861
	PboujMOULINDU_201407_B6ex_YCM824	MOULINDU	44,197	3,702	Species_154	KY623860
	PboujSTPRIVAT_201705_1ID_YAG115	STPRIVAT	44,266	3,870	Species_155	NA
	PboujSTPRIVAT_201705_3_YAG187	STPRIVAT	44,266	3,870	Species_155	NA
<i>Proasellus burgundus</i> Henry & Magniez, 1969	Pburgundus BEZE_201707_7ID_YAG416	BEZE	47,471	5,269	Species_229	NA
	Pburgundus BEZE_201707_8_YAG417	BEZE	47,471	5,269	Species_229	NA
<i>Proasellus cantabricus</i> Henry & Magniez, 1968	Pcantabricus BECERRAL_201110_1ID_YAC115	BECERRAL	43,193	-3,585	Species_254	KY623862
	Pcantabricus BECERRAL_201110_3ID_YAC118	BECERRAL	43,193	-3,585	Species_254	KY623863
	Pcantabricus CHICO_201110_1ID_YAC113	CHICO	43,191	-3,581	Species_254	KY623864
	Pcantabricus NACGANDA_201110_1ID_YAC116	NACGANDA	43,194	-3,582	Species_254	KY623865
	Pcantabricus RESCANO_201010_21am_YAB771	RESCANO	43,294	-3,699	Species_254	KY623866
	Pcantabricus RESCANO_201102_3am_YAD340	RESCANO	43,294	-3,699	Species_254	KY623867
	Pcantabricus RESCANO_200906_1ID_YAS976	RESCANO	43,294	-3,699	Species_256	JQ921067
	Pcantabricus RESCANO_200906_2_YLK506	RESCANO	43,294	-3,699	Species_256	JQ921068
	Pcantabricus RESCANO_200906_3_YLK158	RESCANO	43,294	-3,699	Species_256	JQ921069
	Pcantabricus AFF CUBIO_201110_1ID_YAC111	CUBIO	43,276	-3,620	Species_254	KY623868
	Pcantabricus AFF CUBIO_201110_5_YAC398	CUBIO	43,276	-3,620	Species_254	KY623869
	Pcantabricus AFF CUBIO_201405_17am_YAD341	CUBIO	43,276	-3,620	Species_254	KY623870
<i>Proasellus cantabricus</i> Henry & Magniez, 1968	Pcantabricus AFF CUBIO_201405_21am_YAD342	CUBIO	43,276	-3,620	Species_254	KY623871
	PspCuvera1 CUVERA_201108_2_YAF708	CUVERA	43,255	-3,604	Species_254	NA
	PspCuvera1 CUVERA_201108_5_YAC017	CUVERA	43,255	-3,604	Species_254	KY623872
	PspCuvera1 CUVERA_201108_1_YAF647	CUVERA	43,255	-3,604	Species_254	NA
	PspCuvera1 CUVERA_201108_3_YAF705	CUVERA	43,255	-3,604	Species_254	NA
<i>Proasellus cavaticus</i> (Leydig, 1871)	Pcavaticus BURKH3_200905_1ID_YCD875	BURKH3	48,095	7,593	Species_368	KC610465
	Pcavaticus BURKH3_200905_2_YLK670	BURKH3	48,095	7,593	Species_368	JQ921078
	Pcavaticus BURKH3_200905_3_YLK587	BURKH3	48,095	7,593	Species_368	JQ921079
	Pcavaticus CATIEFEN_201605_2ID_YAF119	CATIEFEN	48,222	7,165	Species_368	NA
	Pcavaticus CATIEFEN_201605_3ID_YAF120	CATIEFEN	48,222	7,165	Species_368	NA
	Pcavaticus ROSSFELD_200911_1ID_YLK832	ROSSFELD	48,334	7,631	Species_368	JQ921086
	Pcavaticus SARREGUE_200210_1_YML131 INCONNU	SARREGUE	49,115	7,038	Species_368	KC610482
	Pcavaticus ARBOUSSE_000000_1ID_YCD230	ARBOUSSE	43,823	3,144	Species_369	JQ921087
	Pcavaticus BADROEID_201009_1ex_YAC415	BADROEID	46,703	7,848	Species_369	KY623873
	Pcavaticus BASCHGGR_201506_1ID_YAE740	BASCHGGR	47,250	9,670	Species_369	NA
	Pcavaticus BASCHGGR_201506_3_YAE710	BASCHGGR	47,250	9,670	Species_369	NA
<i>Proasellus cavaticus</i> (Leydig, 1871)	Pcavaticus BELLEGAR_200902_1ID_YAS684	BELLEGAR	46,611	7,276	Species_369	JQ921073
	Pcavaticus BELLEGAR_200902_2_YAS660	BELLEGAR	46,611	7,276	Species_369	JQ921074
	Pcavaticus BELLEGAR_200902_3_YAS661	BELLEGAR	46,611	7,276	Species_369	JQ921075
	Pcavaticus BINGHOHL_201603_1ex_YAF034	BINGHOHL	49,812	11,216	Species_369	NA
	Pcavaticus BOUDUMON_201509_4am_TRINITYDN9096c0g21	BOUDUMON	46,066	4,556	Species_369	NA
	Pcavaticus BOUDUMON_201507_1ID_YAD862	BOUDUMON	46,066	4,556	Species_369	NA
	Pcavaticus BOUDUMON_201507_2_YAE725	BOUDUMON	46,066	4,556	Species_369	NA
	Pcavaticus CENTFONT_200908_1ex_YAC451	CENTFONT	43,759	3,623	Species_369	KY623875
	Pcavaticus CRAMANS_200808_4ID_YCD229	CRAMANS	46,136	5,476	Species_369	JQ921088
	Pcavaticus CROCHET_201307_1ID_YCM771	CROCHET	45,913	5,404	Species_369	KY623876
	Pcavaticus CROCHET_201307_2_YCM408	CROCHET	45,913	5,404	Species_369	KY623877
	Pcavaticus CRUPET_200807_3ID_YAS371	CRUPET	50,354	4,961	Species_369	JQ921080
	Pcavaticus CRUPET_200807_4_YAS372	CRUPET	50,354	4,961	Species_369	JQ921081
	Pcavaticus ELM_200904_1ID_YAS742	ELM	51,805	-3,140	Species_369	JQ921082
	Pcavaticus FONTANIL_200908_1ex_YAC418	FONTANIL	43,753	3,623	Species_369	KY623878
	Pcavaticus GIRSTERK_201609_2ID_YAF650	GIRSTERK	49,790	6,500	Species_369	NA
	Pcavaticus GRASCENS_201603_3ex_YAF077	GRASCENS	43,557	2,914	Species_369	NA
	Pcavaticus GRASCENS_201603_2ex_YAF070	GRASCENS	43,557	2,914	Species_369	NA
	Pcavaticus GROTTAZE_201512_1ID_YAE384	GROTTAZE	46,439	4,761	Species_369	NA
	Pcavaticus GROTTAZE_201512_2_YAE681	GROTTAZE	46,439	4,761	Species_369	NA
	Pcavaticus GROTTAZE_201512_3_YAE432	GROTTAZE	46,439	4,761	Species_369	NA
	Pcavaticus HOMEDE_200908_1ex_YAC419	HOMEDE	44,077	3,060	Species_369	KY623879
	Pcavaticus ISIS_200901_1ex_YAS807	ISIS	43,984	3,598	Species_369	JQ921083
	Pcavaticus JALLIFIE_000000_36ml_YAC701	JALLIFIE	45,075	5,385	Species_369	KY623880
	Pcavaticus JALLIFIE_000000_AS13_YAS13 R3FPCOI	JALLIFIE	45,075	5,385	Species_369	JQ921085
	Pcavaticus LASFONS2_201407_D3ex_YCM632	LASFONS2	43,968	3,550	Species_369	KY623882
	Pcavaticus LASFONS2_201407_G1ex_YCM585	LASFONS2	43,968	3,550	Species_369	KY623883
	Pcavaticus LAUILOCH_201409_1ID_YAD289	LAUILOCH	46,992	8,717	Species_369	KY623884
	Pcavaticus LAUILOCH_201409_3_YAD111	LAUILOCH	46,992	8,717	Species_369	KY623885
	Pcavaticus MILANDRE_201411_1ID_YAD048	MILANDRE	47,485	7,016	Species_369	KY623886
	Pcavaticus MINDOHL_201410_2ID_YAF651	MINDOHL	49,530	5,850	Species_369	NA
	Pcavaticus MITTENA_200906_1ID_YCD380	MITTENA	47,427	11,259	Species_369	JQ921089
	Pcavaticus MITTENA_200910_1ID_YCD418	MITTENA	47,427	11,259	Species_369	JQ921090
	Pcavaticus MITTENA_200910_3ex_YCD699	MITTENA	47,427	11,259	Species_369	JQ921091
<i>Proasellus cavaticus</i> (Leydig, 1871)	Pcavaticus MONTJOU1_201007_1IDex_YAB956	MONTJOU1	44,502	5,089	Species_369	KC610479
	Pcavaticus MUHLBAGR_201603_3_YAE764	MUHLBAGR	49,023	11,618	Species_369	NA
	Pcavaticus MUHLBAGR_201603_1ID_YAE808	MUHLBAGR	49,023	11,618	Species_369	NA
	Pcavaticus NEBLONCO_200807_3ex_YAS497	NEBLONCO	50,414	5,457	Species_369	JQ921092
	Pcavaticus OGOF_200308_38ml_YAS6 R1FSCOI	OGOF	51,824	-3,673	Species_369	KC610480
	Pcavaticus OGOF_200308_4ID_YAS131	OGOF	51,824	-3,673	Species_369	JQ921093
	Pcavaticus OGOF_200402_1IDex_YAS86	OGOF	51,824	-3,673	Species_369	JQ921100
	Pcavaticus OGOF_200402_2IDex_YAS132	OGOF	51,824	-3,673	Species_369	KC610481
	Pcavaticus ROCHE_200902_1_YAS643	ROCHE	45,072	5,498	Species_369	JQ921097
	Pcavaticus SAUVEGAR_000000_AS2_YAS8 R1FSCOI	SAUVEGAR	43,941	3,950	Species_369	KC610484
	Pcavaticus SAUVEGAR_000000_10ID_YAS130	SAUVEGAR	43,941	3,950	Species_369	JQ921099
	Pcavaticus SAUVEGAR_000000_9ID_YAS129	SAUVEGAR	43,941	3,950	Species_369	JQ921098
	Pcavaticus STEREINE_200807_1ID_YAS345	STEREINE	48,645	5,942	Species_369	JQ921094
	Pcavaticus STEREINE_200807_2_YAS346	STEREINE	48,645	5,942	Species_369	JQ921095
	Pcavaticus STEREINE_200807_3_YAS464	STEREINE	48,645	5,942	Species_369	JQ921096
	Pcavaticus STMATERN_201409_1_YAE386	STMATERN	48,366	7,622	Species_369	NA
	Pcavaticus SWILDON_200904_2ID_YAS743	SWILDON	51,259	-2,673	Species_369	JQ921101
	Pcavaticus THAISGR_201412_1ID_YAD055	THAISGR	45,059	5,248	Species_369	KY623888
	Pcavaticus THAISGR_201412_2_YAD056	THAISGR	45,059	5,248	Species_369	KY623889
	Pcavaticus TYNINGS_200612_1_YCD281	TYNINGS	51,304	-2,760	Species_369	JQ921102
	Pcavaticus UNTERACH_201009_2ex_YCD665	UNTERACH	47,792	13,485	Species_369	JQ921103
	Pcavaticus VIS_201106_1ID_YAE388	VIS	43,900	3,483	Species_369	NA
	Pcavaticus WALDEGG_201009_1ex_YCD642	WALDEGG	46,715	7,832	Species_369	JQ921104
	Pcavaticus WEISSEN1_201009_1ID_YCD697	WEISSEN1	47,789	13,552	Species_369	JQ921105
	Pcavaticus WEISSEN1_201009_2_YCD698	WEISSEN1	47,789	13,552	Species_369	JQ921106
	Pcavaticus WEISSEN1_201009_3_YCD704	WEISSEN1	47,789	13,552	Species_369	JQ921107
	Pcavaticus BELMONT_200707_1ID_YAS310	BELMONT	47,009	5,595	Species_370	JQ921076
	Pcavaticus CABORDES_201202_1ID_YAC416	CABORDES	47,207	6,018	Species_370	KY623874

Pcavaticus JOLIOT_201202_1ID_YAC420	JOLIOT	47,201	6,015	Species_370	KY623881
Pcavaticus MONTBAR_201409_1ID_YAE519	MONTBAR	47,011	5,630	Species_370	NA
Pcavaticus FRAITE_200711_1_YAE625	FRAITE	46,525	5,734	Species_371	JQ921072
Pcavaticus BORCASSO_200711_1ID_YAS128	BORCASSO	46,737	5,642	Species_372	JQ921077
Pcavaticus CORVEISS_200711_1_YAB663	CORVEISS	46,243	5,484	Species_373	KM249024
Pcavaticus QUINQUE_200908_1ID_YAC421	QUINQUE	46,418	5,540	Species_373	KY623887
Pcavaticus TOUVIEGR_201507_1ex_YAE387	TOUVIEGR	46,522	5,703	Species_373	NA
Pcavaticus TOUVIEGR_201507_2ex_YAE587	TOUVIEGR	46,522	5,703	Species_373	NA
Pcavaticus TOUVIEGR_201507_3ex_YAE588	TOUVIEGR	46,522	5,703	Species_373	NA
Pcavaticus VALFIN_200300_11ml_YAD301	VALFIN	46,570	5,520	Species_373	KY623890
PcomasiCF AQUALA_200910_1ID_YCD203	AQUALA	38,060	-2,820	Species_183	JQ921108
PcomasiCF AQUALA_200910_2_YCD322	AQUALA	38,060	-2,820	Species_183	JQ921109
PcomasiCF AQUALA_200910_3_YCD400	AQUALA	38,060	-2,820	Species_183	JQ921110
PcomasiCF COTILLAS_201010_2ex_YCD794	COTILLAS	38,433	-2,529	Species_184	KY623780
PgourbaultaeCF CAUTAB_200910_1ex_YLK908	CAUTAB	39,144	-1,047	Species_193	JQ921111
PgourbaultaeCF CAUTAB_200910_4ex_YLK984	CAUTAB	39,144	-1,047	Species_193	JQ921112
PmonodiCF KOKKINI_200903_1ID_YAB240	KOKKINI	39,755	19,925	Species_233	KY624038
PmonodiCF KOKKINI_200903_2_YAB305	KOKKINI	39,755	19,925	Species_233	KY624039
PruffoiCF PELTUSU_200905_1ID_YAB197	PELTUSU	40,728	8,559	Species_177	KY624059
PruffoiCF PELTUSU_200905_2_YAB567	PELTUSU	40,728	8,559	Species_177	KY624060
PruffoiCF MENGIANU_201003_1ID_YAB310	MENGIANU	39,795	9,007	Species_180	KY624057
PruffoiCF MENGIANU_201003_2_YAB594	MENGIANU	39,795	9,007	Species_180	KY624058
PstockiCF SAJA_201010_1ID_YCD845	SAJA	43,121	-4,292	Species_251	JQ921116
PstockiCF SAJA_201010_2_YCD855	SAJA	43,121	-4,292	Species_251	JQ921117
PstockiCF SAJA_201010_3_YCD913	SAJA	43,121	-4,292	Species_251	JQ921118
Pchappuisi ARTZEGI_200906_1ID_YAS874	ARTZEGI	43,019	-2,754	Species_249	JQ921120
Pchappuisi ARTZEGI_200906_2_YLK251	ARTZEGI	43,019	-2,754	Species_250	JQ921121
Pchappuisi ARTZEGI_200906_3_YLK324	ARTZEGI	43,019	-2,754	Species_250	JQ921122
Pchauvinij HAYTZA_200811_1ID_YAS590	HAYTZA	43,222	-1,345	Species_212	JQ921123
Pchauvinij HAYTZA_200811_2_YAS561	HAYTZA	43,222	-1,345	Species_212	JQ921124
Pchauvinij HAYTZA_200811_3_YAS591	HAYTZA	43,222	-1,345	Species_212	JQ921125
Pchauvinij OYANBELT_200811_1_YAS550	OYANBELT	43,059	-1,228	Species_213	JQ921126
Pchauvinij OYANBELT_200811_2_YAS592	OYANBELT	43,059	-1,228	Species_213	JQ921127
Pchauvinij OYANBELT_200811_3ID_YAS559	OYANBELT	43,059	-1,228	Species_213	JQ921128
Pclaudej VINCENT_201011_1ID_YCD783	VINCENT	44,459	1,300	Species_266	JQ921129
Pclaudej VINCENT_201011_2_YCD795	VINCENT	44,459	1,300	Species_266	JQ921130
Pcoiffaiti GROTLULU_201601_1ID_YAE389	GROTLULU	43,464	-1,444	Species_164	NA
Pcoiffaiti GROTLULU_201601_3_YAE421	GROTLULU	43,464	-1,444	Species_164	NA
Pcoiffaiti MIOSLEY_201505_1ex_YAE404	MIOSLEY	44,593	-0,935	Species_164	NA
Pcoiffaiti NAVALLO_200906_1ID_YAS881	NAVALLO	43,561	-1,052	Species_164	JQ921131
Pcoiffaiti NAVALLO_200906_2ID_YAS955	NAVALLO	43,561	-1,052	Species_164	JQ921132
Pcoiffaiti NAVALLO_200906_3_YAS953	NAVALLO	43,561	-1,052	Species_164	JQ921133
Pcoiffaiti POURRUT_201102_1IDex_YAB163	POURRUT	43,858	-1,358	Species_164	KY623891
Pcoiffaiti POURRUT_201102_2ex_YAB326	POURRUT	43,858	-1,358	Species_164	KY623892
Pcoiffaiti PPEYRE_201110_1_YAC012	PPEYRE	43,545	-1,042	Species_164	KY623893
Pcoiffaiti PPEYRE_201110_8_YAC979	PPEYRE	43,545	-1,042	Species_164	KY623894
Pcoiffaiti PPEYRE_201110_9_YAC980	PPEYRE	43,545	-1,042	Species_164	KY623895
Pcomasi VELEZ_200910_1_YLK909	VELEZ	36,852	-4,135	Species_192	JQ921134
Pcomasi VELEZ_200910_2ID_YCD13	VELEZ	36,852	-4,135	Species_192	JQ921135
Pcomasi VELEZ_200910_3_YCD95	VELEZ	36,852	-4,135	Species_192	JQ921136
Pcoxalis SEJNANES_200707_1ID_YAB307	SEJNANES	37,173	9,448	Species_282	KY623941
Pcoxalis SEJNANES_200707_2_YAB306	SEJNANES	37,173	9,448	Species_282	KY623942
Pcoxalis FANEROPD_201610_2ID_YAF283	FANEROPD	35,090	24,850	Species_283	NA
Pcoxalis FANEROPD_201610_3_YAF284	FANEROPD	35,090	24,850	Species_283	NA
Pcoxalis KPM19_201510_1ID_YAE880	KPM19	35,049	25,223	Species_283	NA
Pcoxalis KPM19_201510_2_YAE877	KPM19	35,049	25,223	Species_283	NA
Pcoxalis PC33ZARO_201610_1ID_YAF273	PC33ZARO	35,140	24,912	Species_283	NA
Pcoxalis PC33ZARO_201610_3_YAF274	PC33ZARO	35,140	24,912	Species_283	NA
Psp1 PC33ZARO_201500_1_YAE682	PC33ZARO	35,140	24,912	Species_283	NA
Pcoxalis CANNYDRIL_201503_3_YAD789	CANNYDRIL	38,712	20,708	Species_287	NA
Pcoxalis CANNYDRIL_201503_2ID_YAD953	CANNYDRIL	38,712	20,708	Species_287	NA
Pcoxalis CANNYDRIL_201503_1ID_YAD788	CANNYDRIL	38,712	20,708	Species_287	NA
Pcoxalis CHRISII_201203_1ID_YAE127	CHRISII	39,590	19,901	Species_287	NA
Pcoxalis CHRISII_201203_2_YAE225	CHRISII	39,590	19,901	Species_287	NA
Pcoxalis DROSATO2_201203_3_YAD352	DROSATO2	39,735	19,741	Species_287	KY623910
Pcoxalis DROSATO2_201203_1ID_YAD377	DROSATO2	39,735	19,741	Species_287	KY623909
Pcoxalis GRMEGALI_201203_1ID_YAE128	GRMEGALI	39,777	19,891	Species_287	NA
Pcoxalis GRMEGALI_201203_3_YAE241	GRMEGALI	39,777	19,891	Species_287	NA
Pcoxalis KALLIGONI_201503_1ID_YAD797	KALLIGONI	38,820	20,711	Species_287	NA
Pcoxalis KALLIGONI_201503_3_YAD783	KALLIGONI	38,820	20,711	Species_287	NA
Pcoxalis KOSTANO_201203_2ID_YAD331	KOSTANO	39,211	20,718	Species_287	KY623959
Pcoxalis KOSTANO_201203_1ID_YAD389	KOSTANO	39,211	20,718	Species_287	NA
Pcoxalis MPAKSES_201203_1ID_YAD338	MPAKSES	39,343	20,436	Species_287	KY623933
Pcoxalis MPAKSES_201203_2_YAD353	MPAKSES	39,343	20,436	Species_287	KY623934
Pcoxalis PANAGHIAI_201503_1ID_YAD785	PANAGHIAI	38,739	20,606	Species_287	NA
Pcoxalis PANAGHIAI_201503_2ID_YAD800	PANAGHIAI	38,739	20,606	Species_287	NA
Pcoxalis POZZANOR_201203_1ID_YAD339	POZZANOR	39,816	19,843	Species_287	KY623935
Pcoxalis POZZANOR_201203_2_YAD355	POZZANOR	39,816	19,843	Species_287	KY623936
Pcoxalis SORKRAN2_201303_1ID_YAD378	SORKRAN2	38,167	20,510	Species_287	KY623944
Pcoxalis SORKRAN2_201303_2_YAD357	SORKRAN2	38,167	20,510	Species_287	KY623945
Psp_gr_coxalis ZOODPYGI_201503_1ID_YAD780	ZOODPYGI	38,816	20,705	Species_287	NA
Psp_gr_coxalis ZOODPYGI_201503_2ID_YAD781	ZOODPYGI	38,816	20,705	Species_287	NA
Pcoxalis MONREALE_201110_1ID_YAD350	MONREALE	38,010	13,251	Species_288	KY623931
Pcoxalis MONREALE_201110_2_YAD332	MONREALE	38,010	13,251	Species_288	KY623932
Pcoxalis PIANADEG_201305_1ID_YAE157	PIANADEG	37,985	13,273	Species_288	NA
Pcoxalis PIANADEG_201305_3_YAE140	PIANADEG	37,985	13,273	Species_288	NA
Pcoxalis ACQUAGRA_201407_1ID_YAE227	ACQUAGRA	37,952	13,612	Species_289	NA
Pcoxalis ACQUAGRA_201407_3_YAE124	ACQUAGRA	37,952	13,612	Species_289	NA
Pcoxalis ANTONIO_201405_1ID_YAE143	ANTONIO	37,718	13,482	Species_289	NA
Pcoxalis ANTONIO_201405_3_YAE163	ANTONIO	37,718	13,482	Species_289	NA
Pcoxalis FAVARA_201203_1ID_YAD346	FAVARA	37,825	13,811	Species_289	KY623913
Pcoxalis FAVARA_201203_2_YAD333	FAVARA	37,825	13,811	Species_289	KY623914
Psp_gr_coxalis ABBCERDA_201505_1ID_YAE835	ABBCERDA	37,910	13,830	Species_289	NA
Psp_gr_coxalis ABBCERDA_201505_2_YAE765	ABBCERDA	37,910	13,830	Species_289	NA
Pcoxalis AMUNT_200803_1_YAS229	AMUNT	42,469	3,092	Species_290	JQ921173
Pcoxalis AMUNT_200803_3_YCD316	AMUNT	42,469	3,092	Species_290	JQ921174
Pcoxalis BEDIZZOL_201608_1ID_YAF447	BEDIZZOL	45,498	10,402	Species_290	NA
Pcoxalis BEDIZZOL_201608_3_YAF655 igHCO2198	BEDIZZOL	45,498	10,402	Species_290	NA
Pcoxalis BELENICI_201107_1ID_YAB905	BELENICI	42,810	18,027	Species_290	KY623901
Pcoxalis BELENICI_201107_2_YAB962	BELENICI	42,810	18,027	Species_290	KY623902
Pcoxalis BENAJO_200910_7m_YAC973	BENAJO	36,754	-5,202	Species_290	KY623951
Pcoxalis BENAJO_200910_8m_YAC974	BENAJO	36,754	-5,202	Species_290	KY623952
Pcoxalis BENAJO1_200910_1ID_YCD238	BENAJO1	36,754	-5,202	Species_290	JQ921143
Pcoxalis BENAJO1_200910_2_YCD231	BENAJO1	36,754	-5,202	Species_290	JQ921144

*Proasellus coxalis* (Dolfs, 1892)

Pcoxalis BUNAR_200402_1ID_YLK400	BUNAR	42,712	17,892	Species_290	JQ921151
Pcoxalis BUNAR_200402_2_YLK581	BUNAR	42,712	17,892	Species_290	JQ921152
Pcoxalis BUNAR_200402_3_YAB013	BUNAR	42,712	17,892	Species_290	JQ921153
Pcoxalis CABRIER_201006_3_YAC712	CABRIER	43,745	3,561	Species_290	NA
Pcoxalis CABRIER_201006_1ID_YAB352	CABRIER	43,745	3,561	Species_290	KY623903
Pcoxalis CALPONT_200904_1ID_YAS738	CALPONT	43,827	11,302	Species_290	JQ921146
Pcoxalis CALPONT_200904_2_YAS798	CALPONT	43,827	11,302	Species_290	JQ921147
Pcoxalis CARLOMAG_201107_1ID_YAE150	CARLOMAG	40,136	18,481	Species_290	NA
Pcoxalis CARLOMAG_201107_2_YAE126	CARLOMAG	40,136	18,481	Species_290	NA
Pcoxalis CASAMOZA_200810_1ID_YAS469	CASAMOZA	42,519	9,442	Species_290	JQ921148
Pcoxalis CASAMOZA_200810_2_YAS406	CASAMOZA	42,519	9,442	Species_290	JQ921149
Pcoxalis CASAMOZA_200810_3_YAS470	CASAMOZA	42,519	9,442	Species_290	JQ921150
Pcoxalis CENTFONT_201006_1ID_YAC469	CENTFONT	43,759	3,623	Species_290	KY623904
Pcoxalis COFANO_200904_1ID_YAB242	COFANO	38,019	12,515	Species_290	KY623905
Pcoxalis COFANO_200904_2_YAB284	COFANO	38,019	12,515	Species_290	KY623906
Pcoxalis COPERTO_201004_1ID_YAE164	COPERTO	41,518	12,993	Species_290	NA
Pcoxalis COPERTO_201004_3_YAE148	COPERTO	41,518	12,993	Species_290	NA
Pcoxalis DOVOZNI_200510_1ID_YAB244	DOVOZNI	45,532	13,605	Species_290	KY623907
Pcoxalis DOVOZNI_200510_2_YAB286	DOVOZNI	45,532	13,605	Species_290	KY623908
Pcoxalis DOVOZNI2_201405_5arn_YAF687	DOVOZNI2	45,523	13,619	Species_290	NA
Pcoxalis DOVOZNI2_201405_6arn_YAF688	DOVOZNI2	45,523	13,619	Species_290	NA
Pcoxalis FALCONIE_201003_1ID_YAB249	FALCONIE	38,711	13,196	Species_290	KY623911
Pcoxalis FALCONIE_201003_2_YAB309	FALCONIE	38,711	13,196	Species_290	KY623912
Pcoxalis FGARIGLI_201203_1_YAD334	FGARIGLI	41,393	13,876	Species_290	KY623915
Pcoxalis FGARIGLI_201203_2_YAD347	FGARIGLI	41,393	13,876	Species_290	KY623916
Pcoxalis FSARNOI_201503_1ex_YAE141	FSARNOI	40,813	14,616	Species_290	NA
Pcoxalis FSARNOI_201503_2ex_YAE125	FSARNOI	40,813	14,616	Species_290	NA
Pcoxalis GAJA_200804_1ID_YLK575	GAJA	43,197	2,420	Species_290	JQ921171
Pcoxalis HPAULHAN_201605_1ex_YAF026	HPAULHAN	43,531	3,477	Species_290	NA
Pcoxalis ILERHINI_201409_3_YAD051	ILERHINI	48,252	7,675	Species_290	KY623921
Pcoxalis ILERHINI_201409_2_YAD052	ILERHINI	48,252	7,675	Species_290	KY623920
Pcoxalis IRCHIOSP_201304_1ID_YAE223	IRCHIOSP	41,873	15,810	Species_290	NA
Pcoxalis ISALLE_201106_2_YAD345	ISALLE	40,368	9,386	Species_290	KY623922
Pcoxalis ISALLE_201106_1ID_YAD390	ISALLE	40,368	9,386	Species_290	NA
Pcoxalis IZVORPUC_201006_1ID_YAB245	IZVORPUC	45,176	13,777	Species_290	KY623923
Pcoxalis IZVORPUC_201006_2_YAB282	IZVORPUC	45,176	13,777	Species_290	KY623924
Pcoxalis JAREK_200708_1ID_YAB248	JAREK	45,476	13,619	Species_290	KY623925
Pcoxalis LOPAR_200508_1ID_YLK449	LOPAR	44,831	14,732	Species_290	JQ921154
Pcoxalis LOPAR_200508_2_YLK595	LOPAR	44,831	14,732	Species_290	JQ921155
Pcoxalis MACCHIA_201005_1ID_YAE145	MACCHIA	41,999	11,979	Species_290	NA
Pcoxalis MACCHIA_201005_2_YAE129	MACCHIA	41,999	11,979	Species_290	NA
Pcoxalis MACCHIA_201005_3_YAE294	MACCHIA	41,999	11,979	Species_290	NA
Pcoxalis MAKIRINA_200506_2_YAB332	MAKIRINA	43,805	15,650	Species_290	KY623930
Pcoxalis MARCIA_201204_1ID_YAE154	MARCIA	42,007	13,023	Species_290	NA
Pcoxalis MARCIA_201204_2_YAE232	MARCIA	42,007	13,023	Species_290	NA
Pcoxalis MILLAS_200803_1ID_YAS468	MILLAS	42,681	2,706	Species_290	JQ921156
Pcoxalis NAUERNAA_200201_1ID_YLK319	NAUERNAA	52,464	4,764	Species_290	JQ921157
Pcoxalis ORFES_200906_1ID_YAS877	ORFES	42,171	2,869	Species_290	JQ921158
Pcoxalis ORFES_200906_2_YLK163	ORFES	42,171	2,869	Species_290	JQ921159
Pcoxalis PICLASSE_201403_1ID_YAE142	PICLASSE	44,342	12,283	Species_290	NA
Pcoxalis PICLASSE_201403_2_YAE134	PICLASSE	44,342	12,283	Species_290	NA
Pcoxalis PIFUCINO_201205_1ID_YAE139	PIFUCINO	42,011	13,609	Species_290	NA
Pcoxalis PIFUCINO_201205_2_YAE131	PIFUCINO	42,011	13,609	Species_290	NA
Pcoxalis POZMERRO_201210_1ID_YAE149	POZMERRO	42,039	12,680	Species_290	NA
Pcoxalis POZMERRO_201210_2_YAE133	POZMERRO	42,039	12,680	Species_290	NA
Pcoxalis QUERCIO_200904_1ID_YAS755	QUERCIO	43,845	11,321	Species_290	JQ921165
Pcoxalis RAAN_200800_3ID_YAS498	RAAN	55,999	12,766	Species_290	JQ921166
Pcoxalis RAAN_200800_8ID_YLK745	RAAN	55,999	12,766	Species_290	JQ921167
Pcoxalis RAAN_200800_9_YLK817	RAAN	55,999	12,766	Species_290	JQ921168
Pcoxalis RNVESSY_201403_1ID_YAE135	RNVESSY	46,176	6,158	Species_290	NA
Pcoxalis RNVESSY_201403_3_YAE151	RNVESSY	46,176	6,158	Species_290	NA
Pcoxalis RNVESSY_201403_2_YAF582	RNVESSY	46,176	6,158	Species_290	NA
Pcoxalis SANDONAC_200903_1ID_YAB246	SANDONAC	40,452	17,926	Species_290	KY623939
Pcoxalis SANDONAC_200903_2_YAB308	SANDONAC	40,452	17,926	Species_290	KY623940
Pcoxalis SCASCINA_201409_1ID_YAE168	SCASCINA	44,456	8,844	Species_290	NA
Pcoxalis SCASCINA_201409_2_YAE156	SCASCINA	44,456	8,844	Species_290	NA
Pcoxalis SCIROCCO_201407_3_YAE176	SCIROCCO	38,108	13,325	Species_290	NA
Pcoxalis SCIROCCO_201407_1ID_YAE172	SCIROCCO	38,108	13,325	Species_290	NA
Pcoxalis SEYMAXTH_201204_1ID_YAC450	SEYMAXTH	46,188	6,184	Species_290	KY623943
Pcoxalis SLOVENIA_000000_PA5_DQ144752 NCBI	SLOVENIA	46,104	14,999	Species_290	DQ144752
Pcoxalis SLOVENIA_000000_PA4_DQ144777 NCBI	SLOVENIA	46,104	14,999	Species_290	DQ144777
Pcoxalis SORGFNE_201407_2_YAE152	SORGFNE	38,103	13,313	Species_290	NA
Pcoxalis SORGFNE_201407_1ID_YAE136	SORGFNE	38,103	13,313	Species_290	NA
Pcoxalis STENSAN_200811_2ID_YAS471	STENSAN	56,428	12,873	Species_290	JQ921172
Pcoxalis STSEVERA_201304_1ID_YAE153	STSEVERA	42,886	9,472	Species_290	NA
Pcoxalis STSEVERA_201304_2_YAE355	STSEVERA	42,886	9,472	Species_290	NA
Pcoxalis TDISUIO_201204_1ID_YAE167	TDISUIO	41,291	13,858	Species_290	NA
Pcoxalis TDISUIO_201204_2_YAE146	TDISUIO	41,291	13,858	Species_290	NA
Pcoxalis TDISUIO_201204_3_YAE297	TDISUIO	41,291	13,858	Species_290	NA
Pcoxalis TLEMCENP_201505_4ID_YAD787	TLEMCENP	34,817	-1,376	Species_290	NA
Pcoxalis TLEMCENP_201505_6_YAD799	TLEMCENP	34,817	-1,376	Species_290	NA
Pcoxalis VOLTURSP_201509_2_YAE707	VOLTURSP	41,640	14,080	Species_290	NA
Pcoxalis VRANA_200409_1ID_YAB247	VRANA	44,115	15,229	Species_290	KY623946
Pcoxalis VRANA_200409_2ID_YAB281	VRANA	44,115	15,229	Species_290	KY623947
Pcoxalis AVENA_200903_1ID_YAB566	AVENA	40,060	16,156	Species_291	KY623899
Pcoxalis AVENA_200903_2ID_YAB285	AVENA	40,060	16,156	Species_291	KY623900
Pcoxalis BEITED_200909_1ID_YLK352	BEITED	33,703	35,471	Species_293	JQ921141
Pcoxalis BEITED_200909_2_YLK348	BEITED	33,703	35,471	Species_293	JQ921142
Pcoxalis BISRI_200909_1ID_YLK590	BISRI	33,580	35,535	Species_293	JQ921145
Pcoxalis TRICHONI_200404_1ID_YLK450	TRICHONI	38,603	21,556	Species_295	JQ921160
Pcoxalis TRICHONI_200404_2_YLK577	TRICHONI	38,603	21,556	Species_295	JQ921161
Pcoxalis HERODOTO_201303_1ID_YAD372	HERODOTO	37,683	20,831	Species_297	KY623918
Pcoxalis HERODOTO_201303_2ID_YAD359	HERODOTO	37,683	20,831	Species_297	KY623919
Pcoxalis LIMNIAVI_201303_2_YAD373	LIMNIAVI	38,172	20,711	Species_297	KY623927
Pcoxalis LIMNIAVI_201303_1ID_YAD354	LIMNIAVI	38,172	20,711	Species_297	KY623926
Pcoxalis AGFLOROSI_201503_1ID_YAE890	AGFLOROS	37,169	22,025	Species_298	NA
Pcoxalis AGFLOROSI_201503_2ID_YAD955	AGFLOROS	37,169	22,025	Species_298	NA
Pcoxalis AGFLOROSI_201503_3_YAD784	AGFLOROS	37,169	22,025	Species_298	NA
Pcoxalis PH15CHAN_201510_1ID_YAF202	PH15CHAN	37,592	21,811	Species_298	NA
Pcoxalis PLOW6AGI_201510_2_YAE739	PLOW6AGI	37,166	22,020	Species_298	NA
Pcoxalis PLOW7AGI_201510_1ID_YAE847	PLOW7AGI	37,170	22,023	Species_298	NA
Pcoxalis PLOW7AGI_201510_2_YAE848	PLOW7AGI	37,170	22,023	Species_298	NA
Pcoxalis LKAIAFA_201303_2_YAD375	LKAIAFA	37,494	21,618	Species_299	KY623929
Pcoxalis LKAIAFA_201303_1ID_YAD360	LKAIAFA	37,494	21,618	Species_299	KY623928

*Proasellus coxalis* (Dolfs, 1892)

