



HAL
open science

**La diversité biologique en Nouvelle-Calédonie-Kanaky :
caractérisation à partir des crustacés dulçaquicoles,
recherche des facteurs explicatifs majeurs et
pérennisation des connaissances pour leur (ré-)utilisation**

Coline Royaux

► **To cite this version:**

Coline Royaux. La diversité biologique en Nouvelle-Calédonie-Kanaky : caractérisation à partir des crustacés dulçaquicoles, recherche des facteurs explicatifs majeurs et pérennisation des connaissances pour leur (ré-)utilisation. Biodiversité et Ecologie. Sorbonne Université, 2024. Français. NNT : 2024SORUS285 . tel-04837022v1

HAL Id: tel-04837022

<https://theses.hal.science/tel-04837022v1>

Submitted on 13 Dec 2024 (v1), last revised 7 Jan 2025 (v2)

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.

Sorbonne Université

École Doctorale Sciences de la Nature et de l'Homme – ED 227

UMR BOREA 8067 / Équipe BIOPAC

&

Pôle National de Données de Biodiversité

La diversité biologique en Nouvelle-Calédonie-Kanaky

*Caractérisation à partir des Branchiopodes et Calanoides
dulçaquicoles, recherche des facteurs explicatifs majeurs et
pérennisation des connaissances pour leur (ré-)utilisation*

Coline Royaux

Thèse de doctorat de Biologie des populations et écologie

Dirigée par : Professeur Nicolas Rabet & Docteur Yvan Le Bras

Présentée et soutenue publiquement le 4 novembre 2024

Devant un jury composé de

Mme Le Bris, Nadine	Professeure des universités, Sorbonne Université, Paris	Présidente
Mme Arnaud-Haond, Sophie	Chercheuse, Ifremer, Sète	Rapportrice
M. Piscart, Christophe	Directeur de recherche, CNRS, Rennes	Rapporteur
Mme Després, Laurence	Professeure des universités, Université Grenoble Alpes, Grenoble	Examinatrice
M. Le Bras, Yvan	Ingénieur de recherche, MNHN, Concarneau	Co-encadrant
M. Rabet, Nicolas	Professeur des universités, Sorbonne Université, Paris	Directeur

Avant-propos

L'appellation Nouvelle-Calédonie-Kanaky a été choisie ici par soucis de neutralité politique vis-à-vis des habitants et habitantes de l'archipel. En effet, en respect du droit des peuples à l'autodétermination (Charte des Nations Unies) et des accords de Nouméa (1998), l'utilisation de l'association des emblèmes français et kanaks est soutenue par l'état français. Cette association permet de reconnaître les identités multiples présentes sur l'archipel unies dans un « destin commun » (accords de Nouméa : préambule & art. 1.5, 1998).

Ce manuscrit rassemble les recherches effectuées dans le cadre de ma thèse en biologie évolutive, des populations et écologie. Cette thèse a été réalisée entre 2021 et 2024 à Sorbonne Université au sein de l'Unité Mixte de Recherche Biologie des Organismes et Écosystèmes Aquatiques (UMR 8067, BOREA) dans l'équipe Biodiversité, plasticité, adaptation et conservation : des espèces aux communautés (BIOPAC) et du Pôle National de Données de Biodiversité (PNDB). Cette thèse a bénéficié d'un financement de l'Ecole Doctorale Science de la Nature et de l'Homme (ED 227) et du programme Erasmus+ Gallantries via un appel à projet auquel le PNDB a répondu et géré par le biais de son ancienne unité de rattachement, l'UAR PatriNat (2020-1-NL01-KA203-064717).



With the support of the
Erasmus+ Programme
of the European Union

La thèse est présentée en huit sections : l'introduction générale, le matériel et méthode général, les trois chapitres, la discussion générale, les références pour l'ensemble du manuscrit et les annexes. Chaque chapitre est débuté par un court contexte et résumé. Le premier chapitre contient deux articles, des résultats complémentaires et une discussion dédiée. Le second chapitre contient un article ainsi qu'un encadré (B) présentant un travail en cours. Le dernier chapitre contient une description du cadre de travail ainsi qu'un article. Le manuscrit est rédigé en français hormis les articles qui sont rédigés en anglais.

Dans ce manuscrit, les polices d'écriture utilisées sont des typographies en licence ouverte distribuées à travers le site : <https://www.design-research.be/by-womxn/>. Pour les titres, la police Squatina a été dessinée par Louise Sansoldi, Anaëlle Turc et Julia Veljkovic. Pour le corps du texte, la police Quando a été dessinée Joana Correia.

Résumé

Les systèmes insulaires sont considérés comme des modèles à petite échelle permettant l'étude de systèmes de plus grande envergure. L'isolement relatif de ces systèmes a permis à la recherche de mieux comprendre la dispersion, la spéciation et l'évolution de la biodiversité. Bien que de nombreuses îles aient servi de refuge à d'anciennes lignées, les systèmes insulaires sont particulièrement sensibles aux perturbations et leur biodiversité représente une grande proportion des extinctions connues. Étant donné l'état alarmant de la biodiversité sur les îles et dans le monde, il est nécessaire de veiller à ce que les connaissances produites puissent être rapidement utilisées pour l'action directe et d'assurer que la recherche sur la biodiversité représente un soutien pour les acteurs de la conservation. Des efforts considérables sont nécessaires afin d'assurer la pérennité des connaissances produites via leur transparence, leur accessibilité et leur réutilisation. L'archipel de Nouvelle-Calédonie-Kanaky, connue pour abriter une diversité endémique exceptionnelle ainsi que des roches riches en métaux lourds (ultramafiques) largement exploitées, est un exemple de biodiversité insulaire menacée. Les mares fonctionnent comme un système insulaire et leur biodiversité est mal connue sur l'archipel. De nombreuses espèces de crustacés sont inféodées à certaines mares en fonction de facteurs tels que la turbidité, le pH et le régime trophique. Au sein des crustacés, les groupes des Branchiopodes et Calanoides ont été plus particulièrement investigués dans cette thèse.

Afin d'étudier la diversité spécifique, les espèces ont été identifiées à l'aide de la morphologie et, pour certains taxons, de la génétique. De plus, des phylogénies basées sur un ou plusieurs marqueurs génétiques ont été calculées pour plusieurs genres. En ce qui concerne la diversité génétique intra-spécifique des Branchiopodes et Calanoides de mares, des analyses de génétique des populations ont été utilisées. A partir des identifications d'espèces, des assemblages ont été délimités à l'aide d'un réseau espèces-sites partitionné. Ensuite, des modèles statistiques ont été calculés pour tenter de définir les facteurs environnementaux déterminant la répartition de certaines espèces et des assemblages d'espèces. Sur la base de ces modèles, la distribution la plus probable des espèces et des assemblages a été établie. Tous les protocoles, analyses et résultats ont été partagés de manière

transparente, accessible et réutilisable afin de garantir que le processus scientifique puisse être appliqué dans divers contextes biologiques.

Les résultats de ces analyses ont permis de mettre en évidence une structuration de la diversité des Branchiopodes et Calanoides des mares, de la génétique aux assemblages d'espèces, est fortement influencée par les massifs ultramafiques. En effet, sur les massifs ultramafiques, la diversité génétique intraspécifiques est plus élevée et géographiquement structurée, indiquant une faible connectivité dans ces régions. En outre, un endémisme localisé a été démontré pour plusieurs genres dans les massifs ultramafiques. Ensuite, sur la base de la répartition des espèces, deux assemblages principaux ont été délimités et les principaux facteurs de répartition de ces assemblages ont été déterminés comme étant la précipitation et les types de roches (ultramafiques ou non). Une plus grande richesse spécifique a été constatée dans l'assemblage qui a des affinités avec les massifs non ultramafiques et les zones à faibles précipitations, tandis qu'un taux plus élevé d'espèces rares a été observé dans l'assemblage présent sur les massifs ultramafiques et zones à fortes précipitations. Certaines localités ont semblé intermédiaires en termes de composition d'espèces, suggérant l'existence d'autres assemblages peu échantillonnés. En outre, plusieurs espèces ont été identifiées pour la première fois sur l'archipel, dont certaines sont nouvelles pour la science.

L'étude de systèmes insulaires imbriqués démontre la pertinence des Branchiopodes et Calanoides de mares en tant que modèle biologique pour l'évaluation de la biodiversité. Ceci souligne la valeur de la mise à disposition de connaissances prêtes à l'emploi.

Abstract

Insular systems are considered as small-scale laboratories for understanding larger scale mechanisms. Indeed, their relative isolation from more continuous systems has helped biodiversity scientists to have a better understanding of dispersion, speciation and evolution. Although it is acknowledged that many islands have acted as refugia for ancient lineages, insular systems are particularly sensitive to perturbations and insular biotas represent a large portion of historically recorded extinctions. Considering the alarming state of biodiversity on islands and worldwide, there is a need to ensure that knowledge produced can be rapidly implemented in direct actions and establish the research on biodiversity as a significant support to stakeholders. This will require significant efforts to ensure the long-term sustainability of the knowledge produced through transparency, accessibility and reusability. The New Caledonia-Kanaky archipelago is an example of threatened insular biodiversity as it is known for hosting outstanding endemic diversity along with extensively capitalised heavy metal-rich rocks called ultramafic massifs. Ponds are a type of insular system and are poorly known on the archipelago regarding their inhabiting biodiversity. Many crustacean species are deeply affiliated to certain ponds depending on the environment characteristics such as turbidity, pH and trophic regime. Among crustaceans, Branchiopod and Calanoid groups are more specifically investigated.

To investigate specific diversity, species were identified using morphological and, for a few taxa, genetic identification. Regarding intra-specific genetic diversity in pond Branchiopods and Calanoids, population genetics methods such as haplotype networks, Analysis of Molecular Variance and genetic population differentiation tests were performed. Phylogenies based on one or several genetic markers were computed for several genera. Species assemblages were delineated using a species-site network on which a clustering algorithm was computed. An attempt to determine main drivers for the distribution of some species and of delineated species assemblages has been made using statistical models. Prediction of most probable distribution of these species and delineated assemblages was made based on these models. Finally, all protocols, analyses, and results were made available in a transparent, accessible and reusable manner to ensure that the

scientific process could be applied in similar or different biological contexts by researchers or other publics.

In this thesis, the structuration of ponds Branchiopods and Calanoids diversity from genetics to species assemblages is shown to be highly shaped by the ultramafic massifs. Indeed, intra-specific diversity in several species endemic to ultramafic massifs is higher and is geographically structured, hinting for low connectivity on ultramafic massifs. Additionally, short-range endemism was also evidenced for several genera on ultramafic massifs. Then, based on species distributions, two main species assemblages were delineated and main drivers of the distribution of these assemblages were determined to be precipitation levels and rock types (ultramafic or non-ultramafic). Higher global species richness was evidenced in the assemblage that has affinity with non-ultramafic massifs and low precipitation areas while higher rate of rare species was shown in the assemblage preferring ultramafic massifs and high precipitation areas. Some sampling locations seemed to be intermediate regarding species composition, suggesting other assemblages could exist on the archipelago but were insufficiently sampled. Additionally, several species were identified for the first time on the archipelago, some of which were entirely new to science.

Through the study of nested insular systems, the relevance of ponds Branchiopods and Calanoids as a suitable biological model for biodiversity evaluation is evidenced which highlights the value of providing ready-to-use knowledge for researchers and stakeholders.

Remerciements

Une thèse est une aventure qui passe par nombreuses étapes et qui s'appuie sur de nombreuses personnes, que je souhaite remercier dans les paragraphes qui suivent.

Tout d'abord, évidemment, je tiens à remercier très chaleureusement mes deux encadrants, **Nicolas Rabet** et **Yvan Le Bras**. Vous m'avez apporté un soutien que je n'aurais pas pu imaginer, vous étiez toujours là. La confiance sans limite que vous m'avez montrée l'un comme l'autre a eu le don de me tranquilliser dans les moments de doute. J'ai eu beaucoup de chance d'avoir un encadrement aussi amical, agréable et fluide, je ne pouvais pas rêver mieux.

Nicolas, merci de m'avoir proposé un stage sans hésiter en 2019 en me proposant un sujet aussi passionnant et de m'avoir encouragée ensuite à faire de la recherche. Je suis toujours impressionnée par ton esprit résolument ingénieux et tourné vers la recherche, tes idées sont d'une originalité et d'une pertinence qui force à l'admiration. Merci de les avoir partagées avec moi et de m'avoir soutenue dans les miennes.

Yvan, merci de m'avoir proposé un stage avec toi en 2020, et d'avoir continué à me faire confiance et à m'encourager depuis. Ton énergie débordante, ton ouverture d'esprit sur les thématiques de recherches et tes idées farfelues sont pour moi ta marque de fabrique. Je suis admirative de ton adaptabilité et de ta vivacité d'esprit qui, de mon point de vue, font de toi un preux ingénieur au service d'une recherche vertueuse.

Tarik Mezianne, directeur d'unité, et **Nadia Améziane** et **Guillaume Massé**, cheffe et chef de la station successifs, pour m'avoir accueilli au sein de BOREA et à la station de biologie marine de Concarneau.

Phillipe Keith, responsable de l'équipe BIOPAC, pour m'avoir accueilli dans l'équipe. **Fabienne Audebert** ou « tata Fab' », co-responsable de l'équipe BIOPAC, pour son authenticité, sa sympathie et toujours avoir une histoire rocambolesque à raconter.

Les membres de mon CSI, **Clara Lord**, **Julien Gasparini** et **Sarah Cohen-Boulakia**. La bienveillance, l'attention et les conseils que vous m'avez prodigués ont été très précieux.

Jean-Baptiste Mihoub qui depuis bien des années est un vrai soutien et même un mentor. Impossible d'énumérer tout ce que j'ai pu faire en grande partie grâce à toi, à commencer par cette thèse ! Tout ça pour dire que je suis très reconnaissante d'avoir croisé ton chemin et d'avoir bénéficié de tes conseils.

Boris Leroy qui est impressionnant de bienveillance tout en partageant ses connaissances, propose toujours des idées originales et est plein de bons conseils. Et tout ça en étant hyper sympa.

Céline Bonillo, qui donne toujours d'excellent conseils et est toujours prête à nous faire profiter de son expérience. **Frédéric Busson**, pour son support sur plusieurs tâches de la thèse et son accueil toujours chaleureux lors de mes visites parisiennes. **Myriam Georges**, qui n'a rien lâché et m'a beaucoup aidé pour les manipulations de biologie moléculaire malgré les nombreux négatifs ! **Géraldine Toutirais** pour sa sympathie et son aide à l'utilisation du Microscope Electronique à Balayage.

Nicolas Charpin et **Nathalie Mary** pour les nombreux échantillons et images envoyées de l'autre bout du monde, j'espère vous retourner la faveur à travers mes travaux en continuant les échanges.

Angel Arnold, **Gabrielle Ourous**, **Agathe Moulin** et **Lucie Lenglet** pour leur aide déterminante dans les tris d'échantillons, j'espère vous avoir appris quelques petits trucs utiles et que vous gardez un bon souvenir de vos stages et volontariats. Merci également à **Félix Navarro** pour sa participation aux tris des échantillons et **Oscar Petit** pour son aide sur le traitement des sédiments.

Jean-François Cart pour avoir accepté que j'illustre ma thèse avec ses images.

La team élargie du Pôle National de Données de Biodiversité composée de **Pauline Seguineau** (merci pour ton stage très facile à encadrer), **Octave Cheminée**, **Thibaud Glinez**, **Julien Sananikone** et souvent en vadrouille avec **Olivier Norvez**. Mais aussi **Elie Arnaud** et **Marie Jossé** (merci pour mon premier encadrement de stage qui a été très agréable avec toi) qui ont quitté l'équipe trop tôt !

La team Galaxy Europe et spécialement **Saskia Hiltemann, Helena Rasche** et **Anne Fouilloux** qui m'ont beaucoup aidé sur la rédaction de l'article.

Pendant ces 3 ans, j'ai eu la chance de naviguer entre Concarneau et Paris, et à chaque endroit ses personnes qui partageaient mon quotidien et qui l'ont enrichi.

Andreas et **Vincent**, vous ne vous connaissez pas mais vous êtes les analogues concarnois et parisien de ma thèse ! Vous êtes les meilleurs copains du labo et de la station, les conseillers, les vanneurs et j'en passe. (Mention spéciale Vincent pour ton aide précieuse au moment du concours !!!)

A Concarneau, les discussions quotidiennes autour d'un café (ou plus, souvent) avec **Marc Eléaume, Juliette Baron, Caroline Doose, Stéphane Hersent, Pauline Poisson, Victor Vacher, Stéphanie Auzoux-Bordenave, Françoise Denis, Alexandre Desparmet, Solène Avignon, Frank David, Antoine Chabrolle, Cyril Gallut, Isabelle Le Viol, Aïcha Badou, Cassandra Treyvaud, Clément Jourdan, Nadine Le Bris, Kévin Barré, Baptiste Lebourg, Anaïs Pessato, Cédric Hubas, Christian Kerbirou** et tous les collègues de la station marine.

Léa Mariton, merci d'avoir apprécié mes râlements et jeux de mots tout au long de ta thèse.

Jérôme Fournier, Martin Thibault, Fabien Verniest (nain-chameau) et **Andreas Ravache** qui, en plus de faire partie du bureau le plus solaire de la station, ont toujours su répondre à mes questions, que ce soit sur mes analyses ou sur tout le reste.

A Paris, j'ai pu bénéficier des bonjours, des sourires et des gâteaux du mardi de **Marion Mennesson, Juliette Ravaux, Sophie Berland, Gaël Denys, Corinne Guchereau, Céline Ellien, Nicolas Gasco, Alexis Martin, Charlotte Chazeau, Clara Peron, Mélyne Hautecoeur, Jonathan Pfliger, Nicolas Gasco, Romain Causse, Patrice Pruvost, Jules Selles, et Valentin de Mazencourt.**

Au-delà de l'environnement de travail, j'ai pu compter sur de nombreuses personnes dans ma vie personnelle.

Marina qui a tout traversé avec moi depuis trois ans dans la vie comme dans cette thèse, de la lettre de motivation pour répondre à l'annonce jusqu'à la (presque) fin de cette épopée. Les mots ne sont pas suffisants pour exprimer toute la gratitude et l'affection que je te porte. Je me réjouis tant de t'imaginer qui arpente les étendues infiniment blanches à la recherche de bêtes à plume ou à nageoire à observer !

Mes fantastiques brujas, la vie avec vous est si douce c'est difficile d'imaginer mieux. Merci de m'avoir fait oublier les soucis de la thèse tant de fois, j'ai hâte de la suite de notre vie ensemble.

Mes parents, bien sûr. Merci **titi** et **nathou** pour votre amour et votre soutien sans condition.

Jérôme, rencontre tardive dans la thèse mais une des plus significatives. Je n'aurais jamais assez de mercis pour le calme, la fête, le courage, la confiance, les restos et tout le reste. Je suis une veinarde.

Victoire, ma filleule et ses parents, merci d'apporter toute cette légèreté dans ma vie, même à distance.

Oscar, pour ses (très nombreux) encouragements et tout le reste. J'ai hâte de te voir devenir un fier apprenti chercheur.

Marguerite, parce qu'elle a demandé à être dans mes remerciements et que je la trouve trop chouette.

Vincent et **Andreas**, je vous remets dans le perso parce que je sais que vous allez râler sinon.

Je ne citerais pas tou.tes les ami.es qui m'ont suivi de près ou de loin dans cette aventure mais merci pour tout.

Communications et stages

Publications incluses dans la thèse

Chapitre I

Royaux C, Charpin N, Rabet N (2024). Genetic variability of New Caledonian *Boeckella* de Guerne & Richard, 1889 (Copepoda: Calanoida), with the description of a new species. *Journal of Crustacean Biology*, **44**, 1–17. <https://doi.org/10.1093/jcbiol/ruae001> ; <https://hal.science/hal-04681440>

Royaux C, Charpin N, Mary N, Busson F, Bonillo C, Rabet N (soumis). Remarkable genetic and specific diversification in *Boeckella* from New Caledonia suggest strong micro-endemism. *Freshwater Biology*.

Chapitre II

Royaux C, Leroy B, Mary N, Charpin N, Le Bras Y, Rabet N (en préparation). Major drivers of community assemblages in ponds of a biodiversity hotspot.

Chapitre III

Royaux C, Mihoub J-B, Jossé M, Pelletier D, Norvez O, Reecht Y, Fouilloux A, Rasche H, Hiltemann S, Batut B, Eléaume M, Segueineau P, Massé G, Amossé A, Bissery C, Lorrilliere R, Martin A, Bas Y, Virgoulay T, Chambon V, Arnaud E, Michon E, Urfer C, Trigodet E, Delannoy M, Loïs G, Julliard R, Grüning B, Le Bras Y (accepté). Guidance framework to apply best practices in ecological data analysis: Lessons learned from building Galaxy-Ecology. Recommandé par *Peer Community In Ecology* <https://doi.org/10.32942/X2G033>

Autres publications

Pelletier D, Roos D, Bouchoucha M, Schohn T, Roman W, Gonson C, Bockel T, Carpentier L, Preuss B, Powell A, Garcia J, Gaboriau M, Cadé F, **Royaux C**, Le Bras Y, Reecht Y (2021) A Standardized Workflow Based on the STAVIRO Unbaited Underwater Video System for Monitoring Fish and Habitat Essential Biodiversity Variables in Coastal Areas. *Frontiers in Marine Science*, **8**, 689280. <https://doi.org/10.3389/FMARS.2021.689280/BIBTEX>

Pelletier D, Rouxel J, Fauvarque O, Hanon D, Gestalin JP, Lebot M, Dreano P, Furet E, Tardivel M, Bras Y Le, **Royaux C**, Leguen G (2021) Kosmos: An open source underwater video lander for monitoring coastal fishes and

habitats. *Sensors*, **21**, 7724. <https://doi.org/10.3390/S21227724/S1>

Afgan E, Nekrutenko A, Grüning BA, Blankenberg D, Goecks J, Schatz MC, Ostrovsky AE, Mahmoud A, Lonie AJ, Syme A, Fouilloux A, Bretaudeau A, Nekrutenko A, Kumar A, Eschenlauer AC, Desanto AD, Guerler A, Serrano-Solano B, Batut B, Grüning BA, Langhorst BW, Carr B, Raubenolt BA, Hyde CJ, Bromhead CJ, Barnett CB, **Royaux C**, Gallardo C, Blankenberg D, Fornika DJ, Baker D, Bouvier D, Clements D, De Lima Morais DA, Tabernero DL, Lariviere D, Nasr E, Afgan E, Zambelli F, Heyl F, Psomopoulos F, Coppens F, Price GR, Cuccuru G, Corguillé G Le, Von Kuster G, Akbulut GG, Rasche H, Hans-Rudolf H, Eguinoa I, Makunin I, Ranawaka IJ, Taylor JP, Joshi J, Hillman-Jackson J, Chilton JM, Kamali K, Suderman K, Poterlowicz K, Yvan LB, Lopez-Delisle L, Sargent L, Bassetti ME, Tangaro MA, Van Den Beek M, Cech M, Bernt M, Fahrner M, Tekman M, Föll MC, Schatz MC, Crusoe MR, Roncoroni M, Kucher N, Coraor N, Stoler N, Rhodes N, Soranzo N, Pinter N, Goonasekera NA, Moreno PA, Videm P, Melanie P, Mandreoli P, Jagtap PD, Gu Q, Weber RJM, Lazarus R, Vorderman RHP, Hiltemann S, Golitsynskiy S, Garg S, Bray SA, Gladman SL, Leo S, Mehta SP, Griffin TJ, Jalili V, Yves V, Wen V, Nagampalli VK, Bacon WA, De Koning W, Maier W, Briggs PJ (2022) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic acids research*, **50**, W345–W351. <https://doi.org/10.1093/NAR/GKAC247>

Hiltemann S, Rasche H, Gladman S, Hotz HR, Larivière D, Blankenberg D, Jagtap PD, Wollmann T, Bretaudeau A, Goué N, Griffin TJ, **Royaux C**, Bras Y Le, Mehta S, Syme A, Coppens F, Droesbeke B, Soranzo N, Bacon W, Psomopoulos F, Gallardo-Alba C, Davis J, Föll MC, Fahrner M, Doyle MA, Serrano-Solano B, Fouilloux AC, van Heusden P, Maier W, Clements D, Heyl F, Grüning B, Batut B (2023) Galaxy Training: A powerful framework for teaching! *PLOS Computational Biology*, **19**, e1010752. <https://doi.org/10.1371/JOURNAL.PCBI.1010752>

Communications orales

Présentations en congrès

Royaux C (2023). « From EBV operationalization pilot to GBiOS, a french BON point of view » ([vidéo](#)) basé sur un travail en collaboration avec Norvez O, Le Bras Y, Jossé M, Seguineau P, Genthon T, Sananikone J, Cheminée O, Julliard R, Delavaud A, Eleaume M, Pavoine S, Pelletier D, Mihoub J-B. *Geo BON Global conference*, 10-13 octobre, Montréal, Canada.

Royaux C & Le Bras Y (2022). « Open Science for Better FAIRness: A biodiversity virtual research environment point of view » ([résumé](#)) basé sur un travail en collaboration avec Arnaud E, Sananikone J, Jossé M, Madelin M, Pelletier D, Norvez O. *Biodiversity Information Science and Standards* (TDWG), 17-21 octobre, en ligne.

Royaux C (2022). « From Biodiversity Observation Networks to Datasets and Workflows Supporting Biodiversity Indicators » ([résumé](#)) basé sur un travail en collaboration avec Arnaud E, Sananikone J, Jossé M, Pavoine S, Pelletier D, Norvez O, Mihoub J-B, Le Bras Y. *Biodiversity Information Science and Standards* (TDWG), 17-21 octobre, en ligne.

Royaux C & Jossé M (2022). « Galaxy-E : A practical way to propose accessible, reproducible and transparent analysis of Ecology data » basé sur un travail en collaboration avec Le Bras Y. *SORTEE Conference 2022*, 11 juillet, en ligne.

Animation d'ateliers en congrès

Royaux C, Norvez O, Le Bras Y (2023). « From Raw Heterogeneous Biodiversity Data to Operational Indicators Based on Essential Variables » ([résumé](#)) basé sur un travail en collaboration avec Jossé M, Segueineau P, Genthon T, Sananikone J, Cheminée O, Julliard R, Delavaud A, Eleaume M, Pavoine S, Pelletier D, Mihoub J-B. *Geo BON Global conference*, 10-13 octobre, Montréal, Canada.

Royaux C, Mihoub J.B, Pavoine S, Norvez O & Le Bras Y (2022). « From raw biodiversity data to operational indicators through Essential Biodiversity Variables » ([résumé W12](#)). *Conference on Ecological Sciences (SFE² GfÖ EEf)*, 21-25 novembre, Metz, France.

Royaux C, Jossé M, Arnaud E, Sananikone J, Norvez O & Le Bras Y (2022). « Galaxy-E Hackathon ». *SORTEE Conference 2022*, 11 juillet, en ligne.

Enseignements et formations

Royaux C (2023). Cours de 2 heures dans l'Unité d'Enseignement « Biodiversité en milieu marin côtier », Master 2 Systématique, Évolution, Paléontologie du Muséum National d'Histoire Naturelle. Novembre, Concarneau, France.

Royaux C, Norvez O & Le Bras Y (2023). Formation Challenge Green Data for Health « Initiation à Galaxy-Ecology » ([vidéo](#)). Ministère de la Transition Écologique, 24 mai, en ligne.

Royaux C (2022). TD de 2 heures dans l'Unité d'Enseignement « Population Genetics and genomics », Master 2 Biodiversité, Ecologie, Evolution de l'Université de Rennes 1. Décembre, Rennes, France.

Stages encadrés

Master 1

Lucie Lenglet, « Inventaire et analyse de la diversité de crustacés planctoniques des mares d'eau douce de Nouvelle-Calédonie ». Avril – mai 2022, Master Biodiversité, Écologie, Évolution de Sorbonne Université.

En co-encadrement avec Nicolas Rabet.

Gabrielle Ourous, « Le zooplancton des mares temporaires dulcicoles de Nouvelle-Calédonie ». Avril – mai 2022, Master Science de la Mer de l'Université de Toulon. En co-encadrement avec Nicolas Rabet.

Angel Arnold, « Endémisme des sididaes de Nouvelle Calédonie ». Avril – mai 2023, Master Systématique, Évolution, Paléontologie du Muséum National d'Histoire Naturelle. En co-encadrement avec Nicolas Rabet.

Master 2

Pauline Segineau, « Mise en place d'un workflow Galaxy pour l'éco-régionalisation de la mer Dumont D'Urville ». Février – Juillet 2023, Master Bio-informatique de l'Université Clermont Auvergne. En co-encadrement avec Marc Eléaume, Yvan Le Bras et Alexis Martin.

Tancrède Bourgeois, « Etude de l'endémisme des crustacés composant le plancton d'eau douce de l'archipel de Belep en comparaison avec Grande-Terre (Nouvelle-Calédonie) ». Mars – août 2023, Master Systématique, Évolution, Paléontologie du Muséum National d'Histoire Naturelle. En co-encadrement avec Nicolas Rabet.

*A mes quatre grands-parents qui sont tous à l'origine de ma
curiosité pour la nature.*

*Anne-Marie pour nous avoir emmenés dans la forêt tropicale
de Fumay à d'innombrables reprises.*

*Joël pour avoir toujours pris le temps de faire des visites
bilingues du potager pour sa petite fille.*

*Framboise pour nous avoir montré des expositions, livres et
jeux qui ont rendu tous les sujets intéressants.*

*Et bien sûr, Androu pour nous avoir passionnés de la
biodiversité du marais des hauts buttés dès le plus jeune âge.*

Sommaire

Avant-propos	I
Résumé	III
Abstract	V
Remerciements	VII
Communications et stages	XI
Sommaire	XVII
Introduction	1
I. L'insularité & la perte de biodiversité	3
II. La perte de connaissances sur la biodiversité	4
III. La Nouvelle-Calédonie-Kanaky, un laboratoire de la biodiversité	7
IV. Les mares d'eau douce : les îles dans l'île	11
V. Les crustacés des mares	14
VI. Objectifs et problématiques	15
Matériel & méthodes	17
I. Échantillons	19
I. 1. Repérage des localités et échantillonnage	19
I. 2. Traitement des échantillons de sédiment	21
I. 2. i. Tri et identification des formes de résistance	21
I. 2. ii. Élevages	22
I. 3. Traitement des échantillons au filet et issus des élevages	22
I. 3. i. Tri et identification des espèces	22
I. 3. ii. Délimitation des assemblages et modélisation	24
I. 3. iii. Extraction de l'ADN	24
I. 3. iv. Séquençage du barcode	25
I. 3. v. Séquençage shotgun de nouvelle génération (NGS)	26
I. 3. vi. Analyses phylogénétiques	26
I. 3. vii. Structuration génétique des populations	27
II. Réutilisation des connaissances : exploration des moyens de pérennisation ...	28
II. 1. Définition des bonnes pratiques	29
II. 2. Le cycle de vie de la recherche	30
II. 3. Le Plan de Gestion de Données	30
II. 4. Le partage des protocoles et données	31
II. 5. La mise à disposition des workflows de nettoyage des données et des analyses	33
II. 6. L'écriture et les publications scientifiques	34
II. 7. Collaborations et communications	34

Chapitre I	35
I. Résumé	37
II. Les espèces de crustacés d'eau douce	39
III. Le cas du genre <i>Boeckella</i>	42
Genetic variability of New Caledonian <i>Boeckella</i> De Guerne & Richard, 1889 (Copepoda: Calanoida), with description of a new species	42
Abstract	42
Introduction	43
Materials and methods	45
Results	50
Discussion	65
Remarkable genetic and specific diversification in <i>Boeckella</i> from New Caledonia suggest strong micro-endemism	71
Abstract	71
Introduction	72
Material and Methods	74
Results	78
Discussion	87
IV. Quelques cas parmi les Branchiopodes	95
IV. 1. Le genre <i>Streptocephalus</i>	95
IV. 2. Le genre <i>Eulimnadia</i>	97
IV. 3. Le genre <i>Triops</i>	101
IV. 4. Le genre <i>Latonopsis</i>	105
IV. 5. Le genre <i>Lynceus</i>	109
V. Discussion	116
Chapitre II	121
I. Résumé	123
II. La distribution des assemblages d'espèces	125
Major drivers of community assemblages in ponds of a biodiversity hotspot.125	
Abstract	125
Introduction	125
Material and methods	128
Results	133
Discussion	142
Chapitre III	157
I. Résumé	159
II. Cadre de travail transparent, reproductible et (ré)utilisable	160
II. 1. Planifier	160
II. 1. i. Mise en œuvre	160
II. 1. ii. Apports pour la pérennisation	161
II. 2. Collecter	161
II. 2. i. Mise en œuvre	161
II. 2. ii. Apports pour la pérennisation	162
II. 3. Organiser	162
II. 3. i. Mise en œuvre	162
II. 3. ii. Apports pour la pérennisation	163

II. 4. Nettoyer et Analyser	164
II. 5. Ecrire et publier	165
II. 5. i. Mise en œuvre	165
II. 5. ii. Apports pour la pérennisation	166
III. Cadre pour les bonnes pratiques de développement des processus analytiques.	167
Guidance framework to apply best practices in ecological data analysis: Lessons learned from building Galaxy-Ecology	167
Abstract	167
Introduction	168
Guidelines for best practices	173
Entering a new dimension: the Galaxy-E initiative example	177
Discussion and limitations	183
Discussion générale, conclusion et perspectives	187
I. Principaux résultats	189
II. Eléments de conservation	190
II. 1. Les milieux d'eau douce	190
II. 2. Les milieux ultramafiques	192
II. 3. Les méthodes d'analyses	193
III. Perspectives	194
III. 1. Partage des connaissances et publications	194
III. 2. Études complémentaires	195
III. 3. Au-delà des modèles d'études	197
Références	199
Annexe A : Chapitre I - Partie II - Article I	234
Annexe B : Chapitre I - Partie II - Article II	242
Annexe C : Chapitre I - Partie IV	250
Annexe D : Chapitre II	256
Annexe E : Chapitre III	260

<< Cela devient de plus en plus évident. Je ne peux plus le nier. Je suis petit. >> - Plankton, Bob l'Eponge

Introduction

I. L'insularité & la perte de biodiversité

L'insularité est un processus emblématique de l'étude de la biodiversité. L'étude du phénomène est au point de départ de la théorie de l'évolution de Charles Darwin, principe fondateur de la biologie évolutive en tant que discipline (Darwin, 1859). En effet, les becs des pinsons des îles Galápagos présentant une forme différente en fonction des ressources disponibles sur l'île habitée représentent l'exemple le plus emblématique de la théorie de Charles Darwin. D'autre part, les oiseaux font aussi partie des représentants de la vulnérabilité des écosystèmes insulaires. En effet, chez les oiseaux, plus de 90% des extinctions reportées historiquement sont des espèces insulaires (Fernández-Palacios *et al.*, 2021). Ce chiffre impressionnant est représenté par des espèces d'oiseaux géants disparus il y a environ 1000 ans tels que *Vorombe titan* disparu de Madagascar ou encore *Sylviornis neocaledoniae* disparu de Nouvelle-Calédonie-Kanaky (Kier *et al.*, 2009). En dehors des oiseaux, 68% des espèces éteintes de plantes vasculaires et 71% des espèces éteintes d'arthropodes étaient insulaires. Cette vulnérabilité du biotope insulaire prend son origine dans l'isolation des habitats, la taille réduite par rapport au continent et la fragmentation naturelle de leur distribution. En effet, un système insulaire est défini comme un habitat favorable enclavé dans un habitat défavorable (Whittaker *et al.*, 2023). Cette définition correspond également à d'autres modèles d'habitats comme les pics montagneux et les milieux aquatiques ponctuels, dont les mares. Tous ces milieux insulaires sont soumis à des processus de colonisation par le vivant et de flux démographiques similaire et, de fait, à une vulnérabilité du même ordre. L'étude des mares dans un archipel consiste à explorer des systèmes insulaires imbriqués qui présentent à la fois des caractéristiques insulaires décuplées et, de fait, des menaces plus élevées.

Les inquiétudes concernant l'érosion de la biodiversité ont pris une dimension planétaire dès les années 1960. Elles ont notamment été représentées par de multiples alertes de la communauté scientifique et par le best-seller « Silent Spring » écrit en 1962 par Rachel Carson et faisant état d'un printemps silencieux dénué d'une grande partie des populations d'oiseaux. Dès 1973, la Convention sur le commerce international des espèces de faune et de flore sauvages menacées d'extinction (CITES, 1973) est signée par 80 pays. Celle-ci reconnaît l'importance de la faune et la flore sauvage ainsi que la nécessité pour les administrations des pays signataires d'agir urgemment pour la protéger. Plus récemment, l'ONU définissait 17 objectifs pour un développement durable mondial dont plus de la

moitié sont liés directement ou indirectement à l'environnement et la biodiversité (<https://sdgs.un.org/fr/goals>).

En effet, les preuves d'un déclin de la biodiversité et des dommages aux écosystèmes par l'humain s'accumulent depuis plus de 60 ans. Or, le 15 septembre 2020, le Global Biodiversity Outlook 5 publié par la convention sur la diversité biologique et les nations unies a déterminé que "aucun des Objectifs d'Aichi (2011-2020) pour la biodiversité ne sera entièrement réalisé". Ce constat décourageant, peut aussi nous pousser à repenser les approches de la recherche en biodiversité pour élargir son impact sur les politiques publiques et internationales.

II. La perte de connaissances sur la biodiversité

En parallèle de la perte de biodiversité, la perte de connaissances en recherche telle que représentée par Michener *et al.* (1997 ; fig. I.1) est alarmante pour la science comme pour la biodiversité. Cette perte de connaissance est représentée dans le temps avec pour cause les oublis naturels de certains détails, de potentiels incidents causant une perte de données ou encore la difficulté d'accès à la connaissance stockée dans la mémoire d'une personne partie à la retraite. Ces pertes peuvent être considérables et occasionner de véritables barrières à la compréhension des dynamiques du vivant dans le passé ou sur le long terme (Hughes *et al.*, 2021). Pour éviter cela, la pérennisation représente le fait de rendre les connaissances à l'épreuve du temps (Borgman, 2020). Par exemple, la capacité d'accéder à des données publiées ou encore de faire fonctionner un logiciel après plusieurs années. La taxonomie est un des domaines les plus affectés par cette perte de connaissance, notamment par le manque de spécialistes de ce domaine. En effet, ne serait-ce que dans les collections de muséums, de nombreuses espèces sont mal étiquetées ou mal identifiées (Drew, 2011). Ce fait se transpose aux études de terrain et aux données retrouvées sur les plateformes en ligne telles que le Global Biodiversity Information Facility (GBIF ; <https://www.gbif.org/>) pour les occurrences d'espèces ou encore GenBank (NCBI ; <https://www.ncbi.nlm.nih.gov/>) pour les données de génétique (Gaiji *et al.*, 2013; Pentinsaari *et al.*, 2020). Les mauvaises identifications sont particulièrement délétères pour les études sur la biodiversité et peuvent mener à des conclusions erronées voire des recommandations de conservation inefficaces mais peuvent aussi diminuer la crédibilité accordée aux études réalisées dans le passé et dans le futur (Rodrigues *et al.*, 2022).

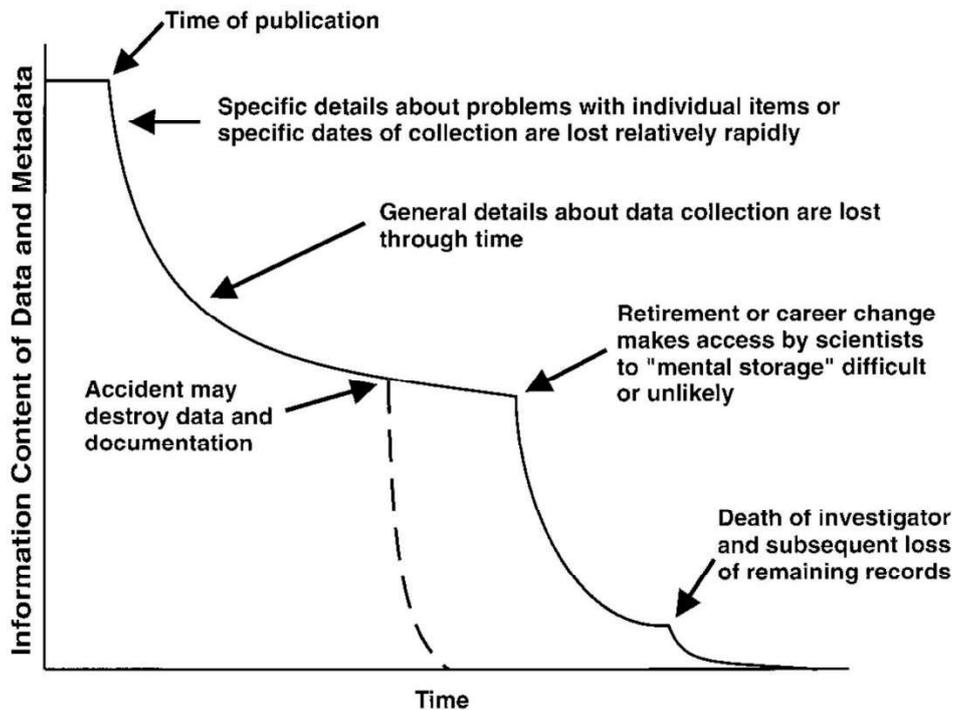


Figure I.1 – Quantité de connaissances dans le temps à partir du moment de publication d'un article d'après Michener *et al.*, 1997 (fig. 1, pp. 332).

Or, cette perte de connaissance n'existe pas que dans le temps mais aussi à mesure que les barrières linguistiques, institutionnelles et de pratique s'accumulent. En effet, accéder aux détails non indiqués dans un article publié sera plus aisé si la personne ayant écrit l'article occupe le bureau voisin ou est dans le même laboratoire. Bien que ces constats semblent évidents, le décalage existant entre les connaissances générées et leur application pour la surveillance ou dans les réglementations est d'autant plus problématique dans un contexte urgent de crise de la biodiversité. Par exemple, il est démontré que la diversité génétique est une composante centrale à la résilience d'une espèce en danger et que la perte de diversité génétique chez les populations est d'une magnitude bien plus élevée que celle de la diversité spécifique (Hughes *et al.*, 1997; Manel *et al.*, 2020). Cependant, la prise en compte de la diversité génétique dans les décisions pour la conservation reste encore anecdotique. Bien que les décisions pour la conservation soient dans les faits rarement du ressort de la recherche, certains leviers existent pour donner un sens aux connaissances acquises en passant du savoir à l'action (Graham *et al.*, 2006; Knight *et al.*, 2008; Lauber *et al.*, 2011).

Tout d'abord, le progrès en recherche est, par définition, basé sur des travaux passés. Les fondations d'un projet de recherche sont construites à partir des recherches précédentes et afin de construire sereinement sur ces fondations, il est fondamental d'accéder à suffisamment d'informations afin de s'assurer de la

qualité de celles-ci. La transparence dans la recherche est cruciale pour limiter les zones grises et la perte de connaissance. En effet, une bonne transparence sur la construction du projet, les méthodes, les données générées et les analyses permet d'augmenter le niveau de confiance accordé par les pairs mais aussi par le grand public (Powers & Hampton, 2019).

Ensuite, l'augmentation constante du nombre de méthodes et de technologies avancées pour l'acquisition, le stockage et l'analyse des données représente un défi pour la recherche qui doit se maintenir à jour (Hampton *et al.*, 2017). Le partage de connaissances de manière accessible est donc important pour maintenir une recherche véritablement collaborative et constructive. De plus, cette accessibilité peut permettre d'élargir le potentiel d'un travail de recherche en alimentant d'autres travaux ou en favorisant l'émergence de pistes pour l'application de la recherche (Touchon & McCoy, 2016; Gownaris *et al.*, 2022).

En outre, la transparence et l'accessibilité des connaissances alimentent les possibilités de réutilisation de celles-ci. La réutilisation directe des connaissances produites est dans un premier temps bénéfique pour la personne ayant produit ces connaissances. Par exemple, il est courant de devoir réutiliser un protocole ou un script analytique créé dans le cadre d'un projet de recherche mais si leur développement n'a pas été pensé pour être réutilisé, une forte inertie peut être générée pour reprendre en main les détails du protocole ou actualiser le script. Une perte de temps similaire peut-être générée au moment de collaborer avec de nouvelles personnes (*e.g.*, stages, collaborations scientifiques). En effet, des connaissances directement réutilisables permettent de passer moins de temps à discuter d'aspects techniques au profit de temps passé directement sur la recherche en elle-même (Borgman, 2020).

L'attention aux bonnes pratiques dans les projets de recherche peuvent grandement bénéficier à une meilleure compréhension de la biodiversité. Par ce biais, la recherche en biodiversité et en conservation peut se placer comme un levier majeur pour tendre vers une société qui se place au cœur des écosystèmes qui sont nécessaires au maintien de celle-ci.

La poursuite d'une recherche véritablement transparente, accessible et réutilisable de façon pérenne est d'autant plus importante pour un territoire comme la Nouvelle-Calédonie-Kanaky. Bien que l'archipel soit pourvu d'une université et de plusieurs laboratoires, une grande partie de la recherche sur la

biodiversité du Pays, dont cette thèse, est effectuée depuis l'Europe, distante de plus de 12 000 km, ou encore depuis l'Australie, distante de plus de 1 000 km. La délocalisation d'un projet de recherche n'est pas un problème en soi, mais l'éloignement a le potentiel de diminuer grandement les impacts positifs d'un travail de recherche sur le territoire considéré (Haelewaters *et al.*, 2021). Cette problématique peut être en grande partie évitée à travers une mise à disposition pérenne des connaissances permettant aux acteurs (*i. e.*, issus de la recherche, de l'expertise, des institutions ou parmi les citoyens) du territoire de se les approprier.

III. La Nouvelle-Calédonie-Kanaky, un laboratoire de la biodiversité

La Nouvelle-Calédonie-Kanaky (NCK) est un archipel particulièrement singulier à étudier. Les plus vieilles îles sont souvent d'origine continentale comme Madagascar qui s'est séparée de l'Afrique il y a 160 millions d'années (Ma) et de l'ensemble Inde-Seychelles il y a 80 Ma. C'est également le cas pour l'île principale de la NCK, Grande Terre, qui s'est séparée de l'Australie il y a environ 105 Ma. Cependant, les variations des niveaux de la mer et de l'altitude des continents au cours des temps géologiques peuvent occasionner des submersions périodiques de la croûte continentale. Aujourd'hui, le continent Zealandia est majoritairement submergé et Grande Terre et la Nouvelle-Zélande représentent sa majeure partie émergée. Il a été démontré que Grande Terre a été submergée sur la période de -75 Ma à -60 Ma (Maurizot & Campbell, 2020). De ce fait, la biodiversité présente sur l'île l'a recolonisée a posteriori de la même façon qu'une île océanique (*i. e.* île n'ayant jamais été reliée au continent et formée par des roches volcaniques ou des formations coralliennes). Du fait des submersions et de la distance à l'Australie (plus de 1000 km ; fig. I.2), on ne retrouve pas d'amphibiens ou de mammifères terrestres natifs sur Grande Terre (hormis les chauves-souris). D'autres submersions ont pu avoir lieu entre -34 Ma et -25 Ma mais les preuves géologiques d'une submersion ou d'une émergence à cette période manquent. Cependant, des découvertes en phylogénie moléculaire datent certains des clades terrestres de l'archipel avant et entre -34 Ma et -25 Ma (Grandcolas *et al.*, 2008; Nattier *et al.*, 2017). Ces résultats suggèrent un maintien de terres émergées dans la zone durant cette période même dans le cas de submersions intermittentes d'une partie ou de l'entièreté de Grande Terre (Maurizot & Campbell, 2020). Après -34 Ma, un évènement tectonique a provoqué un charriage du manteau terrestre en surface ayant recouvert toute l'île. Après altération, il a formé les massifs très riches en

métaux lourds (appelés ultramafiques) présents essentiellement au sud de Grande Terre (fig. I.2).

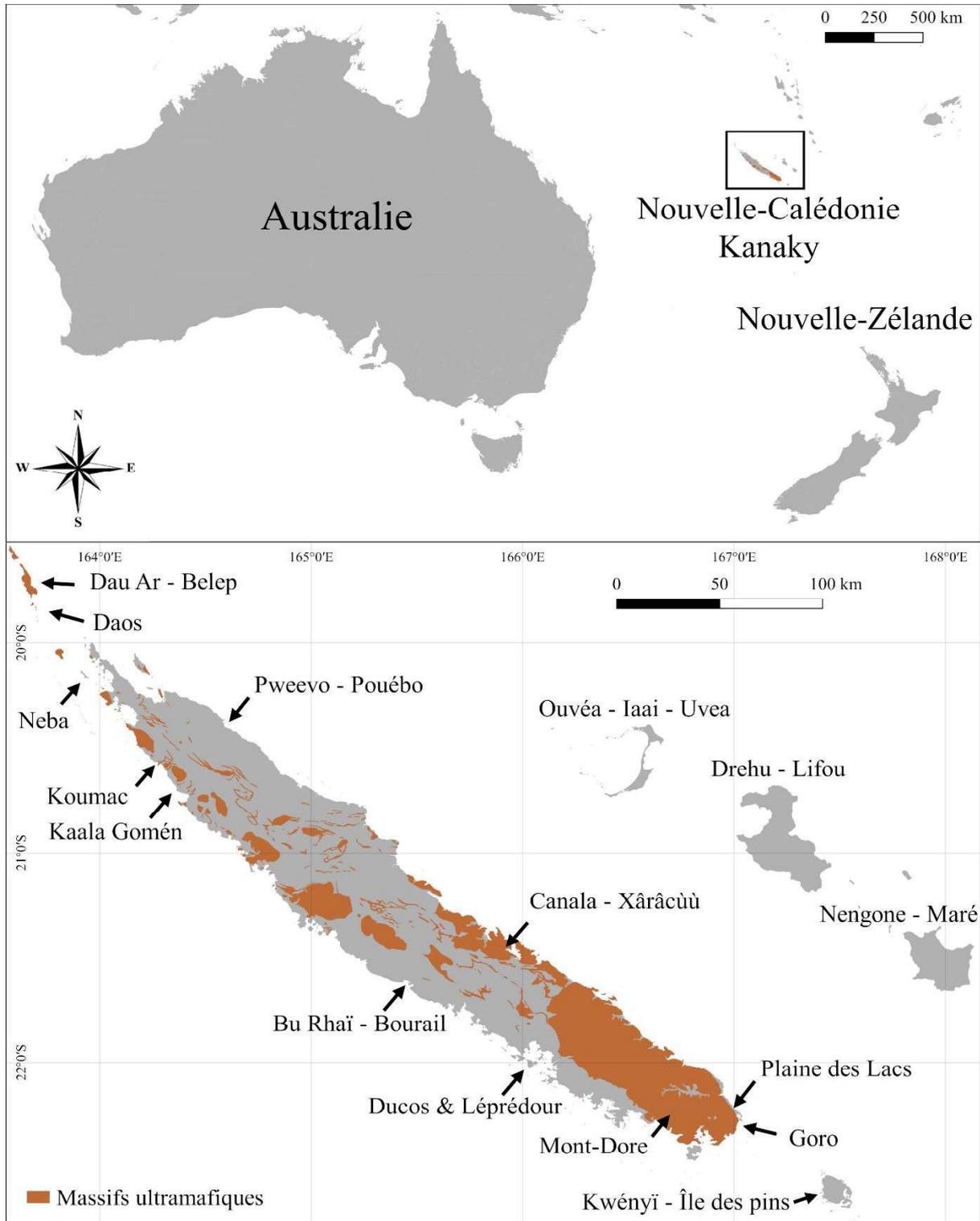


Figure I.2 – En haut, carte situant la Nouvelle-Calédonie-Kanaky par rapport à l’Australie et la Nouvelle-Zélande. En bas, carte de la Nouvelle-Calédonie-Kanaky situant les régions évoquées dans la thèse et les massifs ultramafiques en orange foncé.

Ces massifs contiennent notamment une grande quantité de nickel (20-30% des ressources mondiales) et représentent un intérêt minier majeur. En NCK, les massifs ultramafiques sont exploités depuis la fin du 19^e siècle (Richer de Forges & Pascal, 2008; Pitoiset, 2016) et il est estimé que les réserves sont suffisantes pour continuer cette exploitation pendant au moins un siècle supplémentaire (Batterbury & Kowasch, 2024). En effet, l'économie du Pays est en grande partie basée sur l'exploitation du nickel (90 - 95% des exportations selon L'Huillier & Jaffré, 2010 et Batterbury & Kowasch, 2024). Cependant, les effets de l'exploitation minière sur les écosystèmes et la santé humaine est indéniable et cause notamment une déforestation massive, une augmentation de l'érosion, une détérioration de la qualité de l'eau et une fragmentation des habitats (L'huillier *et al.*, 2010; Desoutter & Bertaud, 2019; Schneider *et al.*, 2020; Boula *et al.*, 2022). Malgré cela, du fait de l'importance de l'industrie minière pour l'économie de l'archipel, la protection des espaces a été globalement ignorée jusqu'aux années 1990 (Toussaint, 2018; Rodary, 2024). En effet, l'impact d'une diminution brutale de l'activité minière sur l'économie et les personnes habitant l'archipel ne peut pas être ignoré dans la recommandation de mesures environnementales.