	Pcoxalis PYRGYRO_201303_2ID_YAD356	PYRGYRO	38,318	20,566	Species_300	KY623938
	Pcoxalis PYRGYRO_201303_1ID_YAD374	PYRGYRO	38,318	20,566	Species_300	KY623937
	Pcoxalis FLATANO_201405_1ID_YAD348	FLATANO	40,609	15,449	Species_301	KY623917
<i>Proasellus (coxalis) coxalis gabriellae</i> (Margalef, 1950)	Pcoxalis_gabriellae MAJORCA_201412_2ID_YAE226	MAJORCA	39,765	2,711	Species_290	NA
	Pcoxalis_gabriellae MAJORCA_201412_3_YAE123	MAJORCA	39,765	2,711	Species_290	NA
<i>Proasellus coxalis nanus</i> Sket, 1991	Pcoxalis_nanus AMAVRISP_201603_1ID_YAE763	AMAVRISP	34,850	32,870	Species_285	NA
	Pcoxalis_nanus AMAVRISP_201603_3_YAE720	AMAVRISP	34,850	32,870	Species_285	NA
<i>Proasellus (coxalis) sp.</i>	Psp_gr_coxalis APHRODSP_201603_1ID_YAE737	APHRODSP	34,760	32,450	Species_284	NA
	Psp_gr_coxalis APHRODSP_201603_3_YAE709	APHRODSP	34,760	32,450	Species_284	NA
<i>Proasellus (coxalis) coxalis septentrionalis</i> (Herbst, 1956)	Pcoxalisseptentrionalis LAUCHA1_200406_1ID_YLK812	LAUCHA1	51,396	11,987	Species_290	JQ921169
	Pcoxalisseptentrionalis LAUCHA2_200406_4_YCD91	LAUCHA2	51,382	11,954	Species_290	JQ921170
<i>Proasellus n. sp. (DABARSKA)</i>	PspDabarska DABARSKA_200611_1FID_YAB167	DABARSKA	44,767	16,667	Species_310	KY624080
	PspDabarska DABARSKA_200611_2_YAB322	DABARSKA	44,767	16,667	Species_310	KY624081
<i>Proasellus deminutus</i> (Sket, 1959)	Pdeminutus MALECN2_201009_1ID_YAF602	MALECN2	46,554	15,696	Species_399	NA
	Pdeminutus MALECN2_201009_2_YCD646	MALECN2	46,554	15,696	Species_399	JQ921178
	Pdeminutus MALECN2_201009_3_YCD656	MALECN2	46,554	15,696	Species_399	JQ921179
	Pdeminutus MALECN2_201009_2ex_YAF603	MALECN2	46,554	15,696	Species_399	NA
	Pdeminutus MALECN2_201009_3ex_YCD700	MALECN2	46,554	15,696	Species_399	JQ921183
	Pdeminutus ROSNJA_201009_1ID_YCD537	ROSNJA	46,483	15,756	Species_399	JQ921180
	Pdeminutus ROSNJA_201009_2_YCD645	ROSNJA	46,483	15,756	Species_399	JQ921181
	Pdeminutus ROSNJA_201009_3_YCD660	ROSNJA	46,483	15,756	Species_399	JQ921182
	Pdeminutus ROSNJE_201307_1ID_YAF709	ROSNJE	46,483	15,753	Species_399	NA
	Pdeminutus CATEZSAV_201504_1ID_YAE438	CATEZSAV	45,894	15,632	Species_404	NA
	Pdeminutus CATEZSAV_201504_2_YAE439	CATEZSAV	45,894	15,632	Species_404	NA
	Pdeminutus CATEZSAV_201504_3_YAE390	CATEZSAV	45,894	15,632	Species_404	NA
	Pdeminutus DOVJEZ_200910_1ID_YLK819	DOVJEZ	46,116	14,475	Species_404	JQ921175
	Pdeminutus DOVJEZ_200910_4ex_YCD99	DOVJEZ	46,116	14,475	Species_404	JQ921176
	Pdeminutus DOVJEZ_200910_8ex_YCD70	DOVJEZ	46,116	14,475	Species_404	JQ921177
	Pdeminutus OBREZJE_201504_1ID_YAE391	OBREZJE	45,849	15,707	Species_404	NA
	Pdeminutus OBREZJE_201504_2_YAE440	OBREZJE	45,849	15,707	Species_404	NA
	Pdeminutus PODGRAD_201009_1ID_YAC404	PODGRAD	46,079	14,665	Species_404	KY623960
<i>Proasellus diana</i> Pesce & Argano, 1985	Pdiana CLITUNNO_200904_1ID_YAS756	CLITUNNO	42,833	12,768	Species_277	JQ921184
	Pdiana CLITUNNO_200904_2_YAS815	CLITUNNO	42,833	12,768	Species_277	JQ921185
	Pdiana CLITUNNO_200904_3_YAS825	CLITUNNO	42,833	12,768	Species_277	JQ921186
<i>Proasellus ebrencis</i> Henry & Magniez, 1992	Pebrensis CERECEDA_200906_1_YLK152	CERECEDA	42,800	-3,494	Species_259	JQ921187
	Pebrensis CERECEDA_201405_11am_YAD343	CERECEDA	42,800	-3,494	Species_259	KY623961
	Pebrensis CERECEDA_201405_4am_YAF607	CERECEDA	42,800	-3,494	Species_259	NA
	Pebrensis MARFIL_201010_2cyto_YAF596	MARFIL	42,861	-3,593	Species_259	NA
	Pebrensis MARFIL_200906_1ID_YLK505	MARFIL	42,861	-3,593	Species_259	JQ921188
	Pebrensis RUDRON_201010_1IDex_YCD848	RUDRON	42,702	-3,918	Species_259	JQ921070
	Pebrensis RUDRON_201010_2ex_YAB327	RUDRON	42,702	-3,918	Species_259	KY623962
	Pebrensis RUDRON_201010_3ex_YCD912	RUDRON	42,702	-3,918	Species_259	JQ921071
<i>Proasellus escolai</i> Henry & Magniez, 1982	Pescolai DEIFONTE_200910_1ID_YCD141	DEIFONTE	37,329	-3,587	Species_181	JQ921192
	Pescolai DEIFONTE_200910_2_YCD315	DEIFONTE	37,329	-3,587	Species_181	JQ921193
	Pescolai DEIFONTE_201010_3cyto_YAB068	DEIFONTE	37,329	-3,587	Species_181	KY623963
	Pescolai DEIFONTE_201110_20_YAC983	DEIFONTE	37,329	-3,587	Species_181	KY623964
	Pescolai DEIFONTE_201110_21_YAC984	DEIFONTE	37,329	-3,587	Species_181	KY623965
	Pescolai SALAR_200910_1ID_YCD173	SALAR	37,141	-4,066	Species_189	JQ921189
	Pescolai SALAR_200910_2_YCD227	SALAR	37,141	-4,066	Species_189	JQ921190
	Pescolai SALAR_201010_3cyto_YCD984	SALAR	37,141	-4,066	Species_189	JQ921191
<i>Proasellus espanoli</i> Henry & Magniez, 1982	Pespanoli ALCAZAR_200910_1ID_YCD146	ALCAZAR	37,176	-3,167	Species_186	JQ921194
	Pespanoli ALCAZAR_200910_2_YCD211	ALCAZAR	37,176	-3,167	Species_186	JQ921195
	Pespanoli ALCAZAR_200910_3_YCD373	ALCAZAR	37,176	-3,167	Species_186	JQ921196
<i>Proasellus ezzu</i> Argano & Campanero, 2004	Pezzu FUNTEZZA_201504_1ID_YAD779	FUNTEZZA	40,213	8,559	Species_175	NA
	Pezzu FUNTEZZA_201504_2_YAD772	FUNTEZZA	40,213	8,559	Species_175	NA
	Pezzu FUNTEZZA_201504_3_YAD773	FUNTEZZA	40,213	8,559	Species_175	NA
<i>Proasellus faesulanus</i> Messana & Caselli, 1995	Pfaesulanus MUGNONE_200904_1ID_YAS760	MUGNONE	43,819	11,293	Species_273	JQ921197
	Pfaesulanus MUGNONE_200904_2_YAS832	MUGNONE	43,819	11,293	Species_273	JQ921198
	Pfaesulanus MUGNONE_200904_3_YAS844	MUGNONE	43,819	11,293	Species_273	JQ921199
<i>Proasellus franciscoli</i> (Chappuis, 1955)	Pfranciscoli BANDITO_200907_1ID_YLK963	BANDITO	44,292	7,427	Species_377	JQ921203
	Pfranciscoli BANDITO_200907_3_YCD96	BANDITO	44,292	7,427	Species_377	JQ921204
	Pfranciscoli BOSSEA_200801_2ID_YAS204	BOSSEA	44,242	7,840	Species_378	JQ921200
	Pfranciscoli BOSSEA_200801_3_YAS205	BOSSEA	44,242	7,840	Species_378	JQ921201
<i>Proasellus franciscoli</i> (Chappuis, 1955)	Pfranciscoli LANZOTOR_201608_1ID_YAF781	LANZOTOR	45,269	7,484	Species_378	NA
	Pfranciscoli LANZOTOR_201608_3_YAF780	LANZOTOR	45,269	7,484	Species_378	NA
	Pfranciscoli ORSO_200801_1ID_YAS207	ORSO	44,119	7,874	Species_378	JQ921202
<i>Proasellus gjorgjevicii litoralis</i> (Karaman, 1933)	Pgjorgjevicilitoralis OHRIDMAC_000000_gli3_DQ305136 NCBI	OHRIDMAC	41,037	20,719	Species_231	DQ305136
	Pgjorgjevicilitoralis OHRIDMAC_000000_gli1_DQ305137 NCBI	OHRIDMAC	41,037	20,719	Species_231	DQ305137
	Pgjorgjevicilitoralis OSOJ1_200909_1ID_YLK751	OSOJ1	40,951	20,776	Species_231	JQ921205
	Pgjorgjevicilitoralis OSOJ2_200909_2ID_YAC709	OSOJ2	40,951	20,776	Species_231	KY623966
<i>Proasellus gjorgjevicii gjorgjevicii</i> (Karaman, 1933)	Pgjorgjevicitypicus OHRIDMAC_000000_gly2_DQ305131 NCBI	OHRIDMAC	41,037	20,719	Species_231	DQ305131
	Pgjorgjevicitypicus OHRIDMAC_000000_gly3_DQ305132 NCBI	OHRIDMAC	41,037	20,719	Species_231	DQ305132
	Pgjorgjevicitypicus OHRIDMAC_000000_gly4_DQ305133 NCBI	OHRIDMAC	41,037	20,719	Species_231	DQ305133
	Pgjorgjevicitypicus OHRIDMAC_000000_gly5_DQ305134 NCBI	OHRIDMAC	41,037	20,719	Species_231	DQ305134
	Pgjorgjevicitypicus OHRIDMAC_000000_gli2_DQ305135 NCBI	OHRIDMAC	41,037	20,719	Species_231	DQ305135
<i>Proasellus grafi</i> Henry & Magniez, 2003	Pgrafi CUEVAVAL_200906_2_YLK514	CUEVAVAL	43,300	-3,420	Species_257	JQ921206
	Pgrafi SOURVALL_200906_1_YLK159	SOURVALL	43,300	-3,420	Species_257	JQ921207
<i>Proasellus granadensis</i> Henry & Magniez, 2003	Pgranadensis PILAS_200910_1IDex_YCD311	PILAS	37,046	-4,063	Species_187	JQ921211
	Pgranadensis PILAS_200910_2ex_YCD127	PILAS	37,046	-4,063	Species_187	JQ921212
	Pgranadensis PILAS_200910_3ex_YCD216	PILAS	37,046	-4,063	Species_187	JQ921213
	Pgranadensis PILAS_201010_17_YAC988	PILAS	37,046	-4,063	Species_187	KY623968
	Pgranadensis PILAS_201010_13_YAC996	PILAS	37,046	-4,063	Species_187	KY623967
<i>Proasellus guipuzcoensis</i> Henry & Magniez, 2003	Pguipuzcoensis UBAO_200906_1ID_YLK99	UBAO	43,003	-2,405	Species_236	JQ921208
	Pguipuzcoensis UBAO_200906_2_YLK405	UBAO	43,003	-2,405	Species_236	JQ921209
	Pguipuzcoensis UBAO_200906_3_YLK592	UBAO	43,003	-2,405	Species_236	JQ921210
<i>Proasellus hercegovinensis</i> (Karaman, 1933)	Phercegovinensis LEDENJA_200508_2_YCD384 R2FSCOI	LEDENJA	43,524	18,665	Species_323	JQ921214
	Phercegovinensis BJELUSNI_200008_1ID_YLK589	BJELUSNI	42,845	17,978	Species_326	JQ921215
	Phercegovinensis BJELUSNI_200008_2ID_YLK814	BJELUSNI	42,845	17,978	Species_326	JQ921216
	Phercegovinensis BJELUSNI_201107_3ID_YAC657	BJELUSNI	42,845	17,978	Species_326	KY623972
	Phercegovinensis BJELUSNI_201107_10_YAC658	BJELUSNI	42,845	17,978	Species_326	KY623969
	Phercegovinensis BJELUSNI_201107_12_YAC659	BJELUSNI	42,845	17,978	Species_326	KY623970
	Phercegovinensis BJELUSNI_201107_13_YAC660	BJELUSNI	42,845	17,978	Species_326	KY623971
<i>Proasellus hermaliensis</i> (Arcangeli, 1938)	Phermaliensis CHAMPP_200807_1_YAS342	CHAMPP	50,310	4,889	Species_211	KY623973
	Phermaliensis CHAMPP_200807_2ID_YAS365	CHAMPP	50,310	4,889	Species_211	JQ921217
	Phermaliensis CRUPET_200807_5ex_YAS496	CRUPET	50,354	4,961	Species_211	JQ921218
	Phermaliensis LUNETTE_200807_1ID_YAS363	LUNETTE	50,126	5,252	Species_211	JQ921219
	Phermaliensis LUNETTE_200807_2_YAS364	LUNETTE	50,126	5,252	Species_211	JQ921220
	Phermaliensis LUNETTE_200807_3_YAS466	LUNETTE	50,126	5,252	Species_211	JQ921221
	Phermaliensis MILMONDI_200807_1_YAS360	MILMONDI	50,375	5,038	Species_211	KY623974
	Phermaliensis NEBLONPR_200807_1ID_YAS361	NEBLONPR	50,416	5,455	Species_211	JQ921222
	Phermaliensis NEBLONPR_200807_2_YAS362	NEBLONPR	50,416	5,455	Species_211	JQ921223
<i>Proasellus ibericus</i> (Braga, 1946)	Pibericus BARRIO_201010_1ID_YCD514	BARRIO	41,846	-8,568	Species_199	JQ921224
	Pibericus BARRIO_201010_2_YCD571	BARRIO	41,846	-8,568	Species_199	JQ921225
	Pibericus BARRIO_201010_3ID_YCD589	BARRIO	41,846	-8,568	Species_199	JQ921226

	Pibericus BARRIO_201010_12_YAC600	BARRIO	41,846	-8,568	Species_199	KY623977
	Pibericus BARRIO_201010_13_YAC601	BARRIO	41,846	-8,568	Species_199	KY623978
	Pibericus BARRIO_201010_10_YAC594	BARRIO	41,846	-8,568	Species_199	KY623975
	Pibericus BARRIO_201010_11_YAC595	BARRIO	41,846	-8,568	Species_199	KY623976
	Pibericus BARROCAS_201504_1ID_YAD752	BARROCAS	41,952	-8,471	Species_199	NA
	Pibericus BARROCAS_201504_2_YAD803	BARROCAS	41,952	-8,471	Species_199	NA
	Pibericus BARROCAS_201504_3_YAD753	BARROCAS	41,952	-8,471	Species_199	NA
<i>Proasellus intermedius intermedius</i> (Sket, 1965)	Pintermedius NIMIS_201405_1IDex_YCM691	NIMIS	46,189	13,252	Species_400	KY623987
	Pintermedius NIMIS_201405_3ex_YCM654	NIMIS	46,189	13,252	Species_400	KY623988
	Pintermedius BABJAJAM_200703_2_YAD384	BABJAJAM	46,125	13,712	Species_401	KY623989
	Pintermedius BABJAJAM_200703_1_YAD393	BABJAJAM	46,125	13,712	Species_401	NA
	Pintermedius DREIJCA_200908_1ID_YLK317	IDREIJCA	46,095	13,834	Species_401	JQ921235
	Pintermedius IDRIJA_200904_1ID_YLK710	IDRIJA	46,154	13,774	Species_401	JQ921227
	Pintermedius IDRIJA_200904_2_YLK725	IDRIJA	46,154	13,774	Species_401	JQ921228
	Pintermedius ISONZO_201208_1ex_YAD050	ISONZO	45,809	13,427	Species_401	KY623981
	Pintermedius ISONZO_201208_2ex_YAD278	ISONZO	45,809	13,427	Species_401	KY623982
	Pintermedius ISONZO_201208_3ex_YAD115	ISONZO	45,809	13,427	Species_401	KY623983
	Pintermedius JAMIANO_201403_1ID_YAD053	JAMIANO	45,818	13,583	Species_401	KY623984
	Pintermedius JAMIANO_201403_3_YAD054	JAMIANO	45,818	13,583	Species_401	KY623986
	Pintermedius JAMIANO_201403_2_YAD114	JAMIANO	45,818	13,583	Species_401	KY623985
	Pintermedius BELLUNO_201504_1ID_YAD960	BELLUNO	46,134	12,226	Species_402	NA
	Pintermedius GROTFOOS_201203_2ex_YAD122	GROTFOOS	46,254	12,814	Species_403	KY623980
	Pintermedius GROTFOOS_201203_1ex_YAD057	GROTFOOS	46,254	12,814	Species_403	KY623979
<i>Proasellus intermedius meridionalis</i> Stoch, 1988	Pintermediusmeridionalis COMARIE_200908_1ID_YCD573	COMARIE	45,844	13,539	Species_401	JQ921229
	Pintermediusmeridionalis COMARIE_200908_2_YCD654	COMARIE	45,844	13,539	Species_401	JQ921230
	Pintermediusmeridionalis COMARIE_200908_3_YCD649	COMARIE	45,844	13,539	Species_401	JQ921231
	Pintermediusmeridionalis ZALOSCE_201008_1ID_YCD510	ZALOSCE	45,888	13,743	Species_401	JQ921232
	Pintermediusmeridionalis ZALOSCE_201008_2_YCD648	ZALOSCE	45,888	13,743	Species_401	JQ921233
	Pintermediusmeridionalis ZALOSCE_201008_3_YCD643	ZALOSCE	45,888	13,743	Species_401	JQ921234
<i>Proasellus istrianus</i> (Stammer, 1932)	Pistrianius PIVKAJAM_201202_2ID_YAC800	PIVKAJAM	45,804	14,205	Species_269	NA
	Pistrianius USTJE_201211_1ex_YAD163	USTJE	45,864	13,872	Species_270	KY623993
	Pistrianius PIVKAJAM_201202_1ID_YAE420	PIVKAJAM	45,804	14,205	Species_271	NA
	Pistrianius GLINSCIC_200701_1ID_YLK711	GLINSCIC	45,621	13,872	Species_272	JQ921239
	Pistrianius GLINSCIC_200701_2_YLK829	GLINSCIC	45,621	13,872	Species_272	JQ921240
	Pistrianius GLINSCIC_200701_3_YLK834	GLINSCIC	45,621	13,872	Species_272	JQ921241
	Pistrianius KLUZA_201007_1ID_YCD608	KLUZA	45,600	13,854	Species_272	JQ921242
	Pistrianius KLUZA_201007_2_YCD930	KLUZA	45,600	13,854	Species_272	JQ921243
	Pistrianius KLUZA_201007_3_YCD708	KLUZA	45,600	13,854	Species_272	JQ921244
	Pistrianius STEPAN1_201009_1_YCD388	STEPAN1	45,550	13,856	Species_272	JQ921236
	Pistrianius STEPAN2_201009_2ID_YCD647	STEPAN2	45,552	13,855	Species_272	JQ921237
	Pistrianius STEPAN2_201009_3_YCD662	STEPAN2	45,552	13,855	Species_272	JQ921238
	Pistrianius TRENTOVI_201104_1_YAD059	TRENTOVI	45,887	13,582	Species_272	KY623990
	Pistrianius TRENTOVI_201104_3_YAD129	TRENTOVI	45,887	13,582	Species_272	KY623992
	Pistrianius TRENTOVI_201104_2_YAD285	TRENTOVI	45,887	13,582	Species_272	KY623991
<i>Proasellus jaloniacus</i> Henry & Magniez, 1978	Pjaloniacus JALON_200910_1ID_YCD142	JALON	38,757	-0,105	Species_239	JQ921245
	Pjaloniacus JALON_200910_2_YCD282	JALON	38,757	-0,105	Species_239	JQ921246
	Pjaloniacus JALON_200910_3_YCD395	JALON	38,757	-0,105	Species_239	JQ921247
	Pjaloniacus JALON_201110_34_YAC989	JALON	38,757	-0,105	Species_239	KY623994
	Pjaloniacus JALON_201110_35_YAC990	JALON	38,757	-0,105	Species_239	KY623995
<i>Proasellus karamani</i> (Remy, 1934)	Pkaramani ASLZ6_201410_3_YAE879	ASLZ6	42,227	19,067	Species_123	NA
	Pkaramani DJATO_201107_1ID_YAB614	DJATO	42,835	19,907	Species_123	KY623998
	Pkaramani DJATO_201107_2ID_YAB630	DJATO	42,835	19,907	Species_123	KY623999
	Pkaramani DJATO_201107_3ID_YAB816	DJATO	42,835	19,907	Species_123	KY624000
	Pkaramani LJUTAMOU_201107_1ID_YAB904	LJUTAMOU	42,536	18,377	Species_123	KY624005
	Pkaramani LJUTAMOU_201107_2ID_YAB959	LJUTAMOU	42,536	18,377	Species_123	KY624006
<i>Proasellus karamani</i> (Remy, 1934)	Pkaramani ASLZ2A_201410_3_YAE878	ASLZ2A	42,839	18,951	Species_124	NA
	Pkaramani VUKOVA_200504_3_YAD571	VUKOVA	42,773	18,949	Species_124	NA
	Pkaramani VUKOVA_200504_1_YAD392	VUKOVA	42,773	18,949	Species_124	NA
	Pkaramani BORACKO_201107_1IDex_YAB612	BORACKO	43,555	18,031	Species_125	KY623996
	Pkaramani BORACKO_201107_2ID_YAB629	BORACKO	43,555	18,031	Species_125	KY623997
	Pkaramani KLJUC_200008_1ID_YAB320	KLJUC	43,093	18,485	Species_126	KM249026
	Pkaramani KLJUC_201107_1ID_YAC085	KLJUC	43,093	18,485	Species_126	KY624002
	Pkaramani KLJUC_200008_2_YAB321	KLJUC	43,093	18,485	Species_126	KY624001
	Pkaramani KLJUC_201107_23_YAC605	KLJUC	43,093	18,485	Species_126	KY624004
	Pkaramani KLJUC_201107_20_YAC598	KLJUC	43,093	18,485	Species_126	KY624003
	Pkaramani ZALOMSKA_200305_1_YLK308	ZALOMSKA	43,181	18,122	Species_126	JQ921248
<i>Proasellus lagari</i> Henry & Magniez, 1982	Plagarij CRUZ_200910_1ID_YCD210	CRUZ	38,101	-1,828	Species_194	JQ921249
	Plagarij CRUZ_200910_2_YCD144	CRUZ	38,101	-1,828	Species_194	JQ921250
	Plagarij CRUZ_200910_3_YCD372	CRUZ	38,101	-1,828	Species_194	JQ921251
<i>Proasellus lescherae</i> Henry & Magniez, 1978	Plescherae CFIAQUERRET_200906_2ex_YLK404	AQUERRET	42,895	-1,554	Species_218	JQ921113
	Plescherae CFIAQUERRET_200906_3_YLK403	AQUERRET	42,895	-1,554	Species_218	JQ921114
	Plescherae MONZON_200906_1ID_YLK137	MONZON	41,916	0,176	Species_214	JQ921259
	Plescherae MONZON_200906_2_YLK179	MONZON	41,916	0,176	Species_214	JQ921260
	Plescherae MONZON_200906_3_YLK180	MONZON	41,916	0,176	Species_214	JQ921261
	Plescherae SIPAN_200906_1ID_YLK98	SIPAN	42,183	-0,266	Species_214	JQ921263
	Plescherae SIPAN_200906_2_YLK504	SIPAN	42,183	-0,266	Species_214	JQ921264
	Plescherae SIPANFON_200906_1_YLK186	SIPANFON	42,183	-0,266	Species_214	JQ921262
	Plescherae BERGANT_200910_1ID_YCD155	BERGANT	40,669	-0,193	Species_215	JQ921257
	Plescherae BERGANT_200910_2_YCD148	BERGANT	40,669	-0,193	Species_215	JQ921258
	Plescherae MIJARES_200910_1ID_YCD143	MIJARES	40,030	-0,350	Species_216	JQ921267
	Plescherae MIJARES_200910_2_YCD212	MIJARES	40,030	-0,350	Species_216	JQ921268
	Plescherae PEQUENO_200910_1ID_YCD195	PEQUENO	40,024	-0,354	Species_216	JQ921269
	Plescherae PEQUENO_200910_2_YCD313	PEQUENO	40,024	-0,354	Species_216	JQ921270
	Plescherae LUMBIER_200906_1ID_YLK114	LUMBIER	42,685	-1,303	Species_217	JQ921254
	Plescherae LUMBIER_200906_2_YLK325	LUMBIER	42,685	-1,303	Species_217	JQ921255
	Plescherae LUMBIER_200906_3_YLK249	LUMBIER	42,685	-1,303	Species_217	JQ921256
	Plescherae ALGAS_200910_1ID_YCD205	ALGAS	40,909	0,249	Species_219	JQ921252
<i>Proasellus lescherae</i> Henry & Magniez, 1978	Plescherae ALGAS_200910_2_YCD933	ALGAS	40,909	0,249	Species_219	JQ921253
	Plescherae MATARRA_200910_1ID_YCD279	MATARRA	40,930	0,104	Species_219	JQ921265
	Plescherae MATARRA_200910_2_YCD197	MATARRA	40,930	0,104	Species_219	JQ921266
	Plescherae TASTAV1_200910_2ex_YCD199	TASTAV1	40,887	0,105	Species_219	JQ921271
<i>Proasellus ligusticus</i> Bodon & Argano, 1982	Pligusticus MADIRA_200904_1ID_YAS764	MADIRA	44,236	9,639	Species_173	JQ921272
	Pligusticus MADIRA_200904_2_YAS809	MADIRA	44,236	9,639	Species_173	JQ921273
	Pligusticus MADIRA_200904_3_YAS829	MADIRA	44,236	9,639	Species_173	JQ921274
	Pligusticus SORBA_200904_1ID_YAS751	SORBA	44,248	9,478	Species_260	JQ921275
	Pligusticus SORBA_200904_2_YAS810	SORBA	44,248	9,478	Species_260	JQ921276
	Pligusticus SORBA_200904_3_YAS830	SORBA	44,248	9,478	Species_260	JQ921277
<i>Proasellus lusitanicus</i> (Frade, 1938)	Plusitanicus TALIMA_200901_1IDex_YAB193	TALIMA	39,988	-8,385	Species_203	KY624009
	Plusitanicus TALIMA_200901_2ex_YAB879	TALIMA	39,988	-8,385	Species_203	KY624010
	Plusitanicus ALVIELA_200708_1ID_YAB243	ALVIELA	39,446	-8,712	Species_204	KY624007
	Plusitanicus CONTENDA_200701_1ID_YAB250	CONTENDA	39,533	-8,699	Species_204	KY624008
<i>Proasellus margalefi</i> Henry & Magniez, 1982	Pmargalefi MEREMIL_200910_1ID_YCD241	MEREMIL	39,598	-0,849	Species_240	JQ921280
	Pmargalefi MEREMIL_200910_2_YCD225	MEREMIL	39,598	-0,849	Species_240	JQ921281
	Pmargalefi ALBENTOS_200910_1ID_YCD86	ALBENTOS	40,122	-0,764	Species_242	JQ921278

	Pmargalefi ALBENTOS_200910_2_YCD169	ALBENTOS	40,122	-0,764	Species_242	JQ921279
	Pmargalefi CF CASCAN2_200910_1_ID_YCD240	CASCAN2	40,204	-1,130	Species_241	JQ921026
	Pmargalefi CF CASCAN2_200910_2_YCD167	CASCAN2	40,204	-1,130	Species_241	JQ921027
	Pmargalefi CF EBRON_200910_1_YCD226	EBRON	40,145	-1,316	Species_241	JQ921115
<i>Proasellus meijersae</i> Henry & Magniez, 2003	Pmeijersae MUNDO_200910_1_ID_YCD154	MUNDO	38,484	-2,362	Species_182	JQ921282
	Pmeijersae MUNDO_200910_2_YCD266	MUNDO	38,484	-2,362	Species_182	JQ921283
	Pmeijersae MUNDO_200910_3_YCD375	MUNDO	38,484	-2,362	Species_182	JQ921284
<i>Proasellus meridianus</i> (Racovitz, 1919)	Pmeridianus 2BUSTA_201110_1m_YAF598	2BUSTA	43,350	-4,182	Species_243	NA
	Pmeridianus BUSTA_200906_1ID_YAS882	BUSTA	43,349	-4,181	Species_243	JQ921308
	Pmeridianus BUSTA_200906_3_YLK112	BUSTA	43,349	-4,181	Species_243	JQ921309
	Pmeridianus BUSTA_200906_14m_YAC982	BUSTA	43,349	-4,181	Species_243	KY624014
	Pmeridianus PIEZA_200906_1_YAS878	PIEZA	43,004	-3,485	Species_244	JQ921316
	Pmeridianus PIEZA_200906_3_YLK174	PIEZA	43,004	-3,485	Species_244	JQ921317
	Pmeridianus CONCHA_201010_1ID_YAB773	CONCHA	43,560	-6,190	Species_245	KY624015
	Pmeridianus CONCHA_201010_2_YAB766	CONCHA	43,560	-6,190	Species_245	KY624016
	Pmeridianus ANDILLE2_201405_1_YAE433	ANDILLE2	46,481	0,333	Species_246	NA
<i>Proasellus meridianus</i> (Racovitz, 1919)	Pmeridianus ARGATXA_201102_1IDex_YAB196	ARGATXA	43,359	-2,658	Species_246	KY624011
	Pmeridianus ARGATXA_201102_2ex_YAB313	ARGATXA	43,359	-2,658	Species_246	KY624012
	Pmeridianus ARTENAC_201301_1ID_YAC804	ARTENAC	43,272	1,577	Species_246	KY624013
	Pmeridianus BALANOS_201505_1ID_YAD790	BALANOS	44,606	-0,984	Species_246	NA
	Pmeridianus BALANOS_201505_2ID_YAD959	BALANOS	44,606	-0,984	Species_246	NA
	Pmeridianus BARBOTT_200808_1_YAS373	BARBOTT	46,057	-0,976	Species_246	JQ921285
	Pmeridianus BARBOTT_200808_2_YAS376	BARBOTT	46,057	-0,976	Species_246	JQ921286
	Pmeridianus BOISFONT_200808_1_YAS461	BOISFONT	46,151	-0,774	Species_246	JQ921287
	Pmeridianus FAIL_201010_12ID_YCD494	FAIL	40,601	-7,984	Species_246	JQ921293
	Pmeridianus FAIL_201010_13_YCD538	FAIL	40,601	-7,984	Species_246	JQ921294
	Pmeridianus FAIL_201010_14_YCD578	FAIL	40,601	-7,984	Species_246	JQ921295
	Pmeridianus FCIBIOUX_201408_1ex_YAF044	FCIBIOUX	46,075	0,365	Species_246	NA
	Pmeridianus FCONDEON_201506_1_YAE542	FCONDEON	45,409	-0,140	Species_246	NA
	Pmeridianus FCONDEON_201506_2_YAE712	FCONDEON	45,409	-0,140	Species_246	NA
	Pmeridianus FEYSSAC_201104_1ID_YAB394	FEYSSAC	45,673	2,141	Species_246	KY624017
	Pmeridianus FEYSSAC_201104_2_YAB593	FEYSSAC	45,673	2,141	Species_246	KY624018
	Pmeridianus FEYSSAC_201104_3_YAB437	FEYSSAC	45,673	2,141	Species_246	KY624019
	Pmeridianus FONTENIL_201604_1ID_YAE734	FONTENIL	47,755	1,617	Species_246	NA
	Pmeridianus FONTENIL_201604_2_YAE706	FONTENIL	47,755	1,617	Species_246	NA
	Pmeridianus FONTSERG_201209_1ID_YAC801	FONTSERG	46,405	0,744	Species_246	KY624020
	Pmeridianus FOTILLET_201505_1ID_YAE538	FOTILLET	45,748	0,044	Species_246	NA
	Pmeridianus FPINQUET_201506_1_YAF128	FPINQUET	45,245	0,435	Species_246	NA
	Pmeridianus FPINQUET_201506_2_YAE714	FPINQUET	45,245	0,435	Species_246	NA
	Pmeridianus FPINQUET_201606_4_YAF130	FPINQUET	45,245	0,435	Species_246	NA
	Pmeridianus FPINQUET_201606_5_YAF707	FPINQUET	45,245	0,435	Species_246	NA
	Pmeridianus FTSEREIN_201405_1ex_YAE406	FTSEREIN	46,405	0,744	Species_246	NA
	Pmeridianus GDBOIT_200812_1_YAC461	GDBOIT	46,044	-0,483	Species_246	KY624021
	Pmeridianus GEORGES_201009_1ID_YCD600	GEORGES	44,890	1,619	Species_246	JQ921296
	Pmeridianus GEORGES_201009_2_YCD811	GEORGES	44,890	1,619	Species_246	JQ921297
	Pmeridianus GEORGES_201009_3_YCD664	GEORGES	44,890	1,619	Species_246	JQ921298
	Pmeridianus GERESLIT_201506_1_YAE545	GERESLIT	46,093	-0,727	Species_246	NA
<i>Proasellus meridianus</i> (Racovitz, 1919)	Pmeridianus ILESNOU2_201406_7am_YAF690	ILESNOU2	45,817	5,090	Species_246	NA
	Pmeridianus ILESNOU2_201406_9am_YAF689	ILESNOU2	45,817	5,090	Species_246	NA
	Pmeridianus ILESNOU2_201406_10am_TRINITYDYN18915c0g11	ILESNOU2	45,817	5,090	Species_246	NA
	Pmeridianus ILESNOUV_200804_1ID_YAS247	ILESNOUV	45,816	5,101	Species_246	JQ921305
	Pmeridianus ILESNOUV_200804_2_YAS265	ILESNOUV	45,816	5,101	Species_246	JQ921306
	Pmeridianus ILESNOUV_200804_3_YAS460	ILESNOUV	45,816	5,101	Species_246	JQ921307
	Pmeridianus JEANMART_201506_1ID_YAE662	JEANMART	45,264	0,168	Species_246	NA
	Pmeridianus JEANMART_201506_2_YAF131	JEANMART	45,264	0,168	Species_246	NA
	Pmeridianus JEANMART_201506_3_YAE590	JEANMART	45,264	0,168	Species_246	NA
	Pmeridianus LAFINOU_201011_1ID_YCD780	LAFINOU	44,880	1,614	Species_246	JQ921299
	Pmeridianus LAFINOU_201011_2_YCD789	LAFINOU	44,880	1,614	Species_246	JQ921300
	Pmeridianus LAFINOU_201011_4_YAF704	LAFINOU	44,880	1,614	Species_246	NA
	Pmeridianus LAREILLE_201606_1ID_YAF029	LAREILLE	45,219	1,135	Species_246	NA
	Pmeridianus LAREILLE_201606_2_YAF134	LAREILLE	45,219	1,135	Species_246	NA
	Pmeridianus LAREILLE_201606_3_YAF135	LAREILLE	45,219	1,135	Species_246	NA
	Pmeridianus LDOUSSAY_201408_1ex_YAF030	LDOUSSAY	46,853	0,264	Species_246	NA
	Pmeridianus LIGLET_201406_1ID_YAD302	LIGLET	46,525	1,064	Species_246	KY624022
	Pmeridianus LOMBARD_201011_1ID_YCD847	LOMBARD	44,882	1,616	Species_246	JQ921301
	Pmeridianus LOMBARD_201011_2_YCD852	LOMBARD	44,882	1,616	Species_246	JQ921302
	Pmeridianus MARTEL_201102_1IDex_YAB165	MARTEL	44,964	1,594	Species_246	KY624023
	Pmeridianus MAUVINIE_201506_1ID_YAE544	MAUVINIE	45,775	-0,962	Species_246	NA
	Pmeridianus MAUVINIE_201506_2_YAE523	MAUVINIE	45,775	-0,962	Species_246	NA
	Pmeridianus MONDANE_201102_1IDex_YAB164	MONDANE	44,849	1,335	Species_246	KY624024
	Pmeridianus MOULINCH_200905_1ID_YLK184	MOULINCH	44,762	-0,635	Species_246	JQ921288
	Pmeridianus MOULINCH_200905_2_YLK541	MOULINCH	44,762	-0,635	Species_246	JQ921289
	Pmeridianus MOULINCH_200905_3_YLK513	MOULINCH	44,762	-0,635	Species_246	JQ921290
	Pmeridianus NAUERNAA_200201_1ID_YAS510	NAUERNAA	52,464	4,764	Species_246	JQ921310
	Pmeridianus NAUERNAA_200201_2_YAS687	NAUERNAA	52,464	4,764	Species_246	JQ921311
	Pmeridianus NAUERNAA_200201_3_YLK508	NAUERNAA	52,464	4,764	Species_246	JQ921312
	Pmeridianus OUESSAN_201101_1ID_YAB254	OUESSAN	48,457	-5,096	Species_246	KY624025
	Pmeridianus OUESSAN_201101_2_YAB328	OUESSAN	48,457	-5,096	Species_246	KY624026
	Pmeridianus PADIDEBA_201011_1ID_YCD917	PADIDEBA	44,858	1,750	Species_246	JQ921303
	Pmeridianus PADIDEBA_201011_2cyto_YAB008	PADIDEBA	44,858	1,750	Species_246	JQ921304
	Pmeridianus PEROUSE_201405_1ex_YAD317	PEROUSE	46,750	0,798	Species_246	KY624027
	Pmeridianus PLAULT_200808_1ID_YAS602	PLAULT	46,747	1,029	Species_246	JQ921291
	Pmeridianus PLAULT_200808_2_YAS624	PLAULT	46,747	1,029	Species_246	JQ921292
	Pmeridianus POSTLAKE_200912_1ID_YLK714	POSTLAKE	50,686	-3,418	Species_246	JQ921313
	Pmeridianus POSTLAKE_200912_2_YLK743	POSTLAKE	50,686	-3,418	Species_246	JQ921314
	Pmeridianus POSTLAKE_200912_3_YLK728	POSTLAKE	50,686	-3,418	Species_246	JQ921315
	Pmeridianus RAYSSAGU_201409_1_YAD062	RAYSSAGU	43,910	2,223	Species_246	KY624028
	Pmeridianus RIBELLE_200808_1_YAS462	RIBELLE	46,752	1,036	Species_246	JQ921318
	Pmeridianus RINGERE_201403_1ex_YAE407	RINGERE	46,617	0,214	Species_246	NA
	Pmeridianus ROUSSIER_201505_1ID_YAD778	ROUSSIER	46,927	-1,111	Species_246	NA
	Pmeridianus ROUSSIER_201505_2ID_YAD802	ROUSSIER	46,927	-1,111	Species_246	NA
	Pmeridianus SANZAYCH_201406_1ex_YAD118	SANZAYCH	46,960	-0,441	Species_246	KY624029
	Pmeridianus SAVONRU_200812_1ID_YAS601	SAVONRU	47,346	0,552	Species_246	JQ921322
	Pmeridianus SAVONRU_200812_2_YAS594	SAVONRU	47,346	0,552	Species_246	JQ921323
	Pmeridianus SAVONRU_200812_3_YAS609	SAVONRU	47,346	0,552	Species_246	JQ921324
	Pmeridianus SCDULIEN_201502_1ex_YAE423	SCDULIEN	46,026	0,201	Species_246	NA
	Pmeridianus SCFAVREL_201407_1ex_YAE539	SCFAVREL	46,578	-0,122	Species_246	NA
<i>Proasellus meridianus</i> (Racovitz, 1919)	Pmeridianus SOUGEAL_000000_1ID_YAS327	SOUGEAL	48,511	-1,506	Species_246	JQ921319
	Pmeridianus SOUGEAL_000000_2_YAS331	SOUGEAL	48,511	-1,506	Species_246	JQ921320
	Pmeridianus SOUGEAL_000000_3ID_YAS622	SOUGEAL	48,511	-1,506	Species_246	JQ921321
	Pmeridianus VARENNES_201401_1ex_YAE527	VARENNES	46,338	0,276	Species_246	NA
	Pmeridianus VARENSOU_201503_1ID_YAD776	VARENSOU	44,157	1,894	Species_246	NA
	Pmeridianus VARENSOU_201503_2ID_YAD777	VARENSOU	44,157	1,894	Species_246	NA
	Pmeridianus VARENSOU_201503_3ID_YAD782	VARENSOU	44,157	1,894	Species_246	NA

<i>Proasellus micropectinatus</i> Baratti & Messina, 1990	Pmicropectinatus MARGHE_200904_1ID_YAS744	MARGHE	44,272	9,657	Species_267	JQ921325
	Pmicropectinatus MARGHE_200904_2_YAS783	MARGHE	44,272	9,657	Species_267	JQ921326
	Pmicropectinatus MARGHE_200904_3_YAS800	MARGHE	44,272	9,657	Species_267	JQ921327
	Pmicropectinatus POGLISPA_201504_1ID_YAE161	POGLISPA	44,228	9,688	Species_268	NA
	Pmicropectinatus POGLISPA_201504_3_YAE162	POGLISPA	44,228	9,688	Species_268	NA
<i>Proasellus minoicus</i> Pesce & Argano, 1980	Pminoicus FTANAXOS_201410_2_YAD068	FTANAXOS	35,308	24,841	Species_307	KY624031
	Pminoicus FTANAXOS_201410_1ID_YAD067	FTANAXOS	35,308	24,841	Species_307	KY624030
<i>Proasellus monodi</i> (Strouhal, 1942)	Pmonodij EPISKOPI_201203_2ID_YAD313	EPISKOPI	39,558	20,867	Species_232	KY624034
	Pmonodij EPISKOPI_201203_3_YAD290	EPISKOPI	39,558	20,867	Species_232	KY624035
	Pmonodij LLAMIA_201303_1ID_YAE401	LLAMIA	38,099	21,369	Species_234	NA
	Pmonodij LLAMIA_201303_3_YAE559	LLAMIA	38,099	21,369	Species_234	NA
	Pmonodij TRICHONI_200404_1ex_YAB239	TRICHONI	38,603	21,556	Species_234	KY624036
	Pmonodij TRICHONI_200404_1ID_YAB866	TRICHONI	38,603	21,556	Species_234	KY624037
<i>Proasellus montalenti</i> Stoch, Valentino & Volpi, 1996	Pmontalenti FTNINFA_201407_1ID_YAD337	FTNINFA	37,780	12,874	Species_289	KY623953
	Pmontalenti FTNINFA_201407_2_YAD363	FTNINFA	37,780	12,874	Species_289	KY623954
	Pmontalenti LERCARA_201405_1ID_YAD351	LERCARA	37,710	13,579	Species_289	KY623955
	Pmontalenti LERCARA_201405_2_YAD336	LERCARA	37,710	13,579	Species_289	KY623956
	Pmontalenti STNINFA_201407_2ID_YAE447	STNINFA	37,780	12,898	Species_289	NA
<i>Proasellus navarrensis</i> Henry & Magniez, 2003	Pnavarrensis NACEDERO_200906_1ID_YAS977	NACEDERO	42,977	-1,918	Species_255	JQ921328
	Pnavarrensis NACEDERO_200906_2_YLK376	NACEDERO	42,977	-1,918	Species_255	JQ921329
	Pnavarrensis NACEDERO_200906_3_YLK378	NACEDERO	42,977	-1,918	Species_255	JQ921330
<i>Proasellus nollii</i> (Karaman, 1952)	Pnollii BURKH1_200905_1_YLK307	BURKH1	48,922	7,599	Species_156	JQ921331
	Pnollii BURKH1_200905_3ex_YAC803	BURKH1	48,922	7,599	Species_156	KY624061
	Pnollii OPFING1_200905_1_YLK212	OPFING1	48,001	7,756	Species_156	JQ921333
	Pnollii OPFING2_200905_1_YLK511	OPFING2	48,007	7,764	Species_156	JQ921334
	Pnollii OPFING3_200905_1ID_YLK290	OPFING3	48,009	7,767	Species_156	JQ921335
	Pnollii ROSSFELD_200911_1ID_YLK820	ROSSFELD	48,334	7,631	Species_156	JQ921332
<i>Proasellus notenboomii</i> Henry & Magniez, 1981	Pnotenboomii TAFNA_201506_1ID_YAE303	TAFNA	34,700	-1,300	Species_279	NA
	Pnotenboomii TAFNA_201506_2_YAE298	TAFNA	34,700	-1,300	Species_279	NA
	Pnotenboomii KHEMISHY_201511_1ID_YAE413	KHEMISHY	34,627	-1,586	Species_280	NA
	Pnotenboomii KHEMISHY_201511_3_YAE428	KHEMISHY	34,627	-1,586	Species_280	NA
<i>Proasellus ortizi</i> Henry & Magniez, 1992	Portizii JIVERO2_200906_1ID_YAS899	JIVERO2	43,304	-3,572	Species_252	JQ921362
	Portizii JIVERO2_200906_2_YLK323	JIVERO2	43,304	-3,572	Species_252	JQ921363
	Portizii JIVERO2_200906_3_YLK113	JIVERO2	43,304	-3,572	Species_252	JQ921364
	Portizii CHICO_201110_1ID_YAC112	CHICO	43,191	-3,581	Species_253	KY624040
	Portizii NACGANDA_201110_1ID_YAC114	NACGANDA	43,194	-3,582	Species_253	KY624041
<i>Proasellus oviedensis</i> Henry & Magniez, 2003	Poviedensis QUINTANA_200906_1_YLK182	QUINTANA	43,400	-4,771	Species_237	JQ921365
<i>Proasellus parvulus</i> (Sket, 1960)	Pparvulus OTOVEC_201009_1ID_YCD536	OTOVEC	45,592	15,167	Species_163	JQ921366
	Pparvulus OTOVEC_201009_2_YCD661	OTOVEC	45,592	15,167	Species_163	JQ921367
	Pparvulus STOBE_201009_1ID_YCD534	STOBE	45,604	15,171	Species_163	JQ921368
	Pparvulus STOBE_201009_11m_YCM361	STOBE	45,604	15,171	Species_163	KY624042
	Pparvulus STOBE_201009_12m_YCM362	STOBE	45,604	15,171	Species_163	KY624043
	Pparvulus STOBE_201009_13m_YCM363	STOBE	45,604	15,171	Species_163	KY624044
<i>Proasellus parvulus</i> (Sket, 1960)	Pparvulus STOBE_201009_14_YCM364	STOBE	45,604	15,171	Species_163	KY624045
	Pparvulus STOBE_201009_15f_YCM365	STOBE	45,604	15,171	Species_163	KY624046
	Pparvulus STOBE_201009_16f_YCM366	STOBE	45,604	15,171	Species_163	KY624047
<i>Proasellus patrizii</i> (Arcangeli, 1952)	Ppatrizii GRCOLORU_201504_1ID_YAE309	GRCOLORU	40,819	8,826	Species_178	NA
	Ppatrizii GRCOLORU_201504_2ID_YAE137	GRCOLORU	40,819	8,826	Species_178	NA
	Ppatrizii GRCOLORU_201504_3_YAE138	GRCOLORU	40,819	8,826	Species_178	NA
<i>Proasellus pavani</i> (Arcangeli, 1942)	Ppavani CVALMALA_201504_1ID_YAD969	CVALMALA	45,717	10,234	Species_308	NA
	Ppavani CVALMALA_201504_2ID_YAD967	CVALMALA	45,717	10,234	Species_308	NA
	Ppavani NIMIS_201405_1IDex_YCM644	NIMIS	46,189	13,252	Species_313	KY624048
	Ppavani NIMIS_201405_2IDex_YCM693	NIMIS	46,189	13,252	Species_313	KY624049
	Ppavani NIMIS_201405_3ex_YCM653	NIMIS	46,189	13,252	Species_313	KY624050
	Ppavani STERPO_201008_1ID_YCD533	STERPO	45,903	13,041	Species_313	JQ921369
	Ppavani STERPO_201008_2_YCD663	STERPO	45,903	13,041	Species_313	JQ921370
	Ppavani STERPO_201008_3_YCD659	STERPO	45,903	13,041	Species_313	JQ921371
<i>Proasellus coxalis peyerimhoffii</i> (Racovitza, 1919)	Ppeyerimhoffii TLEMCCEN_200905_1ID_YAS861	TLEMCCEN	34,773	-1,251	Species_290	JQ921162
	Ppeyerimhoffii TLEMCCEN_200905_2_YLK100	TLEMCCEN	34,773	-1,251	Species_290	JQ921163
	Ppeyerimhoffii TLEMCCEN_200905_3_YLK582	TLEMCCEN	34,773	-1,251	Species_290	JQ921164
<i>Proasellus pribenicensis</i> Flasarova, 1977	Ppribenicensis PRIBENIK_201110_1ID_YAC011	PRIBENIK	48,384	22,003	Species_210	KY624051
	Ppribenicensis PRIBENIK_201110_2_YAC390	PRIBENIK	48,384	22,003	Species_210	KY624052
	Ppribenicensis STRAZME_201105_1ex_YAF045	STRAZME	48,381	21,858	Species_210	NA
<i>Proasellus racovitzae</i> Henry & Magniez, 1972	Ppracovitzae BERNATAS_200811_3ID_YCD371	BERNATAS	42,982	0,899	Species_198	JQ921372
	Ppracovitzae BERNATAS_200811_15f_YAC995	BERNATAS	42,982	0,899	Species_198	KY624053
	Ppracovitzae BERNATAS_200811_16f_YCM371	BERNATAS	42,982	0,899	Species_198	KY624054
<i>Proasellus rectangularis</i> Afonso, 1982	Prectangulatus MONTEMOR_201010_1ID_YCD548	MONTEMOR	38,654	-8,226	Species_207	JQ921373
	Prectangulatus MONTEMOR_201010_2_YCD539	MONTEMOR	38,654	-8,226	Species_207	JQ921374
	Prectangulatus MONTEMOR_201010_3_YCD544	MONTEMOR	38,654	-8,226	Species_207	JQ921375
<i>Proasellus rectus</i> Afonso, 1982	Prectus EVORA_201110_8arn_YAF685	EVORA	38,604	-7,873	Species_208	NA
	Prectus EVORA_201110_9arn_YAF609	EVORA	38,604	-7,873	Species_208	NA
	Prectus EVORA_201010_1ID_YCD572	EVORA	38,604	-7,873	Species_209	JQ921376
	Prectus EVORA_201010_2_YCD966	EVORA	38,604	-7,873	Species_209	JQ921377
	Prectus EVORA_201010_3_YCD575	EVORA	38,604	-7,873	Species_209	JQ921378
<i>Proasellus remyi</i> (Monod, 1932)	Prremyij NAUMOHRI_200404_1ID_YLK309	NAUMOHRI	40,910	20,748	Species_230	JQ921382
<i>Proasellus remyi acutangulus</i> (Karaman, 1953)	Prremyiacutangulus OHRIDMAC_000000_rac1_DQ305127 NCBI	OHRIDMAC	41,037	20,719	Species_230	DQ305127
	Prremyiacutangulus OHRIDMAC_000000_rac2_DQ305128 NCBI	OHRIDMAC	41,037	20,719	Species_230	DQ305128
	Prremyiacutangulus OHRIDMAC_000000_rac3_DQ305129 NCBI	OHRIDMAC	41,037	20,719	Species_230	DQ305129
	Prremyiacutangulus OSOJ2_200909_1ID_YLK851	OSOJ2	40,951	20,776	Species_231	JQ921379
	Prremyiacutangulus OSOJ3_200909_2ID_YLK906	OSOJ3	40,951	20,776	Species_231	JQ921380
<i>Proasellus remyi nudus</i> (Karaman, 1953)	Prremyinus OHRIDMAC_000000_0_DQ305130 NCBI	OHRIDMAC	41,037	20,719	Species_231	DQ305130
<i>Proasellus remyi remyi</i> (Monod, 1932)	Prremyiremyi OHRIDMAC_000000_ma2_DQ305122 NCBI	OHRIDMAC	41,037	20,719	Species_230	DQ305122
	Prremyiremyi OHRIDMAC_000000_ma3_DQ305123 NCBI	OHRIDMAC	41,037	20,719	Species_230	DQ305123
	Prremyiremyi OHRIDMAC_000000_ma4_DQ305124 NCBI	OHRIDMAC	41,037	20,719	Species_230	DQ305124
	Prremyiremyi OHRIDMAC_000000_ma5_DQ305125 NCBI	OHRIDMAC	41,037	20,719	Species_230	DQ305125
	Prremyiremyi OHRIDMAC_000000_rje1_DQ305126 NCBI	OHRIDMAC	41,037	20,719	Species_230	DQ305126
	Prremyiremyi TRPEJCA_200909_1_YLK938	TRPEJCA	40,961	20,783	Species_230	JQ921381
<i>Proasellus rouchi</i> Henry, 1980	Prrouchij MESCLA_200904_1ID_YAS167	MESCLA	43,913	7,186	Species_388	JQ921383
	Prrouchij MESCLA_200904_2ID_YAS159	MESCLA	43,913	7,186	Species_388	JQ921384
	Prrouchij MESCLA_200904_5_YAS168	MESCLA	43,913	7,186	Species_388	JQ921385
<i>Proasellus ruffoi</i> Argano & Campanero, 2004	Pruffoi POCODINU_201504_1ID_YAD747	POCODINU	40,487	8,519	Species_176	NA
	Pruffoi POCODINU_201504_2ID_YAD957	POCODINU	40,487	8,519	Species_176	NA
	Pruffoi FRUNCU_201209_1ID_YAD082	FRUNCU	40,564	9,678	Species_179	KY624055
	Pruffoi FRUNCU_201209_2_YAD083	FRUNCU	40,564	9,678	Species_179	KY624056
<i>Proasellus slavus</i> (Remy, 1948)	Pslavus GLANBACH_201009_1ID_YCD508	GLANBACH	47,810	13,022	Species_156	JQ921400
	Pslavus GLANBACH_201009_2_YCD707	GLANBACH	47,810	13,022	Species_156	JQ921401
	Pslavus GLANBACH_201009_3_YCD814	GLANBACH	47,810	13,022	Species_156	JQ921402
	Pslavus GOLLING_201009_1ID_YCD512	GOLLING	47,582	13,171	Species_156	JQ921403
	Pslavus GOLLING_201009_2_YCD735	GOLLING	47,582	13,171	Species_156	JQ921404
	Pslavus GOLLING_201009_3_YCD737	GOLLING	47,582	13,171	Species_156	JQ921405
	Pslavus LAHAU_200703_1ID_YAS386	LAHAU	48,184	16,522	Species_156	JQ921393
	Pslavus LAHAU_200703_2_YAS384	LAHAU	48,184	16,522	Species_156	JQ921394
	Pslavus LUTTER_200911_1ex_YAC389	LUTTER	48,348	7,578	Species_156	KY624062
	Pslavus MITTEN47_201504_1ex_YAF046	MITTEN47	47,412	11,253	Species_156	NA