Ces constats sur la politique environnementale de la NCK sont d'autant plus importants quand on considère qu'elle est qualifiée de point chaud de la biodiversité. Depuis 2009, le Code minier adopté par le gouvernement de l'archipel inclus de nombreuses régulations de l'activité minière. Parmi elles, la norme de revégétaliser les terrains miniers est appliquée mais peu efficace. En effet, la restauration des écosystèmes est grandement complexifiée par le caractère endémique (*i. e.* distribution d'un taxon restreinte à un habitat ou espace géographique donné) de près de 90% de la flore des massifs ultramafiques (Jaffré *et al.*, 1994, 2010; Rodary, 2024). En NCK, les massifs ultramafiques ont largement conditionné le développement du vivant et sont connus pour accueillir un phénomène d'endémisme à courte distance appelé micro-endémisme, endémisme localisé ou encore endémisme restreint (Caesar *et al.*, 2017). De fait, l'endémisme est un concept échelle-dépendant. Une espèce peut être endémique d'un type de milieu (*e.g.*, forêts tropicales humides), d'un hémisphère, d'un continent, d'une région ou encore d'un sommet montagneux. Le micro-endémisme est un type d'endémisme qui désigne des taxons à la distribution particulièrement restreinte. Le seuil à partir duquel une espèce est considérée micro-endémique dépend de l'objet d'étude. En NCK, ce phénomène de micro-endémisme est constaté chez les

plantes (Lannuzel *et al.*, 2021; Isnard & Jaffré, 2024), les sauterelles (Nattier *et al.*, 2013) ou encore les geckos (Hudel *et al.*, 2020).

Malgré sa surface réduite (18 827 km², proche de celle des Pyrénées), une grande diversité de paysages est observée en NCK et constitue une des raisons avancées pour expliquer l'amplitude des phénomènes d'endémisme et de micro-endémisme. En effet, on retrouve sur l'archipel des forêts sèches, des forêts humides, des prairies (souvent d'origine anthropiques), des maquis miniers ou encore des mangroves (fig. I.3).

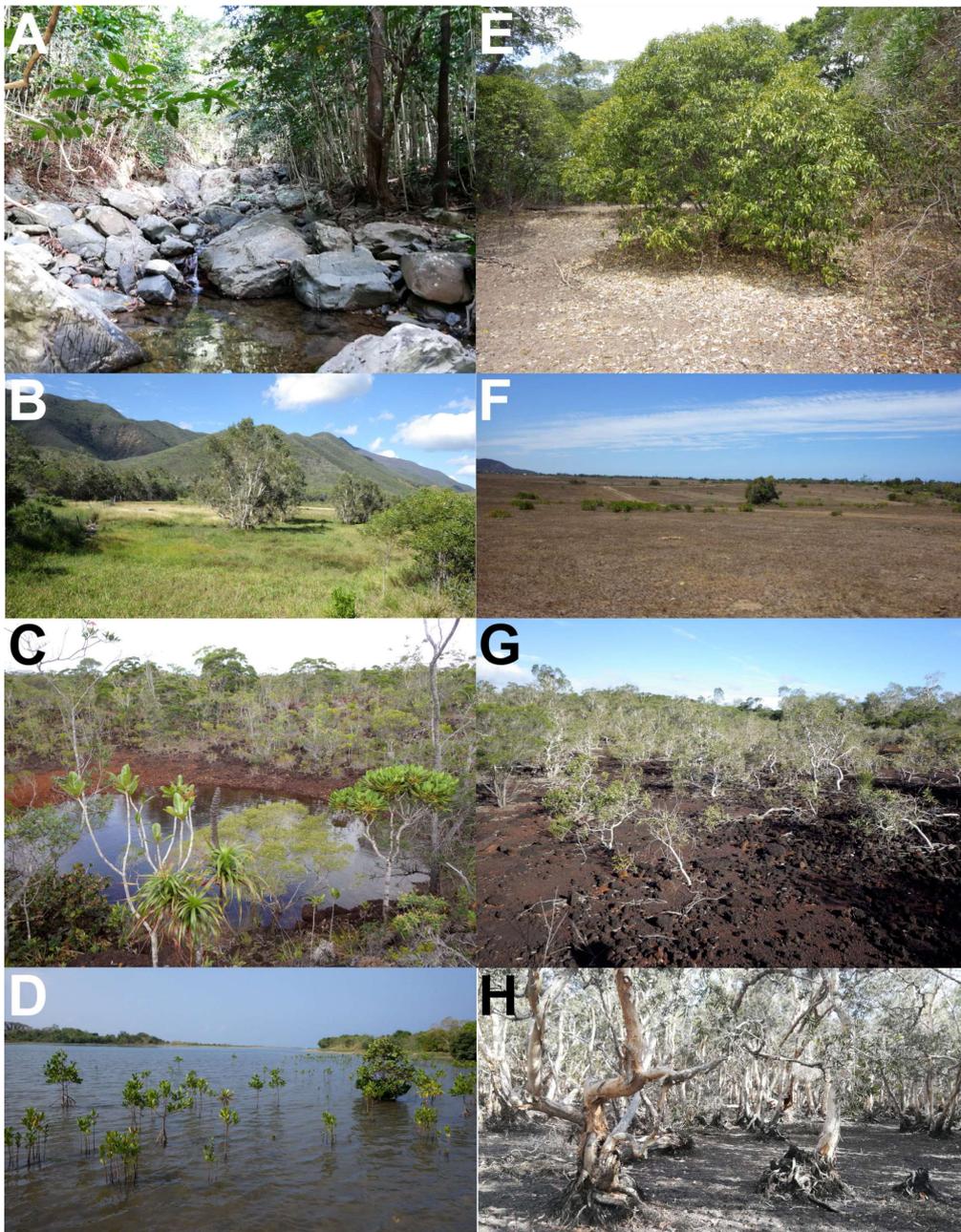


Figure I.3 – Paysages de Nouvelle-Calédonie-Kanaky. A-D habitats en zone à fortes précipitations, E-H habitats en zone à faible précipitations. A : forêt humide. B, F : prairie. C, G : maquis minier. D : mangrove. E : forêt sèche. H : marais à niaoulis.

A l'échelle de l'archipel, le micro-endémisme semble être qualifié lorsque plusieurs espèces d'un même genre sont retrouvées dans des aires de distribution voisines et non recouvrantes. Parmi les exemples de micro-endémisme présentés ci-dessus, la plupart sont retrouvés au sud de Grande Terre sur le plus large massif ultramafique de l'archipel. Dans cette région de l'île, on retrouve notamment de nombreuses étendues d'eau douces aux caractéristiques très diversifiées.

IV. Les mares d'eau douce : les îles dans l'île

Dans cette thèse, la définition la plus large pouvant être donnée aux habitats étudiés est « milieu aquatique dulcicole (*i. e.* d'eau douce), lentique (*i. e.* avec un renouvellement de l'eau lent), et temporaire (*i. e.* connaissant un assèchement) ». Or, ces termes se rapportent autant aux mares qu'aux lacs et marécages.

Les mares sont communément définies comme étant des milieux peu profonds et de petite taille ; les lacs comme étant étendus ou profonds ; et les marécages comme ayant un couvert végétal important. Cependant, les critères et seuils utilisés pour distinguer une mare, un lac et un marécage ne sont pas consensuels dans la littérature. D'après la Convention Ramsar sur les zones humides (4^e édition, 2010), les mares ont une surface inférieure à 10 hectares. D'après Biggs *et al.* (2005), les mares mesurent entre 1 m² et 2 hectares. D'après Richardson *et al.* (2022), les mares ont une surface de moins de 5 hectares, une profondeur de moins de 5 m et moins de 30% de leur surface est recouvertes de végétation.

Dans cette thèse, on s'intéresse aux espèces de crustacés habitant ces milieux et il est important de définir un milieu en fonction du modèle biologique étudié. En effet, l'influence de la taille du milieu aquatique sur les espèces de crustacés retrouvées dans ce milieu semble plutôt indirecte (Fryer, 1985; Bagella *et al.*, 2010). Dans les faits, la taille du milieu va partiellement conditionner le caractère temporaire et la composition en prédateur qui vont à leur tour influencer la distribution des espèces de crustacés (Kuczyńska-Kippen & Pronin, 2018; Coccia *et al.*, 2024). Ainsi, il semble plus cohérent de définir les milieux étudiés par un facteur avec une influence directe sur les crustacés, ici la présence de prédateurs et le caractère temporaire du milieu. Les milieux étudiés ici ont un point commun majeur : l'absence de poissons qui font partie des prédateurs majeurs des espèces étudiées (Brysiewicz *et al.*, 2017; Stefanoudis *et al.*, 2017). En effet, le critère de la taille n'a pas été déterminant quant au choix des localités échantillonnées mais la majorité mesurent moins de 2 hectares. Pour ces raisons et par convenance, il

semble cohérent de rassembler les milieux étudiés sous un terme commun, « mare ». En Nouvelle-Calédonie-Kanaky, de nombreuses mares sont qualifiées de « dolines ». Les dolines sont des figures karstiques qui correspondent classiquement à des dépressions d'effondrement résultant de l'érosion de roches calcaires. Il en existe sur l'archipel notamment à Koumac mais le terme de doline (pseudo-doline) sur l'archipel désigne des dépressions sur substrats ultramafiques dans une configuration de pseudo-karst. Les dolines en NCK sont directement associées avec une végétation environnante qui tranche avec celle du maquis minier.

Les milieux aquatiques lenticulaires peuvent connaître des assèchements partiels ou complets, cycliques ou ponctuels. En fonction de son assèchement ou non, un milieu aquatique peut être qualifié de temporaire ou de permanent. Cependant, pour ces termes aussi, la question de seuil n'est pas résolue. En effet, le caractère temporaire ou permanent d'un milieu se place plutôt sur une échelle continue. Une mare est-elle permanente si plus de 50 % ou 95% de son eau s'assèche annuellement ? Un lac est-il temporaire s'il s'assèche intégralement tous les 3 ans ou tous les 25 ans ? De plus, une surveillance régulière est nécessaire pour déterminer l'assèchement ou non d'un milieu. En effet, la nature temporaire d'un milieu observé à sec est indéniable mais un milieu observé en eau peut être temporaire ou permanent.

La nature temporaire ou non d'un milieu peut être définie par de nombreux facteurs tels que le temps entre deux assèchements (hydropériode), le temps entre deux mises en eaux, le nombre d'assèchements annuels ou encore les variations du volume d'eau, de la profondeur et de la surface dans le temps. Ces facteurs dépendent de variables de l'environnement telles que la pluviométrie, la fréquence des épisodes pluvieux, l'humidité ambiante, le couvert végétal, l'exposition au soleil, le type de substrat ou encore la topographie.

Les caractéristiques d'assèchement des milieux soulèvent d'autres questions telles que la taille d'un milieu temporaire. En utilisant les critères évoqués, une même localité pourrait être qualifiée de mare à son niveau le plus faible, de marécage au début de sa remise en eau et de lac à son niveau le plus élevé.

Comme expliqué plus tôt, les milieux aquatiques ponctuels constituent des modèles d'insularité, les dynamiques de colonisation sont similaires aux îles terrestres et leur assèchement pourrait correspondre aux phénomènes d'immersions ponctuelles des îles décrits auparavant, mais sur un temps plus court

(Whittaker *et al.*, 2023). Ces milieux aquatiques fonctionnent en réseaux plus ou moins connectés entre eux en fonction notamment de la distance, du passage d'animaux ou encore des écoulements d'eau.

Bien qu'occupant une surface terrestre réduite (<1%), les milieux d'eau douce accueillent 10% de la biodiversité mondiale (Strayer & Dudgeon, 2010). Au-delà d'accueillir une grande biodiversité, ces environnements sont vitaux pour de nombreuses espèces terrestres, y compris l'humain (Tickner *et al.*, 2020). En effet, les milieux d'eau douce accueillent des espèces qui sont strictement aquatiques tel que les poissons ou certains crustacés et des espèces amphibiens tel que certaines plantes, coléoptères ou amphibiens mais connaissent également le passage d'espèces terrestres comme les mammifères ou les oiseaux. De nombreuses espèces strictement aquatiques, dont de nombreux crustacés, ont la particularité de produire des structures de résistances comme stratégie de dispersion et de survie aux assèchements (Incagnone *et al.*, 2015). Ces structures déshydratées et dormantes peuvent être produites à différent stade du développement sous forme d'œufs, de graines, de spores ou encore d'éphippie (*i.e.*, un ou plusieurs œufs enveloppés dans une poche constituée par la carapace de la femelle). Elles sont associées à la « diapause » soit la capacité à ralentir le métabolisme et le développement afin de survivre à des conditions environnementales défavorables (Strachan *et al.*, 2015). Les formes de résistances dispersent entre milieux aquatiques par débordement, par l'action du vent ou encore par zoochorie (*i.e.*, dispersion par un vecteur animal) via notamment les amphibiens ou encore les mammifères et oiseaux venant chasser ou boire dans le milieu (Incagnone *et al.*, 2015).

Les mares sont des milieux très hétérogènes, elles présentent par exemple des caractéristiques de turbidité, d'accès à la lumière, de température, de pH, de composition en éléments dissous et de couvert végétal très variées. De ce fait, les organismes strictement aquatiques habitant ces milieux sont très inféodés à ces différentes caractéristiques (Williams, 2007; Tavernini, 2008).

Par ailleurs, les environnements d'eau douce connaissent une vitesse de disparition trois fois plus élevée que les forêts (Gardner & Finlayson, 2020) et les mares sont d'autant plus vulnérables qu'elles constituent des systèmes insulaires, plus sensibles aux perturbations en général (Whittaker *et al.*, 2023). De plus, elles sont parmi les milieux d'eau douce les moins étudiés signifiant que l'évaluation de

leur dégradation pourrait être grandement sous-estimée (Faghihinia *et al.*, 2021; Hill *et al.*, 2021).

V. Les crustacés des mares

Les trois super-classes de crustacés, Allotriocarida, Multicrustacea, et Oligostraca, sont représentées dans les mares temporaires d'eau douce. Il existe une grande diversité de stratégies de diapause chez les crustacés de ces milieux afin de survivre aux assèchements : œufs de résistance, enkystage des individus, ou encore refuge dans des sédiments humides (Strachan *et al.*, 2015). La stratégie de survie à la sécheresse la plus répandue chez les crustacés semble être les œufs de résistance (Incagnone *et al.*, 2015). Certains œufs de résistance peuvent rester viable pendant plusieurs décennies voire plusieurs siècles (Hairston *et al.*, 1995; Sichlau *et al.*, 2011). Selon les espèces, les œufs de résistances vont nécessiter un assèchement obligatoire ou non pour éclore et les individus vont arriver à maturité à des moments différents. Par conséquent, les espèces retrouvées vont être conditionnées notamment par la durée de l'assèchement et la durée de mise en eau des milieux étudiés. En effet, si une espèce ne parvient pas à compléter un cycle de reproduction avant le ré-assèchement du milieu, elle ne pourra pas survivre durablement dans le milieu si cette situation se reproduit fréquemment. Les espèces de crustacés retrouvées dans ces milieux ont des cycles de vie très diversifiés et vont se succéder dans l'environnement à partir de la date de mise en eau. De ce fait, la composition en crustacés retrouvée dans une localité va dépendre également du temps écoulé depuis que la mare est en eau (Boix *et al.*, 2001; Olmo *et al.*, 2016).

Parmi les crustacés retrouvés dans les eaux douces, quelques rares espèces sont largement étudiées comme la Daphnie *Daphnia pulex* étudiée notamment pour son impressionnante plasticité phénotypique en présence de prédateurs (Hammill *et al.*, 2008). Les crustacés d'eau douce occupent de nombreuses places dans les réseaux trophiques et dans les interactions avec d'autres espèces. En effet, certaines espèces de copépodes sont prédatrices comme *Mesocyclops meridianus* (Alekseev, 2020) et d'autres sont parasites comme *Neoergasilus japonicus* (Harada, 1930), certaines espèces de cladocères sont brouteuses comme *Daphnia pulex* (Mandal *et al.*, 2018) et certaines espèces de Spinicaudates sont filtreuses comme *Lynceus insularis* (Olesen *et al.*, 2016). Cependant, ces groupes restent largement méconnus (De Meester *et al.*, 2023; Macedo *et al.*, 2024) malgré leurs nombreuses fonctions écosystémiques et leur potentiel de bioindicateur des milieux (Seminara

et al., 2015; Parmar *et al.*, 2016; Kuczyńska-Kippen & Pronin, 2018). De plus, parmi les 30 espèces des groupes étudiés qui ont été évaluées par l'IUCN (International Union for Conservation of Nature and Natural Resources), 26 sont considérées vulnérables, en danger ou en danger critique (www.iucnredlist.org).

VI. Objectifs et problématiques

Dans le premier chapitre, la diversité spécifique et intra-spécifique des crustacés d'eau douce de NCK est investiguée à travers les problématiques : Quelle diversité est retrouvée chez les Calanoides et les Branchiopodes des mares d'eau douce ? Comment se distribuent les phénomènes d'endémisme chez ces groupes ? Tout d'abord, les taxons présents dans les échantillons prélevés sont identifiés à l'aide d'approches d'identification morphologique et génétique, tandis que ceux précédemment signalés mais n'ayant pas été retrouvés sont répertoriés. Ensuite, la structure de la diversité génétique au sein de ces espèces est analysée à partir des données génétiques recueillies sur certaines des espèces identifiées.

Le second chapitre se consacre à la description et à l'explication de la distribution observée des espèces de crustacés en NCK. D'une part, la distribution de certaines des espèces les plus fréquentes en fonction de variables environnementales est investiguée grâce à des modèles statistiques, en vue de déterminer leur probabilité de présence à l'échelle de l'archipel. D'autre part, les assemblages d'espèces sont recherchés pour déterminer quelles espèces cohabitent dans les mêmes mares. A partir de ces analyses, plusieurs questions sont soulevées : existe-t-il un seul assemblage homogène ou plusieurs assemblages distincts sur l'archipel ? Quelles sont les caractéristiques de la diversité spécifique dans ce ou ces assemblage(s) ? Dans le cas de plusieurs assemblages, les conditions environnementales associées à chaque type d'assemblage sont déterminées par modélisation statistiques. A partir de ces modèles, les régions où la probabilité d'observer chaque assemblage est la plus élevée sont caractérisées.

Le troisième et dernier chapitre s'appuie sur l'organisation et les connaissances acquises dans les deux premiers afin de contribuer à répondre à la question suivante : Comment s'organise le flux d'informations de la description d'une espèce, d'un milieu ou d'un écosystème jusqu'à la compréhension de son écologie pour optimiser sa conservation in fine et rendre ces connaissances utilisables ?

Cette thèse se place dans un contexte riche allant de la diversité morphologique et génétique à la structure des assemblages d'espèces. En plus de chercher à enrichir la connaissance et la compréhension des écosystèmes de la NCK, l'objectif est de rendre ces connaissances pérennes et réutilisables pour le territoire considéré mais aussi plus largement.

Matériel & méthodes

Cette section décrit les méthodes utilisées et le matériel d'étude disponible au cours de la thèse. Elle a pour but de fournir davantage de détails sur les choix méthodologiques, de présenter une version du matériel et méthode homogénéisée à l'échelle du projet de recherche et fournit davantage de détails. De ce fait, certaines parties pourraient se répéter dans les chapitres du manuscrit.

I. Échantillons

I. 1. Repérage des localités et échantillonnage

En préparation des phases de récolte sur le terrain, 2049 localités d'intérêt ont été identifiées sur images satellites avec Google Earth Pro 7.1.2 (fig. M.1). Ici, une localité est définie comme une mare unique et délimitée. Ponctuellement, deux mares peuvent déborder l'une dans l'autre mais sont distinctes la plupart du temps.

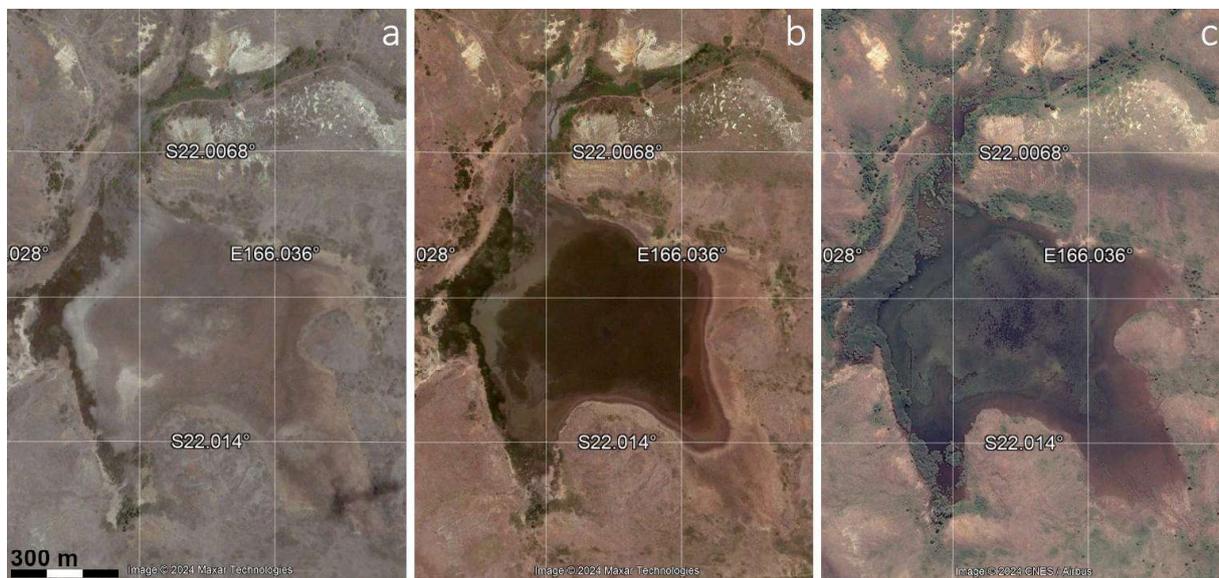


Figure M.1 – Images satellite Google Earth d'une localité temporaire « La Mare aux Canards » (ETNC001) à sec en janvier 2007 (a) et en eau en novembre 2009 (b) et août 2013 (c).

La majorité des localités étudiées dans cette thèse ont été visitées en 2016, 2017 et 2018 pendant les campagnes hydrobiologiques « La Planète Revisitée » en Nouvelle-Calédonie-Kanaky (NCK). Au total, 471 localités ont été visitées durant ces campagnes dont 28 ont été visitées sur deux années (tab. M.1). Malgré le repérage approfondi effectué via les images satellites, l'accès aux zones d'intérêt est parfois difficile ou n'est pas autorisé et les mares les plus petites ou entourées d'un couvert végétal important ne sont pas toujours visibles par satellite. Par conséquent, l'échantillonnage effectué a été essentiellement opportuniste pendant les campagnes « La Planète Revisitée ».

Tableau M.1 – Récapitulatif des campagnes d'échantillonnages « La Planète Revisitée ».

Année	Dates	Localités visitées	Sédiment récoltés	Echantillons au filet
2016	05/11 – 26/11	N = 134 HYNC800-HYNC932	N = 90	N = 29
2017	22/11 – 08/12	N = 249 HYNC2800-HYNC2899 HYNC2600-HYNC2748	N = 213	N = 36
2018	04/06 – 17/06	N = 116 HYNC4000-HYNC4115	N = 47	N = 67

Entre 2020 et 2023, le bureau d'études « ETHYCO » et l'association « Vies d'Ô douces » ont fait parvenir des échantillons de 63 nouvelles localités et de 6 qui avaient déjà été échantillonnées.

Durant ces phases de terrain, deux types d'échantillons ont été récoltés, des échantillons prélevés dans les mares en eau à l'aide d'un filet à plancton et des récoltes de sédiments prélevés dans les mares sèches (tab. M.1 ; fig. M.2). Le filet à plancton utilisé a une maille en nylon de 200 µm avec un filtre à l'extrémité, un diamètre de 25 cm et mesure 55 cm de long. Selon la quantité d'eau, le filet à plancton est passé depuis le bord de la mare ou charrié en nageant dans la mare. Le contenu du filtre est ensuite transféré dans un tube pour être préservé dans de l'éthanol à 80-90% et conservé dans un réfrigérateur à 4°C le plus tôt possible. Selon la taille de la mare, le sédiment est prélevé en un ou plusieurs points sur environ 3-4 cm de profondeur et 100-400 cm². Il est ensuite stocké dans des sacs de congélation soigneusement fermés et étiquetés.

Les échantillons sont transportés par avion (arrivée en quelques semaines) ou par bateau via des conteneurs pouvant rester stockés au port pour des durées variables (arrivée en plusieurs mois).

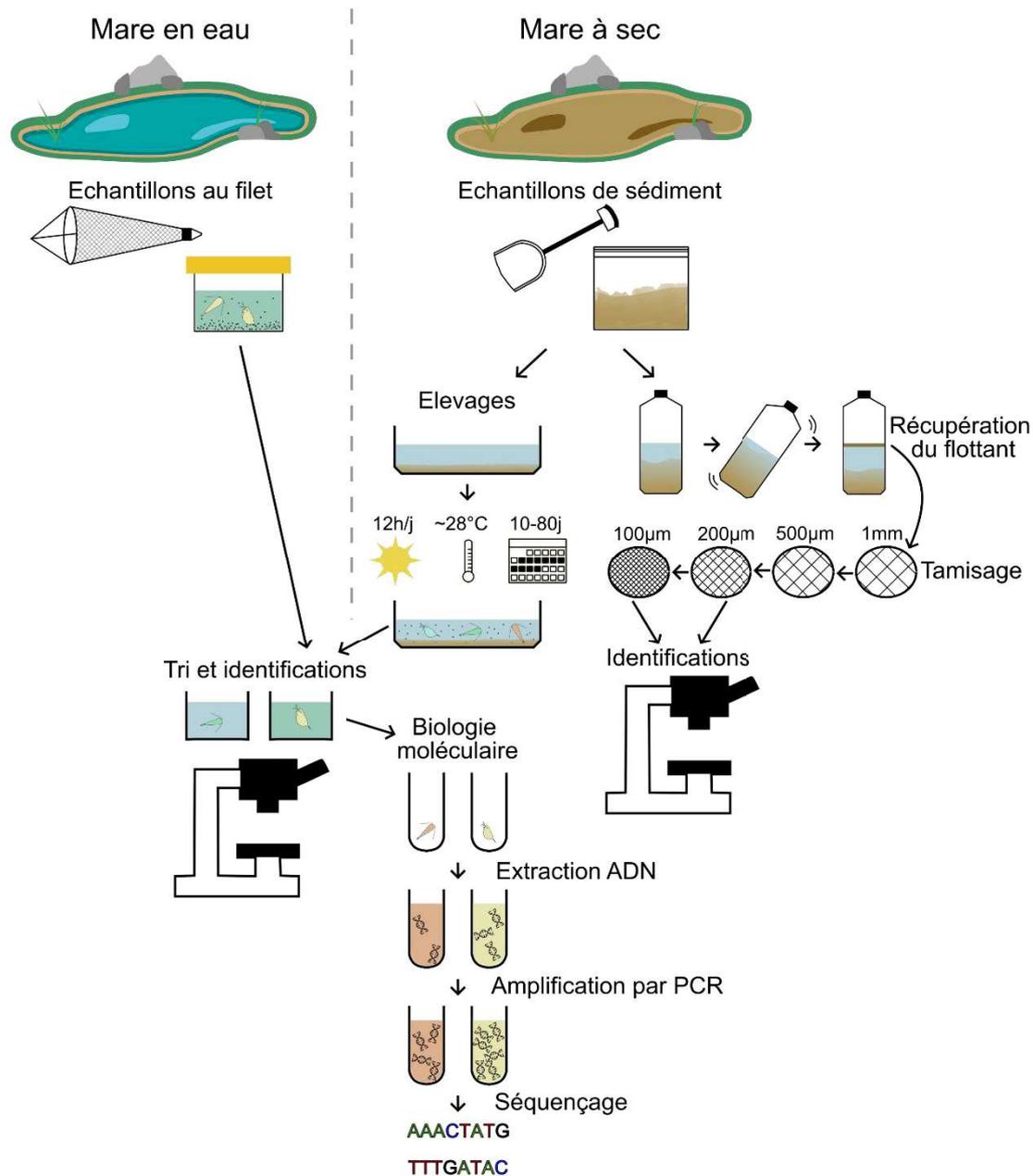


Figure M.2 – Résumé graphique des méthodes utilisées en fonction des types d'échantillons.

I. 2. Traitement des échantillons de sédiment

I. 2. i. Tri et identification des formes de résistance

Après la phase d'échantillonnage, les sédiments de mares sont triés afin de récupérer les formes de résistances qui y sont stockées. La première étape consiste à placer une masse de terre précise (50-200 g en fonction de la quantité de terre à disposition) dans une bouteille plastique rigide à goulot étroit (*e.g.*, Badoit® 1L). L'eau est ajoutée dans la bouteille en la remplissant à moitié. L'ensemble est secoué,

plusieurs fois durant 1 heure afin de décoller les œufs des sédiments (en mimant un orage violent). Les œufs étudiés flottent au milieu d'autres formes de résistance et de nombreux débris. Les structures sont séparées par filtration sur des tamis de 100, 200, 500 micromètres et 1 mm. L'identification des œufs se fait sur les tamis de 100 et 200 micromètres (fig. M.2).

I. 2. ii. Élevages

Les sédiments prélevés sont utilisés pour la mise en place d'élevages. En effet, comme expliqué précédemment, les crustacés étudiés produisent des formes de résistances stockées dans les sédiments qui leurs permettent de maintenir les populations d'une mise en eau à l'autre. Contrairement aux échantillonnages ponctuels qui donnent accès aux espèces présentes à un instant donné, cette démarche permet d'avoir une visibilité sur les espèces qui se succèdent dans le temps.

Les sédiments sont réhydratés dans des bacs avec de l'eau de pluie filtrée. Les élevages sont installés avec des pompes à air, sous exposition solaire ou des lampes spectre solaire avec minuteries de 12h et maintenus dans un environnement avec une température ambiante moyenne de 28°C. Les élevages sont maintenus entre 10 et 80 jours (fig. M.2). En fonction des espèces observées ou non, des prélèvements réguliers sont effectués pendant toute la durée de l'élevage afin d'assurer la détection d'un maximum d'espèces. Lorsque l'élevage est arrêté, la totalité de l'eau est filtrée et le sédiment est laissé à sécher pour être conservé ou remis en élevage ultérieurement.

I. 3. Traitement des échantillons au filet et issus des élevages

I. 3. i. Tri et identification des espèces

Les échantillons au filet et issus d'élevages contiennent les mues, œufs, stades larvaires et individus adultes et juvéniles de crustacés. Ils contiennent également des individus de mollusques, insectes, amphibiens, éponges, vers, cnidaires, rotifères à divers stades de développements, des graines, des débris organiques et inorganiques.

Chaque échantillon est trié séparément sous loupe binoculaire dans un premier temps. Les individus non-crustacés sont identifiés au niveau du phylum ou de la classe et transférés dans des tubes en verre ou en plastique. Les débris organiques et inorganiques sont mis de côté pendant l'identification des individus

de crustacés pour être revérifiés avant d'être jetés. Les organismes appartenant au groupe des crustacés sont identifiés au niveau taxonomique le plus précis possible sans dissection sous loupe binoculaire (50×) et sous microscope (500×) si nécessaire. Pour chaque échantillon, tous les individus identifiés d'un même taxon sont rassemblés dans un tube en verre avec bouchon à vis et étiqueté (tab. M.2). L'éthanol à 80-90% est renouvelé dans chaque tube. Afin d'éviter l'évaporation de l'alcool, tous les tubes d'un même taxon sont rassemblés dans un pot rempli d'éthanol en attendant la mise en collection.

Tableau M.2 - Contenu des étiquettes apposées aux tubes triés

Echantillon filet		Echantillon élevage	
Obligatoire	Optionnel	Obligatoire	Optionnel
Nom d'espèce / taxon	Identifiant d'échantillon	Nom d'espèce / taxon	Date d'échantillonnage à la localité
Identifiant de localité		Identifiant d'élevage	Date de début d'élevage
Coordonnées de la localité		Coordonnées de la localité	
Date d'échantillonnage		Date d'échantillonnage	
Collecteur		Collecteur du sédiment	
		Identifiant d'échantillon	

Les clés de Stingelin (1915), Smirnov & Timms (1983), Timms (1985), El-moor Loureiro (1997), Korovchinsky (2001), et Quinlan & Bayly (2017) ont été utilisées pour l'identification des espèces. D'autres ressources décrivant des espèces non répertoriées dans ces clés telles que Defaye (1998) et Olesen *et al.* (2016) sont utilisées en complément.

Le premier tri a été réalisé principalement par sept étudiantes et étudiants en stage ou volontariat, Lucie Lenglet, Gabrielle Ourous, Félix Navarro, Tancrede Bourgeois, Enzo Beule, Agathe Moulin et Angel Arnold. Afin d'homogénéiser les identifications, chaque tube d'individus isolés est repris et réétiqueté si nécessaire.

Parmi les groupes de crustacés identifiés dans les échantillons, certains taxons n'ont pas pu être étudiés par manque de temps. En effet, une partie des espèces d'Ostracodes, d'Harpacticoides et de Cyclopoïdes nécessitent d'être disséquées pour être formellement identifiées. Les individus de ces groupes mesurent pour la plupart moins de 3 mm, leur dissection est très technique et particulière à chaque groupe. Ce travail prend énormément de temps et le risque d'endommagement des individus est élevé. La plupart des espèces de ces groupes étant retrouvés en faible abondance dans les échantillons, l'endommagement de quelques individus

signifierait une potentielle impossibilité d'identification future. Il a donc été choisi de ne pas étudier la totalité des groupes afin de se concentrer sur des taxons ayant une délimitation taxonomique adéquate dans le temps de ce projet. En somme, les identifications se sont concentrées sur les individus de Branchiopodes et de Calanoides.

I. 3. ii. Délimitation des assemblages et modélisation

A partir des échantillons triés en totalité, des assemblages d'espèces ont été délimités et des modèles statistiques ont permis de déterminer les facteurs environnementaux principaux influant sur la répartition de ces assemblages sur l'île. Les méthodes et les résultats obtenus sont détaillés dans le chapitre II.

I. 3. iii. Extraction de l'ADN

La diversité génétique a été étudiée chez le genre de Copépode *Boeckella* et chez les genres de Branchiopodes *Lynceus*, *Eulimnadia*, *Streptocephalus*, *Triops* et *Latonopsis*. L'ADN est extrait des tissus à partir d'un prélèvement d'une partie d'un individu, d'un individu complet ou de plusieurs individus selon la taille des organismes et la méthode de séquençage. Les extractions ont été réalisées avec le robot Macherey-Nagel NucleoSpin® 96 avec le NucleoSpin® 96 Tissue Core Kit ou manuellement avec le QIAamp® DNA Micro Kit. Les protocoles détaillés pour l'extraction de l'ADN sont disponibles sur [protocols.io](https://dx.doi.org/10.17504/protocols.io.36wgqnm9ygk5/v1) : <https://dx.doi.org/10.17504/protocols.io.36wgqnm9ygk5/v1> ; <https://dx.doi.org/10.17504/protocols.io.j8nlk4x7wg5r/v2>. Au total, 482 extractions d'ADN ont été réalisées (tab. M.3).

Tableau M.3 - Bilan des extractions ADN et séquençages réussis pour chaque genre étudié

Classe	Genre	Extraits ADN	Séquençages réussis		
			COI Sanger	COI pool	NGS
Branchiopoda	<i>Streptocephalus</i>	35	28	0	1
	<i>Triops</i>	7	5	0	2
	<i>Latonopsis</i>	43	1	6	3
	<i>Lynceus</i>	87	73	0	7
	<i>Eulimnadia</i>	85	32	2	0
Copepoda	<i>Boeckella</i>	225	160	1	6
Total		482	299	9	19

I. 3. iv. Séquençage du barcode

Le séquençage Sanger ou le séquençage en pool d'amplicons est utilisé pour obtenir un fragment de 658 pb (paires de bases) du gène mitochondrial de la Cytochrome Oxydase I (COI). Le gène du COI est bien connu notamment du fait de son utilisation pour le barcoding et metabarcoding chez les métazoaires. Par sa vitesse de divergence, ce gène permet d'évaluer en partie la diversité génétique au sein d'une même espèce mais aussi d'apporter des éléments pour la distinction génétique entre espèces.

Les échantillons d'ADN extraits d'un individu unique sont soumis au processus de réaction en chaîne par polymérase (PCR ; polymerase chain reaction). Ce processus d'amplification permet d'augmenter la concentration d'un fragment d'ADN donné, ici un fragment du gène COI. A partir de l'ADN extrait, de nombreuses copies du fragment souhaité sont produites en chaîne à l'aide notamment de deux courts brins d'ADN (amorces) délimitant le début et la fin du fragment visé et d'une enzyme polymérase thermostable (Taq polymérase). L'ensemble des réactifs est soumis à plusieurs cycles de températures permettant la réaction en chaîne (fig. M.2). Afin de vérifier le bon fonctionnement (ou non) des PCR, les produits de la réaction sont soumis à une électrophorèse en gel d'agarose. Les protocoles détaillés pour les PCR et les électrophorèses sont disponibles sur protocols.io : <https://dx.doi.org/10.17504/protocols.io.n92ld8q57v5b/v1> ; <https://dx.doi.org/10.17504/protocols.io.261ge4y4jv47/v3>.

D'une part, une fraction des produits de PCR sont envoyés tels quels au laboratoire Eurofins Scientific afin d'effectuer le séquençage Sanger. Les résultats de ces séquençages sont nettoyés avec le flux automatisé (workflow) que j'ai développé sur la plateforme Galaxy et en accès libre : https://usegalaxy.fr/u/coline_royaux/w/sanger1--from-ab1-to-aligned-fasta-imported-from-uploaded-file. Chaque séquence obtenue est ensuite vérifiée et corrigée à la main si nécessaire.

D'autre part, certains produits de PCR de qualité médiocre, sont préparés pour être envoyés en séquençage en pool d'amplicons. Le pool d'amplicon est une méthode de séquençage à bas coût utilisant la technologie du séquençage de nouvelle génération (NGS ; next-generation sequencing). Cette méthode consiste en une mise en commun des séquençages par le dépôt de produits de PCR de plusieurs origines dans un même tube à séquencer. Par exemple, des produits de PCR d'une

espèce de plante, d'une espèce de poisson et d'une espèce de crustacé peuvent être rassemblés dans un même tube pour être séquencés. En sortie du séquençage, le fichier associé au tube contiendra des fragments de séquence des trois espèces qui pourront être différenciées par cartographie à l'aide de séquences de référence. Il est donc important de ne pas rassembler les produits de PCR d'individus appartenant à des taxons génétiquement proches afin d'éviter une confusion entre les séquences. Les séquençages NGS pour les pools d'amplicons ont été réalisés à l'Institut du Cerveau et de la Moelle épinière (ICM, Hôpital Pitié Salpêtrière). Les séquences obtenues ont été cartographiées avec Geneious 11.1.4 et l'outil Bowtie2 2.4.2 sur la plateforme Galaxy (usegalaxy.eu).

I. 3. v. Séquençage shotgun de nouvelle génération (NGS)

Le NGS est utilisé pour obtenir un plus large spectre de gènes pouvant aller jusqu'aux génomes mitochondrial et nucléaire complets. Or, il est rare d'aller jusqu'aux génomes complets qui nécessitent un investissement (*i.e.*, horaire, monétaire, en stockage, en infrastructure) très important et un intérêt limité pour la taxonomie. Pour les groupes étudiés, l'objectif est d'obtenir les gènes nucléaires du 28S et la séquence complète du gène mitochondrial du COI et du 16S. Ces trois gènes sont ici utilisés pour renseigner sur les liens phylogénétiques entre et au sein des espèces d'un même genre ou de même famille.

Le cas général est d'utiliser un individu unique partiel ou complet par extrait d'ADN. Dans le cas d'individus de très petite taille (< 2 mm) ou endommagés, il est parfois nécessaire d'extraire l'ADN de plusieurs individus dans un même tube afin d'obtenir une concentration en ADN suffisamment élevée afin d'obtenir le seuil requis en quantité d'ADN pour le séquençage NGS.

Les banques ont été préparées avec un kit Nextera XT (Illumina) et séquencées à l'ICM. L'assemblage des séquences et la correction des erreurs de séquençage a été faite via Geneious 11.1.4.

I. 3. vi. Analyses phylogénétiques

Des arbres phylogénétiques ont été calculés pour les genres *Triops*, *Eulimnadia*, *Boeckella* et *Lynceus*.

Après alignement des séquences et vérification de l'absence de codon stop dans la phase de lecture (dans le cas du COI), les séquences sont soumises à un

Gblock sur la plateforme LIRMM (Dereeper *et al.*, 2008). Les estimations du modèle d'évolution ont été calculées pour chaque position de codon avec « Partition finder 2 » pour les analyses bayésiennes sur PhyloSuite 1.2.3 (Lanfear *et al.*, 2016; Zhang *et al.*, 2020; Xiang *et al.*, 2023). L'analyse bayésienne est calculée avec MrBayes 3.2.7 (Ronquist *et al.*, 2012) avec les trois meilleurs modèles pour chaque position de codon. Les arbres phylogénétiques finaux ont ensuite été formaté sur FigTree 1.4.4 (<https://tree.bio.ed.ac.uk/software/figtree>) et Inkscape 1.2.2.

I. 3. vii. Structuration génétique des populations

Un haplotype est défini comme une version d'un gène donné. Deux haplotypes se distinguent par une ou plusieurs mutations dans la séquence de nucléotides de ce gène. Dans cette thèse, on observe uniquement des mutations par substitution c'est-à-dire le remplacement ponctuel d'un nucléotide par un autre (par exemple, Adénine remplacé par Thymine).

La structure génétique des populations a été étudiée pour chacune des espèces pour lesquelles plus de trois versions de gène (haplotype) différentes ont été retrouvées pour le fragment du gène COI. Le test exact de différenciation des populations compare la composition en haplotype de chaque localité et détermine si celle-ci est significativement différente en fonction des localités. Ce test ne prend pas en compte la distance génétique entre chaque haplotype ni la position des mutations qui les distingue. L'analyse de variance moléculaire (AMOVA) compare les distances génétiques observées au sein des localités échantillonnées avec les distances génétiques observées entre les localités. Le réseau d'haplotype est une manière de visualiser les liens entre les haplotypes retrouvés chez une espèce ou une population donnée. Un réseau d'haplotype peut-être calculé par différents algorithmes qui vont chercher à résoudre les relations entre les haplotypes en fonction des mutations qui les différencient. Certains algorithmes se basent uniquement sur les distances entre les séquences (*e.g.*, Minimum Spanning Tree, Minimum Spanning Network) et ne prennent en compte ni la position ni la nature des mutations. A la différence de ces derniers, les algorithmes basés sur les séquences des haplotypes peuvent ponctuellement inférer l'existence d'haplotypes hypothétiques n'ayant pas été échantillonnés (*e.g.*, parcimonie statistique, Median-Joining Network). Du fait de la nature de l'échantillonnage réalisé dans cette thèse, les algorithmes basés sur les séquences ont semblé plus pertinents pour représenter la structure génétique des populations. En effet, les individus séquencés

peuvent avoir été obtenus à travers des élevages ou directement dans l'environnement et certaines zones de l'île n'ont pas pu être accédées pour l'échantillonnage. Il semble donc probable qu'il existe des haplotypes manquants mais nécessaire à l'articulation des réseaux. Les réseaux d'haplotypes ne parviennent pas toujours à résoudre les liens entre les différents haplotypes, surtout quand un grand nombre de mutation existe entre les haplotypes. En effet, il arrive que le nombre de mutations réel observé entre deux haplotypes ne soit pas identique au nombre de mutations représentées sur le réseau d'haplotype. Afin de représenter ce problème, les algorithmes calculent des liens alternatifs. Cependant, ces liens alternatifs ont tendance à rendre les réseaux d'haplotypes illisibles, ils ne sont donc pas représentés.

Les données de séquençage ont été préparées pour l'analyse de la structure des populations avec un flux automatisé (workflow) développé sur la plateforme Galaxy et en accès libre : https://usegalaxy.eu/u/coline_royaux/w/sanger-5--from-hand-corrected-haplotypes-table-to-population-haplotypes-map-and-euclidean-distance-matrix.

Les tests exacts de différenciation de population et les AMOVA ont été réalisés respectivement avec les packages `mmod` 1.3.3 (Winter, 2012) et `ade4` 1.7-22 (Dray & Dufour, 2007) du logiciel R 4.2.2. Les réseaux d'haplotypes ont été calculés par les algorithmes de Median-Joining (Bandelt *et al.*, 1999) et de parcimonie statistique (Templeton *et al.*, 1992) avec le package `pegas` 1.3 (Paradis, 2010) du logiciel R 4.2.2. Les scripts correspondant à ces analyses sont disponibles sur GitHub (https://github.com/ColineRoyaux/population_genetics).

II. Réutilisation des connaissances : exploration des moyens de pérennisation

Par « connaissances », est entendu ici tous les éléments du projet de recherche pouvant aller du cadre de travail aux résultats et conclusions finales. A travers la pérennisation de ces connaissances, l'objectif est aussi d'obtenir un ensemble d'éléments utilisables et réutilisables par une diversité d'acteurs et pour un large spectre de cas d'études. L'essentiel de la méthode pour explorer les leviers disponibles a été de mener cette thèse en effectuant une veille sur les concepts et solutions existantes et accessibles à la recherche pour la mise à disposition pérenne des connaissances générées. L'idée de partir de l'existant est essentiel pour apporter des éléments pertinents, concrets et utilisables en l'état.

II. 1. Définition des bonnes pratiques

La reproductibilité est définie par Plessier (2018) comme la capacité d'obtenir avec un même protocole des résultats similaires dans une étude indépendante (soit une équipe et une configuration expérimentale différente). La reproductibilité des analyses (*i.e.*, reproductibilité computationnelle), telle que défini par Cohen-Boulakia *et al.* (2017), représente la capacité d'analyses distinctes à arriver à des conclusions identiques.

Les principes FAIR ont été définis à l'origine pour le partage des données et métadonnées mais sont valables également pour d'autres objets numériques tels que les logiciels et procédures analytiques (Lamprecht *et al.*, 2019; Goble *et al.*, 2020). Les termes de l'acronyme sont ; « Facile à trouver » pour les humains comme les ordinateurs ; « Accessible » avec une démarche définie pour y accéder ; « Interopérable » afin d'être utilisable en interaction avec d'autres données ou analyses ; et « Réutilisable » dans un contexte identique ou non (Wilkinson *et al.*, 2016).

Les principes FAIR ont été complétés par les principes CARE pour les données collectées sur les territoires ou les populations indigènes (Carroll *et al.*, 2020). Ils sont applicables plus largement afin d'assurer que les données et objets numériques partagés puissent être utilisables par et sous la gouvernance des territoires et populations considérées (Carroll *et al.*, 2021). En effet, dans l'objectif de passer de la recherche à l'action, il est important de prendre en compte les populations qui pourront s'en emparer et y prendre part. Les termes de l'acronymes sont ; « bienfait Collectif » par le développement collaboratif d'issues équitables ; « Autorité de contrôle » par la reconnaissance des droits et intérêts de populations concernées et la possibilité de gouvernance de ces populations sur les objets numériques en question ; « Responsabilité » par l'entretien de relations positives avec les populations des territoires et l'expansion du potentiel d'action de ces populations ; et « Ethiques » par la minimisation de potentiels effets délétères dans les utilisations présentes et futures des objets numériques.

Les 5 V du « big data » définissent les caractéristiques du « big data », soit l'augmentation exponentielle de la quantité de données produites et analysées générant des difficultés logistiques. Ils permettent d'adresser et d'anticiper les problématiques liées au « big data ». La « Vitesse » se réfère à la vitesse de production et de traitement des données ; le « Volume » se réfère à la quantité de

données ; la « Valeur » représente l'apport et les bénéfices de l'acquisition et l'analyse des données ; la « Variété » représente la diversité de types de données ; la « Véracité » représente la qualité des données à travers leur précision, leur crédibilité ou encore l'éthique de leur collecte et de leur analyse.

II. 2. Le cycle de vie de la recherche

Afin d'assurer un regard global sur la totalité du projet de recherche, il est possible de s'appuyer sur la notion de cycle de vie de la recherche. Le cycle de vie de la recherche décrit les étapes d'un projet de recherche qui s'inscrivent dans une boucle partant de la planification du projet à la publication des résultats (fig. M.3).

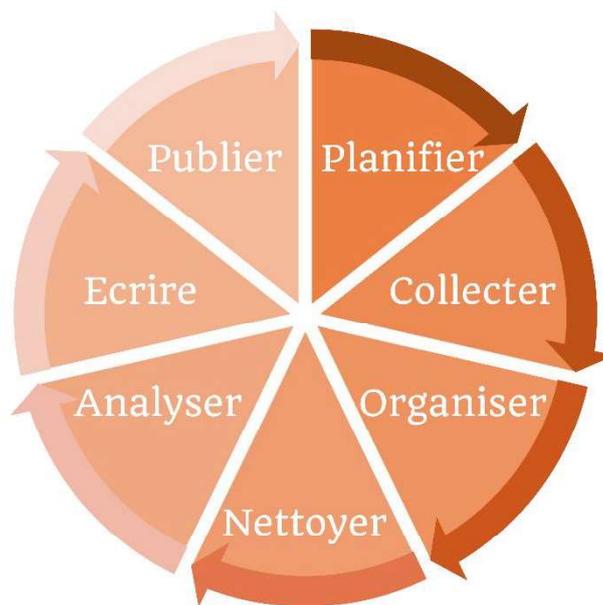


Figure M.3 – Le cycle de vie de la recherche (Borgman, 2020)

II. 3. Le Plan de Gestion de Données

Le Plan de Gestion de Données (PGD ; Data Management Plan ; DMP) est un document décrivant un projet de recherche dans le détail avec une attention particulière aux données produites. Ce document peut être requis par les institutions ou lors d'une demande de financement et contient généralement les éléments suivants :

- Mots clés
- Description du projet de recherche
- Description du financement
- Contacts
- Contributeurs

- Description des données
- Distribution des données
- Coûts de gestion des données
- Licence
- Hébergeur des données
- Sécurité et sensibilité des données
- Ressources techniques utilisées

Au-delà d'un document réglementaire, le PGD contient les éléments fondamentaux auxquels réfléchir en amont de la construction d'un projet de recherche. Pour les personnes porteuses du projet de recherche et rédigeant le PGD, il permet de structurer les données de façon adaptée et de s'assurer leur cohérence avec le projet de recherche en amont de la récolte de données. Il sera ensuite possible de revenir sur le document pour vérifier certaines informations ou créer de nouvelles versions si nécessaire. Le versionnage permet de conserver une forme de flexibilité du projet de recherche tout en documentant ses évolutions. Pour les personnes contribuant au projet de recherche, il permet notamment de mettre en avant leurs contributions et de saisir les nombreux détails inhérents au projet de recherche. Pour les externes au projet de recherche, il est un document de référence permettant de savoir comment le projet a été construit, où aller chercher les données ou encore sous quelle licence elles sont partagées.

De nombreuses ressources existent pour accompagner la rédaction d'un plan de gestion de données complet et utile telles que la plateforme DoRANum, l'article écrit par Michener (2015) stipulant dix règles pour la rédaction d'un PGD et l'article de Gajbe *et al.* (2021) répertoriant et évaluant 14 outils de rédaction d'un PGD.

II. 4. Le partage des protocoles et données

Bien que les méthodes soient toujours associées aux résultats dans les publications scientifiques, les limitations de taille des articles entraînent bien souvent une impossibilité de reproduire dans le détail les méthodes décrites. De plus, la plupart des revues ayant un système d'accès libre payant, les méthodes ne sont pas toujours accessibles. La plateforme protocols.io est une solution permettant de décrire en détails les protocoles utilisés. La description des protocoles est facilitée par l'interface qui permet d'organiser les protocoles en étapes et sous-étapes, de les annoter et de communiquer les mesures de sécurité à appliquer. Chaque personne

ayant contribué au développement ou au test du protocole peut être référencée et toute personne ayant testé le protocole après le partage du protocole peut indiquer s'il a fonctionné ou non (fig. M.4). Ainsi, les protocoles sont accessibles et faciles à trouver directement par leur nom et les mots clés associés. Chaque protocole partagé est associé à un identifiant d'objet numérique (digital object identifier ; DOI).