	Pslavus MITTENB_200910_1ID_YCD389	MITTENB	47,427	11,259	Species_156	JQ921395	
	Pslavus MITTENB_200910_2_YCD705	MITTENB	47,427	11,259	Species_156	JQ921396	
	Pslavus MUNICHA_201503_1ex_YAE402	MUNICHA	48,207	11,637	Species_156	NA	
Proasellus slavus (Remy, 1948)	Pslavus MUNICHB_201503_1ID_YAF047	MUNICHB	48,206	11,637	Species_156	NA	
	Pslavus MUNICHC_201503_2_YAE403	MUNICHC	48,207	11,633	Species_156	NA	
	Pslavus MUNICHC_201503_2ex_YAE452	MUNICHC	48,207	11,633	Species_156	NA	
	Pslavus NEVESIN_200504_1_YLK375	NEVESIN	49,193	8,113	Species_156	JQ921392	
	Pslavus STMATERN_201409_3arn_YAF604	STMATERN	48,366	7,622	Species_156	NA	
	Pslavus STMATERN_201409_5arn_YAF686	STMATERN	48,366	7,622	Species_156	NA	
	Pslavus SWABIANA_200904_1ID_YCD381	SWABIANA	48,533	10,285	Species_156	JQ921397	
	Pslavus SWABIANA_200904_2_YCD740	SWABIANA	48,533	10,285	Species_156	JQ921398	
	Pslavus SWABIANA_200904_3_YCD736	SWABIANA	48,533	10,285	Species_156	JQ921399	
	Pslavus CVEN_201108_1ID_YAE541	CVEN	46,551	16,209	Species_157	NA	
	Pslavus DROBTINC_201009_1ID_YCD509	DROBTINC	46,697	15,846	Species_157	JQ921389	
	Pslavus DROBTINC_201009_2_YCD650	DROBTINC	46,697	15,846	Species_157	JQ921390	
	Pslavus DROBTINC_201009_3_YCD644	DROBTINC	46,697	15,846	Species_157	JQ921391	
	Pslavus MALECN1_201009_1ex_YCD601	MALECN1	46,554	15,695	Species_157	JQ921406	
	Pslavus PRVOVERZ_201009_1ID_YCD511	PRVOVERZ	46,586	16,169	Species_157	JQ921407	
	Pslavus PRVOVERZ_201009_2_YCD817	PRVOVERZ	46,586	16,169	Species_157	JQ921408	
	Pslavus PRVOVERZ_201009_3_YCD706	PRVOVERZ	46,586	16,169	Species_157	JQ921409	
	Pslavus PTUJ_201107_1ID_YAE520	PTUJ	46,391	15,928	Species_157	NA	
	Pslavus PTUJ_201107_3_YAE554	PTUJ	46,391	15,928	Species_157	NA	
	Pslavus RACE_201307_1ID_YAD041	RACE	46,403	15,704	Species_157	KY624063	
	Pslavus STARSE_201307_1ID_YAE521	STARSE	46,472	15,768	Species_157	NA	
	Pslavus ZLATOLI2_201307_1ID_YAE558	ZLATOLI2	46,458	15,790	Species_157	NA	
	Pslavus CEPKI_201009_1ID_YCD490	CEPKI	45,543	13,834	Species_159	JQ921386	
Pslavus CEPKI_201009_2ID_YCD658	CEPKI	45,543	13,834	Species_159	JQ921387		
Pslavus CEPKI_201009_3_YCD653	CEPKI	45,543	13,834	Species_159	JQ921388		
Proasellus slavus histriae (Sket, 1963)	Pslavushistriae KLUZA_201007_1ID_YCD495	KLUZA	45,600	13,854	Species_158	JQ921410	
	Pslavushistriae KLUZA_201007_2_YCD710	KLUZA	45,600	13,854	Species_158	JQ921411	
Pslavushistriae KLUZA_201007_3_YCD741	KLUZA	45,600	13,854	Species_158	JQ921412		
Proasellus slovenicus (Sket, 1957)	Pslavenicus MOBISAHT_200607_1_YAF069	MOBISAHT	45,708	14,790	Species_337	NA	
	Psp EBENTAL_201407_3_YAD157	EBENTAL	45,729	14,873	Species_337	KY624066	
	Psp EBENTAL_201407_2_YAD145	EBENTAL	45,729	14,873	Species_337	KY624065	
	Psp EBENTAL_201407_1_YAD063	EBENTAL	45,729	14,873	Species_337	KY624064	
	Psp MOBISAHT_201407_1_YAD121	MOBISAHT	45,708	14,790	Species_337	KY624067	
	Psp MOBISAHT_201407_2_YAD277	MOBISAHT	45,708	14,790	Species_337	KY624068	
	Pslavenicus JAMAPOD_200409_1_YLK369	JAMAPOD	45,819	15,099	Species_338	JQ921413	
	Pslavenicus JAMAPOD_200409_2ID_YLK727	JAMAPOD	45,819	15,099	Species_338	JQ921414	
	Pslavenicus JAMAPOD_200409_3_YCD73	JAMAPOD	45,819	15,099	Species_338	JQ921415	
	Proasellus solanasi Henry & Magniez, 1972	Psolanasij GATO_201010_1ID_YCD925	GATO	36,728	-5,238	Species_188	KC610504
Psolanasij GATO_201010_2_YAB355		GATO	36,728	-5,238	Species_188	KY624071	
Psolanasij GATO_201010_3_YAB015		GATO	36,728	-5,238	Species_188	KY624072	
Psolanasij GATO_201010_11m_YAC977		GATO	36,728	-5,238	Species_188	KY624069	
Psolanasij GATO_201010_12m_YAC978		GATO	36,728	-5,238	Species_188	KY624070	
Psp CALORESP_201601_3_YAE723		CALORESP	40,490	15,210	Species_247	NA	
Pnsp1 Sospel SOSP_201409_1IDex_YAD286		SOSP	43,893	7,421	Species_376	KY624110	
Pnsp1 Sospel SOSP_201409_4ID_YAD151		SOSP	43,893	7,421	Species_376	NA	
Psp2 PC45KALA_201500_1_YAE683		PC45KALA	38,015	22,061	Species_296	NA	
PdeminutusAFF CANTARA_200801_1IDex_YAS329		CANTARA	44,121	7,876	Species_348	JQ921017	
PdeminutusAFF CANTARA_200801_2_YAS305	CANTARA	44,121	7,876	Species_348	JQ921018		
PdeminutusAFF CANTARA_200801_4ID_YAS465	CANTARA	44,121	7,876	Species_348	JQ921019		
Proasellus sp. aff. escolai Henry & Magniez, 1982	PescolaiAFF JESUS_200910_1ID_YCD201	JESUS	38,132	-2,969	Species_196	JQ921020	
	PescolaiAFF JESUS_200910_2_YCD194	JESUS	38,132	-2,969	Species_196	JQ921021	
	PescolaiAFF JESUS_200910_3_YCD379	JESUS	38,132	-2,969	Species_196	JQ921022	
	PescolaiAFF COTILLAS_201010_1ID_YAD427	COTILLAS	38,433	-2,529	Species_197	NA	
	PescolaiAFF COTILLAS_201010_3_YAD415	COTILLAS	38,433	-2,529	Species_197	NA	
	Proasellus sp. aff. hypogeus (Racovitza, 1922)	PhypogeusAFF ABDELKAD_201512_1ID_YAE731	ABDELKAD	36,763	5,102	Species_278	NA
		PhypogeusAFF ABDELKAD_201512_2ID_YAE703	ABDELKAD	36,763	5,102	Species_278	NA
	Proasellus n. sp. (STORARDA)	PintermediusAFF STORARDA_201504_2ID_YAE169	STORARDA	44,807	9,819	Species_398	NA
		PintermediusAFF STORARDA_201504_1ID_YAE159	STORARDA	44,807	9,819	Species_398	NA
	Proasellus sp. aff. lagari Henry & Magniez, 1982	PlagariAFF PERAL_200910_1ID_YCD172	PERAL	37,079	-2,101	Species_195	JQ921023
PlagariAFF PERAL_200910_2_YCD198		PERAL	37,079	-2,101	Species_195	JQ921024	
Proasellus sp. aff. minoicus Pesce & Argano, 1980	PlagariAFF PERAL_200910_3_YCD378	PERAL	37,079	-2,101	Species_195	JQ921025	
	PminoicusAFF ARKADICA_201610_2_YAF294	ARKADICA	35,324	24,647	Species_303	NA	
Proasellus sp. aff. minoicus Pesce & Argano, 1980	PminoicusAFF ARKADICA_201610_3_YAF317	ARKADICA	35,324	24,647	Species_303	NA	
	PminoicusAFF ARKADIOU_201410_1ID_YAD070	ARKADIOU	35,324	24,648	Species_303	KY624032	
	PminoicusAFF ARKADIOU_201410_3_YAD155	ARKADIOU	35,324	24,648	Species_303	KY624033	
	PminoicusAFF KPM28_201610_1ID_YAF319	KPM28	35,311	24,837	Species_306	NA	
	Pcoxalis KPM28_201510_1ex_YAE887	KPM28	35,311	24,837	Species_306	NA	
	Psp AXOSPT1_201610_2ID_YAF314	AXOSPT1	35,309	24,854	Species_306	NA	
	Proasellus sp. aff. notenboomii Henry & Magniez, 1981	PnotenboomiiAFF SIKKAKOU_201509_1ID_YAE351	SIKKAKOU	34,867	-1,261	Species_281	NA
		PnotenboomiiAFF SIKKAKOU_201509_2_YAE306	SIKKAKOU	34,867	-1,261	Species_281	NA
	Proasellus sp. aff. n. sp. (MARTINO)	PnspMartinoAFF MONFIEIS_201004_1ID_YAB150	MONFIEIS	44,360	7,266	Species_381	KY624093
		PnspMartinoAFF MONFIEIS_201004_2_YAB314	MONFIEIS	44,360	7,266	Species_381	KY624094
Proasellus sp.	Pnsp12bK 102K_201107_1ID_YAB645	102K	43,750	39,810	Species_408	KY624073	
	Pnsp12bK 12bK_201107_1ID_YAB679	12BK	43,750	39,810	Species_408	KY624074	
Proasellus sp.	Psp140K 140K_201107_1_YAB858	140K	43,610	39,930	Species_387	KY624075	
	PnspAlvielaj ALVIELA_201010_1ID_YCD513	ALVIELA	39,446	-8,712	Species_205	JQ921337	
Proasellus n. sp. (ALVIELA)	PnspAlvielaj ALVIELA_201010_3_YCD641	ALVIELA	39,446	-8,712	Species_205	JQ921338	
	Proasellus n. sp. (AQUERRETA)	PnspAquerretaj AQUERRET_200906_1IDex_YAS978	AQUERRET	42,895	-1,554	Species_235	JQ921339
PnspAquerretaj AQUERRET_200906_4ex_YCD87		AQUERRET	42,895	-1,554	Species_235	JQ921340	
Proasellus n. sp. (ARDA)	PnspArdaj STORARDA_201504_1_YAE160	STORARDA	44,807	9,819	Species_248	NA	
	PnspArdaj STORARDA_201504_2_YAE231	STORARDA	44,807	9,819	Species_248	NA	
Proasellus n. sp. (ARPAON)	PnspArpaon ARPAON_201109_1ID_YAB954	ARPAON	44,300	3,663	Species_200	KY623782	
	PnspArpaon ARPAON_201407_C4_YCM588	ARPAON	44,300	3,663	Species_200	KY623784	
Proasellus n. sp. (BERTOLA)	PnspArpaon ARPAON_201407_D4_YCM589	ARPAON	44,300	3,663	Species_200	KY623785	
	PnspArpaon ARPAON_201407_A31D_YCM633	ARPAON	44,300	3,663	Species_200	KY623783	
	PnspArpaon CASSAGNA_201008_1_YAB868	CASSAGNA	44,271	3,730	Species_200	KY623786	
	PnspArpaon CASSAGNA_201008_2_YAB878	CASSAGNA	44,271	3,730	Species_200	KY623787	
	PnspArpaon CASSAGNA_201407_B2ID_YCM595	CASSAGNA	44,271	3,730	Species_200	KY623788	
	PnspArpaon CASSAGNA_201408_B5_YCM607	CASSAGNA	44,271	3,730	Species_200	KY623789	
	PnspArpaon COCURES_201506_1ID_YAE424	COCURES	44,352	3,616	Species_200	NA	
	PnspArpaon COCURES_201506_3_YAE412	COCURES	44,352	3,616	Species_200	NA	
	PnspArpaon GOUDECHE_201407_C1ID_YCM593	GOUDECHE	44,350	3,784	Species_200	KY623790	
	PnspArpaon GOUDECHE_201407_G3_YCM594	GOUDECHE	44,350	3,784	Species_200	KY623791	
	PnspArpaon GOUDECHE_201407_H3_YCM608	GOUDECHE	44,350	3,784	Species_200	KY623792	
	PnspArpaon MALLEPIQ_201303_1ID_YCM642	MALLEPIQ	43,707	2,346	Species_200	KY623793	
	PnspArpaon MALLEPIQ_201303_3_YCM643	MALLEPIQ	43,707	2,346	Species_200	KY623795	
	PnspArpaon MALLEPIQ_201303_2ID_YCM656	MALLEPIQ	43,707	2,346	Species_200	KY623794	
	Proasellus n. sp. (BERTOLA)	PnspBertolaj FONTTOSI_201504_1ID_YAD962	FONTTOSI	45,562	8,555	Species_165	NA
		PnspBertolaj FONTTOSI_201504_2ID_YAD963	FONTTOSI	45,562	8,555	Species_165	NA
	Proasellus n. sp. (BOREON)	PnspBoreon BOREON_200904_1ID_YAS784	BOREO	44,079	7,251	Species_342	JQ921341
		PnspBoreon BOREON_200904_2_YAS808	BOREO	44,079	7,251	Species_342	JQ921342



	PnspBoreon BOREON_200904_3_YAS813	BOREO	44,079	7,251	Species_342	JQ921343
	PnspBoreon PELAOUVE_201509_1ID_YAE437	PELAOUVE	44,035	7,295	Species_342	NA
	PnspBoreon PELAOUVE_201509_3_YAE394	PELAOUVE	44,035	7,295	Species_342	NA
	PnspBoreon STSMART2_200904_1_YLK570	STSMART2	44,083	7,245	Species_342	JQ921344
	PnspBoreon SAUTINEE_200904_2_YCD383	SAUTINEE	44,081	7,105	Species_343	JQ921345
	PnspBoreon VIONENE_201509_1ex_YAE415	VIONENE	44,083	7,102	Species_343	NA
	PnspBoreon VIONENE_201509_7arn_TRINITYDYN71804c4g2i1	VIONENE	44,083	7,102	Species_343	NA
	PnspBoreon VIONENE_201509_3ex_YAE449	VIONENE	44,083	7,102	Species_343	NA
Proasellus n. sp. (CALATSID)	Psp CALATSID_201610_1_YAF286	CALATSID	35,123	26,150	Species_304	NA
	Psp CALATSID_201610_2ID_YAF282	CALATSID	35,123	26,150	Species_304	NA
	Psp CALATSID_201610_3_YAF287	CALATSID	35,123	26,150	Species_304	NA
Proasellus n. sp. (CALORESP)	Psp CALORESP_201601_1_YAE724	CALORESP	40,490	15,210	Species_302	NA
	Psp CALORESP_201601_2_YAE834	CALORESP	40,490	15,210	Species_302	NA
Proasellus n. sp. (CHAMOIS)	PnspChamois CHAMOIS_201208_1_YAC476	CHAMOIS	44,052	6,699	Species_382	KY624076
	PnspChamois CHAMOIS_201308_2ID_YCM392	CHAMOIS	44,052	6,699	Species_382	KY624077
	PnspChamois CHAMOIS_201308_3_YCM413	CHAMOIS	44,052	6,699	Species_382	KY624078
Proasellus n. sp. (GABRIELE)	PspGabriele GABRIELE_201407_2_YAD072	GABRIELE	38,103	13,311	Species_275	KY624087
	PspGabriele GABRIELE_201407_1ID_YAD071	GABRIELE	38,103	13,311	Species_275	KY624086
Proasellus n. sp. (HFELETTTO)	PnspHfeletto HFELETTTO_201608_2ID_YAF775	HFELETTTO	45,309	7,755	Species_339	NA
	PnspHfeletto HFELETTTO_201608_1ID_YAF723	HFELETTTO	45,309	7,755	Species_339	NA
Proasellus n. sp. (JABRONH2)	PnspJabronH2 JABRONH2_201510_1_YAE451	JABRONH2	44,172	5,915	Species_385	NA
Proasellus n. sp. (KURUZOVICA)	PnspKurzovica KURUZOVICA_200611_1ID_YAB330	KURUZOVICA	44,994	15,651	Species_162	KY624088
Proasellus n. sp. (LENTEGI)	PnspLentegi LENTIGI_201010_1ID_YCD784	LENTIGI	36,832	-3,684	Species_191	JQ921346
	PnspLentegi LENTIGI_201010_2_YCD793	LENTIGI	36,832	-3,684	Species_191	JQ921347
	PnspLentegi LENTIGI_201010_3_YCD987	LENTIGI	36,832	-3,684	Species_191	JQ921348
Proasellus n. sp. (LESDUYES)	PspLesduyes LESDUYES_201007_1_YAC448	LESDUYES	44,041	6,103	Species_386	KY624089
	Psp VINONVER_201509_7_YAE531	VINONVER	43,724	5,794	Species_386	NA
Proasellus n. sp. (MAGLIA)	PnspMaglia MAGLIA_201507_2_YAF073	MAGLIA	43,972	7,475	Species_346	NA
	PnspMaglia BRUGIA_201108_1ex_YAB908	BRUGIA	43,957	7,502	Species_375	KY624090
	PnspMaglia MAGLIA_200908_1ID_YLK849	MAGLIA	43,972	7,475	Species_375	JQ921349
Proasellus n. sp. (MAGLIA)	PnspMaglia TORRACA_201409_2_YAD080	TORRACA	43,882	7,433	Species_384	KY624092
	PnspMaglia TORRACA_201409_1ID_YAD079	TORRACA	43,882	7,433	Species_384	KY624091
Proasellus n. sp. (MARTINO)	PnspMartino CRISSOLO_200801_1ID_YAS178	CRISSELO	44,698	7,142	Species_380	JQ921350
	PnspMartino CRISSOLO_200801_2ID_YAS199	CRISSELO	44,698	7,142	Species_380	JQ921351
	PnspMartino CRISSOLO_200801_3_YAS200	CRISSELO	44,698	7,142	Species_380	JQ921352
Proasellus n. sp. (MAZANDRAN)	PnspMazandran MAZANDRAN_200910_1_YCD42	MAZANDRAN	36,664	51,181	Species_407	JQ921336
Proasellus n. sp. (MECINA)	PnspMecina MECINA_201010_1ID_YAB864	MECINA	36,988	-3,150	Species_185	KY623896
	PnspMecina MECINA_201010_2_YAB331	MECINA	36,988	-3,150	Species_185	KY623897
	PnspMecina MECINA_201010_3_YAB985	MECINA	36,988	-3,150	Species_185	KY623898
Proasellus n. sp. (MESCLA)	PnspMescla MESCLA_200904_20_YAS198	MESCLA	43,913	7,186	Species_383	JQ921353
	PnspMescla MESCLA_200904_21_YAS206	MESCLA	43,913	7,186	Species_383	JQ921354
Proasellus n. sp. (MESKLASP)	Psp MESKLASP_201610_1ID_YAF311	MESKLASP	35,400	23,956	Species_305	NA
	Psp MESKLASP_201610_2ID_YAF315	MESKLASP	35,400	23,956	Species_305	NA
Proasellus n. sp. (MONTELIMAR)	PnspMontelimar ROUMONTE_201205_2_YAD414	ROUMONTE	44,552	4,748	Species_374	NA
Proasellus n. sp. (SANGUINIERS)	PnspSanguiniers SANGUINI_200908_1ID_YLK929	SANGUINI	44,252	6,773	Species_379	JQ921355
Proasellus n. sp. (SOSPTEL)	PnspSospel SOSPTEL_200908_1ID_YCD396	SOSPTEL	43,893	7,421	Species_344	JQ921356
	PnspSospel SOSPTEL_200908_4_YCD397	SOSPTEL	43,893	7,421	Species_344	JQ921357
	PnspSospel TORRACA_201409_1IDex1_YAD078	TORRACA	43,882	7,433	Species_344	KY624095
	PnspSospel TORRACA_201409_2ex2_YAD081	TORRACA	43,882	7,433	Species_344	KY624097
	PnspSospel TORRACA_201409_1IDex2_YAD144	TORRACA	43,882	7,433	Species_344	KY624096
Proasellus n. sp. (SPUZ)	PnspSpuz SPUZONCL_201107_2ID_YAC119	SPUZONCL	42,528	19,219	Species_122	KY624099
	PnspSpuz SPUZONCL_201107_1ID_YAB916	SPUZONCL	42,528	19,219	Species_122	KY624098
Proasellus n. sp. (STYLOSS2)	Psp STYLOSS2_201610_2_YAF279	STYLOSS2	35,434	24,126	Species_294	NA
	Psp STYLOSS2_201610_3_YAF280	STYLOSS2	35,434	24,126	Species_294	NA
	Psp STYLOSS2_201610_1_YAF318	STYLOSS2	35,434	24,126	Species_294	NA
Proasellus n. sp. (TENDE)	PnspTende TENDEAV_201108_1ID_YAB909	TENDEAV	44,073	7,597	Species_345	KY624102
	PnspTende TENDEAV_201108_2_YAC005	TENDEAV	44,073	7,597	Species_345	KY624103
	PnspTende FOUIGE_201108_2ex_YAC013	FOUIGE	44,045	7,577	Species_347	KY624101
	PnspTende FOUIGE_201108_1ex_YAC004	FOUIGE	44,045	7,577	Species_347	KY624100
Proasellus n. sp. (VERCLAUSE)	PnspVerclause VAUCLUS_200703_2ex_YLK602	VAUCLUS	43,919	5,132	Species_363	JQ921359
	PnspVerclause VAUCLUS_200703_5ex_YLK723	VAUCLUS	43,919	5,132	Species_363	JQ921360
	PnspVerclause BUECHAVA_201007_1ex_YCD810	BUECHAVA	44,202	5,921	Species_364	KY624104
	PnspVerclause LAPENNE_201002_1ID_YLK923	LAPENNE	44,242	5,227	Species_364	JQ921361
	PnspVerclause MONTMAUR_201003_6ex_YAB770	MONTMAUR	44,560	5,892	Species_364	KY624105
Proasellus n. sp. (VERCLAUSE)	PnspVerclause OUVSABLE_201204_3ID_YAC400	OUVSABLE	44,195	4,984	Species_364	KY624107
	PnspVerclause OUVSARR_201204_4ID_YAC401	OUVSARR	44,125	4,937	Species_364	KY624108
	PnspVerclause OUVSTAUB2_201204_1ID_YAC399	OUVSTAUB2	44,293	5,414	Species_364	KY624109
	PnspVerclause OUVAISO_201204_2ID_YAC438	OUVAISO	44,244	5,048	Species_364	KY624106
	PnspVerclause VERCLAUS_201002_1ID_YLK919	VERCLAUS	44,380	5,426	Species_364	JQ921358
Proasellus n. sp. 1 (HFELETTTO)	Pnsp1Hfeletto HFELETTTO_201608_4ID_YAF584	HFELETTTO	45,309	7,755	Species_341	NA
	Pnsp1Hfeletto HFELETTTO_201608_6_YAF697	HFELETTTO	45,309	7,755	Species_341	NA
Proasellus n. sp. 2 (CUVERA)	PspCuvera2 CUVERA_201108_4_YAB676	CUVERA	43,255	-3,604	Species_258	KY624079
Proasellus spelaes (Racovitz, 1922)	Pspelaes ETXANKO_200811_1_YAS522	ETXANKO	43,124	-0,971	Species_223	KC610466
	Pspelaes ETXANKO_200811_2ID_YAB907	ETXANKO	43,124	-0,971	Species_223	KC610505
Proasellus strouhali (Karaman, 1955)	Pstrouhali PONTMOUL_200707_1_YAS135	PONTMOUL	47,318	6,361	Species_365	JQ921430
	Pstrouhali SCENOIRE_200707_2_YAS134	SCENOIRE	47,321	6,439	Species_365	JQ921431
	Pstrouhali SCENOIRE_200707_4ID_YAS133	SCENOIRE	47,321	6,439	Species_365	JQ921432
	Pstrouhali BRETEMIN_201507_1ID_YAD861	BRETEMIN	46,210	4,514	Species_366	NA
	Pstrouhali BRETEMIN_201509_5arn_TRINITYDYN85300c2g1i1	BRETEMIN	46,210	4,514	Species_366	NA
	Pstrouhali BRETEMIN_201507_2_YAD961	BRETEMIN	46,210	4,514	Species_366	NA
Proasellus strouhali (Karaman, 1955)	Pstrouhali COLOMBIE_201705_2ex_YAG102	COLOMBIE	45,049	4,709	Species_366	NA
	Pstrouhali COLOMBIE_201705_3ex_YAG060	COLOMBIE	45,049	4,709	Species_366	NA
	Pstrouhali COQUILLE_200807_1ID_YAS328	COQUILLE	47,645	4,766	Species_366	JQ921419
	Pstrouhali COQUILLE_200807_2ID_YAS333	COQUILLE	47,645	4,766	Species_366	JQ921420
	Pstrouhali CRECEY_200804_1ID_YAS248	CRECEY	47,563	5,126	Species_366	JQ921421
	Pstrouhali CRECEY_200804_2_YAS261	CRECEY	47,563	5,126	Species_366	JQ921422
	Pstrouhali EMPURAHY_201705_1ex_YAG188	EMPURAHY	45,008	4,617	Species_366	NA
	Pstrouhali EMPURAHY_201705_2ex_YAG040	EMPURAHY	45,008	4,617	Species_366	NA
	Pstrouhali EMPURAHY_201705_3ex_YAG076	EMPURAHY	45,008	4,617	Species_366	NA
	Pstrouhali FONTMAIN_201006_1ID_YAB021	FONTMAIN	47,094	6,058	Species_366	KC610485
	Pstrouhali GAGGENAU_200703_1ID_YAS85	GAGGENAU	48,803	8,327	Species_366	JQ921423
	Pstrouhali GAGGENAU_200703_2_YAS87	GAGGENAU	48,803	8,327	Species_366	JQ921424
	Pstrouhali GAGGENAU_200703_3_YAS88	GAGGENAU	48,803	8,327	Species_366	JQ921425
	Pstrouhali GAISBER_201009_1ID_YCD592	GAISBER	47,848	13,332	Species_366	JQ921428
	Pstrouhali GLANBACH_201009_1ID_YCD594	GLANBACH	47,810	13,022	Species_366	JQ921435
	Pstrouhali GOLLING_201009_1ID_YCD593	GOLLING	47,582	13,171	Species_366	JQ921436
	Pstrouhali GRANBAUM_201006_1ID_YAB009	GRANBAUM	47,043	6,238	Species_366	JQ921439
	Pstrouhali LAMASTRE_201705_1ID_YAG075	LAMASTRE	44,989	4,565	Species_366	NA
	Pstrouhali LAMASTRE_201705_3_YAG039	LAMASTRE	44,989	4,565	Species_366	NA
	Pstrouhali MARBRENERI_200809_1_YAS387	MARBRENERI	46,058	4,563	Species_366	JQ921426
	Pstrouhali MARBRENERI_200809_2ID_YAS393	MARBRENERI	46,058	4,563	Species_366	JQ921427
	Pstrouhali MERLEVEN_201208_1ID_YCM412	MERLEVEN	47,727	-3,266	Species_366	KY624111
	Pstrouhali MONTBAR_201409_2IDex_YAE385	MONTBAR	47,011	5,630	Species_366	NA
	Pstrouhali NORGES_200700_2_YAS136	NORGES	47,407	5,079	Species_366	JQ921429

	Pstrouhal NORGES_200700_4ID_YAC828	NORGES	47,407	5,079	Species_366	KY624112
	Pstrouhal SONNLEI_201009_1_YCD667	SONNLEI	47,850	13,321	Species_366	KC610486
	Pstrouhal SOURCFIN_201006_1ID_YAB016	SOURCFIN	46,893	5,805	Species_366	JQ921437
	Pstrouhal SOURLOUE_201006_1ID_YAB010	SOURLOUE	47,011	6,299	Species_366	JQ921438
	Pstrouhal UNTERACH_201009_1ID_YCD591	UNTERACH	47,792	13,485	Species_366	JQ921433
	Pstrouhal UNTERACH_201009_3_YCD666	UNTERACH	47,792	13,485	Species_366	JQ921434
	Pstrouhal CF GRANCUIS_201006_2ID_YAB020	GRANCUIS	46,879	5,815	Species_366	JQ921119
<i>Proasellus synaselloides</i> (Henry, 1963)	Psynaselloides STANDRE_201507_1ID_YAE300	STANDRE	43,963	6,514	Species_391	NA
	Psynaselloides STANDRE_201507_2_YAE307	STANDRE	43,963	6,514	Species_391	NA
	PnspPontclot PONTCLOT_201507_1ID_YAE308	PONTCLOT	44,077	6,568	Species_391	NA
	PnspPontclot PONTCLOT_201507_2ID_YAE301	PONTCLOT	44,077	6,568	Species_391	NA
	Psynaselloides CHATEVER_200704_4ID_YAC477	CHATEVER	43,499	6,026	Species_392	KY624115
	Psynaselloides DUYESAMO_201509_1ID_YAE393	DUYESAMO	44,054	6,111	Species_392	NA
	Psynaselloides HYPODRAC_201509_1ID_YAE441	HYPODRAC	44,675	6,075	Species_392	NA
	Psynaselloides HYPODRAC_201509_2_YAE442	HYPODRAC	44,675	6,075	Species_392	NA
	Psynaselloides HYPODRAC_201509_3_YAE416	HYPODRAC	44,675	6,075	Species_392	NA
	Psynaselloides JABRON_201003_2_YAB022	JABRON	44,168	5,923	Species_392	KC610487
	Psynaselloides JABRON_201003_3_YAB779	JABRON	44,168	5,923	Species_392	KC610488
	Psynaselloides LESDUYES_201003_1ID_YAC402	LESDUYES	44,041	6,103	Species_392	KY624119
	Psynaselloides MONTMAUR_201003_1ID_YAD376	MONTMAUR	44,560	5,892	Species_392	KY624121
	Psynaselloides MONTMAUR_201003_3_YAB768	MONTMAUR	44,560	5,892	Species_392	KC610490
	Psynaselloides SENEZASS_201002_1ID_YLK924	SENEZASS	43,915	6,408	Species_392	JQ921440
	Psynaselloides SENEZASS_201002_2_YLK921	SENEZASS	43,915	6,408	Species_392	JQ921441
	Psynaselloides VANSON_201003_1ID_YAC403	VANSON	44,139	6,004	Species_392	KY624124
	Psynaselloides VARSOURCE_201410_1_YAD154	VARSOURCE	44,238	6,752	Species_392	KY624125
	Psynaselloides VARSOURCE_201410_3_YAD329	VARSOURCE	44,238	6,752	Species_392	KY624126
<i>Proasellus synaselloides</i> (Henry, 1963)	Psynaselloides VAUCLUS_200703_1_YAS718	VAUCLUS	43,919	5,132	Species_392	JQ921442
	Psynaselloides VAUCLUS_200703_3_YAS63	VAUCLUS	43,919	5,132	Species_392	JQ921443
	Psynaselloides VAUCLUS_200703_4ID_YAS65	VAUCLUS	43,919	5,132	Species_392	JQ921444
	Psynaselloides VINONVER_201509_1ID_YAE443	VINONVER	43,724	5,794	Species_392	NA
	Psynaselloides VIOLETTE_200707_1ID_YAS123	VIOLETTE	45,937	5,532	Species_392	JQ921447
	Psynaselloides VIOLETTE_200707_2ID_YAS124	VIOLETTE	45,937	5,532	Species_392	JQ921448
	Psynaselloides EYGUAM1_201007_3ex_YAC462	EYGUAM1	44,366	5,465	Species_393	KY624116
	Psynaselloides EYGUAM7_201007_4IDex_YAC444	EYGUAM7	44,380	5,426	Species_393	KY624117
	Psynaselloides EYGUAM8_201003_1ID_YAC802	EYGUAM8	44,344	5,771	Species_394	KY624118
	Psynaselloides LUSCROIX_201003_1ID_YAC414	LUSCROIX	44,657	5,711	Species_394	KY624120
	Psynaselloides MONTMAUR_201003_2_YAB778	MONTMAUR	44,560	5,892	Species_394	KC610489
	Psynaselloides CF LAPENNE_201002_1ID_YCD84	LAPENNE	44,242	5,227	Species_394	JQ921449
	Psynaselloides CF LAPENNE_201002_3ID_YAB860	LAPENNE	44,242	5,227	Species_394	KC610493
	Psynaselloides CF LAPENNE_201002_4ID_YAB316	LAPENNE	44,242	5,227	Species_394	KC610494
	Psynaselloides SABRE_201101_1_YAB148	SABRE	43,369	5,918	Species_395	KC610491
	Psynaselloides THEOULE_200810_2ID_YAS651	THEOULE	43,264	5,960	Species_396	JQ921445
	Psynaselloides THEOULE_200700_3_YAS161	THEOULE	43,264	5,960	Species_396	KC610492
	Psynaselloides THEOULE_200700_4_YAS137	THEOULE	43,264	5,960	Species_396	JQ921446
	Psynaselloides OUVSARR_201204_455ex_YAC411	OUVSARR	44,125	4,937	Species_397	KY624122
	Psynaselloides OUVSARR_201204_458ex_YAC453	OUVSARR	44,125	4,937	Species_397	KY624123
	PspDrobeaur DROBEAUR_201204_W0748ex_YAC697	DROBEAUR	44,553	5,525	Species_398	KY624113
	PspDroeur DROEURRE_201204_W0848ex_YAC698	DROEURRE	44,741	4,947	Species_399	KY624114
<i>Proasellus valdensis</i> (Chappuis, 1948)	Pvaldensis BALME_200709_1_YAS29	BALME	45,852	5,339	Species_367	JQ921452
	Pvaldensis BALME_200709_19ml_YML019 INCONNU	BALME	45,852	5,339	Species_367	JQ921450
	Pvaldensis BALME_200709_20ml_YSC70	BALME	45,852	5,339	Species_367	JQ921451
	Pvaldensis BENOITE_000000_1_YAS914	BENOITE	45,708	6,042	Species_367	KC610467
	Pvaldensis BENOITE_000000_2_YAS917	BENOITE	45,708	6,042	Species_367	KC610468
	Pvaldensis BENOITE_000000_3_YSC33	BENOITE	45,708	6,042	Species_367	JQ921453
	Pvaldensis BERNARD_200802_1_YAS941	BERNARD	46,102	6,780	Species_367	JQ921468
	Pvaldensis BERNARD_200802_2_YAB358	BERNARD	46,102	6,780	Species_367	KC610469
	Pvaldensis BERNARD_200802_3_YAS935	BERNARD	46,102	6,780	Species_367	KC610470
	Pvaldensis CAVALE_200511_27ml_YSC72	CAVALE	45,665	5,990	Species_367	JQ921454
	Pvaldensis CAVALE_200511_28ml_YML088 F1FSCOI	CAVALE	45,665	5,990	Species_367	JQ921455
<i>Proasellus valdensis</i> (Chappuis, 1948)	Pvaldensis CAVALE_200511_29ml_YML029 INCONNU	CAVALE	45,665	5,990	Species_367	JQ921456
	Pvaldensis CHRISTOP_200511_1ID_YSC36	CHRISTOP	45,438	4,438	Species_367	KC610471
	Pvaldensis CLAUDE_200504_24ml_YML024 INCONNU	CLAUDE	46,364	5,878	Species_367	JQ921484
	Pvaldensis CLAUDE_200504_25ml_YML025 INCONNU	CLAUDE	46,364	5,878	Species_367	KC610472
	Pvaldensis CLAUDE_200504_26ml_YSC73	CLAUDE	46,364	5,878	Species_367	JQ921485
	Pvaldensis CUVES_200906_1ID_YAS597	CUVES	45,209	5,659	Species_367	JQ921480
	Pvaldensis CUVES_200906_2_YAS650	CUVES	45,209	5,659	Species_367	JQ921481
	Pvaldensis CUVES_200906_3_YSC43	CUVES	45,209	5,659	Species_367	JQ921482
	Pvaldensis DOUVERA_200908_1ID_YSC35	DOUVERA	46,318	5,834	Species_367	JQ921469
	Pvaldensis DOUVERA_200908_2_YSC29	DOUVERA	46,318	5,834	Species_367	JQ921470
	Pvaldensis DOUVERA_200908_3_YSC30	DOUVERA	46,318	5,834	Species_367	JQ921471
	Pvaldensis FALCON_200806_1_YAS939	FALCON	45,842	5,547	Species_367	JQ921458
	Pvaldensis FALCON_200806_2_YAS933	FALCON	45,842	5,547	Species_367	KC610473
	Pvaldensis FALCON_200806_3_YAS562	FALCON	45,842	5,547	Species_367	KC610474
	Pvaldensis FARDELE_200704_1f_YCD306	FARDELE	46,025	6,773	Species_367	JQ921462
	Pvaldensis FARDELE_200704_2_YAS921	FARDELE	46,025	6,773	Species_367	KC610475
	Pvaldensis FARDELE_200704_3_YSC87	FARDELE	46,025	6,773	Species_367	NA
<i>Proasellus valdensis</i> (Chappuis, 1948)	Pvaldensis FARDELW_200602_17ml_YSC97	FARDELW	46,025	6,773	Species_367	JQ921459
	Pvaldensis FARDELW_200602_18ml_YSC98	FARDELW	46,025	6,773	Species_367	JQ921460
	Pvaldensis FARDELW_200602_22ml_YSC69	FARDELW	46,025	6,773	Species_367	JQ921461
	Pvaldensis FONTAGNY_200602_2_YSC41	FONTAGNY	46,071	6,805	Species_367	KC610476
	Pvaldensis FOULESC_200710_1ID_YAS154	FOULESC	46,377	5,896	Species_367	JQ921463
	Pvaldensis FOULESC_200710_3ID_YAS152	FOULESC	46,377	5,896	Species_367	JQ921464
	Pvaldensis FROIDIER_201307_2_YCM852	FROIDIER	45,958	5,547	Species_367	KY624128
	Pvaldensis FROIDIER_201307_1ID_YCM407	FROIDIER	45,958	5,547	Species_367	KY624127
	Pvaldensis FURON_200808_1_YAS588	FURON	45,167	5,612	Species_367	JQ921457
	Pvaldensis GERMINAL_201612_1ID_YAF828	GERMINAL	45,794	6,108	Species_367	NA
	Pvaldensis GERMINAL_201612_2_YAF653 jg LCO1490	GERMINAL	45,794	6,108	Species_367	NA
	Pvaldensis GRODIAU_201501_1ID_YAE430	GRODIAU	45,958	6,283	Species_367	NA
	Pvaldensis HUGBIS_200709_2_YAS30	HUGBIS	46,032	5,769	Species_367	JQ921465
	Pvaldensis HUGBIS_200709_3_YAS833	HUGBIS	46,032	5,769	Species_367	JQ921466
	Pvaldensis HUGBIS_200709_4_YAS31	HUGBIS	46,032	5,769	Species_367	JQ921467
	Pvaldensis HYPOROCH_201508_1ID_YAE414	HYPOROCH	45,072	5,497	Species_367	NA
	Pvaldensis HYPOROCH_201508_2_YAE453	HYPOROCH	45,072	5,497	Species_367	NA
	Pvaldensis JOSEPHGR_201607_1ID_YAF728	JOSEPHGR	45,781	5,285	Species_367	NA
	Pvaldensis JOSEPHGR_201607_2_YAF758	JOSEPHGR	45,781	5,285	Species_367	NA
	Pvaldensis LONGIROD_201406_1ID_YAD042	LONGIROD	46,530	6,231	Species_367	KY624129
	Pvaldensis MORREARCE_201006_1_YDE008	MORREARCE	46,512	6,028	Species_367	KC610477
	Pvaldensis MOUDOUBS_201006_1_YDE007	MOUDOUBS	46,705	6,209	Species_367	KC610478
	Pvaldensis PLAGNE_200511_33ml_YSC74	PLAGNE	45,440	5,904	Species_367	JQ921472
	Pvaldensis PLAGNE_200511_34ml_YSC100	PLAGNE	45,440	5,904	Species_367	JQ921473
	Pvaldensis PLEUREUS_200509_1_YSC67	PLEUREUS	46,012	6,759	Species_367	JQ921476
	Pvaldensis PLEUREUS_200509_13ml_YSC94	PLEUREUS	46,012	6,759	Species_367	JQ921474
	Pvaldensis PLEUREUS_200509_14ml_YSC95	PLEUREUS	46,012	6,759	Species_367	JQ921475
	Pvaldensis SALLES_200509_7_YSC99	SALLES	46,001	6,754	Species_367	JQ921479

	Pvaldensis[SALLES_200509_15ml_YSC96	SALLES	46,001	6,754	Species_367	JQ921477
	Pvaldensis[SALLES_200509_16ml_YSC68	SALLES	46,001	6,754	Species_367	JQ921478
	Pvaldensis[SCESIXT_200704_1_YAS916	SCESIXT	46,057	6,784	Species_367	JQ921483
	Pvaldensis[STAUPRE_201411_1ex_YAD147	STAUPRE	45,398	5,674	Species_367	KY624130
	Pvaldensis[STAUPRE_201411_2ex_YAD066	STAUPRE	45,398	5,674	Species_367	KY624131
<i>Proasellus vandeli</i> (Magniez & Henry, 1969)	Pvandeli[LAMINA_200811_1_YAS566	LAMINA	43,145	-0,926	Species_222	JQ921488
	Pvandeli[CHAPGEN_200811_3ID_YLK510	CHAPGEN	43,083	-0,873	Species_225	JQ921486
	Pvandeli[CHAPGEN_200811_4_YSC880]INCONNU	CHAPGEN	43,083	-0,873	Species_225	JQ921487
	Pvandeli[UTHURRIA_201005_1ID_YCD849	UTHURRIA	43,234	-0,917	Species_226	JQ921489
	Pvandeli[UTHURRIA_201005_2_YCD922	UTHURRIA	43,234	-0,917	Species_226	JQ921490
	Pvandeli[UTHURRIA_201005_3_YCD981	UTHURRIA	43,234	-0,917	Species_226	JQ921491
<i>Proasellus vulgaris</i> (Sket, 1965)	Pvulgaris[PRADIELI_201405_1IDex_YCM935	PRADIELI	46,289	13,269	Species_405	KY624132
	Pvulgaris[PRADIELI_201405_2ex_YCM936	PRADIELI	46,289	13,269	Species_405	KY624133
	Pvulgaris[PRADIELI_201405_3ex_YCM934	PRADIELI	46,289	13,269	Species_405	KY624134
	Pvulgaris[DOVJEZ_200910_1ex_YLK922	DOVJEZ	46,116	14,475	Species_406	JQ921492
	Pvulgaris[DOVJEZ_200910_1ID_YLK747	DOVJEZ	46,116	14,475	Species_406	JQ921494
	Pvulgaris[DOVJEZ_200910_2ex_YLK926	DOVJEZ	46,116	14,475	Species_406	JQ921493
<i>Proasellus walteri</i> (Chappuis, 1948)	Pwalteri[BURKH1_200905_1ex_YCD391	BURKH1	48,092	7,599	Species_350	JQ921498
	Pwalteri[ROSSFELD_200911_1ID_YLK750	ROSSFELD	48,334	7,631	Species_350	JQ921516
	Pwalteri[ROSSFELD_200911_2_YCD280	ROSSFELD	48,334	7,631	Species_350	JQ921517
	Pwalteri[STMATERN_201409_1ID_YAE435	STMATERN	48,366	7,622	Species_350	NA
	Pwalteri[AINR07AM_201507_1ex_YAF974	AINR07AM	46,009	5,305	Species_351	NA
<i>Proasellus walteri</i> (Chappuis, 1948)	Pwalteri[AINR07AM_201507_2ex_YAG054	AINR07AM	46,009	5,305	Species_351	NA
	Pwalteri[AINR29AM_201507_4_YAG196	AINR29AM	45,889	5,232	Species_351	NA
	Pwalteri[AINR31AV_201507_6ID_YAF975	AINR31AV	45,860	5,226	Species_351	NA
	Pwalteri[AZERGUES_200700_1_YAS263	AZERGUES	45,929	4,551	Species_351	JQ921495
	Pwalteri[AZERGUES_200700_2ID_YAS302	AZERGUES	45,929	4,551	Species_351	JQ921496
	Pwalteri[BESSEGES_201407_A1ID_YCM591	BESSEGES	44,297	4,075	Species_351	KY624140
	Pwalteri[BESSEGES_201407_E2_YCM613	BESSEGES	44,297	4,075	Species_351	KY624141
	Pwalteri[BESSEGES_201407_F2_YCM620	BESSEGES	44,297	4,075	Species_351	KY624142
	Pwalteri[CEZEBAGN_201605_1ID_YAF049	CEZEBAGN	44,164	4,634	Species_351	NA
	Pwalteri[CEZEBAGN_201605_3_YAF036	CEZEBAGN	44,164	4,634	Species_351	NA
	Pwalteri[CHAMPO1_201704_2ID_YAF971	CHAMPO1	45,135	5,012	Species_351	NA
	Pwalteri[CHAMPO1_201704_2ex_YAF972	CHAMPO1	45,135	5,012	Species_351	NA
	Pwalteri[CHAMPO1_201704_3_YAF987	CHAMPO1	45,135	5,012	Species_351	NA
	Pwalteri[COUXHY_201705_1ID_YAG046	COUXHY	44,733	4,634	Species_351	NA
	Pwalteri[COUXHY_201705_2_YAG059	COUXHY	44,733	4,634	Species_351	NA
	Pwalteri[DROAOUST_201103_15_YAC291	DROAOUST	44,714	5,058	Species_351	KY624149
	Pwalteri[DROEURRE_201204_W0798_YAE395	DROEURRE	44,741	4,947	Species_351	NA
	Pwalteri[DROEURRE_201204_W0806_YAE411	DROEURRE	44,741	4,947	Species_351	NA
	Pwalteri[DROLABAT_201004_1_YAB824	DROLABAT	44,512	5,622	Species_351	KC610495
	Pwalteri[DROLABAT_201004_2_YAB772	DROLABAT	44,512	5,622	Species_351	KC610496
	Pwalteri[DROLIVRO_201004_1_YDE010	DROLIVRO	44,762	4,841	Species_351	KC610497
	Pwalteri[DROLIVRO_201004_2_YAB812	DROLIVRO	44,762	4,841	Species_351	KC610498
	Pwalteri[ENTRECHA_201007_2_YAC396	ENTRECHA	44,239	5,162	Species_351	KY624150
	Pwalteri[FALGIERE_201407_H4ID_YCM590	FALGIERE	44,128	3,892	Species_351	KY624152
	Pwalteri[FALGIERE_201407_G5_YCM619	FALGIERE	44,128	3,892	Species_351	KY624151
	Pwalteri[GRIGNY1_201704_1ID_YAG057	GRIGNY1	45,605	4,796	Species_351	NA
	Pwalteri[GRIGNY1_201704_2_YAF966	GRIGNY1	45,605	4,796	Species_351	NA
	Pwalteri[HBUGUEIR_201609_1ID_YAF576	HBUGUEIR	44,077	4,108	Species_351	NA
	Pwalteri[HBUGUEIR_201609_3_YAF699	HBUGUEIR	44,077	4,108	Species_351	NA
	Pwalteri[HCOLLIAS_201605_2_YAF124	HCOLLIAS	43,954	4,483	Species_351	NA
	Pwalteri[HCOLLIAS_201605_1ID_YAF123	HCOLLIAS	43,954	4,483	Species_351	NA
	Pwalteri[HCOLLIAS_201605_3_YAF125	HCOLLIAS	43,954	4,483	Species_351	NA
	Pwalteri[HMASCAIN_201605_1ID_YAF115	HMASCAIN	44,112	3,859	Species_351	NA
	Pwalteri[HMASCAIN_201605_3_YAF116	HMASCAIN	44,112	3,859	Species_351	NA
	Pwalteri[LAPENNE_201002_2ex_YLK935	LAPENNE	44,242	5,227	Species_351	JQ921520
	Pwalteri[LAVABREI_201609_1ID_YAF624	LAVABREI	44,168	4,053	Species_351	NA
	Pwalteri[MIRIBEL_201704_1ID_YAF965	MIRIBEL	45,810	5,062	Species_351	NA
	Pwalteri[MIRIBEL_201704_2_YAF993	MIRIBEL	45,810	5,062	Species_351	NA
	Pwalteri[MONTJOU1_201002_1ID_YLK937	MONTJOU1	44,502	5,089	Species_351	JQ921503
	Pwalteri[MONTJOU1_201002_2_YLK928	MONTJOU1	44,502	5,089	Species_351	JQ921504
	Pwalteri[OUBVEDAR_201204_472_YAC393	OUBVEDAR	44,053	4,930	Species_351	KY624161
	Pwalteri[OUVSABLE_201204_431_YAC391	OUVSABLE	44,195	4,984	Species_351	KY624162
	Pwalteri[OUVSARR_201204_441_YAC392	OUVSARR	44,125	4,937	Species_351	KY624163
	Pwalteri[OUVVAISO_201007_1_YAC394	OUVVAISO	44,244	5,048	Species_351	KY624164
	Pwalteri[ROUBCRUP_201003_1ID_YAC395	ROUBCRUP	44,571	5,159	Species_351	KY624165
	Pwalteri[ROUBCRUP_201003_2_YAC481	ROUBCRUP	44,571	5,159	Species_351	KY624166
	Pwalteri[ROUBSAUZ_201003_2_YDE009	ROUBSAUZ	44,591	4,819	Species_351	KC610499
	Pwalteri[ROUBSAUZ_201003_3_YAB811	ROUBSAUZ	44,591	4,819	Species_351	KC610500
	Pwalteri[SOUSPIRO_200700_2ID_YAS158	SOUSPIRO	43,908	5,090	Species_351	JQ921518
	Pwalteri[SOUSPIRO_200700_4_YAS164	SOUSPIRO	43,908	5,090	Species_351	JQ921519
<i>Proasellus walteri</i> (Chappuis, 1948)	Pwalteri[STMAURHY_201704_1ID_YAG061	STMAURHY	44,299	4,569	Species_351	NA
	Pwalteri[STMAURHY_201704_2_YAF970	STMAURHY	44,299	4,569	Species_351	NA
	Pwalteri[VERCLAUS_201002_1IDex_YLK968	VERCLAUS	44,380	5,426	Species_351	JQ921501
	Pwalteri[VERCLAUS_201002_2ex_YLK927	VERCLAUS	44,380	5,426	Species_351	JQ921502
	Pwalteri[VIOLETTE_200707_1ID_YAS463	VIOLETTE	45,937	5,532	Species_351	JQ921513
	Pwalteri[VIOLETTE_200707_2ID_YAS126	VIOLETTE	45,937	5,532	Species_351	JQ921514
	Pwalteri[VIOLETTE_200707_3ID_YAS127	VIOLETTE	45,937	5,532	Species_351	JQ921515
	Pwalteri[BERTHEZE_201407_D5_YCM599	BERTHEZE	44,092	3,636	Species_352	KY624137
	Pwalteri[BERTHEZE_201407_E5_YCM600	BERTHEZE	44,092	3,636	Species_352	KY624138
	Pwalteri[BERTHEZE_201407_G4ID_YCM615	BERTHEZE	44,092	3,636	Species_352	KY624139
	Pwalteri[CHINIER_201407_B1ID_YCM592	CHINIER	44,077	3,662	Species_352	KY624143
	Pwalteri[CHINIER_201407_B3ID_YCM612	CHINIER	44,077	3,662	Species_352	KY624144
	Pwalteri[CHINIER_201407_C3_YCM601	CHINIER	44,077	3,662	Species_352	KY624145
	Pwalteri[COULAROU_201407_F1ID_YCM604	COULAROU	43,984	3,623	Species_352	KY624147
	Pwalteri[COULAROU_201407_E3_YCM602	COULAROU	43,984	3,623	Species_352	KY624146
	Pwalteri[COULAROU_201407_F3_YCM614	COULAROU	43,984	3,623	Species_352	KY624148
	Pwalteri[HASPIRAN_201609_1ID_YAF588	HASPIRAN	43,567	3,472	Species_352	NA
	Pwalteri[HISSENAC_201609_1ID_YAF642	HISSENAC	43,845	3,710	Species_352	NA
	Pwalteri[HISSENAC_201609_3_YAF701	HISSENAC	43,845	3,710	Species_352	NA
	Pwalteri[HYPOPRAT_201407_D1ID_YCM603	HYPOPRAT	44,011	3,675	Species_352	KY624155
	Pwalteri[HYPOPRAT_201407_A4_YCM586	HYPOPRAT	44,011	3,675	Species_352	KY624153
	Pwalteri[HYPOPRAT_201407_B4_YCM587	HYPOPRAT	44,011	3,675	Species_352	KY624154
	Pwalteri[PTDIABLE_201609_1ex_YAF589	PTDIABLE	43,703	3,558	Species_352	NA
	Pwalteri[BELMONT_200707_1ID_YAS64	BELMONT	47,009	5,595	Species_353	JQ921497
	Pwalteri[COLLIER_200804_1ID_YAS246	COLLIER	46,067	4,578	Species_353	JQ921510
	Pwalteri[COLLIER_200804_2_YAS262	COLLIER	46,067	4,578	Species_353	JQ921511
	Pwalteri[COLLIER_200804_3_YAS301	COLLIER	46,067	4,578	Species_353	JQ921512
	Pwalteri[COQUILLE_200807_1ID_YAS326	COQUILLE	47,645	4,766	Species_353	JQ921499
	Pwalteri[COQUILLE_200807_2ID_YAS332	COQUILLE	47,645	4,766	Species_353	JQ921500
	Pwalteri[LONGAMON_200804_1ID_YAS245	LONGAMON	47,287	5,079	Species_353	JQ921505
	Pwalteri[LONGAMON_200804_2_YAS264	LONGAMON	47,287	5,079	Species_353	JQ921506
	Pwalteri[LONGAMON_200804_3_YAS279	LONGAMON	47,287	5,079	Species_353	JQ921507

	Pwalteri MONTBAR_200707_1_YAS61	MONTBAR	47,011	5,630	Species_353	JQ921508
	Pwalteri MONTBAR_200707_2_YAS62	MONTBAR	47,011	5,630	Species_353	JQ921509
	Pwalteri NARDIABL_200707_1_ID_YAG104	NARDIABL	47,317	6,369	Species_353	NA
	Pwalteri NARDIABL_200707_2_ID_YAG051	NARDIABL	47,317	6,369	Species_353	NA
	Pwalteri NORGES_200807_9_ID_YAC397	NORGES	47,407	5,079	Species_353	KY624160
	Pwalteri GROSPHY_201704_1_ID_YAF969	GROSPHY	44,416	4,288	Species_354	NA
	Pwalteri GROSPHY_201704_2_YAF983	GROSPHY	44,416	4,288	Species_354	NA
	Pwalteri AUJAC_201407_H5ID_YCM609	AUJAC	44,359	3,996	Species_355	KY624136
	Pwalteri AUJAC_201407_A6_YCM598	AUJAC	44,359	3,996	Species_355	KY624135
	Pwalteri VIELVIC_201407_A2ID_YCM605	VIELVIC	44,408	3,949	Species_355	KY624167
	Pwalteri MEYRASHY_201704_1_ID_YAG049	MEYRASHY	44,670	4,270	Species_356	NA
	Pwalteri MEYRASHY_201704_2_YAF968	MEYRASHY	44,670	4,270	Species_356	NA
	Pwalteri SAMPZON_201705_1_ID_YAG111	SAMPZON	44,424	4,353	Species_357	NA
	Pwalteri SAMPZON_201705_2_YAG077	SAMPZON	44,424	4,353	Species_357	NA
	Pwalteri SAMPZON_201705_3_YAG107	SAMPZON	44,424	4,353	Species_357	NA
	Pwalteri STPRIVAT_201705_1_ID_YAG078	STPRIVAT	44,266	3,870	Species_358	NA
	Pwalteri STPRIVAT_201705_2_YAG106	STPRIVAT	44,266	3,870	Species_358	NA
	Pwalteri STPRIVAT_201705_3_YAG042	STPRIVAT	44,266	3,870	Species_358	NA
<i>Proasellus walteri</i> (Chappuis, 1948)	Pwalteri STPCOSTE_201705_2_YAG103	STPCOSTE	44,164	3,974	Species_359	NA
	Pwalteri STPCOSTE_201705_3_YAG044	STPCOSTE	44,164	3,974	Species_359	NA
	Pwalteri HMASCAIN_201605_7_YAF973	HMASCAIN	44,112	3,859	Species_360	NA
	Pwalteri HMASCAIN_201605_8_YAF977	HMASCAIN	44,112	3,859	Species_360	NA
	Pwalteri HMASCAIN_201605_9_YAF976	HMASCAIN	44,112	3,859	Species_360	NA
	Pwalteri HMASCAIN_201605_2ID_YAF052	HMASCAIN	44,112	3,859	Species_360	NA
	Pwalteri HSAUMANE_201605_1ex_YAF139	HSAUMANE	44,119	3,763	Species_361	NA
	Pwalteri HSAUMANE_201605_2IDex_YAF028	HSAUMANE	44,119	3,763	Species_361	NA
	Pwalteri LEMARTIN_201407_D2ID_YCM616	LEMARTIN	44,157	3,841	Species_362	KY624156
	Pwalteri LEMARTIN_201407_G2_YCM606	LEMARTIN	44,157	3,841	Species_362	KY624157
	Pwalteri MOULINDU_201407_A5ID_YCM597	MOULINDU	44,197	3,702	Species_362	KY624158
	Pwalteri MOULINDU_201407_C6_YCM622	MOULINDU	44,197	3,702	Species_362	KY624159
<i>Proasellus wolff</i> Dudich, 1925	Pwolff FTCIANE_201405_2_YAD335	FTCIANE	37,042	15,235	Species_286	KY623957
	Pwolff IRMINIO_201307_1ID_YAD362	IRMINIO	36,789	14,602	Species_286	KY623958
	Psp_gr_coxalis ARODESSP_201603_1ID_YAE721	ARODESSP	34,940	32,400	Species_286	NA
	Psp_gr_coxalis ARODESSP_201603_3_YAE762	ARODESSP	34,940	32,400	Species_286	NA
<i>Proasellus meridianus xavieri</i> (Braga, 1956)	Pxavieri FERMENTE_201010_1ID_YCD492	FERMENTE	40,552	-8,499	Species_246	JQ921521
	Pxavieri FERMENTE_201010_2_YCD540	FERMENTE	40,552	-8,499	Species_246	JQ921522
	Pxavieri FERMENTE_201010_3_YCD577	FERMENTE	40,552	-8,499	Species_246	JQ921523
<i>Proasellus n. sp.</i> (ETCHEBERRIGARAY)	Pnsp Etcheberrigaray ETCHEVE_200811_1ex_YAS563	ETCHEVE	43,138	-1,200	Species_224	JQ921416
	Pnsp Etcheberrigaray ETCHEVE_200811_2ex_YAS504	ETCHEVE	43,138	-1,200	Species_224	JQ921417
	Pnsp Etcheberrigaray ETCHEVE_200811_3IDex_YAS576	ETCHEVE	43,138	-1,200	Species_224	JQ921418
	Pnsp Etcheberrigaray ETCHEVE_201102_12_YCM367	ETCHEVE	43,138	-1,200	Species_224	KY624082
	Pnsp Etcheberrigaray ETCHEVE_201102_13_YCM368	ETCHEVE	43,138	-1,200	Species_224	KY624083
	Pnsp Etcheberrigaray ETCHEVE_201102_14_YCM369	ETCHEVE	43,138	-1,200	Species_224	KY624084
	Pnsp Etcheberrigaray ETCHEVE_201102_15_YCM370	ETCHEVE	43,138	-1,200	Species_224	KY624085
<i>Protelsonia hungarica hungarica</i> Méhely, 1924	PROThungarica ABALIG7_201009_1ID_YAB153	ABALIG7	46,137	18,117	Species_4	KY624168
	PROThungarica MANFA2_201012_1ID_YAB194	MANFA2	46,149	18,210	Species_4	KY624169
	PROThungarica VADETE4_201012_1FID_YAB168	VADETE4	46,122	18,114	Species_4	KY624170
<i>Remasellus parvus</i> (Steeves, 1964)	REMParvus THUNDER_000000_0_AY566472 NCBI	THUNDER	30,416	-83,214	Species_106	AY566472
<i>Stenasellus breuili</i> Racovitz, 1924	Sbreuili TURGOYE_201504_3am_TRINITYDIN1340c3q21	ITURGOYE	43,362	-2,702	Species_34	NA
	Sbreuili TURGOYE_201504_1_YAD951	ITURGOYE	43,362	-2,702	Species_34	NA
	Sbreuili TURGOYE_201504_2_YAE117	ITURGOYE	43,362	-2,702	Species_34	NA
	Sbreuili ALLI_200906_1ID_YLK176	ALLI	42,988	-1,893	Species_35	JQ921607
	Sbreuili ALLI_200906_2_YLK391	ALLI	42,988	-1,893	Species_35	JQ921608
	Sbreuili ALLI_200906_3_YLK349	ALLI	42,988	-1,893	Species_35	JQ921609
<i>Stenasellus buili</i> Remy, 1949	Sbuili CAUNENG_197112_13_YLK79	CAUNENG	42,869	2,874	Species_6	JQ921613
	Sbuili CAUNENG_197112_2_YLK20	CAUNENG	42,869	2,874	Species_6	JQ921614
	Sbuili BAGATELL_201407_E1ex_YCM611	BAGATELL	43,991	3,623	Species_7	KY624171
	Sbuili CENTFONT_200908_1_YLK831	CENTFONT	43,759	3,623	Species_7	JQ921610
	Sbuili FONTANIL_200908_1ID_YLK810	FONTANIL	43,753	3,623	Species_7	JQ921611
	Sbuili FONTANIL_200908_2_YLK815	FONTANIL	43,753	3,623	Species_7	JQ921612
	Sbuili HISSENAC_201609_1ex_YAF719	HISSENAC	43,845	3,710	Species_7	NA
	Sbuili LASFONS2_201407_C2ex_YCM634	LASFONS2	43,968	3,550	Species_7	KY624172
	Sbuili LASFONS2_201407_H1ex_YCM610	LASFONS2	43,968	3,550	Species_7	KY624173
	Sbuili PTDIABLE_201609_1ID_YAF593	PTDIABLE	43,703	3,558	Species_7	NA
	Sbuili PTDIABLE_201609_2_YAF664	PTDIABLE	43,703	3,558	Species_7	NA
<i>Stenasellus galhanoae</i> Braga, 1962	Sgalhanoae MORA_201010_1ID_YCD781	MORA	37,786	-6,833	Species_42	JQ921615
	Sgalhanoae MORA_201010_2_YAB780	MORA	37,786	-6,833	Species_42	JQ921616
	Sgalhanoae MORA_201010_2_YAB865	MORA	37,786	-6,833	Species_43	KY624174
<i>Stenasellus nuragicus</i> Argano, 1968	Snuragicus GRGIOVAN_201410_1FID_YAE132	GRGIOVAN	39,337	8,628	Species_53	NA
<i>Stenasellus racovitzai</i> Razzutti, 1925	Sracovitzai ASINAIA_200810_1_YLK81	ASINAIA	42,493	9,295	Species_36	KY624177
	Sracovitzai BERNACCI_200810_1ID_YLK73	BERNACCI	42,546	9,137	Species_36	JQ921619
	Sracovitzai BERNACCI_200810_2_YLK70	BERNACCI	42,546	9,137	Species_36	JQ921620
	Sracovitzai CASALUNA_200810_1ID_YLK52	CASALUNA	42,388	9,259	Species_36	JQ921623
	Sracovitzai CASALUNA_200810_2_YLK66	CASALUNA	42,388	9,259	Species_36	JQ921624
	Sracovitzai FIUMALTO_200810_1_YLK67	FIUMALTO	42,445	9,473	Species_36	JQ921618
	Sracovitzai CASAMOZA_200810_1_YLK44	CASAMOZA	42,519	9,442	Species_37	JQ921621
	Sracovitzai CASAMOZA_200810_2ID_YLK69	CASAMOZA	42,519	9,442	Species_37	JQ921622
	Sracovitzai SOLENZA_200908_1ID_YLK573	SOLENZA	41,866	9,388	Species_38	JQ921625
	Sracovitzai SOLENZA_200908_2_YLK672	SOLENZA	41,866	9,388	Species_38	JQ921626
	Sracovitzai BUDANESE_201509_1_YAE705	BUDANESE	42,850	11,120	Species_39	NA
	Sracovitzai BUDANESE_201509_2_YAE733	BUDANESE	42,850	11,120	Species_39	NA
<i>Stenasellus n. sp.</i> (ACSASSAR)	Ssp ACSASSAR_200905_1_YAD312	ACSASSAR	40,718	8,605	Species_31	KY624178
	Ssp ACSASSAR_200905_2_YAD233	ACSASSAR	40,718	8,605	Species_31	KY624179
<i>Stenasellus n. sp.</i> (BUEMARI)	Ssp BUEMARI_200905_1_YAD315	BUEMARI	40,247	9,624	Species_32	KY624180
	Ssp FRUNCU_201209_1_YAD234	FRUNCU	40,564	9,678	Species_33	KY624181
<i>Stenasellus virei</i> sp. aff. <i>buchneri</i> Stammer, 1936	Svireibuchneri AFF CONCHA_201010_1ID_YCD809	CONCHA	43,560	-6,190	Species_10	JQ921627
	Svireibuchneri AFF CONCHA_201010_2_YCD868	CONCHA	43,560	-6,190	Species_10	JQ921628
	Svireibuchneri AFF CONCHA_201010_2ex_YCD857	CONCHA	43,560	-6,190	Species_10	JQ921572
<i>Stenasellus n. sp.</i> (MEXILHOE)	Ssp Mexilhoe MEXILHOE_200909_1ID_YAB151	MEXILHOE	37,161	-8,613	Species_41	KY624175
	Ssp Mexilhoe MEXILHOE_200909_2FID_YAB391	MEXILHOE	37,161	-8,613	Species_41	KY624176
<i>Stenasellus n. sp.</i> (MOGON)	Ssp Mogon MOGON_200910_1ID_YAB166	MOGON	38,070	-3,037	Species_40	KY624182
<i>Stenasellus virei</i> Dollfus, 1897	Ssp Nacason NACASON_201110_1ex_YAC214	NACASON	43,202	-3,587	Species_14	KY624184
	Svirei OUEILNEE_200811_1IDex_YLK496	OUEILNEE	43,150	-0,405	Species_11	JQ921657
	Svirei OUEILNEE_200811_2ex_YCD125	OUEILNEE	43,150	-0,405	Species_11	JQ921658
	Svirei MONBEQUI_201304_1ex_YCM393	MONBEQUI	43,893	1,208	Species_19	KY624187
	Svirei LAALFRAN_201300_1ID_YCM394	LAALFRAN	41,592	-0,767	Species_27	KY624185
	Svirei LAALFRAN_201300_2_YCM395	LAALFRAN	41,592	-0,767	Species_27	KY624186
<i>Stenasellus virei angelieri</i> Magniez, 1968	Svireiangelier ALBANYA_200906_1ID_YLK131	ALBANYA	42,305	2,720	Species_5	JQ921629
	Svireiangelier ORFES_200906_1FID_YLK417	ORFES	42,171	2,869	Species_5	JQ921630
	Svireiangelier ORFES_200906_2_YLK402	ORFES	42,171	2,869	Species_5	JQ921631
<i>Stenasellus virei boui</i> Magniez, 1968	Svireiboui NERTNAPP_196802_1_YLK31	NERTNAPP	42,945	1,175	Species_21	JQ921632
	Svireiboui NERTNAPP_196802_2_YLK13	NERTNAPP	42,945	1,175	Species_21	JQ921633
<i>Stenasellus virei buchneri</i> Stammer, 1936	Svireibuchneri FUESENTA_200906_1FID_YLK110	FUESENTA	43,346	-5,483	Species_9	JQ921638
	Svireibuchneri FUESENTA_200906_2_YLK377	FUESENTA	43,346	-5,483	Species_9	JQ921639

	Svreibuchneri FUESENTA_200906_3_YLK373	FUESENTA	43,346	-5,483	Species_9	JQ921640
	Svreibuchneri JIVERO2_200906_1FID_YLK713	JIVERO2	43,304	-3,572	Species_15	JQ921641
	Svreibuchneri SOURVALL_200906_1_YLK211	SOURVALL	43,300	-3,420	Species_15	JQ921642
	Svreibuchneri SOURVALL_200906_2FID_YLK673	SOURVALL	43,300	-3,420	Species_15	JQ921643
	Svreibuchneri BURGUI_201110_1ID_YAC213	BURGUI	42,704	-1,016	Species_16	KY624189
	Svreibuchneri BECERRAL_2011110_1ID_YAC215	BECERRAL	43,193	-3,585	Species_25	KY624188
	SspChico CHICO_201110_1_YAC211	CHICO	43,191	-3,581	Species_25	KY624183
	Svreibuchneri CERECEDA_200906_1ID_YLK252	CERECEDA	42,800	-3,494	Species_26	JQ921635
	Svreibuchneri CERECEDA_200906_2FID_YLK576	CERECEDA	42,800	-3,494	Species_26	JQ921636
	Svreibuchneri CERECEDA_200906_3_YLK580	CERECEDA	42,800	-3,494	Species_26	JQ921637
<i>Stenasellus virei buchneri</i> Stammer, 1936	Svreibuchneri RIBERO_200906_1FID_YLK418	RIBERO	43,038	-3,480	Species_26	KY624190
	Svreibuchneri TOJOS_201010_1ID_YCD911	TOJOS	43,159	-4,252	Species_30	JQ921644
	Svreibuchneri TOJOS_201010_2_YCD860	TOJOS	43,159	-4,252	Species_30	JQ921645
	Svreibuchneri TOJOS_201010_3_YCD929	TOJOS	43,159	-4,252	Species_30	JQ921646
<i>Stenasellus virei hussoni</i> Magniez, 1968	Svireihussoni BERNATAS_201011_1FIDex_YAB781	BERNATAS	42,982	0,899	Species_20	KY624191
	Svireihussoni COUMES_200802_1_YLK54	COUMES	43,017	1,153	Species_20	JQ921647
	Svireihussoni COUMES_200802_2_YLK68	COUMES	43,017	1,153	Species_20	JQ921648
	Svireihussoni PUJOS_196403_1_YLK19	PUJOS	43,034	0,822	Species_20	JQ921649
<i>Stenasellus virei margalefi</i> Magniez, 1996	Svireimargalefi MIJARES_200910_1ID_YCD325	MIJARES	40,030	-0,350	Species_13	JQ921650
<i>Stenasellus virei nobrei</i> Braga, 1942	Svireinobrei ALGAS_200910_1ID_YCD351	ALGAS	40,909	0,249	Species_23	JQ921651
	Svireinobrei BERGANT_200910_1ID_YCD326	BERGANT	40,669	-0,193	Species_23	JQ921652
	Svireinobrei MONZON_200906_1ID_YLK115	MONZON	41,916	0,176	Species_23	JQ921653
	Svireinobrei MONZON_200906_2_YLK416	MONZON	41,916	0,176	Species_23	JQ921654
	Svireinobrei MONZON_200906_3_YLK497	MONZON	41,916	0,176	Species_23	JQ921655
	Svireinobrei ALBIA_200807_1_YLK56	ALBIA	42,982	-3,100	Species_24	JQ921617
	Svireinobrei TRILLO_200910_1ID_YCD364	TRILLO	40,691	-2,583	Species_29	JQ921656
<i>Stenasellus virei rouchi</i> Magniez, 1996	Svireirouchi ALTURA_200910_1ID_YCD320	ALTURA	43,839	-5,509	Species_12	JQ921660
	Svireirouchi ALBENTOS_200910_1ID_YCD328	ALBENTOS	40,122	-0,764	Species_13	JQ921659
	Svireirouchi PEQUENO_200910_1FID_YCD327	PEQUENO	40,124	-0,354	Species_13	JQ921663
	Svireirouchi MATARRA_200910_1ID_YCD352	MATARRA	40,930	0,104	Species_23	JQ921662
	Svireirouchi EBRON_200910_1ID_YCD350	EBRON	40,145	-1,316	Species_28	JQ921661
<i>Stenasellus virei virei</i> Dollfus, 1897	Svireivirei COBALLE1_200906_1FID_YLK253	COBALLE1	43,186	-5,387	Species_8	JQ921664
	Svireivirei PERPETUA_200906_1FID_YLK715	PERPETUA	41,463	1,391	Species_17	KY624194
	Svireivirei CREUS_200906_1ID_YLK291	CREUS	41,349	1,362	Species_18	JQ921667
	Svireivirei CREUS_200906_2_YLK289	CREUS	41,349	1,362	Species_18	JQ921668
	Svireivirei BERNATAS_201110_1ID_YAC216	BERNATAS	42,982	0,899	Species_20	KY624192
	Svireivirei GOUELDI_201110_1ID_YAC212	GOUELDI	42,982	0,898	Species_20	KY624193
	Svireivirei PROMILHA_000000_1_YLK12	PROMILHA	43,005	2,919	Species_22	JQ921666
	Svireivirei RIBERO_200906_1FID_YLK153	RIBERO	43,038	-3,480	Species_26	JQ921665
	Svireivirei RIBERO_200906_2_YLK292	RIBERO	43,038	-3,480	Species_26	KY624195
	Svireivirei RIBERO_200906_3_YLK316	RIBERO	43,038	-3,480	Species_26	KY624196
	Svireivirei TRILLO_200910_1ID_YCD365	TRILLO	40,691	-2,583	Species_29	JQ921669
<i>Synasellus bragaianus</i> Henry & Magniez, 1987	SYbragaianus YUSOREIN_200906_1_YLK111	YUSOREIN	43,022	-4,866	Species_56	JQ921596
	SYbragaianus YUSOREIN_200906_2ex_YLK401	YUSOREIN	43,022	-4,866	Species_56	JQ921597
	SYbragaianus DEVALICE_200906_1_YLK247	DEVALICE	43,115	-4,769	Species_58	JQ921595
	SYbragaianus DEVANARG_200906_1ex_YLK183	DEVANARG	43,330	-4,568	Species_59	JQ921592
	SYbragaianus DEVAPANE_200906_1ID_YLK97	DEVAPANE	43,311	-4,610	Species_59	JQ921593
	SYbragaianus HERMIDA_200906_1_YLK244	HERMIDA	43,250	-4,605	Species_59	JQ921594
<i>Synasellus mejersae</i> Henry & Magniez, 1987	SYmejsae PENDONES_200906_1ID_YCD348	PENDONES	43,132	-5,247	Species_55	JQ921598
	SYmejsae PENDONES_200906_2ID_YCD349	PENDONES	43,132	-5,247	Species_55	JQ921599
	SYmejsae PENDONES_200906_3ID_YCD399	PENDONES	43,132	-5,247	Species_55	JQ921600
<i>Synasellus n. sp. 1 (ABOIM)</i>	SYnspAboim1 ABOIM_201504_11arnex_YAF106	ABOIM	41,917	-8,447	Species_63	NA
	SYnspAboim1 ABOIM_201504_5am_YAF057	ABOIM	41,917	-8,447	Species_63	NA
	SYnspAboim1 ABOIM_201010_1ID_YCD872	ABOIM	41,917	-8,447	Species_64	JQ921601
	SYnspAboim1 ABOIM_201010_4ID_YAB011	ABOIM	41,917	-8,447	Species_64	JQ921602
	SYnspAboim1 ABOIM_201504_13arnex_YAF109	ABOIM	41,917	-8,447	Species_64	NA
<i>Synasellus n. sp. 2 (ABOIM)</i>	SYnspAboim2 ABOIM_201010_3_YAB767	ABOIM	41,917	-8,447	Species_66	KY624197
	SYnspAboim2 ABOIM_201504_9arnex_YAF104	ABOIM	41,917	-8,447	Species_66	NA
	SYnspAboim2 ABOIM_201504_10arnex_YAF105	ABOIM	41,917	-8,447	Species_66	NA
	SYnspAboim2 ABOIM_201504_12arnex_YAF108	ABOIM	41,917	-8,447	Species_66	NA
<i>Synasellus n. sp. (BARRIO)</i>	SYnspBarrio BARRIO_201110_1ID_YAC220	BARRIO	41,846	-8,568	Species_65	KY624198
<i>Synasellus n. sp. (MEZIO)</i>	SYspMezio MEZIO_201110_1ex_YAC219	MEZIO	41,282	-8,308	Species_120	KY624202
<i>Synasellus n. sp. (TOJOS)</i>	SYnspTojos TOJOS_201010_1ID_YCD792	TOJOS	43,159	-4,252	Species_57	JQ921603
	SYnspTojos TOJOS_201010_2_YCD813	TOJOS	43,159	-4,252	Species_57	JQ921604
<i>Synasellus n. sp. (YUSO)</i>	SYnspYuso YUSOREIN_200906_3ex_YAB304	YUSOREIN	43,022	-4,866	Species_169	KY624203
<i>Synasellus n. sp. 1 (CONCHA)</i>	SYnspConcha1 CONCHA_201010_1ID_YAB776	CONCHA	43,560	-6,190	Species_171	KY624199
<i>Synasellus n. sp. 2 (CONCHA)</i>	SYnspConcha2 CONCHA_201010_2ID_YAB777	CONCHA	43,560	-6,190	Species_170	KY624200
<i>Synasellus n. sp. 3 (CONCHA)</i>	SYnspConcha3 CONCHA_201010_3_YAB393	CONCHA	43,560	-6,190	Species_172	KY624201