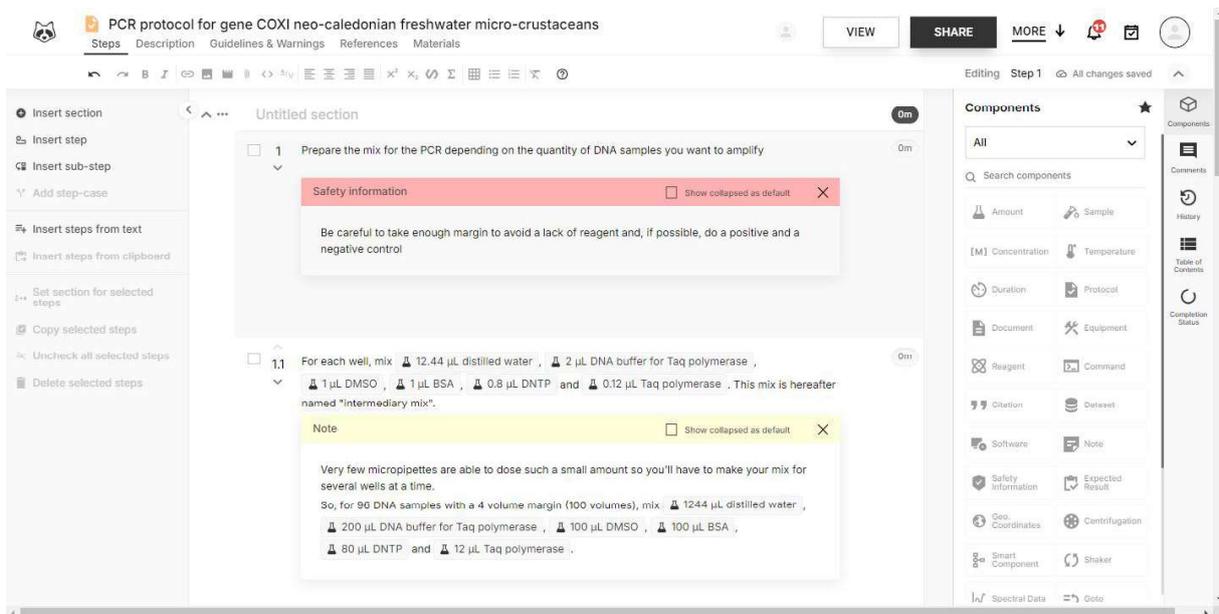


Figure M.4 – interface d'édition des protocoles sur la plateforme www.protocols.io.

La tenue d'un cahier de laboratoire est une bonne pratique recommandée pour avoir un suivi quotidien des expérimentations réalisées. Le cahier de laboratoire rédigé dans le cadre de cette thèse sera mis à disposition sur la plateforme d'archives ouvertes HAL.

Le partage des données est un sujet largement traité dans les textes réglementaires, les ressources institutionnelles et la littérature scientifique et grise. La diversité des types de données, des modes d'acquisition et de curation, des formats de données et des utilisations des données engendre une grande complexité quant à la manière de partager les données. Les nombreux standards et les recommandations existantes seront toujours sujets à des adaptations liées aux particularités des projets de recherches.

La description des métadonnées (*i.e.*, données sur les données) est cruciale à l'accessibilité, la transparence, la réutilisation et la pérennité des données. Les métadonnées sont un levier majeur pour atteindre un niveau suffisant dans les principes FAIR (Wilkinson *et al.*, 2016). Une partie des métadonnées est

notamment écrite dans le Plan de Gestion de Données (*e.g.*, description du projet de recherche, des collaborateurs, financements). La description détaillée des champs (*e.g.*, le contenu des différents fichiers, leur organisation, les unités de mesures utilisées, le format des colonnes de fichiers tabulés) permet dans un premier temps de faire savoir aux personnes travaillant sur le projet le contenu des données et comment les alimenter (Michener, 2015). La diffusion des métadonnées dans un standard lisible par les ordinateurs est largement recommandée (Wilkinson *et al.*, 2016). Michener *et al.* (1997) ont créé l'Ecological Metadata Language (EML), un standard de métadonnées dédié à l'écologie mais utilisable plus largement pour les sciences de l'environnement. L'EML est utilisé par de nombreux standards de données comme le Darwin core, inspire ou encore la norme iso19115.

Dans le cas de cette thèse, il a été choisi de partager en licence ouverte (CC-by 4.0) d'une part les données sans nettoyage, telles que stockées sur les appareils et services de stockage distant (cloud) et, d'autre part, les données nettoyées telles qu'utilisées en entrée des flux analytiques. De cette façon, il est possible de comparer les données dites « brutes » avec les données utilisées en entrée des analyses pour avoir connaissance des corrections et filtres appliqués. Ceci en conservant une reproduction des analyses facilitée par la disponibilité d'une version des données filtrée et prête à être analysées. Les données sont partagées à travers la plateforme institutionnelle data.InDoRES. Les données de séquençage sont partagées également à travers la plateforme d'archivage GenBank du « National Center for Biotechnology Information » (NCBI).

II. 5. La mise à disposition des workflows de nettoyage des données et des analyses

Un « workflow » est un enchaînement de plusieurs étapes analytiques distinctes. Il débute souvent par le formatage et nettoyage des données. Par exemple, les étapes analytiques peuvent être des calculs, modélisations ou représentations graphiques.

Les workflows de nettoyage des données et d'analyses sont tous partagés à travers la plateforme Galaxy (usegalaxy.eu ; usegalaxy.fr). Lorsque ceux-ci ne sont pas disponibles sur la plateforme Galaxy, ils sont partagés sous forme de scripts à travers la plateforme GitHub (github.com/ColineRoyaux) et des représentations schématisées des workflows sont jointes aux publications. Les méthodes de partage

des processus analytiques ont été développées au cours de la thèse et sont décrites en profondeur dans le chapitre III.

II. 6. L'écriture et les publications scientifiques

Tous les articles sont rendus accessibles à travers la plateforme d'archives ouvertes HAL. Les manuscrits non-révisés et en cours de révision (pre-print) des articles suivants sont et seront déposés sur la plateforme d'archives dédiée aux travaux d'écologie et d'évolution EcoEvoRxiv (ecoevorxiv.org).

Le second article écrit dans cette thèse a été révisé via le processus Peer Community In (PCI). PCI est un processus de révision par les pairs gratuit et en accès ouvert. L'organisation est à ce jour composée de 17 PCI disciplinaires (*e.g.*, Ecology, Evolutionary Biology, Neuroscience). Un PCI disciplinaire est composé d'un comité d'experts reconnus dans la discipline et de « recommenders » qui ont un rôle similaire aux éditeurs des journaux traditionnels. Lors de la soumission d'un manuscrit à un PCI disciplinaire, les auteurs ont le choix parmi une liste de « recommenders » qui vont devoir décider de gérer ou non la révision du manuscrit. Si un « recommender » accepte, il aura la charge de composer le comité de révision du manuscrit et de gérer les différents tours de révisions. Lorsque l'article est révisé et accepté, il est dit « recommandé par PCI » et la totalité du processus de révision (*i.e.*, commentaires des pairs et réponses) est en accès libre. Par la suite, il pourra éventuellement être proposé dans des journaux traditionnels ou au Peer Community Journal. Ce processus permet la publication ouverte de travaux de recherche révisés par les pairs sans engager de frais pour les lecteurs ni les auteurs et instituts.

II. 7. Collaborations et communications

Afin de garantir une recherche utile et utilisable pour une diversité d'acteurs (*e.g.*, citoyens, associations, bureaux d'étude, décideurs, établissements publics), il est fondamental de construire des collaborations extra-recherche et de travailler à du matériel dédié à des publics non-chercheurs. Depuis les campagnes d'échantillonnage en NCK, des collaborations pérennes ont été mises en place avec l'association Vies d'Ô douces et le bureau d'étude Etyco. Le matériel dédié aux publics non-chercheurs peut prendre diverses formes à discuter avec ces publics selon leurs besoins. Dans le but d'étendre les collaborations et d'échanger sur le matériel pouvant être utile aux acteurs du territoire, plusieurs demandes de financement pour des missions sur l'archipel ont été faites.

Chapitre I

La diversité de crustacés dans les mares d'eau
douce de Nouvelle-Calédonie-Kanaky

I. Résumé

Dans ce premier chapitre, l'objectif est de faire un état des lieux de la diversité chez les crustacés des eaux douces stagnantes (*i.e.* mares) de Nouvelle-Calédonie-Kanaky (NCK). La diversité y est étudiée des points de vue morphologique et génétique.

Dans un premier temps, la diversité spécifique chez les crustacés des eaux douces stagnantes de NCK a été investiguée à travers le tri des échantillons récoltés au filet. Avant le tri des échantillons, 50 espèces de crustacés étaient connues des mares de l'archipel (tab. 1.1).

Les identifications effectuées pendant cette thèse ont permis d'ajouter 23 espèces qui n'avaient pas encore été signalées sur l'archipel. Ces résultats ont également permis de confirmer la nécessité de clarifier le statut taxonomique de certains cladocères dont *Ilyocryptus* et les Chydoridae. En effet, ces animaux correspondent à des espèces définies morphologiquement ayant une répartition cosmopolite. De fait, elles sont suspectées de constituer des complexes d'espèces cryptiques (*i.e.*, espèces dissociées d'un point de vue génétique, reproductif, écologique ou géographique mais indissociables par la morphologie). De plus, il est à noter que la distribution des espèces *Triops intermedius*, *Eulimnadia sp. A*, et *Streptocephalus archeri* est plus large que celle décrite dans ce chapitre. En effet, les analyses des échantillons de sédiment ont permis d'identifier leurs œufs de résistances dans de nombreuses localités supplémentaires présentées dans le second chapitre (encadré B).

Dans un second temps, le travail s'est concentré sur le genre de copépodes *Boeckella*. Le premier article écrit dans le cadre de cette thèse s'est concentré sur la description d'une nouvelle espèce, nommée *B. sibleti*, associée à la diversité génétique de la seule autre espèce connue de ce groupe dans l'archipel, *B. spinogibba*, décrite par Danielle Defaye (1998). Dans cet article, la remarquable diversité génétique retrouvée chez *B. spinogibba* et la proximité géographique de ses populations avec la population type de *B. sibleti* semble indiquer un phénomène d'endémisme à petite échelle. Cette hypothèse est confirmée par l'article suivant qui signale quatre nouvelles populations de *Boeckella* génétiquement et morphologiquement distinctes de *B. spinogibba* et *B. sibleti*. Dans ces deux articles, l'arbre phylogénétique du genre *Boeckella* à l'échelle mondiale est complété par les données génétique des populations de *Boeckella* retrouvées en NCK.

Tableau 1.1 - Espèces de Branchiopodes, Copépodes et Ostracodes répertoriées en NCK dans les eaux douces stagnantes (**en gras** : espèces endémiques).

Classe	Ordre	Espèce	Référence	
Branchiopoda	Notostraca	<i>Triops longicaudatus intermedius</i>	Longhurst, 1955	
	Anomopoda	<i>Chydorus eurynotus</i>	Timms, 1985	
		<i>Ephemeroporus barroisi</i>	Timms, 1985	
		<i>Dunhevedia crassa</i>	Timms, 1985	
		<i>Coronatella rectangula</i>	Timms, 1985	
		<i>Alonella clathratula</i>	Timms, 1985	
		<i>Karualona karua</i>	Timms, 1985	
		<i>Armatalona macrocopa</i>	Timms, 1985	
		<i>Simocephalus exspinosus australiensis</i>	Timms, 1985	
		<i>Simocephalus acutirostratus</i>	Timms, 1985	
		<i>Ceriodaphnia rigaudi</i>	Stingelin, 1915	
		<i>Daphnia cephalata</i>	Timms, 1985	
		<i>Daphnia longicephala</i>	Timms, 1985	
		<i>Daphnia carinata mirabilis</i>	Stingelin, 1915	
		<i>Ilyocryptus spinifer</i>	Timms, 1985	
		<i>Macrothrix spinosa</i>	Timms, 1985	
		<i>Diaphanosoma sarsi</i>	Timms, 1985	
		<i>Latonopsis brehmi dedeckkeri</i>	Timms, 1985	
			Laevicaudata	<i>Lynceus insularis</i>
	Copepoda	Cyclopoida	<i>Eucyclops neocaledoniensis</i>	Alekseev, 2019
<i>Gonicyclops arenicola</i>			Holyńska & Ślugocki, 2023	
<i>Mesocyclops medialis</i>			Holyńska & Ślugocki, 2023	
<i>Tropocyclops prasinus</i>			Holyńska & Ślugocki, 2023	
<i>Microcyclops varicans</i>			Holyńska & Ślugocki, 2023	
<i>Thermocyclops decipiens</i>			Holyńska & Ślugocki, 2023	
<i>Oithona nana</i>			Dussart, 1984	
Harpacticoida		<i>Phyllognathopus sp.</i>	Dussart, 1984	
		<i>Parastenocaris digitophora</i>	Dussart, 1984	
		<i>Parastenocaris sp.</i>	Dussart, 1984	
		<i>Attheyella ablatifurcata</i>	Hamond, 1987	
Calanoida		<i>Boeckella spinogibba</i>	Defaye, 1998	
		<i>Dussartopages fagesi</i>	Dussart, 1986	
Ostracoda	Podocopida	<i>Penthesilenula brasiliensis</i>	Martens <i>et al.</i> , 2019	
		<i>Vestalenula marmonieri</i>	Martens <i>et al.</i> , 2019	
		<i>Cyprideis australiensis</i>	Martens <i>et al.</i> , 2019	
		<i>Cyprideis consobrina</i>	Martens <i>et al.</i> , 2019	
		<i>Kennethia major</i>	Martens <i>et al.</i> , 2019	
		<i>Cyprinotus cingalensis</i>	Martens <i>et al.</i> , 2019	
		<i>Cyprinotus drubea</i>	Martens <i>et al.</i> , 2019	
		<i>Cypris granulata</i>	Martens <i>et al.</i> , 2019	
		<i>Eucypris wolffhügeli</i>	Martens <i>et al.</i> , 2019	
		<i>Strandesia rouxi</i>	Martens <i>et al.</i> , 2019	
		<i>Strandesia sanoamuangae</i>	Kisseih <i>et al.</i> , 2020	
		<i>Strandesia mehesi</i>	Kisseih <i>et al.</i> , 2020	
		<i>Candonocypris novaezelandiae</i>	Martens <i>et al.</i> , 2019	
		<i>Stenocypris major</i>	Martens <i>et al.</i> , 2019	
		<i>Stenocypris marginata</i>	Martens <i>et al.</i> , 2019	
		<i>Cypridopsis sarasini</i>	Martens <i>et al.</i> , 2019	
		<i>Caledromus robinsmithi</i>	Martens <i>et al.</i> , 2019	
		<i>Lithocypris peyia</i>	Martens <i>et al.</i> , 2019	

Bien que des analyses comprenant un plus large spectre de gènes soient nécessaire pour conclure, les espèces de *Boeckella* retrouvées en NCK semblent se placer en groupe frère de tous les autres *Boeckella* du monde. Les populations néo-calédoniennes de *Boeckella* pourraient constituer un élément central pour expliquer la distribution naturelle de ce genre localisé aussi en Australie, Nouvelle-Zélande, Amérique du Sud, Antarctique et îles sub-Antarctique. La diversité génétique remarquable observée chez les *Boeckella* néo-calédoniens font du genre *Boeckella*, avec potentiellement 6 espèces, un cas particulièrement prononcé de micro-endémisme. Ce phénomène est presque commun sur les massifs ultramafiques de NCK mais reste globalement incompris et nécessite des plus amples investigations. Ce genre pourrait constituer un modèle intéressant pour l'étude de la diversification génétique et spécifique dans les systèmes insulaires imbriqués.

Enfin, des analyses similaires ont été réalisées sur les genres de Branchiopodes *Triops*, *Streptocephalus*, *Lynceus*, *Latonopsis* et *Eulimnadia*. Chez les espèces de ces genres retrouvés en NCK semblent se distinguer trois cas distincts. D'une part, les espèces retrouvées uniquement sur les massifs ultramafiques semblent avoir des distributions restreintes à l'échelle de l'île et présentent une nette structure géographique de leur diversité génétique. D'autre part, les espèces retrouvées uniquement hors des massifs ultramafiques sont majoritairement signalées également en Australie et ne semblent pas présenter de structuration géographique de leur diversité génétique. Enfin, une espèce non décrite, *Lynceus sp. A*, est retrouvée majoritairement en dehors des massifs ultramafiques mais présente une population sur les massifs ultramafiques. Cette espèce est également retrouvée en Australie et présente une structuration géographique de sa diversité génétique qui semble liée à la nature du sol. En effet, les haplotypes retrouvés chez la population présente sur les massifs ultramafiques ne sont pas retrouvés dans les autres localités de l'archipel ou en Australie.

II. Les espèces de crustacés d'eau douce

Comme évoqué en matériel et méthodes, 3 taxons parmi les crustacés connus en NCK n'ont pas été étudiés, la classe Ostracoda et les ordres Cyclopoida et Harpacticoida. Sans ces trois taxons, 21 espèces sont reportées en NCK chez les taxons étudiés (Branchiopoda et Calanoida). Les tris partiels et complets de 180 échantillons (114 prélevés au filet et 66 élevages) sur l'archipel ont permis d'identifier 23 nouvelles espèces (tab. 1.2). Parmi ces espèces, 15 étaient déjà décrites par ailleurs et 8 sont potentiellement nouvelles pour la science.

Tableau 1.2 - Espèces de crustacés détectées dans les échantillons triés partiellement ou entièrement (+ : espèce non reportée en NCK à l'origine ; **en gras** : espèces suspectées endémiques ; ? : Potentiel problème d'identification). Abréviations type de mare : HUM – Hors ultramafique strict ; UM – Ultramafique strict ; P – Partout.

Classe	Ordre	Famille	Espèce	Mare	
Branchiopoda	Anostraca	Streptocephalidae	+ <i>Streptocephalus archeri</i>	HUM	
		Thamnocephalidae	+ <i>Branchinella australiensis</i> (id. œufs)	HUM	
	Notostraca	Triopsidae	<i>Triops longicaudatus intermedius</i>	HUM	
	Anomopoda	Chydoridae	<i>Chydorus eurynotus</i>	P	
+ <i>Chydorus pubescens</i>			HUM		
+ <i>Chydorus herrmanni</i>			P		
+ <i>Rhynchochydorus australiensis</i>			UM		
<i>Ephemeroporus barroisi</i>			UM		
<i>Dunhevedia crassa</i>			HUM		
+ <i>Leberis davidi</i>			P		
+ <i>Alona rigidicaudis</i>			P		
<i>Coronatella rectangula</i>			UM		
<i>Alonella clathratula</i>			UM		
<i>Karualona karua</i>			HUM		
<i>Armatalona macrocopa</i>			UM		
+ <i>Dadaya macrops</i>			HUM		
			Daphniidae	<i>Simocephalus exspinosus australiensis</i>	P
				<i>Simocephalus acutirostratus</i>	HUM
				<i>Ceriodaphnia rigaudi</i>	P
				<i>Daphnia cephalata</i>	HUM
		<i>Daphnia longicephala</i>	HUM		
	Ilyocryptidae	? <i>Ilyocryptus spinifer - sordidus - timmsi</i>	P		
	Macrothricidae	<i>Macrothrix spinosa</i>	HUM		
		+ <i>Macrothrix sp. A</i>	UM		
		+ <i>Macrothrix triserialis</i>	HUM		
	Moinidae	+ <i>Moina micrura</i>	P		
	Sididae	+ <i>Diaphanosoma australiensis</i>	P		
		+ <i>Diaphanosoma unguiculatum</i>	HUM		
		<i>Latonopsis dedeckkeri</i>	UM		
		+ <i>Latonopsis australis</i>	HUM		
	Laevicaudata	Lynceidae	<i>Lynceus insularis</i>	UM	
			+ <i>Lynceus sp. A</i>	P	
			+ <i>Lynceus sp. B</i>	UM	
	Spinicaudata	Limnadiidae	+ <i>Eulimnadia sp. A</i>	HUM	
			+ <i>Australimnadia sp.</i> (id. œufs)	HUM	
Copepoda	Calanoida	Centropagidae	<i>Boeckella spinogibba</i>	UM	
			+ <i>Boeckella sibleti</i>	UM	
			+ <i>Boeckella sp. A</i>	UM	
			+ <i>Boeckella sp. B</i>	UM	
			+ <i>Boeckella sp. C</i>	UM	
			+ <i>Boeckella sp. D</i>	UM	

Les espèces *Branchinella australiensis* et *Australimnadia sp.* ont été identifiées uniquement par la détection de leurs œufs dans les sédiments. Deux espèces connues de l'archipel n'ont pas été ré-observées dans les échantillons triés, le Branchiopode *Daphnia carinata mirabilis* et le Calanoïde *Dussartopages fagesi*. Ces deux espèces n'ont pas été signalées depuis leur premier signalement sur l'archipel.

D'après les analyses effectuées par Angel Arnold dans son stage, la majorité des individus du genre *Diaphanosoma* trouvés dans les échantillons étudiés correspondent partiellement à la description de *D. sarsi*. De plus, ils semblent morphologiquement proches de *D. australiensis* qui a été décrite plus tard (Arnold *et al.*, *in prep*). Cette espèce semble avoir une distribution similaire à *Lynceus sp. A* qui sont toutes deux retrouvées sur tous les types de massifs en NCK et en Australie. D'autre part, l'espèce *D. unguiculatum* a été identifiée dans une seule localité.

La taxonomie de la famille des Chydoridae connaît de nombreux réarrangements depuis le début des années 2010 (Van Damme & Dumont, 2008; Sinev & Kobayashi, 2012; Gu *et al.*, 2022). Bien que de légères différences morphologiques aient pu être constatées ponctuellement, les identifications se sont cantonnées aux espèces connues de ce taxon en attendant une révision taxonomique.

L'espèce *Ilyocryptus spinifer* a été originellement reportée en NCK par Stingelin (1915). Elle n'y a pas été ré-indiquée depuis et, malgré une révision du genre par Kotov & Elias-Gutierrez (2009), les morphologies affiliées à *I. sordidus*, *I. spinifer* et *I. timmsi* sont très proches. Des morphologies légèrement différentes ont été identifiées dans les échantillons étudiés mais il n'a pas été possible d'identifier formellement les individus prélevés.

III. Le cas du genre *Boeckella*

Genetic variability of New Caledonian *Boeckella* De Guerne & Richard, 1889 (Copepoda: Calanoida), with description of a new species

Coline Royaux | Nicolas Charpin | Nicolas Rabet

Publié, *Journal of Crustacean Biology*
<https://doi.org/10.1093/jcbiol/ruae001>

Abstract

Thirteen new freshwater populations of the copepod genus *Boeckella* De Guerne & Richard, 1889 were found during three expeditions to New Caledonia ('La Planète Revisitée,' 2016-2018). The 12 populations from the Plaine des Lacs, which show remarkable genetic diversity among themselves, were identified as *B. spinogibba* Defaye, 1998, the only species of its genus known from New Caledonia until now. The sole exception, the population from Mont-Dore 22 km further east, appeared genetically and morphologically distinct from the others and is described herein as a new species. The two species are distinguished from each other by the shapes of the male and female P5, female Th5, and body colour. A previously published key is amended to separate the species. Our concatenated COI+28S phylogeny places the two New Caledonia species as a new branch within *Boeckella*, distinct from the branches consisting of South American, Antarctic and Australasian species. All 13 *Boeckella* populations inhabit the extreme south of New Caledonia, an area known for endemism and high heavy metal concentrations in the soil. Extensive mining activity in this metal-rich area, begun in 1873, is now leading to conflict with conservation goals. By using aerial photographs to trace the fate of the pond that is the type locality of *B. spinogibba*, we confirmed that it disappeared between 2007 and 2014 as a result of the expansion of the Goro nickel mine.

Keywords

Boeckella spinogibba, conservation, Crustacea, micro-endemism, freshwater zooplankton, mining perturbation, morphology, phylogeny, population genetics

Introduction

The main island of New Caledonia (Grande Terre) separated from Australia around 80 mya. It later became submerged until approximately 37 mya, when uplift caused the re-emersion of Grande Terre along with rocks from the upper mantle. Laterization of these mantellic rocks formed ultramafic massifs abundant in heavy metals across the island (Dubois *et al.*, 1974; Pelletier, 2007; Grandcolas *et al.*, 2008). Ultramafic massifs are mainly located in the southeast of New Caledonia (fig. 1.1.1), a site of high endemism (Nattier *et al.*, 2013; Isnard *et al.*, 2016). Intensive mining of, e.g., nickel and cobalt, constitutes a major threat to the area's biodiversity (L'Huillier & Jaffré, 2010; Pitoiset, 2016).

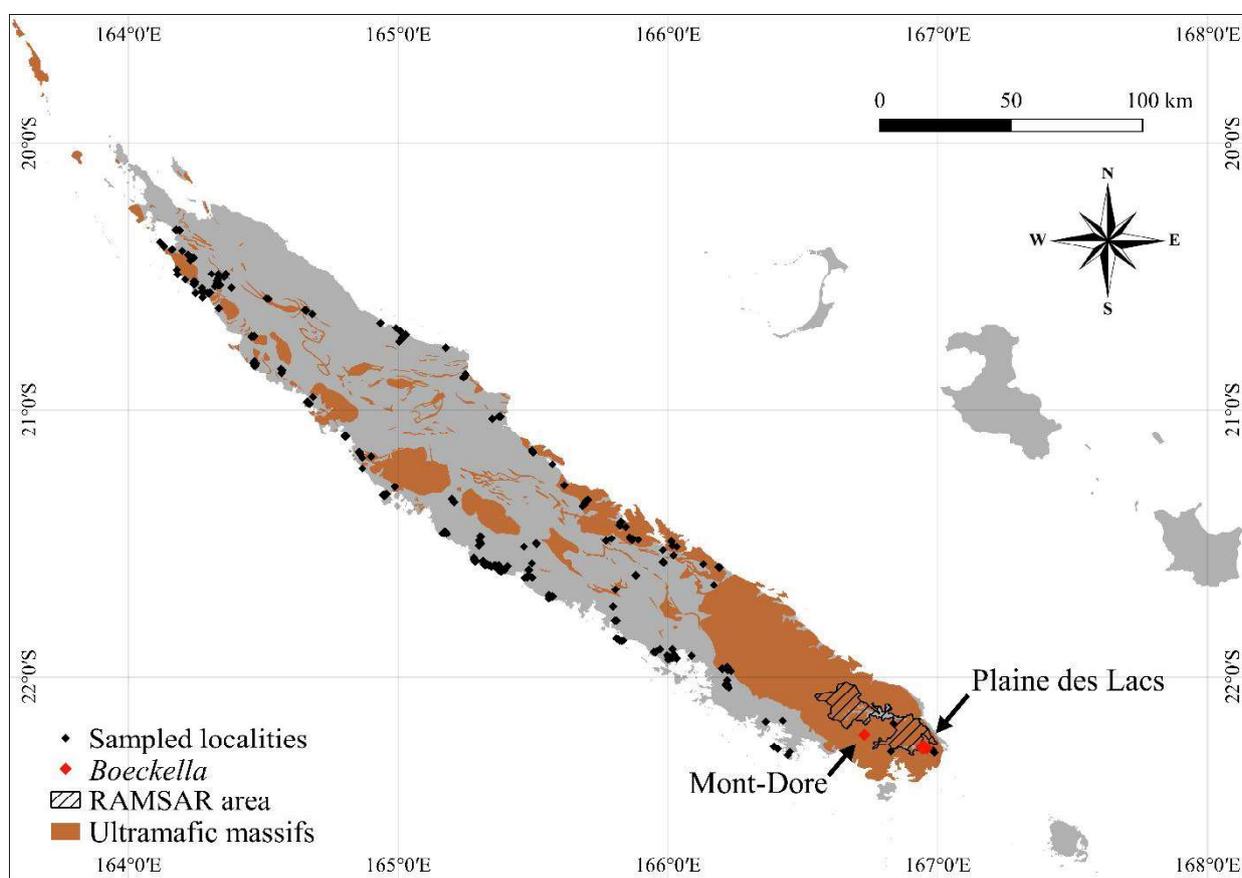


Figure 1.1.1. Localities sampled in New Caledonia during the three 'La Planète Revisitée' expeditions (2016–2018). Red represents localities where *Boeckella* was found.

New Caledonia is globally acknowledged to be a major biodiversity hotspot despite its relatively small surface area (18,576 km²). The fauna of its lentic freshwater bodies is, however, relatively little-known. Freshwater ponds are diverse environments; the turbidity of the water, size of the pool, flora outside and inside, and average water temperature are some of the characteristics that vary over time and space and condition the fauna inhabiting them. The fauna of New Caledonia

freshwater ponds includes such invertebrates as the coleopteran *Phalilus oberthuri* Guignot, 1935, the ostracod *Cyprinotus cingalensis* Brady, 1886, and the branchiopod *Lynceus insularis* Olesen, Pöllabauer, Sigvardt & Rogers, 2016 [Olesen *et al.*, 2016] (Inventaire National du Patrimoine Naturel, 2023). When such bodies are temporary, their cyclic drying-out prevents most members of the aquatic fauna from colonizing them permanently so that most fishes and some macrocrustaceans are excluded. As planktonic crustaceans in temporary freshwater can produce diapausing eggs, they often represent a prominent fraction of biodiversity in these environments (Brock *et al.*, 2003).

The diversity of freshwater zooplankton remains inadequately studied globally, even more so in New Caledonia. To fill in knowledge gaps there, many temporary ponds and sinkholes all across Grande Terre were sampled during three ‘La Planète Revisitée’ expeditions between 2016 and 2018 (fig. 1.1.1). Preliminary examinations of these samples identified many previously unknown populations of freshwater crustaceans. More specifically, the copepod genus *Boeckella* De Guerne & Richard, 1889 was exclusively found on ultramafic massifs known for hosting high endemism. Only one species of *Boeckella* had previously been reported in New Caledonian temporary freshwater ponds: *B. spinogibba* Defaye, 1998. It was described and reported from a single locality on the ultramafic massifs of the south of Grande Terre (Defaye, 1998) and has not been further studied or sought for since.

We aimed to describe part of the taxonomic and genetic diversity of *Boeckella* in New Caledonia. This genus is currently composed of 42 valid species, including 23 from Australasia and 19 from South America, various sub-Antarctic islands, and Antarctica (Supplementary material fig. AS1; Maturana *et al.*, 2021; Walter & Boxshall, 2021). Extensive sampling conducted between 2016 and 2018 resulted in the identification of new populations of *Boeckella* in temporary ponds, one of which is geographically isolated in Mont-Dore, approximately 22 km away from the others, which inhabit ponds in the Plaine des Lacs. The morphological and genetic features of the *Boeckella* populations from both locations were investigated, including a population genetics analysis based on the cytochrome C oxidase I (COI) gene in these populations. The left antennule (A1) of *B. spinogibba* is redescribed. Some of these populations are in ponds located in the RAMSAR area defined by the Convention on Wetlands (fig. 1.1.1) and very close to the Goro mine, the largest nickel mine in New Caledonia. This facility is an open-pit complex where nickel and cobalt minerals are extracted. It has notably expanded since the description of *B.*

spinogibba in 1998, and the effect of such expansion on the fate of the type locality of *B. spinogibba* is here tracked through time.

Materials and methods

Study sites and sampling

The survey of 469 ponds on Grande Terre was conducted between 2016 and 2018. Ponds filled with water were sampled using a plankton net (NHBS, Bonn, Germany) with a frame diameter of 25 cm and a 55 cm-long bag of 200 µm nylon mesh with a filter at the tip. The net was either tossed and pulled from the edge of the pond or towed while swimming in the pond. Samples were preserved in 80–90% ethanol at 4 °C as soon as possible to avoid alteration of genetic material. Individuals of *Boeckella* were found in 13 temporary ponds. Twelve of these were in the Plaine des Lacs (figs. 1.1.1, 1.1.2A–E) in a small area of 2.5 km² at around 290 m elevation situated between 22°16.2'S and 22°15'S latitude and between 166°56.4'E and 166°57.6E longitude (sites HYNC918, HYNC920, HYNC925, HYNC927, HYNC931, HYNC4102, HYNC4103, HYNC4104, HYNC4105, HYNC4109, HYNC4110 and HYNC4111) (Table 1.1.1). The remaining pond was in Mont-Dore (site HYNC928, 22°12.9'S, 166°43.6'E, 117 m elevation; figs. 1.1.1, 1.1.2F–G).

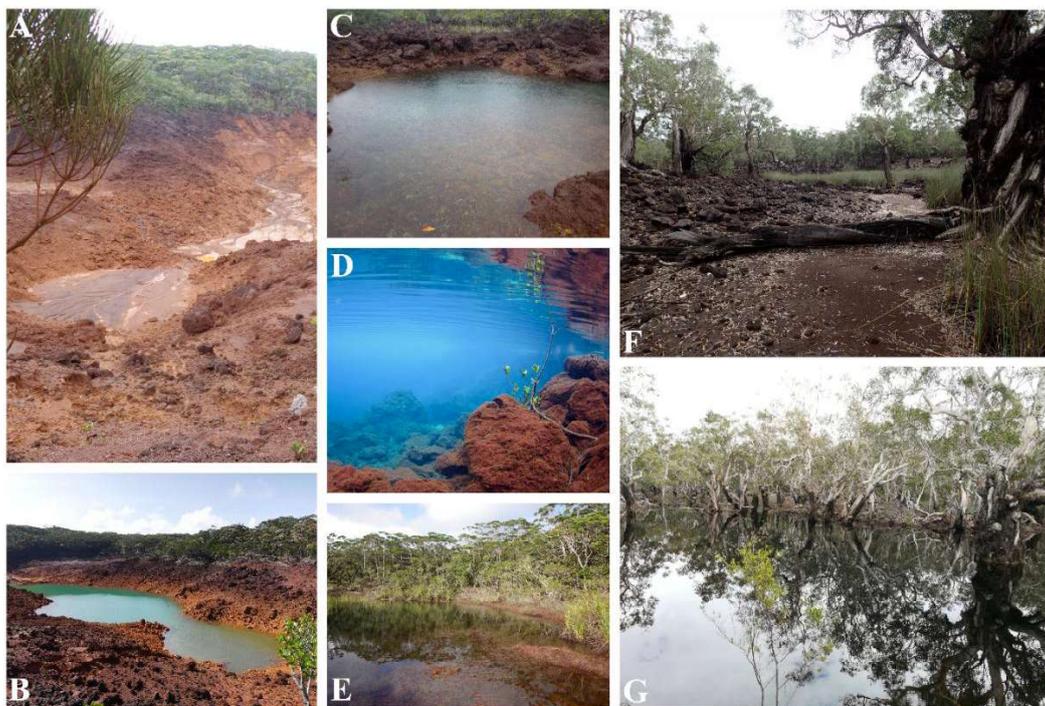


Figure 1.1.2. Sampled ponds in the Plaine des Lacs where species of *Boeckella* were found. Pond HYNC918 dry in November 2016 (A); filled in June 2018 (B). Pond HYNC920 in November 2016 (C); and in June 2018 (D). Pond HYNC931 in June 2018 (E). Sampled pond (HYNC928) in Mont-Dore where *Boeckella sibleti* sp. nov. was found (type locality). Dry in November 2016 (F); filled in June 2018 (G).

Table 1.1.1. *Boeckella* sampling sites, coordinates and aliases.

Locality (pond) single identifier	Area in New Caledonia	Latitude	Longitude	Other names given to the locality
HYNC918	Plaine des Lacs	22°16.1982S	166°57.5412E	HYNC4101
HYNC920	Plaine des Lacs	22°15.7692S	166°57.126E	HYNC4005
HYNC925	Plaine des Lacs	22°15.6642S	166°57.1554E	HYNC4107 ; Lac en long
HYNC927	Plaine des Lacs	22°15.726S	166°57.1938E	HYNC4108
HYNC931	Plaine des Lacs	22°15.5502S	166°56.901E	HYNC4106 ; DT53
HYNC4102	Plaine des Lacs	22°15.5556S	166°56.4846E	
HYNC4103	Plaine des Lacs	22°16.125S	166°56.4996E	
HYNC4104	Plaine des Lacs	22°16.101S	166°56.8638E	
HYNC4105	Plaine des Lacs	22°15.7542S	166°56.865E	
HYNC4109	Plaine des Lacs	22°15.7182S	166°57.201E	
HYNC4110	Plaine des Lacs	22°15.7686S	166°57.207E	
HYNC4111	Plaine des Lacs	22°15.7176S	166°57.2382E	Doline croissante
HYNC928	Mont-Dore	22°12.93S	166°43.662E	HYNC4004

Morphological analysis

Specimens were dissected under a binocular dissecting microscope (SZX7, Olympus, Tokyo, Japan) and microscopically observed in glycerol or 96% ethanol. Photographs serving as the basis for drawings and measurements were taken with a digital camera (MC170 HD, Leica, Wetzlar, Germany) attached to an optical compound microscope (BX51, Olympus); the drawings and measurements were made using Inkscape 1.2.2 (The Inkscape Project; <https://inkscape.org>) and Gimp 2.10 (GNOME Project; <https://www.gimp.org/>).

Some specimens from the Plaine des Lacs and Mont-Dore were prepared for SEM examination and underwent sequential dehydration in 96% and 100% ethanol followed by critical point-drying in liquid CO₂ (Emitech K850, Quorum Technologies, Lewes, UK). After mounting on flat aluminium stubs (12.7 mm diameter) with double-sided conductive carbon adhesive tape, specimens were sputter-coated with platinum (EM ACE600, Leica) and examined at 15 kV accelerating voltage and at 90–5500× magnification in a scanning electron microscope (SU3500, Hitachi, Tokyo, Japan).

Abbreviations used in the text: Th (1–5), thoracic segment; A1, antennules; Ur (1–5), urosomites; P (1–5), legs; B1, coxa; B2, basis; Re (1–3), exopod; Ri (1–3), endopod.

All the sampled material except for specimens used destructively for SEM or sequencing is preserved in ethanol and deposited in the Muséum national d’Histoire naturelle, Paris (see Supplementary material fig. AS2 for registration numbers of samples).

DNA extraction and PCR

Apart from site HYNC4105, where only a few individuals of *Boeckella* were found, DNA was extracted from 5–10 individuals from each pond where this genus was detected (90 individuals in total). Lysis was performed on whole bodies, and DNA was extracted from tissues by the NucleoSpin® 96 robot using the NucleoSpin® 96 Tissue Core Kit (Macherey-Nagel, Düren, Germany).

In order to amplify the mitochondrial gene COI (cytochrome C oxidase I), Folmer primers LCO1490 and HCO2198 (Folmer *et al.*, 1994) were used with the following PCR (polymerase chain reaction) cycle: 5 min at 94 °C; 40 cycles of 30 sec at 94 °C, 30 sec at 50 °C, 1 min at 72 °C, and 8 min at 72 °C. To ensure proper amplification of the sequences, PCR products were subjected to gel electrophoresis. The expected size of amplified DNA strands was 658 bp without primers. Detailed protocols are available on protocols.io (DNA extraction: <https://dx.doi.org/10.17504/protocols.io.j8nlk4x7wg5r/v2>; PCR: <https://dx.doi.org/10.17504/protocols.io.261ge4y4jv47/v3>).

Sequencing and phylogenetic analysis

Properly amplified DNA sequences were sent to Eurofins Scientific (Luxembourg City, Luxembourg) for Sanger sequencing (Sanger *et al.*, 1977). The obtained sequences were cleaned, complementary strands were aligned, and consensus sequences were determined using Galaxy (<https://ecology.usegalaxy.eu>; The Galaxy Community, 2022).

To obtain more genomic information on individuals from New Caledonia, library preps from DNA amplified from one individual from the Plaine des Lacs (Site HYNC931; DNA sample identifier M156) and 50 individuals from Mont-Dore (DNA sample identifier M223) were performed with a Nextera XT kit (Illumina;

<https://www.illumina.com/>) using fragmentation and Illumina index adapter ligation. Equimolar pools of each library were established. Qualification and quantification of the final library was established before sequencing on Illumina Miseq at Plateforme iGenSeq, Institut du Cerveau (ICM) (Hôpital Pitié Salpêtrière, Paris) with 2×25 million reads cartridge of 300 bases each (30–45 libraries per run).

All raw sequences were assembled and sequencing errors corrected using Geneious (Geneious® 11.1.4, Dotmatics; <https://www.geneious.com/>). The consensus sequences are available in GenBank (see Supplementary material fig. AS3 for NCBI accession numbers). In addition, gene concatenations were performed by adding the assembled genes from the present study to the concatenation (ribosomal RNA large subunit 28S+COI) published by Maturana *et al.* (2021) and retrieved in Genbank. In total, the phylogeny was computed on 40 individual sequence concatenations (Supplementary material fig. AS4). Apart from the two species from New Caledonia, 16 other species of *Boeckella* were included in the phylogeny. *Calamoecia ampulla* (Searle, 1911) and *Parabroteas sarsi* (Daday, 1901) were used as an external group, as Adamowicz *et al.* (2010) demonstrated *Calamoecia* Brady, 1906 and *Parabroteas* Mrázek, 1901 formed a monophyletic group closest to *Boeckella* in the family Centropagidae. Genes were independently aligned with Geneious alignment in Geneious 11.1.4 (Dotmatics). The 28S was delimited by hand. The COI sequences were aligned and verified for the absence of a stop codon in the reading phase. The COI alignment was then subjected to a Gblock performed with the default options on the LIRMM web platform (Dereeper *et al.*, 2008). The concatenation was realized on PhyloSuite v1.2.3 (Zhang *et al.*, 2020; Xiang *et al.*, 2023). Evolution models were estimated with distinct codon positions for COI using ‘model finder’ in PhyloSuite 1.2.3 (Kalyaanamoorthy *et al.*, 2017; Zhang *et al.*, 2020; Xiang *et al.*, 2023).

The concatenated 28S+COI phylogeny was constructed from a Bayesian analysis performed with MrBayes 3.2.7 (Ronquist *et al.*, 2012) using the models TIM+I+G (28S), TIM+G, TRN+I, and F81+I for each codon position (COI). Maximum likelihood analysis was performed with IQTree 2.2.0 (Minh *et al.*, 2020) using the models TIM+I+G (28S), TIM+G, TRN+I, and F81+I for each codon positions (COI), always with 1,000 bootstrap resamplings. The multilocus phylogenetic tree was formatted using FigTree 1.4.4 (<https://tree.bio.ed.ac.uk/software/figtree>) and Inkscape 1.2.2 (The Inkscape Project).

Two species-delimitation models were performed on the bPTP server (Bayesian Poisson Tree Processes; <https://species.h-its.org/>), one on the concatenated 28S+COI phylogenetic tree and one on a COI phylogenetic tree (Supplementary material fig. AS5). Both analyses were calibrated with 100,000 MCMC generations and exclusion of the outgroup.

Population structure of Boeckella spinogibba

To assess the genetic structure of the populations of *Boeckella spinogibba* found in the Plaine des Lacs, an exact test of population differentiation and an analysis of molecular variance (AMOVA) were performed using R 4.1.2 (CRAN) packages mmod 1.3.3 (Winter, 2012) and ade4 1.7-18 (Dray & Dufour, 2007), respectively. Significance of the AMOVA was computed through randomised matrix permutations. An alignment was also produced using Galaxy (mafft method, tool align seqs 1.9.1.0; Katoh et al., 2002) in order to compute a median-joining haplotype network (Bandelt *et al.*, 1999) using R package pegas 1.2 (Paradis, 2010). Median-joining haplotype networks represent genetic links between haplotypes. The median-joining algorithm occasionally computes hypothetical haplotypes (referred to as ‘median vectors’), that could potentially be unsampled sequences, to resolve the network. The network was computed using several tolerance (epsilon) parametrisations in order to obtain several possible representations of the links between haplotypes. The tolerance parameter set in a median-joining network affects the search for median vectors. The augmentation of the tolerance parameter beyond 2 had little impact on the organisation of the network.

All sequence cleaning and analyses conducted in Galaxy have been implemented as reusable workflows at: https://usegalaxy.fr/u/coline_royaux/w/sanger1--from-ab1-to-aligned-fasta-imported-from-uploaded-file; https://usegalaxy.eu/u/coline_royaux/w/sanger-5--from-hand-corrected-haplotypes-table-to-population-haplotypes-map-and-euclidean-distance-matrix.

History of type locality of Boeckella spinogibba

The coordinates of the type locality of *B. spinogibba* provided by Defaye (1998) are precise within about 2 km². Aerial photographs of the region in question were downloaded from the Observatoire de l’environnement Nouvelle-Calédonie (www.oeil.nc), and satellite images were obtained from Google Earth Pro 7.3.2. The

description of the surrounding environment by Defaye (1998) and these aerial and satellite images allowed us to find the pond and monitor its condition from 1976 to 2014.

Results

Taxonomy

Order Calanoida G.O. Sars, 1903

Family Centropagidae Giesbrecht, 1892

Boeckella De Guerne & Richard, 1889

Boeckella spinogibba Defaye, 1998

(fig. 1.1.3)

Material examined: 2 males examined in SEM from *B. spinogibba* in pond HYNC931 (MNHN-UI-2017-2310, MNHN-UI-2017-2311); 5 males examined in photonic microscope from *B. spinogibba* in pond HYNC931 (4 individuals; MNHN-UI-2017-2310 and MNHN-UI-2017-2311) and in pond HYNC920 (1 individual; MNHN-UI-2017-2303).

Morphological remarks: On the right A1 of males, Defaye (1998) described five types of sensory structure: aesthetascs (fig. 1.1.3A–C, G); thick setae (fig. 1.1.3B, C); thin setae (fig. 1.1.3C, G); setae with hairs (fig. 1.1.3F); and pectinate setae (fig. 1.1.3D, E). Detailed examination with SEM of specimens of *B. spinogibba* in pond HYNC931 of Plaine des Lacs (Table 1.1.2) showed that the distribution of these sensory structures on the right antennule of males is different from the original description by Defaye (1998). Namely, the original description made no mention of the thin seta on segments 3, 12, 17, and 19; the supplementary pectinate seta and the seta with hairs on segment 19; and the two thin setae on segment 20. No hairs were observed on the thin setae of segments 3, 5, 9, and 14, contrary to the original description, which mentioned such structures (Defaye, 1998).

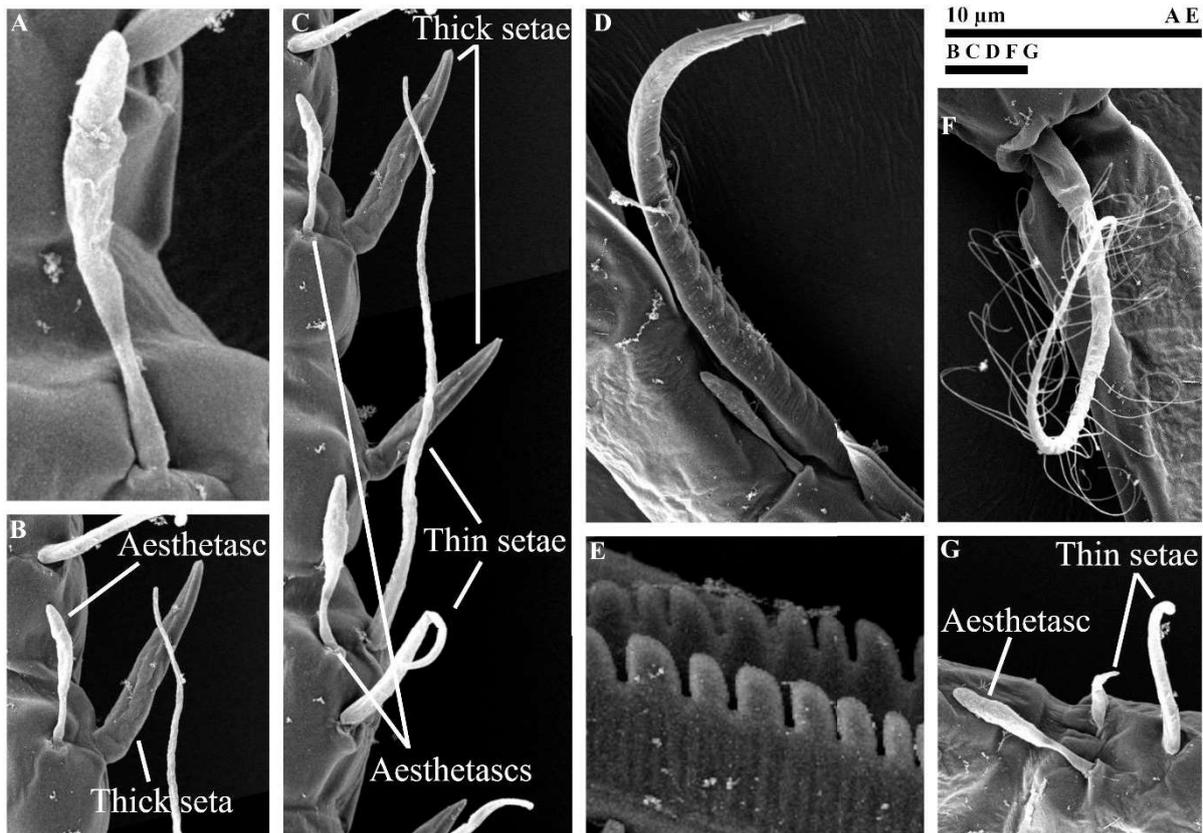


Figure 1.1.3. Types of sensory structures on male right antennule of *Boeckella spibogibba*. Lanceolate aesthetasc on segment 7 (A). Thick seta next to aesthetasc on segment 11 (B); long and short thin setae with two aesthetascs and two thick seta on segments 9–11 (C); pectinate seta on segment 19 (D); pectinate edge of pectinate seta on segment 18 (E); long seta on segment 19 with thin setules on distal third (F); short seta and tiny thin seta next to aesthetasc on segment 12 (G).

Table 1.1.2. Sensory armature of each segment (seg.) of the right antennules of male *Boeckella spinogibba* and *Boeckella sibleti* sp. nov.

Seg.	Structures	Seg.	Structures	Seg.	Structures	Seg.	Structures
1	1 aesthetasc 1 thin seta	7	2 thin setae 1 aesthetasc	13	1 thin seta 1 aesthetasc	19	2 pectinate setae 1 aesthetasc 1 thin seta 1 long, setulose seta
2	3 thin setae 1 aesthetasc	8	1 thin seta 1 thick seta	14	2 thin setae 1 aesthetasc	20	2 thin setae 2 long, setulose setae
3	2 thin setae 1 aesthetasc	9	1 aesthetasc 2 thin setae	15	2 thin setae 1 aesthetasc	21	2 long, setulose setae
4	2 thin setae	10	1 thick seta	16	2 thin setae 1 aesthetasc	22	3 long, setulose setae 2 thin setae 1 aesthetasc
5	2 thin setae 1 aesthetasc	11	1 thick seta 1 aesthetasc	17	1 pectinate seta 1 thin seta		
6	1 thin seta	12	2 thin setae 1 aesthetasc	18	1 pectinate seta 1 thin seta		

Boeckella sibleti sp. nov.

(figs. 1.1.4–1.1.6)

Material examined: Holotype, MNHN-UI-2017-2299, male. Allotype, MNHN-UI-2017-2300, female. Paratypes, MNHN-UI-2017-2298, 25 males; MNHN-UI-2017-2297, 20 females. Topotypes, MNHN-UI-2017-2301, MNHN-UI-2017-2302, >1,000 males (length, without furcal setae, 1.00–1.14 mm) and females (length, without furcal setae, 1.07–1.28 mm).

Type locality: HYNC928, a pond near the Pirogues River in Mont-Dore, Southern Province, New Caledonia, 22°12.9'S, 166°43.68'E, elevation 117 m, collected 5 June 2018. The pond occupies a pseudo-karstic formation called a 'doline' or 'sinkhole' (fig. 1.1.2F-G). Its water is rather heavily tannic and the surrounding extensive area of shrubby vegetation includes a large number of niaouli trees (bark tea tree, *Melaleuca quinquenervia*, Blake).

Etymology: Named in honour of Jean-Philippe Siblet, a fervent defender of nature conservation.

Colouration: Body red in both sexes.

Male: Holotype, MNHN-UI-2017-2299, length without furcal setae 1.13 mm. Body elongated as a tapered droplet. Cephalothorax 6-segmented with cephalosome longest, 5 free thoracic segments (Th1, closest to cephalosome, to Th5, closest to urosome) successively shorter towards urosome (fig. 1.1.4A).

Antennules (A1) asymmetrical as in *B. spinogibba* (Defaye, 1998: figs. 1F, 4C); left one 25-segmented, reaching tips of caudal rami; right one 22-segmented, slightly longer than left, with geniculations between 18, 19th and 19, 20th segments (fig. 1.1.4B). Mouthparts as in *B. spinogibba* (Defaye, 1998: figs. 2, 3A, 5A).

Th5 (fig. 1.1.5A) symmetrical, with each 'wing' ending in pointed lobe nearly reaching posterior end of first urosomite. Wing ending with small terminal spine and bearing second small spine towards posterior margin.

Natatory legs P1 to P4 identical to those of *B. spinogibba* (Defaye, 1998: fig. 5B–E).

Urosome with 5 urosomites successively decreasing in length towards furca from Ur1 to Ur4 but with Ur5 longer than Ur4 (fig. 1.1.4A, 1.1.5A).

P5 (fig. 1.1.5B–D) asymmetrical, with left limb's coxa (B1) twice as long as that of right limb. Left P5: Basis (B2) with medially swollen, finely denticular protrusion on inner margin, overlapping proximal margin of exopod (Re1), latter with distally denticulate terminal lateral spine. Endopod (Ri) atrophied but bearing fine hairs, inserted beneath B2's protrusion before Re1. Re2 with proximolateral distally denticulate spine. Re3 elongated, spine-like with slight medial curve, denticulate, pointed at apex. Right P5: B2 without medial protrusion, Ri elongated, digitiform, swollen at midlength, medially curved, rounded terminally. Re1 approximately half as long as Re2, bearing distal spine armed with approximately uniformly-sized denticles. Re2 with small medial bulge on inner margin and large, distally denticulate distolateral spine of almost equal length to segment. Re3 elongated, medially curved at midlength with slight medial bulge, ending in pointed apex.