**Annexe 4 :**

**Liste des MOTUs et séquences utilisés pour construire  
la phylogénie calibrée dans le temps**

MOTU	COI	16S	28S	FAST2	MOTU	COI	16S	28S	FAST2
Species_1	NA	NA	NA	NA	Species_65	KY624198	NA	NA	NA
Species_4	KY624170	NA	NA	NA	Species_66	KY624197	NA	NA	NA
Species_5	JQ921629	JQ921839	KC610458	NA	Species_67	JQ921540	KC610272	JQ921866	NA
Species_7	JQ921611	NA	NA	NA	Species_68	JQ921524	KC610275	KC610421	NA
Species_8	JQ921664	JQ921852	KC610460	NA	Species_69	JQ921558	JQ921687	JQ921873	NA
Species_9	JQ921639	KC610331	KC610459	NA	Species_70	KY623739	JQ921688	JQ921874	NA
Species_10	JQ921627	JQ921838	JQ922012	NA	Species_71	JQ921562	JQ921685	KC610420	NA
Species_11	JQ921658	KC610329	KC610457	NA	Species_72	KY623737	NA	NA	NA
Species_12	JQ921660	JQ921850	JQ922018	NA	Species_74	JQ921529	JQ921673	NA	NA
Species_13	JQ921650	JQ921849	JQ922017	NA	Species_75	JQ921571	JQ921690	JQ921876	NA
Species_14	KY624184	NA	NA	NA	Species_76	JQ921530	JQ921674	JQ921864	NA
Species_15	JQ921642	KC610332	JQ922013	NA	Species_77	JQ921534	KC610271	JQ921865	NA
Species_16	KY624189	NA	NA	NA	Species_78	KY623734	NA	NA	NA
Species_17	KY624194	NA	NA	NA	Species_79	JQ921528	JQ921672	JQ921863	NA
Species_18	JQ921667	JQ921854	KC610461	NA	Species_80	JQ921535	NA	NA	NA
Species_20	JQ921648	KC610333	JQ922029	NA	Species_81	NA	NA	NA	NA
Species_22	JQ921666	JQ921853	JQ922020	NA	Species_82	NA	NA	NA	NA
Species_24	JQ921617	JQ921834	JQ922016	NA	Species_83	JQ921555	JQ921684	JQ921870	NA
Species_25	KY624183	NA	NA	NA	Species_84	JQ921549	KC610273	NA	NA
Species_26	JQ921635	KC610334	JQ922014	NA	Species_85	JQ921554	KC610274	NA	NA
Species_27	KY624185	NA	NA	NA	Species_86	JQ921525	JQ921671	JQ921862	NA
Species_28	JQ921661	JQ921851	JQ922019	NA	Species_87	JQ921544	JQ921679	NA	NA
Species_29	JQ921669	JQ921855	JQ922021	NA	Species_88	JQ921547	JQ921681	JQ921869	NA
Species_30	JQ921645	JQ921844	JQ922015	NA	Species_89	JQ921545	JQ921680	JQ921868	NA
Species_31	KY624178	NA	NA	NA	Species_90	JQ921575	AF259534	JQ921877	NA
Species_32	KY624180	NA	NA	NA	Species_91	NA	NA	NA	NA
Species_33	KY624181	NA	NA	NA	Species_92	NA	NA	NA	NA
Species_34	NA	NA	NA	NA	Species_93	NA	NA	NA	NA
Species_35	JQ921607	KC610327	NA	NA	Species_100	NA	NA	NA	NA
Species_36	JQ921618	KC610328	KC610455	NA	Species_101	NA	NA	NA	NA
Species_37	JQ921622	KM248976	JQ922011	NA	Species_112	NA	NA	NA	NA
Species_38	JQ921626	JQ921837	KC610456	NA	Species_113	NA	NA	NA	NA
Species_39	NA	NA	NA	NA	Species_114	NA	NA	NA	NA
Species_41	KY624176	NA	NA	NA	Species_115	NA	NA	NA	NA
Species_42	JQ921616	JQ921833	JQ922009	NA	Species_116	NA	NA	NA	NA
Species_43	KY624174	NA	NA	NA	Species_117	NA	NA	NA	NA
Species_53	NA	NA	NA	NA	Species_118	NA	NA	NA	NA
Species_55	JQ921598	KC610336	JQ922025	NA	Species_119	NA	NA	NA	NA
Species_56	JQ921596	JQ921858	JQ922024	NA	Species_120	KY624202	NA	NA	NA
Species_57	JQ921603	KC610337	KC610464	NA	Species_121	NA	NA	NA	NA
Species_58	JQ921595	JQ921857	JQ922022	NA	Species_122	KY624098	NA	NA	NA
Species_59	JQ921592	KC610335	KC610462	NA	Species_123	NA	NA	NA	NA
Species_60	JQ921537	JQ921676	JQ922027	NA	Species_125	KY623996	NA	NA	NA
Species_61	JQ921541	JQ921678	KC610419	NA	Species_126	JQ921248	JQ921766	JQ921948	NA
Species_62	JQ921568	JQ921689	JQ921875	NA	Species_128	KY623768	NA	NA	NA
Species_63	NA	NA	NA	NA	Species_129	KY623765	NA	NA	NA
Species_64	JQ921602	JQ921860	KC610463	NA	Species_130	NA	NA	JQ921882	NA
Species_131	NA	NA	NA	NA	Species_155	NA	NA	NA	NA
Species_132	NA	NA	NA	NA	Species_156	JQ921332	KC610206	KC610393	NA
Species_133	NA	NA	KC610425	NA	Species_157	NA	KC610193	KC610394	NA
Species_134	NA	KC610280	JQ921881	NA	Species_158	JQ921411	KC610212	KC610398	NA
Species_135	KY623755	NA	NA	NA	Species_159	JQ921386	JQ921806	JQ921986	NA
Species_136	NA	NA	NA	NA	Species_160	KY623733	NA	NA	NA
Species_137	NA	NA	NA	NA	Species_161	DQ144778	NA	DQ144737	NA
Species_138	NA	NA	NA	NA	Species_162	KY624088	NA	NA	NA
Species_140	NA	NA	DQ144747	NA	Species_163	JQ921366	JQ921800	KC610355	NA
Species_141	NA	NA	NA	NA	Species_164	JQ921132	JQ921736	KC610342	NA
Species_143	NA	NA	DQ144746	NA	Species_165	NA	NA	NA	NA
Species_146	NA	NA	NA	NA	Species_166	NA	NA	NA	NA
Species_147	NA	NA	NA	NA	Species_167	NA	NA	NA	NA
Species_148	NA	NA	NA	NA	Species_168	NA	NA	NA	NA
Species_149	NA	NA	NA	NA	Species_169	KY624203	NA	NA	NA
Species_151	NA	NA	NA	NA	Species_170	KY624200	NA	NA	NA
Species_152	JQ921065	JQ921718	KC610339	NA	Species_171	KY624199	NA	NA	NA
Species_153	NA	NA	NA	NA	Species_172	KY624201	NA	NA	NA
Species_154	NA	NA	NA	NA	Species_173	JQ921273	NA	KC610445	NA
Species_174	KY623851	KC610289	KC610431	NA	Species_207	JQ921373	JQ921803	JQ921983	NA
Species_175	NA	NA	NA	NA	Species_208	NA	NA	NA	NA
Species_176	NA	NA	NA	NA	Species_209	JQ921376	JQ921804	JQ921984	NA
Species_177	KY624059	NA	NA	NA	Species_210	NA	NA	NA	NA
Species_178	NA	NA	NA	NA	Species_211	JQ921220	NA	KC610349	NA
Species_179	KY624055	NA	NA	NA	Species_212	JQ921125	KC610294	JQ921915	NA
Species_180	KY624057	NA	NA	NA	Species_213	JQ921126	JQ921734	JQ921916	NA
Species_181	JQ921192	JQ921750	JQ921932	NA	Species_214	JQ921262	KC610308	KC610444	NA
Species_182	JQ921284	KC610313	KC610448	NA	Species_215	JQ921258	KC610307	JQ921951	NA
Species_183	JQ921110	JQ921727	JQ921909	NA	Species_216	JQ921268	JQ921771	KC610443	NA
Species_184	KY623780	NA	NA	NA	Species_217	JQ921254	JQ921772	JQ921950	NA
Species_185	KY623896	NA	NA	NA	Species_218	JQ921114	KC610310	JQ921911	NA
Species_186	JQ921194	JQ921751	JQ921933	NA	Species_219	JQ921266	KC610309	NA	NA
Species_187	KY623968	KC610100	KC610347	NA	Species_220	JQ921035	KC610287	JQ921888	NA
Species_188	KY624069	KC610322	KC610452	NA	Species_221	JQ921033	KC610288	JQ921887	NA
Species_189	JQ921189	JQ921749	JQ921931	NA	Species_222	JQ921488	JQ921822	JQ922001	NA
Species_190	JQ921052	JQ921712	JQ921895	NA	Species_223	KC610505	KC610323	KC610360	NA
Species_191	JQ921346	JQ921790	JQ921966	NA	Species_224	KY624082	JQ921811	NA	NA
Species_192	JQ921134	JQ921737	JQ921919	NA	Species_225	JQ921486	KC610324	KC610454	NA
Species_193	JQ921112	JQ921728	KC610346	NA	Species_226	JQ921489	JQ921823	KC610368	NA
Species_194	JQ921249	KC610306	JQ921949	NA	Species_227	JQ921038	JQ921706	KC610430	NA
Species_195	JQ921023	JQ921700	KC610442	NA	Species_228	JQ921037	JQ921705	JQ921889	NA
Species_196	JQ921021	KC610299	JQ921884	NA	Species_229	NA	NA	NA	NA
Species_197	NA	KC610298	KC610439	NA	Species_230	DQ305138	DQ305107	KC610451	NA
Species_198	KY624054	KC610321	KM249002	NA	Species_231	JQ921205	DQ305114	KC610358	NA
Species_199	KY623976	JQ921762	JQ921944	NA	Species_232	KY624035	NA	NA	NA
Species_200	KY623788	NA	NA	NA	Species_233	KY624039	NA	NA	NA
Species_201	NA	NA	NA	NA	Species_234	KY624036	NA	NA	NA
Species_202	JQ921028	JQ921702	JQ921886	NA	Species_235	JQ921340	JQ921786	JQ921968	NA
Species_203	KY624010	NA	NA	NA	Species_236	JQ921208	KC610101	JQ921940	NA
Species_204	KY624008	NA	NA	NA	Species_237	JQ921365	JQ921799	JQ921979	NA
Species_205	JQ921337	JQ921785	JQ921967	NA	Species_238	JQ921061	JQ921717	JQ921899	NA
Species_206	KY623854	JQ921710	JQ921893	NA	Species_239	JQ921245	JQ921765	JQ921947	NA
Species_240	JQ921280	JQ921776	JQ921958	NA	Species_273	JQ921197	KC610098	JQ921934	NA
Species_241	JQ921027	JQ921701	KC610447	NA	Species_274	KY623779	NA	NA	NA
Species_242	JQ921278	JQ921775	KC610446	NA	Species_275	KY624087	NA	NA	NA
Species_243	NA	KC610314	JQ921961	NA	Species_276	KY623796	NA	NA	NA
Species_244	JQ921316	JQ921781	JQ921962	NA	Species_277	JQ921184	JQ921746	KC610438	NA
Species_245	KY624015	NA	NA	NA	Species_278	NA	NA	NA	NA

Species_246	NA	NA	KC610353	NA	Species_279	NA	NA	NA	NA
Species_247	NA	NA	NA	NA	Species_280	NA	NA	NA	NA
Species_248	NA	NA	NA	NA	Species_281	NA	NA	NA	NA
Species_249	JQ921120	JQ921732	JQ921914	NA	Species_282	KC610501	KC610284	NA	NA
Species_250	JQ921121	NA	KC610341	NA	Species_283	NA	NA	NA	NA
Species_251	JQ921116	JQ921730	JQ921913	NA	Species_284	NA	NA	NA	NA
Species_252	JQ921364	JQ921798	JQ921978	NA	Species_285	NA	NA	NA	NA
Species_253	KY624040	NA	NA	NA	Species_286	NA	NA	NA	NA
Species_254	NA	NA	NA	NA	Species_287	KY623959	NA	NA	NA
Species_255	JQ921328	JQ921783	JQ921964	NA	Species_288	KY623932	NA	NA	NA
Species_256	JQ921067	JQ921719	KC610434	NA	Species_289	NA	NA	NA	NA
Species_257	JQ921207	KC610302	JQ921938	NA	Species_290	NA	NA	KM249014	NA
Species_258	KY624079	NA	NA	NA	Species_291	KY623900	NA	NA	NA
Species_259	NA	KC610296	JQ921930	NA	Species_292	NA	NA	NA	NA
Species_260	JQ921275	KC610312	JQ921956	NA	Species_293	JQ921141	KC610282	KC610426	NA
Species_261	JQ921055	KC610290	JQ921896	NA	Species_294	NA	NA	NA	NA
Species_262	JQ921059	JQ921715	NA	NA	Species_295	JQ921161	JQ921741	DQ144751	NA
Species_263	NA	KC610292	KC610432	NA	Species_296	NA	NA	NA	NA
Species_264	JQ921057	KM248975	JQ921897	NA	Species_297	KY623919	NA	NA	NA
Species_265	JQ921056	KC610291	JQ921898	NA	Species_298	NA	NA	NA	NA
Species_266	JQ921129	JQ921735	JQ921917	NA	Species_299	KY623928	NA	NA	NA
Species_267	JQ921327	NA	KC610354	NA	Species_300	KY623937	NA	NA	NA
Species_268	NA	NA	NA	NA	Species_301	KY623917	NA	NA	NA
Species_269	NA	NA	NA	NA	Species_302	NA	NA	NA	NA
Species_270	KY623993	NA	NA	NA	Species_303	NA	NA	NA	NA
Species_271	NA	NA	NA	NA	Species_304	NA	NA	NA	NA
Species_272	JQ921239	JQ921764	KC610441	NA	Species_305	NA	NA	NA	NA
Species_306	NA	NA	NA	NA	Species_339	NA	NA	NA	NA
Species_307	KY624030	NA	NA	NA	Species_340	JQ921579	KC610277	KC610422	NA
Species_308	NA	NA	NA	NA	Species_341	NA	NA	NA	NA
Species_309	JQ921050	JQ921711	JQ921894	NA	Species_342	JQ921344	KC610316	JQ921969	NA
Species_310	KY624080	NA	NA	NA	Species_343	NA	JQ921788	NA	NA
Species_311	KY623842	NA	NA	NA	Species_344	KY624095	NA	JQ921975	NA
Species_312	KY623813	NA	NA	NA	Species_345	KY624102	NA	NA	NA
Species_313	JQ921370	JQ921801	JQ921981	NA	Species_346	NA	NA	NA	NA
Species_314	NA	NA	NA	NA	Species_347	KY624100	NA	NA	NA
Species_315	KY623816	NA	NA	NA	Species_348	JQ921018	KC610295	JQ921925	NA
Species_316	KY623823	NA	NA	NA	Species_349	JQ921576	KC610278	KC610423	NA
Species_317	KY623826	NA	NA	NA	Species_350	JQ921516	JQ921829	JQ922006	NA
Species_318	KY623820	NA	NA	NA	Species_351	NA	NA	JQ922004	NA
Species_319	KY623799	NA	NA	NA	Species_352	NA	NA	NA	NA
Species_320	KY623808	NA	NA	NA	Species_353	NA	KC610258	KC610415	NA
Species_321	KY623838	NA	NA	NA	Species_354	NA	NA	NA	NA
Species_322	KY623839	NA	NA	NA	Species_355	KY624136	NA	NA	NA
Species_323	JQ921214	KC610303	JQ921941	NA	Species_356	NA	NA	NA	NA
Species_324	KY623812	NA	NA	NA	Species_357	NA	NA	NA	NA
Species_325	NA	NA	NA	NA	Species_358	NA	NA	NA	NA
Species_326	JQ921215	JQ921760	JQ921942	NA	Species_359	NA	NA	NA	NA
Species_327	KY623845	NA	NA	NA	Species_360	NA	NA	NA	NA
Species_328	KY623803	NA	NA	NA	Species_361	NA	NA	NA	NA
Species_329	NA	KC610285	KC610428	NA	Species_362	KY624156	NA	NA	NA
Species_330	KY623837	NA	NA	NA	Species_363	JQ921359	KC610319	JQ921976	NA
Species_331	KY623821	NA	NA	NA	Species_364	KY624104	KC610320	JQ921977	NA
Species_332	KY623828	NA	NA	NA	Species_365	JQ921432	KC610226	JQ921995	NA
Species_333	KY623832	NA	NA	NA	Species_366	NA	NA	KC610400	NA
Species_334	KY623822	NA	NA	NA	Species_367	NA	NA	KC610453	NA
Species_335	KC610503	KC610286	KC610429	NA	Species_368	NA	JQ921724	KC610340	NA
Species_336	NA	NA	NA	NA	Species_369	NA	NA	KC610370	NA
Species_337	NA	NA	NA	NA	Species_370	JQ921076	NA	NA	NA
Species_338	JQ921413	JQ921810	JQ921990	NA	Species_371	JQ921072	JQ921721	NA	NA
Species_372	JQ921077	JQ921723	JQ921905	NA	Species_405	KY624133	NA	NA	NA
Species_373	NA	KC610293	NA	NA	Species_406	JQ921493	KC610325	JQ922003	NA
Species_374	NA	NA	NA	NA	Species_407	JQ921336	JQ921789	JQ921970	NA
Species_375	KY624090	JQ921791	JQ921971	NA	Species_408	KY624073	NA	NA	NA
Species_376	NA	NA	NA	NA	Species_395	KC610491	KC610236	NA	NA
Species_377	JQ921204	KC610300	JQ921936	NA	Species_396	JQ921445	KC610240	JQ921996	NA
Species_378	JQ921200	KC610301	KC610440	NA	Species_397	KY624123	NA	NA	NA
Species_379	JQ921355	KC610318	JQ921974	NA	Species_398	NA	NA	NA	NA
Species_380	JQ921350	JQ921792	JQ921972	NA	Species_399	NA	JQ921745	KC610437	NA
Species_381	KY624093	NA	NA	NA	Species_400	KY623987	NA	NA	NA
Species_382	KY624077	NA	NA	NA	Species_401	JQ921235	JQ921763	JQ921945	NA
Species_383	JQ921353	KC610317	JQ921973	NA	Species_402	NA	NA	NA	NA
Species_384	KY624091	NA	NA	NA	Species_403	KY623980	NA	NA	NA
Species_385	NA	NA	NA	NA	Species_404	JQ921175	JQ921744	NA	NA
Species_386	KY624089	NA	NA	NA					
Species_387	KY624075	NA	NA	NA					
Species_388	JQ921383	JQ921805	JQ921985	NA					
Species_389	KY624113	NA	NA	NA					
Species_390	KY624114	NA	NA	NA					
Species_391	NA	NA	NA	NA					
Species_392	JQ921443	NA	KC610410	NA					
Species_393	KY624117	NA	NA	NA					
Species_394	KC610489	KC610234	NA	NA					





## Annexe 5 :

### Tableau récapitulatif du séquençage de transcriptomes

Ce tableau récapitule les espèces pour lesquelles des transcriptomes ont été séquencés. Les transcriptomes séquencés avant 2013 ont été fait avec des pool d'individus dont les numéros sont indiqués. Après 2013, chaque transcriptomes correspond à un individu. Les dates de chaque étapes ont été notés afin de pouvoir détecter des étapes de potentielle cross-contamination qui ont été éliminées des jeux de données.

Species	Code	Assembled individuals	Sampling_date	RNA_extraction	PoolARN	Add_oneindividual_in_pool	Library	Sequencing_run
Paragonensis	PArE	PE:(2, 4, 10, 13, 15)	201102	15/06/2011	07/11/2011	NA	03/09/2012	05/2012
Pjaloniacus	PJJ	PE:(23, 24, 26, 27, 29, 30, 32)	201102	06/09/2011	07/12/2011	NA	03/08/2012	05/2012
Pbeticus	PBS	PE:(3, 5, 8, 9, 10, 11, 25)	201102	15/03/2011; 01/04/2011	13/04/2011	NA	2011	09/2011
Pspelaeus	PSSE	PE:(11, 12, 13, 14, 15, 17, 18 (pool 20ndiv))	201102	21/06/2011; 23/06/2011	07/12/2011	NA	03/08/2012	05/2012
Phercegovinensis	PHeZ	PE:(4, 5, 6, 7, 8, 9)	201107	13/01/2012	14/02/2012	NA	26/11/2012	02/2013
Pkaramani	PKK	PE:(5, 6, 7, 8, 10, 9)	201107	19/01/2012	13/02/2012	18/07/2012	26/11/2012	02/2013
Pslovenicus	PSloP5	PE:5	201107	22/09/2015	NA	NA	17/10/2016	S16273
Pparvulus	PPrS	PE:(3, 10 (pool de 6indiv))	201107	13/01/2012; 20/02/2012	29/10/2012	NA	26/11/2012	02/2013
Pgranadensis	PGrP	PE: CROH2	201110	13/11/2012	NA	NA	26/10/2015	S15223
Pmeridianus	PMLL	PE:(4, 5, 6, 7, 8, 9)	201110	19/01/2012	13/02/2012	11/08/2012	26/11/2012	02/2013
Ppeltaus	BPM	PE:(3, 4, 5, 6, 7, 9)	201110	01/05/2012	13/02/2012	17/07/2012	10/01/2012	10/2012
Pibeticus	PibB	PE:(1, 2, 3, 4, 6, 7)	201110	05/01/2012; 19/01/2012	14/02/2012	NA	13/09/2012	10/2012
Psolanasi	PSoG	PE:(1, 2, 3, 4, 5, 7)	201110	06/12/2011; 10/07/2012	13/02/2012	11/08/2012	26/11/2012	02/2013
Pmagalefi	PMaF	PE:(1, 3, 6, 7 (pool 6indiv))	201110	06/12/2011; 20/02/2012	11/08/2012	NA	26/11/2012	02/2013
Pproxalis	PCoB	PE:(1, 2, 3, 4, 5, 6)	201110	12/07/2011	14/02/2012	NA	26/11/2012	02/2013
Pcoiffaiti	PCoiP	PE:(1, 2, 3, 4, 5, 6)	201110	12/12/2011; 05/01/2012	13/02/2012	17/07/2012	26/11/2012	02/2013
Parthrodius	PAL	PE:(1, 2, 4, 5, 7, 6)	201110	12/12/2011; 19/01/2012	14/02/2012	18/07/2012	13/09/2012	10/2012
Bmolinae	BMoV	PE:(4, 5, 6, 9, 14, 16 (pool 8indiv))	201110	13/01/2012; 19/01/2012; 20/02/2012	28/02/2012	NA	13/09/2012	10/2012
Pracovitzai	PRBe	PE:(5, 6, 7, 8, 9, 10)	201110	19/12/2011; 19/01/2012	13/02/2012	29/10/2012	26/11/2012	02/2013
Passaforensis	PAsA	PE:(4, 5, 6, 7, 8, 10)	201110	19/12/2011; 24/02/2012	13/02/2012	11/08/2012	26/11/2012	02/2013
Prectus	PRuE	PE:(4, 5, 6, 7, 9, 8)	201110	19/12/2011; 19/01/2012	14/02/2012	11/08/2012	26/11/2012	02/2013
Pcavaticus	PCBF	PE:(10, 11, 12, 13, 15, 16)	201111	11/02/2012	11/08/2012	NA	26/11/2012	02/2013
Pcantabricus	PCanC	PE:14	201405	06/10/2014	NA	NA	07/01/2014	S15148
Pontizi	POJ	PE: 32	201405	06/10/2014	NA	NA	07/01/2014	S14095
Pistriamus	PIS	PE:4	201405	05/11/2015	NA	NA	18/06/2015	S15148
Pgrafi	PGC	PE: 7	201405	06/11/2014	NA	NA	07/01/2014	S14095
Pebrensis	PEF	PE: 9	201405	16/06/2014	NA	NA	07/07/2014	S14095
Pescolai	PEcD	PE:25	201405	06/03/2014	NA	NA	26/10/2015	S15223
Pmeridianus	PMB	PE:10	201406	18/06/2014	NA	NA	07/01/2014	S15148
Pwalteri	PWM5	PE:5	201409	18/09/2014	NA	NA	11/01/2014	S14184
Pwalteri	PWRo5	PE:5	201409	18/09/2014	NA	NA	11/01/2014	S14184
Pboui	PBS4	PE:4	201409	24/10/2014	NA	NA	11/01/2014	S14184
Pcavaticus	PCMI5	PE:5	201409	28/09/2016	NA	NA	10/11/2016	S16240
Pslavus	PSIRo11	PE:11	201409	29/09/2014	NA	NA	11/01/2014	S14184
Pwalteri	PWH13	PE:13	201409	31/10/2014	NA	NA	11/01/2014	S14184
Pwalteri	PWV16	PE:6	201409	31/10/2014	NA	NA	11/01/2014	S14184
Pboui	PBF5	PE:5	201409	10/07/2014	NA	NA	11/01/2014	S14184
Palbigensis	PAL6	PE:6	201409	10/08/2014	NA	NA	11/01/2014	S14184
Pnsparpaon	PspAG1	PE:1	201409	10/08/2014	NA	NA	11/01/2014	S14184

Bcortesi	BCoP	PE:16	201504	05/11/2015	NA	NA	06/08/2015	S15148
Pmeridianus	PMAX	PE:3	201504	05/11/2015	NA	NA	06/08/2015	S15148
Bfrontellum	BFA	PE:5	201504	15/06/2015	NA	NA	26/06/2015	S15148
Synsp	SypA	PE:14	201504	16/10/2015	NA	NA	26/10/2015	S15223
Aaquaticus	AAP1	PE:3	201504	22/05/2015	NA	NA	17/06/2015	S15148
Pmicropectinatus	PMiM	PE:6	201504	22/05/2015	NA	NA	26/06/2015	S15148
Sbreuili	SBrI	PE:3	201504	22/05/2015	NA	NA	26/06/2015	S15148
Pmeridianus	PMC	PE:10	201504	22/06/2015	NA	NA	25/06/2015	S15148
Pnsphoreon	ISPB	PE:7	201509	23/09/2015	NA	NA	26/10/2015	S15223
Pstrouhali	PSIB	PE:5	201509	23/09/2015	NA	NA	26/10/2015	S15223
Pcavaticus	PCBo	PE:4	201509	24/09/2015	NA	NA	26/10/2015	S15223
Pnsphoreon	PspV1	PE:7	201509	24/09/2015	NA	NA	26/10/2015	S15223
Pstrouhali	PSiN	PE:6	201509	10/01/2015	NA	NA	26/10/2015	S15223
Gheyli	GHM2	PE:2	201603	22/03/2016	NA	NA	10/05/2016	S16240
Pcavaticus	PCG6	PE:6	201604	09/12/2016	NA	NA	26/09/2016	S16157
Jpraehirsuta	JPL1	PE:1	201606	16/09/2015	NA	NA	10/11/2016	S16240
Akosswigi	AKT6	PE:6	201608	16/09/2015	NA	NA	19/10/2016	S16157
Bskopljensis	BSSV1	PE:1	201608	16/09/2015	NA	NA	10/03/2016	S16157
Ppavani	PPvN2	PE:2	201608	16/09/2015	NA	NA	10/11/2016	S16240
Aaquaticus	AAD3	PE:3	201610	10/10/2016	NA	NA	17/10/2016	S16273
Pcoxalis	PCoAx5	PE:5	201610	27/10/2016	NA	NA	12/05/2016	S16273
Pcoxalis	PCoZa5	PE:5	201610	27/10/2016	NA	NA	12/05/2016	S16273
Pcoxalis	PCoCd9	PE:9	201610	30/11/2016	NA	NA	12/05/2016	S16273
Pburgundus	PBuB2	PE:2	201707	23/08/2017	NA	NA	10/05/2017	S17275
Phermallensis	PHL5	PE:5	201707	10/02/2017	NA	NA	10/05/2017	S17275
Blagari	BLaA4	PE:4	201709	10/02/2017	NA	NA	10/05/2017	S17275
Pleschetae	PLeC5	PE:5	201709	10/02/2017	NA	NA	10/05/2017	S17275
Pleschetae	PLeM4	PE:4	201709	10/02/2017	NA	NA	10/05/2017	S17275



**Annexe 6 :**

**Article : Life History Traits Impact the Nuclear  
Rate of Substitution but Not the Mitochondrial  
Rate in Isopods**

# Life History Traits Impact the Nuclear Rate of Substitution but Not the Mitochondrial Rate in Isopods

Nathanaëlle Saclier,<sup>\*1</sup> Clémentine M. François,<sup>1</sup> Lara Konecny-Dupré,<sup>1</sup> Nicolas Lartillot,<sup>2</sup> Laurent Guéguen,<sup>2</sup> Laurent Duret,<sup>2</sup> Florian Malard,<sup>1</sup> Christophe J. Douady,<sup>1,3</sup> and Tristan Lefébure<sup>\*,1</sup>

<sup>1</sup>Univ Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5023, ENTPE, Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés, Villeurbanne, France

<sup>2</sup>Univ Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, Villeurbanne, France

<sup>3</sup>Institut Universitaire de France, Paris, France

**\*Corresponding authors:** E-mails: nathanaelle.saclier@univ-lyon1.fr; tristan.lefebure@univ-lyon1.fr.

**Associate editor:** Keith Crandall

## Abstract

The rate of molecular evolution varies widely among species. Life history traits (LHTs) have been proposed as a major driver of these variations. However, the relative contribution of each trait is poorly understood. Here, we test the influence of metabolic rate (MR), longevity, and generation time (GT) on the nuclear and mitochondrial synonymous substitution rates using a group of isopod species that have made multiple independent transitions to subterranean environments. Subterranean species have repeatedly evolved a lower MR, a longer lifespan and a longer GT. We assembled the nuclear transcriptomes and the mitochondrial genomes of 13 pairs of closely related isopods, each pair composed of one surface and one subterranean species. We found that subterranean species have a lower rate of nuclear synonymous substitution than surface species whereas the mitochondrial rate remained unchanged. We propose that this decoupling between nuclear and mitochondrial rates comes from different DNA replication processes in these two compartments. In isopods, the nuclear rate is probably tightly controlled by GT alone. In contrast, mitochondrial genomes appear to replicate and mutate at a rate independent of LHTs. These results are incongruent with previous studies, which were mostly devoted to vertebrates. We suggest that this incongruence can be explained by developmental differences between animal clades, with a quiescent period during female gametogenesis in mammals and birds which imposes a nuclear and mitochondrial rate coupling, as opposed to the continuous gametogenesis observed in most arthropods.

**Key words:** substitution rate, longevity, generation time, isopods.

## Introduction

The rate at which DNA accumulates substitutions varies widely among lineages (Lynch 2010; Bromham et al. 2015; Allio et al. 2017). Correlations have been observed between many life history traits (LHTs) and the rate of substitution, to the point that one could use ancestral rate reconstruction based on phylogenetic methods to infer ancestral LHTs (Lartillot and Delsuc 2012; Wu et al. 2017). Nevertheless, the mechanism at the origin of these correlations and the question of their universality remain unclear (Caccone et al. 2004; Baer et al. 2007; Allio et al. 2017).

Two broad categories of hypotheses have been proposed to explain the link between LHT and the rate of molecular evolution: hypotheses where the substitution rate depends on the DNA replication rate and hypotheses where it is controlled by the activity of the mitochondrial respiratory chain. In the first category, a well-established hypothesis is the generation time (GT) hypothesis (Li et al. 1987) which proposes that a species with shorter GT copies its genome more often per time unit, thereby accumulating more DNA replication errors. Correlations between GT and substitution rate have

been observed in mammals (Ohta 1993; Bromham et al. 1996), birds (Mooers and Harvey 1994), protostomians (Thomas et al. 2010), and plants (Smith and Donoghue 2008). Hypotheses belonging to the second category rely on the assumption that the mitochondrial respiratory chain generates mutagenic reactive oxygen species (ROS). As a consequence species that generate more ROS are expected to accumulate more DNA damage, and in particular damage to the mitochondrial genome (Barja 2002). Two influences—one direct and one indirect—of mitochondrial respiratory chain activity on the substitution rate have then been proposed. First, species with high metabolic rate (MR) produce more ROS, accumulate more DNA damage, and so should have a higher substitution rate. Indeed, some studies have reported a correlation between MR and substitution rate (Martin et al. 1992; Gillooly et al. 2005). Second, Nabholz et al. (2008) proposed that mitochondrial mutations will be differentially counter-selected depending on the species longevity. The mitochondrial theory of aging proposes that mutations in the mitochondrial genome lead to a chain reaction by affecting the efficiency of the respiratory chain, thus leading to an increase in the production of ROS and a

correlative acceleration of aging (Kujoth et al. 2007). To counteract this effect and ultimately delay aging, selection is expected to act in long-lived species to reduce the mitochondrial mutation rate. One expected outcome of this indirect effect is a negative correlation between substitution rate and longevity. These two last hypotheses imply a direct causal link between the MR, the production of ROS and ultimately the mutation rate. Several studies have challenged the link between MR and ROS production (Barja 2007) but also the ROS influence on the mitochondrial mutation rate (Larsson 2010; Kennedy et al. 2013). While the former category of hypotheses (the DNA replication rate hypotheses) extends to all organelles and their genomes, the second set (the mitochondrial respiratory chain hypotheses) mostly applies to the mitochondrial genome (Galtier, Jobson, et al. 2009), as ROS are produced in the mitochondria where they directly induce damage to the mitochondrial genome. Using experimental mutagenesis on rat's mitochondria, Richter et al. (1988) showed that mutations induced by ROS are 16 times more frequent in the mitochondrial genome than the nuclear genome.

While these questions are not new, we still lack a clear understanding of why substitution rates are not clock-like across taxa and which traits, if any, are responsible for the violation of a strict molecular clock. Trait covariation, a focus on a very restricted set of taxa, and a reduced number of studied loci, mostly mitochondrial, are but a few roadblocks preventing a better understanding of these questions, as we discuss below. First, disentangling the influence of different biological traits on the substitution rate is complicated by rampant trait covariation. In particular, GT, MR, and longevity are often strongly correlated along the slow-fast life history continuum (Jeschke and Kokko 2009). Second, most studies have been devoted to mammals and birds, two endothermic groups of amniotes, which calls into question the universality of the observed patterns and proposed hypotheses. Third, most studies are also based on few loci (most between one to six genes, with a few reaching 15 genes). Within a genome, substitution rates vary extensively due to variation in mutation rate (Wolfe et al. 1989; Ellegren et al. 2003) or selection pressure (Duret and Mouchiroud 2000) along the chromosomes. Estimates of rate variation could be further blurred by this intragenomic variation, in particular when a small number of loci are considered. Finally, most studies are based on mitochondrial genes (though there are exceptions, see Caccone et al. 2004; Nikolaev et al. 2007; Welch et al. 2008; Lartillot and Delsuc 2012). The mitochondrial genome has a very distinct evolutionary dynamic compared with the nuclear genome. Because it is maternally inherited, haploid, and has specific mutational exposure, repair mechanisms and replication machineries, the mitochondrial genome evolves under a very different effective population size, and at different mutation and recombination rates. As such, the usage of the mitochondrial genome as a marker to infer population histories and divergences is debated (Galtier, Nabholz, et al. 2009; Allio et al. 2017). At the very least, substitution rate variation observed in the context of the mitochondrial genome should not be generalized to the nuclear genome. To sum up, the

respective influence of LHT on the nuclear and mitochondrial substitution rates requires further investigation.

In this study, we used the transition to the subterranean habitat to disentangle the influence of the replication rate and the activity of the mitochondrial respiratory chain on the mitochondrial and nuclear substitution rates. We used the unique features of the aquatic isopods of the Asellidae family: within this family, many species have independently become subterranean, offering an opportunity to study independent replicates of the same evolutionary transition from surface to groundwater habitats. This transition triggered an important modification of the LHT of the underground species, including an extension of their lifetime (5–10 times), a delay of sexual maturity and a decrease in the number of offspring (Henry 1976). As a response to food scarcity, subterranean organisms also tend to display low MRs (Hüppop 1987; Renault et al. 2002). Thanks to recent phylogenetic (Morvan et al. 2013) and genomic (Francois et al. 2016; Lefébure et al. 2017) studies, the phylogeny of Asellidae is now well-resolved and many nuclear genes have been sequenced for 13 independent surface/subterranean species pairs. We complemented these nuclear data sets with mitochondrial genome assemblies and tested the predictions of the DNA replication rate and mitochondrial respiratory chain hypotheses. Namely, if the activity of the mitochondrial respiratory chain is the main driver of change to the substitution rate, we expect a lower mitochondrial rate in the long-lived and low-metabolism subterranean species with little to no variation on the nuclear rate. Conversely, if the DNA replication rate is a major factor, we this time expect a decrease of the substitution rate in subterranean species that is of the same amplitude in the mitochondrial and nuclear genomes.

## Results

### Nuclear and Mitochondrial Data Set

For 13 pairs of surface and subterranean species, 386 one-to-one nuclear orthologous gene family alignments were taken from a previous study (Francois et al. 2016). Four gene families displaying evidence of contamination were discarded. A phylogeny of the 26 species was built with PhyML v3.0 (Guindon et al. 2010) by using the 382 nuclear genes. Except for two nodes with values at 97.4% and 98%, all nodes have a bootstrap value of 100% (supplementary fig. S1, Supplementary Material online). Nearly complete mitochondrial genomes for these 13 species pairs were de novo assembled using a combination of genomic and RNA-seq reads. Only the control region consistently failed to be assembled. The de novo mitochondrial genomes are syntenous and display the same gene order as the one already described in Asellidae and in the more distantly related isopod Oniscidea (Kilpert and Podsiadlowski 2006; Kilpert et al. 2012). Difficulty in assembling the mitochondrial control region in isopods is a well-known problem that has been interpreted as the result of an atypical mtDNA composed of linear monomers associated with circular “head-to-head” dimers (Raimond et al. 1999; Doublet et al. 2012). Twelve of the 13 mitochondrial protein coding genes (*atp8* was too small to be identified with



certainty) were identified for each species and their DNA sequences were concatenated and used to estimate a global mitochondrial substitution rate.

### Effect of the Transition to Subterranean Habitat on Substitution Rates

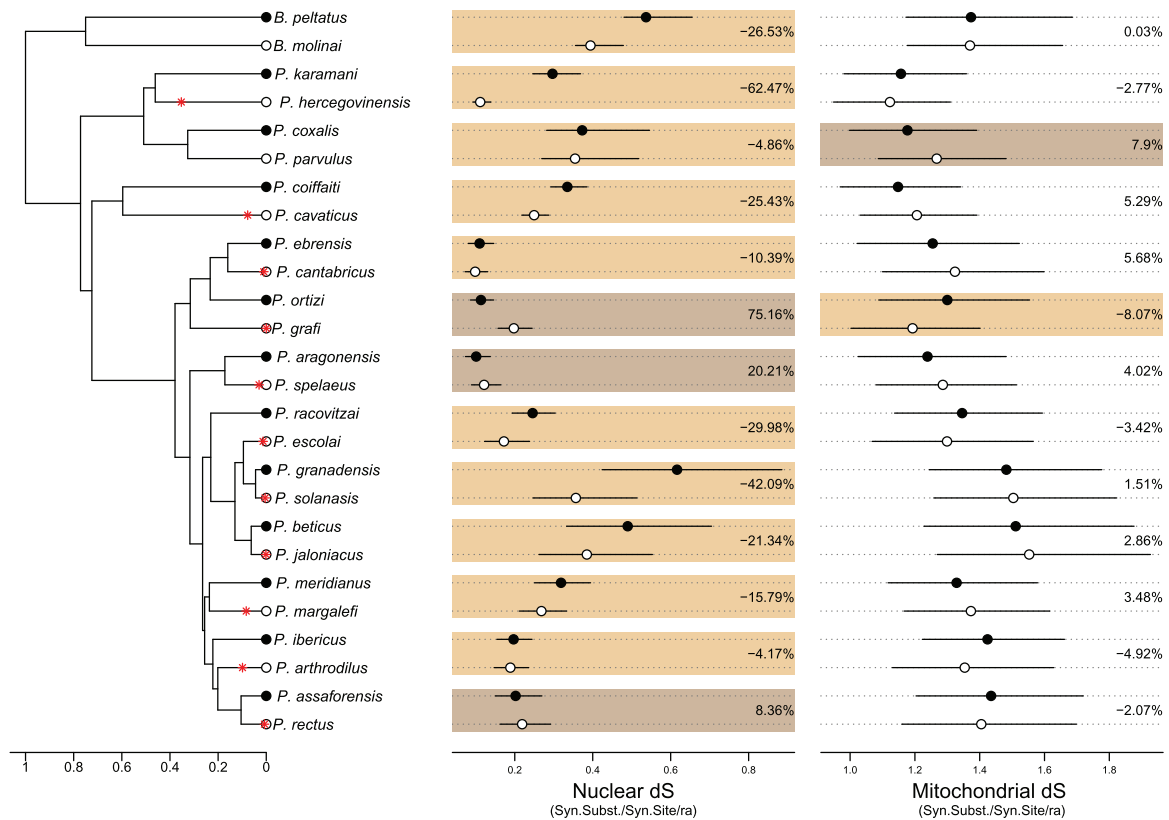
For each of the 26 species, we computed a rate of synonymous substitution ( $d_s$ ) per unit of time using the Bayesian software program CoEvol (Lartillot and Poujol 2012) independently for the nuclear (382 genes) and for the mitochondrial concatenations (12 genes, [supplementary table S1, Supplementary Material](#) online). If we ignore selection for codon usage, a transcriptome-wide  $d_s$  is expected to give a good estimate of the mutation rate. Correlations between the rate of molecular evolution and habitat were tested using PGLS (Phylogenetic Least Square) models. Moreover, inside each pair of species, we computed the posterior probability of an increase of the  $d_s$  in surface species (later names pp) as the frequency of samples from the CoEvol Monte-Carlo chain where the surface species has a higher  $d_s$  than the subterranean species. In the nuclear compartment, the  $d_s$  was significantly lower in subterranean species (PGLS, [table 1, fig. 1](#)), with a significant subterranean  $d_s$  decrease in 10 pairs out of 13 ([fig. 1, pp > 0.90](#)). However, in the mitochondrial genome, we observed no effect of the ecological status on the  $d_s$  (PGLS, [table 1](#)) with only one pair supporting an increase in  $d_s$  in the subterranean species (pp = 1), another supporting a decrease (pp = 0), and the remaining 11 pairs showing no significant  $d_s$  variation ( $0.10 < \text{pp} < 0.90$ , [fig. 1](#)). To quantify the apparent decoupling between the mitochondrial and nuclear rates, we estimated the ratio of mitochondrial  $d_s$  to nuclear  $d_s$  ( $d_{s_{\text{mito}}}/d_{s_{\text{nuc}}}$ ). Mitochondrial genes undergo a synonymous substitution rate 15.71 times faster than the nuclear genes but with some substantial variation across taxa (from  $2.5\times$  to  $34.9\times$ , [supplementary table S2, Supplementary Material](#) online). The variation in  $d_{s_{\text{mito}}}/d_{s_{\text{nuc}}}$  was mostly explained by the nuclear contribution, with a mitochondrial  $d_s$  much more stable than the nuclear  $d_s$  (coefficient of variation of 8.96% against 51.67% in the mitochondrial and nuclear genome, respectively).

### Impact of the Colonization Time

The comparative analysis presented in the last section, based on surface/subterranean independent species pair and using the present day ecological status in a regression model, implicitly amounts to assuming that the speciation event between the two species and the ecological transition are concomitant. However, in reality, speciation might have occurred long before the ecological transition. Some subterranean species have intermediate traits (microphthalmia, partial depigmentation) suggesting that they colonized the subterranean environment only recently. This timing difference between sets of pairs can reduce our power to detect a rate variation induced by the shift to a subterranean habitat. For any given species, our substitution rate estimate is an average across the terminal branch of the phylogeny leading to that species, that is, for a period of time that extends to the last observed speciation event. If a species became subterranean

recently, relative to this speciation event, and only then evolved a different rate of substitution, its estimated substitution rate on the whole branch will be marginally influenced by this late colonization event. Quantitatively, the magnitude of the observed rate variation will depend directly on the proportion of time along the terminal branch corresponding to underground life (hereafter called the relative colonization time, RCT). This RCT can range from 0 (surface species) to 1 (subterranean species which colonized groundwater when diverging from the paired surface species). If RCT is small, the observed rate variation will be small. We estimated RCT using the pseudogenization of an opsin locus, as described in [Lefébure et al. \(2017\)](#). Concomitantly with the transition to a subterranean habitat, there is a regression of the ocular system that is commonly associated with a loss of function of the opsin genes ([Niemi et al. 2012](#)). We estimated RCT for 23 out of 26 species for which we were able to assemble an opsin gene. Briefly, we assume a two-state model for the opsin gene: a surface and a subterranean state. Before the habitat transition the opsin ratio of nonsynonymous to synonymous substitution rates ( $d_N/d_S$ ) is very low and mainly influenced by purifying selection. After the habitat transition, the opsin gene evolves under a neutral process with a  $d_N/d_S$  ratio equal to one. The estimated opsin  $d_N/d_S$  of a subterranean species is therefore a weighted mean of these two  $d_N/d_S$ , with weights depending on the ratio between the colonization and the speciation time. From this, we can then estimate the proportion of the RCT and directly test its correlation with  $d_s$ .

The estimated values of RCT confirmed that our naive classification into surface and subterranean species hides a more complex scenario. Indeed, some species have colonized the subterranean habitat quickly after speciation whereas others colonized the subterranean habitat much later on (e.g., 2.2% and 76.5% of the terminal branch for *Proasellus jaloniacus* and *Proasellus hercegovinensis*, respectively, [fig. 1, supplementary table S1, Supplementary Material](#) online). In the nuclear genome, we recovered a negative correlation between  $d_s$  and RCT ( $r^2 = 0.43$ , [table 1, fig. 2](#)). We then tested if a model with the RCT explained a significantly greater part of the  $d_s$  variance than a model with only the binary description of the habitat. A model comparison with both explanatory variables to a simpler model with only the binary ecological status favored the former more complex model (LRT  $P$  value = 0.0049,  $L$  ratio = 7.921). On the other hand, there was no correlation between the mitochondrial  $d_s$  and RCT. The  $d_{s_{\text{mito}}}/d_{s_{\text{nuc}}}$  was significantly correlated to RCT ([table 1](#)). On the basis of RCT, species can be empirically split into three groups (see [supplementary table S1, Supplementary Material](#) online): 1) the “true surface” species with RCT < 1%—these are the six pigmented and ocululated species and, surprisingly, the subterranean species *Proasellus grafi*; 2) the “true subterranean” species for which RCT > 10%—these are the depigmented and anophthalmous species; and 3) the intermediate states ( $1\% < \text{RCT} < 10\%$ ) with partially depigmented and microphthalmous species. When only using the habitat status, these intermediate species blur the signal and reduce the estimated correlation because the contrast between the



**FIG. 1.** Substitution rate of the nuclear (center) and mitochondrial (right) genomes in 13 Aselloidea species pairs (left) each composed of one surface species (black circle) and one subterranean species (white circle). The chronogram (left) time scale is arbitrarily defined so that the age of the root is equal to 1. Whenever an opsin gene could be recovered in a subterranean species, a red star on the chronogram indicates the estimated colonization time using opsin gene nonfunctionalization. Rates are synonymous substitutions per synonymous site ( $d_s$ ) relative to the root age ( $ra$ ) with their 95% credibility interval as estimated by CoEvol. For each pair, percentage indicates the variation of the substitution rate in the subterranean species relative to the surface species. When this variation has a posterior probability of being different from 0 that is  $>0.9$ , the species pair is highlighted with a dark (rate increase) or light (rate decrease) colored box.

**Table 1.** Phylogenetic Least Square (PGLS) Regression of Nuclear Synonymous Substitution Rate ( $d_s$ ), Mitochondrial  $d_s$ , the  $d_s$  Ratio of the Two Compartments and the Number of Mitochondrial Synonymous Transversions (TvS) Against the Ecological Status (surface—subterranean), the Proportion of the Relative Colonization Time (RCT), and the Metabolic Rate (estimated by the RNA/protein ratio).

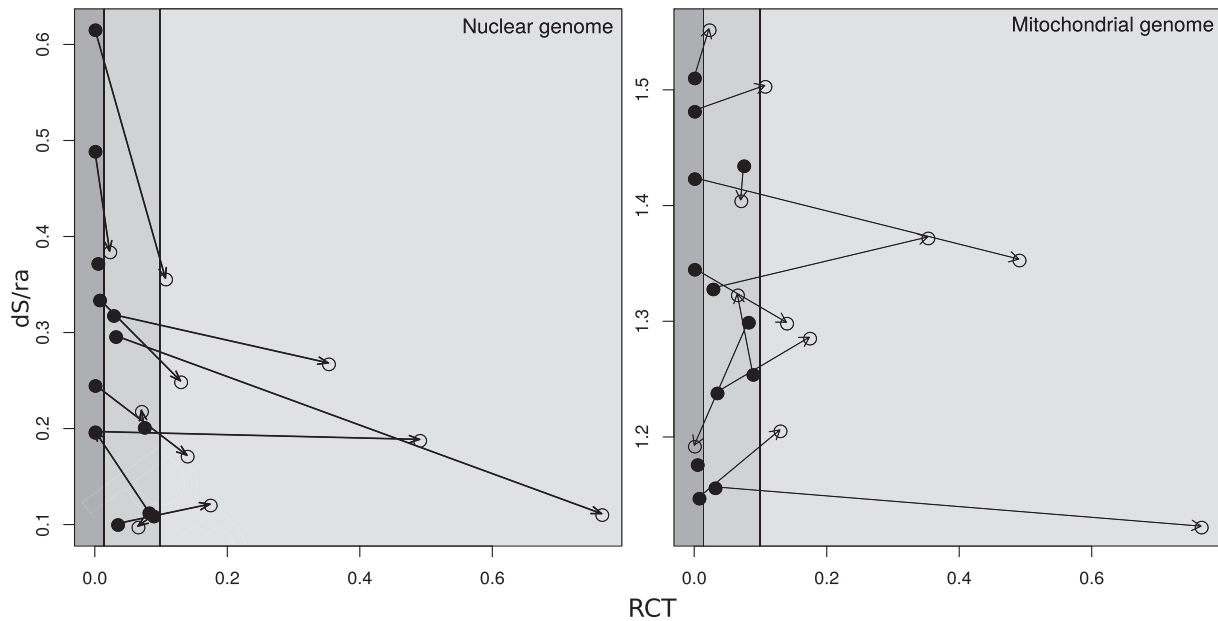
Dependent Variable	Explanatory Variable	Slope Sign	L ratio	P value	Pseudo $R^2$ (Coxsnell)
Nuclear $d_s/ra$	Ecological status	-	5.516	0.019	0.191
	RCT	-	13.066	3.10 <sup>-4</sup>	0.433
	RNA/protein ratio	+	11.413	7.10 <sup>-4</sup>	0.510
Mitochondrial $d_s/ra$	Ecological status	+	0.042	0.838	0.002
	RCT	-	0.337	0.562	0.015
	RNA/protein ratio	-	0.771	0.380	0.047
$d_{s_{mito}}/d_{s_{nucl}}$ ratio	Ecological status	+	2.733	0.098	0.100
	RCT	+	8.498	0.004	0.309
	RNA/protein ratio	-	9.258	0.002	0.439
Mitochondrial TvS	Ecological status	+	0.037	0.847	0.001
	RCT	-	0.023	0.881	0.001
	RNA/protein ratio	-	0.342	0.559	0.021

NOTE.—Each line corresponds to one likelihood ratio test between the models with and without the given explanatory variable.

surface and the subterranean species in these pairs is less strong and is diluted compared with the other pairs. Indeed, the three species pairs that show an inverse variation in the nuclear  $d_s$  ( $d_{s_{subterranean}} \geq d_{s_{surface}}$ ) each contain at least one species with this intermediate state (figs. 1 and 2).

### RNA/Protein Ratio

Under the hypotheses that the activity of the respiratory chain directly influences the mutation rate, the MR is expected to be positively correlated to the mitochondrial  $d_s$ . To test this hypothesis, we used the RNA/protein ratio.



**Fig. 2.** Relationship between synonymous substitution rate relative to the root age ( $d_s/ra$ ) and the proportion of the terminal branch that is subtterranean (RCT), for the nuclear (left) and mitochondrial (right) genomes. Subterranean species are depicted with white circles and surface species with black circles. The two species of one pair are linked with a black line. Different gray levels indicate species for which the subtterranean lifestyle represents <1% of the terminal branch (left) or >10% of the terminal branch (right).

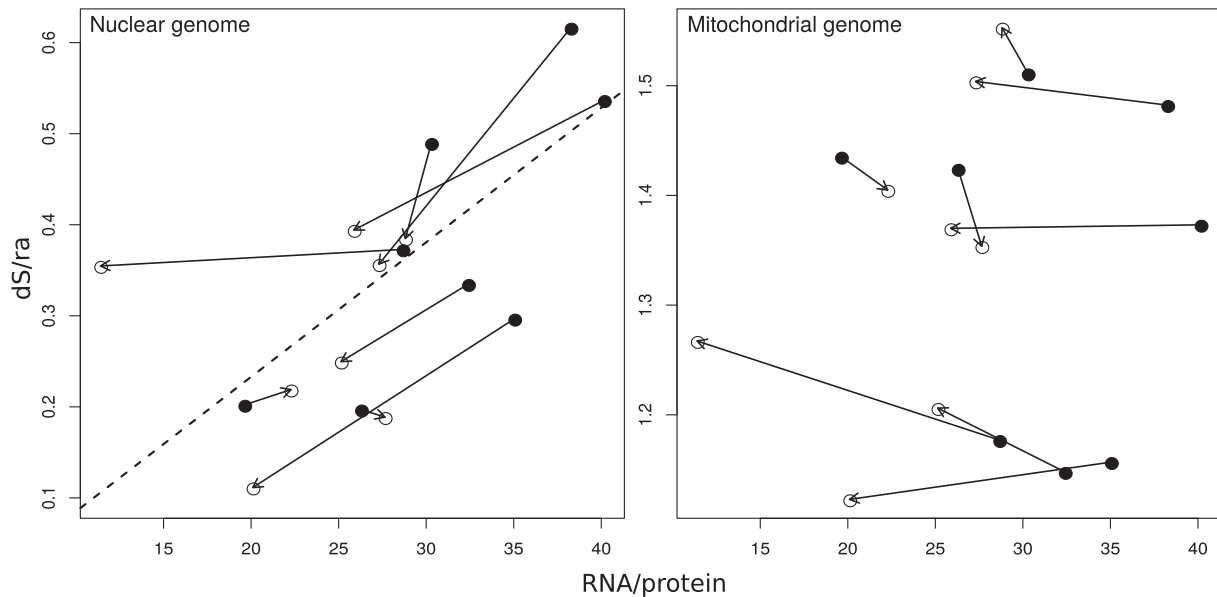
This ratio is a good proxy of the number of active ribosomes per synthesized protein (Cox 2003; Karpinets et al. 2006) and when measured at the organism level it reflects the overall speed at which proteins are renewed. A positive correlation between the RNA/protein and growth rate has been observed in unicellular organisms (Neidhardt and Magasanik 1960; Leick 1968; Cox 2003; Karpinets et al. 2006), in rotifers (Wojewodzic et al. 2011), cephalopods (Houlihan et al. 1990; Pierce et al. 1999), ray-finned fish (Mathers et al. 1992; Peragon et al. 2001), copepoda (Wagner et al. 2001), cladocera (McKee and Knowles 1987), and even in plants (Xing et al. 2016). Though different from the more commonly used resting oxygen consumption per gram of tissue, this ratio should also be positively correlated with the MR. For example, induced hypermetabolism in mammals is associated with a higher RNA/protein ratio (Tata et al. 1962, 1963). That said, the exact relationship between the RNA/protein ratio, growth rate and MR have yet to be thoroughly tested in Isopods. The RNA/protein ratio was found to be lower in subtterranean species (fig. 3, PGLS  $P$ -value = 0.03). However, no significant correlation was found between this ratio and the mitochondrial  $d_s$  (fig. 3, table 1, PGLS test). On the other hand, the RNA/protein ratio was positively correlated with the nuclear  $d_s$  and negatively with the  $d_{s_{mito}}/d_{s_{nuc}}$  ratio (PGLS test, table 1).

### Confounding Factors and Analytical Bias

#### Translational Selection

Subterranean species have smaller long-term effective population sizes (Lefebure et al. 2017). While effective population size does not impact the rate of neutral substitution, it can theoretically impact the  $d_s$ . Indeed, effective population size affects the efficacy of selection, including translational

selection on codon usage, which in turn has an impact on the synonymous rate (Duret and Mouchiroud 1999). For that reason, differences in effective population size might differentially impact the  $d_s$  of subtterranean and surface species (Chamary et al. 2006). For each data set, we checked for a change in the intensity of translational selection by computing for each codon the difference in the relative synonymous codon usage ( $\Delta RSCU$ ) between the 10% most highly expressed and the 10% most lowly expressed nuclear genes (supplementary table S3, Supplementary Material online). If translational selection is strong, the most highly expressed genes should have a higher frequency of optimal codons (i.e., codons corresponding to the most frequent tRNA). We considered that codons with high and positive  $\Delta RSCU$  (>0.05) are the optimal codons. Because there are only 13 protein-coding mitochondrial genes, RSCU computation is not reliable for mitochondrial genes. As a substitute, we computed the ENC, the effective number of codons (ENC; Wright 1990). This number varies between 20 (indicating that only a single codon is used for each amino acid) and 61 (indicating that all synonymous codons are used with equal frequency for any one amino-acid). It should be noted that ENC does not directly detect translational selection, however, it is expected to reflect a codon usage bias that can be linked to translational selection. For the nuclear genes, despite a slight trend for an excess of optimal codons in surface than in subtterranean species it was not significant (PGLS,  $P$ -value = 0.18). Nonetheless this trend cannot explain the observed pattern, as stronger translational selection in surface species should reduce their  $d_s$  not increase it. For the mitochondrial genes, ENC ranges between 43 and 55, except for *Bragasellus peltatus* which has an ENC of 38 (supplementary table S1, Supplementary Material online). These values are



**Fig. 3.** Relationship between synonymous substitution rate relative to root age ( $d_s/ra$ ) and the RNA/protein ratio for the nuclear (left) and mitochondrial (right) genomes. Subterranean species are depicted with white circles and surface species with black circles. The two species of one pair are linked with a black line. The dotted line indicates the linear regression estimated using a PGLS.

high (Wright 1990), indicating that there is a weak mitochondrial codon usage bias. Moreover, ENC scores are not correlated with the species habitat (PGLS,  $P$ -value = 0.38). Altogether, there is no evidence that selection on synonymous codon usage is biasing the estimated nuclear and mitochondrial  $d_s$ .

#### Ultraviolet (UV) Radiation

The complete absence of UV radiation in subterranean habitats could impact the mutation rate, independently of LHTs variation. UV radiation causes photo-excitation between pyrimidine dimers, which twists the DNA molecule. DNA and RNA polymerases cannot transcribe/copy the damaged bases and instead replace them by two thymine bases (Brash 2015). Thus, any variation in the mutation rate caused by different UV exposures will be accompanied by a concomitant variation in the frequency of the CpT, TpC, CpC, or TpT dinucleotides. To detect a bias in dinucleotide composition, we computed the  $Z_{score}$  score statistic proposed by Palmeira et al. (2006). Briefly, this method computes the excess of each dinucleotide in synonymous positions compared with a neutral model based on nucleotide frequencies. We detected no difference in the composition in the four pyrimidine dinucleotides between the surface and subterranean species (fig. 4, Wilcoxon,  $P$ -value > 0.05), suggesting that the absence of UV in subterranean habitat is not the factor reducing the subterranean nuclear  $d_s$ .

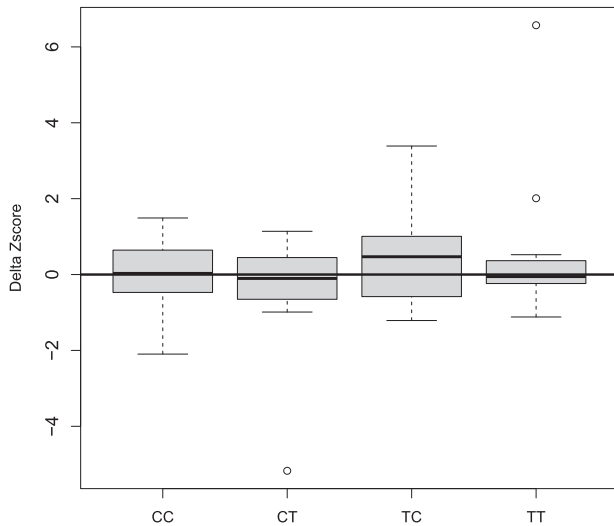
#### Saturation

As mitochondrial genomes evolve faster than nuclear genomes, our estimate of the mitochondrial substitution rate is more prone to be biased by saturation (multiple substitutions at certain nucleotide positions). Stronger saturation in the mitochondrial genome could potentially mask

variation in  $d_s$  that would otherwise be observable in a non-saturated data-set like the nuclear one. In other words, the observed decoupling between the nuclear and mitochondrial  $d_s$  may be the consequence of saturation of the mitochondrial rate. To estimate the level of saturation, we compared the observed divergence (number of differences observed between two sequences) to the patristic divergence along the phylogenetic tree, estimated using a complex model of evolution (GTR + G + I model). The latter estimate of divergence will at least partially compensate for the effect of saturation whereas the former will not (Philippe and Forerre 1999). While mitochondrial third position transitions are quickly saturated, this is not the case for the third position transversions, which vary linearly with the distance between species (supplementary fig. S3, Supplementary Material online). Consequently, to test if saturation hides mitochondrial rate variation, we used the synonymous transversions. As seen with the  $d_s$ , the number of mitochondrial synonymous transversions is not correlated with habitat status nor with colonization time (table 1).

#### Statistical Power

The number of analyzed mitochondrial genes is much lower than the number of nuclear genes (12 vs. 382). The observed lack of mitochondrial rate variation could therefore be the result of a reduced power in the mitochondrial case. By considering only synonymous transversions, this difference in data size is even more pronounced. To test if differences in data size alone might explain the observed differences between the nuclear and mitochondrial genomes, we performed 1,000 resamplings of 26,566 codons in the nuclear alignment to obtain the same number of substitutions in the two compartments (on average 504 synonymous transversions on a terminal branch). We then computed the mean



**Fig. 4.** Delta  $Z_{score}$  ( $Z_{score_{surface}} - Z_{score_{subterranean}}$ ) for the four dinucleotide combinations sensitive to UV rays. The  $Z_{score}$  (Palmeira et al. 2006) computes the excess of each dinucleotide in synonymous positions compared with a neutral model based on nucleotide frequencies. If the absence of UV exposure in subterranean habitat had a direct influence on the mutation rate, we would observe positive delta  $Z_{score}$ .

synonymous transversion contrast  $[S = \frac{1}{13} \sum_{i=1}^{13} (\log(\frac{TVS_{surface_i}}{TVS_{subterranean_i}}))]$  for each nuclear sample and for the concatenation of mitochondrial genes, so as to estimate the probability that the mitochondrial genes behave like a subsample of the nuclear genes. This resampling shows that, despite a strong reduction in data size, surface species still display a higher rate of nuclear molecular evolution than subterranean species ( $S_{mean} = 0.253$ , 95% CI = 0.251–0.254, fig. 5). The mitochondrial contrast is marginally negative and outside the nuclear distribution ( $S = -0.04$ ). Thus, the absence of rate variation in the mitochondrial genome is not linked to a reduced statistical power.

## Discussion

### Reconsideration of the Main Hypotheses

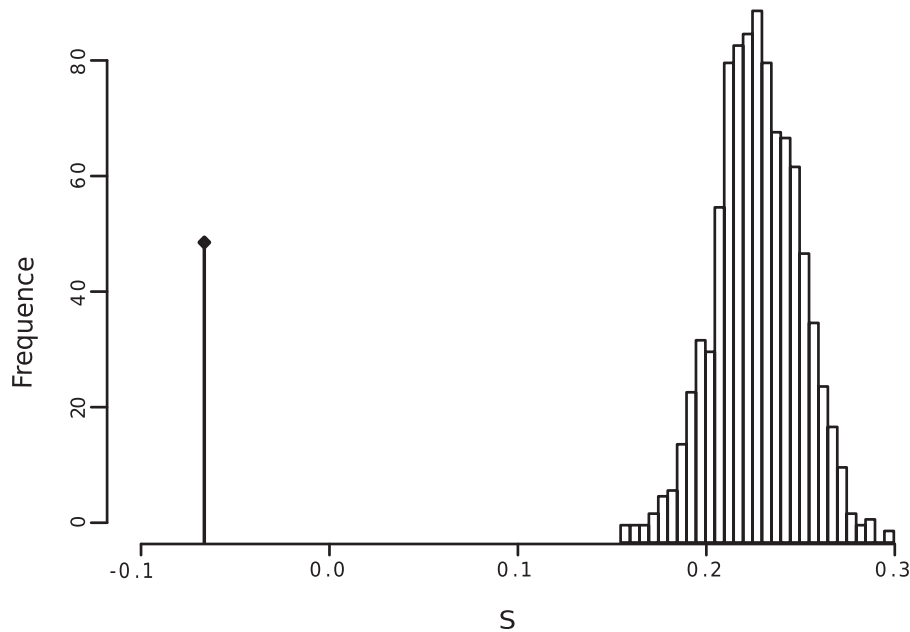
Subterranean isopod species evolved a lower nuclear substitution rate but kept a steady mitochondrial rate. This observation does not correspond to any of the predictions borne out of the two main bodies of hypotheses proposed in the literature, namely those linked to the rate of duplication and those linked to the activity of the respiratory chain, at least in their commonly accepted forms. The two hypotheses linked to the activity of the respiratory chain are based on the idea that the ROS production influences the rate of mutation. Because ROS are produced in the mitochondria, these two hypotheses would both imply a more pronounced decrease in  $d_s$  in the mitochondrial genome than in the nuclear genome. Under the rate of duplication hypothesis one would expect a decrease in  $d_s$  in subterranean species relative to surface species of the same magnitude in the two genomic compartments. We saw no correlation between the RNA/protein ratio—a proxy for MR—and the mitochondrial  $d_s$ . These results are robust to the effect of different colonization

timings, to potential confounding factors (differences in codon usage bias and UV exposure) and to methodological bias (differences in the level of saturation and data size). Surprisingly, whereas the RNA/protein ratio is not correlated with the mitochondrial  $d_s$ , it is strongly correlated with the nuclear  $d_s$ . Although this method needs further testing to be applied to these isopods, absence of correlation for the mitochondrial  $d_s$  is in accordance with previous studies, which found no correlation between MR and mitochondrial substitution rate (Lanfear et al. 2007). A direct impact of the MR on the nuclear genome but not on the mitochondrial one is highly unlikely because the mutagenic agents, here the ROS, are produced in the mitochondria and should impact it first. As the MR is often strongly correlated to GT (Martin and Palumbi 1993), the correlation between RNA/protein ratio and the nuclear  $d_s$  suggests that the GT—rather than the MR—does indeed impact the rate of substitution but that it does so only in the nuclear genome. In addition, it supports the view that the activity of the respiratory chain, either directly through the production of ROS and indirectly through longevity, has little to no impact on either the mitochondrial or the nuclear substitution rates. These results are in accordance with a previous study which showed that mutations due to ROS are negligible compared with other sources of mutation such as errors of the  $\gamma$  polymerase (Larsson 2010).

### Decoupling of the Two Genomes by Independent Replication Cycles

Contrary to the predictions raised by earlier hypotheses, mitochondrial  $d_s$  does not appear to be impacted by the same factors that impact nuclear  $d_s$ . We found that the  $d_{s_{mito}}/d_{s_{nuc}}$  ratio varies extensively across species, showing an increase in subterranean species. This result is not a byproduct of saturation, a lack of statistical power, or other confounding factors (UV and translational selection). Allio et al. (2017) also reported extensive ratio of mitochondrial to nuclear mutation rate variation among animal phyla, with a major shift between nonvertebrates and vertebrates ( $\sim 6$  vs.  $\sim 20$ , respectively). However, our results point to a variation in  $d_{s_{mito}}/d_{s_{nuc}}$  over a much smaller phylogenetic scale. We observed substantial variation across species (3.3–40.3) of the same magnitude as that reported by Allio et al. (2017) across phyla (supplementary table S2, Supplementary Material online). Instead of different  $d_{s_{mito}}/d_{s_{nuc}}$  per phyla, we propose that there is a complete and permanent decoupling of the mitochondrial and nuclear rates across all taxa. In our data set, but also in Allio et al.'s data, the mitochondrial rate varied very little. Therefore, decoupling in these two data sets is apparently controlled by variation in the nuclear rate alone, which we suggest is driven by a GT effect.

To accommodate this observation, we propose a model in which the nuclear and mitochondrial  $d_s$  are decoupled, fundamentally because these two genomes have independent replication cycles. Available data in Asellidae (Henry 1976) indicate a general lengthening of all stages across the life cycle (lengthening of the embryonic stages, puberty postponement, and increased lifespan) in subterranean species. On the basis of the principle that most mutations occur during



**Fig. 5.** Testing if the observed uncoupling between the nuclear and mitochondrial rates is due to a data-set size difference by subsampling the nuclear data-set. The statistic  $\left[ S = \frac{1}{13} \sum_{i=1}^{13} \left( \log \left( \frac{TvS_{surface_i}}{TvS_{subterranean_i}} \right) \right) \right]$  was computed on 1,000 resampling of 26,566 codons of the nuclear alignment (in white) and on the mitochondrial alignment (black line).  $TvS_{surface_i}$  is the number of synonymous transversions estimated on the terminal branch leading to the surface species in the pair  $i$  and  $TvS_{subterranean_i}$  is the analogous estimates in the subterranean species of the pair.

genome replication (Gangloff et al. 2017), we can then assume that a lengthening of the GT leads to a decrease in the cell duplication rate, inducing a decrease in the nuclear substitution rate. If the mitochondrial replication rate does not change accordingly, then this could explain why GT as not effect on the mitochondrial substitution rate. We propose that the number of nuclear genome replication per unit of time reduces in the subterranean species as a by-product of a longer GT, but that the number of mitochondrial genome replication per unit of time is stable. In other words, the rate of mitochondrial replication does not seem to be correlated to the rate of germ stem cell mitosis in these species.

### Gametogenesis May Control Rate Decoupling

The differential impact of LHT on the two metazoan genomic compartments was unexpected and, to the best of our knowledge, has not been explicitly described in previous studies. Thus far, most studies have tested only small numbers of nuclear genes and were almost exclusively devoted to the study of mammals or birds. These two endothermic taxa present particularities in their gametogenesis. In these organisms, primary oocytes are formed during embryonic development where they begin meiosis and are stopped in prophase I. Then, oocytes stay in a quiescent state until just before ovulation. Over this time, mitochondria replicate very little or not at all (de Paula et al. 2013), imposing a synchronization of replication of the mitochondrial and nuclear genomes. If the rate of mutation is mainly governed by the rate of genome replication, and even if the nuclear and mitochondrial replication rates are independent, the observed coupling between the two rates in mammals and birds could be a result of this imposed gametic synchronization. In contrast, in arthropods,

gametogenesis is continuous: primitive germ cells (germ stem cells) undergo mitosis during the entire life of the organism, and a small fraction of these stem cells periodically commence meiosis (Charniaux-Cotton 1973). As such, gametogenesis does not impose replication synchronization, allowing a complete uncoupling of the mitochondrial and nuclear mutation rates in arthropods. It should be noted that spermatogenesis in mammals is similar to arthropod gametogenesis. However, as mitochondria transmission is strictly maternal, this should not influence the mitochondrial mutation rate.

### Aging and Mutation Rate

The absence of any influence of the activity of the respiratory chain, whether direct (MR hypothesis) or indirect (longevity hypothesis), on the mitochondrial  $d_S$  suggests that mitochondrial metabolism has little influence on the germline mutation rate. These results are in accordance with the point of view that the mitochondrial theory of aging implies an oversimplified link between MR, ROS production and mutation (Lapointe and Hekimi 2010). Indeed, the absence of an influence of the activity of the respiratory chain on the germline mutation rate can be explained by two nonexclusive hypotheses. First, an increased longevity is not necessarily accompanied by better ROS management. Indeed, Jobson et al. (2010) showed no consistent variation in  $d_N/d_S$  in genes linked to oxidative stress in long-lived mammals. On the other hand, they found a lower  $d_N/d_S$  in genes coding for lipid composition of cellular membrane. Membrane fatty acid composition modulates the resistance to oxidative damage. Thus, a strategy to increase possible lifespan may not be to reduce ROS production and damage, but instead, to improve tissue resistance. This strategy would have no impact on the rate of

molecular evolution and may explain the absence of a correlation between the RNA/protein ratio and the mitochondrial  $d_s$ . This is particularly likely as the impact of ROS on the mitochondrial mutation rate seems to be marginal (Larsson 2010). Second, a differential impact of ROS in somatic cells between long- and short-lived species does not necessarily mean that this difference exists in germinal cells. The longevity hypothesis is based on the mitochondrial theory of aging. This theory proposes that ROS produced by metabolic activity over the life of an organism would cause mutations in the mitochondrial genome, which would then lead to a senescent phenotype. The longevity hypothesis makes the assumption that these mechanisms are ubiquitous, namely that the selective pressure to decrease the somatic mutation rate leads to an overall decrease in the germinal mutation rate. However, the activity of the respiratory chain is lower in germinal cells than in somatic cells (de Paula et al. 2013), and thus germinal cells have less ROS exposure. It is therefore possible that the mechanism in place in somatic cells to reduce the impact of ROS is not active in the germline. Thus, the somatic mutation rate can be decoupled from the germinal mutation rate and respond differently to LHT variation. Altogether, we propose that ROS have only a limited influence on the substitution rate.

## Conclusion

We have showed that in isopods, increased GT in subterranean species is correlated with decreased nuclear substitution rate but that neither the activity of the respiratory chain nor the GT or longevity influenced the mitochondrial substitution rate. We propose that this complete decoupling between the nuclear and mitochondrial rate is linked to the fact that these two compartments are not subjected to the same replication schedule during development. Most prior studies have been based on few genes, rarely nuclear, and have been biased toward mammals and birds which have peculiar gametogenesis and metabolism. The hypotheses that were developed from these studies (namely the GT, MR, and longevity hypotheses) fit poorly with our observations, which puts their universality into question. We suggest that the rate of substitution is mainly controlled by the number of duplications until fertilization per unit of time which will depend on the type of gametogenesis, the organelle and the GT.

## Materials and Methods

### Data Acquisition

#### Nuclear Genes

Three hundred and eighty six one-to-one orthologous nuclear genes from 13 species pairs (sensu Felsenstein 2004) were retrieved from a study by Francois et al. (2016) (<http://dx.doi.org/10.5281/zenodo.53830>; last accessed October 06, 2018). This data set was built using de novo assembled transcriptomes (ENA project PRJEB14193) followed by gene family definition and ortholog sequence extraction using a tree pattern (Dufayard et al. 2005). Four genes presenting suspicious gene tree topologies or branch lengths were filtered out

to avoid potential contaminants (FAM000151\_1, FAM000919\_1, FAM000990\_2, FAM003078\_1).

#### Mitochondrial Genes

Mitochondrial genes were not present amongst the 382 genes obtained above. Indeed, owing to a different genetic code in invertebrate mitochondria, mitochondrial ORFs were systematically missed by the ORF caller (Transdecoder) used by Francois et al. (2016). We reconstructed mitochondrial genomes using both the de novo transcriptome assemblies from Francois et al. (2016) and low coverage genome sequencing reads available for 22 species from Lefébure et al. (2017). Since mitochondrial genomes are present in multiple copies per cell, a very low sequencing effort is sufficient to yield a high coverage for mitochondrial sequences. Using the DNA-seq reads, mitogenomes were assembled with MITObim (Hahn et al. 2013) using the COI gene as a seed to build the complete mitogenome. For the four species without DNA-seq reads (*P. ebrensis*, *P. cantabricus*, *P. ortizi*, and *P. graffi*), large mitochondrial contigs were built with MITObim by using RNA-seq reads and the mitochondrial genes from the closest possible species (taken from the previous 22 species with assembled mitogenomes) as seeds. The mitochondrial genome from this closest species was used as a reference to map the RNA contigs and to assemble complete mitogenomes. Mitochondrial genomes were annotated using the MITOS web server (Bernt et al. 2013, supplementary fig. S2, Supplementary Material online). We recovered 12 mitochondrial protein-coding genes (*cytb*, *cox1*, *cox2*, *cox3*, *atp6*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4l*, *nad5*, *nad6*) for each of the 26 species. Because of its short size, *atp8* was difficult to identify and delimit, and was thus not included in subsequent analyses. Mitochondrial genes were aligned with PRANK (Löytynoja and Goldman 2008) and sites ambiguously aligned were removed with Gblocks (Castresana 2000). The 382 nuclear and 12 mitochondrial gene alignments used in this study have been deposited on Zenodo (<https://doi.org/10.5281/zenodo.1409532>; last accessed October 06, 2018).

#### Rate of Molecular Evolution

We used the synonymous substitution rate ( $d_s$ ) as a proxy for the mutation rate. Genes were concatenated independently for the nuclear and the mitochondrial compartment. Phylogenetic tree was built with the 382 nuclear genes with PhyML v3.0 (Guindon et al. 2010) under a GTR + G + I model with 500 bootstrap replicates and was rooted using the *Bragasellus* lineage as an outgroup. Synonymous substitution rates per unit of time were estimated using CoEvol (Lartillot and Poujol 2012). This software program implements a Muse and Gaut codon model (Muse and Gaut 1994), with Brownian variation in  $d_s$  and  $\omega = d_N/d_s$  along the tree. Bayesian inference and reconstruction of the history of variation in  $d_s$  and  $d_N/d_s$  along the tree is conducted by Markov Chain Monte Carlo (MCMC). Two independent chains were run, and were stopped after checking for convergence by eye and with the tracecomp program included in the Coevol package (effective sample size > 300

and discrepancy between chains  $< 0.1$  for all statistics). Chains were stopped after 29,713 generations (burn-in excluded) for mitochondrial genomes, and 4,001 generations (burn-in excluded) for nuclear genomes. The age of the root was arbitrarily set to 1, resulting in synonymous substitution rate estimates that are relative to the root age ( $d_S/ra$ ). The 26 assembled mitogenomes were deposited to the ENA under the accession number from ERZ724045 to ERZ724070.

## Colonization Time

The relative colonization time (RCT), which is the proportion of the branch corresponding to a subterranean lifestyle was estimated using the nonfunctionalization of the opsin I gene locus as in Lefébure et al. (2017). Assuming a two-state model of evolution with one surface opsin  $d_N/d_S$  estimated using opsins from surface species, and one subterranean opsin  $d_N/d_S$  equal to 1, we estimated the colonization time ( $t$ ) as a function of the speciation time ( $T$ ) and the estimated opsin  $d_N/d_S$  measured on a given branch leading to a subterranean species (supplementary table S1, Supplementary Material online):

$$RCT = \frac{t}{T} = \frac{d_N/d_{S_{subterranean}} - d_N/d_{S_{surface}}}{1 - d_N/d_{S_{surface}}}$$

$d_N/d_S$  was estimated using the CodeML program in the PAML package (Yang 2007).  $d_N/d_{S_{surface}}$  was computed using the mean of opsin  $d_N/d_S$  of the six occluded and pigmented species (*P. ibericus*, *P. beticus*, *P. meridianus*, *P. coxalis*, *P. coiffaiti*, and *P. karamani*). We retrieved opsin sequences for 19 species from Lefébure et al. (2017) and complemented this data set with four additional species (*P. ebrensis*, *P. cantabricus*, *P. ortizi*, and *P. graffi*), using a combination of Sanger sequencing, transcriptome assemblies and genome sequencing reads. In accordance with Lefébure et al. (2017), this opsin locus was not found in the two species of the genus *Bragasellus* and in one *Proasellus* subterranean species (*P. parvulus*). The opsin alignment used in this study has been deposited on Zenodo (<https://doi.org/10.5281/zenodo.1409532>).

## RNA/Protein Ratio

As in Lefébure et al. (2017), we used the RNA/protein ratio as a proxy for MR. For 16 species, we recovered the RNA/protein ratio from Lefébure et al. (2017). Briefly, for 10 individuals per species, total RNA and proteins were extracted with a TRI-Reagent protocol (Molecular Research Center). Total RNA was quantified by fluorimetry (Qubit; Life technologies) and total protein was obtained using the Bicinchoninic acid assay (Smith et al. 1985). The RNA/protein ratio was then estimated by the total RNA normalized by the total protein biomass (supplementary table S1, Supplementary Material online).

## Confounding Factors and Analytical Bias

### Synonymous Codon Usage

The RSCU corresponds to the number of times a given codon is observed, relative to the number of times this codon would

be observed under a uniform synonymous codon usage (i.e., with all the codons for a given amino-acid having the same probability). Optimal codons are defined by a higher RSCU value in highly expressed transcripts (Duret and Mouchiroud 1999). Thus, optimal codons were identified by subtracting, for each codon, the RSCU averaged over the 10% transcripts having the lowest expression level to the RSCU averaged over the 10% most highly expressed transcripts. Expression of each contig was estimated with RSEM (Li and Dewey 2011). The 10% most highly expressed and the 10% most lowly expressed contigs were selected based on TPM (transcript per million) estimates (Li et al. 2010; Wagner et al. 2012). RSCU indices were calculated using the seqinR R package (Charif and Lobry 2007). The ENC was computed for each species on the concatenation of the 12 mitochondrial genes with the vha R package (Wallau et al. 2016).

### UV Radiation

To test if UV radiation affects the rate of molecular evolution of surface species, we estimated dinucleotide frequencies using the  $Z_{score}$  statistic proposed by Palmeira et al. (2006) in the seqinR package (Charif and Lobry 2007). We computed this statistic on the concatenation of the 382 nuclear genes under a “syncodon” model with 1,000 permutations (supplementary table S4, Supplementary Material online). This method considers the nucleotide frequency and the codon structure. Because changes in the second position of the codon are never synonymous, only changes in dinucleotides in the first and the third position of the codon are taken into account. Thus, this method produces results similar to what would be measured in neutrally evolving regions.

### Saturation and Power

To reduce the influence of saturation on  $d_S$  estimates, we computed synonymous transversions with the Bio++ suite (Dutheil and Boussau 2008). A nonhomogenous model (NY68 model) model was first applied to the mitochondrial alignment with BppML and then the MapNH program (Version 1.1.1) of the TestNH package (Guéguen and Duret 2018) was used to reconstruct the ancestral states and to estimate the number of synonymous transversions on each branch (supplementary table S5, Supplementary Material online). The relation between synonymous transversion number and ecology, colonization time and RNA/protein ratio was tested with a PGLS.

### Comparative Analyses

Correlations between the rate of molecular evolution and habitat, RCT and RNA/protein ratio were tested by a Phylogenetic Generalised Least Squares (PGLS, Martins and Hansen 1997). Tests were performed by using the nlme (Pinheiro et al. 2014) and ape packages (Paradis et al. 2004) in R (R Core Team 2014). The output chronogram built by CoEvol on the 382 nuclear genes was used to correct for phylogenetic inertia. Second, for each pair composed of a surface and a subterranean species, and based on the output of Coevol, we computed the posterior probability of an



increase of the  $d_s$  in surface species as the percentage of point of the Monte-Carlo chain where the surface species has a higher  $d_s$  than the subterranean species.

## Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

## Acknowledgments

This work was supported by the Agence Nationale de la Recherche (ANR-15-CE32-0005 Convergenomix, France) and the CNRS (APEGE No. 70632, PEPS ExoMod 2014 and Enviromics 2014, France). We gratefully acknowledge support from the CNRS/IN2P3 Computing Center (Lyon/Villeurbanne—France), for providing a significant amount of the computing resources needed for this work. We thank Florian Mermillot-Blondin and Laurent Simon for their help in the field and for the RNA/protein ratio measurement. We are grateful to Benoît Nabholz, Nicolas Galtier, Marie Semon, and Carine Rey for helpful discussions. We also thank Laura Grice for her comments and suggestions in the latter stages of manuscript preparation, as well as two anonymous reviewers for their helpful comments.

## References

- Allio R, Donega S, Galtier N, Nabholz B. 2017. Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Mol Biol Evol.* 34(11):2762–2772.
- Baer CF, Miyamoto MM, Denver DR. 2007. Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nat Rev Genet.* 8(8):619–631.
- Barja G. 2007. Mitochondrial oxygen consumption and reactive oxygen species production are independently modulated: implications for aging studies. *Rejuvenation research*, 10(2):215–224.
- Barja G. 2002. Endogenous oxidative stress: relationship to aging, longevity and caloric restriction. *Ageing Res Rev.* 1(3):397–411.
- Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsche G, Pütz J, Middendorf M, Stadler PF. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol.* 69(2):313–319.
- Brash DE. 2015. UV signature mutations. *Photochem Photobiol.* 91(1):15–26.
- Bromham L, Hua X, Lanfear R, Cowman PF. 2015. Exploring the relationships between mutation rates, life history, genome size, environment, and species richness in flowering plants. *Am Nat.* 185(4):507–524.
- Bromham L, Rambaut A, Harvey PH. 1996. Determinants of rate variation in mammalian DNA sequence evolution. *J Mol Evol.* 43(6):610–621.
- Caccone A, Gentile G, Burns CE, Sezzi E, Bergman W, Ruelle M, Saltonstall K, Powell JR. 2004. Extreme difference in rate of mitochondrial and nuclear DNA evolution in a large ectotherm, Galápagos tortoises. *Mol Phylogenet Evol.* 31(2):794–798.
- Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 17(4):540–552.
- Chamary JV, Parmley JL, Hurst LD. 2006. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat Rev Genet.* 7(2):98.
- Charif D, Lobry JR. 2007. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: Bastolla U, Porto M, Roman HE, Vendruscolo M, editors. *Structural Approaches to Sequence Evolution*. Berlin, Heidelberg: Springer Berlin Heidelberg, 207–232.
- Charniaux-Cotton. 1973. Description et contrôle de l'ovogenèse chez les Crustacés supérieurs. *Ann Biol Anim Biochim Biophys.* 13:21–37.
- Cox RA. 2003. Correlation of the rate of protein synthesis and the third power of the RNA: protein ratio in *Escherichia coli* and *Mycobacterium tuberculosis*. *Microbiology* 149(Pt 3):729–737.
- Doublet V, Raimond R, Grandjean F, Lafitte A, Souty-Grosset C, Marcadé I. 2012. Widespread atypical mitochondrial DNA structure in isopods (Crustacea, Peracarida) related to a constitutive heteroplasmy in terrestrial species. *Genome* 55(3):234–244.
- Dufayard JF, Duret L, Penel S, Gouy M, Rechenmann F, Perrière G. 2005. Tree pattern matching in phylogenetic trees: automatic search for orthologs or paralogs in homologous gene sequence databases. *Bioinformatics* 21(11):2596–2603.
- Duret L, Mouchiroud D. 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc Natl Acad Sci U S A.* 96(8):4482–4487.
- Duret L, Mouchiroud D. 2000. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Mol Biol Evol.* 17(1):68–070.
- Dutheil J, Boussau B. 2008. Non-homogeneous models of sequence evolution in the Bio++ suite of libraries and programs. *BMC Evol Biol.* 8:1.
- Ellegren H, Smith NG, Webster MT. 2003. Mutation rate variation in the mammalian genome. *Curr Opin Genet Dev.* 13(6):562–568.
- Felsenstein J. 2004. *Inferring phylogenies*, volume 2. Sunderland (MA): Sinauer Associates.
- Francois CM, Mermillod-Blondin F, Malard F, Fourel F, Lécuyer C, Douady CJ, Simon L. 2016. Trophic ecology of groundwater species reveals specialization in a low-productivity environment. *Funct Ecol.* 30(2):262–273.
- Galtier N, Jobson RW, et al. 2009. Mitochondrial whims: metabolic rate, longevity and the rate of molecular evolution. *Biol Lett* 5(3):413–6.
- Galtier N, Nabholz B, et al. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol Ecol.* 18(22):4541–4550.
- Gangloff S, Achaz G, Francesconi S, Villain A, Miled S, Denis C, Arcangioli B. 2017. Quiescence unveils a novel mutational force in fission yeast. *eLife* 6:e27469.
- Gillooly JF, Allen AP, West GB, Brown JH. 2005. The rate of DNA evolution: effects of body size and temperature on the molecular clock. *Proc Natl Acad Sci U S A.* 102(1):140–145.
- Guéguen L, Duret L. 2018. Unbiased estimate of synonymous and non-synonymous substitution rates with nonstationary base composition. *Mol Biol Evol.* 35(3):734–742.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59(3):307–321.
- Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res.* 41(13):e129.
- Henry JPA. 1976. Recherches sur les Asellidae hypogés de la lignée cavaticus: Crustacea, Isopoda, Asellota. [s.n.]. [S.l.].
- Houlihan DF, McMillan DN, Agnisola C, Genoio IT, Foti L. 1990. Protein synthesis and growth in *Octopus vulgaris*. *Mar Biol.* 106(2):251–259.
- Hüppop K. 1987. Food-finding ability in cave fish (*Astyanax fasciatus*). *Int J Speleol.* 16(1):4.
- Jeschke JM, Kokko H. 2009. The roles of body size and phylogeny in fast and slow life histories. *Evol Ecol.* 23(6):867–878.
- Jobson RW, Nabholz B, Galtier N. 2010. An evolutionary genome scan for longevity-related natural selection in mammals. *Mol Biol Evol.* 27(4):840–847.
- Karpinets TV, Greenwood DJ, Sams CE, Ammons JT. 2006. RNA: protein ratio of the unicellular organism as a characteristic of phosphorous and nitrogen stoichiometry and of the cellular requirement of ribosomes for protein synthesis. *BMC Biol.* 4:30.
- Kennedy SR, Salk JJ, Schmitt MW, Loeb LA. 2013. Ultra-sensitive sequencing reveals an age-related increase in somatic mitochondrial

- mutations that are inconsistent with oxidative damage. *PLOS Genet.* 9(9):e1003794.
- Kilpert F, Held C, Podsiadlowski L. 2012. Multiple rearrangements in mitochondrial genomes of Isopoda and phylogenetic implications. *Mol Phylogenet Evol.* 64(1):106–117.
- Kilpert F, Podsiadlowski L. 2006. The complete mitochondrial genome of the common sea slater, *Ligia oceanica* (Crustacea, Isopoda) bears a novel gene order and unusual control region features. *BMC Genomics.* 7:241.
- Kujoth GC, Bradshaw PC, Haroon S, Prolla TA. 2007. The role of mitochondrial DNA mutations in mammalian aging. *PLoS Genetics.* 4:161–173.
- Lanfear R, Thomas JA, Welch JJ, Brey T, Bromham L. 2007. Metabolic rate does not calibrate the molecular clock. *Proc Natl Acad Sci U S A.* 104(39):15388–15393.
- Lapointe J, Hekimi S. 2010. When a theory of aging ages badly. *Cell Mol Life Sci.* 67(1):1–8.
- Larsson NG. 2010. Somatic mitochondrial DNA mutations in mammalian aging. *Annu Rev Biochem.* 79:683–706.
- Lartillot N, Delsuc F. 2012. Joint reconstruction of divergence times and life-history evolution in placental mammals using a phylogenetic covariance model. *Evolution* 66(6):1773–1787.
- Lartillot N, Poujol R. 2011. A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. *Mol Biol Evol.* 28(1):729–744.
- Lefébure T, Morvan C, Malard F, et al. (17 co-authors). 2017. Less effective selection leads to larger genomes. *Genome Res.* 27:1016–1028.
- Leick V. 1968. Ratios between contents of DNA, RNA and protein in different micro-organisms as a function of maximal growth rate. *Nature* 217(5134):1153–1155.
- Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 12:1.
- Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN. 2010. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* 26(4):493–500.
- Li WH, Tanimura M, Sharp PM. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *J Mol Evol.* 25(4):330–342.
- Lynch M. 2010. Evolution of the mutation rate. *Trends Genet.* 26(8):345–352.
- Löytynoja A, Goldman N. 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* 320(5883):1632–1635.
- Martin AP, Naylor GJP, Palumbi SR. 1992. Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. *Nature* 357(6374):153.
- Martin AP, Palumbi SR. 1993. Body size, metabolic rate, generation time, and the molecular clock. *Proc Natl Acad Sci U S A.* 90(9):4087–4091.
- Martins EP, Hansen TF. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am Nat.* 149(4):646–667.
- Mathers EM, Houlihan DF, Cunningham MJ. 1992. Nucleic acid concentrations and enzyme activities as correlates of growth rate of the saithe *Pollachius virens*: growth-rate estimates of open-sea fish. *Mar Biol.* 112(3):363–369.
- McKee MJ, Knowles CO. 1987. Levels of protein, RNA, DNA, glycogen and lipid during growth and development of *Daphnia magna* Straus (Crustacea: cladocera). *Freshw Biol.* 18(2):341–351.
- Mooers AO, Harvey PH. 1994. Metabolic rate, generation time, and the rate of molecular evolution in birds. *Mol Phylogenet Evol.* 3(4):344–350.
- Morvan C, Malard F, Paradis E, Lefébure T, Konecny-Dupré L, Douady CJ. 2013. Timetree of aselloidea reveals species diversification dynamics in groundwater. *Syst Biol.* 62(4):512–522.
- Muse SV, Gaut BS. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol.* 11(5):715–724.
- Nabholz B, Glémin S, Galtier N. 2008. Strong variations of mitochondrial mutation rate across mammals—the longevity hypothesis. *Mol Biol Evol.* 25(1):120–130.
- Neidhardt FC, Magasanik B. 1960. Studies on the role of ribonucleic acid in the growth of bacteria. *Biochim Biophys Acta.* 42:99–116.
- Niemiller ML, Near TJ, Fitzpatrick BM. 2012. Delimiting species using multilocus data: diagnosing cryptic diversity in the Southern Cavefish, *Typhlichthys Subterraneus* (teleostei: Amblyopsidae). *Evolution* 66(3):846–866.
- Nikolaev SI, Montoya-Burgos JL, Popadin K, Parand L, Margulies EH, et al. 2007. Life-history traits drive the evolutionary rates of mammalian coding and noncoding genomic elements. *Proc Natl Acad Sci U S A.* 104(51):20443–20448.
- Ohta T. 1993. An examination of the generation-time effect on molecular evolution. *Proc Natl Acad Sci U S A.* 90(22):10676–10680.
- Palmeira L, Guéguen L, Lobry JR. 2006. UV-targeted dinucleotides are not depleted in light-exposed prokaryotic genomes. *Mol Biol Evol.* 23(11):2214–2219.
- Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20(2):289–290.
- de Paula WBM, Lucas CH, Agip A-NA, Vizcay-Barrena G, Allen JF. 2013. Energy, ageing, fidelity and sex: oocyte mitochondrial DNA as a protected genetic template. *Philos Trans R Soc B.* 368(1622):20120263.
- Peragon J, Barroso JB, García-Salguero L, de la Higuera M, Lupianez JA. 2001. Growth, protein-turnover rates and nucleic-acid concentrations in the white muscle of rainbow trout during development. *Int J Biochem Cell Biol.* 33(12):1227–1238.
- Philippe H, Forterre P. 1999. The rooting of the universal tree of life is not reliable. *J Mol Evol.* 49(4):509–523.
- Pierce GJ, Key LN, Boyle PR, Siegert KJ, Gonçalves JM, Porteiro FM, Martins HR. 1999. RNA concentration and the RNA to protein ratio in cephalopod tissues: sources of variation and relationship with growth rate. *J Exp Mar Biol Ecol.* 237(2):185–201.
- Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. 2014. nlme: linear and nonlinear mixed effects models. R package version 3.1-117.
- R Core Team. 2014. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Raimond R, Marcadé I, Bouchon D, Rigaud T, Bossy JP, Souty-Grosset C. 1999. Organization of the large mitochondrial genome in the isopod *Armadillidium vulgare*. *Genetics* 151(1):203–210.
- Renault D, Hervant F, Vernon P. 2002. Comparative study of the metabolic responses during food shortage and subsequent recovery at different temperatures in the adult lesser mealworm, *Alphitobius diaperinus* (Coleoptera: tenebrionidae). *Physiol Entomol.* 27(4):291–301.
- Richter C, Park JW, Ames BN. 1988. Normal oxidative damage to mitochondrial and nuclear DNA is extensive. *Proc Natl Acad Sci U S A.* 85(17):6465–6467.
- Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD, Fujimoto EK, Goeke NM, Olson BJ, Klenk DC. 1985. Measurement of protein using bicinchoninic acid. *Anal Biochem.* 150(1):76–85.
- Smith SA, Donoghue MJ. 2008. Rates of molecular evolution are linked to life history in flowering plants. *Science* 322(5898):86–89.
- Tata JR, Ernster L, Lindberg O. 1962. Control of basal metabolic rate by thyroid hormones and cellular function. *Nature* 193:1058–1060.
- Tata JR, Ernster L, Lindberg O, Arrhenius E, Pedersen S, Hedman R. 1963. The action of thyroid hormones at the cell level. *Biochem J.* 86:408–428.
- Thomas JA, Welch JJ, Lanfear R, Bromham L. 2010. A generation time effect on the rate of molecular evolution in invertebrates. *Mol Biol Evol.* 27(5):1173–1180.
- Wagner GP, Kin K, Lynch VJ. 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* 131(4):281–285.
- Wagner MM, Campbell RG, Boudreau CA, Durbin EG. 2001. Nucleic acids and growth of *Calanus finmarchicus* in the laboratory under

- different food and temperature conditions. *Mar Ecol Prog Ser.* 221:185–197.
- Wallau GL, Capy P, Loreto E, Le Rouzic A, Hua-Van A. 2016. VHICA, a new method to discriminate between vertical and horizontal transposon transfer: application to the mariner family within *Drosophila*. *Mol Biol Evol.* 33(4):1094–1109.
- Welch JJ, Bininda-Emonds OR, Bromham L. 2008. Correlates of substitution rate variation in mammalian protein-coding sequences. *BMC Evol Biol.* 8:53.
- Wojewodzic MW, Rachamim T, Andersen T, Leinaas HP, Hessen DO. 2011. Effect of temperature and dietary elemental composition on RNA/protein ratio in a rotifer. *Funct Ecol.* 25(5):1154–1160.
- Wolfe KH, Sharp PM, Li WH. 1989. Mutation rates differ among regions of the mammalian genome. *Nature* 337(6204):283–285.
- Wright F. 1990. The 'effective number of codons' used in a gene. *Gene* 87(1):23–29.
- Wu Y, Wang H, Hadly EA. 2017. Invasion of ancestral mammals into dim-light environments inferred from adaptive evolution of the phototransduction genes. *Sci Rep.* 7:46542.
- Xing W, Shi Q, Liu H, Liu G. 2016. Growth rate, protein: rNA ratio and stoichiometric homeostasis of submerged macrophytes under eutrophication stress. *Knowl Manage Aquat Ecosyst.* 417:25.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24(8):1586–1591.