Female: Allotype, MNHN-UI-2017-2300. Larger than male, length without furcal setae 1.23 mm.

Body shape slender, elongated. Cephalothoracic tagmosis as in male (fig. 1.1.6A). Mouthparts as in *B. spinogibba* (Defaye, 1998: figs. 2, 3A).

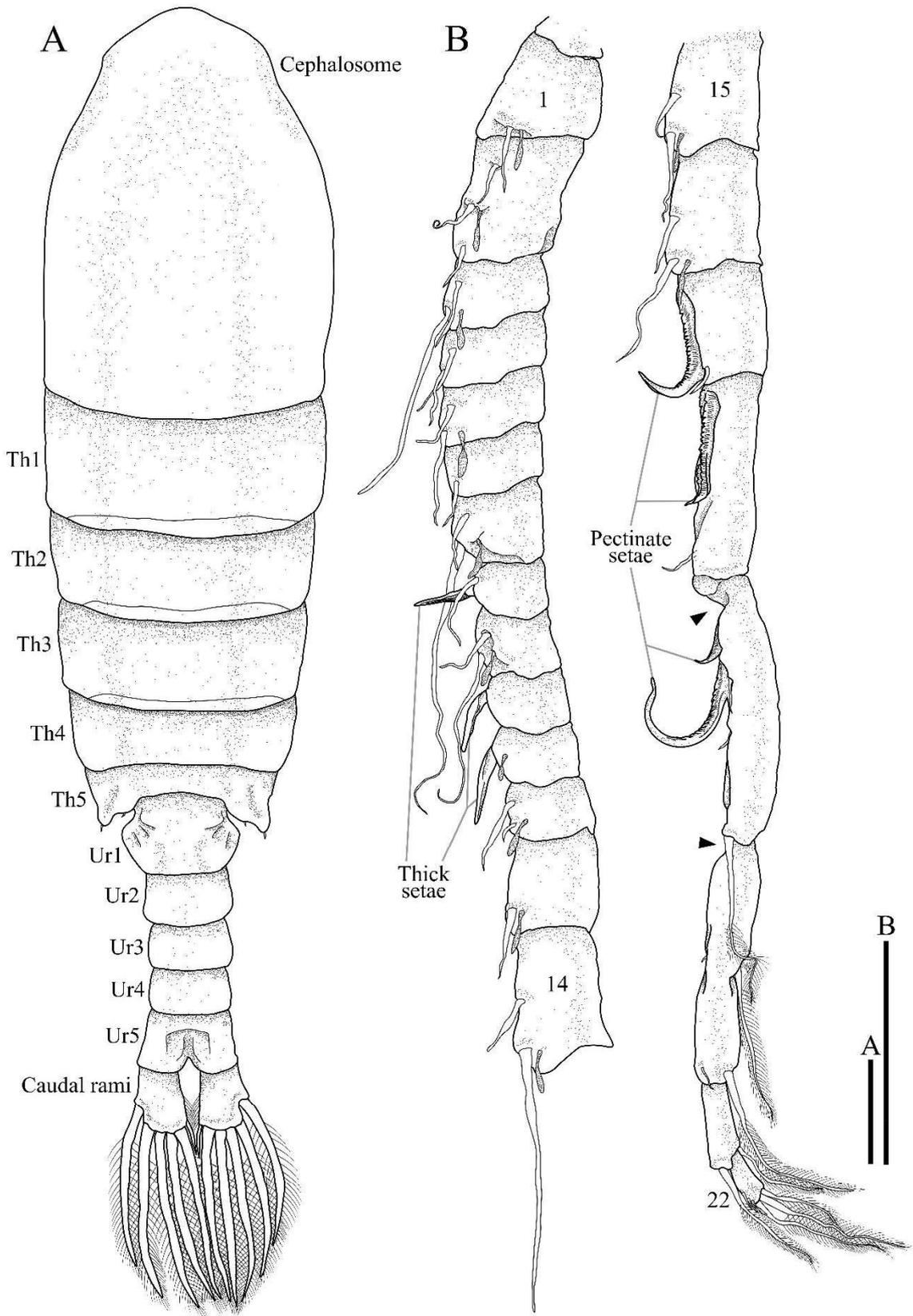


Figure 1.1.4. *Boeckella sibleti* sp. nov., male. Whole body without antennules, posterior view (A); right antennule, with thick setae and pectinate setae labelled, aesthetascs stippled, and long, setulose setae drawn with hairs, all others are thin setae; geniculation points indicated by black arrows (B). Scale bars = 100 μ m.

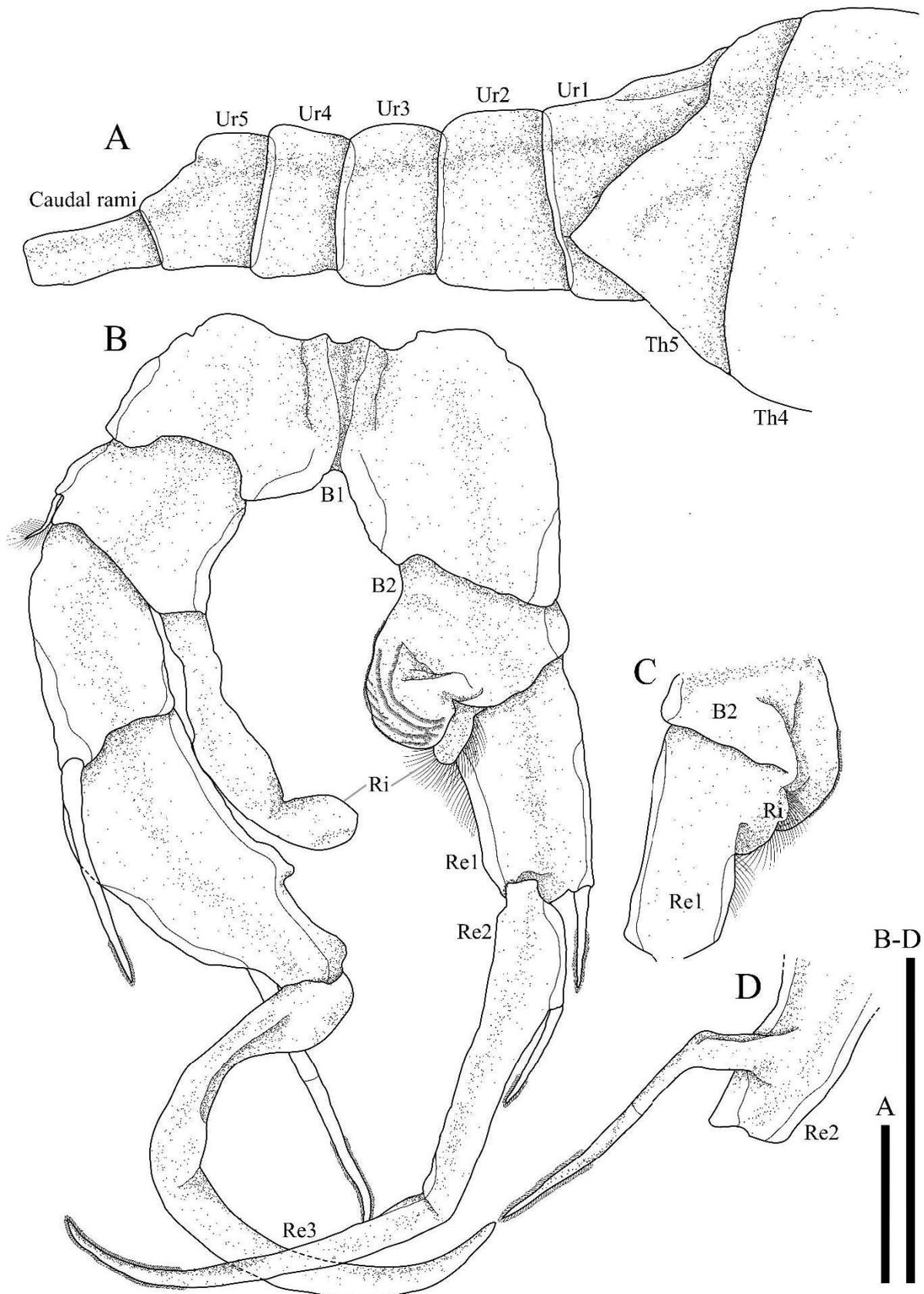


Figure 1.1.5. *Boeckella sibleti* sp. nov., male. Th4, Th5, and urosome, viewed from right side (A); P5, anterior view (B); left B2, Ri, and Re1, posterior view (C); distal part of left Re2, posterior view (D).

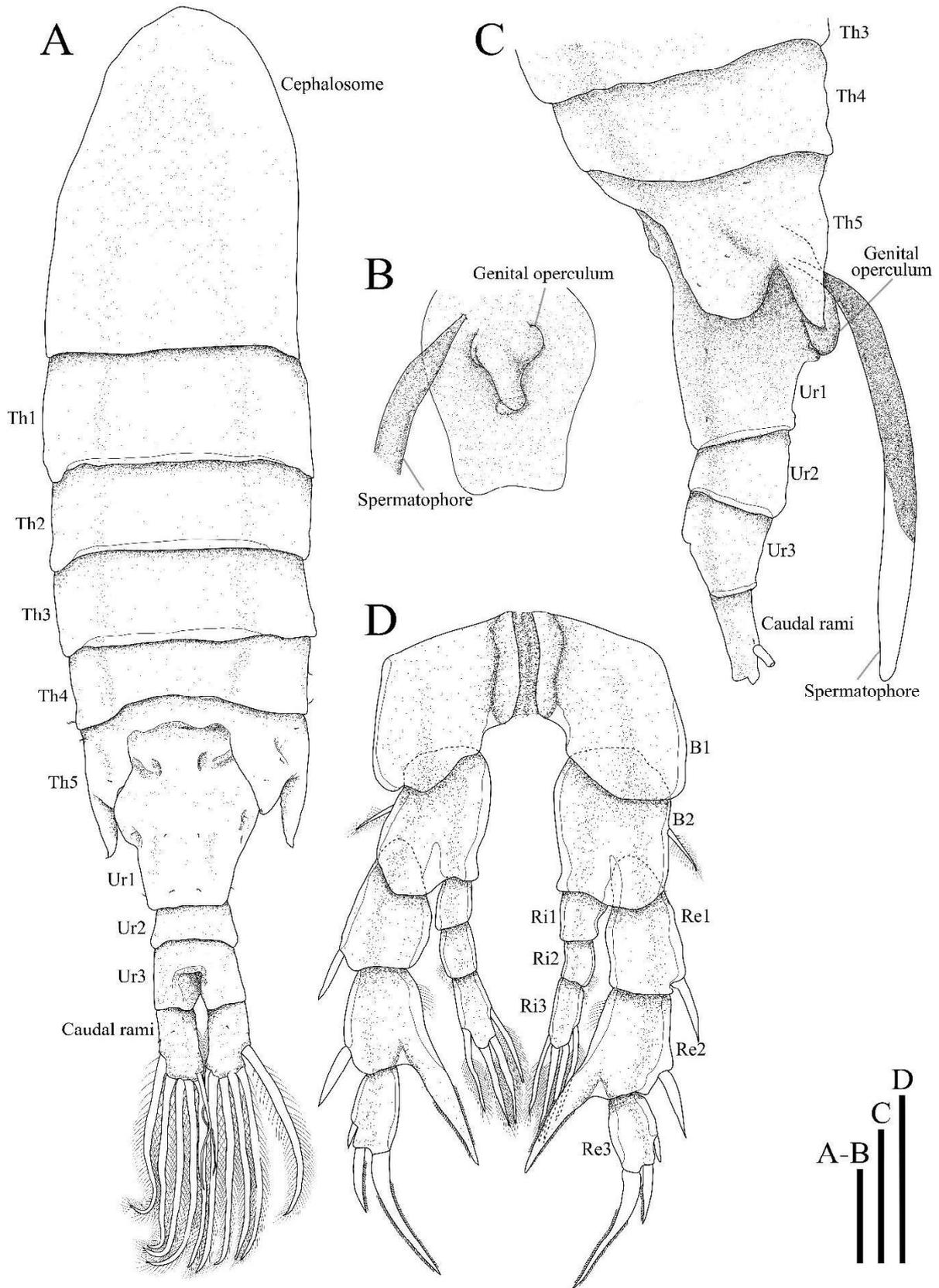


Figure 1.1.6. *Boeckella sibleti* sp. nov., female. Whole body without antennules, posterior view (A); genital operculum with part of spermatophore, anterior view (B); Th3, Th4, Th5, and urosome with spermatophore and genital operculum, viewed from right side (C); P5, anterior view (D). Scale bars = 100 μ m.

Antennules 25-segmented, extending beyond caudal rami, identical to those of *B. spinogibba* (Defaye, 1998: fig. 1F).

Left, right lateral ‘wings’ of Th5 ending in 2 lobes (fig. 1.1.6A, C), and each lobe bearing small spine on outer margin.

Natatory legs P1–P4 identical to those of *B. spinogibba* (Defaye, 1998: fig. 5B–E).

Urosome 3-segmented. Urosomite 1 (Ur1) longest, widest on medial transversal axis, narrowing towards anterior and posterior margins from posterior view (fig. 1.1.6A), asymmetrical anterior genital operculum, latter with wide anterior part, thinner posterior part in anterior view (fig. 1.1.6B). Ur1 also bearing spermatophore (fig. 1.1.6C).

P5 (fig. 1.1.6D) symmetrical, distolateral spines of Re1-2 equal in length. Distomedial spiniform process of Re2 extending beyond Re3, Re3 armed with small distolateral spine and 2 longer terminal spines, medial one of which longest and medially curved. Ri 3-segmented; Ri3 bearing 4 spines with fine hairs.

Remarks: The body of *B. sibleti* **sp. nov.** is red, whereas that of *B. spinogibba* is blue. The body shape of males is similar in both species, and the mouthparts and swimming legs appear identical between the species. The main differences are found in P5 of both sexes (fig. 1.1.7), and, for the female, Th5 (see Table 1.1.3).



Figure 1.1.7. SEM images of male copulatory organ (P5) of *Boeckella sibleti* sp. nov. (A) and of *B. spinogibba* (B). Major differences listed in Table 1.1.3 are indicated with a white arrow and a numeric identifier.

Table 1.1.3. Morphology-based differential diagnosis of *Boeckella spinogibba* and *B. sibleti* sp. nov. See Scheihing *et al.* (2010: 47, fig. 2) for abbreviations of parts of legs. Numbers (no.) given to differences shown in Figure 1.1.8 of text are provided for male P5.

		<i>Boeckella spinogibba</i>		<i>Boeckella sibleti</i>		Numbers in Figure 1.1.8		
Female	Th5	Right	Ending in two lobes of same length	Ending in two lobes with ventral lobe slightly longer				
		Left	Ending in one lobe with a slight swelling bearing a small spine on the dorsal side	Ending in two lobes				
	P5	Endopodite 3	3 spines	4 spines				
Male	P5	Right	Short spines on Re1 and Re2	Longer spines on Re1 and Re2		no. 1, 2		
			Ri thin and long	Ri thick and long		no. 3		
			Re3 slightly curved	Re3 highly curved like a hook		no. 4		
			No slit	Inner margin of Re2 with a small slit		no. 5		
		Left		Round process on B2 rather flat and bonded to B2	Round process on B2 rather plump and detached from B2		no. 6	

Both *B. spinogibba* and *B. sibleti* **sp. nov.** bear three spines on the Re3 of the female P5 (according to Defaye, 1998: 697), this is ‘the usual setation pattern of [this exopodite in] *Boeckella*.’ By contrast, most of the Australian species of *Boeckella* (*B. pilkililli* Quinlan & Bayly, 2017, *B. montana* Bayly, 1964, *B. opaqua* Fairbridge, 1945, and *B. triarticulata* Thomson, 1883) bear five to seven spines on this segment (Bayly, 1964; Quinlan & Bayly, 2017).

The female P5 in *B. spinogibba* has three spines on the Ri3, whereas the females of *B. sibleti* **sp. nov.** have four spines. By contrast, Australasian species usually bear four to six spines on the Ri3 in females (Bayly, 1964; cf. *B. pilkililli* Quinlan & Bayly, 2017 and *B. shieli* Bayly, 1985; Quinlan & Bayly, 2017: fig. 4D; Bayly, 1985: fig. 1E).

Phylogenetic analysis

Among the 90 DNA samples submitted to PCR, 84 were successfully amplified, sequenced, and cleaned. The haplotypes and NCBI accession numbers are provided in Supplementary material fig. AS3 (OQ117983–OQ118066). Pairwise genetic distances between each of the B_spi_COI01–B_spi_COI13 haplotypes and the single Mont-Dore haplotype (B_sib_COI01) ranged 8.51% –9.73%. Pairwise genetic distances among the B_spi_COI01–B_spi_COI13 haplotypes of 74 individuals were 2.9% or less. A unique haplotype was found at Mont-Dore in 10 individuals (fig. 1.1.8, Supplementary material fig. AS2).

		<i>B. spinogibba</i> haplotypes from the Plaine des Lacs												
		B_spi_COI01	B_spi_COI02	B_spi_COI03	B_spi_COI04	B_spi_COI05	B_spi_COI06	B_spi_COI07	B_spi_COI08	B_spi_COI09	B_spi_COI10	B_spi_COI11	B_spi_COI12	B_spi_COI13
<i>B. spinogibba</i> haplotypes from the Plaine des Lacs	B_spi_COI02	2.13												
	B_spi_COI03	2.89	1.67											
	B_spi_COI04	2.43	1.22	1.98										
	B_spi_COI05	2.28	0.76	1.82	1.06									
	B_spi_COI06	2.74	2.13	1.37	2.43	2.28								
	B_spi_COI07	2.43	0.91	1.98	1.22	0.15	2.43							
	B_spi_COI08	2.58	1.37	2.13	0.15	1.22	2.58	1.37						
	B_spi_COI09	1.98	0.15	1.82	1.37	0.91	2.28	1.06	1.52					
	B_spi_COI10	0.61	2.13	2.89	2.43	2.28	2.43	2.43	2.58	1.98				
	B_spi_COI11	0.46	1.98	2.74	2.28	2.13	2.58	2.28	2.43	1.82	0.15			
	B_spi_COI12	2.28	0.15	1.82	1.37	0.76	2.28	0.91	1.52	0.30	2.28	2.13		
	B_spi_COI13	2.28	0.15	1.82	1.37	0.91	2.28	1.06	1.52	0.30	2.28	2.13	0.30	
	<i>B. sibleti</i> sp. nov. haplotype from Mont-Dore		8.51	8.97	8.66	9.57	9.12	8.66	8.97	9.73	9.12	8.51	8.66	9.12

Figure 1.1.8. Matrix of cytochrome C oxydase I (COI) haplotype uncorrected genetic distance (%). All haplotypes of *Boeckella spinogibba* (Plaine des Lacs) and *B. sibleti* sp. nov. (Mont-Dore) are represented.

The consensus sequences for 28S were established with 2,087 reads for *B. sibleti* sp. nov. (M223; NCBI accession: OR612784) and 545 reads for *B. spinogibba* (M156; NCBI accession: OR612785).

The concatenated tree (28S+COI) based on the concatenation used in Maturana *et al.* (2021) has been reproduced with the additional concatenated sequences from New Caledonia (fig. 1.1.9, Supplementary material fig. AS4). One concatenated sequence, identified as ‘Hap_19’ in Maturana *et al.* (2021), was excluded from the original data because, after verification, this sequence appeared likely to be a chimeric concatenation of two distinct species. Originally identified as *B. brevicaudata* (Brady, 1875), this sequence has been corrected in NCBI (accession MZ757027.3) since the publication of Maturana *et al.* (2021) and is now identified as *B. vallentini* (Scott, 1914). The final alignment obtained includes 1045 bp (673 for 28S, 372 for COI). The results are very similar to those obtained by Maturana *et al.* (2021) except that, in addition to the Australia/New Zealand group and the group of *Boeckella* from South America, Antarctica, and various sub-Antarctic

Islands, the two New Caledonian species together form a third robust monophyletic group within this genus.

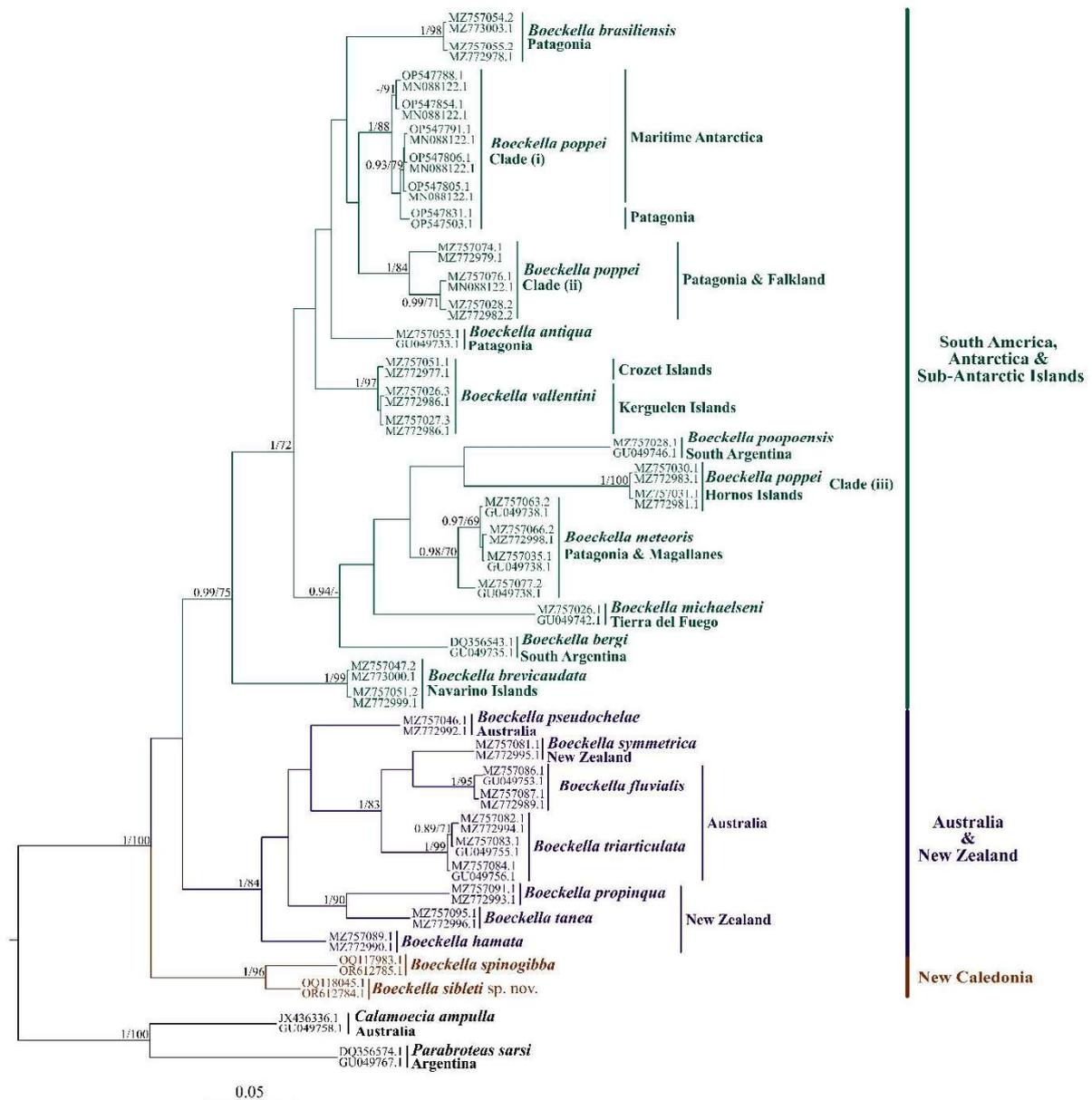


Figure 1.1.9. Concatenated phylogeny (28S+COI) of *Boeckella* in New Caledonia, other *Boeckella* species from South America, Antarctica, Australia, and New Zealand, and the outgroup *Calamoecia ampulla* - *Parabroteas sarsi* (nodes: probability support values according to Bayesian/bootstrap value for Maximum Likelihood).

The raw results from the two bPTP analyses are available in Supplementary material Figure AS6), and the COI phylogenetic tree used for one of the analyses is also available there (Supplementary material fig. AS5). *Boeckella spinogibba* and *B. sibleti* sp. nov. are distinct in the two models, with good support.

Population structure

The AMOVA test performed on populations of *B. spinogibba* from the Plaine des Lacs showed these populations had a genetic variance rate of 17.24% within ponds and 82.76% among ponds ($P = 0.001$). Exact tests of population differentiation showed that the *Boeckella* populations were generally significantly genetically differentiated among themselves.

The haplotype network of *B. spinogibba* (fig. 1.1.8, 10) shows the mutations, distribution, and abundance of haplotypes in the sampled localities. The network presented in Figure 1.1.10 has a tolerance parameter of 2; this network was chosen for its closeness to the distance matrix (fig. 1.1.8). For the sake of clarity, the network is presented without alternative links, and six median vectors have been removed as they do not directly interact with several of the sampled haplotypes (see Supplementary material fig. AS7 for the network with all alternative links and median vectors). The network on Supplementary material fig. AS7 has a tolerance parameter of 1. Its main difference with Figure 1.1.10 concerns the link between haplotypes B_spi_COI03 and B_spi_COI06. The median-joining network was unable to resolve the links between these two haplotypes and the rest of the network as regards the real number of mutations between the sequences (e.g., 18 mutations between B_spi_COI01 and B_spi_COI06 versus 22 mutations displayed on fig. 1.1.10 and 30 displayed in Supplementary material fig. AS7).

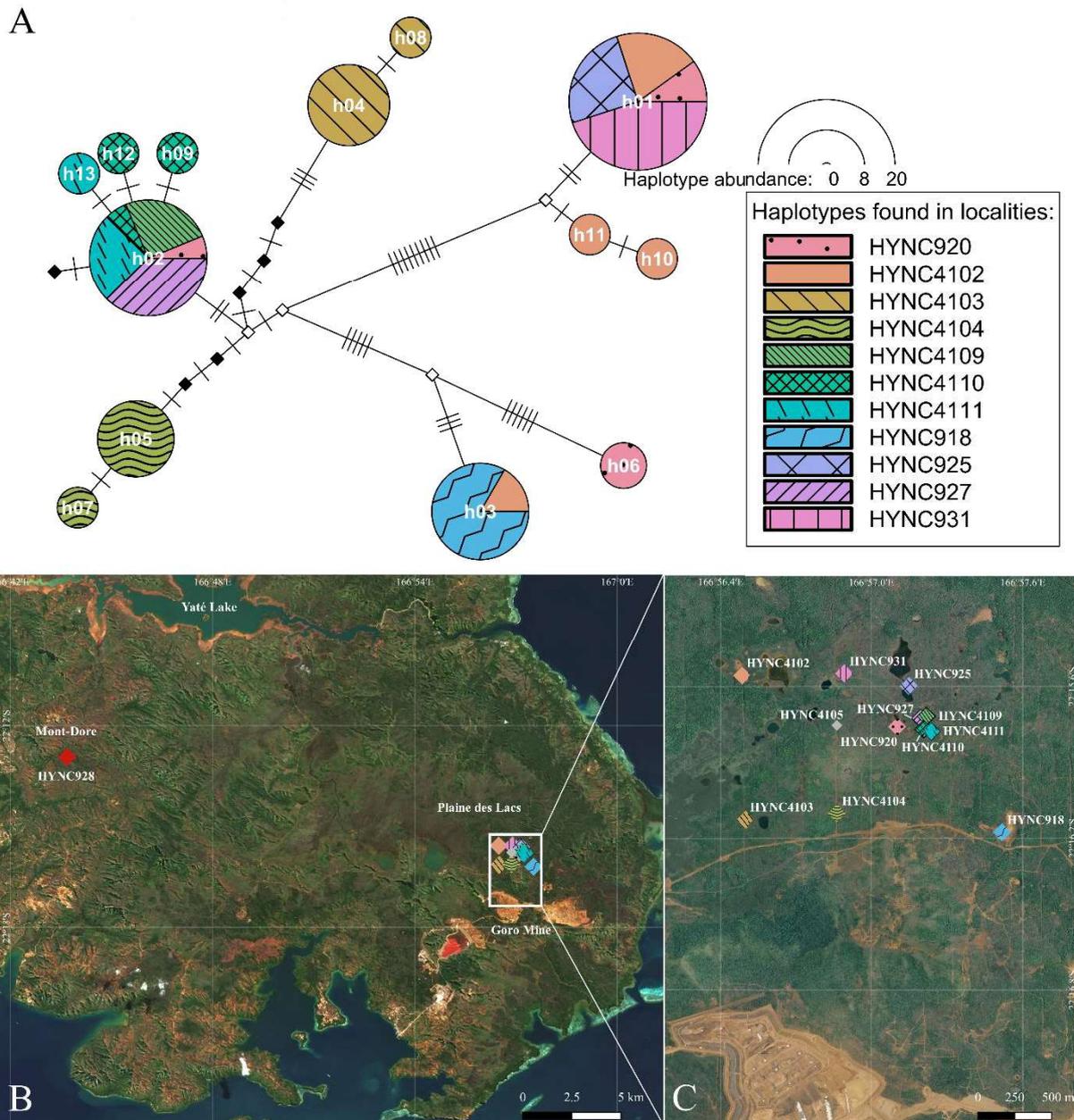


Figure 1.1.10. Haplotype network of 74 individuals of *Boeckella spinogibba* from 11 ponds in the Plaine des Lacs; haplotype names, originally of the form B_spi_COIXX, are abbreviated to hXX for the purpose of this representation (A). Satellite images and locations of sampled localities where species of *Boeckella* were found (B). Closer view of satellite images of locations where *B. spinogibba* was found in the Plaine des Lacs (outlined by a white rectangle in fig. 1.1.10B). Localities are represented with colours and patterns corresponding to those in the haplotype network (fig. 1.1.10A), except locality HYNC4105, which is represented in grey as no sequences were obtained from copepods from this pond (C).

The pond at site HYNC4102 displayed the highest haplotype richness (four haplotypes), whereas six other ponds contained only one haplotype each (sites HYNC918, HYNC925, HYNC927, HYNC928, HYNC931, and HYNC4109; Supplementary material fig. AS2). The median-joining model inferred four main haplotypes, missed in the sampling (white diamonds in fig. 1.1.10), being needed to articulate the structure of the network: one between haplotypes B_spi_COI01 and

B_spi_COI11; one between B_spi_COI02, B_spi_COI04, and B_spi_COI05; one between B_spi_COI03 and B_spi_COI06; and one at the centre of the network between the B_spi_COI03–B_spi_COI06 and the B_spi_COI01–B_spi_COI11–B_spi_COI10 haplotype complexes. The other missing haplotypes inferred by the model support alternative links (black diamonds in fig. 1.1.10, Supplementary material fig. AS7).

History of the type locality of Boeckella spinogibba

Defaye (1998: 687) described the type locality of *B. spinogibba* at the time of collection in 1996 as ‘a small pond in the Plaine des Lacs [...] on the route between Grand Lac and the fork to Parc Boisé, the pond nearest to the fork.’ As the area has been damaged by the Goro mine, roads on recent satellite images (fig. 1.1.11C) are very different from those in 1996. Aerial photographs nonetheless allowed the pond to be located at 22°17.652’S, 166°58.044’E, while also showing that it was filled in by the mine between 2007 and 2014 (filled with water: fig. 1.1.11A, B; filled with tailings from the mine: fig. 1.1.11C).

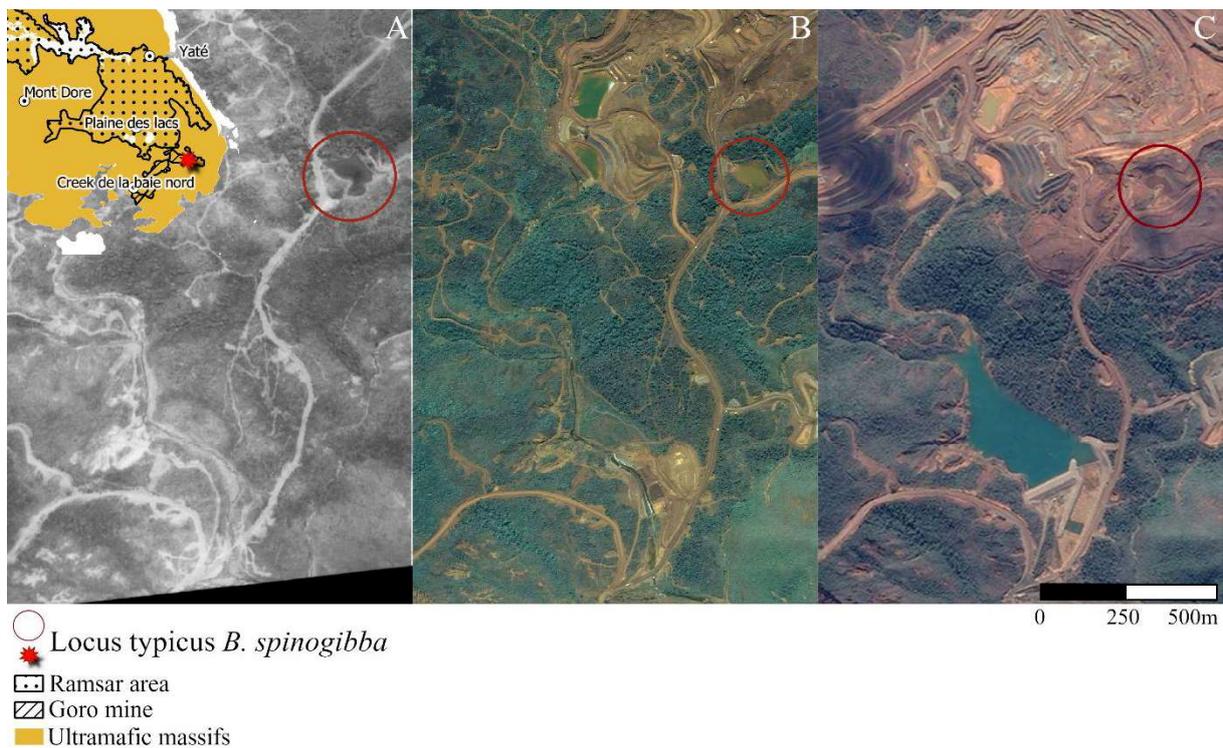


Figure 1.1.11. Aerial photographs and satellite images of the type locality of *Boeckella spinogibba* in the Goro mine area, New Caledonia in 1992 (A), 2007 (B), and 2014 (C).

Discussion

Boeckella in New Caledonia

The genus *Boeckella* is generally considered a Gondwanan taxon (Bayly & Morton, 1978; Bayly, 1992a; Maturana *et al.*, 2021) with three centres of distribution. It occurs in South America from the Andes Mountains in Colombia to the plains of southern Chile and Argentina, on sub-Antarctic islands and in coastal Antarctica; and in Australasia, especially in South Australia and New Zealand (Maturana *et al.*, 2019). *Boeckella triarticulata* Thomson, 1883 in Mongolia (Bayly, 1992a) and Italy (Alfonso & Belmonte, 2008) is the only species present outside the Gondwanan area. Our concatenated COI+28S phylogeny places the two New Caledonian species as a monophyletic branch in the phylogeny of the genus, distinct from other such groups composed of South American and Australian species (fig. 1.1.9). This suggests that the evolutionary history of the genus is more complex than previously suspected and merits further investigation.

No further information has been published on *Boeckella* in New Caledonia since the original observation of *B. spinogibba* from a single pond by Defaye (1998). We, however, found several populations of the species in the Plaine des Lacs which permitted the redescription of the left male antennule (A1) (fig. 1.1.3) and the acquisition of genomic information on the species. We also found a population of *Boeckella* at Mont-Dore, 22.5 km from the nearest originally sampled population of *B. spinogibba*, that showed 8-9.6% genetic divergence from the latter in the COI gene and clear morphological differentiation, particularly in the male fifth leg, the copulatory organ (fig. 1.1.7, Table 1.1.3) and its different body colour. According to the principles of integrative taxonomy (Padial *et al.*, 2010), several criteria, including evidence of genetic, morphological, and geographical differentiation, must be fulfilled to confirm that the species are distinct. Geographically, although the present two species are separated from each other, they are yet close enough to possibly indicate micro-endemism, something that is common in New Caledonia among, for example, plants and insects (Caesar *et al.*, 2017) and freshwater invertebrates such as gastropods (Haase & Bouchet, 1998). Such phenomenon seems to be particularly prominent on ultramafic soils (Pillon *et al.*, 2021), which is consistent with the unusual environmental conditions in New Caledonia, notably a high pH and high concentrations of heavy metals in the surface waters (Marques *et al.*, 2008). In New Caledonia, *Boeckella* has been found only in temporary freshwater ponds on ultramafic soils despite a fairly large amount of sampling in

other types of ponds. This is a peculiar habitat, with water that is toxic for most plant and animal species and particularly poor in nutrients (Kierczak *et al.*, 2021). We still lack enough data on the distribution of *Boeckella* in New Caledonia to confirm the extent of its micro-endemism.

The body colouration of copepods is mainly due to carotenoids, and freshwater copepods tend to have higher carotenoid concentrations than marine copepods. As highlighted in the review by Vilgrain *et al.* (2023), body colouration can vary widely in individuals depending on biotic and abiotic factors, making its study challenging. *Boeckella spinogibba* is blue, which is typical for copepods occurring at the water surface where this colour may be useful for camouflage, especially so for *B. spinogibba*, which lives in ponds with bright blue water (fig. 1.1.2D). Red colouration like that of *B. sibleti* **sp. nov.** occurs in copepods from all ecosystems (marine and freshwater, shallow and deep) and can provide many physiological and ecological advantages, including, protection from UV light, antioxidant activity, and higher reproduction rates (Vilgrain *et al.*, 2023).

Genetic divergence can occur over time when connectivity and the genetic flows that depend on it are no longer maintained (Rundle & Nosil, 2005). Indeed, poor connectivity between different populations can lead to speciation (Schluter, 2001). The small size and weight of freshwater zooplankton permits their transportation, either as eggs, larvae, or adult individuals, by other animals, mostly wading birds, which may get eggs stuck in their feathers or ingest them accidentally (Incagnone *et al.*, 2015; Fontaneto, 2019; Parry *et al.*, 2024). In New Caledonia, *Boeckella* has only been found in ultramafic ponds, which are mostly oligotrophic if not dystrophic, with low prey density. Wading birds occur where there is a high density of prey (Bryant, 1979; Herteux *et al.*, 2019), which greatly limits the potential for dispersion of zooplanktonic species from these ultramafic ponds by birds. Moreover, *B. spinogibba* showed high genetic variability within a very small, 2.5 km² area, with up to 2.9% COI differentiation between individuals from different ponds and with highly structured populations in terms of haplotype distribution (fig. 1.1.8, 10). AMOVA revealed that 83% of the observed genetic variance is attributed to variance among ponds, with only 17% being within ponds. The exact test also identified significant genetic divergence among ponds. As it has also been observed in *B. poppei* populations from Signy Islands, South Orkney Islands (Maturana *et al.*, 2020), these results tend to confirm weak connectivity for this genus in New Caledonia, which can support the hypothesis of an allopatric speciation between *B. spinogibba* and *B. sibleti* **sp. nov.**

Updated key by Quinlan & Bayly (2017)

Quinlan & Bayly (2017) updated the key of Australasian *Boeckella* by including *B. spinogibba* at couplet 18, using characters that are similar in *B. sibleti* **sp. nov.** as well. To incorporate the new species in the key, couplet 18 is left unchanged but it now points towards a newly created couplet 24 for the two New Caledonian species, not just to *B. spinogibba*.

‘18. Left Ri small, reduced to a triangular plate and hidden by a rounded process covered in thick spiniform hairs; left B1 noticeably longer than right B1’..... 24

24. Right Re3 with acute claw-like, rounded shape with bulge at mid-length disto-lateral spine of Re1 approximately two-thirds as long as Re2 *B. sibleti* **sp. nov.**

Right Re3 slightly curved; disto-lateral spine of Re1 less than one-third as long as Re2 *B. spinogibba* Defaye, 1998

Impact of Goro mine

The pond that was the type locality of *B. spinogibba* was destroyed and filled in between 2007 and 2014 as a result of expansion of the Goro mine (fig. 1.1.11). This event exemplifies the threat that mines pose for ponds and other freshwaters inside and outside their enclosures. Along with the destruction of the environment to build roads and other infrastructure and to extract minerals, mines in New Caledonia produce clouds of fine tailings that fill ponds outside the exploited perimeter. The open-pit complex of the Goro mine is very close to the present populations of *B. spinogibba* and the southern border of the RAMSAR area. According to Jeanpert *et al.* (2016), about 8,500 ponds occupied the 148 km² area around the future mining site in 2006 before the construction of the mine. Supplementary studies are needed in such areas, in particular in the RAMSAR area, where ponds and many other habitats are threatened by mining activities (Pascal *et al.*, 2008).

Perspectives and limitations

New Caledonia has been extensively sampled, but ultramafic soils are often less accessible. Our discovery of a new species in Mont-Dore suggests that other still unknown species may be found in unexplored parts of the Plaine des Lacs and elsewhere in New Caledonia.

The concatenated phylogeny computed in this study was mainly constructed from sequences concatenated by other workers, and the sources did not specify which individual copepods provided the sequences. Hence, there is a possibility that COI and 28S sequences of different individuals were concatenated for some portion of the Australian, New Zealand, Antarctic, and South American species. Because the mitochondrial marker (COI) we used only provides information about the maternal lineage and only two nuclear (28S) sequences were retrieved from shotgun sequencing, further exploration of the genetic structure of *Boeckella* populations in New Caledonia, notably using other genes, is desirable. Microsatellite sequences of populations and species could also be used to test the hypothesis of isolation by distance and to help assess the dispersal potential of *Boeckella* in New Caledonia (Teske *et al.*, 2018).

Acknowledgements

The New Caledonia Hydrobiological expeditions 2016–2018 were conducted under the ‘La Planète Revisitée’ program implemented by the Muséum national d’Histoire naturelle (MNHN), Paris in partnership with the Conservatoire d’Espaces naturels (CEN). We deeply thank Pascale Joannot, head of the expeditions program at MNHN, and Phillippe Bouchet, the expedition leader. These expeditions were operated under permits issued by New Caledonia’s Province Sud (under APA_NCPS_2017_028) and Province Nord, and the organizers thank Emmanuel Coutures and Isabelle Jurquet (Province Sud) and Jean-Jérôme Cassan and Yannick Monlouis (Province Nord) for their help in obtaining these permits. We thank Sébastien Faninoz and Alice Leblond. We also thank Céline Bonillo for her help with DNA extractions and PCR, and Marion Mennesson for the idea of using a haplotype network and AMOVA to represent the genetic diversity of *Boeckella spinogibba*. We are grateful to Géraldine Toutirais (Plateau technique de Microscopie Électronique (PTME) du Muséum national d’Histoire naturelle) for her help in the acquisition of SEM images. We also thank Valentin de Mazancourt for sharing photographs of the dry Mont-Dore pond and Claudia Maturana for sharing genetic concatenation data and for providing much advice about her methodology. Finally, we thank the editors and anonymous reviewers for improvements to the manuscript. Author contributions: CR (text, laboratory manipulations, data cleaning and analysis), N.R (sampling, phylogeny, SEM images, some field photography, and detailed review of the text); NC (sampling and information and photos of habitats). Funding for molecular biology manipulations were granted to NR by Sorbonne Université through the Emergence (2017-2018) call for projects. Funding for ‘La Planète Revisitée’ expeditions were provided by the Gouvernement de la Nouvelle-Calédonie, Province Sud, Province Nord, Office des Postes et Télécommunications (OPT), Maison de la Nouvelle-Calédonie, and the Ministère Français des Outre-mer.

Data availability statement

AB1 files from sequencing and datasets generated during and/or analysed during the current study are available in the Data.InDoRes repository [<https://doi.org/10.48579/PRO/DGVWPW>].

Remarkable genetic and specific diversification in *Boeckella* from New Caledonia suggest strong micro-endemism

Royaux Coline | Charpin Nicolas | Mary Nathalie | Busson Frédéric | Bonillo Céline | Rabet Nicolas

Soumis, Freshwater Biology

Abstract

The relative isolation of islands from mainland makes insular biodiversity both fascinating and particularly sensitive to perturbations. New Caledonia is a tropical archipelago hosting among the highest observed levels of endemism that knows many threats notably induced by mining activities. Several areas of the archipelago comprises numerous freshwater ponds and sinkholes that functions as insular systems. In these environments, planktonic crustaceans such as the copepod genus *Boeckella* are abundant. To date, only two species of *Boeckella* have been described in New Caledonia: *B. spinogibba* and *B. sibletii*. In this study, the high genetic diversity with geographic structuration found in *Boeckella* populations of the archipelago highlights a potentially low dispersion capacity. Considering all *Boeckella* populations in New Caledonia are identified on soils with high heavy metal concentrations that are extensively exploited by mining activity since 1873, lentic freshwater zooplanktonic populations are facing a significant threat to their conservation. Additionally, four new potential species are discussed, tripling the specific diversity of the genus previously known in the archipelago. The narrow distributions of these species of *Boeckella* might be associated to a commonly observed phenomenon in New Caledonia: micro-endemism. All these observations tend to validate most insular theories on highly isolated insular systems. Finally, a phylogenetic analysis of the genus using two genetic markers (28S, and Cytochrome Oxidase I), tends to confirm that the species from the archipelago might form a single clade that is a sister group to all other *Boeckella* species worldwide, which changes the evolutionary history known of the genus.

Keywords

DNA sequences, crustaceans, ponds, insularity, ultramafic massifs

Introduction

Biodiversity on islands is extensively studied in biology and biogeography since the premises of ecology as a discipline (Darwin, 1859; MacArthur & Wilson, 1963; Gorman, 2012; Whittaker *et al.*, 2023). The plurality of islands' age, size, climate, geology and history makes of each a unique open-air laboratory permitting to explain not only island-related dynamics but also large-scale processes (Berry, 1983; Kueffer & Kinney, 2017). On insular systems, processes such as bottle-necks and founder effects tend to decrease species and genetic diversity locally but relative reproductive isolation from mainland often lead to higher endemism (Chen & He, 2009; Veron *et al.*, 2019). As well as terrestrial islands, lentic freshwater ecosystems represent aquatic islands separated by an uninhabitable dry sea and are submitted to similar ecological processes regarding diversity and dispersion. In this context, the study of pond biodiversity in Nouvelle-Calédonie (New Caledonia, Kanaky) involves studying nested insular systems.

Biodiversity in freshwater ponds is highly dependent on several characteristics of the environment such as *e.g.* the potential cyclic drying and overflowing of the ponds or the composition, turbidity, temperature, and pH of the water. These characteristics are the results of environmental factors such as *e.g.* precipitation, soil type, surrounding vegetation, and potential anthropogenic perturbations (Bagella *et al.*, 2010; Feld *et al.*, 2016). The high abundance of freshwater zooplankton in these lentic freshwaters renders them suitable as bio-indicators of the environment (Seminara *et al.*, 2015; Leonardo *et al.*, 2016; Pattanayak *et al.*, 2020). The establishment of efficient and trustable bio-indicators requires a deep knowledge of the environment and its biodiversity. However, ponds remain some of the less studied freshwater environment in scientific literature, especially in tropical environments (Faghihinia *et al.*, 2021).

Grande Terre, the main island of New Caledonia, is a continental island that separated from Australia approximately 105 Mya during the Gondwana break-up (Maurizot & Campbell, 2020). Gondwana is the supercontinent formed 600 Mya by today's South America, Antarctica, Australia, Zealandia and Africa continents and the Arabian and Indian plates. Since its separation from mainland, the island has known several submersion-emersion episodes and there is strong evidence of continuous emersion since 25 Mya (Maurizot & Campbell, 2020). Given that no terrestrial mammals native to the archipelago are known to exist apart from bats

(Chazeau, 1993; Affholder *et al.*, 2024), it is posited that the native biota recolonised the archipelago following the last emersion and subsequently diverged.

Grande Terre and several surrounding islands host heavy metal-rich massifs called “ultramafic massifs”. These environments are known to host significant endemism on the archipelago with notably 74-90% of endemic plant species (Jaffré *et al.*, 2010; Nattier *et al.*, 2013; Isnard *et al.*, 2016; Pillon *et al.*, 2021). Their high concentrations in heavy metals also represent an important resource and extensive mining activity exists on the archipelago since the 19th century which leads to biodiversity conservation conflicts. To cope for such conflicts, a Ramsar (convention on wetlands) area has been notably created in the ultramafic massifs of the south (fig. 1.2.1).

The copepod genus *Boeckella* is widely distributed across the southern hemisphere (South America, Australia, Zealandia, Antarctica and sub-antarctic islands). In literature, *Boeckella* is considered as an old taxon with Gondwanan origins despite its absence from South Africa (Adamowicz *et al.*, 2010; Maturana *et al.*, 2021). Two species of *Boeckella* are described from Grande Terre, *B. spinogibba* and *B. sibleti*, both are endemic to the island and exclusively reported from ultramafic massifs (Defaye, 1998; Royaux *et al.*, 2024).

In this study, a large sampling of ponds in New Caledonia permits to evaluate the specific and genetic diversity of the genus *Boeckella* in the archipelago. The different species are identified using morphologic, genetic and geographic criteria. Additionally, the genetic structure of sampled *B. spinogibba* and *B. sibleti* populations permits to get further insights regarding connectivity in ultramafic ponds. Finally, the origin of the genus is discussed in the light of scientific literature and our phylogenetic analysis.

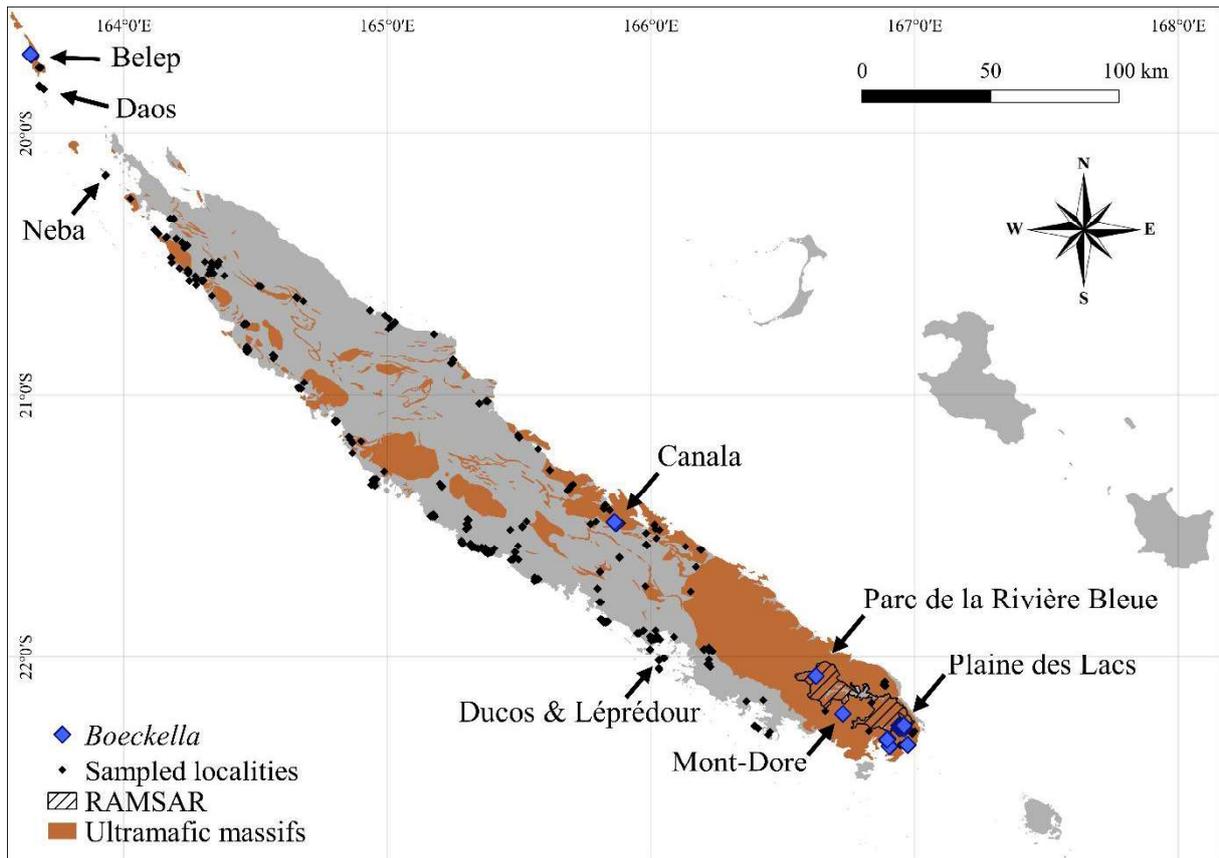


Figure 1.2.1 – Sampled localities in New Caledonia between 2016 and 2023. *Boeckella* individuals were found in localities in large blue squares.