---

## Annexe 7 :

### Article : Bedrock radioactivity influences the rate and spectrum of mutation.

Nathanaëlle Saclier<sup>1</sup>, Patrick Chardon<sup>2</sup>, Florian Malard<sup>1</sup>, Lara Konecny-Dupré<sup>1</sup>, David Eme<sup>1,6</sup>, Arnaud Bellec<sup>1,4</sup>, Vincent Breton<sup>2</sup>, Laurent Duret<sup>3</sup>, Tristan Lefébure<sup>1</sup>, Christophe J. Douady<sup>1,5</sup>

<sup>1</sup>Univ Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5023, ENTPE, Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés, F-69622 Villeurbanne, France

<sup>2</sup>LPC, Université Clermont Auvergne, CNRS/IN2P3 UMR6533, Clermont-Ferrand, France

<sup>3</sup>Univ Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622 Villeurbanne, France

<sup>4</sup>Univ Lyon, Université Jean Moulin Lyon 3, CNRS UMR 5600 Environnement Ville Société, F-69362, Lyon, France

<sup>5</sup>Institut Universitaire de France, F-75005 Paris, France

<sup>6</sup>Current address : IFREMER, unité Ecologie et Modeles pour l'Halieutique, Nantes, France

## Abstract

All organisms on Earth are exposed to low doses of natural radioactivity but some habitats are more radioactive than others. Yet, documenting the influence of natural radioactivity on the evolution of biodiversity is challenging. Here, we addressed whether organisms living in naturally more radioactive habitats accumulate more mutations across generations using 14 species of waterlice living in subterranean habitats with contrasted levels of radioactivity. We found that the mitochondrial and nuclear mutation rates across a waterlouse species' genome increased on average by 60 and 30%, respectively, when radioactivity increased by a factor of three. We also found a positive correlation between the level of radioactivity and the probability of G to T (and complementary C to A) mutations, a hallmark of oxidative stress. We conclude that even low doses of natural bedrock radioactivity influence the mutation rate through the likely accumulation of oxidative damage, in particular in the mitochondrial genome.

## Introduction

Natural radioactivity is the main natural source of exposure to ionizing radiations on Earth. Natural radioactivity is generated by cosmic radiation or by radionuclides released from the bedrock. While levels of cosmic radiation fluctuates over time due to cosmic events such as supernovae or solar flares, bedrock radioactivity remained mainly stable until 2 billion years ago, when it began to slowly decrease (Karam and Leslie, 2005). Bedrock radioactivity depends on the nature of the rocks which extensively varies spatially (e.g Ielsch et al., 2017). While few extremely naturally radioactive sites such as the India Kerala and Iranian Ramsar region have been monitored for their impact on the human mutation rate (Forster et al., 2002; Masoomi et al., 2006) or on plant physiology (Saghirzadeh et al., 2008), the influence of regional variation in baseline natural radioactivity on the evolution of biodiversity is still unknown (Møller and Mousseau, 2012).

Natural radioactivity can modify the rate of molecular evolution by increasing the rate of mutations. Ionizing radiations can damage DNA directly by breaking the DNA sugar-phosphate backbone, or indirectly by the radiolysis of water in cells which results in the production of reactive oxygen species (ROS), which are mutagenic. These impacts have been well characterized after exposure to high doses of radioactivity (Dubrova et al., 1996; Ziegler et al., 1993) but the multi-generational impact of exposure to low doses of radioactivity is poorly known. Some authors (Tubiana et al., 2006, 2009) propose that DNA repair may completely counteract the effect of ionizing radiations for doses below 0.1 Gy, suggesting that low doses have no biological impact. However, *in vitro* exposure to very low dose of ionizing radiation increases the mutation rate of mammalian cells (Vilenchik and Knudson, 2000), suggesting that the repair system is not activated at low radiation doses, thus facilitating the accumulation of mutations that may be transmitted across generations. While some studies found an increase in the number of mutations in the offsprings of exposed people (Dubrova et al., 1996; Forster et al., 2002), others studies found the opposite (Satoh C et al., 1996; Czeizel et al., 1991), leaving open the question of the transmission of mutations generated by low doses of radioactivity. Studying the long-term mutational impact of natural radioactivity is challenging because it raises a number of methodological difficulties. On the one hand, experimentally exposing multiple generations of multicellular organisms to low doses of radiation would require years of experimentation and complex experimental controls. On the other hand, the main obstacles to *in naturae* studies are (i) the organisms' mobility, which prevents certainty that a population was exposed to the same natural radioactivity for many generations, and (ii) confounding factors such as ultraviolet radiation from the sun. Here we overcome these difficulties by coupling *in situ* radioactivity characterizations with the distinctive bio-ecological characteristics of subterranean waterlice within a phylogenetic comparative framework. Subterranean waterlice are never exposed to UV

radiation, live in contrasted bedrock set-ups and have very limited dispersal capacity (Eme et al., 2018), allowing us to make the assumption that different species have persisted in different but nearly constant radioactive habitats for numerous generations.

## Results and Discussion

In order to build a robust and powerful comparative design aimed for testing the influence of natural radioactivity on the mutation rate, we first prospected for closely related subterranean species living in contrasted radioactive set-ups. Using the map of bedrock uranium content in France (Ielsch et al., 2017), we prospected areas with low and high radioactivity. From this large survey (58 sites with waterlice), we selected 14 sites roundwith contrasted levels of  $\alpha$  radioactivity which were inhabited by closely related groundwater waterlice species. We paid special attention that  $\alpha$  radioactivity differed at least by a factor of three between two habitats, each containing a closely related species (Figure 1, Table S1). On average low level of radioactivity was around 0.357 Bq/g of dry sediment and high level around 1.259 Bq/g of dry sediment. Based on transcriptome sequencing and *de novo* assembly, we used a phylogenetic approach to estimate nuclear and mitochondrial synonymous substitution rates (*i.e.* the rate of mutations which are fixed). While experimental approaches allow to measure the impact of radioactivity on somatic mutations or on the transmission of mutations across few generations, this phylogenetic approach allows us to measure the impact of natural radiation on the germinal mutation rate over the course of a species history.

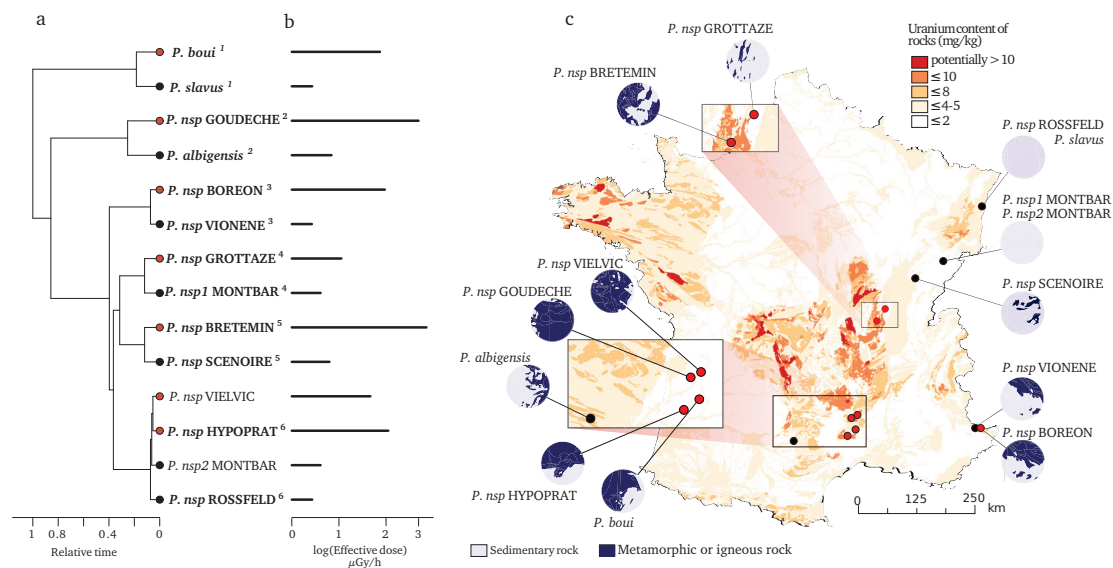


Figure 1: Species and locations selected to study the impact of bedrock radioactivity on the mutation rate and spectrum. 14 species with contrasted bedrock radioactivity exposure were selected (black dot: low exposure, red dot: high exposure). Based on their phylogenetic history (a), we further selected 6 monophyletic pairs of closely related species to compare their mutational spectrum (Vielvic and Montbar are excluded because of unresolved phylogeny, pairs are indicated using superscript numbers). Received dose of radioactivity was measured on sediments of the sampled sites (b). For each site, the areal proportion of low-radioactivity sedimentary rocks and high-radioactivity metamorphic and igneous rocks in a radius of 15 km around the sampling ( $\lambda 15$ ) site is represented with circles next to the map (c).

Using the 14 selected species and locations, we tested whether there was a significant positive relationship between natural radioactivity and the long-term mutation rate. The latter was estimated using the synonymous substitution rate ( $d_S$ ) calculated on the terminal branches of the phylogenetic tree tracing the history of these 14 species. The  $d_S$  is the rate at which silent mutations accumulate in protein coding genes and, when calculated using many different loci and in the absence of a strong synonymous codon usage bias (see methods), is an estimator of the average mutation rate across a species' genome (Kimura, 1983). We computed  $d_S$  using 769 one-to-one nuclear and 13 mitochondrial orthologous protein-coding genes shared by all 14 species. At each sampling site, we measured the global  $\alpha$  radioactivity and the activity of all radio-elements in the sediment. The analysis of the composition in radionuclides at each site reveals that 2 sites (BRETEMIN and BOREON) show a disruption of the secular equilibrium in the U-238 chain. This suggests that nearby industrial activities (e.g. lead mines) have modified the natural radioactivity at these two sites. As these industrial activities are very recent (since 1950), their impact on the substitution rate, which is measured on a much longer time scale, is unlikely. We therefore did not use these two sites to test the correlation between  $d_S$  and any site-specific radioactivity measurement (however, see next paragraph for a regional measurement). The  $d_S$  is positively correlated with the alpha radioactivity in the nuclear genome as well as in the mitochondrial genome (Table 1, Figure 2). A linear model predicts a  $d_S$  increase of 31.8% in the nuclear genome and 56.5% in the mitochondrial genome between species living in low (on average 0.357 Bq/g of dry sediment) and high radioactivity (on average 1.259 Bq/g of dry sediment). We also modeled the biologically effective dose of radioactivity received by each species (in  $\mu\text{Gy/h}$ , Figure 1). This measure takes into account the transfer coefficient from environment to biota and the radio-toxicity of each radio-element for a crustacean model (ERICA tool V1.2.1 Brown et al., 2016). Again, we found positive correlations between the  $d_S$  and the received dose of radioactivity (Table 1).

Table 1: Phylogenetic generalized least square (PGLS) regressions of the nuclear synonymous substitution rate ( $d_S$ ) and mitochondrial  $d_S$  against the  $\alpha$  radioactivity measured in the sediments ( $\alpha$  radio.), the received dose (RD) of radioactivity modeled with the ERICA tool, and the surface of metamorphic and igneous bedrock within a 15 km radius around the sampling sites ( $\lambda 15$ ).  $\alpha$  radioactivity and RD were log transformed to fit with the linear model assumptions.

	Nuclear $d_S$				Mitochondrial $d_S$				N taxa
	Slope	L. Ratio	p. value	$R^2$	Slope	L. Ratio	p. value	$R^2$	
log( $\alpha$ radio.)	0.034	5.995	<b>0.014</b>	0.393	0.506	7.895	<b>0.005</b>	0.482	12
log(RD)	0.038	6.51	<b>0.011</b>	0.419	0.491	5.981	<b>0.015</b>	0.392	12
$\lambda 15$	0.076	9.039	<b>0.003</b>	0.476	1.097	11.680	<b>0.001</b>	0.566	14

NOTE — Each line corresponds to one likelihood ratio test between the models with and without the given explanatory variable.

As previously explained, the measured radioactivity at two sites overestimates the radioactivity level to which the organisms have been exposed for many generations because it is influenced by recent human activities. Moreover, while most species collected in highly radioactive habitats were from metamorphic or igneous formations, two species were from sedimentary formations. Contrary to metamorphic and igneous formations, radioactivity in sedimentary formations is often observed in restricted localities (Ielsch et al., 2017) and can show large variations at the meter scale. A single radiation measurement may not therefore accurately represent the average radiation that a species is exposed to. To account for this variability as well as to include the two human-impacted sites into the regression analysis, we calculated the areal proportion of metamorphic and igneous rock within a 15 km radius around each site (later called  $\lambda 15$ , Figure 1). This proportion was used as a proxy for the long-term regional radioactive exposure because

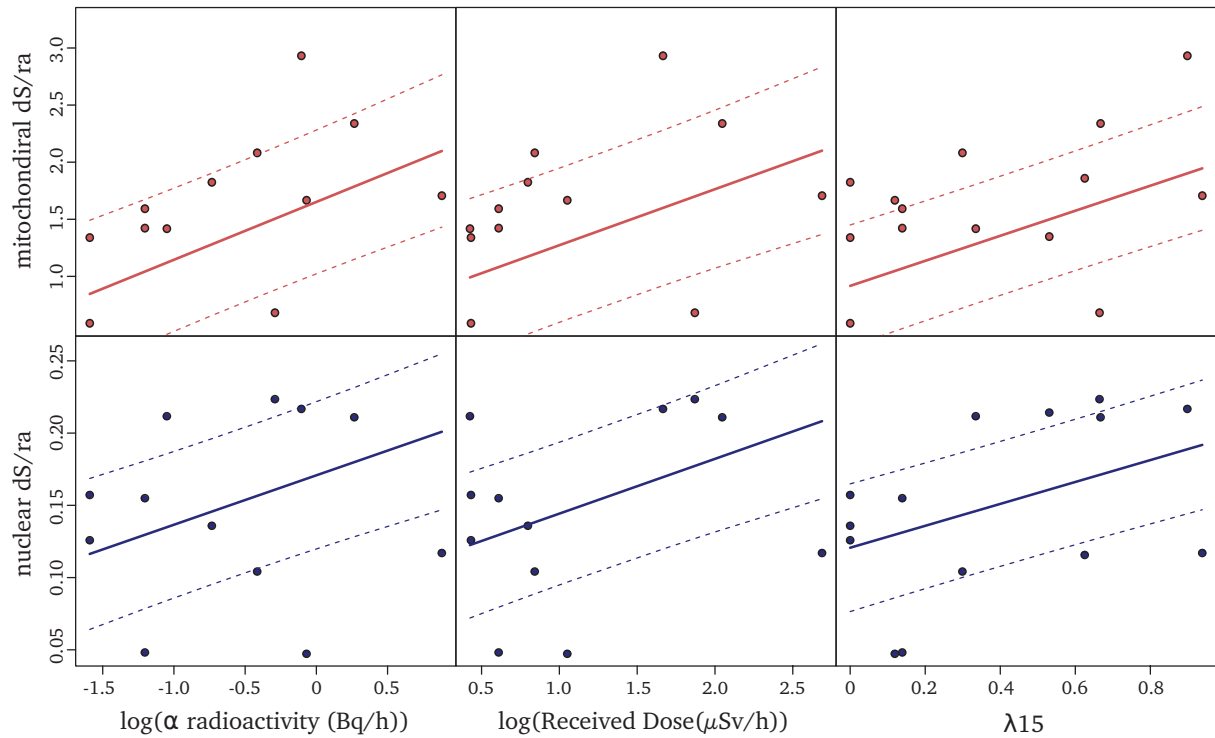


Figure 2: Relationships between synonymous substitution rate relative to the root age ( $d_S/ra$ ) and radioactivity measured either as the alpha radioactivity measured in sediments (left), the received dose (middle) or the proportion of igneous and metamorphic rock in a radius of 15 km around the sampling sites,  $\lambda_{15}$  (right). Each dot represents a species.  $d_S$  measured on mitochondrial genes are depicted in red and  $d_S$  measured on nuclear genes in blue. The fit of the PGLS model is indicated with a line and the confidence interval of the correlation is indicated with dashed lines.

105 the average linear distribution range of a groundwater crustacean is 30 km (Eme et al., 2018). We found a positive and stronger correlation between the  $d_S$  and  $\lambda_{15}$  in both genomes (Table 1, Figure 2,  $n=14$  species). The linear model predicts that the nuclear and mitochondrial  $d_S$  of a species living in a metamorphic formation ( $>50\%$  of metamorphic and igneous rocks) are on average 34.4% and 61.3% higher, respectively, than those of a species living in a sedimentary  
 110 formation ( $<50\%$  of metamorphic or igneous rocks).

As bedrock radioactivity is positively correlated with the mutation rate, the underlying question is whether radioactivity also modifies the mutational spectrum, that is, the specific types of mutations that tend to occur. To address this question, we reconstructed the mutational spectrum of 6 independent pairs of species, each composed of two species located in low and high  
 115 bedrock radiation set-ups, respectively, with a minimum of 3X increase in the received dose of radioactivity between the two species (Figure 1). Briefly, we first estimated species polymorphism across a set of 2490 one-to-one orthologous genes by sequencing transcriptomes for eight individuals per species. After ancestral sequence reconstruction, we then identified mutations that occurred in each species and computed the relative proportion of each type of mutation  
 120 (from A to T, A to C, ...), pooling together complementary mutations (e.g.  $p(C \rightarrow A) + p(G \rightarrow T) = p(C:G \rightarrow A:T)$ ). We found that bedrock radioactivity was correlated with three types of mutations (Table 2, Figure 3): while the proportions of A:T  $\rightarrow$  T:A and A:T  $\rightarrow$  G:C mutations decreased with increasing radioactivity, the proportion of C:G  $\rightarrow$  A:T mutations increased with bedrock radioactivity, regardless of the radioactivity characterisation (PGLS,  $p < 0.05$ , Table 2). Selection  
 125 is unlikely to be responsible for all these correlations as they are also observed when the data



set is limited to mutations found at the third, usually redundant, codon position (Figure S2, Table S2). While we found no explanation in the literature for the decrease of A:T→T:A and A:T→G:C mutations, C:G→A:T mutations, and in particular the G to T mutation triggered by the formation of 8-oxoguanine, are a hallmark of oxidative damage (Shibutani et al., 1991).  
 130 Similarly, high artificial doses of ionizing radiation were found to increase oxidative damage (Einor et al., 2016; Haghdoost et al., 2006). We thus propose that bedrock radiation impacts the rate and spectrum of mutation of in naturae populations through the formation of reactive oxygen species (ROS, unstable molecules that contain oxygen) generated by the radiolysis of water in cells (Riley, 1994; Azzam et al., 2012).

Table 2: Phylogenetic Generalized Least Square (PGLS) regressions of the proportion of each type of mutation against the  $\alpha$  radioactivity measured in sediment ( $\alpha$  radio.), the Received Dose (RD) modeled with ERICA tool, and the areal proportion of metamorphic and igneous rock within a 15 km radius ( $\lambda 15$ ).  $\alpha$  radioactivity and RD were log transformed to fit with linear model assumptions.  $R^2$  are Cox-Snell pseudo  $R^2$ .

Response variable	Explanatory variable	Slope	L.Ratio	P.value	$R^2$	N
P(A:T>T:A)	log( $\alpha$ radio)	-0.009	6.8187	0.009	0.433	12
	log(RD)	-0.011	9.3479	0.002	0.541	12
	$\lambda 15$	-0.025	6.994	0.008	0.442	14
P(A:T>C:G)	log( $\alpha$ radio)	0.001	0.121	0.728	0.010	12
	log(RD)	0.000	0.015	0.904	0.001	12
	$\lambda 15$	0.001	0.005	0.945	0.000	14
P(A:T>G:C)	log( $\alpha$ radio)	-0.022	7.778	0.005	0.477	12
	log(RD)	-0.019	4.004	0.045	0.284	12
	$\lambda 15$	-0.014	0.394	0.530	0.032	14
P(C:G>G:C)	log( $\alpha$ radio)	0.005	2.137	0.144	0.163	12
	log(RD)	0.006	2.403	0.121	0.181	12
	$\lambda 15$	0.006	0.360	0.549	0.030	14
P(C:G>A:T)	log( $\alpha$ radio)	0.013	13.010	0.000	0.662	12
	log(RD)	0.014	12.079	0.001	0.635	12
	$\lambda 15$	0.042	13.791	0.000	0.683	14
P(C:G>T:A)	log( $\alpha$ radio)	0.012	3.323	0.068	0.242	12
	log(RD)	0.010	1.684	0.194	0.131	12
	$\lambda 15$	-0.009	0.183	0.669	0.015	14

135 Radioactive environments can cause oxidative stress in two intertwined ways. First, the ionisation of molecules in the cells can directly affect the DNA structure by breaking sugar phosphate backbones or can affect DNA indirectly through the radiolysis of water which decomposes the  $H_2O$  molecules and create free radicals (Desouky et al., 2015). These free radicals can damage DNA molecules and create mutations. Second, radioactive decay chains also generate heavy  
 140 metals (lead, polonium, etc) which are toxic for cells and also cause oxidative stress (Quinlan et al., 1988; Pinto et al., 2003). Due to the physicochemical association between radioactivity and heavy metals, *in naturae* experiments cannot discriminate heavy metal chemical toxicity from the direct toxicity of radioactive rays. Therefore, this suggests that naturally radioactive environments impact the rate and spectrum of mutation by directly or indirectly increasing the  
 145 level of oxidative stress.

Radiolysis of water alone cannot explain the much higher impact of radioactivity on the mitochondrial mutation rate compared to the nuclear rate. Since there is no reason to argue that radiolysis would not evenly occur within cells, it should impact both genomes similarly. However, differences between the two genomes may explain why the mitochondrial genome is more  
 150 sensitive to radioactivity. First, the mitochondrial genome lacks some repair systems. For example, in a different yet analogous context, UV damage accumulates in the mitochondrial genome

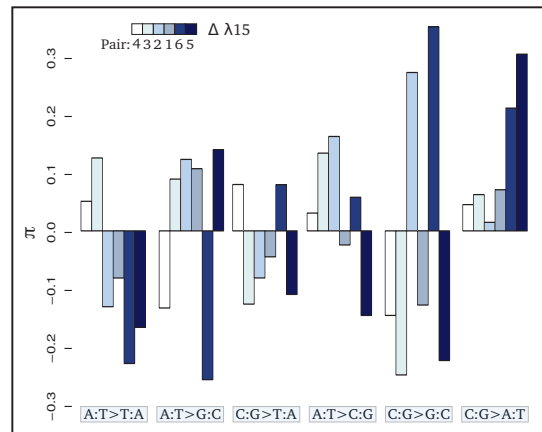


Figure 3: Contrasts ( $\pi$ ) of the relative proportion of each mutation [ $p(i:j \rightarrow k:l)$ ] in each pair of sister species :  $\pi_m = \log \frac{p(i:j \rightarrow k:l)^+}{p(i:j \rightarrow k:l)^-}$  where + and - refer respectively to the species exposed to the higher and lower level of radioactivity in the pair m. Thus, positive bars represent a higher proportion of the given mutation in the species living in the high radioactivity rock. From left to right, bars are in increasing order of difference ( $\Delta$ ) in  $\lambda_{15}$  (the areal proportion of igneous and metamorphic rock in a radius of 15 km around the site) between the two species of each pair. From left to right, mutations are in increasing order of correlation with radioactivity. Numbers below the color scale indicate the species pair number as in Figure 1.

while they are repaired in the nuclear genome (Clayton et al., 1974). Second, the two genomes have very different organisations: while the nuclear genome is compacted into chromatin, the mitochondrial one is organized into nucleoids (Chen and Butow, 2005). While the role of the mitochondrial nucleoids is unclear, the chromatin structure and histone proteins protect the nuclear genome against radiation-induced strand breaks (Ljungman, 1991) and oxidative damage (Ljungman and Hanawalt, 1992). Third, direct radioactivity damage to the mitochondria may increase the activity of the mitochondrial respiratory chain and indirectly increase the production of ROS and DNA damage (Yamamori et al., 2012; Kam and Banati, 2013).

While life history traits are known to be central in controlling the mutation rate in metazoans (Nabholz et al., 2008; Martin and Palumbi, 1993), we show here that natural variation of radioactivity can have a comparable effect. Indeed, we found a minimum increase of around 30% percent of the nuclear mutation rate (60% in mitochondria) for species of waterlice living in the more naturally radioactive habitats made of igneous and metamorphic rocks. This increase is of the same magnitude as that observed when waterlice species evolve a 5-fold increase in generation time, a key life history trait controlling mutation rate in waterlice and metazoans in general (Saclier et al., 2018). As groundwater waterlice ingest sediment (Francois et al., 2016), they are internally exposed to radioactivity, which may cause more mutations than through external exposure only (Sawada, 2007). The influence of environmental radioactivity on mutation rate should therefore be explored across a wider range of organisms with contrasted diets.

Natural radioactivity is often considered to have a negligible biological impact (Tubiana et al., 2006, 2009). Indeed, only a handful of isolated studies support an impact of natural radioactivity on the mutation rate. For instance, a higher mutation rate was observed in the human mitochondrial genome in the Kerala region (Forster et al., 2002) and in satellite sequences of crickets inhabiting cave with high radon concentration (Allegrucci et al., 2015). In this study, by combining a large number of genes with the characteristics of the subterranean waterlice,

namely the absence of UV confounding effect and limited dispersal, within a statistically powerful comparative framework allowing to work on large time scales and with numerous replicates, we found that a mild variation ( $\approx 3.5X$ ) in natural bedrock radioactivity substantially alters the mutation rate, in particular the mitochondrial one. While the universality of this finding warrants corroborative studies in other taxa, it suggests that the influence of natural radioactivity on the evolution of biodiversity may have been overlooked.

## Methods

### Sampling

For 58 sites in France selected on the map of uranium (Ielsch et al., 2017), we collected Asellidae species and sampled about 50 grams of sediment to measure global  $\alpha$  radioactivity (see the following paragraph). Animals and sediments were collected using the Bou-Rouch pumping methods (Bou and Rouch, 1967). Collected species were stored in 96% ethanol at  $-20^{\circ}\text{C}$  and were morphologically and molecularly identified. For molecular identification, DNA was extracted using an optimized chloroform DNA extraction protocol for the Aselloidea (Calvignac et al., 2011). We amplified DNA with primers targeting the 16S mitochondrial rDNA gene. PCR reactions were done following Morvan et al. (2013). PCR products were sequenced in both directions using the same primers as for amplification (GATC Biotech, Konstanz; Eurofins MWG Operon, Ebersberg; SeqLab, Göttingen, Germany; BIOFIDAL, Vaulx-en-Velin, France). Chromatograms were visualized and cleaned using Finch v1.5.0 (Geospiza, Seattle, USA). 16S have been deposited on the European Nucleotide Archive and are available under the accession number from LR214526 to LR214880. Using Eme et al. (2018) molecular species delimitation, each sequence has been assigned to a species. Based on this taxonomic assignment and radioactivity measurement, 14 species were retained for further analyses (Table S1). For these 14 selected species, during a new sampling trip, individuals were flash frozen alive in the field.

### Measures of radioactivity

**$\alpha$  radioactivity.** In order to estimate the global radioactivity in sediments, we measured the  $\alpha$  radioactivity. An  $\alpha$  decay occurs when an atom disintegrates by ejecting an  $\alpha$  particle, *i.e.* a particle made of two neutrons and two protons. The  $\alpha$  radioactivity should be correlated with the global radioactivity in natural systems. For the 58 prospected sites, 3 samples of about 50 grams of sediment were collected in polyethylene bottles.  $\alpha$  radioactivity measurements were made by the LABRADOR service (Institut de Physique Nucléaire de Lyon, France) on proportional counter with the NF ISO 18589-6 standard (Data available on Zenodo, (DOI: 10.5281/zenodo.3356835)).

**Received dose.** In order to estimate the received dose of radiation that is impacting organisms, we collected three samples of 100g of fine sediments ( $<100\mu\text{m}$ ) in each of the selected sites. These sediments were prepared with the NF EN ISO 18589-2 standard and measured by gamma spectrometry in conformity with the NF EN ISO 18589-3 standard using the PRISNA-P analysis platform at the Centre d'Etude Nucléaire de Bordeaux Gradignan (CENBG). This platform is certified by the French Nuclear Security Authority (ASN) for measures of natural radioactivity. Samples were dried in open air, and then dried at  $100^{\circ}\text{C}$ . Matters were packed in a waterproof geometry. Geometries were sealed for one month and then counted for a duration of 86500 seconds on the same chain of measure. The chain used is an ORTEC chain, presenting an

220 efficiency of about 60% and calibrated in May 2016. This chain is equipped with a cosmic veto  
device and located in a half buried laboratory in order to: (i) attenuate the background noise,  
(ii) improve the detection limits and (iii) reduce the measure uncertainty. The activity of the  
main radionuclides were measured in sediment and the activity of the remaining radionuclides  
was deduced based on the hypothesis of a secular equilibrium of the uranium 238 and thorium  
235 chains. As activities of the radionuclides of the uranium 235 decay chain are generally  
low, only measures higher than the decision threshold (according the measure variability) were  
taken into account. When the uranium 235 activity was too low to be measured it has been  
deduced from the uranium 238 activity, using the natural isotopic ratio of 21.6.

The received dose impacting organisms was estimated using the ERICA tool (V1.2.1, Brown  
et al., 2016) with a ‘crustacean’ model. We assumed that organisms stay 10% of their time  
230 on the surface of sediment and 90% inside sediment. All radionuclides available in the tool  
were taken into account (*i.e.*  $U^{238}$ ,  $Th^{234}$ ,  $U^{234}$ ,  $Th^{230}$ ,  $Ra^{226}$ ,  $Pb^{210}$ ,  $Po^{210}$ ,  $U^{235}$ ,  $Th^{231}$ ,  $Pa^{231}$ ,  
 $Th^{227}$ ,  $Th^{232}$ ,  $Ra^{228}$  and  $Th^{228}$ ). We used the distribution coefficients proposed by the ERICA  
tool. Concentration factors proposed by the tool were used when available. If not, we used the  
concentration factor of the closest biogeochemical element available.

235 Two sites (BRETEMIN and BOREON) show a disruption of the secular equilibrium in the  
 $U^{238}$  chain. This suggests that nearby industrial activities (*e.g.* lead mines) have modified the  
natural radioactivity of these two sites. As these industrial activities are very recent (since 1950),  
their impact on the substitution rate which is measured on a much longer time scale is unlikely.  
These two sites were removed from the correlation between  $d_S$  and radioactivity measured with  
240 the global  $\alpha$  radioactivity or with the received dose.

**Proportion of magmatic and igneous rocks in a 15 km radius.** Usign the geological map of  
France (scale : 1/1 000 000, ©BRGM), the areal proportions of magmatic and igneous rocks  
in a radius of 15 km around sampling sites were computed (noted  $\lambda_{15}$ ), 30 km represent the  
distribution range for a subterranean isopod (Eme et al., 2018).

## 245 **Transcriptome Sequencing and Assembly**

**Sequencing.** For each species, we sequenced transcriptomes from 8 individuals. For each in-  
dividual total RNA was isolated using TRI Reagent (Molecular Research Center). Extraction  
quality was checked on a BioAnalyser RNA chip (Agilent Technologies) and RNA concentrations  
were estimated using a Qubit fluorometer (Life Technologies). Prior to any additional analysis,  
250 species identification was corroborated for each individual by sequencing a fragment of the 16S  
gene. Illumina libraries were then prepared using the TruSeq™ RNA Sample Prep Kit v2 (Illu-  
mina). For each species one library was paired-end sequenced using 100 cycles, and the 7 other  
libraries were single-end sequenced using 50 cycles on a HiSeq2500 sequencer (Illumina) at the  
IGBMC GenomEast Platform (Illkirch, France). We obtained around 30 million single-end reads  
255 per individual and 118 million paired-end reads per species.

**Assembly.** Adapters were clipped from the sequences, low quality read ends were trimmed  
(phred score < 30) and low quality reads were discarded (mean phred score < 25 or if remain-  
ing length < 19 bp) using fastq-mcf of the ea-utils package (Aronesty, 2013). Paired-end tran-  
scriptomes were *de novo* assembled using Trinity v2.3.2 (Grabherr et al., 2011). Open reading  
260 frames (ORFs) were identified with TransDecoder (<http://transdecoder.sourceforge.net>). For  
each assembled component, only the most express ORF was retained.

## Families of Orthologous Genes

Gene families were delimited using an all-against-all BLASTP (Altschul et al., 1990) and SiLix (Miele et al., 2011) on the ORFs delimited in the previous step. We then kept gene families containing the 14 species, with only one sequence for each species in order to remove paralogs. We obtained 2490 families hereafter considered as one-to-one orthologous genes. These genes were aligned with PRANK (Löytynoja and Goldman, 2008) using a codon model and sites ambiguously aligned were removed with Gblocks (Castresana, 2000).

## Species Tree and Gene Trees

The 2490 genes were concatenated and a phylogenetic tree (hereafter called the concatenated tree) was built using PhyML v3.0 (Guindon et al., 2010) under a GTR+G+I model with 100 bootstrap replicates and was rooted using the Slavus lineage (*Proasellus boui* and *Proasellus slavus*) as an outgroup (Morvan et al., 2013). Most nodes have a bootstrap value of 100% (Figure S1). Two nodes have values at 84 and 98% in the clade containing *P. nsp VIELVIC*; *P. nsp HYPOPRAT*; *P. nsp MONTBAR* and *P. nsp ROSSFELD*. To check the relationship between these 4 species, we built 2490 individual gene trees with PhyML v3.0 under a GTR+G+I model with 100 bootstrap replicates. 29 gene trees strongly support (bootstraps > 90%) the phylogeny of the concatenated tree for this clade, 208 support other various topologies and the remaining 2253 gene trees do not support any relationship in particular for this clade (bootstraps < 90%). Thus, the phylogeny for this clade remained unresolved, possibly as the consequence of a concomitant speciation process of these 4 species. For approaches with pairs of sister species, as we were unable to resolve the phylogeny for this clade, we selected the species living in the highest level of radioactivity (*P. nsp HYPOPRAT*) and the species living in the lowest level of radioactivity (*P. nsp MONTBAR*) among these 4 species to build a pair, resulting in a total of 6 pairs of sister species (*sensu* Felsenstein, 2004).

## Mitochondrial genes

Mitochondrial genes were not present amongst the 2490 genes obtained above. Indeed, owing to a different genetic code in invertebrate mitochondria, mitochondrial ORFs were systematically missed by the ORF caller (Transdecoder). We reconstructed mitochondrial genomes using the *de novo* transcriptome assemblies. Large mitochondrial contigs were built with MITObim (Hahn et al., 2013) by using RNA-seq reads. These contigs were mapped on the assembled mitochondrial genome from the closest possible species (taken from Saclier et al., 2018), allowing us to assemble them. Mitochondrial genomes were annotated using the MITOS web server (Bernt et al., 2013). We recovered the 13 mitochondrial protein-coding genes. Mitochondrial genes were aligned with PRANK (Löytynoja and Goldman, 2008) and sites ambiguously aligned were removed with Gblocks (Castresana, 2000).

## Rate of molecular evolution

We used the synonymous substitution rate ( $d_S$ ) computed on the terminal branches of the tree as a proxy for the long-term species mutation rate (Kimura, 1983). This proxy is valid in absence of selection on codon usage. To check for the absence of biased codon usage, we computed the effective number of codons on the 2490 orthologous genes (ENC, Wright, 1990). This number varies between 20 (only a single codon is used for each amino acid) and 61 (all synonymous codons are used with equal frequency for each amino-acid). ENC ranged between 49.17 and

50.48 (Table S1), indicating a moderate codon usage bias, more importantly, they do not correlate with alpha radioactivity (PGLS, p.value = 0.6378). Altogether, the  $d_S$  estimation doesn't seem impacted by a strongly biased or variable codon usage.

To compute  $d_S$  we first removed some genes showing a conflicting phylogeny. Including genes supporting different phylogenies in a concatenation amounts to constrain a wrong phylogeny for these genes which may bias  $d_S$  estimations. Indeed imposing a wrong gene tree will tend to generate convergent mutations in terminal branches of the tree. To avoid such bias in our  $d_S$  estimation we used ProfileNJ (Noutahi et al., 2016) with a bootstrap threshold of 90% to compute a cost of reconciliation between the concatenated tree and the gene trees. We kept the gene families with a cost of reconciliation of zero and with sequences long enough for all species (at least a half of the alignment) and removed all other genes, resulting in a set of 769 gene families.  $d_S$  were estimated using CoEvol (Lartillot and Poujol, 2011). This software program implements a Muse and Gaut codon model (Muse and Gaut, 1994), with Brownian variation in  $d_S$  and  $d_N/d_S$  along the tree. Bayesian inference and reconstruction of the history of variation in  $d_S$  and  $d_N/d_S$  along the tree is conducted by Markov Chain Monte Carlo (MCMC). Two independent chains were run, and were stopped after checking for convergence by eye and with the tracecomp program included in the Coevol package (effective sample size > 200 and discrepancy between chains < 0.3). Chains were stopped after 7117 generations (4200 generations excluded as burn-in). The age of the root was arbitrarily set to 1, resulting in synonymous substitution rate estimates that are relative to the root age ( $d_S/ra$ ) (Table S1). In order to ensure that assumptions made by CoEvol on the  $d_S$  evolution along branches don't bias the  $d_S$  estimation,  $d_S$  were also computed with CodeML (Yang, 2007) using a free ratio model and with the Bio++ suite (Dutheil and Boussau, 2008). For the last one, a non-homogeneous model (NY98 model) was first applied to the alignment with BppML and then the MapNH program (Version 1.1.1) of the TestNH package (Guéguen and Duret, 2018) was used to reconstruct the ancestral states to estimate the number of synonymous substitutions on each branch. CoEvol  $d_S$  was highly correlated with CodeML  $d_S$  ( $R^2=0.81$ ) as well as with Mapnh  $d_S$  ( $R^2=0.82$ ). Regarding the correlation with radioactivity, by dividing the CodeML  $d_S$  and the Mapnh  $d_S$  by the divergence time estimated by CoEvol in order to obtain comparable  $d_S$  among all species, we obtained similar results whatever the method used to compute the  $d_S$  (Table S3).

## Mutational spectrum

To compute the mutational spectrum, we used an approach by pairs of sister species. We determined the polymorphism at the population level for each species by mapping the 7 single-end transcriptomes on the assembled paired-end transcriptome with BWA (Li and Durbin, 2009). BAM files were produced with SAMtools (Li et al., 2009), and reads2snps (Gayral et al., 2013) was used to detect polymorphic sites. We then conserved only the 2490 orthologous genes shared by all species to compute the mutational spectrum on the same set of genes.