Material and Methods

Study design

In total, 535 sites were sampled on Grande Terre, Belep, Léprédour, Ducos, Neba and Daos islands between 2016 and 2023 (fig. 1.2.1; tab. 1.2.1). Samples collected between 2016 and 2018 were collected during three “La Planète Revisitée” expeditions. Since 2020, samples are collected by the environmental consulting firm “Ethyc’o” and the organisation “Vies d’Ô douce”. Collection of individuals is made using a plankton net (NHBS, Bonn, Germany) with a frame diameter of 25 cm and a 55 cm-long bag of 200 µm nylon mesh with a filter at the tip. As soon as possible, samples were transferred in 80-90% ethanol at 4°C to avoid DNA alteration.

Table 1.2.1 - Sampled localities where *Boeckella* were found, coordinates and aliases.

Locality single identifier	Area	Longitude (x)	Latitude (y)	Other names given to the location
ETNC109	Plaine des Lacs	166°57.552E	22°15.593S	DP46
ETNC113	Plaine des Lacs	166°53.908E	22°18.567S	DT70
ETNC114	Plaine des Lacs	166°53.692E	22°18.743S	DT73
ETNC115	Plaine des Lacs	166°54.287E	22°20.227S	Doline pilote
ETNC117	Plaine des Lacs	166°57.120E	22°15.743S	DP52
ETNC120	Parc de la Rivière Bleue	166°37.518E	22°4.385S	Pourina ; PPRB
ETNC124	Plaine des Lacs	166°58.402E	22°19.913S	DP60
ETNC210	Belep	163°38.749E	19°41.877S	Belep10
ETNC211	Belep	163°38.895E	19°42.031S	Belep11
HYNC918	Plaine des Lacs	166°57.5412E	22°16.1982S	HYNC4101
HYNC920	Plaine des Lacs	166°57.126E	22°15.7692S	HYNC4005
HYNC925	Plaine des Lacs	166°57.1554E	22°15.6642S	HYNC4107 ; Lac en long
HYNC927	Plaine des Lacs	166°57.1938E	22°15.726S	HYNC4108
HYNC928	Mont-Dore	166°43.662E	22°12.93S	HYNC4004
HYNC931	Plaine des Lacs	166°56.901E	22°15.5502S	HYNC4106 ; DT53
HYNC2604	Canala	165°52.020E	21°29.253S	
HYNC2608	Canala	165°51.657E	21°28.923S	
HYNC4102	Plaine des Lacs	166°56.4846E	22°15.5556S	
HYNC4103	Plaine des Lacs	166°56.4996E	22°16.125S	
HYNC4104	Plaine des Lacs	166°56.8638E	22°16.101S	
HYNC4105	Plaine des Lacs	166°56.865E	22°15.7542S	
HYNC4109	Plaine des Lacs	166°57.201E	22°15.7182S	
HYNC4110	Plaine des Lacs	166°57.207E	22°15.7686S	
HYNC4111	Plaine des Lacs	166°57.2382E	22°15.7176S	Doline croissante

DNA extraction and PCR

DNA was extracted from 1 to 24 individuals from ponds where enough individuals of *Boeckella* were found (127 individuals in total). Whole bodies were used for lysis and DNA was extracted from tissues with the QIAamp® DNA Micro Kit (Qiagen GmbH, Hilden, Germany).

Amplification of the mitochondrial gene COI (cytochrome C oxidase I) were performed using the following PCR (polymerase chain reaction) cycle: 5 min at 94°C; 40 cycles with 30 sec - 1 min at 94°C, 30 sec - 1 min at 45 - 52°C and 1 min at 72°C; and 8 min at 72°C. PCR products were then subjected to gel

electrophoresis to ascertain appropriate amplification. The expected size of DNA strands without primers was 658 bp. Detailed protocols are available on protocols.io (DNA extraction: <https://dx.doi.org/10.17504/protocols.io.36wgqnm9ygk5/v1>; PCR: <https://dx.doi.org/10.17504/protocols.io.n92ld8q57v5b/v1>).

Sequencing and phylogenetic analysis

In total, 76 PCR products with amplified COI sequences were sent to Eurofins Scientific (Luxembourg City, Luxembourg) for Sanger sequencing (Sanger *et al.*, 1977) and one amplified COI sequence was obtained with amplicon pools subjected to Illumina Miseq sequencing at Plateforme iGenSeq, Institut du Cerveau (ICM) (Hôpital Pitié Salpêtrière, Paris) with 2 × 25 million reads cartridge of 300 bases each (30-45 libraries per run). The resulting Sanger sequences were cleaned and consensus sequences were determined using Galaxy (<https://usegalaxy.fr>; The Galaxy Community, 2022) and the amplicon sequence was mapped with *B. spinogibba* – OQ118064.1 using Geneious under standard conditions (Geneious® 11.1.4, Dotmatics; <https://www.geneious.com/>). The Galaxy workflow and history for cleaning sequences and annotations on manual corrections are accessible on the french Galaxy platform (https://usegalaxy.fr/u/coline_royaux/w/sanger1--from-ab1-to-consensus-and-blast-no-final-alignment; https://usegalaxy.fr/u/coline_royaux/h/boeck-sanger-2024-clean).

Additionally, 7 DNA extracts were prepared for NGS (next-generation sequencing). Library preps were performed with a Nextera XT kit (Illumina; <https://www.illumina.com/>) using fragmentation and Illumina index adapter ligation. The final library was qualified and quantified before sequencing on Illumina Miseq at Plateforme iGenSeq, Institut du Cerveau (ICM) (Hôpital Pitié Salpêtrière, Paris) with 2 × 25 million reads cartridge of 300 bases each (30-45 libraries per run). Assembly and sequencing error correction was performed on all raw sequences using Geneious (Geneious® 11.1.4, Dotmatics; <https://www.geneious.com/>).

All 83 COI sequences and 4 28S sequences obtained are available in GenBank (see supporting tab. BS1 for NCBI accessions).

Gene concatenations of ribosomal RNA large subunit 28S and COI from the present study were complemented by concatenations published by Royaux *et al.*

(2024) and Maturana *et al.* (2021). The phylogeny was computed on 44 individual sequence concatenations representing 16 species of *Boeckella* from Australia, New Zealand, South America, Antarctica and sub-antarctic islands, six populations including two already known species and four unstudied populations from Grande Terre and Belep, and two external groups *Calamoecia ampulla* and *Parabroteas sarsi* (supporting tab. BS2). *Calamoecia* and *Parabroteas* form a monophyletic group closest to *Boeckella* in the family Centropagidae (Adamowicz *et al.*, 2010). After alignment and verification of the absence of stop codon in the reading phase, COI sequences were subjected to a Gblock (default options) on the LIRMM web platform (Dereeper *et al.*, 2008). The concatenations of COI and hand-delimited 28S sequences and estimation of evolution models with distinct codon position for COI through “Partition finder 2” for Bayesian analysis and “model finder” for maximum likelihood analysis were performed on PhyloSuite v1.2.3 (Lanfear *et al.*, 2016; Kalyaanamoorthy *et al.*, 2017; Zhang *et al.*, 2020; Xiang *et al.*, 2023).

The Bayesian analysis was computed with MrBayes 3.2.7 (Ronquist *et al.*, 2012) using the models TIM+I+G (28S), K81UF+G, TRN+I, and F81+I for each codon position. The maximum likelihood analysis was performed using 1,000 bootstrap resamplings with IQTree 2.2.0 (Minh *et al.*, 2020) using the models TNe+I (28S), K3Pu+F+R2, and TH+F+I for each codon positions (COI). The multilocus phylogenetic tree was formatted using FigTree 1.4.4 (<https://tree.bio.ed.ac.uk/software/figtree>) and Inkscape 1.2.2 (The Inkscape Project).

Finally, the species delimitation model was computed on the bPTP server (Bayesian Poisson Tree Processes; <https://species.h-its.org/>) using the concatenated 28S+COI phylogenetic tree with 100,000 MCMC generations and exclusion of outgroups.

Morphological analysis

Male fifth pair of legs (P5) is the most important discriminant character in *Boeckella* species (Bayly, 1992b,a). After dissection under a binocular microscope (SZX7, Olympus, Tokyo, Japan), P5 of specimens were prepared for scanning electron microscope (SEM) examination with sequential dehydration in 96% and 100% ethanol followed by critical point-drying in liquid CO₂ (Emitech K850, Quorum Technologies, Lewes, UK). P5 were mounted on flat aluminium stubs (12.7 mm diameter) with double-sided conductive carbon adhesive tape and sputter-

coated with platinum (EM ACE600, Leica) to be examined at 15kV accelerating voltage and at 250-500x magnification in a SEM (SU3500, Hitachi, Tokyo, Japan).

Population structure of Boeckella species

In addition with sequences from the present study, all available sequences of *Boeckella* from Grande Terre were retrieved from GenBank (n = 84, see supporting tab. BS3). Data preparation was made on Galaxy (workflow: https://ecology.usegalaxy.eu/u/coline_royaux/w/population-genetics-data-prep; history: https://ecology.usegalaxy.eu/u/coline_royaux/h/boeck-pop-gen-data-prep). The genetic structure of each *Boeckella* species were analysed with an exact test of population differentiation and an analysis of molecular variance (AMOVA) using R 4.2.2 (CRAN) packages mmod 1.3.3 (Winter, 2012) and ade4 1.7-22 (Dray & Dufour, 2007), respectively. Significance of the AMOVA was computed with the random matrix permutations technique. Haplotype networks were computed with median-joining and statistical parsimony techniques using R package pegas 1.3 (Paradis, 2010). R scripts are available on GitHub (https://github.com/ColineRoyaux/population_genetics). The median-joining algorithm occasionally computes hypothetical haplotypes (named “median vectors”). These median vectors represent potentially unsampled haplotypes that are useful to resolve the network.

Results

Phylogenetic analysis

Consensus sequences obtained with NGS were established with 612554 – 1852528 reads (supporting tab. BS4). The NGS sequences from *B. spinogibba*, *B. sp. A*, *B. sp. C*, and *B. sp. D* permitted to retrieve 28S and COI from single DNA extracts. Along with the 28S sequence, only fragments of the COI sequence were retrieved by NGS for *B. sp. B*, the concatenation for this species consists of the 28S sequence obtained in NGS and the single COI haplotype obtained in Sanger for this species. These sequences were obtained from distinct DNA extracts.

The final alignment for the concatenated phylogeny is composed of 673 bp for 28S and 372 bp for COI, 1045 bp in total. *Boeckella* species from New Caledonia are clearly distinct from Australia/New Zealand and South America/Antarctica/sub-antarctic islands groups and form a robust monophyletic group within the genus (fig. 1.2.2, 1.2.3).

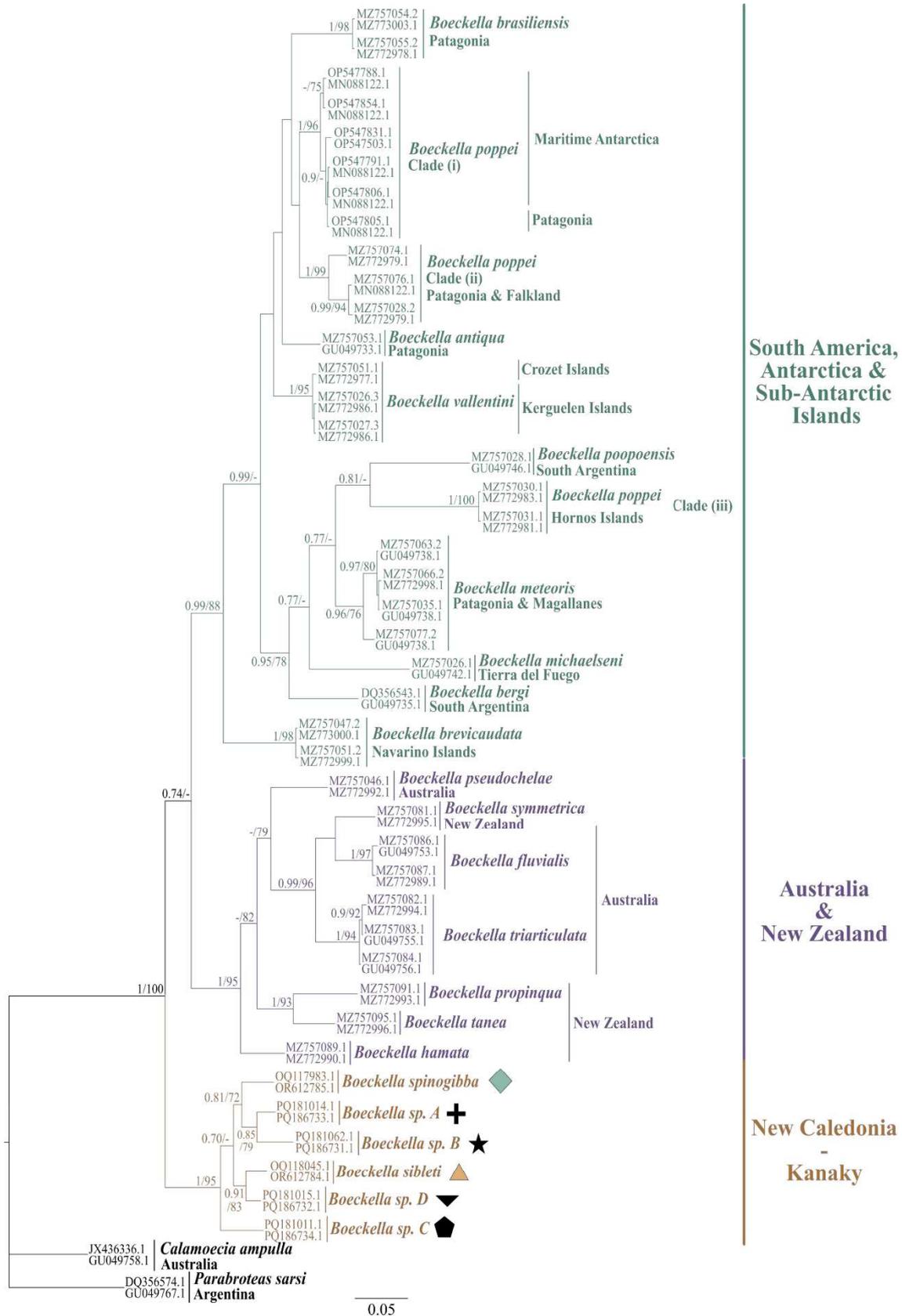


Figure 1.2.2 – Concatenated phylogeny (28S+COI) of the *Boeckella* genus and the outgroup *Calamoecia ampulla* – *Parabroteas sarsi* (nodes: probability support values according to Bayesian/bootstraps value for Maximum Likelihood). Symbols next to *Boeckella* species from New Caledonia corresponds to symbols on fig. 1.2.3.

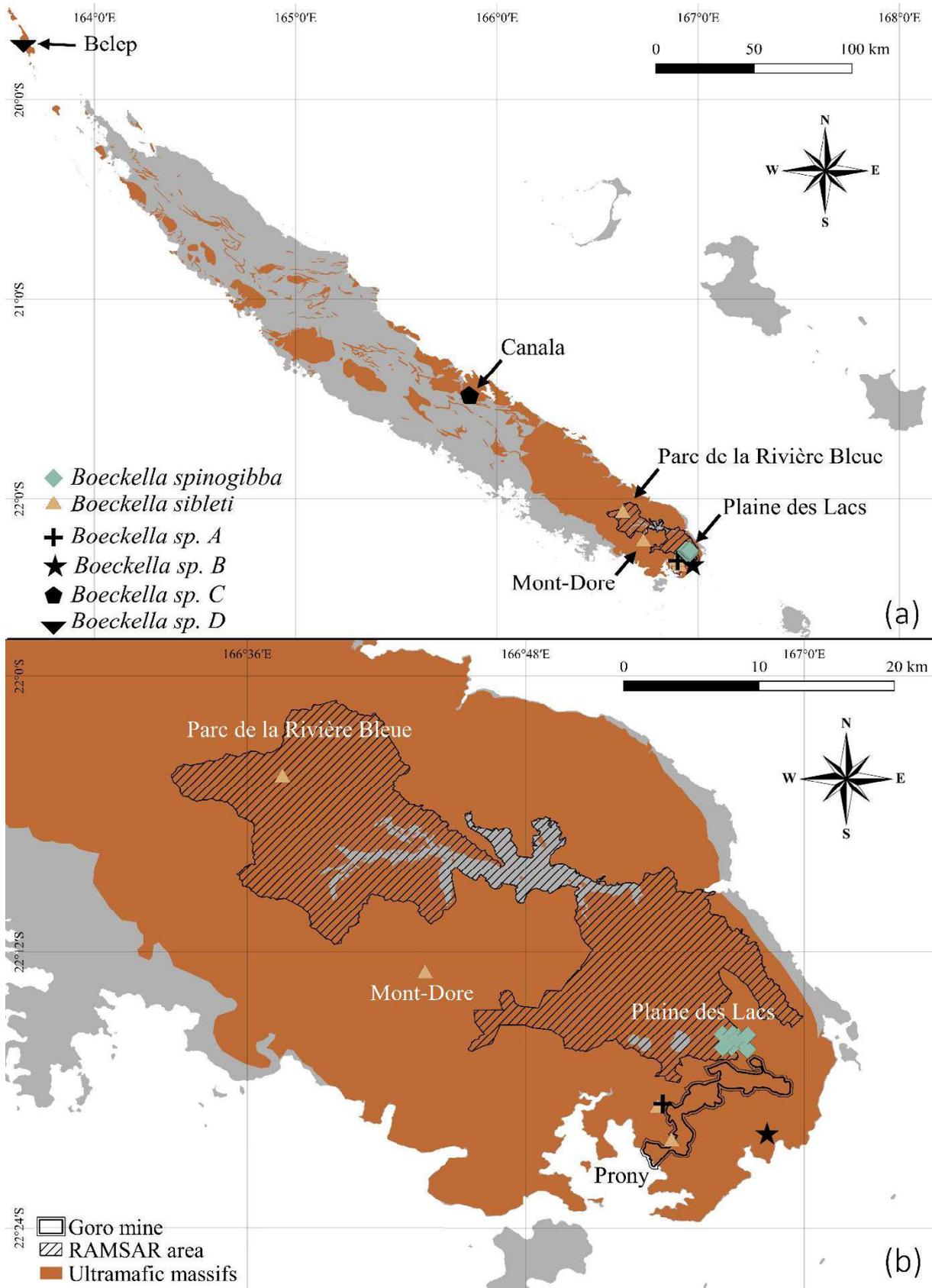


Figure 1.2.3 – Map of sampled localities (a) of all *Boeckella* species on Grande Terre and Belep; (b) and of *Boeckella* species from the south of Grande Terre.

The bPTP analysis determined species *B. spinogibba*, *B. sp. A*, *B. sp. B*, *B. sp. C* and *B. sp. D* as distinct species with good support (see raw output in supporting fig. BS5). Pairwise genetic uncorrected distance between each sampled clade of *Boeckella* for COI and 28S markers are listed in table 1.2.2 (detail between each haplotype in supporting tab. BS6). *B. sp. C* is most distant from all other species in COI marker and *B. sibleti* is most distant from all other species in 28S marker. *B. sibleti* and *B. sp. D* are closest in COI marker and *B. sp. C* and *B. sp. D* are closest in 28S marker.

Table 1.2.2 - Uncorrected distance matrix of sequenced COI fragment and 28S gene in *Boeckella* species from New Caledonia

		<i>Boeckella spinogibba</i>	<i>Boeckella sibleti</i>	<i>Boeckella sp. A</i>	<i>Boeckella sp. B</i>	<i>Boeckella sp. C</i>	<i>Boeckella sp. D</i>
<i>Boeckella spinogibba</i>	COI	0 - 2.89 %					
<i>Boeckella sibleti</i>	COI	8.51 - 10.3 %	0 - 4.41 %				
	28S	0.91 %	0 %				
<i>Boeckella sp. A</i>	COI	9.73 - 10.9 %	7.75 - 8.51 %	0 %			
	28S	0.61 %	1.06 %	0 %			
<i>Boeckella sp. B</i>	COI	9.27 - 9.73 %	6.08 - 7.6 %	6.53 %	0 %		
	28S	0.46 %	1.06 %	0.91 %	0 %		
<i>Boeckella sp. C</i>	COI	10.79 - 12 %	10.3 - 11.4 %	12.16 %	10.49 %	0 %	
	28S	0.46 %	0.76 %	0.76 %	0.91 %	0 %	
<i>Boeckella sp. D</i>	COI	9.12 - 9.88 %	4.86 5.47 %	- 8.21 %	6.23 %	10.33 %	0 %
	28S	0.30 %	0.61 %	0.61 %	0.76 %	0.15 %	0 %

Morphologic analysis

In SEM images, male P5 of sampled *Boeckella* are oriented ventrally except for *B. sp. B* that is oriented dorsally (fig. 1.2.4). Many morphological differences are observable, the most noticeable are the shape of right exopod 3 (indicated by 1 and colored in blue on fig. 1.2.4), the shape and attachment to the leg of the curved process on basipodite 2 (indicated by 2 and colored in red on fig. 1.2.4), and the length of the spine on exopod 1 (indicated by 3 and colored in orange on fig. 1.2.4). The description of each morphological character is presented in table 1.2.3.

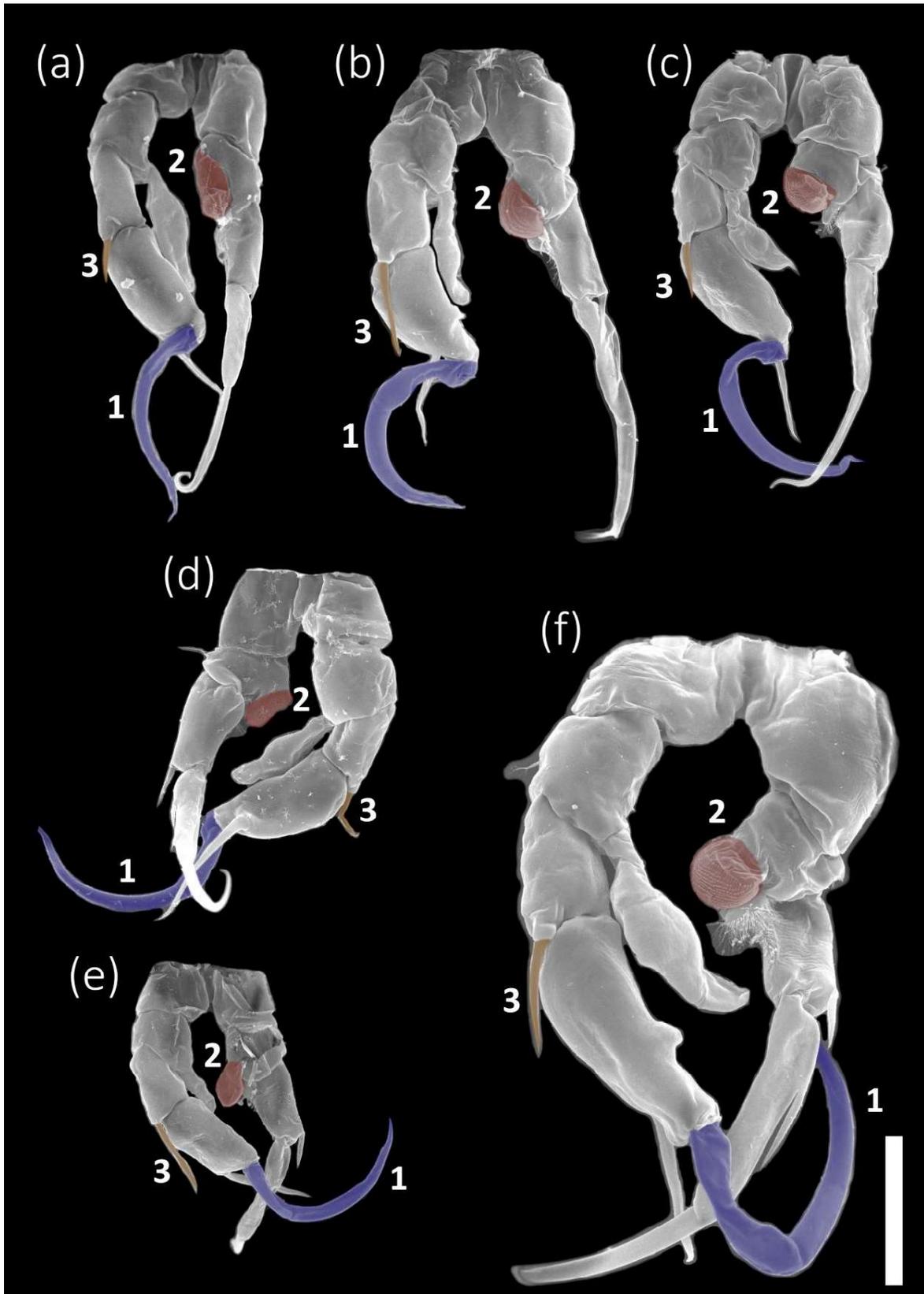


Figure 1.2.4 – Detoured SEM images of male P5 dissected from *Boeckella* species. Right exopod 3 is highlighted in blue and indicated by 1; the curved process on basipodite 2 is highlighted in red and indicated by 2; the spine on exopod 1 is highlighted in orange and indicated by 3. Scale bar = 100µm. (a) *B. spinogibba* from Plaine des Lacs (b) *B. sibleti* from Mont Dore (c) *B. sp. A* from ETNC113 (d) *B. sp. B* from ETNC124 (e) *B. sp. C* from Canala (f) *B. sp. D* from Belep.

Table 1.2.3 – Differential diagnosis based on the morphology of male P5 of *Boeckella*.

	<i>B. spinogibba</i>	<i>B. sibleti</i>	<i>B. sp. A</i>	<i>B. sp. B</i>	<i>B. sp. C</i>	<i>B. sp. D</i>
Right Exopod 3 (fig. 1.2.4: Blue; 1)	slightly curved	highly curved C-shaped	highly curved proximally and slightly curved distally	curved	curved	acute angle U-shaped
Curved process on left Basipodite 2 (fig. 1.2.4: Red; 2)	flat and attached	plump and detached	plump and directed towards posterior end	flat with two bumps	flat and detached, shaped as a pouch	plump and directed towards anterior end
Spine on right Exopod 1 (fig. 1.2.4: Orange; 3)	short	long	short	short	long	long

Populations' structures

In total, 165 COI gene sequences of *Boeckella* were retrieved to analyse the genetic structure of each population (NCBI accession numbers provided in supporting tab. BS1, BS3). One sequence of *B. sibleti* (COI_T1C3_boeck_E115) and one sequences of *B. spinogibba* (T3A7_boeck_E117) were ambiguous between two haplotypes, these sequences were removed from the analyses. The genetic structure of the four undescribed species (A, B, C and D) cannot be analysed at this time, as they are represented by a single haplotype and each were found in only one or two ponds (tab. 1.2.4).

For the species *B. spinogibba*, haplotype richness of 18 was found within 13 ponds (tab. 1.2.4), the genetic differentiation between these haplotypes is up to 2.86% (tab. 1.2.2). Among these haplotypes, 13 have been found in only one pond, the haplotype B_spi_COI02 has been observed in 7 ponds and is the most widespread. The ponds with the highest haplotype richness are HYNC4102 and ETNC109 (n = 4; tab. 1.2.4). According to the AMOVA, *B. spinogibba* populations' genetic variance rate is 15.4% explained by variance within ponds and 84.6% explained by variance between ponds (p = 0.001). The exact test of population differentiation found significant genetic differentiation of *B. spinogibba* populations (p = 0.0005). The median-joining haplotype network of *B. spinogibba* (fig. 1.2.5a; $\epsilon = 2$) has computed 15 median vectors and is presented without alternative links and four median vectors that are not directly linked to a sampled haplotype to ensure its readability (see supporting fig. BS7 for a full representation; parsimony network supporting fig. BS8). The algorithm has not been able to resolve the links between all haplotypes in the network, especially between B_spi_COI01 and all haplotypes

other than B_spi_COI02, B_spi_COI10 and B_spi_COI11 (*e.g.*, 18 mutations between B_spi_COI01 and B_spi_COI06 versus 24 mutations displayed on fig. 1.2.5a). The network represents that haplotypes B_spi_COI05 and B_spi_COI07 are exclusive to pond HYNC4104, with a comparable exclusivity noted for haplotypes B_spi_COI04 and B_spi_COI08 in pond HYNC4103. These two ponds are the most distant from the remainder (fig. 1.2.5a, d).

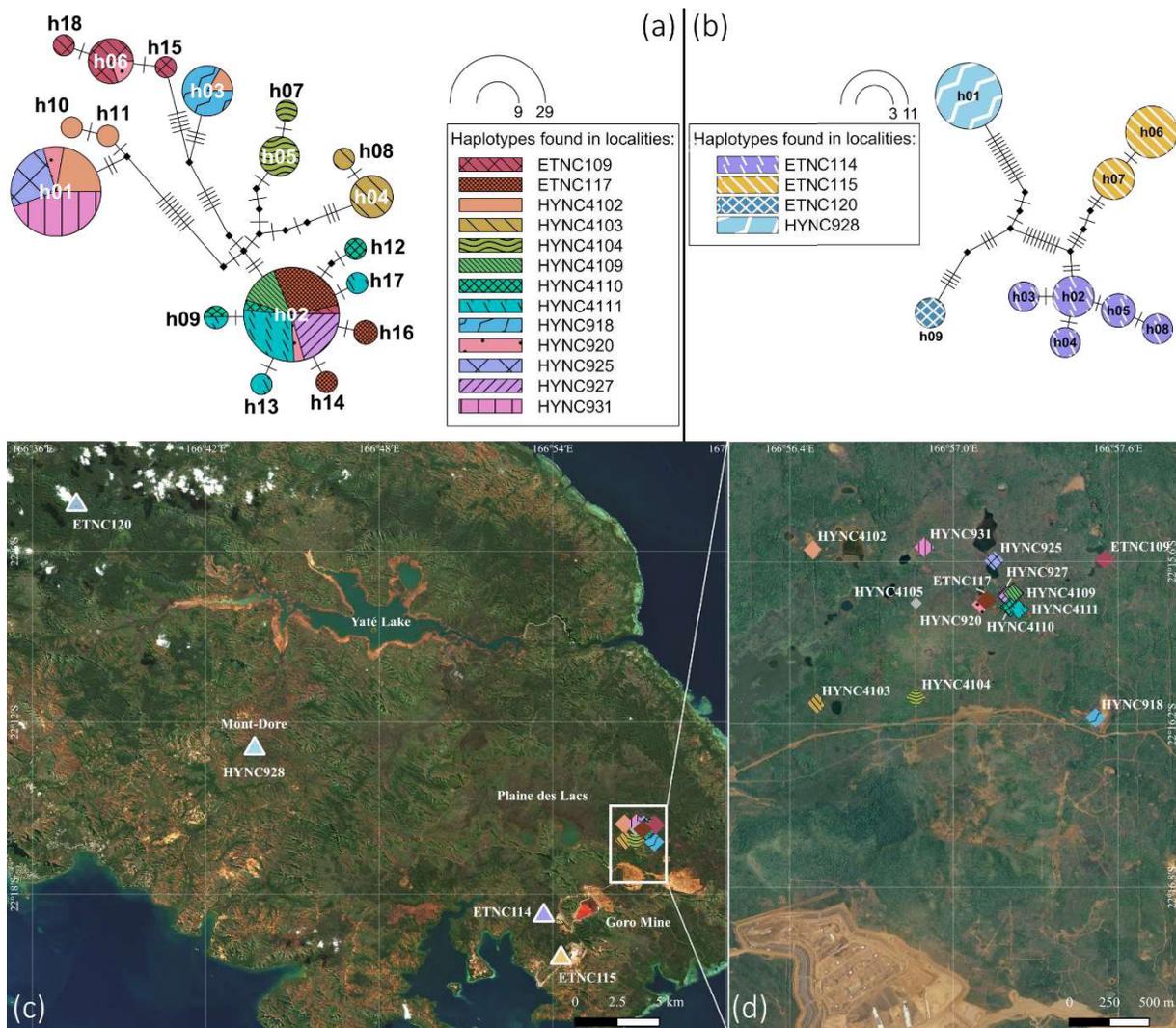


Figure 1.2.5 – (a) Haplotype network of *B. spinogibba*, black squares are median vectors; haplotype names, originally of the form B_spi_COIXX, are abbreviated to hXX for the purpose of this representation. (b) Haplotype network of *B. sibleti*, black squares are median vectors; haplotype names, originally of the form B_sib_COIXX, are abbreviated to hXX for the purpose of this representation. (c - d) Satellite images and sampled localities where *Boeckella* individuals were sampled, squares for *B. spinogibba* and triangles for *B. sibleti*. Localities are represented by colours and patterns corresponding to the haplotype networks (fig. 1.2.5a - b). Locality HYNC4105 is shown in grey as no sequences were obtained from copepods from this pond.

For the species *B. sibleti*, haplotype richness of 8 was found within 4 ponds the genetic differentiation between these haplotypes is up to 4.41%. All haplotypes were found in only one pond. The pond with the highest haplotype richness is ETNC114 (n = 5). According to the AMOVA, *B. sibleti* populations' genetic variance rate is 2.4% explained by variance within ponds and 97.6% explained by variance between ponds (p = 0.001). The exact test of population differentiation found significant genetic differentiation of *B. sibleti* populations (p = 0.0005). The median-joining haplotype network of *B. sibleti* (fig. 1.2.5b; $\epsilon = 2$) has computed ten median vectors and is presented without alternative links and three median vectors

that are not directly linked to a sampled haplotype to ensure its readability (see supporting fig. BS9 for a full representation; parsimony network supporting fig. BS10). The algorithm has not been able to resolve the links between all haplotypes in the network. The differences between the mutations observed in sequences and the mutations displayed on the network are mostly minor apart from mutations between B_sib_COI09 and B_sib_COI07 with 16 observed versus 22 displayed. The network represents haplotypes B_sib_COI01 and B_sib_COI09 that seem to be particularly distant from the remainder. These haplotypes are found in ponds most distant from the others (fig. 1.2.5b, c).

Discussion

Species delimitation

Our results regarding the genetic distance between sampled populations of *Boeckella* suggest *B. spinogibba*, *B. sibleti*, *B. sp. A*, *B. sp. B*, and *B. sp. C* are potential distinct species. The differential diagnosis based on male fifth legs (P5) of each population (tab. 1.2.3) confirmed the delimitation of the species. *B. spinogibba*, found in the Plaine des Lacs, has the less curved right third exopod (fig. 1.2.4a: Blue; 1) and the shortest spine on right first exopod (fig. 1.2.4a: Orange; 3). *B. sibleti*, found from the Goro Mine to the Parc de la Rivière Bleue, has the longest spine on right first exopod (fig. 1.2.4b: Orange; 3) and its right third exopod has a claw-like shape with a slight bump (fig. 1.2.4b: Blue; 1). *B. sp. A*, found west of the Goro Mine, 500 m away from a *B. sibleti* population, has a plump curved process on left second basipodite (fig. 1.2.4c: Red; 2) and a short spine on right first exopod (fig. 1.2.4c: Orange; 3). *B. sp. B*, found south-east of the Goro Mine, has its curved process on left second basipodite with two slight bumps (fig. 1.2.4d: Red; 2). *B. sp. C*, found near Canala on the east coast of Grande Terre, has its curved process on left second basipodite rather flat and detached from the leg as a pouch (fig. 1.2.4e: Red; 2). However, reasonable doubt can be emitted on the distinction between *B. sibleti* and *B. sp. D* due to the genetic closeness between their COI haplotypes. In respect with the principles of integrative taxonomy (Padial *et al.*, 2010), several evidence between genetic, morphologic, geographic, ecologic and reproductive divergence must be demonstrated to posit for a distinct species. Nevertheless, the bPTP analysis determined the two populations as a putative distinct species. Populations of *B. sibleti* and *B. sp. D* have both a slight bump on their right third exopod but shapes are very different (fig. 1.2.4b, f: Blue; 1) and the curved process on left second basipodite is strictly round and more detached from the leg in *B. sp.*

D (fig. 1.2.4f: Red; 2). Finally, relatively to the size of the whole P5, the spine on right first exopod is shorter in *B. sp. D* (fig. 1.2.4f: Orange; 3). In addition with the morphologic and genetic differences, they are present on distinct islands of the archipelago (Grande Terre and Belep respectively) and are distant of 400 km (fig. 1.2.3). These differences are sufficient to determine the populations are different species. The formal description of these four potential species will be provided in subsequent publications.

Micro-endemism

Genetic and specific diversification can have multiple origins that can be *e.g.*, mutation rate, hybridization, reproductive isolation and genetic drift. Divergence and endemism are observed especially for islands that are the most distant from mainland and have been long separated (MacArthur & Wilson, 1963; Whittaker *et al.*, 2023). In this article, high genetic and specific diversification at narrow geographic scale has been found in New Caledonian clade of *Boeckella*. Regarding specific diversification, several species of the genus have been detected in close localities, a phenomenon that can be interpreted as micro-endemism or short-range endemism. Indeed, micro-endemism is a scale-dependant concept. Most of *Boeckella* species have a wide distribution with most distant locations found at more than 100 km when the species is described in over 2 locations (GBIF). In Grande Terre, the most widespread species is *B. sibleti* with its two most distant populations located approximately 40 km apart. Surprisingly, *B. sibleti* and *B. sp. D* that are genetically and morphologically closest (tab. 1.2.2 - 1.2.3; fig. 1.2.4b, f) are located on distinct islands and geographically distant (400 km; fig. 1.2.3a). Even more peculiar, *B. sibleti* and *B. sp. A* are clearly distinct in morphology and genetic (tab. 1.2.2 - 1.2.3, fig. 1.2.4) but have been collected in neighbouring ponds distant of less than 500 m (fig. 1.2.3b). In New Caledonia, micro-endemism has been notably posited in grasshoppers (Nattier *et al.*, 2013), freshwater gastropods (Haase & Bouchet, 1998) and other plant and animal taxa (Caesar *et al.*, 2017; Pillon *et al.*, 2021). Additionally, other taxa such as orthopterans and the genus *Araucaria* may be candidate to a micro-endemism phenomenon in New Caledonia given the high number of recorded endemic species (239 species of orthopterans with 236 strictly endemic to the archipelago; Johanson & Wells, 2019; 20 species of *Araucaria* in the world with 14 endemic to the archipelago; Isnard & Jaffré, 2024).

Genetic diversification and populations structures

We were able to investigate genetic diversification of the COI marker on two species: *B. spinogibba* sampled in 13 ponds, and *B. sibleti* sampled in 4 ponds (tab. 1.2.4). Both species demonstrates very high and geographically structured genetic diversity, and high genetic distances (tab. 1.2.2, 1.2.4; fig. 1.2.5). In comparison, the highest genetic differentiation of the COI marker found in a *Boeckella* species is *B. meteoris* with 5.2% (Supporting data S11). *B. meteoris* is distributed across southern Chile and Argentina in a territory over four times larger than New Caledonia. *B. titicacae*, that is described only from Lake Titicaca and surroundings has a genetic differentiation of 3.3% (Supporting data S11). However, Lake Titicaca measures approximately half the superficiality of New Caledonia, the distribution of *B. titicacae* is likely to be way larger than *B. sibleti* or *B. spinogibba*.

Despite close sampled locations with most distant ponds 2.2 km apart, *B. spinogibba* demonstrates clear geographic structuration in its genetic diversity (fig. 1.2.5a, d). Indeed, two of the most distant ponds have exclusive haplotypes (HYNC4103 and HYNC4104), AMOVA determined most of its genetic variability was explained by variance between ponds and exact tests demonstrated significant population differentiation in haplotype assemblages. *B. sibleti* shows even higher genetic distance, its haplotypes were always found from a unique location (fig. 1.2.5b - c) and AMOVA logically determined almost all of observed genetic differentiation between haplotypes was explained by variance between ponds. These results might be explained by the larger distribution observed in *B. sibleti* populations. Additionally, even if mitochondrial markers are considered unfitted to evaluate isolation by distance (Teske *et al.*, 2018), the haplotypes showed higher genetic distance when found in distant ponds.

Maturana *et al.* (2020) demonstrated population structure at a similar scale on Signy Island in *B. poppei* with the difference that COI haplotypes showed low genetic distance between them. In the article, authors suggest this low genetic distance indicates a recent bottle-neck or (re-)colonisation in the population explained by a complete ice cover of the island (~7000 YBP). Following the same logic and in respect with the New Caledonian climate that is believed to be stable since over 2 Mya (Soltis *et al.*, 2011), it seems likely that New Caledonian populations *B. spinogibba* and *B. sibleti* are older.

Compared to haplotype networks from Maturana *et al.* (2020), haplotype networks computed in our study seem much more complex. So complex that the neighbour-joining algorithm computed a rather high number of hypothetical haplotypes (median vectors) and was still not able to completely resolve relationships between all haplotypes (fig. 1.2.5). For both species, these hypothetical haplotypes tend to be located at the center of the network, suggesting lacks in the dataset. For *B. sibleti*, sampling is holey which can explain the lacks in the haplotype network. *B. spinogibba* has been found in a narrow area that has been extensively sampled, the lacks in the haplotype dataset are more likely to originate from a different unsampled area or from ancient haplotypes that have now disappeared. If valid, this last hypothesis could be linked with the recent opening of the Goro nickel mining facility near *B. spinogibba*'s observed distribution (observable on fig. 1.2.5c - d around latitude 22°18'S and between longitudes 166°52'E - 167°00'E). As demonstrated by Royaux *et al.* (2024), the Goro mine opened in 2010 but prospective works are observable on aerial photographs of the area since 1992 which resulted in the disappearance of the type locality where *B. spinogibba* has been first described. Additionally from perturbations generated by the construction of infrastructures and by extensive soil extraction, the facility processes the laterite soils through a chemical treatment protocol that generates liquid and solid waste. These wastes contaminate freshwaters and impact biodiversity in these environments and in others (Pascal *et al.*, 2008; Boula *et al.*, 2022; Germande *et al.*, 2022).

Conservation

The specific and genetic structuration observed at such low scale indicates low connectivity between ponds. Dispersion of lentic freshwater zooplankton is mostly handled by terrestrial fauna and birds. Indeed, small-sized eggs and individuals get accidentally ingested or stuck on hairs or feathers and are transported in other locations (Incagnone *et al.*, 2015; Fontaneto, 2019; Parry *et al.*, 2024). As discussed by Royaux *et al.* (2024), *Boeckella* individuals have been exclusively found in ultramafic ponds that are oligotrophic and contain low prey density. Consequently, birds visit less ultramafic ponds which slims dispersion opportunities for the genus and other taxa living in these environments (Bryant, 1979; Herteux *et al.*, 2019). Associated with the high perturbations generated by mining activities, low connectivity means populations are less resilient and more sensible to extinction (Hill *et al.*, 2002; Whittaker *et al.*, 2017). This study is focused on a single copepod genus but these concerning observations are likely transposable to organisms with

similar dispersion strategies (*e.g.*, Rotifera, Branchiopoda, Ostracoda). Such results highlight notably the importance of including genetic data in biodiversity assessment to establish relevant connectivity conservation measures (Doerr *et al.*, 2011).

Origin of the Boeckella genus

Regarding the origin of the *Boeckella* genus, several options can be considered. As *Boeckella* is not present in South Africa, in the pervasive hypothesis of its gondwanan origin, the genus would have disappeared from the area after the Gondwana break-up due to environmental changes. Such scenario was validated for marsupials (Beck, 2008) and clusioids (Ruhfel *et al.*, 2016) with the presence of fossil records but no fossil records exist for *Boeckella* to confirm this hypothesis. The datation of the genus by Maturana *et al.* (2021) places the appearance of *Boeckella* 99 Mya with an uncertainty period from 63 Mya to 149 Mya. This covers a period when the Gondwana was whole (approx. 149 – 120 Mya) but is mostly associated with the period of the supercontinent's break-up. However, this datation has been computed on a single marker, which is insufficient to get reliable datation of the genus (Duchêne *et al.*, 2011). Additionally, as this datation is computed on a partial COI marker, it is only valid for the mitochondrial lineage and different results could be obtained with a nuclear marker. In this context, concluding for a Gondwanan origin of the genus seems premature if not misleading.

As discussed by Bayly (1992a), another option would be that the genus arose when South Africa was already separated from South America and Antarctica, starting around 115 Mya. Indeed, after Africa and India drifted northward, South America, Antarctica and Australia remained linked until 85 Mya and relatively close until 50 – 45 Mya (White *et al.*, 2013). This biogeographical scenario is described notably for the Malpighiales plant order (Ruhfel *et al.*, 2016) and is consistent with distribution and molecular datation of other plant and animal taxa (van den Ende *et al.*, 2017). This hypothesis is further supported by the presence of *Boeckella* in Antarctica, which is not the case for the majority of other examples of amphi-pacific distribution that are disjunct in Antarctica as the continent is now mostly ice-covered.

The phylogeny of *Boeckella* suggests that New Caledonian *Boeckella* is an old lineage possibly sister group to all other known *Boeckella* species (fig. 1.2.2). Similarly, the phylogeny of angiosperms places the Amborellales order as an old

lineage that is sister group to all other known flowering plant orders. The Amborellales order has a single living species *Amborella trichopoda* that is recent and endemic to Grande Terre. For the Amborellales order, it has been hypothesised that it was once widespread and went extinct except in Grande Terre where it found refuge after its last emersion (Soltis *et al.*, 2011). Indeed, the New Caledonia archipelago has known a stable climate since the beginning of the Quaternary (Sandel *et al.*, 2011) and other clades such as conifers are believed to have found refuge on the archipelago (Condamine *et al.*, 2017). As it has been hypothesised for the Amborellales order, the *Boeckella* genus may have originated out of Zealandia, went extinct except in New Caledonia where it found refuge before colonising or recolonising South America, Australia, Antarctica and the sub-antarctic islands.

Centropagidae occur in freshwaters but also in marine, brakish, and inland saline waters (Bayly, 1967; Adamowicz *et al.*, 2010). Bayly (1992a) proposed that freshwater centropagids such as *Calamoecia* could originate from a taxa closely related to *Boeckella* as it is the most widespread. Dussart (1986) described *Dussartopages fagesi* from an estuary of Grande Terre; he stated that the species is morphologically closest to *Boeckella* in centropagids and suggested the single species genus *Dussartopages* could be the most primitive lineage of the family. As a final hypothesis, the genus *Boeckella* could originate from Zealandia and, based on observations by Dussart (1986) and Bayly (1992a), be derived from a *Dussartopages* species that colonised inland freshwaters there. Indeed, *Boeckella* species are found exclusively in temperate climates in South America, mostly in temperate with some exceptions for tropical climates in Australia and New Zealand, and exclusively in tropical climates in New Caledonia. This final hypothesis can be supported by the assumption that most lineages have a tropical origin and progressively dispersed towards temperate climates (Brown, 2014).

Perspectives and limitations

A phylogeny including full mitochondrial DNA sequences, and 18S and 28S nuclear markers would help to get further insights and conclude on the evolutionary history of the *Boeckella* genus. However, data is still lacking, notably for Australian species. Additionally, *D. fagesi* has never been indicated since its original description. Its inclusion in future phylogenetic analyses of Centropagidae could have paramount impact on the comprehension of the taxa.

The impressive genetic and specific diversity observed within the *Boeckella* genus in New Caledonia suggests that more sampling could lead to the discovery of new species. This would enable a more comprehensive evaluation of the distribution and improve the understanding of *Boeckella* on the archipelago and elsewhere. Additionally, the formal description of species from the present study and potential future discoveries will be necessary for their inclusion in environmental assessments and conservation policies.

As discussed in Royaux *et al.* (2024), concatenation of genes from several DNA extracts is biased as there is no certainty that both haplotypes existed in a single individual and chimeric individual might be created. In our phylogeny, it is the case for the concatenation representing *B. sp. B*. However, a single COI haplotype was found for this species and COI fragments obtained with NGS also matched this haplotype. Some concatenations by Maturana *et al.* (2021) might also be chimeric individuals living far from each other as we were not able to verify the precise origin of each sequence.

Conclusion

The *Boeckella* genus seems to check all the boxes of emblematic groups from New Caledonia. As the clade Amborellales with all other flowering plants, New Caledonian clade of *Boeckella* seems to form a monophyletic group that is sister to all other representatives of the genus. As the freshwater fish *Galaxias neocaledonicus* (Keith, 2002), and the giant gecko *Rhacodactylus auriculatus* (Hudel *et al.*, 2020), *Boeckella* species are endemic to ultramafic massifs of the archipelago. Finally, as for many animal and plant taxa on the archipelago, *Boeckella* species seems to present short-range endemism (Caesar *et al.*, 2017).

The study of islands (*i.e.* ponds, sinkholes) within islands (*i.e.* Grande Terre, Belep) could represent a powerful tool for understanding insular systems. This study presents a textbook case of insular systems, characterised by high rates of endemism and genetic divergence, with clear geographic structuration.

Acknowledgements

Funding for ‘La Planète Revisitée’ expeditions were provided by the Gouvernement de la Nouvelle-Calédonie, Province Sud, Province Nord, Office des Postes et Télécommunications (OPT), Maison de la Nouvelle-Calédonie, and the Ministère Français des Outre-mer. Funding for ‘Vies d’Ô douce’ expeditions were provided by call for projects ATM from the Muséum national d’Histoire naturelle, by the AAP SUD and the AAP transverse axe MIDI from UMR BOREA.

Funding for molecular biology manipulations and SEM images were granted to NR by Muséum national d’Histoire naturelle through the Action Transversale du Muséum (2022) call for projects.

The New Caledonia Hydrobiological expeditions 2016–2018 were conducted under the ‘La Planète Revisitée’ program implemented by the Muséum national d’Histoire naturelle (MNHN), Paris in partnership with the Conservatoire d’Espaces naturels (CEN). We deeply thank Pascale Joannot, head of the expeditions program at MNHN, and Philippe Bouchet, the expedition leader. These expeditions were operated under permits issued by New Caledonia’s Province Sud (under APA_NCPS_2017_028) and Province Nord, and the organizers thank Emmanuel Coutures and Isabelle Jurquet (Province Sud) and Jean-Jérôme Cassan and Yannick Monlouis (Province Nord) for their help in obtaining these permits. We thank Sébastien Faninoz and Alice Leblond.

Authors thank the “Station de Biologie Marine de Concarneau” and the “Service de Systématique Moléculaire” of the MNHN (UAR 2AD 2700) for laboratory access and assistance, notably for the amplicon pool sequencing service. Authors thank the “Plateau Technique de Microscopie et de Microanalyses” of the MNHN for access to the SEM.

Authors want to deeply thank Myriam Georges for her unrestricted support on the early design and application of molecular manipulations. Many thanks to Geraldine Toutirais and Sylvain Pont for their help in the SEM imagery process.

Finally, authors want to dedicate this work to Loyola Até “Fanfan”.

Data availability statement

AB1 files from sequencing and datasets generated during and/or analysed during the current study are available in the Data.InDoRes repository [LINK IN PROGRESS]. R scripts are available on GitHub (https://github.com/ColineRoyaux/population_genetics).

IV. Quelques cas parmi les Branchiopodes

Les analyses similaires effectuées sur les cinq genres de Branchiopodes *Streptocephalus*, *Eulimnadia*, *Triops*, *Latonopsis* et *Lynceus* sont présentés dans les parties suivantes.

IV. 1. Le genre *Streptocephalus*

L'étude de la morphologie des individus semble indiquer que l'unique espèce de *Streptocephalus* retrouvée sur l'archipel est *S. archeri* (fig. 1.1).



Figure 1.1 – Image d'un *S. archeri* de NCK. Echelle approximative. © J.-F. Cart

Cette espèce est présente dans 17 localités de l'archipel (8 échantillonnées par filet, 8 en élevage et 1 par les deux) et est également la seule retrouvée en Australie (Timms, 2015). *S. archeri* n'a pas été observée sur les massifs ultramafiques (fig. 1.2).

Au total, 28 séquences de COI partiel ont été obtenues par PCR et séquençage Sanger pour le genre *Streptocephalus* en NCK. Une séquence du même gène a été obtenue par NGS. Les historiques Galaxy de nettoyage et préparation des données ADN pour les analyses de structure de population sont accessibles sur la plateforme (https://usegalaxy.fr/u/coline_royaux/h/sanger1-plate-2024-strep ; https://usegalaxy.eu/u/coline_royaux/h/strep-pop-gen-data-prep). La séquence consensus obtenue par NGS a été établie avec 38 séquences lues.

Quatre haplotypes différents ont été retrouvés dans 7 localités (2 par filet et 5 en élevage) avec une distance génétique non corrigée allant jusqu'à 1,22 % (tab. 1.3).

Tableau 1.3 – Matrice des distances non corrigées des fragment de COI séquencés chez *S. archeri*.

Haplotypes	S_arc_COI01	S_arc_COI02	S_arc_COI03
S_arc_COI02	0,15 %		
S_arc_COI03	0,46 %	0,61 %	
S_arc_COI04	0,76 %	0,91 %	1,22 %

L'haplotype le plus répandu, S_arc_COI01, est présent dans 6 localités. Les autres haplotypes sont chacun retrouvés dans une seule localité. Les localités HYNC2703 (filet) et HYNC869 (élevage) sont les plus riches avec 2 haplotypes retrouvés dans chacune (tab. 1.4).

Tableau 1.4 - Distribution des haplotypes de *S. archeri* amplifiés. S : date de collecte du sédiment ; E : date de collecte dans l'élevage.

Identifiant de la localité	Date(s) de collecte	Total	S_arc_COI01	S_arc_COI02	S_arc_COI03	S_arc_COI04	Richesse en haplotypes
HYNC2703	07-12-2017	10	0	8	2	0	2
HYNC4055	12-06-2018	1	1	0	0	0	1
HYNC816	S : 06-11-2016 E : 31-05-2017	1	1	0	0	0	1
HYNC817	S : 06-11-2016 E : 13-08-2017	7	7	0	0	0	1
HYNC818	S : 06-11-2016 E : 16-08-2017	2	2	0	0	0	1
HYNC858	S : 09-11-2016 E : 16-08-2017	1	1	0	0	0	1
HYNC869	S : 09-11-2016 E : 22-06-2017	3	2	0	0	1	2
Total		25	14	8	2	1	

D'après l'AMOVA, les populations de *S. archeri* ont leur variance moléculaire expliquée à 21,9 % par de la variance entre localités et à 78,1 % par de la variance au sein des localités. Le test exact de différenciation des populations a déterminé une différenciation significative entre les populations ($p = 0.0005$). Les réseaux d'haplotypes calculés avec les algorithmes de median-joining et de parcimonie

statistique sont identiques (fig. 1.2). Les liens entre les différents haplotypes séquencés chez *S. archeri* semblent bien résolus, aucun haplotype hypothétique ni liens alternatifs n'ont été détectés.

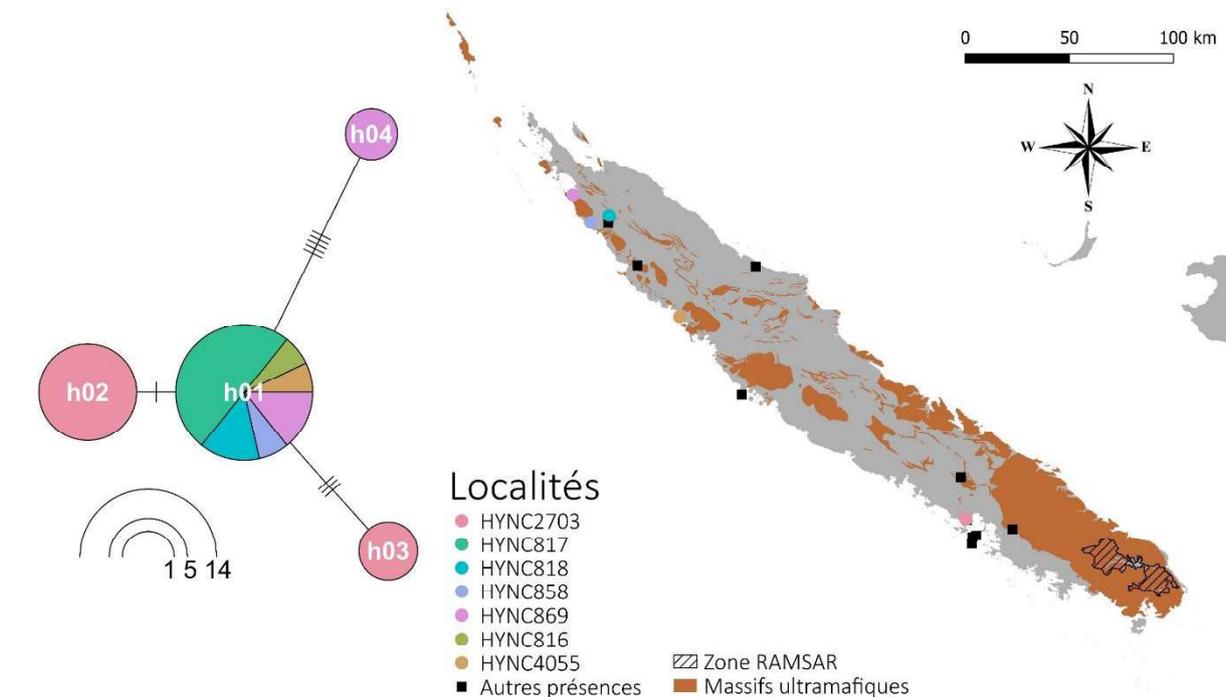


Figure 1.2 – A gauche, réseau d'haplotypes de *S. archeri* calculé en median-joining et parcimonie statistique ; les noms d'haplotypes sont abrégés de S_arc_COIXX en hXX pour alléger la figure. A droite, carte de répartition de l'espèce et de ses haplotypes en NCK.

IV. 2. Le genre *Eulimnadia*

Au total, 34 séquences de COI partiel ont été obtenues par PCR et séquençage Sanger ou pool d'amplicon pour le genre *Eulimnadia* en NCK. L'historique Galaxy de nettoyage des données ADN est accessible sur la plateforme (https://usegalaxy.fr/u/coline_royaux/h/sanger1-plate-2024-eulim). Chez les populations étudiées, 7 haplotypes différents ont été retrouvés dans 11 localités.

La phylogénie est calculée sur 45 séquences individuelles représentant 19 espèces d'*Eulimnadia* de Nouvelle-Calédonie-Kanaky, d'Australie, d'Afrique et du Japon. *E. cylindrova* est utilisée en groupe externe (Bellec & Rabet, 2016 ; annexe tab. CS1). Pour la phylogénie bayésienne les trois modèles utilisés pour chaque position de codon sont F81+I, TIM+G et TRN+G.

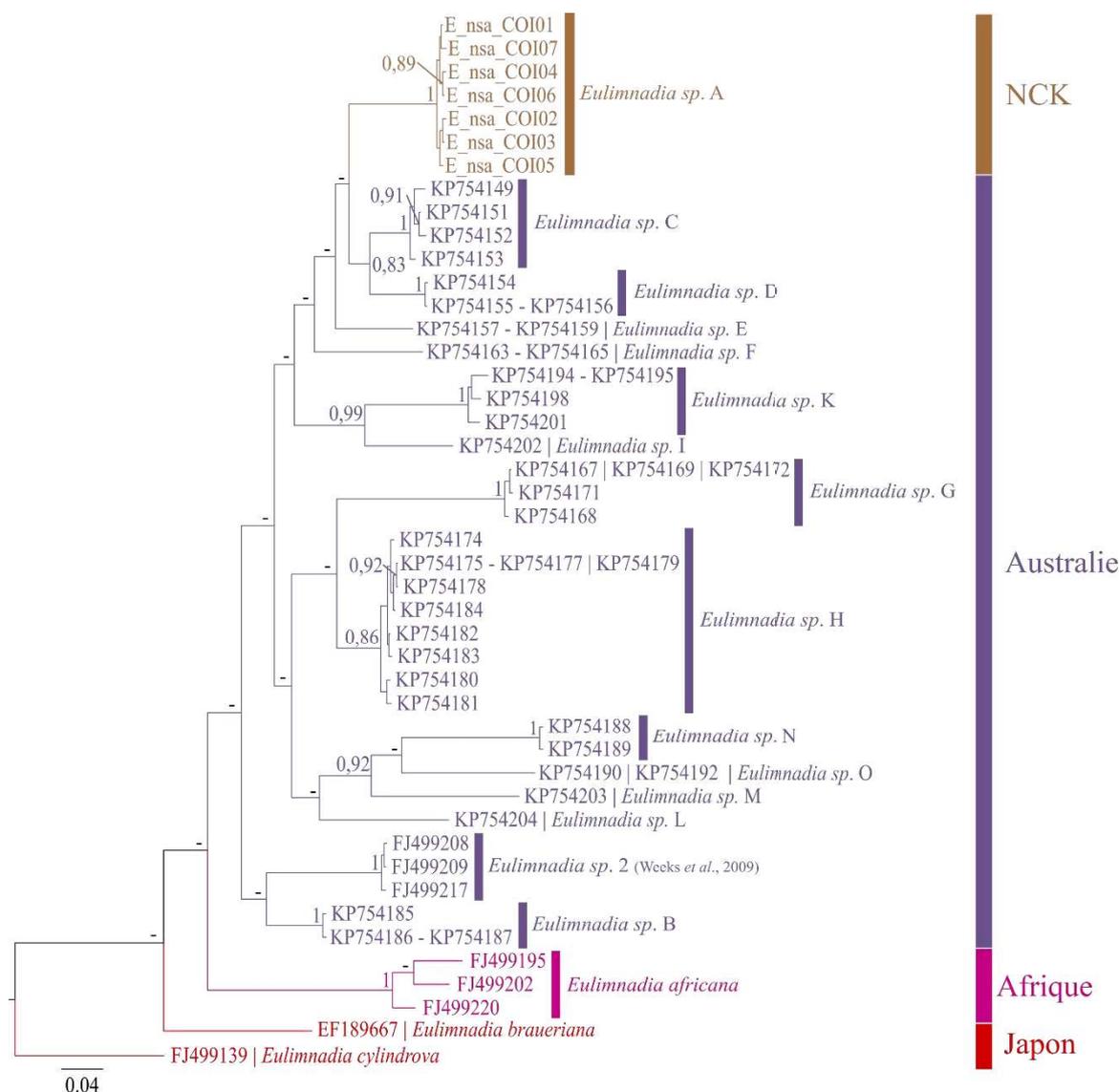


Figure 1.3 – Phylogénie COI du genre *Eulimnadia* (méthode bayésienne, probabilités postérieures indiquées au niveau des noeuds). Les lettres associées aux espèces sont reprises de l'article Schwentner *et al.* (2015).

La phylogénie d'*Eulimnadia* basée sur le gène COI est faiblement supportée globalement (fig. 1.3). Cependant, les haplotypes de NCK forment un groupe bien supporté. D'après la forme des œufs, l'espèce d'*Eulimnadia* retrouvée dans les échantillons étudiés (fig. 1.4) est proche d'*E. australiensis*. Or, une analyse de séquences des gènes nucléaires ITS1 et ITS2 (« Internal Transcribed Spacer » ou Espaceur interne transcrit) d'individus proche de la localité type australienne de cette espèce ont montré une distance génétique de plus de 12% avec les séquences de NCK (Rabet, non publié). L'unique espèce d'*Eulimnadia* présente en NCK semble être une espèce non décrite.



Figure 1.4 – Image d'un *E. sp. A* de NCK. Echelle approximative. © J.-F. Cart

Eulimnadia sp. A est présente dans 28 localités de l'archipel (6 échantillonnées par filet et 22 en élevage). Elle est présente exclusivement hors des massifs ultramafiques et n'a pas (encore) été signalée en Australie (fig. 1.5). La distance génétique non corrigée entre les haplotypes de cette espèce atteint jusqu'à 0,76 % (tab. 1.5).

Tableau 1.5 – Matrice des distances non corrigées des fragments de COI séquencés chez *E. sp.*

Haplotypes	E_nsa_COI01	E_nsa_COI02	E_nsa_COI03	E_nsa_COI04	E_nsa_COI05	E_nsa_COI06
E_nsa_COI02	0,46 %					
E_nsa_COI03	0,30 %	0,15 %				
E_nsa_COI04	0,30 %	0,76 %	0,61 %			
E_nsa_COI05	0,30 %	0,46 %	0,30 %	0,61 %		
E_nsa_COI06	0,15 %	0,61 %	0,46 %	0,15 %	0,46 %	
E_nsa_COI07	0,30 %	0,76 %	0,61 %	0,61 %	0,61 %	0,46 %

L'haplotype le plus répandu, E_nsa_COI01, est présent dans 6 localités. La localité HYNC2807 est la plus riche avec 3 haplotypes retrouvés (tab. 1.6).

Tableau 1.6 - Distribution des haplotypes de *E. sp. A* amplifiés. S : date de collecte du sédiment ; E : date de collecte dans l'élevage.

Identifiant de la localité	Date(s) de collecte	Total	E_nsa_COI01	E_nsa_COI02	E_nsa_COI03	E_nsa_COI04	E_nsa_COI05	E_nsa_COI06	E_nsa_COI07	Richesse en haplotypes
ETNC004	14-01-2021	1	0	0	1	0	0	0	0	1
HYNC2654	06-07-2018	5	0	0	0	1	0	4	0	2
HYNC2690	S : 05-12-2017 E : 08-08-2019	1	0	0	1	0	0	0	0	1
HYNC2807	S : 23-11-2017 E : 15-07-2018	5	1	3	0	0	0	0	1	3
HYNC2874	S : 26-11-2017 E : 15-07-2018	1	0	0	0	0	1	0	0	1
HYNC4084	14-06-2018	4	4	0	0	0	0	0	0	1
HYNC816	S : 06-11-2016 E : 31-05-2017	1	1	0	0	0	0	0	0	1
HYNC828	S : 06-11-2016 E : 30-07-2017	5	5	0	0	0	0	0	0	1
HYNC907	S : 16-11-2016 E : 26-10-2017	1	1	0	0	0	0	0	0	1
HYNC909	17-11-2016	4	4	0	0	0	0	0	0	1
HYNC932	29-11-2016	6	0	0	6	0	0	0	0	1
Total		34	16	3	8	1	1	4	1	

L'historique Galaxy de préparation des données ADN pour les analyses de structure de population est accessible sur la plateforme (https://usegalaxy.eu/u/coline_royaux/h/eulim-population-genetics-data-prep).

D'après l'AMOVA, les populations de *Eulimnadia sp. A* ont leur variance moléculaire expliquée à 69,3 % par de la variance entre localités et à 30,7 % par de la variance au sein des localités. Le test exact de différenciation des populations n'a pas pu être calculé pour cette espèce. L'algorithme median-joining a permis de résoudre correctement les liens entre haplotype en calculant un haplotype hypothétique entre les haplotypes E_nsa_COI01, E_nsa_COI03 et E_nsa_COI05 (fig. 1.5 ; $\epsilon = 2$). Les haplotypes les plus distants génétiquement, E_nsa_COI02 et E_nsa_COI07, sont présents dans la même mare (HYNC2807 ; fig. 1.5).

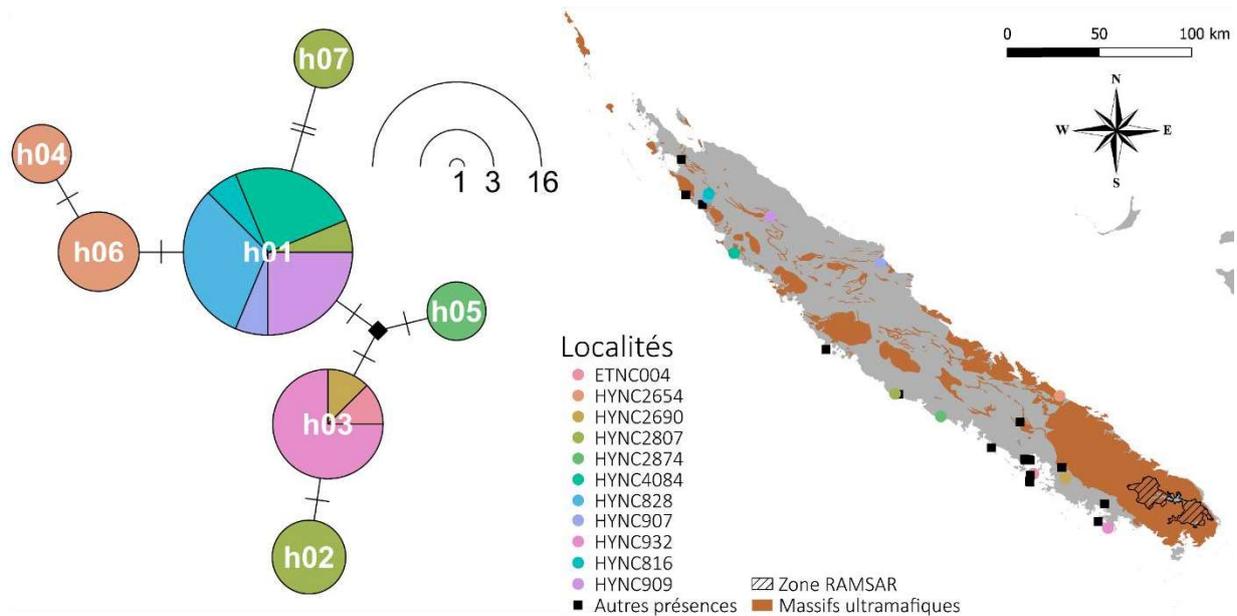


Figure 1.5 – A gauche, réseau d’haplotypes de *Eulimnadia* sp. A calculé en median-joining; l’haplotype hypothétique est représenté par un losange noir ; les noms d’haplotypes sont abrégés de E_nsa_COIXX en hXX pour alléger la figure. A droite, carte de répartition de l’espèce et de ses haplotypes en NCK.

IV. 3. Le genre *Triops*

Au total, 5 séquences de COI partiel ont été obtenues par PCR et séquençage Sanger pour le genre *Triops* en NCK. Deux séquences du même gène ont été obtenues par NGS. L’historique Galaxy de nettoyage des données ADN est accessible sur la plateforme (https://usegalaxy.fr/u/coline_royaux/h/sanger1-plate-2024-triops). Pour cette espèce, 2 haplotypes différents ont été retrouvés dans 3 localités.

La phylogénie est calculée sur 141 séquences individuelles reproduisant la phylogénie publiée par Meusel & Schwentner (2017). *Lepidurus* est le taxon le plus proche de *Triops* (Korn *et al.*, 2013), *Lepidurus apus viridis* (HF911398.1) est utilisée en groupe externe (annexe tab. CS2). Pour la phylogénie bayésienne les trois modèles utilisés pour chaque position de codon sont SYM+G, F81 et GTR+I+G.

La phylogénie du genre *Triops* basée sur le gène COI place les haplotypes de NCK en groupe frère bien supporté de *Triops* sp. A (nommé par Meusel & Schwentner, 2017 ; tab. 1.7). L’unique espèce de *Triops* trouvée en NCK (fig. 1.6) a été décrite sur l’archipel par Longhurst (1955) sous le nom *Triops longicaudatus intermedius*. C’est l’unique signalement de l’espèce sur l’archipel et la localité n’a pas été précisée. Or, la phylogénie distingue clairement les *T. longicaudatus* des *Triops* de NCK. Les haplotypes de NCK et d’Australie (*Triops* sp. A) semblent former un ensemble bien supporté que l’on nomme ici *T. intermedius* (fig. 1.7).

Tableau 1.7 – Distances génétiques du fragment de gène COI entre les espèces de *Triops* incluses dans la phylogénie (noms d'après Meusel & Schwentner, 2017 ; fig. 1.7) et les haplotypes obtenus de NCK.

Espèce	Pays	Distance génétique avec haplotypes de NCK
<i>Triops intermedius</i> (A)	Australie	4.3-6.1 %
<i>Triops longicaudatus</i>	Monde	11.2-11.9 %
<i>Triops newberryi</i>	USA	11.8-12.4 %
<i>Triops sp.</i> (Korn <i>et al.</i> , 2013)	Bolivie	10.9-11 %
<i>Triops sp.</i> (Naganawa, 2018)	Australie	10.1-10.2 %
<i>Triops sp.</i> P	Australie	10.8-10.9 %
<i>Triops sp.</i> H	Australie	10.9-12.2 %
<i>Triops sp.</i> M	Australie	11.4-11.6 %
<i>Triops sp.</i> F	Australie	8.8-10.3 %
<i>Triops sp.</i> I	Australie	10.3-10.7 %
<i>Triops sp.</i> N	Australie	11.1 %
<i>Triops sp.</i> O	Australie	10.4-11.6 %
<i>Triops sp.</i> C	Australie	10.3-12.1 %
<i>Triops sp.</i> T	Australie	7.2-9.3 %
<i>Triops sp.</i> S	Australie	9-9.7 %
<i>Triops sp.</i> Q	Australie	8.5 %
<i>Triops sp.</i> E	Australie	8.6-11 %
<i>Triops sp.</i> L	Australie	7.6-9.2 %
<i>Triops australiensis</i> (D)	Australie	8.6-12.1 %
<i>Triops sp.</i> K	Australie	6.9-8 %
<i>Triops strenuus</i> (B)	Australie	7.6-9.5 %
<i>Triops gracilis</i> (R)	Australie	11.3-11.4 %
<i>Triops sp.</i> G	Australie	11.3-12.41 %
<i>Triops sp.</i> J	Australie	12.3-12.7 %
<i>Lepidurus apus viridis</i>	Australie	18.9 %



Figure 1.6 - Image d'un *T. intermedius* de NCK. Echelle approximative. © N. Charpin

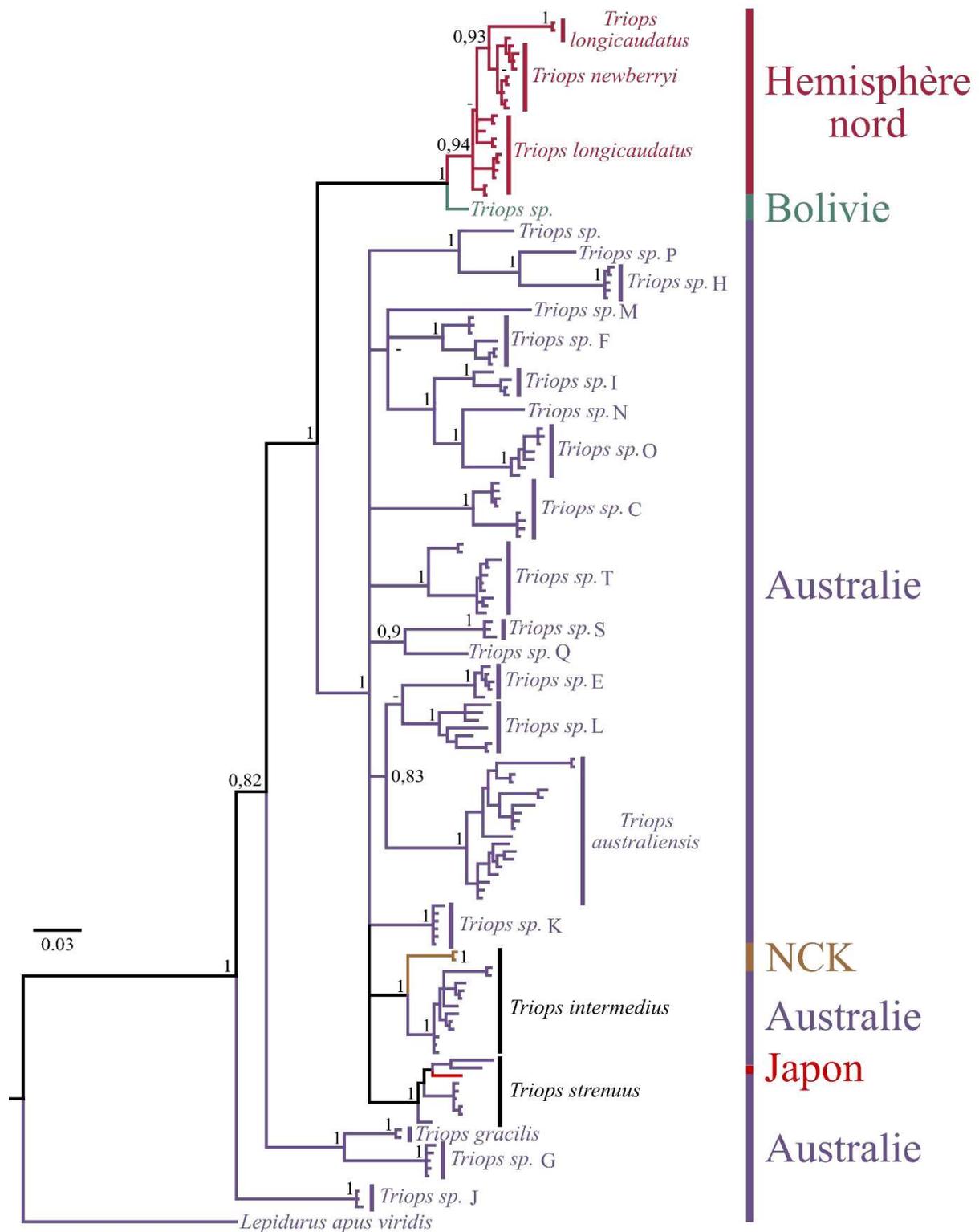


Figure 1.7 – Phylogénie COI du genre *Triops* et du groupe externe *Lepidurus apus viridis* (noeuds: probabilité selon la phylogénie bayésienne). Les lettres associées aux espèces sont reprises de l'article Meusel & Schwentner (2017).

T. intermedius est présente dans 4 localités de l'archipel (3 échantillonnées par filet et 1 en élevage). Elle est présente exclusivement hors des massifs ultramafiques dans les zones les plus sèches de NCK et semble présente en

Australie étant donné sa proximité génétique avec le clade A (fig. 1.7). Une distance génétique non corrigée de 0,15 % (une seule mutation) a été retrouvée entre les deux haplotypes. L'haplotype le plus répandu, T_int_COI01, est présent dans 2 localités (tab. 1.8).

Tableau 1.8 - Distribution des haplotypes de *T. intermedius* amplifiés. S : date de collecte du sédiment ; E : date de collecte dans l'élevage.

Identifiant de la localité	Date(s) de collecte	Total	T_int_COI01	T_int_COI02	Richesse en haplotypes
ETNC001	14-01-2021	5	5	0	1
HYNC2703	07-12-2017	1	0	1	1
HYNC4035	S : 10-06-2018 E : 03-08-2018	1	1	0	1
Total		7	6	1	

Afin de réaliser des réseaux d'haplotypes, les séquences obtenues dans cette thèse ont été complétées par les séquences de *Triops* les plus proches des haplotypes néo-calédoniens selon la phylogénie (fig. 1.7) représentée par les haplotypes h50 et h52-h61 publiés par Meusel et Schwentner (2017) (tab. 1.7). Ces séquences sont obtenues à travers la plateforme GenBank (Accessions NCBI en annexe tab. CS2). L'historique Galaxy de préparation des données ADN pour les analyses de structure de population est accessible sur la plateforme (https://usegalaxy.eu/u/coline_royaux/h/triops-pop-gen-data-prep).

L'AMOVA et le test exact de différenciation des populations n'ont pas pu être calculés pour cette espèce dû au faible nombre d'haplotypes. Les distances génétiques observées entre les haplotypes australiens et néo-calédoniens ne sont valides dans aucun des réseaux d'haplotypes calculés. Par exemple, 26 mutations sont observées entre les séquences de T_int_COI02 et h55 alors que 18 et 19 mutations sont représentées sur les réseaux d'haplotypes en parcimonie statistique (fig. 1.8) et en median-joining, respectivement. Les haplotypes de NCK ne semblent pas présenter de structuration dans leur distribution géographique. En effet, les individus des deux localités les plus éloignées géographiquement, ETNC001 et HYNC4035, présentent le même haplotype, T_int_COI01 (fig. 1.8).

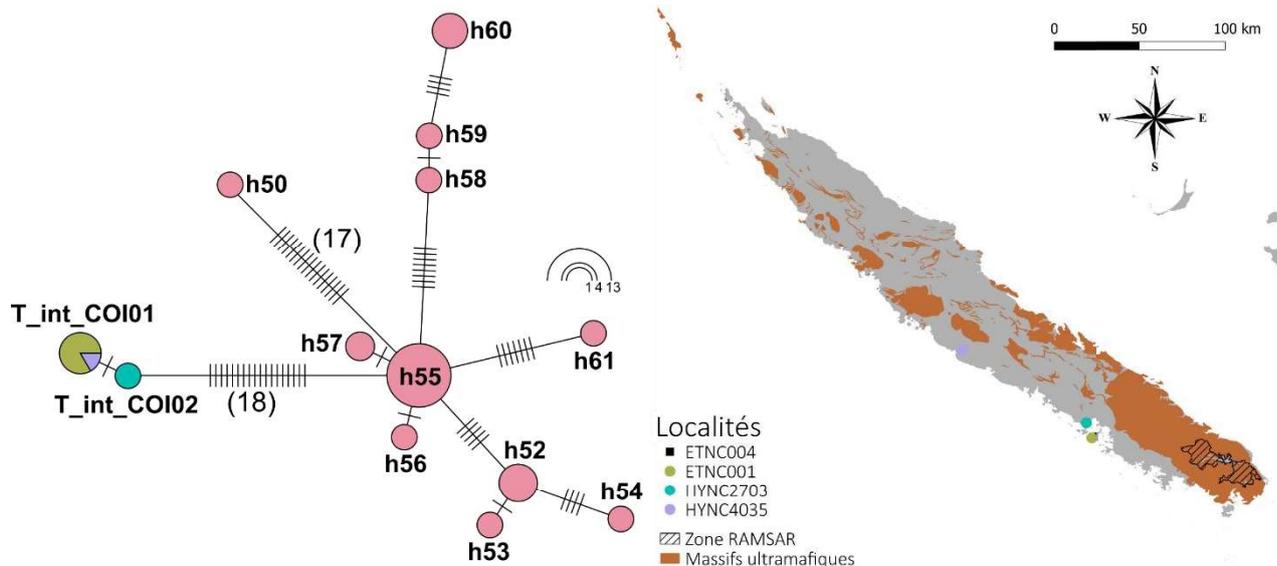


Figure 1.8 – A gauche, réseau d’haplotypes de *Triops intermedius* calculé en parcimonie statistique, les haplotypes australiens sont en rose. A droite, carte de répartition de l’espèce et de ses haplotypes en NCK.

IV. 4. Le genre *Latonopsis*

Latonopsis brehmi dedeckkeri a été décrite comme une sous-espèce sur l’archipel par Timms (1985). La description faite en 1985 est très succincte, le statut taxonomique peu clair (Korovchinsky, 2001) et aucune donnée de génétique n’est disponible sur cette sous-espèce ni sur l’espèce *L. brehmi* d’Australie. Angel Arnold, durant son stage de master 2, a étudié des individus australiens de *L. brehmi* et les a comparés aux individus de NCK. Des différences morphologiques plus prononcées et plus nombreuses ont été identifiées entre les deux groupes d’individus suggérant que les populations néo-calédoniennes formeraient une espèce différente endémique à la NCK (Arnold *et al.*, *in prep*). On considère donc que deux espèces de *Latonopsis* sont retrouvées sur l’archipel : *L. dedeckkeri* et *L. australis* (fig. 1.9).

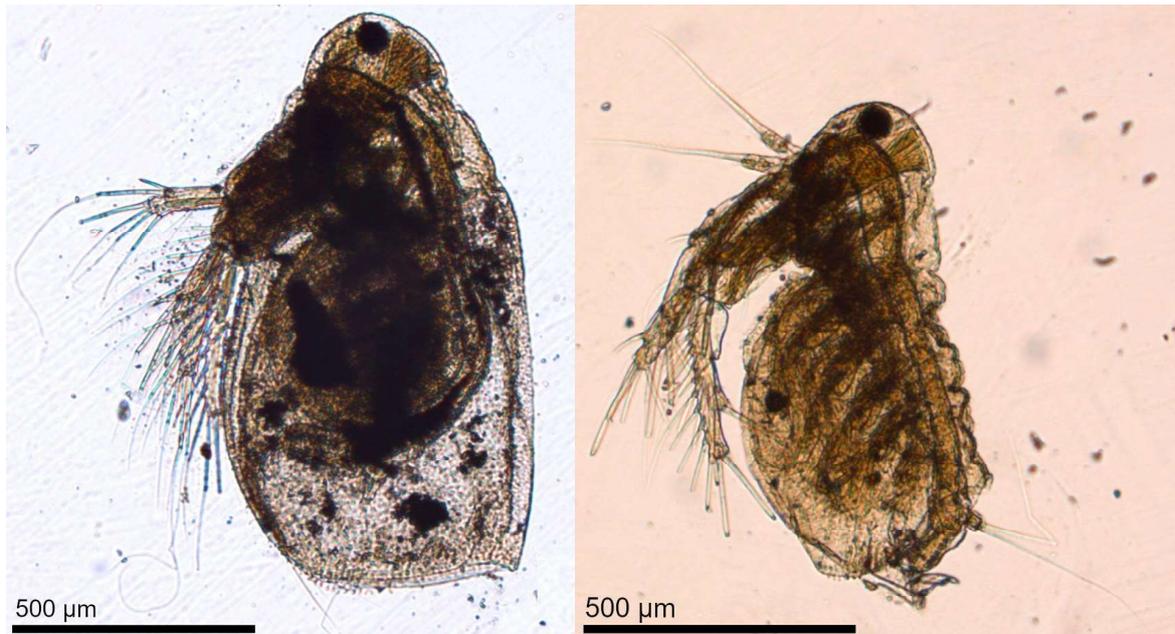


Figure 1.9 – A gauche, image d'un *L. australis* de NCK. A droite, image d'un *L. dedeckkeri*. Prise de vue : Angel Arnold

Au total, 7 séquences de COI partiel ont été obtenues par PCR et séquençage Sanger ou pool d'amplicon pour le genre *Latonopsis* en NCK. Trois séquences du même gène ont été obtenues par NGS. Les historiques Galaxy de nettoyage et préparation des données ADN pour les analyses de structure de population sont accessibles sur la plateforme (https://usegalaxy.fr/u/coline_royaux/h/sanger1-plate-2024--pool39laton ; https://usegalaxy.eu/u/coline_royaux/h/lato-pop-genetics-data-prep).

L. australis est présente dans 7 localités de l'archipel (4 échantillonnées par filet et 3 en élevage) et est également retrouvée en Australie. Cette espèce n'a pas été observée sur les massifs ultramafiques (fig. 1.10). Aucune séquence n'a pu être obtenue pour cette espèce.

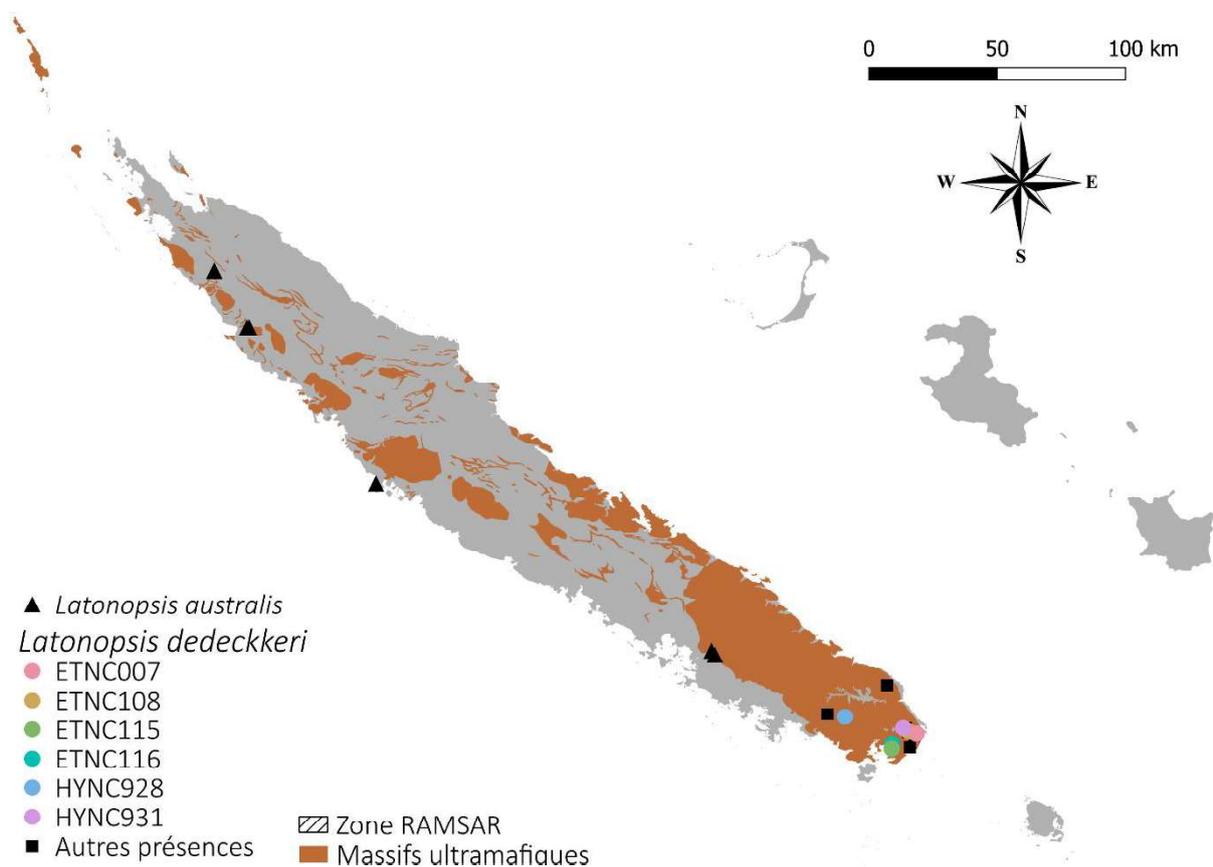


Figure 1.10 – Carte de présence des deux espèces du genre *Latonopsis* trouvées dans les échantillons. Les couleurs des points de présence de *L. dedeckkeri* correspondent aux couleurs du réseau d'haplotypes (fig. XX).

L. dedeckkeri est présente dans 16 localités de l'archipel (toutes échantillonnées par filet). Elle est présente exclusivement sur les massifs ultramafique et n'est pas présente en Australie (fig. 1.10). Pour cette espèce, 7 haplotypes différents ont été retrouvés dans 6 localités avec une distance génétique non corrigée allant jusqu'à 7,9 % (tab. 1.9).

Tableau 1.9 – Matrice des distances non corrigées des fragments de COI séquencés chez *L. dedeckkeri*.

Haplotypes	L_brd_COI01	L_brd_COI02	L_brd_COI03	L_brd_COI04	L_brd_COI05	L_brd_COI06
L_brd_COI02	0,15 %					
L_brd_COI03	1,52 %	1,67 %				
L_brd_COI04	2,58 %	2,43 %	1,98 %			
L_brd_COI05	0,61 %	0,46 %	1,82 %	2,28 %		
L_brd_COI06	7,60 %	7,45 %	7,14 %	7,75 %	7,60 %	
L_brd_COI07	7,75 %	7,60 %	7,29 %	7,90 %	7,75 %	0,15 %

L'haplotype le plus répandu, L_brd_COI01, est présent dans 2 localités. Les autres haplotypes sont chacun retrouvés dans une seule localité. Les localités

ETNC108 et HYNC928 sont les plus riches avec 2 haplotypes retrouvés dans chacune (tab. 1.10).

Tableau 1.10 - Distribution des haplotypes de *L. dedeckkeri* amplifiés.

Identifiant de la localité	Date(s) de collecte	Total	L_brd_COI01	L_brd_COI02	L_brd_COI03	L_brd_COI04	L_brd_COI05	L_brd_COI06	L_brd_COI07	Richesse en haplotypes
ETNC007	08-06-2020	2	2	0	0	0	0	0	0	1
ETNC108	09-06-2020	2	1	1	0	0	0	0	0	2
ETNC115	22-06-2020	1	0	0	0	1	0	0	0	1
ETNC116	22-06-2020	2	0	0	2	0	0	0	0	1
HYNC928	05-06-2018	2	0	0	0	0	0	1	1	2
HYNC931	16-06-2018	1	0	0	0	0	1	0	0	1
Total		10	3	1	2	1	1	1	1	

D'après l'AMOVA, les populations de *L. dedeckkeri* ont leur variance moléculaire expliquée à 98 % par de la variance entre localités et à 2 % par de la variance au sein des localités ($p = 0.002$). Le test exact de différenciation des populations n'a pas pu être calculé pour cette espèce. Aucun des deux algorithmes n'est parvenu à résoudre tous les liens entre les haplotypes. Par exemple, 16 mutations sont observées entre les séquences des haplotypes L_brd_COI01 et L_brd_COI04 tandis que 19 et 23 mutations sont représentées entre les deux haplotypes dans les réseaux calculés par median-joining (fig. 1.11) et parcimonie statistique. Le réseau calculé par median-joining (fig. 1.11 ; $\varepsilon = 2$) est le plus proche de ce qui est observé en termes de mutations. L'algorithme median-joining a calculé 3 haplotypes hypothétiques (median vectors). Les haplotypes L_brd_COI06 et L_brd_COI07 sont les plus distants génétiquement des autres haplotypes (fig. 1.11) et sont retrouvés dans la mare la plus distante géographiquement des autres, HYNC928 (entouré en trait continu sur fig. 1.11). Les haplotypes L_brd_COI03 et L_brd_COI04 (entouré en pointillés sur fig. 1.11) semblent se distinguer légèrement des haplotypes L_brd_COI01, L_brd_COI02 et L_brd_COI05 (entouré en double trait sur fig. 1.11). Ces deux groupes d'haplotypes sont également retrouvés dans des mares séparées géographiquement (entouré par les mêmes motifs sur fig. 1.11).

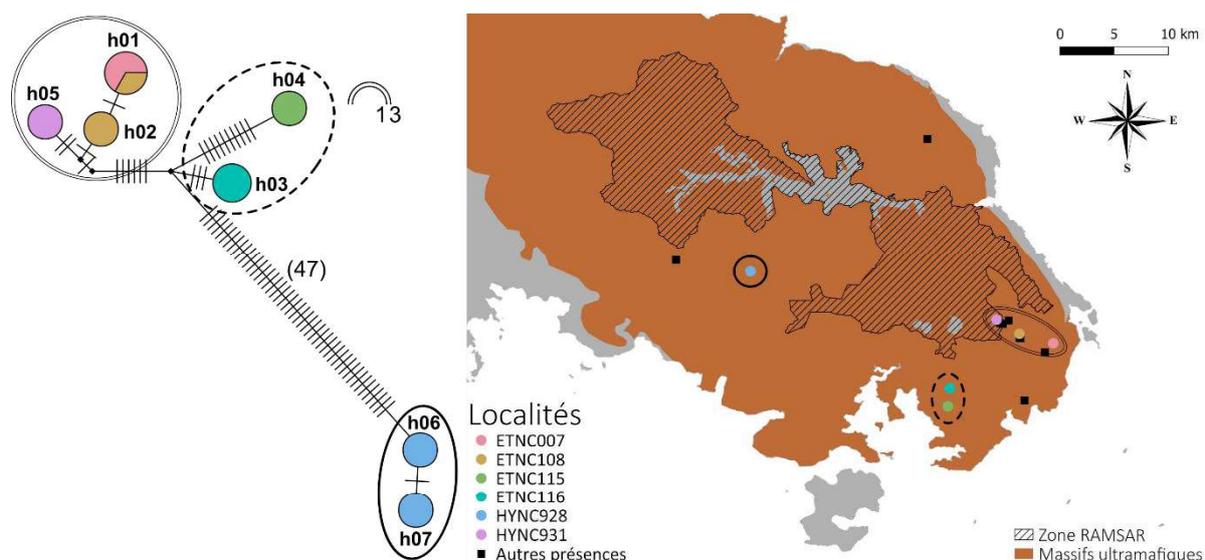


Figure 1.11 – A gauche, réseau d’haplotypes de *L. dedeckkeri* calculé en median-joining ; les haplotypes hypothétiques sont représentés par des losanges noirs ; les noms d’haplotypes sont abrégés de *L_brd_COIXX* en hXX pour alléger la figure. A droite, carte de répartition de l’espèce et de ses haplotypes en NCK.

IV. 5. Le genre *Lynceus*

Au total, 73 séquences de COI partiel ont été obtenues par PCR et séquençage Sanger pour le genre *Lynceus* en NCK. Sept séquences de COI partiel et des marqueurs des sous-unités ribosomiques nucléaire 18S et mitochondrial 16S ont été obtenues par NGS à partir d’individus uniques. Parmi ces sept séquences, une a été obtenue à partir d’un individu prélevé en Australie par B. Timms. L’historique Galaxy de nettoyage des données ADN est accessible sur la plateforme (https://usegalaxy.fr/u/coline_royaux/h/sanger1-plate-2024-lynceus).

Les concaténations des gènes 18S, COI et 16S effectuées avec les données acquises dans cette thèse ont été complétées par des concaténations publiées par Sigvardt *et al.* (2021) (annexe tab. CS3). La phylogénie est calculée sur 20 concaténations de séquences individuelles représentant 17 espèces de *Lynceus* de Nouvelle-Calédonie-Kanaky, d’Australie, d’Asie de l’Est, d’Afrique du Sud, des États-Unis d’Amérique et du Chili. *Triops* et *Cyclestheria* sont utilisés en groupes externes (Sigvardt *et al.*, 2021 ; annexe tab. CS3). Pour la phylogénie bayésienne les trois modèles utilisés pour chaque position de codon sont GTR+G, TVM+I+G et TVM+G pour le COI, GTR+I+G pour le 16S et TRN+I pour le 18S.

La phylogénie du genre *Lynceus* basée sur les gènes 18S, COI et 16S place les haplotypes d’Australie, de Nouvelle-Zélande et de NCK en un groupe monophylétique bien supporté (fig. 1.12). Morphologiquement et phylogénétiquement, les *Lynceus* de NCK semblent former trois groupes distincts

est présente exclusivement sur les massifs ultramafiques et n'est pas présente en Australie (fig. 1.13). Pour cette espèce, 4 haplotypes différents ont été retrouvés dans 6 localités avec une distance génétique non corrigée allant jusqu'à 4,4 % (tab. 1.11). L'haplotype le plus répandu, L_ins_COI02, est présent dans 5 localités. Les autres haplotypes sont chacun retrouvés dans une seule localité. Les localités ETNC007 et HYNC931 sont les plus riches avec 2 haplotypes retrouvés dans chacune (tab. 1.12).

Tableau 1.12 - Distribution des haplotypes de *L. insularis* amplifiés.

Identifiant de la localité	Date(s) de collecte	Total	L_ins_COI01	L_ins_COI02	L_ins_COI03	L_ins_COI04	Richesse en haplotypes
ETNC007	08-06-2020	4	0	1	3	0	2
HYNC4109	16-06-2018	2	0	2	0	0	1
HYNC4111	16-06-2018	1	0	1	0	0	1
ETNC501	23-03-2023	1	0	0	0	1	1
HYNC920	05-06-2018	11	0	11	0	0	1
HYNC931	16-06-2018	10	1	9	0	0	2
Total		29	1	24	3	1	

D'après l'AMOVA, les populations de *L. insularis* ont leur variance moléculaire expliquée à 85,7 % par de la variance entre localités et à 14,3 % par de la variance au sein des localités. Le test exact de différenciation des populations a déterminé une différenciation significative entre les populations ($p = 0.0005$). Les deux algorithmes sont parvenus tous deux à résoudre les liens entre les 4 haplotypes hormis celui entre L_ins_COI03 et L_ins_COI04. En effet, 29 mutations sont observées entre les deux haplotypes tandis que 30 et 36 mutations sont représentées dans les réseaux calculés par median-joining (fig. 1.13) et parcimonie statistique. L'algorithme en median-joining ($\epsilon = 2$) a calculé 3 haplotypes hypothétiques. L'haplotype L_ins_COI04 est le plus éloigné des autres génétiquement et est l'unique haplotype retrouvé dans la mare la plus éloignée géographiquement, ETNC501 (entouré en trait continu sur fig. 1.13). L'haplotype L_ins_COI03 semble également se distinguer génétiquement et est également retrouvé dans une mare isolée, ETNC007 (entouré en double trait sur fig. 1.13). Or, dans cette mare on retrouve également l'haplotype le plus répandu, L_ins_COI02.

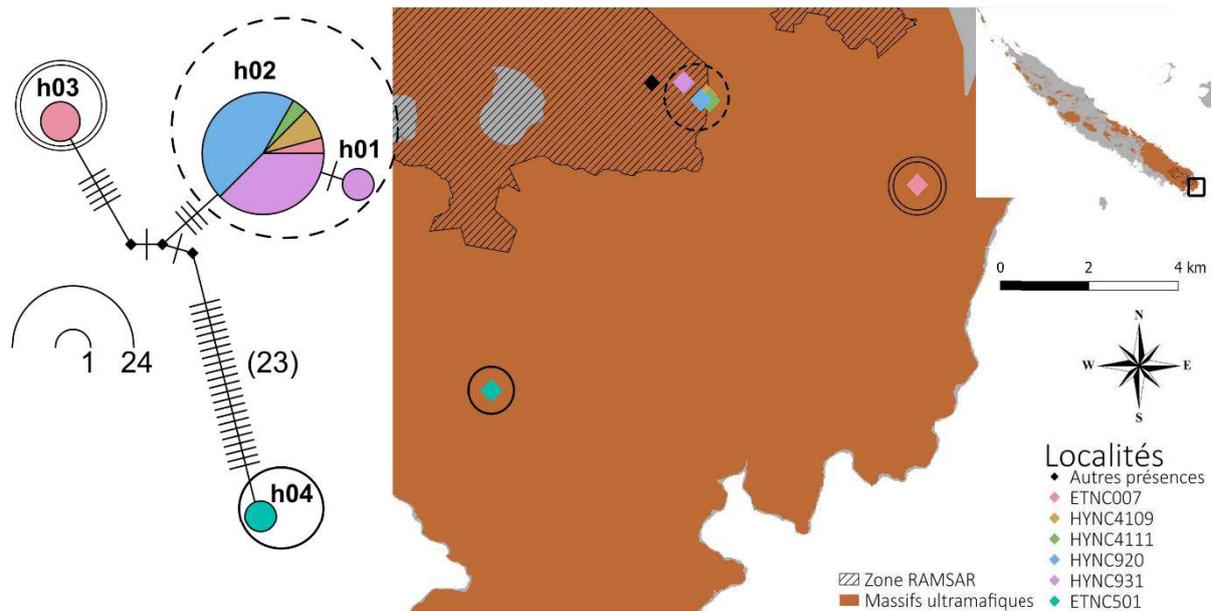


Figure 1.13 – A gauche, réseau d'haplotypes de *L. insularis* calculé en median-joining ; les haplotypes hypothétiques sont représentés par des losanges noirs ; les noms d'haplotypes sont abrégés de L_ins_COIXX en hXX pour alléger la figure. A droite, carte de répartition de l'espèce et de ses haplotypes en NCK.

Lynceus sp. A (fig. 1.14) est présente dans 19 localités de l'archipel (11 échantillonnées par filet et 8 en élevage). Elle est présente majoritairement hors des massifs ultramafiques sauf pour une mare ultramafique, HYNC928 (fig. 1.15). Elle est également présente en Australie. Pour cette espèce, 11 haplotypes différents ont été retrouvés dans 9 localités avec une distance génétique non corrigée allant jusqu'à 3,3 %. Un haplotype différent a été retrouvé chez l'individu prélevé en Australie, il est distant des haplotypes néo-calédoniens de 3,2 - 4,9 % (tab. 1.13). L'haplotype le plus répandu, L_nsa_COI02, est présent dans 3 localités. Les localités HYNC4000 et HYNC928 sont les plus riches avec 3 haplotypes retrouvés dans chacune (tab. 1.14).



Figure 1.14 - Image d'un *L. sp. A* de NCK. Echelle approximative. © J.-F. Cart

Tableau 1.13 - Matrice des distances non corrigées des fragments de COI séquencés chez *L. sp. A* en fonction des régions (UM = ultramafique = massifs riches en métaux lourds).

Région	NCK - UM	NCK - hors UM
NCK - UM	0 - 1,37 %	
NCK - hors UM	2,43 - 3,34 %	0 - 1,37 %
Australie	4,56 - 4,86 %	3,19 - 3,80 %

Tableau 1.14 - Distribution des haplotypes de *L. sp. A* amplifiés. S : date de collecte du sédiment ; E : date de collecte dans l'élevage.

Identifiant de la localité	Date(s) de collecte	Total	L_nsa_COI01	L_nsa_COI02	L_nsa_COI03	L_nsa_COI04	L_nsa_COI05	L_nsa_COI06	L_nsa_COI07	L_nsa_COI08	L_nsa_COI09	L_nsa_COI10	L_nsa_COI11	M106	Richesse en haplotypes
ETNC001	14-01-2021	4	0	0	0	0	0	0	0	0	0	4	0	0	1
HYNC4000	04-06-2018	9	4	1	4	0	0	0	0	0	0	0	0	0	3
HYNC4003	04-06-2018	3	2	0	0	1	0	0	0	0	0	0	0	0	2
HYNC4008	06-06-2018	3	0	3	0	0	0	0	0	0	0	0	0	0	1
HYNC4033	10-06-2018	1	0	0	0	0	0	0	1	0	0	0	0	0	1
HYNC834	S : 07-11-2016 E : 30-07-2017	2	0	0	0	0	1	0	1	0	0	0	0	0	2
HYNC836	S : 07-11-2016 E : 15-08-2017	1	0	0	0	0	1	0	0	0	0	0	0	0	1
HYNC928	05-06-2018	3	0	0	0	0	0	0	0	1	1	0	1	0	3
HYNC932	29-11-2016	6	0	1	0	0	0	5	0	0	0	0	0	0	2
Australie		1	0	0	0	0	0	0	0	0	0	0	0	1	1
Total		33	6	5	4	1	2	5	2	1	1	4	1	1	

D'après l'AMOVA, les populations de *L. sp. A* ont leur variance moléculaire expliquée à 75,4 % par de la variance entre localités et à 24,6 % par de la variance au sein des localités ($p = 0,001$). Le test exact de différenciation des populations a déterminé une différenciation significative entre les populations ($p = 0.0005$). Aucun des deux algorithmes n'est parvenu à résoudre les liens entre les haplotypes observés. Par exemple, 9 mutations sont observées entre les haplotypes L_nsa_COI04 et L_nsa_COI07 tandis que 11 et 8 mutations sont représentées dans les réseaux calculés par median-joining (fig. 1.15; $\epsilon = 2$) et parcimonie statistique. L'algorithme en parcimonie statistique est le plus proche de ce qui est observé en termes de mutations. Les haplotypes trouvés hors des massifs ultramafiques L_nsa_COI01 - L_nsa_COI07 et L_nsa_COI10 forment un groupe génétiquement proche malgré qu'ils soient trouvés dans des localités réparties largement sur l'île. Ce groupe d'haplotypes se place au centre du réseau entre l'haplotype australien (M106) et les haplotypes trouvés dans la mare ultramafique HYNC928 (L_nsa_COI08, L_nsa_COI09 et L_nsa_COI11 ; entourés sur fig. 1.15).