For the two species of a pair, we reconstructed the ancestral sequence using a parsimonious approach. Namely, for each site in the alignment, if the two species had a single shared allele, this allele was considered as ancestral and the other alleles, if they existed, were considered as derived from the ancestral allele. For each species, we estimated the probability of a mutation in their population,  $p(i \rightarrow j | f(i)_{anc})$ , by counting each type of mutation, either on all positions (Table S8) or on third positions, corrected by the ancestral base frequency:

$$p(i \rightarrow j | f(i)_{anc}) = \frac{N(i \rightarrow j)}{N i_{anc}}$$

This probability being dependent on the mutation rate  $\mu$ , we estimated the mutational spectrum by the proportion, when a mutation occurs, of mutation from the base  $i$  to the base  $j$ , noted  $p(i \rightarrow j | \mu, f(i)_{anc})$ :

$$p(i \rightarrow j | \mu, f(i)_{anc}) = p(i \rightarrow j | f(i)_{anc}) * \frac{1}{\sum_{i=\{A,C,G,T\}} \sum_{j=\{A,T,C,G\}} p(i \rightarrow j | f(i)_{anc})}$$

This proportion takes into account the mutation rate and is so comparable across species. We pooled complementary mutations (e.g. A to C with T to G) to increase the counts by mutational categories and improve statistical power.

## 355 Statistical analyses

Correlations between  $d_S$  computed on terminal branch of the tree and the different measures of radioactivity were tested using Phylogenetic Generalised Least Squares models (pGLS Martins and Hansen, 1997) with the nlme (Pinheiro et al., 2007) and ape packages (Paradis et al., 2004) in R (R Core Team, 2014). For the  $\alpha$  radioactivity and effective dose, the two species showing  
360 a disruption in the secular equilibrium were removed. The ultrametric tree built by CoEvol was used to calculate the phylogenetic variance/covariance matrix under a Brownian motion model to take into account the non-independence among species in the PGLS. The same statistical procedure was used to test the correlation between the probability of each type of mutation and the effective dose of radioactivity or proportion of metamorphic or igneous rocks, this time  
365 pruning *P. nsp* VIELVIC and *P. nsp* MONTBAR from the chronogram built by CoEvol. Normality of residuals was checked for all models, log transformation was applied when the normality was rejected (Shapiro test).

## Data availability

16S have been deposited on the European Nucleotide Archive and are available under the accession numbers from LR214526 to LR214880 (<https://www.ebi.ac.uk/ena/data/view/LR214526-LR214880>).  
370

The 2490 alignments and the list of the 769 genes used to compute synonymous substitution rate have been deposited on Zenodo (<https://zenodo.org/deposit/2563829>).

Transcriptome reads have been deposited on the European Nucleotide Archive and are  
375 available under accession numbers from LR536601 to LR536626 in the study ID PRJEB14193 (<https://www.ebi.ac.uk/ena/data/search?query=PRJEB14193>). Number of reads and data used for correlations, namely measures of radionuclides and mutations counts have been deposited on Zenodo (<https://doi.org/10.5281/zenodo.3356835>).

## References

- 380 **Allegrucci G**, Sbordoni V, Cesaroni D. Is Radon Emission in Caves Causing Deletions in Satellite DNA Sequences of Cave-Dwelling Crickets? PLOS ONE. 2015 Mar; 10(3):e0122456. <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0122456>, 10.1371/journal.pone.0122456doi: 10.1371/journal.pone.0122456.
- Altschul SF**, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. 385 Journal of Molecular Biology. 1990 Oct; 215(3):403–410. <http://www.sciencedirect.com/science/article/pii/S0022283605803602>, 10.1016/S0022-2836(05)80360-2doi: 10.1016/S0022-2836(05)80360-2.
- Aronesty E**. Comparison of Sequencing Utility Programs. The Open Bioinformatics Journal. 2013 Jan; 7(1). <https://benthamopen.com/ABSTRACT/TOBIOIJ-7-1>.

- 390 **Azzam EI**, Jay-Gerin JP, Pain D. Ionizing radiation-induced metabolic oxidative stress and prolonged cell injury. *Cancer Letters*. 2012 Dec; 327(1):48–60. <http://www.sciencedirect.com/science/article/pii/S0304383511007592>, 10.1016/j.canlet.2011.12.012doi: 10.1016/j.canlet.2011.12.012.
- Bernt M**, Donath A, Jühling F, Externbrink F, Florentz C, Fritsch G, Pütz J, Midendorff M, Stadler PF. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*. 2013 Nov; 69(2):313–319. <http://www.sciencedirect.com/science/article/pii/S1055790312003326>, 10.1016/j.ympev.2012.08.023doi: 10.1016/j.ympev.2012.08.023.
- 400 **Bou C**, Rouch R. Un nouveau champ de recherches sur la faune aquatique souterraine. . 1967; [https://scholar.google.fr/scholar?hl=fr&as\\_sdt=0%2C5&q=Bou%2C++C.%2C++and++Rouch%2C++R.++1967.++Un++nouveau++champ++de++recherchessur++la++faune++aquatique++souterraine.++C.R.++Hebd.++S%3%A9ances++Acad.Sci.+Ser.+III.+Sci.+Vie%2C265%3A+369%E2%80%9393370&btnG=](https://scholar.google.fr/scholar?hl=fr&as_sdt=0%2C5&q=Bou%2C++C.%2C++and++Rouch%2C++R.++1967.++Un++nouveau++champ++de++recherchessur++la++faune++aquatique++souterraine.++C.R.++Hebd.++S%3%A9ances++Acad.Sci.+Ser.+III.+Sci.+Vie%2C265%3A+369%E2%80%9393370&btnG=).
- 405 **Brown JE**, Alfonso B, Avila R, Beresford NA, Copplestone D, Hosseini A. A new version of the ERICA tool to facilitate impact assessments of radioactivity on wild plants and animals. *Journal of Environmental Radioactivity*. 2016 Mar; 153:141–148. <http://www.sciencedirect.com/science/article/pii/S0265931X15301715>, 10.1016/j.jenvrad.2015.12.011doi: 10.1016/j.jenvrad.2015.12.011.
- 410 **Calvignac S**, Konecny L, Malard F, Douady CJ. Preventing the pollution of mitochondrial datasets with nuclear mitochondrial paralogs (numts). *Mitochondrion*. 2011 Mar; 11(2):246–254. <http://www.sciencedirect.com/science/article/pii/S1567724910001881>, 10.1016/j.mito.2010.10.004doi: 10.1016/j.mito.2010.10.004.
- 415 **Castresana J**. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution*. 2000 Apr; 17(4):540–552. <https://academic.oup.com/mbe/article/17/4/540/1127654>, 10.1093/oxfordjournals.molbev.a026334doi: 10.1093/oxfordjournals.molbev.a026334.
- Chen XJ**, Butow RA. The organization and inheritance of the mitochondrial genome. *Nature Reviews Genetics*. 2005 Nov; 6(11):815–825. <https://www.nature.com/articles/nrg1708>, 10.1038/nrg1708doi: 10.1038/nrg1708, number: 11 Publisher: Nature Publishing Group.
- 420 **Clayton DA**, Doda JN, Friedberg EC. The Absence of a Pyrimidine Dimer Repair Mechanism in Mammalian Mitochondria. *Proceedings of the National Academy of Sciences*. 1974 Jan; 71(7):2777–2781. <http://www.pnas.org/doi/10.1073/pnas.71.7.2777>.
- 425 **Czeizel AE**, Elek C, Susánszky E. The evaluation of the germinal mutagenic impact of Chernobyl radiological contamination in Hungary. *Mutagenesis*. 1991 Jul; 6(4):285–288. <https://academic.oup.com/mutage/article/6/4/285/1548257>, 10.1093/mutage/6.4.285doi: 10.1093/mutage/6.4.285.
- Desouky O**, Ding N, Zhou G. Targeted and non-targeted effects of ionizing radiation. *Journal of Radiation Research and Applied Sciences*. 2015 Apr; 8(2):247–254. <https://doi.org/10.1016/j.jrras.2015.03.003>, 10.1016/j.jrras.2015.03.003doi: 10.1016/j.jrras.2015.03.003.
- 430 **Dubrova YE**, Nesterov VN, Krouchinsky NG, Ostapenko VA, Neumann R, Neil DL, Jeffreys AJ. Human minisatellite mutation rate after the Chernobyl accident. *Nature*. 1996 Apr; 380(6576):683–686. <http://www.nature.com/articles/380683a0>, 10.1038/380683a0doi: 10.1038/380683a0.



- Dutheil J**, Boussau B. Non-homogeneous models of sequence evolution in the Bio++ suite of libraries and programs. *BMC Evolutionary Biology*. 2008 Sep; 8(1):1. <http://bmcevolbiol.biomedcentral.com/doi/10.1186/1471-2148-8-255>doi: 10.1186/1471-2148-8-255.
- Einor D**, Bonisoli-Alquati A, Costantini D, Mousseau TA, Møller AP. Ionizing radiation, antioxidant response and oxidative damage: A meta-analysis. *Science of The Total Environment*. 2016 Apr; 548-549:463–471. <http://www.sciencedirect.com/science/article/pii/S0048969716300274>, 10.1016/j.scitotenv.2016.01.027doi: 10.1016/j.scitotenv.2016.01.027.
- Eme D**, Zgmaïster M, Delić T, Fišer C, Flot JF, Konecny-Dupré L, Pálsson S, Stoch F, Zakšek V, Douady CJ, et al. Do cryptic species matter in macroecology? Sequencing European ground-water crustaceans yields smaller ranges but does not challenge biodiversity determinants. *Ecography*. 2018; 41(2):424–436.
- Felsenstein J**. *Inferring Phylogenies*; 2004.
- Forster L**, Forster P, Lutz-Bonengel S, Willkomm H, Brinkmann B. Natural radioactivity and human mitochondrial DNA mutations. *Proceedings of the National Academy of Sciences*. 2002 Oct; 99(21):13950–13954. <http://www.pnas.org/content/99/21/13950>, 10.1073/pnas.202400499doi: 10.1073/pnas.202400499.
- Francois CM**, Mermillod-Blondin F, Malard F, Fourel F, Lécuyer C, Douady CJ, Simon L. Trophic ecology of groundwater species reveals specialization in a low-productivity environment. *Functional Ecology*. 2016 Feb; 30(2):262–273. <http://doi.wiley.com/10.1111/1365-2435.12484>, 10.1111/1365-2435.12484doi: 10.1111/1365-2435.12484.
- Gayral P**, Melo-Ferreira J, Glémin S, Bierne N, Carneiro M, Nabholz B, Lourenco JM, Alves PC, Ballenghien M, Faivre N, Belkhir K, Cahais V, Loire E, Bernard A, Galtier N. Reference-Free Population Genomics from Next-Generation Transcriptome Data and the Vertebrate–Invertebrate Gap. *PLOS Genetics*. 2013 Apr; 9(4):e1003457. <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1003457>, 10.1371/journal.pgen.1003457doi: 10.1371/journal.pgen.1003457.
- Grabherr MG**, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*. 2011 Jul; 29(7):644–652. <http://www.nature.com/articles/nbt.1883>, 10.1038/nbt.1883doi: 10.1038/nbt.1883.
- Guindon S**, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*. 2010 May; 59(3):307–321. <https://academic.oup.com/sysbio/article/59/3/307/1702850>, 10.1093/sysbio/syq010doi: 10.1093/sysbio/syq010.
- Guéguen L**, Duret L. Unbiased Estimate of Synonymous and Nonsynonymous Substitution Rates with Nonstationary Base Composition. *Molecular Biology and Evolution*. 2018 Mar; 35(3):734–742. <https://academic.oup.com/mbe/article/35/3/734/4705835>, 10.1093/molbev/msx308doi: 10.1093/molbev/msx308.
- Haghdooost S**, Sjölander L, Czene S, Harms-Ringdahl M. The nucleotide pool is a significant target for oxidative stress. *Free Radical Biology and Medicine*. 2006 Aug; 41(4):620–626. <http://www.sciencedirect.com/science/article/pii/S0891584906002954>, 10.1016/j.freeradbiomed.2006.05.003doi: 10.1016/j.freeradbiomed.2006.05.003.

- Hahn C**, Bachmann L, Chevreur B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research*. 2013 Jan; 41(13):e129–e129. <http://nar.oxfordjournals.org/doi/10.1093/nar/gkt371>doi: 10.1093/nar/gkt371.
- Ielsch G**, Cuney M, Buscail F, Rossi F, Leon A, Cushing ME. Estimation and mapping of uranium content of geological units in France. *Journal of Environmental Radioactivity*. 2017 Jan; 166:210–219. <http://www.sciencedirect.com/science/article/pii/S0265931X16301667>, 10.1016/j.jenvrad.2016.05.022doi: 10.1016/j.jenvrad.2016.05.022.
- Kam WWY**, Banati RB. Effects of ionizing radiation on mitochondria. *Free Radical Biology and Medicine*. 2013 Dec; 65:607–619. <http://www.sciencedirect.com/science/article/pii/S0891584913003687>, 10.1016/j.freeradbiomed.2013.07.024doi: 10.1016/j.freeradbiomed.2013.07.024.
- Karam PA**, Leslie SA. Changes in terrestrial natural radiation levels over the history of life. In: *Radioactivity in the Environment*, vol. 7 Elsevier; 2005.p. 107–117. [https://sci-hub.tw/10.1016/S1569-4860\(04\)07011-1](https://sci-hub.tw/10.1016/S1569-4860(04)07011-1), 10.1016/S1569-4860(04)07011-1doi: 10.1016/S1569-4860(04)07011-1.
- Kimura M**. *The Neutral Theory of Molecular Evolution*. Cambridge University Press; 1983. Google-Books-ID: olIoSumPevYC.
- Lartillot N**, Poujol R. A Phylogenetic Model for Investigating Correlated Evolution of Substitution Rates and Continuous Phenotypic Characters. *Molecular Biology and Evolution*. 2011 Jan; 28(1):729–744. <https://academic.oup.com/mbe/article/28/1/729/985883>, 10.1093/molbev/msq244doi: 10.1093/molbev/msq244.
- Li H**, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009 Jul; 25(14):1754–1760. <http://academic.oup.com/bioinformatics/article/25/14/1754/225615>, 10.1093/bioinformatics/btp324doi: 10.1093/bioinformatics/btp324.
- Li H**, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009 Aug; 25(16):2078–2079. <https://academic.oup.com/bioinformatics/article/25/16/2078/204688>, 10.1093/bioinformatics/btp352doi: 10.1093/bioinformatics/btp352.
- Ljungman M**. The Influence of Chromatin Structure on the Frequency of Radiation-Induced DNA Strand Breaks: A Study Using Nuclear and Nucleoid Monolayers. *Radiation Research*. 1991 Apr; 126(1):58–64. <https://www.rrjournal.org/doi/abs/10.2307/3578171>, 10.2307/3578171doi: 10.2307/3578171, publisher: The Radiation Research Society.
- Ljungman M**, Hanawalt PC. Efficient protection against oxidative DNA damage in chromatin. *Molecular Carcinogenesis*. 1992; 5(4):264–269. <http://onlinelibrary.wiley.com/doi/abs/10.1002/mc.2940050406>, 10.1002/mc.2940050406doi: 10.1002/mc.2940050406.
- Löytynoja A**, Goldman N. Phylogeny-Aware Gap Placement Prevents Errors in Sequence Alignment and Evolutionary Analysis. *Science*. 2008 Jun; 320(5883):1632–1635. <http://science.sciencemag.org/content/320/5883/1632>, 10.1126/science.1158395doi: 10.1126/science.1158395.
- Martin AP**, Palumbi SR. Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences*. 1993 May; 90(9):4087–4091. <http://www.pnas.org/content/90/9/4087>, 10.1073/pnas.90.9.4087doi: 10.1073/pnas.90.9.4087.

- Martins EP**, Hansen TF. Phylogenies and the Comparative Method: A General Approach to Incorporating Phylogenetic Information into the Analysis of Interspecific Data. *The American Naturalist*. 1997 Apr; 149(4):646–667. <https://www.journals.uchicago.edu/doi/abs/10.1086/286013>, 10.1086/286013doi: 10.1086/286013.
- Masoomi JR**, Mohammadi S, Amini M, Ghiassi-Nejad M. High background radiation areas of Ramsar in Iran: evaluation of DNA damage by alkaline single cell gel electrophoresis (SCGE). *Journal of Environmental Radioactivity*. 2006 Jan; 86(2):176–186. <http://www.sciencedirect.com/science/article/pii/S0265931X0500247X>, 10.1016/j.jenvrad.2005.08.005doi: 10.1016/j.jenvrad.2005.08.005.
- Miele V**, Penel S, Duret L. Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC Bioinformatics*. 2011 Apr; 12(1):116. <https://doi.org/10.1186/1471-2105-12-116>, 10.1186/1471-2105-12-116doi: 10.1186/1471-2105-12-116.
- Morvan C**, Malard F, Paradis E, Lefébure T, Konecny-Dupré L, Douady CJ. Timetree of Aselloidea Reveals Species Diversification Dynamics in Groundwater. *Systematic Biology*. 2013 Jan; 62(4):512–522. <http://sysbio.oxfordjournals.org.docelec.univ-lyon1.fr/content/62/4/512>, 10.1093/sysbio/syt015doi: 10.1093/sysbio/syt015.
- Muse SV**, Gaut BS. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Molecular Biology and Evolution*. 1994 Sep; 11(5):715–724. <https://academic.oup.com/mbe/article/11/5/715/1008710>, 10.1093/oxfordjournals.molbev.a040152doi: 10.1093/oxfordjournals.molbev.a040152.
- Møller AP**, Mousseau TA. The effects of natural variation in background radioactivity on humans, animals and other organisms. *Biological Reviews*. 2012 Nov; 88(1):226–254. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-185X.2012.00249.x>, 10.1111/j.1469-185X.2012.00249.xdoi: 10.1111/j.1469-185X.2012.00249.x.
- Nabholz B**, Glémin S, Galtier N. Strong Variations of Mitochondrial Mutation Rate across Mammals—the Longevity Hypothesis. *Molecular Biology and Evolution*. 2008 Jan; 25(1):120–130. <http://academic.oup.com/mbe/article/25/1/120/1097840>, 10.1093/molbev/msm248doi: 10.1093/molbev/msm248.
- Noutahi E**, Semeria M, Lafond M, Seguin J, Boussau B, Guéguen L, El-Mabrouk N, Tannier E. Efficient Gene Tree Correction Guided by Genome Evolution. *PLOS ONE*. 2016 Aug; 11(8):e0159559. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0159559>, 10.1371/journal.pone.0159559doi: 10.1371/journal.pone.0159559.
- Paradis E**, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*. 2004 Jan; 20(2):289–290. <http://bioinformatics.oxfordjournals.org.docelec.univ-lyon1.fr/content/20/2/289>, 10.1093/bioinformatics/btg412doi: 10.1093/bioinformatics/btg412.
- Pinheiro J**, Bates D, DebRoy S, Sarkar D, Team RC. Linear and nonlinear mixed effects models. R package version. 2007; 3(57):1–89.
- Pinto E**, Sigaud-kutner TC, Leitao MA, Okamoto OK, Morse D, Colepicolo P. Heavy metal-induced oxidative stress in algae 1. *Journal of phycology*. 2003; 39(6):1008–1018.
- Quinlan GJ**, Halliwell B, Moorhouse CP, Gutteridge JM. Action of lead (II) and aluminium (III) ions on iron-stimulated lipid peroxidation in liposomes, erythrocytes and rat liver microsomal fractions. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism*. 1988; 962(2):196–200.

**R Core Team; R Foundation for Statistical Computing.** R: A language and environment for statistical computing. . 2014; <https://www.R-project.org/>.

570 **Riley PA.** Free Radicals in Biology: Oxidative Stress and the Effects of Ionizing Radiation. *International Journal of Radiation Biology.* 1994 Jan; 65(1):27–33. <https://doi.org/10.1080/09553009414550041>, 10.1080/09553009414550041doi: 10.1080/09553009414550041.

575 **Saclier N,** François CM, Konecny-Dupré L, Lartillot N, Guéguen L, Duret L, Malard F, Douady CJ, Lefébure T. Life History Traits Impact the Nuclear Rate of Substitution but Not the Mitochondrial Rate in Isopods. *Molecular Biology and Evolution.* 2018 Dec; 35(12):2900–2912. <https://academic.oup.com/mbe/article/35/12/2900/5106169>, 10.1093/molbev/msy184doi: 10.1093/molbev/msy184.

580 **Saghirzadeh M,** Gharaati MR, Mohammadi S, Ghiassi-Nejad M. Evaluation of DNA damage in the root cells of *Allium cepa* seeds growing in soil of high background radiation areas of Ramsar – Iran. *Journal of Environmental Radioactivity.* 2008 Oct; 99(10):1698–1702. <http://www.sciencedirect.com/science/article/pii/S0265931X08000544>, 10.1016/j.jenvrad.2008.03.013doi: 10.1016/j.jenvrad.2008.03.013.

585 **Satoh C,** Takahashi N, Asakawa J, Kodaira M, Kuick R, Hanash S M, Neel J V. Genetic analysis of children of atomic bomb survivors. *Environmental Health Perspectives.* 1996 May; 104(suppl 3):511–519. <https://ehp.niehs.nih.gov/doi/abs/10.1289/ehp.96104s3511>, 10.1289/ehp.96104s3511doi: 10.1289/ehp.96104s3511.

**Sawada S.** Cover-up of the effects of internal exposure by residual radiation from the atomic bombing of Hiroshima and Nagasaki. *Medicine, Conflict and Survival.* 2007 Jan; 23(1):58–74. <http://www.tandfonline.com/doi/abs/10.1080/13623690601084617>, 10.1080/13623690601084617doi: 10.1080/13623690601084617.

590 **Shibutani S,** Takeshita M, Grollman AP. Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature.* 1991 Jan; 349(6308):431–434. <http://www.nature.com/articles/349431a0>, 10.1038/349431a0doi: 10.1038/349431a0.

595 **Tubiana M,** Aurengo A, Averbeck D, Masse R. The debate on the use of linear no threshold for assessing the effects of low doses. *Journal of Radiological Protection.* 2006 Sep; 26(3):317–324. <https://sci-hub.tw/10.1088/0952-4746/26/3/n01>, 10.1088/0952-4746/26/3/N01doi: 10.1088/0952-4746/26/3/N01.

600 **Tubiana M,** Feinendegen LE, Yang C, Kaminski JM. The Linear No-Threshold Relationship Is Inconsistent with Radiation Biologic and Experimental Data1. *Radiology.* 2009 Apr; <https://pubs.rsna.org/doi/abs/10.1148/radiol.2511080671>, 10.1148/radiol.2511080671doi: 10.1148/radiol.2511080671.

**Vilenchik MM,** Knudson AG. Inverse radiation dose-rate effects on somatic and germline mutations and DNA damage rates. *Proceedings of the National Academy of Sciences.* 2000 May; 97(10):5381–5386. <http://www.pnas.org/content/97/10/5381>, 10.1073/pnas.090099497doi: 10.1073/pnas.090099497.

605 **Wright F.** The ‘effective number of codons’ used in a gene. *Gene.* 1990 Mar; 87(1):23–29. <http://www.sciencedirect.com/science/article/pii/0378111990904919>, 10.1016/0378-1119(90)90491-9doi: 10.1016/0378-1119(90)90491-9.

610 **Yamamori T,** Yasui H, Yamazumi M, Wada Y, Nakamura Y, Nakamura H, Inanami O. Ionizing radiation induces mitochondrial reactive oxygen species production accompanied by upregulation of mitochondrial electron transport chain function and

mitochondrial content under control of the cell cycle checkpoint. *Free Radical Biology and Medicine*. 2012 Jul; 53(2):260–270. <http://www.sciencedirect.com/science/article/pii/S0891584912002468>, 10.1016/j.freeradbiomed.2012.04.033doi: 10.1016/j.freeradbiomed.2012.04.033.

615 **Yang Z.** PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*. 2007 Aug; 24(8):1586–1591. <http://mbe.oxfordjournals.org.docelec.univ-lyon1.fr/content/24/8/1586>, 10.1093/molbev/msm088doi: 10.1093/molbev/msm088.

620 **Ziegler A,** Leffell DJ, Kunala S, Sharma HW, Gailani M, Simon JA, Halperin AJ, Baden HP, Shapiro PE, Bale AE. Mutation hotspots due to sunlight in the p53 gene of nonmelanoma skin cancers. *Proceedings of the National Academy of Sciences*. 1993 Jan; 90(9):4216–4220. <http://www.pnas.org.docelec.univ-lyon1.fr/content/90/9/4216>, 10.1073/pnas.90.9.4216doi: 10.1073/pnas.90.9.4216.

## Acknowledgments

625 This work was supported by the French program STYGOMICS (CNRS Défi Enviromics), the Zone Atelier Territoire Uranifère, the Grottes d’Azé, the Conseil Départemental de Saône-et-Loire, the Association Culturelle du Site d’Azé and the Agence Nationale de la Recherche (ANR-15-CE32-0005 Convergénomix, France, ANR-17-EURE-0018 H2O’Lyon, France). We gratefully acknowledge support from the CNRS/IN2P3 Computing Center (Lyon/Villeurbanne, France)  
630 for providing a significant amount of the computing resources needed for this work. We thank the Grottes d’Azé and more specifically Lionel Barriquand, the Mercantour national parc (authorization n°2015-251) and more specifically its scientific manager Marie-France Leccia, the Fédération Rhône-Alpes de Protection de la Nature and the owner of Bout du monde mine for giving us access for sampling. We thank Marcel Meysonnier, Aymeric and Audric Berjoan,  
635 Josiane and Bernard Lips, Audrey Brechet, Claude Bou, Benjamin Benti and Léa Dantony for their help in the field. We are grateful to Nicolas Lartillot, Gilles Escarguel, Marie Sémon, Laurent Guéguen, Nicolas Galtier, Benoît Nabholz and Bastien Boussau for helpful discussions. We also thank Laurent Simon and Laura Grice for their suggestions in the latter stages of manuscript preparation.

## 640 Contributions

Douady C., Lefébure T., Chardon P. and Malard F. designed the study. Saclier N. and Lefébure T. wrote the manuscript. Douady C., Lefébure T., Chardon P., Malard F., Saclier N. and Eme D. did the field work (animal and sediment sampling). Malard F. and Eme D. dissected and identified morphologically all sampled species. Lara Konecny-Dupré extracted DNA and RNA  
645 from all samples, made PCR and migrations and prepared library for sequencing. Saclier N. built the phylogeny, computed substitution rate and performed statistical analyses. Duret L. and Saclier N. computed the mutational spectrum. Chardon P. and Breton V. performed all radioactivity measurements and the effective dose estimation. Bellec A. extracted the outcrop cover of low-radioactivity sedimentary rocks and high-radioactivity metamorphic and igneous  
650 rocks in a radius of 15 km around the sampling and made the map of uranium for the figure 1. All authors interpreted the data and contributed to the final manuscript.

## Supplementary information

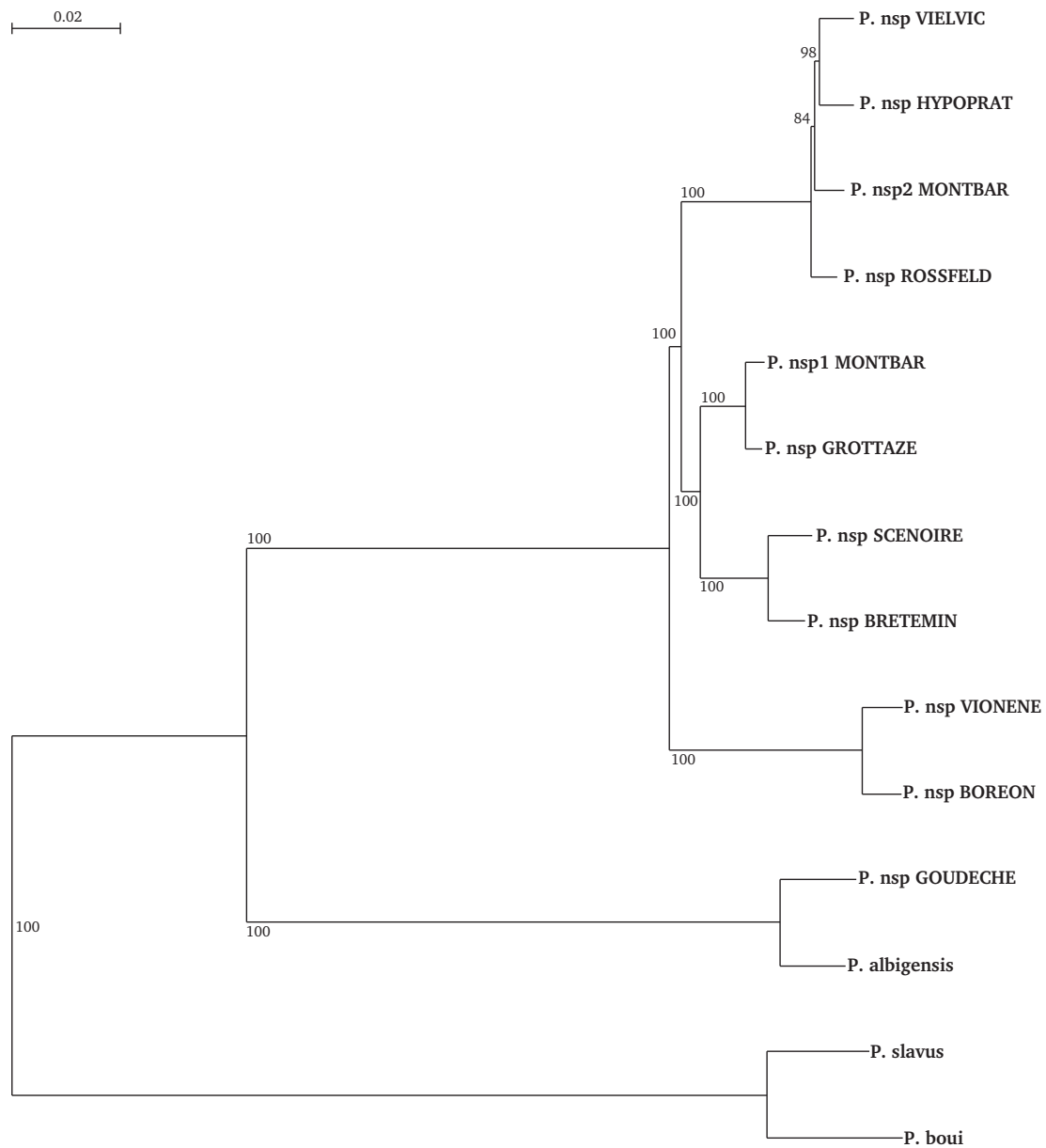


Figure S1: Phylogeny of the 14 species used to compute synonymous substitution rates. Tree was built with the concatenation of the 2490 genes with PhyML3.0, under a GTR+G+I model with 100 bootstrap replicates.

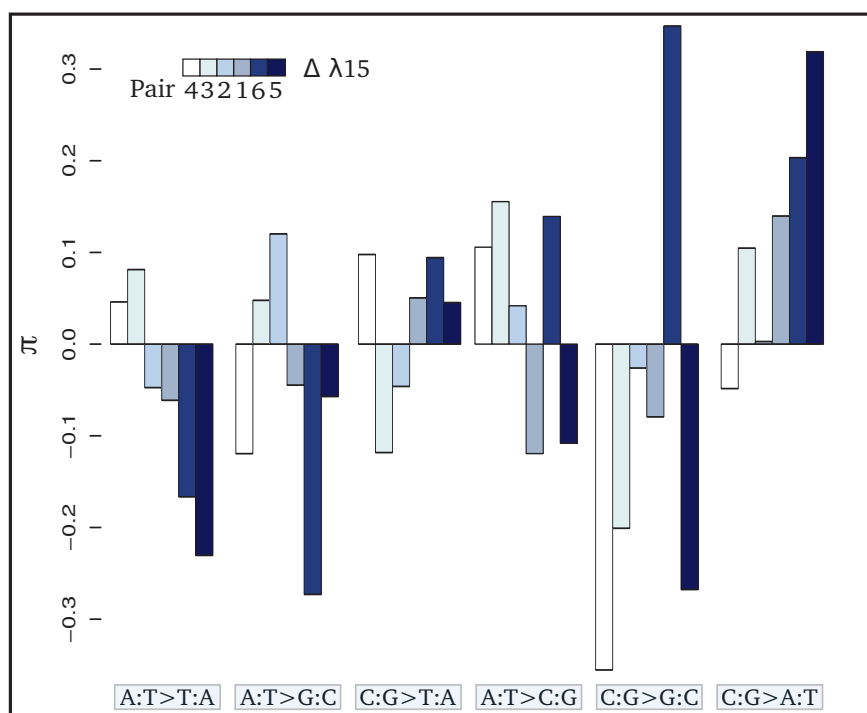


Figure S2: Contrasts ( $\pi$ ) of the relative proportion of each mutation [ $P(i : j \rightarrow k : l)$ ] computed only on third positions in each pair of sister species : [ $\pi_m = \log \frac{P(i:j \rightarrow k:l)_+}{P(i:j \rightarrow k:l)_-}$ ] where + and - refer respectively to the species exposed to the higher and lower level of radioactivity in the pair m. Thus, positive bars represent a higher proportion of the given mutation in the species living in the high radioactivity rock. From left to right, bars are in increasing order of difference ( $\Delta$ ) in  $\lambda_{15}$  (the areal proportion of igneous and metamorphic rock in a radius of 15 km around the site) between the two species of each pair. From left to right, mutations are in increasing order of correlation with radioactivity. Numbers below the color scale indicate the species pair number as in Figure 1.

Table S1: 14 sequenced species with sampling coordinates, synonymous substitution rate relative to the root age for nuclear and mitochondrial genome,  $\alpha$  radioactivity measured on each site, effective dose of radioactivity (in  $\mu\text{Gy/h}$ ), and this effective dose corrected for recent human impact (in  $\mu\text{Gy/h}$ ), non-synonymous substitution rate over synonymous substitution rate ( $d_N/d_S$ ), the areal proportion of magmatic and metamorphic rocks in a radius of 15 km around the sampling point ( $\lambda_{15}$ ), the GC content for all positions or for third positions, the effective number of codon (ENC) and the  $d_S$  computed with the CodeML programm and with the MapNH programm.

Species	RNA Code	$\alpha$ Radio.	lat.	long.	$d_S/\text{ra}$ nucl	$d_S/\text{ra}$ mito.	Dose	
<i>Proasellus nsp2</i> MONTBAR	PWM	0.300	47.01	5.62	0.1549	1.5924	1.84	
<i>Proasellus nsp</i> VIELVIC	PWVi	0.900	44.40	3.94	0.2167	2.9317	5.29	
<i>Proasellus nsp</i> ROSSFELD	PWRo	0.204	48.33	7.63	0.1258	1.3395	1.54	
<i>Proasellus nsp</i> HYPOPRAT	PWH	1.305	44.01	3.67	0.2109	2.3386	7.74	
<i>Proasellus nsp</i> SCENOIRE	PStN	0.480	47.32	6.43	0.1358	1.8241	2.22	
<i>Proasellus nsp</i> BRETEMIN	PStB	2.027	46.20	4.51	0.1156	1.8590	33.8	
<i>Proasellus nsp1</i> MONTBAR	PCMt	0.300	47.01	5.62	0.0481	1.4223	1.84	
<i>Proasellus nsp</i> GROTTAZE	PCG	0.933	46.43	4.76	0.0472	1.6664	2.86	
<i>Proasellus nsp</i> VIONENE	PspVi	0.350	44.08	7.10	0.2117	1.4174	1.53	
<i>Proasellus nsp</i> BOREON	ISPB	2.370	44.07	7.25	0.2142	1.3483	37.2	
<i>Proasellus alibigenis</i>	PAL	0.660	43.91	2.23	0.1042	2.0817	2.32	
<i>Proasellus nsp</i> GOUDECHE	PspAG	2.413	44.35	3.78	0.1169	1.7065	14.7	
<i>Proasellus slavus</i>	slavPWRo	0.204	48.33	7.63	0.1571	0.5893	1.54	
<i>Proasellus boui</i>	PBF	0.748	44.12	3.89	0.2235	0.6813	6.49	
Species	$d_N/d_S$	$\lambda_{15}$	GC	GC3	ENC	$d_S$ CodeML	$d_S$ MapNH	Pair
<i>Proasellus nsp2</i> MONTBAR	0.1526	0.1387	0.4041	0.3297	50.17	0.012413	0.01036720	NA
<i>Proasellus nsp</i> VIELVIC	0.1774	0.8985	0.4040	0.3294	50.18	0.014071	0.01324370	
<i>Proasellus nsp</i> ROSSFELD	0.1503	0.0000	0.4045	0.3301	50.22	0.011268	0.01105270	1
<i>Proasellus nsp</i> HYPOPRAT	0.1809	0.6672	0.4041	0.3295	50.17	0.013782	0.01289470	
<i>Proasellus nsp</i> SCENOIRE	0.1610	0.0000	0.4051	0.3318	50.35	0.018364	0.01817050	2
<i>Proasellus nsp</i> BRETEMIN	0.1534	0.6250	0.40508	0.3318	50.33	0.015513	0.01536320	
<i>Proasellus nsp1</i> MONTBAR	0.1970	0.1387	0.4055	0.3323	50.36	0.007196	0.00622081	3
<i>Proasellus nsp</i> GROTTAZE	0.1850	0.1190	0.4054	0.3321	50.34	0.006391	0.00606323	
<i>Proasellus nsp</i> VIONENE	0.1656	0.3346	0.4062	0.3347	50.48	0.016790	0.01658250	4
<i>Proasellus nsp</i> BOREON	0.1517	0.5306	0.4062	0.3347	50.46	0.017109	0.01674480	
<i>Proasellus alibigenis</i>	0.1369	0.2991	0.3985	0.3163	49.35	0.029319	0.02895040	5
<i>Proasellus nsp</i> GOUDECHE	0.1369	0.9383	0.3979	0.3149	49.27	0.033744	0.03296400	
<i>Proasellus slavus</i>	0.1672	0.0000	0.3969	0.3119	49.17	0.041241	0.04523370	6
<i>Proasellus boui</i>	0.1265	0.6645	0.3975	0.3128	49.19	0.061836	0.05802430	



Table S2: Phylogenetic Least Square (PGLS) regression of mutation probabilities computed on third positions against radioactivity measured as the  $\alpha$  radioactivity measured in sediments, as the effective dose received by organisms (RD) or as the areal proportion of metamorphic and magmatic rocks in a radius of 15 km around the sampled point ( $\lambda_{15}$ ). Each line corresponds to one likelihood ratio test between the models with and without the given explanatory variable.  $\alpha$  radioactivity and received dose were log transformed to fit with linear model assumptions.

Dependant variable	Explanatory variable	Slope	L. ratio	P. value	Pseudo R <sup>2</sup> (Cox-Snell)	N
P(A:T>T:A)	log( $\alpha$ radio.)	-0.007	4.681	0.031	0.323	12
	log(RD)	-0.008	5.902	0.015	0.388	12
	$\lambda_{15}$	-0.023	5.917	0.015	0.389	14
P(A:T>C:G)	log( $\alpha$ radio.)	0.003	0.423	0.515	0.035	12
	log(RD)	0.001	0.038	0.845	0.003	12
	$\lambda_{15}$	0.003	0.068	0.794	0.006	14
P(A:T>G:C)	log( $\alpha$ radio.)	-0.028	11.759	0.001	0.625	12
	log(RD)	-0.025	6.495	0.011	0.418	12
	$\lambda_{15}$	-0.048	5.697	0.017	0.378	14
P(C:G>G:C)	log( $\alpha$ radio.)	0.002	0.606	0.436	0.049	12
	log(RD)	0.003	0.977	0.323	0.078	12
	$\lambda_{15}$	0.003	0.240	0.624	0.020	14
P(C:G>A:T)	log( $\alpha$ radio.)	0.010	8.947	0.003	0.526	12
	log(RD)	0.011	10.561	0.001	0.585	12
	$\lambda_{15}$	0.040	15.208	0.000	0.718	14
P(C:G>T:A)	log( $\alpha$ radio.)	0.019	8.705	0.003	0.516	12
	log(RD)	0.018	5.138	0.023	0.348	12
	$\lambda_{15}$	0.025	1.440	0.230	0.113	14

Similar results were obtained on third positions than on all positions. Namely, probability of mutations from C:G to A:T increases and probability of mutations from A:T to T:A and A:T to G:C decreases with radioactivity.

Table S3: Phylogenetic Least Square (PGLS) regression of  $d_S$  computed with CoEvol, CodeML or mapNH against radioactivity measured as the  $\alpha$  radioactivity measured in sediments, as the effective dose received by organisms or as the areal proportion of metamorphic and magmatic rocks in a radius of 15km around the sampled point ( $\lambda_{15}$ ). Each test corresponds to one likelihood ratio test between the models with and without the given explanatory variable. For  $\alpha$  radioactivity and received dose, sampled sites with a break in the secular equilibrium were removed, resulting in tests with only 12 taxa.

	$d_S$ CoEvol			$d_S$ CodeML			$d_S$ MapNH			N
	Slope	p.value	Cox-Snell R <sup>2</sup>	Slope	p.value	Cox-Snell R <sup>2</sup>	Slope	p.value	Cox-Snell R <sup>2</sup>	
log( $\alpha$ radio.)	0.034	0.014	0.39	0.054	0.021	0.36	0.054	0.018	0.37	12
log(Received Dose)	0.038	0.011	0.42	0.057	0.024	0.35	0.057	0.023	0.35	12
$\lambda_{15}$	0.076	0.003	0.47	0.118	0.006	0.42	0.118	0.005	0.43	14

## Annexe 8 :

## Résultats des PGLS multivariés et des PGLS incluant le temps de colonisation.

TABLE S1 – Résultats des tests PGLS lorsque toutes les variables sont incluses dans le même modèle. Chaque ligne représente le test sur la pente de la covariable à l'intérieur du modèle multivarié.

		data367sp_nucl	data367sp_mito	data367sp_FAST2	data367sp_COI	data64sp_mito	data64sp_nucl
	N	298	117	298	117	41	43
	Taux	global	global	dS	dS	dS	dS
Latitude	Pente	-0.03741	-0.01849	-0.06422	-0.01765	0.02987	-0.04975
	P.value	0.23950	0.19370	0.06520	0.18340	0.25700	0.50780
Température	Pente	0.02363	-0.00317	0.02552	0.01298	0.01646	0.02450
	P.value	0.10290	0.74410	0.10720	0.15320	0.34480	0.62390
AET	Pente	-0.00031	-0.00012	-0.00064	-0.00017	0.00011	0.00022
	P.value	0.32490	0.46960	0.06260	0.26750	0.58760	0.70830
Delta LGM	Pente	0.06019	-0.00288	0.06348	0.03571	-0.05917	0.11874
	P.value	0.24140	0.88080	0.25840	0.04830	0.16230	0.32700

TABLE S2 – Résultats des tests PGLS lorsque le temps de colonisation est inclus comme covariable. Chaque ligne représente un test LRT entre un modèle ne prenant en compte que le temps de colonisation et un modèle prenant en compte ce temps de colonisation et la variable d'intérêt.

		data367sp_nucl	data367sp_FAST2	data64sp_nucl
	N	71	71	26
	Taux	global	dS	dS
Latitude	Pente	0.03003	0.00665	-0.00155
	P.value	0.39640	0.86250	0.94630
	Cox-Snell R <sup>2</sup>	0.01008	0.00042	0.00017
Température	Pente	0.06885	0.07575	0.03471
	P.value	0.06030	0.05580	0.30160
	Cox-Snell R <sup>2</sup>	0.04850	0.05023	0.04022
AET	Pente	-0.00045	-0.00041	0.00023
	P.value	0.26740	0.35110	0.40620
	Cox-Snell R <sup>2</sup>	0.01718	0.01217	0.02618
Delta LGM	Pente	0.00575	-0.04024	0.05442
	P.value	0.95210	0.69750	0.36180
	Cox-Snell R <sup>2</sup>	0.00005	0.00213	0.03148