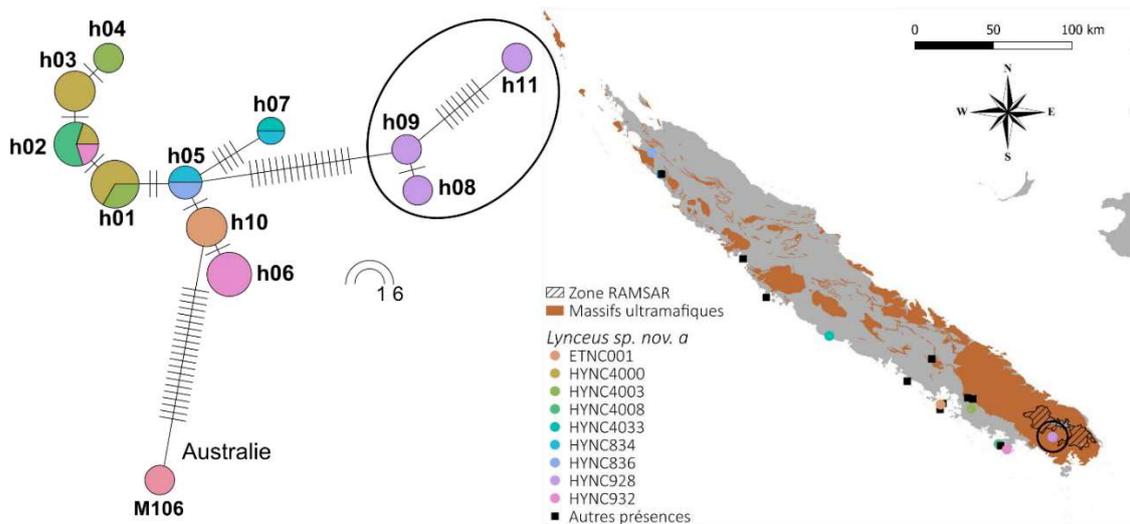


Figure 1.15 - A gauche, réseau d'haplotypes de *L. sp. A* calculé en parcimonie statistique ; les noms d'haplotypes sont abrégés de L_nsa_COIXX en hXX pour alléger la figure ; l'haplotype australien, en rose, est nommé M106. A droite, carte de répartition de l'espèce et de ses haplotypes en NCK.

Lynceus sp. B est présente dans deux localités de l'archipel distantes de moins d'un kilomètre (toutes échantillonnées par filet). Elle est présente exclusivement sur les massifs ultramafiques et n'est pas présente en Australie (fig. 1.16). Pour cette espèce, 4 haplotypes différents ont été retrouvés avec une distance génétique non corrigée allant jusqu'à 0,8 % (tab. 1.13). Aucun haplotype n'est commun aux deux localités. La localité HYNC2608 est la plus riche avec 3 haplotypes retrouvés (tab. 1.15).

Tableau 1.15 - Distribution des haplotypes de *L. sp. B* amplifiés.

Identifiant de la localité	Date(s) de collecte	Total	L_ins_COI01	L_ins_COI02	L_ins_COI03	L_ins_COI04	L_ins_COI05	Richesse en haplotypes
HYNC2604	09-04-2018	2	0	0	1	1	0	2
HYNC2608	09-04-2018	9	7	1	0	0	1	3
Total		11	7	1	1	1	1	

D'après l'AMOVA, les populations de *L. sp. B* ont leur variance moléculaire expliquée à 70,6 % par de la variance entre localités et à 29,4 % par de la variance au sein des localités ($p = 0,03$). Le test exact de différenciation des populations a déterminé une différenciation significative entre les populations ($p = 0.0005$). Les deux algorithmes sont globalement parvenus tous deux à résoudre les liens entre les 5 haplotypes hormis pour deux liens chacun. Le réseau d'haplotype calculé en parcimonie statistique représente deux et quatre mutations supplémentaires entre

l'haplotype L_nsb_COI05 et les haplotypes L_nsb_COI04 et L_nsb_COI04, respectivement. Le réseau d'haplotype calculé en median-joining (fig. 1.16; $\epsilon = 2$) représente deux mutations supplémentaires entre l'haplotype L_nsb_COI05 et les haplotypes L_nsb_COI01 et L_nsb_COI02. Ce dernier a calculé deux haplotypes hypothétiques.

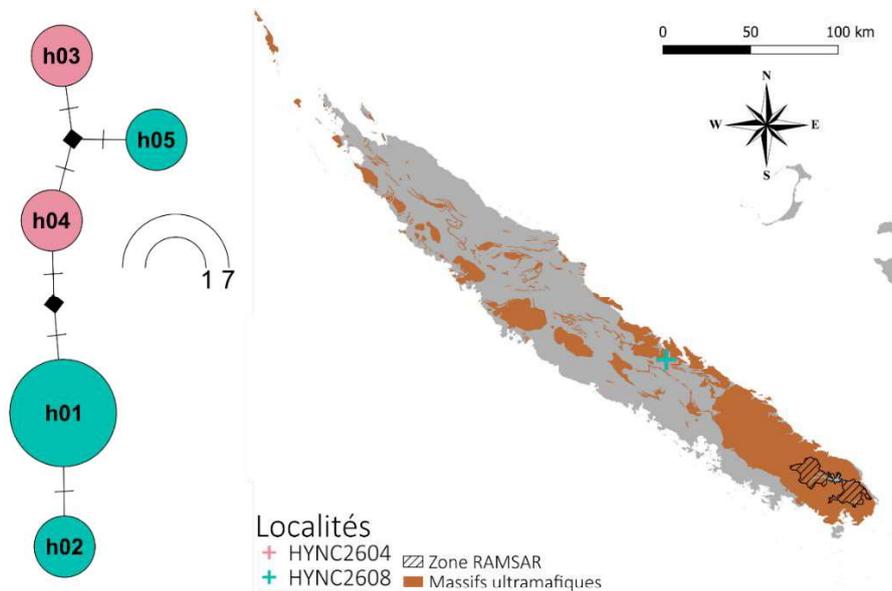


Figure 1.16 – A gauche, réseau d'haplotypes de *L. sp.* B calculé en median-joining ; les noms d'haplotypes sont abrégés de L_nsb_COIXX en hXX pour alléger la figure. A droite, carte de répartition de l'espèce et de ses haplotypes en NCK.

V. Discussion

Comme évoqué en introduction, le biotope néo-calédonien semble distinct entre les massifs riches en métaux lourds (ultramafiques) et les autres types de massifs chez une majorité de groupes (Caesar *et al.*, 2017; Hudel *et al.*, 2020; Pillon *et al.*, 2021). Un tel constat semble valable également pour les espèces retrouvées dans les échantillons étudiés dont la majorité sont inféodés à l'un ou l'autre type de massif (tab. 1.16). En effet, seules 10 espèces sur 41 semblent présentes sur tous les types de massifs de l'archipel. Parmi ces 10 espèces, 5 sont potentiellement ambiguës dans leur identification (famille Chydoridae et genre *Ilyocryptus*). Majoritairement, ces espèces sont retrouvées souvent hors des massifs ultramafiques et ponctuellement (1 à 3 localités) sur les massifs ultramafique.

Au total, 9 espèces nouvelles sont à décrire formellement dont 7 sont appuyées par des éléments de génétique.

Les résultats présentés dans ce chapitre montrent trois scénarios de distribution de la diversité génétique chez les espèces étudiées (tab. 1.16).

Tableau 1.16 – Caractéristiques des espèces étudiées d'un point de vue génétique.

Espèce	Présences	Sols	Haplotypes	Localités haplotypes	Distance génétique max	AMOVA		Structuration
						Inter-loc	Intra-loc	
<i>Latonopsis australis</i>	7	Non-UM	/	/	/	/	/	/
<i>Streptocephalus archeri</i>	17	Non-UM	4	7	1,22 %	21,9 %	78,1 %	Non
<i>Eulimnadia</i> sp.	28	Non-UM	7	11	0,76 %	69,3 %	30,7 %	Non
<i>Triops intermedius</i>	4	Non-UM	2	3	0,15 %	/	/	Non
<i>Lynceus</i> sp. A	19	Non-UM et UM	11	9	3,3 %	75,4 %	24,6 %	Géographique et en fonction des massifs
<i>Lynceus insularis</i>	7 (+ 5)	UM	4	6	4,4 %	85,7 %	14,3 %	Géographique
<i>Lynceus</i> sp. B	2	UM	5	2	0,8 %	70,6 %	29,4 %	Géographique
<i>Latonopsis dedeckeri</i>	16	UM	7	6	7,9 %	98 %	2 %	Géographique
<i>Boeckella spinogibba</i>	14	UM	18	13	2,86 %	84,6 %	15,4 %	Géographique
<i>Boeckella sibleti</i>	4	UM	9	4	4,41 %	97,6 %	2,4 %	Géographique
<i>Boeckella</i> sp. A	1	UM	1	1	/	/	/	Localité unique
<i>Boeckella</i> sp. B	1	UM	1	1	/	/	/	Localité unique
<i>Boeckella</i> sp. C	2	UM	1	1	/	/	/	Localité unique
<i>Boeckella</i> sp. D	2	UM	1	2	/	/	/	Localité unique

Les espèces étudiées présentes hors des massifs ultramafiques, comme *S. archeri*, montrent une diversité et une structuration génétique a priori faible et une large aire de distribution. On observe l'inverse pour les espèces présentes sur les massifs ultramafiques, comme *B. sibleti*, qui montrent une diversité et une structuration génétique élevée en comparaison (tab. 1.16). En plus de ces deux scénarios opposés, un scénario intermédiaire est représenté par *Lynceus sp. A*, chez qui la diversité et la structuration génétique semblent dépendantes du type de sol. En effet, pour cette espèce, la distance génétique observée entre les haplotypes retrouvés en dehors des massifs riches en métaux lourds (hors ultramafique) n'excède pas les 1,4 % (tab. 1.13) alors qu'elle varie entre 2,4 % et 3,3 % quand on les compare aux haplotypes des massifs ultramafiques. Au sein des haplotypes des massifs ultramafiques, on observe également une distance génétique de 1,4 %. Or, ces haplotypes ont été échantillonnés dans une seule mare tandis que les haplotypes trouvés hors des substrats ultramafiques ont été échantillonnés dans 9 mares différentes réparties sur Grande Terre. Il serait donc intéressant d'analyser également la structure génétique des populations de *Diaphanosoma australiensis* afin de les comparer avec *L. sp. A* du fait de leur distribution similaire.

Chez les genres *Triops* et *Eulimnadia*, il est surprenant de ne retrouver qu'une seule espèce génétiquement homogène sur l'archipel. En effet, ces deux lignées sont très diversifiées en Australie (*Triops* 23 clades et *Eulimnadia* 14 clades) avec de nombreuses espèces sympatriques (fig. 1.3, 1.7 ; Schwentner *et al.*, 2015; Meusel & Schwentner, 2017). Ce type de différence peut s'expliquer notamment par l'insularité et une colonisation récente de l'archipel. Les populations des îles ont tendance à présenter une diversité génétique plus faible par l'effet pionnier (Whittaker *et al.*, 2023).

La diversité génétique intra-spécifique est largement étudiée dans les milieux aquatiques (Manel *et al.*, 2020; Tsuji *et al.*, 2020; Garrigue *et al.*, 2022; Duan *et al.*, 2023; Haÿ *et al.*, 2024). Dans les mares d'eau douce temporaires ou non, les crustacés représentent un des modèles les plus utilisés pour étudier la structure génétique des populations (De Meester, 1996; Martin *et al.*, 2021; Cervantes-Martínez *et al.*, 2021; Huang *et al.*, 2022). Une structuration génétique différente en fonction du type de sol comme observé ici n'a à ce jour jamais été évoquée chez ces groupes. Cependant, une structuration géographique de la diversité morphologique a été constatée chez l'espèce d'Anostracé *Branchipodopsis wolffi* présente dans le Sud de l'Afrique. Celle-ci n'a pas été reliée aux caractéristiques de

l'environnement mais aux faibles capacités de dispersion de l'espèce (Brendonck *et al.*, 2017).

La diversité génétique intra-spécifique a été peu étudiée dans les eaux douces de l'archipel néo-calédonien (Mennesson *et al.*, 2018). La diversification génétique a cependant été étudiée chez les plantes, qui ont un mode de dispersion similaire aux crustacés étudiés ici. En effet, de la même manière que les œufs de résistances, les graines sont des formes de résistances qui sont également des moyens efficaces de dispersion par le vent, le courant, ou encore les animaux (Van Damme & Sinev, 2013; Incagnone *et al.*, 2015). Contrairement à ce qui est observé chez les espèces de crustacés strictement inféodées aux milieux ultramafiques, peu de structuration géographique a été observée chez l'espèce de plante carnivore *Nepenthes vieillardii* (Kurata *et al.*, 2008). Cependant, de nombreux groupes de plantes ultramafiques montrent une structuration géographique de leur diversité génétique (*e.g.*, *Araucaria*, *Spiraeanthemum*, *Amborella* ; Pillon *et al.*, 2009; Poncet *et al.*, 2013; Ruhsam *et al.*, 2016). Chez le genre de pin *Araucaria*, l'analyse de la diversité génétique chez *A. montana*, *A. rulei* et *A. scopulorum* a démontré une structuration géographique claire (Gaudeul *et al.*, 2014). *A. rulei* et *A. scopulorum* sont strictement cantonnées aux massifs ultramafiques tandis que *A. montana* est présente sur les massifs ultramafiques mais aussi hors de ceux-ci. Comme observé chez *Lynceus sp. A*, la diversité génétique est structurée particulièrement vis-à-vis du type de sol et les populations hors des massifs ultramafiques semblent montrer une diversité génétique plus faible. D'après Gaudeul *et al.* (2014), la divergence génétique retrouvée entre les deux types de sols pourrait être due à une adaptation chez certaines graines à germer sur l'un ou l'autre type de sol, entraînant une diminution des flux de gènes entre eux. Cette même hypothèse peut être posée pour *Lynceus sp. A*. Cependant, comme évoqué dans les articles sur le genre *Boeckella* (Royaux *et al.*, 2024 ; Royaux *et al.*, soumis), dans les milieux aquatiques, la structuration géographique pourrait ne pas être due directement au type de sol. En effet, du fait de la toxicité des niveaux en métaux lourds, les mares ultramafiques sont oligotrophes et peu de proies y sont disponibles pour les prédateurs, notamment les oiseaux limicoles. Ceux-ci visiteraient moins les mares ultramafiques, ce qui créerait un différentiel entre la capacité de dispersion des populations des massifs ultramafiques et celles hors de ceux-ci. Ces hypothèses peuvent également poser la question d'une spéciation en cours chez *Lynceus sp. A* entre les populations d'Australie et de la côte ouest de Grande Terre et la population des massifs ultramafiques (HYNC928).

En effet, les hypothèses évoquées pourraient également expliquer la structuration de la diversité spécifique qui semble structurée de manière similaire. On observe que les genres strictement hors des massifs ultramafiques (*Eulimnadia*, *Triops*, *Streptocephalus*) ont une espèce unique en NCK et la majorité sont aussi retrouvées en Australie. La seule exception est *Eulimnadia sp. A*. Or, cette espèce pourrait être présente en Australie mais avoir été confondue avec *Eulimnadia australiensis* du fait de la forme similaire de leurs œufs (Timms, 2016). Par opposition, aucune espèce présente sur les massifs ultramafiques n'est retrouvée en Australie et les groupes *Lynceus* et *Boeckella* montrent une diversité remarquable. Ces groupes présentent des caractéristiques de l'endémisme que l'on pourrait appeler micro-endémisme (*i.e.*, endémisme à petite échelle, endémisme intra-insulaire) pour certaines espèces (*Lynceus insularis*, *Lynceus sp. B*, *Latonopsis dedeckkeri*, toutes les espèces de *Boeckella*). L'endémisme et le micro-endémisme en NCK est particulièrement observé sur les massifs ultramafiques chez la plupart des groupes (Nattier *et al.*, 2013; Caesar *et al.*, 2017; Pillon *et al.*, 2021). Or, chez les Gastéropodes des rivières de l'archipel, le micro-endémisme semble réparti par régions ou bassins versants plutôt qu'en fonction du type de massif (Haase & Bouchet, 1998). En effet, ces groupes ne présentent pas de forme de résistance et la dispersion est assurée par les larves ou les adultes entraînant qu'une dispersion entre bassins versants est très rare sauf pour les espèces amphihalines qui passent par la mer. Ici, bien que les groupes étudiés dans cette thèse aient un mode de dispersion similaire entre elles (Incagnone *et al.*, 2015), la distribution de la diversité spécifique et génétique semble conditionnée par diverses contraintes (*e.g.*, toxicité, milieux oligotrophes) liées notamment à la nature de la roche mère dans les zones occupées par les populations. La diversité et la structuration géographique trouvée aux niveaux spécifiques et génétiques chez les crustacés étudiés semble indiquer une très faible capacité de dispersion dans les massifs ultramafiques de l'archipel. Comme discuté dans les articles sur le genre *Boeckella* (Royaux *et al.*, 2024 ; Royaux *et al.*, soumis), la forte perturbation liée à l'exploitation minière et des terres riches en métaux lourds associée à cette faible capacité de dispersion rends la situation potentiellement critique pour la conservation des espèces étudiées.

Chapitre II

La distribution des espèces et assemblages de crustacés
d'eau douce de Nouvelle-Calédonie-Kanaky

I. Résumé

Sur la base des espèces identifiées dans le chapitre précédent, ce second chapitre porte sur les facteurs influençant la répartition des espèces retrouvées sur l'archipel de NCK. Les principaux résultats ont été obtenus sur la distribution des assemblages d'espèces, soit les groupes d'espèces qui sont retrouvés ensembles au sein de mêmes localités (voir l'article en partie II). Malgré de nombreux essais pour produire un modèle de la répartition individuelle de chaque espèce, ceux-ci se sont avérés infructueux (voir encadré B).

Dans le chapitre précédent, il a pu être constaté qu'il semble exister une scission entre les caractéristiques des espèces présentes exclusivement sur les massifs ultramafiques, exclusivement hors des massifs ultramafiques et sur tous les types de sols (*Lynceus sp. A*, *Latonopsis dedeckkeri*). Ici, l'objectif est d'investiguer les caractéristiques des assemblages d'espèces en NCK, de vérifier si une scission similaire est observée à ce niveau d'organisation et si d'autres facteurs environnementaux impactent la répartition des assemblages.

En plus des massifs ultramafiques, l'archipel offre une grande diversité de conditions environnementales auxquelles sont également attribuées les taux élevés de micro-endémisme (Nattier *et al.*, 2013; Gourmelon *et al.*, 2016; Isnard & Jaffré, 2024). En effet, la diversité des sols et la pluviométrie annuelle variant entre 840 et 3340 mm par an en fonction des régions de l'archipel conditionnent en grande partie les écosystèmes néo-calédoniens. Cette variabilité de pluviométrie a certainement une incidence sur l'hydropériode (*i.e.*, durée ou fréquence de mise en eau) qui conditionne en partie la présence des espèces dans les milieux aquatiques (Schneider & Frost, 1996; Tavernini, 2008; Olmo *et al.*, 2016).

D'après les analyses des échantillons triés, deux assemblages principaux ont été retrouvés en NCK à partir d'un réseau biogéographique et d'un algorithme de partitionnement hiérarchique. A travers l'utilisation de modèles d'ensembles, la distribution de l'assemblage le plus répandu a pu être associée à l'absence de massifs ultramafiques et à des niveaux de précipitation inférieurs à 2000 mm/an. Le second assemblage a été associé à la présence de massifs ultramafiques et à des niveaux de précipitations supérieurs à 1700 mm/an. Bien qu'une plus grande diversité spécifique ait été estimée pour l'assemblage le plus répandu, l'assemblage le moins répandu semble présenter une plus haute proportion d'espèces rares.

Les modèles calculés ont permis d'estimer les régions où chaque assemblage a la plus haute probabilité d'être retrouvé. Il existe des zones de frontières entre chaque assemblage qui semblent présenter une faible probabilité de présence pour chacun des assemblages. L'hypothèse qu'un assemblage n'ait pas été détecté est proposée. En effet, plusieurs localités semblent présenter une proportion équilibrée d'espèce de chacun des deux assemblages délimités (localités HYNC928 et ETNC120). Cet assemblage non détecté pourrait représenter une zone tampon entre les deux assemblages principaux. Deux autres assemblages mineurs ont été détectés par l'algorithme de partitionnement dont l'un, retrouvé dans la région de Canala, pourrait représenter un assemblage bien distinct mais sous-échantillonné. Ces résultats sont détaillés et discutés dans l'article en préparation présenté en partie II de ce chapitre.

II. La distribution des assemblages d'espèces

Major drivers of community assemblages in ponds of a biodiversity hotspot

Royaux Coline | Leroy Boris | Mary Nathalie | Charpin Nicolas | Le Bras Yvan | Rabet Nicolas

En préparation

Abstract

Despite being present on all continents including Antarctica, ponds remain one of the least studied surface freshwater ecosystem. Ponds are particularly diverse in their physicochemical characteristics which is reflected in the biological assemblages inhabiting them. On the New Caledonia archipelago, acknowledged to be a global biodiversity hotspot, lentic freshwater habitats are numerous. The archipelago is subjected to extremely diverse environmental conditions notably divided along two approximate axes. The north-east has high precipitation rates (around 3000 mm/y) for the most part while the south-west has low precipitation rates (around 900mm/y). On a practically perpendicular axis, the south-east has soils with high heavy metal concentration while it is low in the north-west. We analysed the crustacean species inhabiting small lentic freshwaters and delineated two major community assemblages with network clustering analyses. We then modelled the drivers of the distribution of these communities and identified heavy metal soils and annual precipitation rates as the major drivers. Our models allowed us to project the distribution of most probable areas for each assemblage and suggest the existence of a buffer area between the two assemblages.

Introduction

Small lentic freshwater habitats, here designated as ponds, are ubiquitous throughout all climatical regions and cover a greater area of the Earth surface than lakes (Williams, 2007; Downing *et al.*, 2006). Despite this prevalence, ponds are significantly more fragmented in their distribution. They represent a diverse and dynamic habitat that exhibits considerable spatial and temporal heterogeneity. Indeed, a number of variables, including pH, trace element concentration, turbidity and deepness, are dependent on a range of geographical characteristics, such as soil type and topography, as well as temporal factors, such as recent evaporation rates and precipitation levels (Williams, 2007; Tavernini, 2008). These environments are

nonetheless understudied in comparison to other surface freshwater environments such as lakes or rivers, especially in tropical regions (Faghihinia *et al.*, 2021; Hill *et al.*, 2021).

A large proportion of ponds worldwide undergo cyclic drying which has profoundly structuring influence on the biota. The persistence of populations in temporary ponds implies necessary drought survival traits, such as diapausing eggs (Brock *et al.*, 2003; Williams, 2007; Incagnone *et al.*, 2015). Furthermore, the habitat duration, also called hydroperiod, is a crucial factor influencing populations and communities inhabiting temporary ponds (Schneider & Frost, 1996; Tavernini, 2008; Olmo *et al.*, 2016). For example, a species that attains maturity 13 days after the pond is flooded such as *Lynceus bififormis* would not be observed in a temporary pond with a 10-day hydroperiod (Wang *et al.*, 2014). Species inhabiting ponds are notably micro-crustaceans. Continental micro-crustacean species are facing many knowledge shortfalls (Hortal *et al.*, 2015). As modelled by Macedo *et al.* (2024), many species from this group are not described, especially in temporary waters. In New Caledonia, many species of micro-crustaceans are known from less than 6 locations, if not a single one (Longhurst, 1955; Timms, 1985; Olesen *et al.*, 2016).

Community assemblages formed in these environments by crustacean species are highly reliant on all previously cited variables of the environment they live in (Feld *et al.*, 2016). A community assemblage is defined as a multi-specific group of taxonomically related populations that occur in a given space (Stroud *et al.*, 2015). In New Caledonia, lentic freshwater crustacean community assemblages are notably composed of the endemic longtail tadpole shrimp sub-species *Triops longicaudatus intermedius* (Longhurst, 1955), of the recently described copepod *Boeckella sibleti* (Royaux *et al.*, 2024), of the globally distributed Ostracoda species *Candonocypris novaezelandiae* (Martens *et al.*, 2019), and of the remarkable Australian-New Caledonian water flea *Daphnia longicephala* (Timms, 1985).

Nouvelle-Calédonie (New Caledonia, Kanaky) is an archipelago located in the southwest Pacific Ocean. Its main island, Grande Terre, is continental and separated from Australia approximately 105 Mya (Maurizot & Campbell, 2020). The archipelago is known for its outstanding biodiversity and has one of the highest proportions of endemic species on earth (Veron *et al.*, 2019). In New Caledonia, the annual precipitation rates exhibit a considerable range over the territory, from 887 to 3339 mm, which results to markedly disparate conditions in the numerous permanent and temporary ponds that are distributed over the archipelago. The

density of lentic freshwaters is notably high in the extreme south of Grande Terre in the Plaine des Lacs and the Parc Provincial de la Rivière Bleue, which has been designated a “wetland of international importance” under the Ramsar Convention (Jeanpert *et al.*, 2016). This area is also notable for its heavy metal-rich soils, which originate from the alteration of ultramafic massifs. Such massifs cover approximately one-third of the archipelago and are considered as a significant factor in environmental filtering there (Zakardjian *et al.*, 2023; Isnard & Jaffré, 2024). Indeed, a significant proportion of endemic species of the archipelago are known to be exclusively found in ultramafic environments (Nattier *et al.*, 2013; Pillon *et al.*, 2021). However, the high concentrations of notably nickel, cobalt, and chromium in ultramafic rocks and soils are an important economic resource for the archipelago. The mining activity on ultramafic massifs is present in Grande Terre and surrounding islands since the 19th century, which has many impacts on landscapes and ecosystems (Pascal *et al.*, 2008; Boula *et al.*, 2022; Germande *et al.*, 2022). Freshwaters are particularly sensitive to mining and other anthropogenic perturbations, which have led them to be considered among the most vulnerable environments in terms of biodiversity (Tickner *et al.*, 2020). Indeed, in New Caledonia, many ponds are notably found in and around the Goro mining facility in the Plaine des Lacs. This facility extracts nickel from soils with a rather low heavy metal concentration. The process inevitably produces many chemical residuals which pollutes freshwaters (Germande *et al.*, 2022). In addition to evident impact through the excavation of soils, it creates inert wastes in the form of fine dust that is filling local freshwaters (Boula *et al.*, 2022). These threats are particularly significant for pond micro-crustaceans that are inhabiting freshwater islands on dry oceans they cannot cross on their own (Incagnone *et al.*, 2015). They have never been studied from the perspective of species assemblages on the archipelago.

Based on 70 samples of crustaceans from freshwater ponds on the archipelago, we outline main community assemblages present in these environments using a biogeographical network. The sampling completeness and diversity estimates at several diversity orders are evaluated to account for the accuracy of the assemblage delineation and their composition in rare, frequent and highly frequent species. The results of these analyses could provide insight into the extent of the knowledge gap that remains to be addressed regarding these species. Based on the assemblages delineations, we tested for the affinity of the distinct assemblages to ultramafic massifs, annual precipitation rates and distance to freshwaters in association with the sampling bias from the study using ensemble models. Finally,

based on the ensemble models, we predicted the probable distribution of each assemblage on the archipelago.

Material and methods

Sampling and identification

Samples were collected on the New Caledonian archipelago during three “La Planète Revisitée” campaigns between 2016 and 2018, and by the environmental consulting firm “Ethyco” and the organisation “Vies d’Ô douces” between 2020 and 2023. During these sampling expeditions, 535 locations were visited. However, a significant proportion of these were found to be dry, and thus 197 were sampled using a plankton net (NHBS, Bonn, Germany) with a frame diameter of 25 cm and a 55 cm-long bag of 200 µm nylon mesh with a filter at the tip. The content of the filter is preserved in 80-90% ethanol.

In total, 70 randomly selected samples from 65 locations were entirely sorted to identify the species of crustaceans (fig. 2.1; supplementary tab. DS1). Identifications under binocular microscope (and optical microscope when required) were based on keys by Quinlan & Bayly (2017), Korovchinsky (2001), Elmoor-Loureiro (1997), Timms (1985), Smirnov & Timms (1983), and Stingelin (1915). All individuals of the same species and from the same sample were pooled in clean 80-90% ethanol in a glass tube with a screw cap and identification were verified by 2-3 different persons.

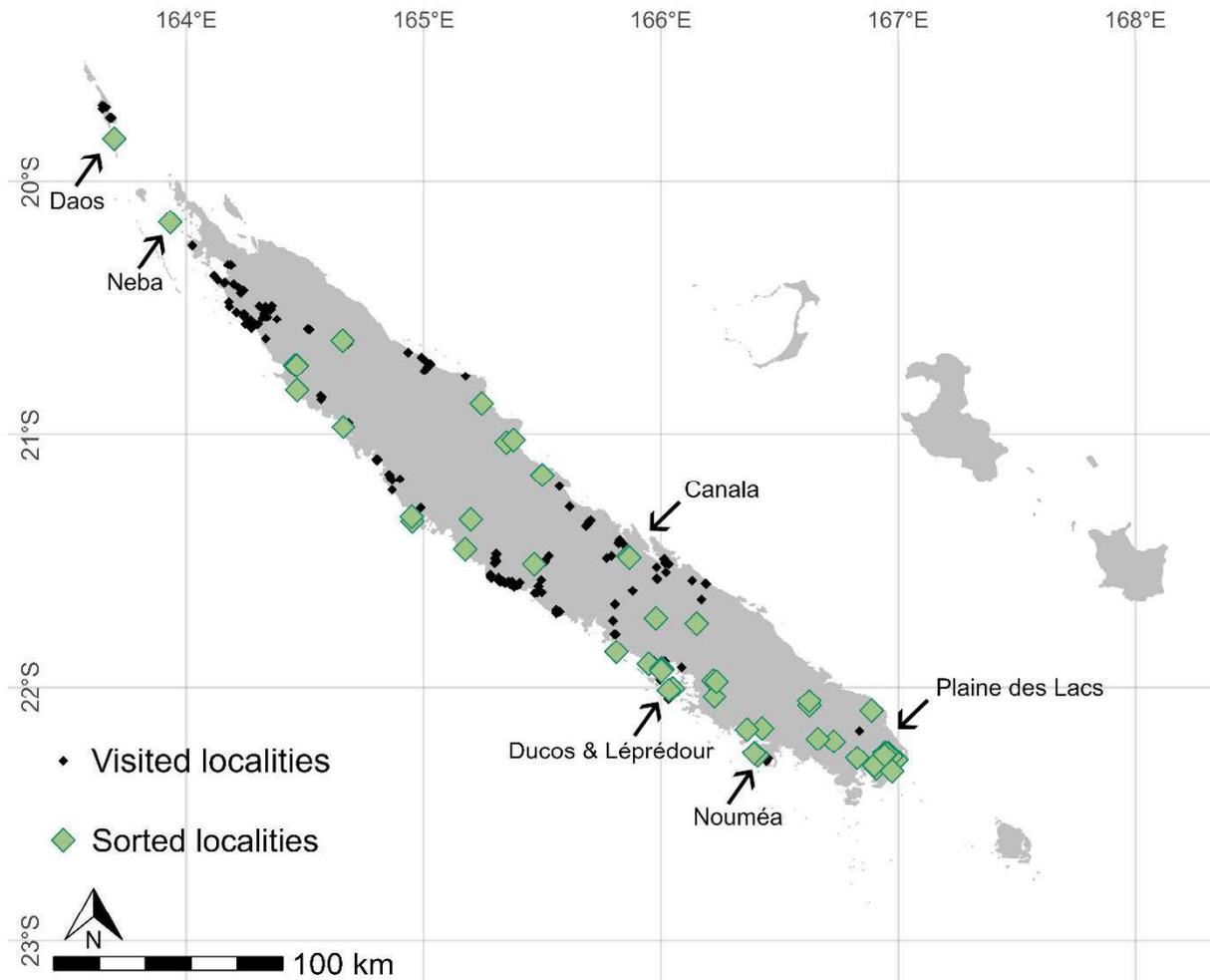


Figure 2.1 - Visited and sorted localities.

Community assemblage delineation

For the community assemblage delineation analysis, several filters were applied on raw data. Identifications of class Ostracoda and of orders Cyclopoida and Harpacticoida were removed as the formal identification requires damaging individuals with micro-dissections. The taxonomy of the family Chydoridae has known many changes and re-arrangements in the past decades (Sinev, 2004; Van Damme & Dumont, 2008; Gu *et al.*, 2022). As most of the identified Chydoridae species in the samples were not recently revised, identifications of this family were ambiguous and were removed from the data. Additionally, individuals identified from the genus *Ilyocryptus* seemed to present traits from two morphologically close species *I. sordidus* and *I. spinifer* (Kotov & Elias-Gutierrez, 2009). As the two were undistinguishable, identifications of this genus were filtered as well. Finally, identifiable moult and eggs were not considered as a presence of the related species in the analysis.

To cluster samples into communities, we used a bipartite species-sample network which is designed with nodes and links, each node can represent a species or a sample. When a species occurs in a particular sample, the node of the species is bound with the node of the sample through a link. In this kind of bipartite network, direct species-species or sample-sample links are not allowed. Assemblages were delimited on the network with a hierarchical clustering algorithm with 1000 trials and 1000 draw of random seeds using the R 4.2.2 package *bioregion* 1.1.1 and the Map equation algorithm *infomap* 2.7.1 (Rosvall & Bergstrom, 2008; Vilhena & Antonelli, 2015; Bloomfield *et al.*, 2018; Leroy *et al.*, 2019). The clustering algorithm is delineating assemblages with high intra-group and low inter-group connectivity. It may detect several hierarchical levels of community assemblage delineations (or clusters) in the network and associates each node (species and samples) with an assemblage for each level. We chose these methods because they are reliable even with low species richness and are preserving the identity of the species as opposed to methods based on dissimilarity metrics such as β diversity (Leroy *et al.*, 2019).

The final network was handled with *gephi* 0.9.2 (Bastian *et al.*, 2009) and the *forceatlas2* algorithm that groups interconnected nodes together and separates less connected nodes. The final representation of the network was reworked for better clarity using *Inkscape* 1.2.2 (The Inkscape Project).

Sampling completeness and diversity estimates

For each delineated assemblage, sampling completeness and diversity estimates were computed using the interpolation and extrapolation for three dimensions of biodiversity method based on Hill numbers (Hill, 1973; Chao *et al.*, 2020; Ramiro-Sánchez *et al.*, 2023). Hill numbers characterise the diversity of species with equal frequencies depending on the q parameter and, although typically computed on abundance data, they can also be computed on occurrence data as this is the case here. Hence, each order $q = 0$, $q = 1$, and $q = 2$ corresponds to the estimation of the total species richness, the number of frequent species and the number of highly frequent species respectively (see Chao *et al.*, 2020). Estimates of Hill numbers at the three orders were computed on the whole dataset and on delimitations based on the previously delineated assemblages using the R 4.2.2 packages *iNEXT.3D* 1.0.4 and *iNEXT.4steps* 1.0.0 (Chao *et al.*, 2024). Sampling completeness estimation is critical to validate the delineated assemblages and identify lacks in

the dataset. Estimated Hill numbers permit to better comprehend each delineated assemblage and compare their characteristics.

Community assemblage distribution modelling

The relationships between the distributions of the assemblages and their environment were modelled in order to identify the main drivers influencing these distributions and to predict the most probable distribution area for each assemblages accross the archipelago. A given assemblage is considered to be present in a location when at least one of its samples has been associated with it, and absent when none of its samples has been associated with it. When two samples from a same location are associated with two different assemblages, the two community assemblages are considered present. Four response variables were tested to model the potential distribution of each assemblage. The presence of ultramafic soils is known to be highly structuring for species and community assemblages on the New Caledonian archipelago (Nattier *et al.*, 2013; Pillon *et al.*, 2021; Isnard & Jaffré, 2024). The annual precipitation amount is considered here as a proxy of habitat duration of freshwater ecosystems (Tavernini, 2008) and the distance to freshwater as a proxy of habitat presence probability. Finally, the samples examined in this article were not destined to such analyses and the sampled habitats exhibited varying degrees of accessibility, resulting in localised patchy sampling. Consequently, it seemed critical to take account of sampling density to model the potential distribution of delimited assemblages.

The presence of ultramafic massifs (fig. 2.2a; peridotite v23-09-2021; <https://georep-dtsi-sgt.opendata.arcgis.com/datasets/dtsi-sgt::massifs-de-p%C3%A9ridotites-au-50-000-2>), the annual precipitation amount (fig. 2.2b; CHELSA 2.1, <https://chelsa-climate.org/bioclim/>) and the distance to freshwater based on the global maximum water extent (fig. 2.2d; 1984-2021, European Commission, <https://global-surface-water.appspot.com/>; Pekel *et al.*, 2016) were found on online open data repositories. The sampling density (kernel) of all campaigns was computed with R 4.2.2 package MASS 7.3-58.1 (fig. 2.2c).

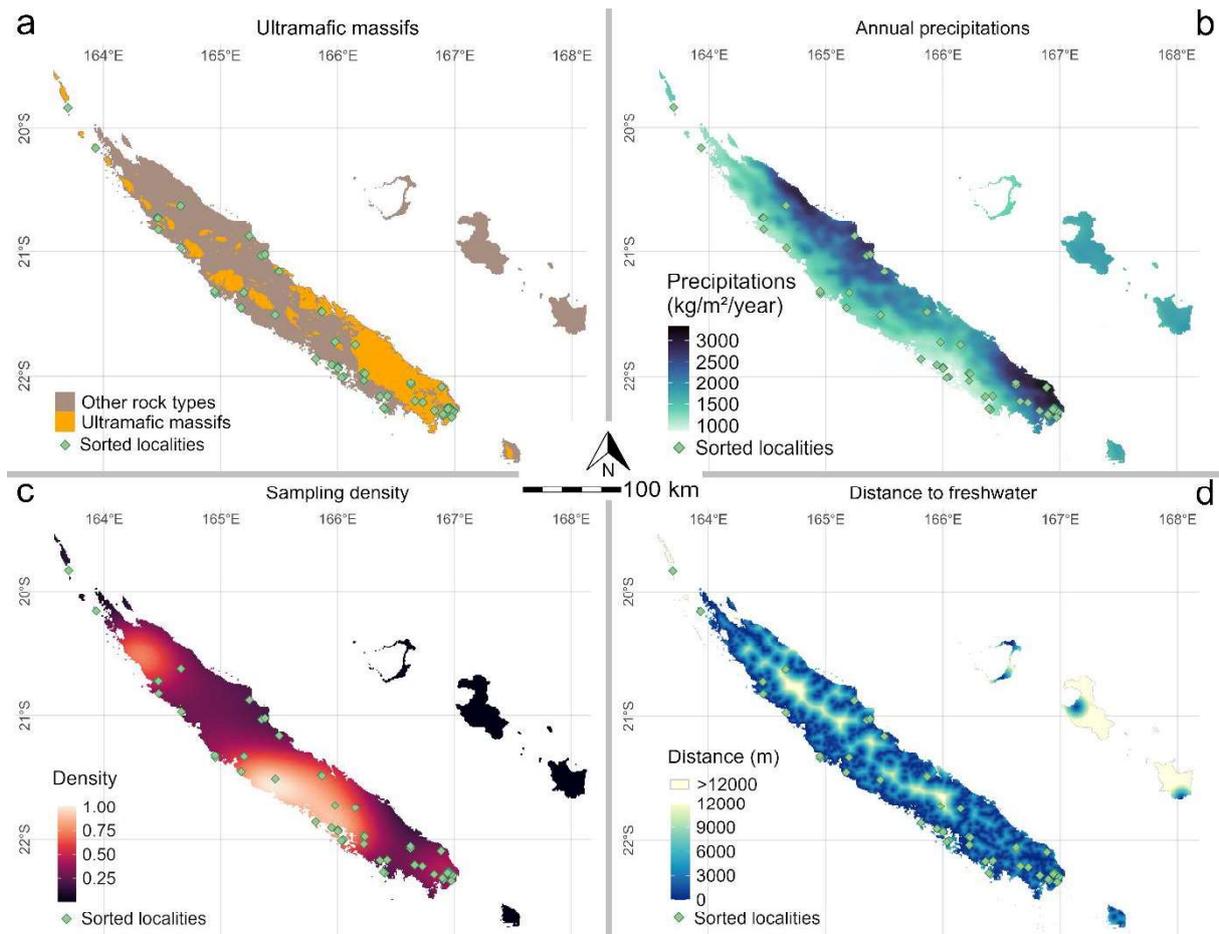


Figure 2.2 - Map of sorted localities and response variables tested for the community assemblage distribution models. Presence-absence of ultramafic massifs (a); annual precipitation modelled by CHELSA (b); expeditions' sampling kernel density (c); and distance to freshwater derived from the global maximum water extent (d).

All four variables were uncorrelated according to Pearson's coefficient (supplementary fig. DS2). The final environmental stack had 0.0083×0.0083 degree resolution. Community assemblage distribution models were calibrated and computed on geographically filtered occurrences using mostly R 4.2.2 packages terra 1.7-71 and biomod2 4.2-4. Models were computed with down-sampled Random Forest (RF; Breiman, 2001; Valavi *et al.*, 2021) and down-weighted Generalised Boosted Regression Modelling (GBM; Ridgeway, 1999), and extreme Gradient Boosting (XGBOOST; Chen & Guestrin, 2016) algorithms.

The down-sampling and down-weighting were used to balance the statistical contribution of presences and absences in the models (Guisan *et al.*, 2017; Valavi *et al.*, 2021). For each assemblage, mean presence probability of all individual models permitted to define a binary map representing the most probable assemblage at each pixel. Additionally, a map of the mean presence probability of the most probable assemblage has been produced to estimate the level of certainty of the binary map.

Each model was performed with five runs of random cross-validation calibrated on 80% of the dataset. The evaluation of each model is based on 20% of the dataset using TSS, ROC and Jaccard estimates.

Encadré A : Méthode d'évaluation des modèles

Les modèles calculés sont évalués par validation croisée (cross-validation). Cette méthode consiste à calibrer le modèle sur un pourcentage des données (*e.g.*, 80%) et à estimer la validité du modèle sur le reste des données (*e.g.*, 20%). Il existe plusieurs stratégies de validation croisée telles que la validation croisée en blocs (Roberts *et al.*, 2017) ou le tirage aléatoire des données qui seront utilisées pour la calibration et l'évaluation qui est la stratégie utilisée ici. Le modèle va être exécuté plusieurs fois et les données utilisées pour la calibration et l'évaluation vont être tirées aléatoirement à chaque exécution. De ce fait, la multiplication des exécutions de modèle va permettre de calibrer et évaluer le même modèle avec différentes parties des données à chaque fois. Les métriques utilisées pour l'évaluation sont le TSS (True Skill Statistic ; Allouche *et al.*, 2006), le ROC (Receiver Operating Characteristic ; Hanley & McNeil, 1982) et l'indice de Jaccard (Jaccard, 1901). En fonction des valeurs de ces métriques calculées pour chaque exécution de modèle, les modèles sont sélectionnés ou non pour être inclus dans le modèle d'ensemble.

The ODMAP protocol has been applied for reporting the models and is available in supplementary material (Zurell *et al.*, 2020). The complete analytical workflow from data preparation and community assemblage delineation to final community assemblage distribution models is displayed on supplementary fig. DS3 [IN PREP] and associated scripts is available on GitHub (<https://github.com/ColineRoyaux/DM>).

Results

Identification and community assemblage delineation

In total, 24 species of 13 genera were identified in the 70 samples (tab. 2.1). The highest species richness is found in HYNC4002 with 6 species detected and most locations have a species richness of 1 (n=32) or 2 (n=17).

Table 2.1 - Species identified in sorted samples, their occurrence count, assemblage associated with them using the network clustering algorithm and the assemblages associated with the samples in which they were observed.

Class	Family	Species	N presences	Associated assemblage	Observed in samples from	Known distribution
Branchiopoda	Streptocephalidae	<i>Streptocephalus archerii</i>	8	Assemb. 1	Assemb. 1	Australia
	Triopsidae	<i>Triops longicaudatus intermedium</i>	1	Assemb. 1	Assemb. 1	Australia; New Caledonia
	Daphniidae	<i>Simocephalus exspinosus australiensis</i>	4	Assemb. 1	Assemb. 1 & 2	Australia; New Caledonia
		<i>Simocephalus acutirostratus</i>	3	Assemb. 1	Assemb. 1	Southern hemisphere
		<i>Ceriodaphnia rigaudi</i>	10	Assemb. 1	Assemb. 1	Southern hemisphere
		<i>Daphnia cephalata</i>	3	Assemb. 1	Assemb. 1	Indopacific area
		<i>Daphnia longicephala</i>	2	Assemb. 1	Assemb. 1	Australia; New Caledonia
	Macrothricidae	<i>Echinisca sp. A</i>	5	Assemb. 2	Assemb. 2	New Caledonia
		<i>Echinisca triserialis</i>	1	Assemb. 1	Assemb. 1	Southern hemisphere
		<i>Macrothrix spinosa</i>	10	Assemb. 1	Assemb. 1	Southern hemisphere
Moinidae	<i>Moina micrura</i>	6	Assemb. 1	Assemb. 1	Global	
Sididae	<i>Diaphanosoma unguiculatum</i>	1	Assemb. 1	Assemb. 1	Australia; New Caledonia	
	<i>Diaphanosoma australiensis</i>	12	Assemb. 1	Assemb. 1 & 2	Australia; New Caledonia	
	<i>Latonopsis brehmi dedekkeri</i>	13	Assemb. 2	Assemb. 2	New Caledonia	
	<i>Latonopsis australis</i>	4	Assemb. 1	Assemb. 1	Indopacific area	
Lynceidae	<i>Lynceus insularis</i>	5	Assemb. 2	Assemb. 2	New Caledonia	
	<i>Lynceus sp. A</i>	10	Assemb. 1	Assemb. 1 & 2	Australia; New Caledonia	
	<i>Lynceus sp. B</i>	2	Assemb. 3	Assemb. 3	New Caledonia	
Limnadiidae	<i>Eulimnadia sp. A</i>	8	Assemb. 1	Assemb. 1	New Caledonia	
Maxillopoda (Copepoda)	Centropagidae	<i>Boeckella spinogibba</i>	19	Assemb. 2	Assemb. 2	New Caledonia
		<i>Boeckella sibletii</i>	2	Assemb. 2	Assemb. 2	New Caledonia
		<i>Boeckella sp. A</i>	1	Assemb. 4	Assemb. 4	New Caledonia
		<i>Boeckella sp. B</i>	1	Assemb. 2	Assemb. 2	New Caledonia
		<i>Boeckella sp. C</i>	2	Assemb. 3	Assemb. 3	New Caledonia

The bipartite species-sample network had 94 nodes (24 species and 70 samples) and 132 links. The infomap algorithm determined two hierarchical levels of clustering in the network. The first level distinguished the network in four clusters, each cluster is considered as a community assemblage in the article (fig. 2.3). The second level distinguished the network in 15 clusters which means an average species richness of less than 2 for each cluster. The delineation at the second level cannot be considered as a community assemblage delineation and is not further investigated.

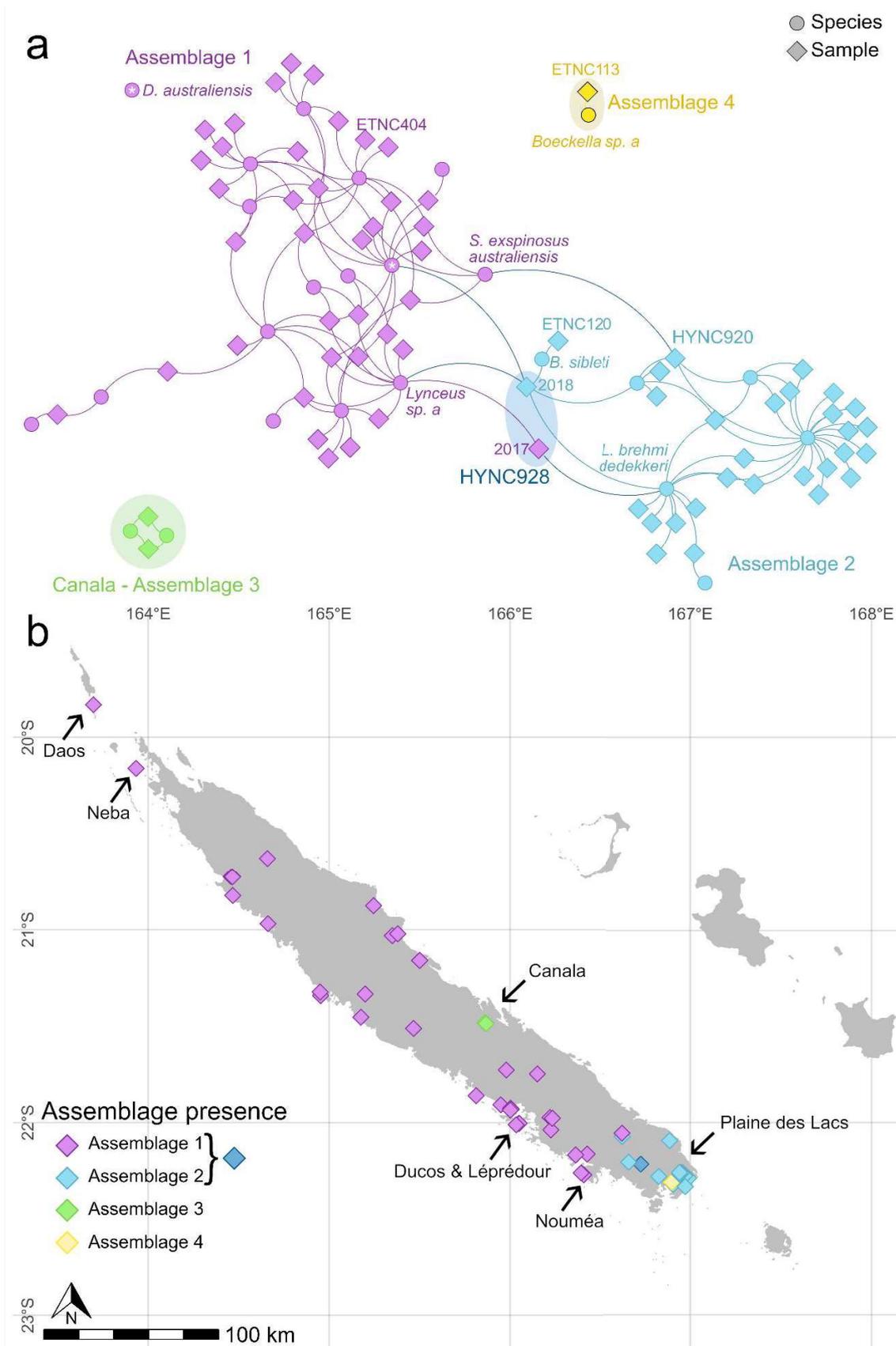


Figure 2.3 – Species-sample network colored with each assemblage delineated by the clustering algorithm at the first level. Intermediate location HYN928 is highlighted in dark blue. Isolated assemblages 3 and 4 are highlighted in green and yellow respectively (a). Presence map of each assemblage delineated by the clustering algorithm (b).

At the first level of clustering, two major clusters are delineated (assemblage 1 and 2 on fig. 2.3 and tab. 2.1). Assemblage 1 is composed of 15 species in 38 samples and occur in most of Grande Terre, Ducos, Léprédour, Neba and Daos islands. Assemblage 2 is composed of 6 species in 29 samples and occur in the southernmost area of Grande Terre, in and around the Plaine des Lacs. The two community assemblages are mostly separated geographically (fig. 2.3b). However, the HYNC928 location sampled in 2017 and 2018, highlighted by a dark blue oval on fig. 2.3a, appears to be intermediate as species from both assemblages are observed in equilibrated proportions with one species of each assemblage in 2017 and two species of each assemblage in 2018. The sample from 2018 contains a fifth species, *Boeckella sibleti*, that is associated with assemblage 2. However, this species is observed only from one other sample, ETNC120 sampled in 2020. ETNC120 is also associated with assemblage 2 and *B. sibleti* is the only species observed in the sample. Consequently, the affiliation of *B. sibleti* with assemblage 1 or 2 seems ambiguous which implies the affiliation of ETNC120 to assemblage 2 is ambiguous as well. Additionally, the sampled location closest to ETNC120, ETNC404 sampled in 2023, has been associated with assemblage 1 (less than 3 km distance, located around coordinates 22°S, 166.6°E).

The two community assemblages are related by four links between the 2018 sample of HYNC928 and species *Diaphanosoma australiensis* and *Lynceus sp. A*; between the 2017 sample of HYNC928 and species *Latonopsis brehmi dedekkeri*; and between HYNC920 and species *Simocephalus exspinosus australiensis*. The three species are the only ones found in samples affiliated with several assemblages and have all been associated with assemblage 1 by the clustering algorithm (tab. 2.1).

Assemblage 3, highlighted in green on fig. 2.3a, is composed of two species (*Lynceus sp. B* and *Boeckella sp. C*) in two samples and assemblage 4, highlighted in yellow on fig. 2.3a, of one species (*Boeckella sp. A*) in one sample. Despite being relatively geographically close to other locations (fig. 2.3b), both assemblages are completely isolated and disconnected from any other part of the network (assemblage 3 and 4 on fig. 2.3). As for assemblage 2, assemblages 3 and 4 are located on ultramafic massifs.

Sampling completeness and diversity estimates

Sampling completeness and diversity estimates are meaningful when sufficient samples are provided. As assemblages 3 and 4 are based respectively on only two and one samples, the estimators were computed only for assemblages 1 and 2. For both community assemblages, sampling completeness increased with diversity order, highlighting undetected diversity especially for rare species (tab. 2.2). However, most frequent species were seemingly detected in assemblages 1 and 2 (fig. 2.4).

Hill numbers estimated a higher diversity in assemblage 1 for each diversity order. However, assemblage 2 appears to have a higher proportion of rare species when comparing estimates at $q = 0$ with $q = 1$ and $q = 2$ (63.8-75.3% of estimated infrequent species for assemblage 2 and 40.9-49.9% for assemblage 1; tab. 2.2).

Table 2.2 - Sampling completeness and Hill numbers diversity estimates at three q orders.

Assemblage	Diversity order	Sampling Completeness	Observed diversity	Estimated diversity	Undetected species	% of undetected species
1	$q = 0$; all species	67%	$D_0 = 16$	23.79	7.79	32.7%
	$q = 1$; frequent species	95%	$D_1 = 12.24$	$\hat{D}_1 = 14.04$	1.8	12.8%
	$q = 2$; super-frequent species	100%	$D_2 = 10.63$	$\hat{D}_2 = 11.92$	1.29	10.8%
2	$q = 0$; all species	54%	$D_0 = 9$	$\hat{D}_0 = 16.72$	7.72	46.2%
	$q = 1$; frequent species	91%	$D_1 = 5.25$	$\hat{D}_1 = 6.22$	0.96	15.5%
	$q = 2$; super-frequent species	100%	$D_2 = 3.92$	$\hat{D}_2 = 4.14$	0.22	5.3%

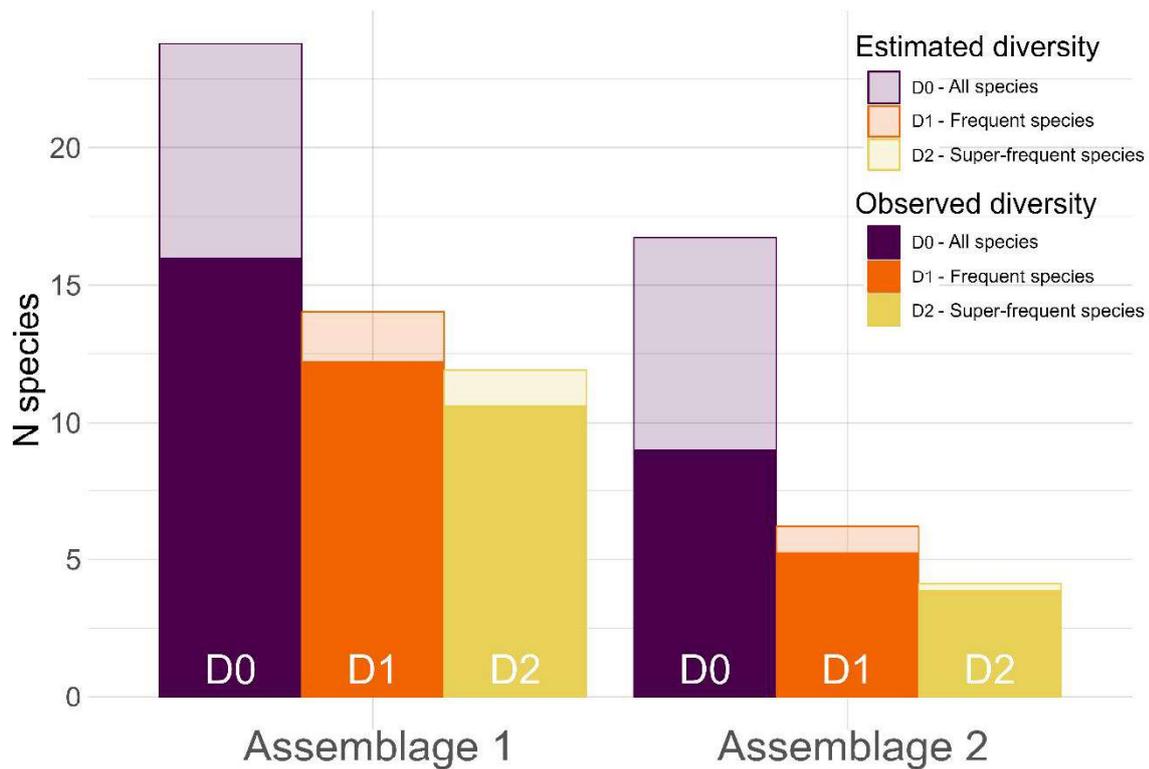


Figure 2.4 – Observed and estimated diversity at three q orders.

Community assemblage distribution modelling

The full model computed on all four variables (ultramafic massifs, annual precipitations, sampling density and distance to freshwater) determined low variable importance for sampling density and distance to freshwater. Hence, two other models were computed, one with two variables, ultramafic massifs, and annual precipitations and one with three variables, sampling bias, ultramafic massifs, and annual precipitations. This last model with three variables showed the highest evaluation metrics and had the most consensual responses curves (fig. 2.5).

According to this last model, assemblage 1 is more likely to be observed in the absence of ultramafic massifs and at lower annual precipitation rates (highest probability between 1000 and 1500 kg/m²/y), little variations are observed on response curves for sampling density (fig. 2.5a). Assemblage 2 is more likely to be observed in the presence of ultramafic soils and at higher annual precipitation rates (highest probability between 2250 and 3340 kg/m²/y), this assemblage responded with higher observation probability around 0.4 kernel density from our sampling (fig. 2.5b).

Assemblage 1 appears to be the most probable community assemblage across the majority of the archipelago, except for the southernmost region of Grande Terre, where assemblage 2 is observed with a higher probability (fig. 2.5c). Figure 2.5c depicts the most probable assemblage in terms of colour overlaid by its presence probability represented in black of varying opacity. The lower the probability of the most probable assemblage, the higher the opacity of the black. In other words, the less it is possible to identify which community assemblage is most probable on the map, the less it is probable to actually observe this assemblage in the area. Indeed, areas with the lowest probabilities are situated predominantly at frontiers between both assemblages. Additionally, the majority of the area where assemblage 2 is most probable has a relatively low probability.

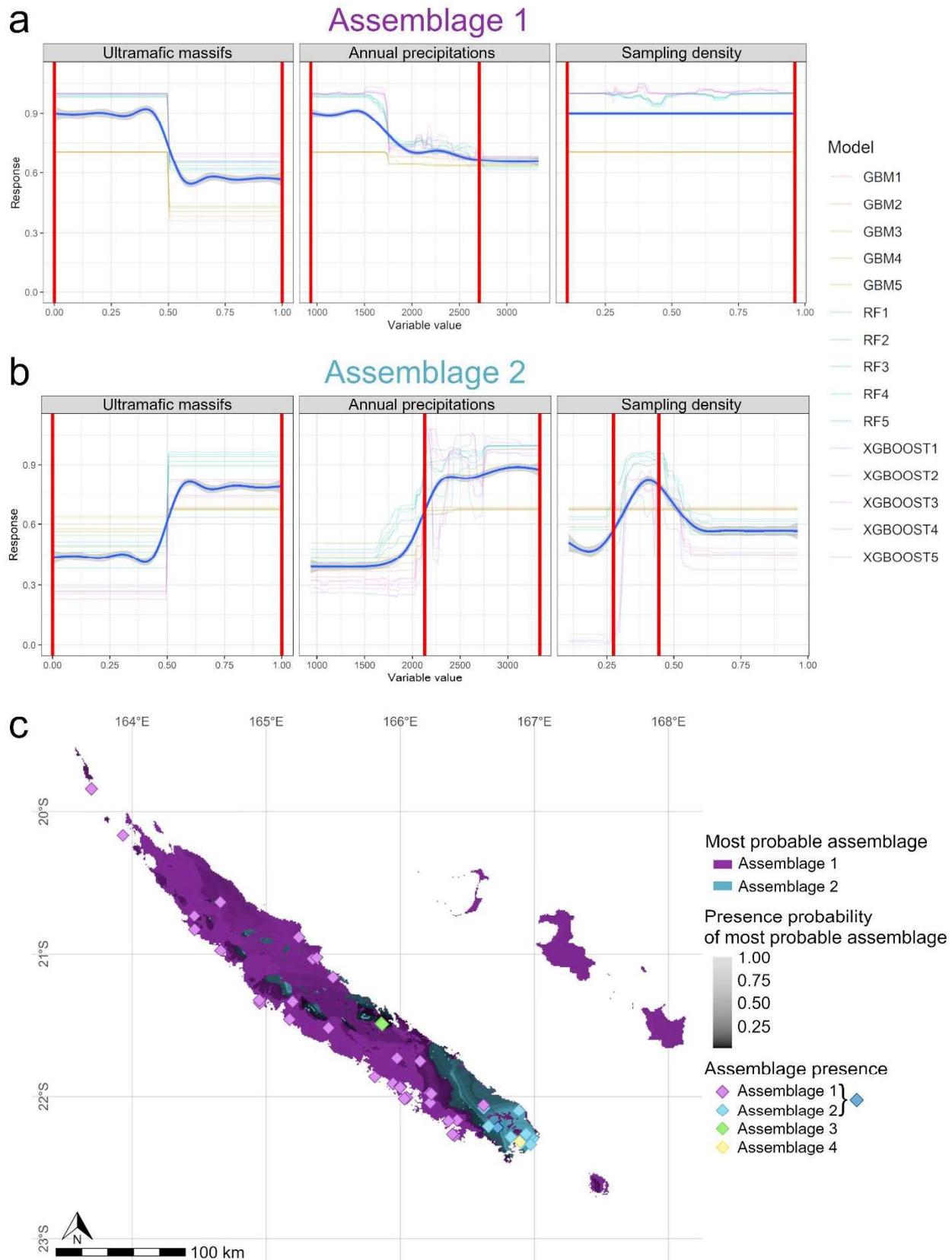


Figure 2.5 - Models' response plots of assemblages 1 and 2 for each selected variable. Variable ranges for observed presences of community assemblages is represented by vertical red lines (a, b). Binary map of most probable assemblage at each pixel (in purple and blue) overlaid by the mean presence probability of the most probable assemblage in black of varying opacity (c).

Discussion

The delineated communities and computed models demonstrated that the distribution of lentic freshwater crustacean community assemblages seems to be primarily driven by the presence or absence of ultramafic soils and, to a lesser extent, by annual precipitation. Furthermore, the incorporation of sampling bias from our study permitted to correct the response inconsistencies between models.

Effects of ultramafic massifs and precipitations

Links between ultramafic soils and the floral, fungal and microbial diversity in New Caledonia has been extensively studied (Isnard *et al.*, 2016; Gourmelon *et al.*, 2016; Pillon *et al.*, 2021; Amir *et al.*, 2023; Isnard & Jaffré, 2024). For example, a difference in pollinator composition has been demonstrated between ultramafic and non-ultramafic environments of the archipelago (Zakardjian *et al.*, 2023). Numerous faunal taxa on the archipelago are known to be exclusively observed from ultramafic soils (Keith, 2002; Berman & Andersen, 2012; Nattier *et al.*, 2013; Hudel *et al.*, 2020; Hrivniak *et al.*, 2023). However, no formal demonstration of this phenomenon using analytical tools has been made prior to this study.

Adaptations of plants to the high heavy metal concentrations in ultramafic soils are well known (*e.g.* mycorrhiza, accumulation organs). In lentic ecosystems, the effects of heavy metal are studied but mostly as a stressor (Faghihinia *et al.*, 2021). In New Caledonia, the concentration of heavy metals is not the result of anthropogenic pollution but is naturally present in the environment since over 25 Mya (Chevillotte *et al.*, 2006; Maurizot & Campbell, 2020). All species associated with assemblage 2 are, to present knowledge, endemic to New Caledonia while only one species associated with assemblage 1 is endemic to the archipelago. Higher proportion of endemic species is also found in plants, pollinators and grasshoppers on the archipelago (Nattier *et al.*, 2013; Isnard *et al.*, 2016; Zakardjian *et al.*, 2023; Isnard & Jaffré, 2024). The difference in heavy metal concentration itself can suffice as a factor for adaptative speciation that could explain the distinction between the two assemblages. However, more indirect effects could also explain the observed differences between ultramafic and non-ultramafic environments. Indeed, ultramafic and non-ultramafic environments are very different. Ultramafic environments appear to have lower fertility rates, lower species richness and higher habitat heterogeneity (Isnard *et al.*, 2016; Pillon *et al.*, 2021). Similarly, lentic environments on ultramafic soils show low primary productivity and low

species richness (Bargier *et al.*, 2018) which could notably lower the frequency of terrestrial and aerial macrofauna visits and consequently dispersion opportunities (Herteux *et al.*, 2019).

The appearing preference of assemblage 1 for lower precipitation rates could indicate species functioning in ponds with shorter habitat duration. On average without distinction between shared socioeconomic pathways and general circulation model, higher precipitation rates are predicted on the archipelago at both 2041-2070 and 2071-2100 horizons (CHELSA). It could indicate a wider spread of assemblage 2 in future decades. However, as assemblage 2 seems to be ultramafic obligate, a new potentially non-native assemblage might colonise the archipelago as well.

Diversity within assemblages

As with observed diversity, estimated total diversity is higher in assemblage 1 but there is a higher proportion of rare species in assemblage 2 that is exclusively located on ultramafic massifs (tab. 2.1). Regarding assemblage 1, it is notable that, except one potential endemic species, all species associated with it are also reported in Australia (Smirnov & Timms, 1983; Timms, 2015). In this context, a comparable assemblage could be identified in Australia which would deepen the knowledge on these assemblages and the links between Australian and New Caledonian biodiversity. Should a comparable assemblage be observed in Australia, it may be of interest to investigate the diversity levels there and ascertain whether the island biogeography assumption that richness is lower on isolated islands is verified within such assemblages (Whittaker *et al.*, 2023).

Only three of the 24 total species were found in samples from both assemblages which implies these assemblages exhibit notable divergence from one another. Additionally, the two assemblages are geographically well separated in the archipelago. Assemblage 1 comprises more than half of the network nodes and is distributed over almost two-thirds of New Caledonia while assemblage 2 is distributed over a significantly smaller area. In community assemblages 1 and 2, the sampling appears to be incomplete especially for rare species. Nevertheless, frequent species are adequately sampled which is the most crucial to delineate community assemblages.

Several hypotheses can be drawn from the observation that only species classified as belonging to assemblage 1 were sometimes observed in samples from assemblage 2 and not the other way around. Assemblage 2 could comprise only specialists to ultramafic environments, which would preclude them from successfully colonising locations in non-ultramafic environments as observed for most ultramafic plant species in New Caledonia (Isnard *et al.*, 2016). The species observed in both assemblages would be generalists. However, these species seem to be anecdotal in samples classified in assemblage 2, which is not consistent for truly generalist species. Furthermore, the models predict a low presence probability for both assemblages at the borders between these assemblages. Considering the ambiguous belonging of notably location HYNC928 and species *B. sibleti* to assemblage 1 or 2, a buffer area or environmental gradient may exist where the species from the two community assemblages can coexist in a limited extent. Such gradient between ultramafic and non-ultramafic environment has been posited for plant assemblages (Enright *et al.*, 2001). Indeed, on ultramafic massifs of the archipelago, three major vegetal formations are found; low-altitude maquis (< 1000 m), high-altitude maquis (> 1000 m), and dense forests (Jaffre, 1996). Low-altitude maquis would be related with most of locations associated with assemblage 2. Only one pond has been sampled in high-altitude maquis in Pic Ningua (ETNC401) where a single species has been identified, *Moina micrura*, this pond has been classified in assemblage 2. Interestingly, the intermediate location HYNC928 (tab. 2.1, fig. 2.5) and ETNC120 are both found in ultramafic dense forest assemblages, and both are located on areas where the models predict low presence probabilities for the two assemblages.

Despite the impossibility to model their potential distribution, assemblages 3 and 4 are both present on ultramafic massifs which could indicate that they represent inadequately sampled sub-assemblages of assemblage 2. Ultramafic massifs show high complexity and heterogeneity (Chevillotte *et al.*, 2006; Nattier *et al.*, 2013; Isnard *et al.*, 2016) which could explain sub-divisions of the assemblage in these environments. Regarding assemblage 4, such hypothesis seems probable as it is located on an area where assemblage 2 is highly probable. However, regarding assemblage 3, it is notable that the presence probability of assemblages 1 and 2 are both low in the area. Indeed, locations associated with assemblage 3 are present on low-precipitations ultramafic massifs which are differing with conditions apparently preferred by assemblages 1 and 2.

Conservation

Ponds disappear at accelerating rates which is sufficient to be alarming (Rhazi *et al.*, 2012). Additionally, knowledge about their biota is scarce in tropical environments such as New Caledonia which could indicate an even worse situation in these areas as demonstrated for tropical vascular plants (Junk, 2002; Vamosi & Vamosi, 2008; Faghihinia *et al.*, 2021).

Among the families included in this analysis (*i.e.*, Streptocephalidae, Triopsidae, Daphniidae, Macrothricidae, Moinidae, Sididae, Lynceidae, Limnadidae, Centropagidae), only two species signalled in New Caledonia have not been sampled in this study, the Centropagidae *Dussartopages fagesi* and, the Daphniidae *Daphnia carinata mirabilis*. However, at least seven undetected species were estimated in each assemblage by the interpolation and extrapolation method. This number is indicative and insufficient to conclude for the exact number of unknown species on the archipelago. However, it is clear that most of the global biodiversity is unknown and New Caledonia should be no exception (Macedo *et al.*, 2024).

Future predictions indicate wetter conditions that are potentially detrimental for assemblage 1. However, it is widespread on the archipelago and might also be present in Australia. Large distribution is an advantaging factor regarding extinction risks (McKinney, 1997). Assemblage 2 seems to face higher extinction risks with its lower species richness and narrower observed distribution. Additionally, the apparent preference of assemblage 2 for ultramafic massifs makes it face more immediate threat that is mining activity (Pascal *et al.*, 2008).

In this study, known species are associated with one or several community assemblages with distinctive features. On the field, the identification of several species from a particular assemblage could now be used to infer threats faced by the studied environment and apply adapted and effective conservation strategies. Models based on similar methods as the ones presented in this article have demonstrated their impact on conservation policies and applications on the field (Guisan *et al.*, 2013).

Limitations and perspectives

The sampling is a major limitation in this study for several reasons. The sampling has taken place during seven years with an interruption in 2019. The first three years of sampling were exploratory as the location of many ponds were unknown.

Despite methodical identification of interest locations on satellite images, many ponds of small size are undetectable with this method and were sampled opportunistically. Additionally, some locations have been visited several times but the vast majority has only one sample which means the analysis is made between ponds sampled in different years, thus different environmental conditions. Unfortunately, many of the sampled ponds are difficult to access and, if temporary, might be dry most of the year which makes sampling unpredictable to some extent.

The lack of knowledge or the complexity of identification of several crustacean species living in the studied environment led the community assemblages to be inferred from a largely filtered set of taxonomic groups among crustaceans. Indeed, the family Chydoridae seemed to be the most diverse in our samples and a global taxonomic clarification of this group would undoubtedly be of major interest. Such clarifications have been made in recent years for the New Caledonian Ostracod species and adding species from this class to the analysis would be a future amelioration to make (Martens *et al.*, 2019).

As highlighted in the introduction, temporary lentic freshwaters are dynamic systems. Indeed, depending on the length and periodicity of the hydroperiod, assemblages in these environments will be very different. All species found in these environments show differing life spans, drought survival strategies, have different egg hatching time after flooding and attain maturity at different times. Precipitations have undoubtedly a major impact on hydroperiod. However, many other factors such as soil granulometry, surrounding vegetation, connexion to phreatic zones, air humidity and sun exposure affects the hydroperiod from local to regional scales. Indeed, water seasonality (European Commission, <https://global-surface-water.appspot.com>; Pekel *et al.*, 2016) is close to a measure of hydroperiod and shows low correlation with precipitation ranges (see supplementary tab. DS2). Unfortunately, when investigating for the integration of water seasonality in the models, the resolution was too coarse to detect the studied ponds and most sampled locations were considered as never flooded in the data. Consequently, models presented in this study seemingly lack one of the most important factor structuring communities that is hydroperiod. Knowing the exact hydroperiod of a pond requires a daily surveillance of the location through direct observation or field sensors which would be a major enhancement in the presented models.

The incorporation of sampling density as an independent variable in the models is not the most satisfying way of dealing with sampling bias. Indeed, the

sampling density from this study drives to some extent the displayed projection of the potential distribution of each assemblage. However, as the variable importance computed by the models for sampling density was the lowest for both assemblages, its impact on projections seemed reasonably low. Additionally, the projected potential distribution of each assemblages seemed biologically coherent with our observations and with what has been posited for other taxa on the archipelago. Nevertheless, it would be of interest to try other bias correction techniques such as environmental filtering or the draw of background points to compare model performance and projections (Varela *et al.*, 2014; Barber *et al.*, 2022; Leroy, 2023).

From location similarity to statistical approach, many methods other than biogeographical network exist for the delineation of community assemblages (Kreft & Jetz, 2010; Woolley *et al.*, 2020). Comparing the outputs of such methods with results from this study would certainly add depth to the conclusions made in this article. The choice of using a biogeographical network has been made for its standardized framework requiring few methodological choices (Vilhena & Antonelli, 2015).

Similarly, there are many methods for community distribution models. Ensemble models as used in this study have been deeply reviewed in the last decade and many methodological choices are required (Franklin & Miller, 2010; Araújo *et al.*, 2019; Valavi *et al.*, 2021). Many of these choices are hard to standardise and no consensus on best decision practice exist for the large majority of them (Feng *et al.*, 2019a; Leroy, 2023). For instance, regarding community distribution models, part of the literature recommends modelling the potential distribution of each species individually and construct Stacked or Joint Species Distribution Models (Guisan & Rahbek, 2011; Warton *et al.*, 2015). However, efficiently applying such models requires a deep knowledge of each species of the assemblage and satisfying calibration of each species model implying a multiplication of hard-to-standardise methodological choices. Additionally, in the case of the environments studied in this article, the high proportion of endemic species and their restricted observed area implied the impossibility to model the potential distribution of most species.

Encadré B : Modéliser la distribution des espèces de crustacés d'eau douce

Travail en cours

Contexte

De la même façon que pour les assemblages d'espèces, il a été essayé de modéliser la distribution des espèces de crustacés d'eau douce. Or, malgré le tri complet de 70 échantillons prélevés au filet, la majorité des espèces étudiées sont retrouvées dans moins de 4 localités. Un nombre d'occurrences insuffisant pour assurer une modélisation suffisamment représentative.

Cependant, les échantillons de sédiment permettent de récupérer les œufs de résistance des espèces étudiées et de les identifier. Au-delà de communiquer une information sur la présence d'une espèce donnée, l'absence d'œufs d'une espèce permet de déterminer que celle-ci est absente avec une certitude raisonnable. De ce fait, la présence et l'absence des espèces sont caractérisées ici par l'utilisation du tri des sédiments prélevés dans les mares. Contrairement à ce qui est constaté pour les échantillons au filet, les échantillons de sédiment permettent d'obtenir un nombre de présence suffisamment élevé pour pouvoir modéliser la répartition des espèces.

Matériel et méthodes

Cinq espèces sont identifiables à l'aide de leurs œufs de résistance, *Streptocephalus archeri*, *Eulimnadia sp. A*, *Branchinella sp.*, *Australimnadia sp.*, et *Triops intermedius* (fig. 2.6).

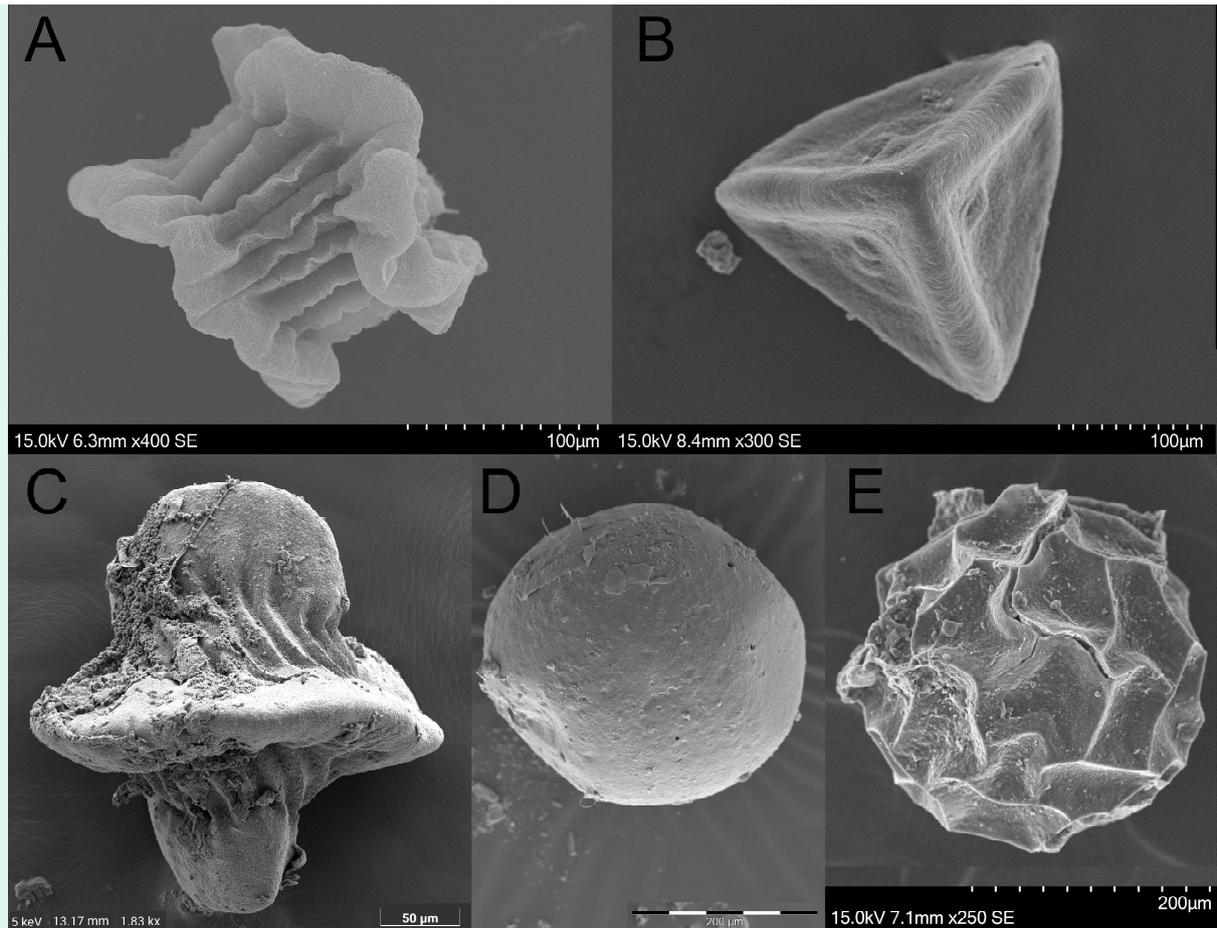


Figure 2.6 – Images MEB des œufs de résistance. A, *Eulimnadia* sp. A. B, *Streptocephalus archeri*. C, *Australimnadia* sp. D, *Triops longicaudatus* de Saint Barthélemy (très similaire à *T. intermdius*). E, *Branchinella* sp.

Au total, 368 échantillons de sédiments de 351 localités ont été triés sur Grande Terre (fig. 2.7).

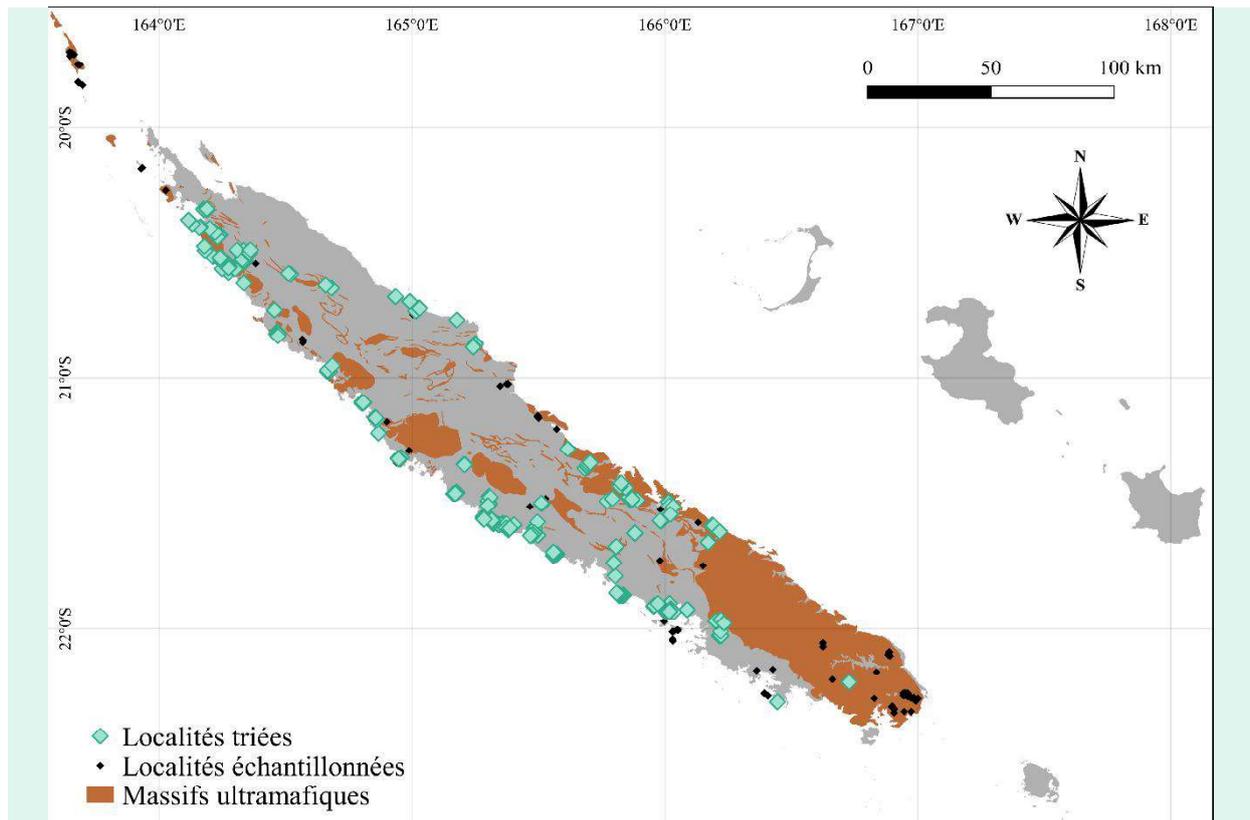


Figure 2.7 – Localités dont le sédiment a été trié.

Les variables utilisées pour modéliser la probabilité de présence de ces espèces sont la quantité de précipitation annuelle, la distance à l'eau douce et la présence de massifs ultramafiques tels que dans l'article présenté ci-dessus (fig. 2.2). Les modèles utilisés sont identiques à l'article ci-dessus.

Résultats préliminaires et interprétation

En comparaison avec les données issues des échantillons au filet, le tri des échantillons de sédiment permet d'étendre l'aire de répartition des espèces étudiées.

Les œufs de résistance de l'espèce *Eulimnadia sp. A* ont permis de déterminer la présence de l'espèce dans 128 localités (218 absences ; fig. 2.8).

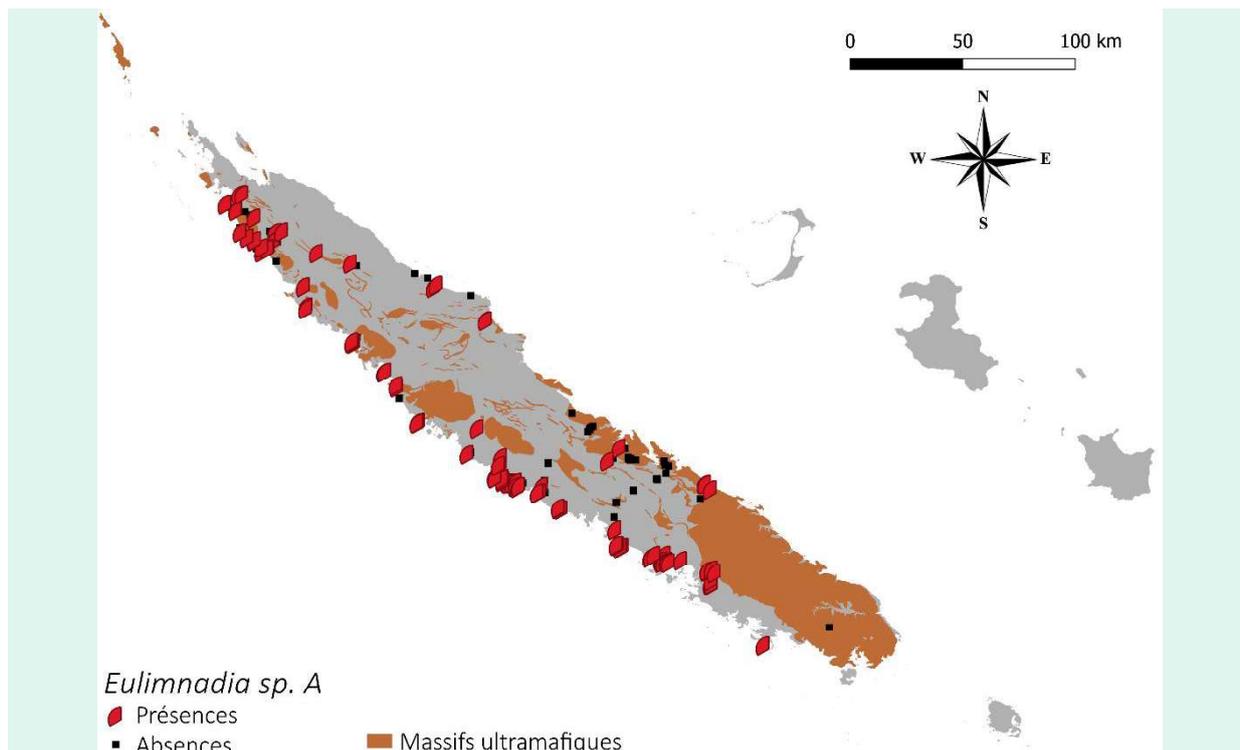


Figure 2.8 – Présences et absences de *Eulimnadia sp. A* à partir des détections d'œufs de résistance dans les sédiments.

Les œufs de résistance de l'espèce *Streptocephalus archeri* ont permis de déterminer la présence de l'espèce dans 100 localités (245 absences ; fig. 2.9).

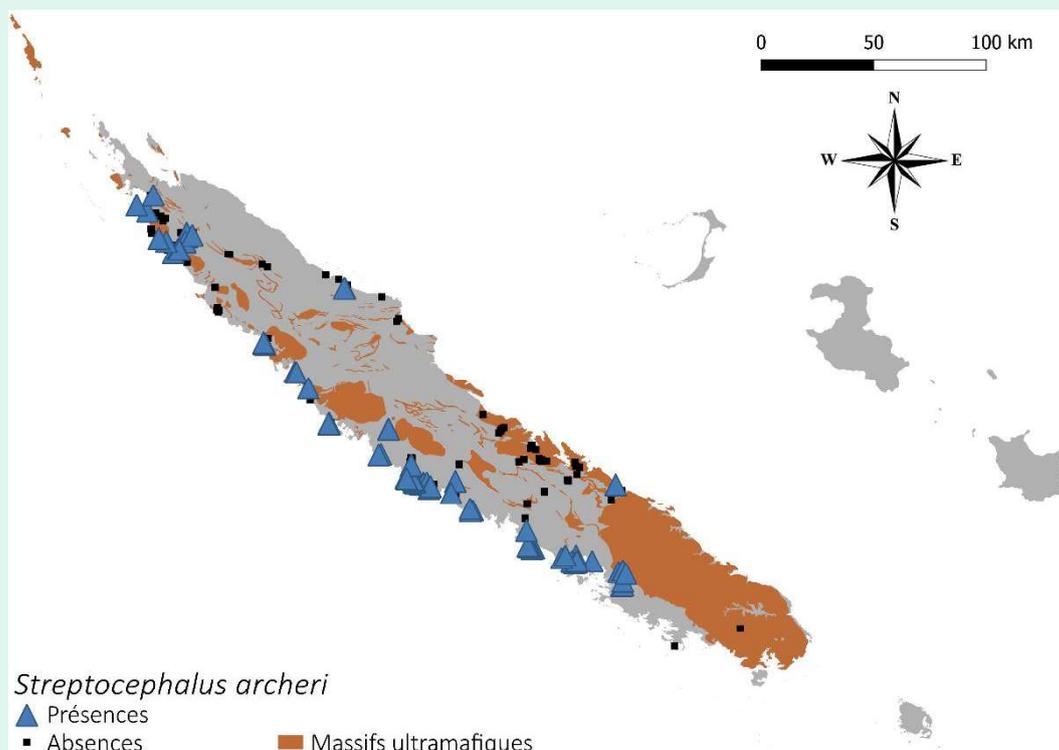


Figure 2.9 – Présences et absences de *Streptocephalus archeri* à partir des détections d'œufs de résistance dans les sédiments.

Les œufs de résistance de *Branchinella sp.* et *Australimnadia sp.* ont chacun été trouvé dans une seule localité. Les œufs de résistance de l'espèce *Triops intermedius* ont permis de déterminer la présence de l'espèce dans 6 localités (337 absences ; fig. 2.10). Pour cette espèce, les 6 présences ne sont pas suffisantes pour utiliser un modèle. En effet, ces 6 points de présences sont très proches et sont donc sujets à des conditions environnementales très similaires voire identiques. Les modèles vont donc avoir peu de valeurs différentes par variable environnementale sur lesquelles se baser pour déterminer la probabilité de présence de l'espèce sur l'archipel. De ce fait, les modèles vont déterminer qu'il y a une haute probabilité de présence de l'espèce sur les valeurs de variables environnementales de ces points et une faible probabilité sur toutes les autres valeurs. Ce phénomène est appelé le surapprentissage ou surajustement (over-fitting). Ce surajustement existera quel que soit le type de variable environnementale prise en compte (*e.g.*, pluviométrie, altitude, couleur des maisons environnantes) et les modèles ne pourront pas déterminer quelles sont les variables les plus importantes pour l'espèce. En définitive, il sera donc impossible de déterminer si le modèle est valide ou non.

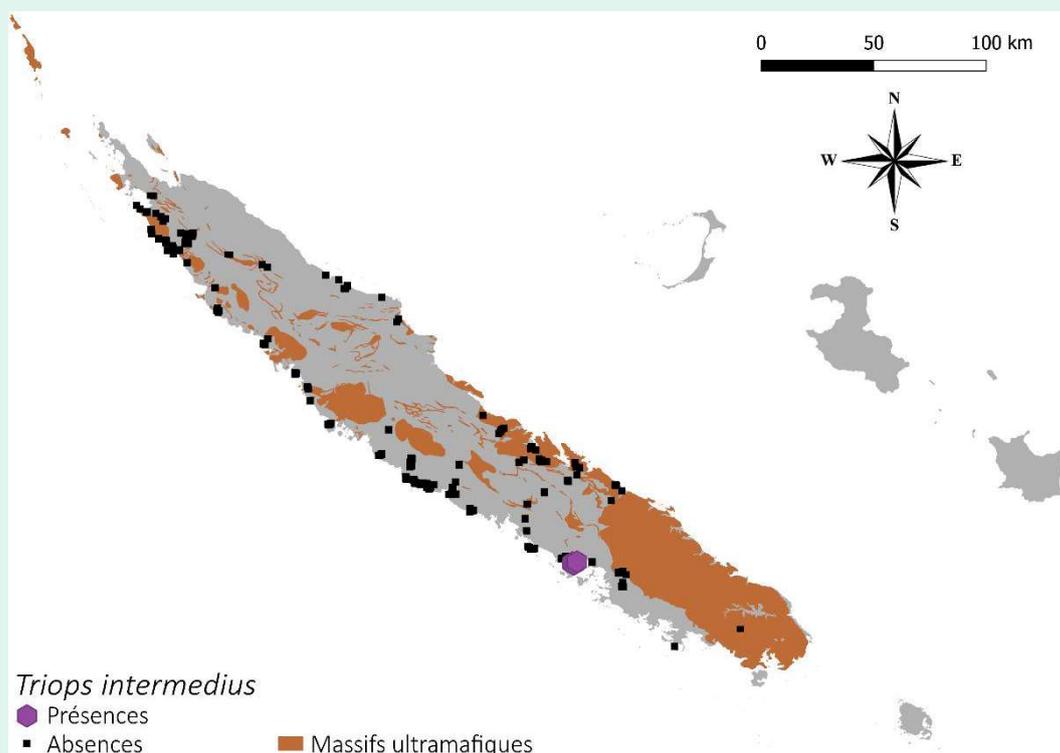


Figure 2.10 – Présences et absences de *Triops intermedius* à partir des détections d'œufs de résistance dans les sédiments.

Afin de modéliser les probabilités de présence des espèces *E. sp. A* et *S. archeri*, des essais ont été faits avec diverses paramétrisations (*e.g.*, variables

centrées – réduites, sous-échantillonnage, filtration géographique). Or, les évaluations des modèles n'ont jamais été améliorées par les différentes paramétrisations quelle que soit le type de validation croisée utilisée (TSS < 0.6).

Les courbes de réponses obtenues pour les deux espèces sont très différentes en fonction des modèles (e.g., courbes variants entre une réponse de 0 à 1 et indissociables entre les niveaux de précipitations 1000 et 1500 mm/an chez les deux espèces ; fig. 2.11). Cette absence de consensus entre les modèles permet de confirmer que le modèle n'est pas valide.

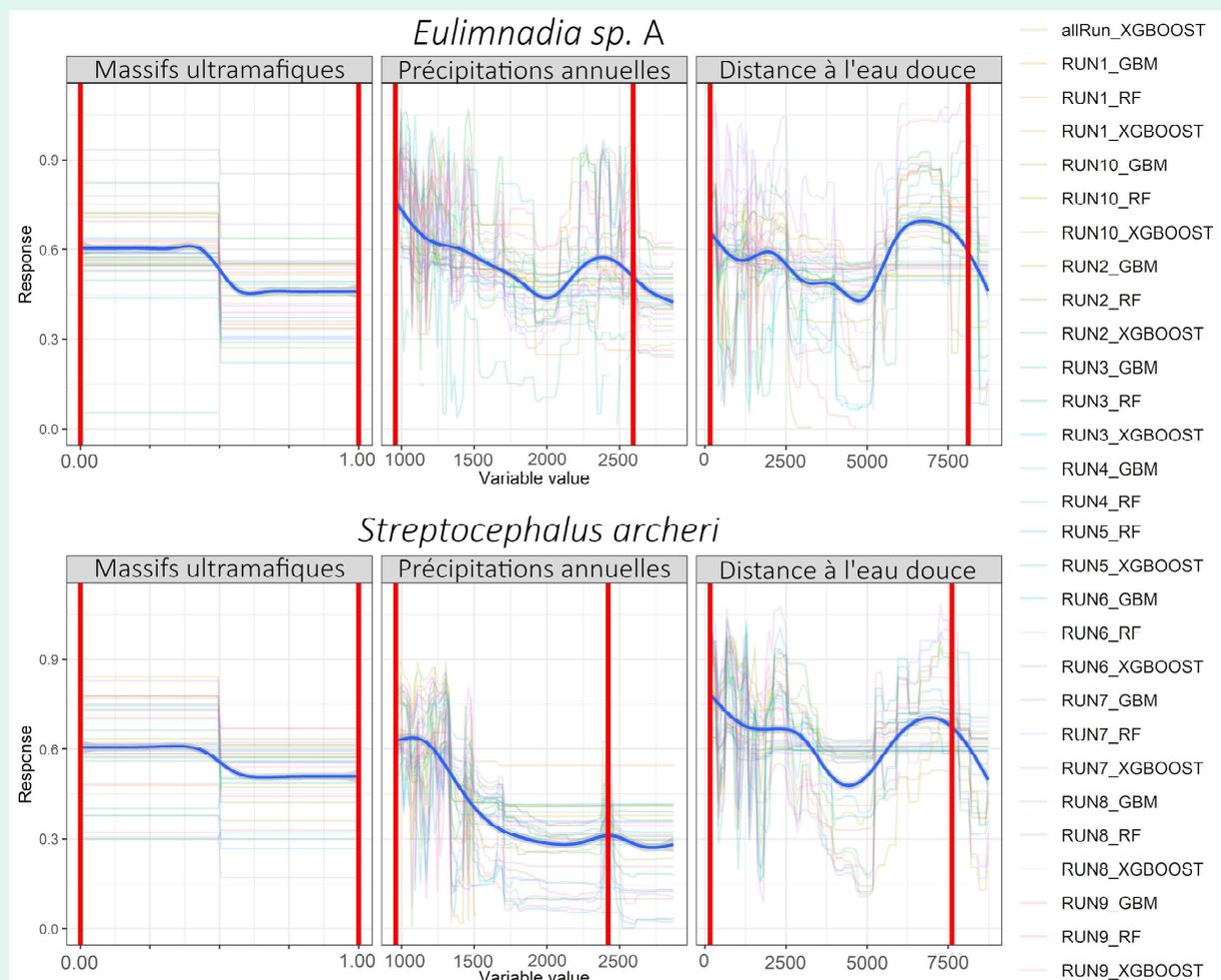


Figure 2.11 – Courbes de réponses de chaque variable environnementale calculées par les modèles pour *Eulimnadia sp. A* et *Streptocephalus archeri*.

Par extension, les projections des probabilités de présence des espèces calculées par les modèles sont tout aussi incohérentes biologiquement. En effet, il semble peu probable qu'une espèce soit distribuée selon une succession de bandes de faible et haute probabilité (fig. 2.12).

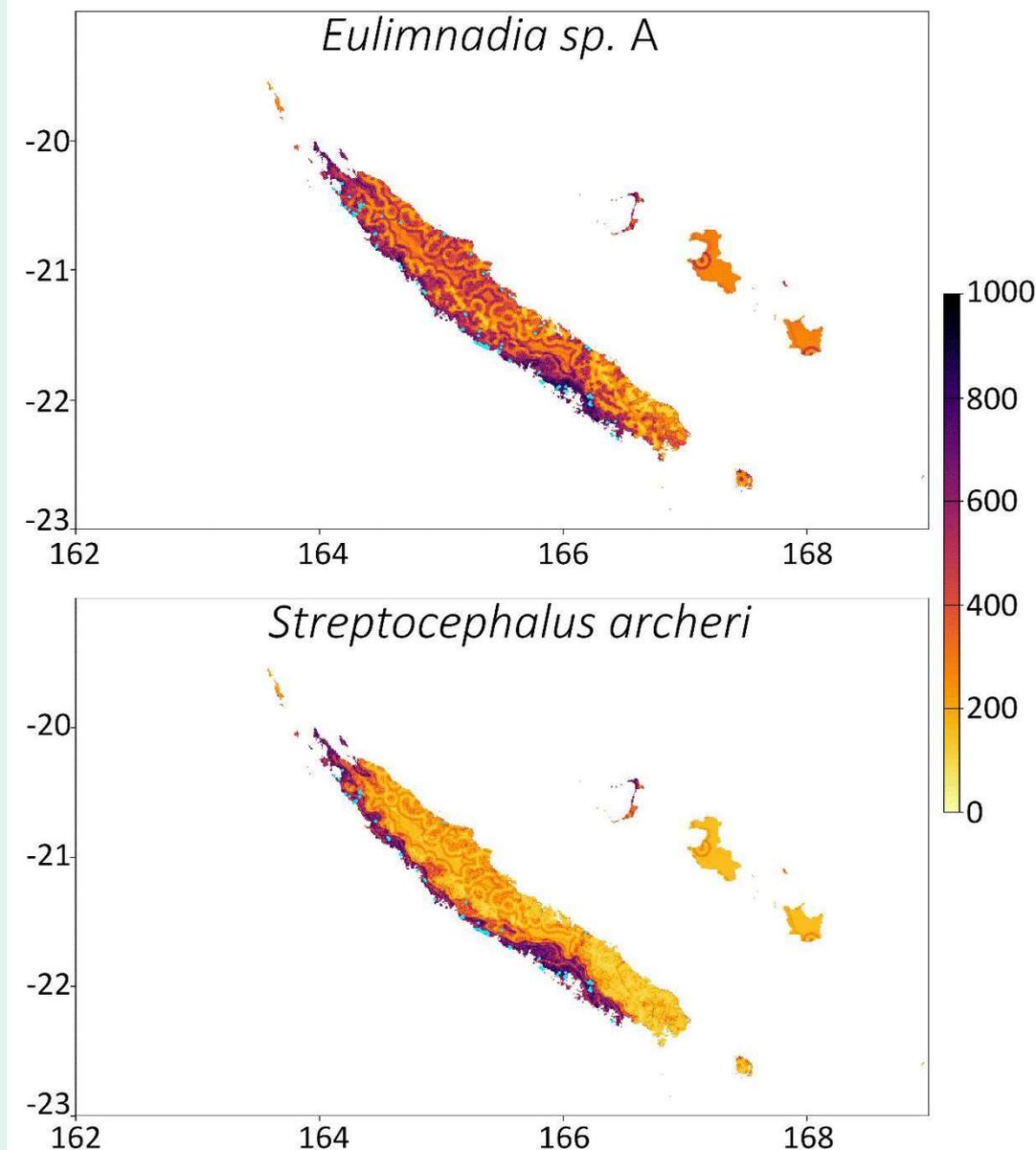


Figure 2.12 – Cartes de probabilité de présence calculées par les modèles des espèces *Eulimnadia sp. A* et *Streptocephalus archeri*.

Perspectives

De la même façon que pour les assemblages d'espèce, il est probable que l'hydropériode soit une variable particulièrement déterminante pour la distribution de ces espèces. L'absence de ce facteur pourrait être une des raisons qui expliquent la mauvaise qualité des modèles calculés.

La méthode de modélisation utilisée dans ce chapitre (connue sous les noms « modèles d'ensembles », « modèles de distribution » ou encore « modèles de niche ») est très largement utilisée et commentée dans la littérature depuis 20 ans (Guisan & Thuiller, 2005; Guisan *et al.*, 2013; Pollock *et al.*, 2020; Valavi *et al.*,

2021). Cette méthode demande de faire de nombreux choix pour la préparation des données, la paramétrisation des modèles, leur sélection, etc. (Leroy, 2023). De ce fait, de nombreuses possibilités existent pour tenter de corriger les modèles présentés dans cet encadré. La perspective principale est ici de continuer à expérimenter différentes stratégies afin d'obtenir un modèle suffisamment proche de la réalité pour être utile. A titre d'exemple, l'utilisation de pseudo-absences (*i.e.*, points artificiels où l'absence de l'espèce est supposée) ou la filtration environnementale des données pourrait être des pistes à explorer.

Comme précisé dans le chapitre I, certaines de ces espèces sont ou pourraient être également présentes en Australie (cela est déjà avéré pour au moins 3 d'entre elles et probable pour une quatrième). Il serait intéressant de comparer ou compléter les données de présences de NCK avec celles d'Australie pour obtenir une vision plus holistique de la niche de ces espèces.

Chapitre III

La pérennité des connaissances pour leur
(ré-)utilisation dans la recherche et l'expertise

I. Résumé

Dans ce dernier chapitre, l'objectif est de trouver des pistes d'organisation des projets de recherche en biodiversité pour obtenir des connaissances pouvant s'actionner simplement pour la conservation de la biodiversité et être réutilisées pour d'autres thématiques de recherche.

Le travail autour de ce chapitre est intervenu à toutes les étapes de la thèse, de l'acquisition de la donnée à la rédaction d'articles. Les lignes directrices ont été essentiellement de se baser sur l'existant et de proposer des solutions sobres qui avantagent les futurs ré-utilisateurs des connaissances mais aussi le projet de recherche en cours. Ceci passe par la recherche d'approches, d'outils et services utilisables par le plus grand nombre. En effet, il s'agit ici de faire émerger des pistes pour maximiser la transparence, la reproductibilité, la ré-utilisation et la pérennité de différentes parties du processus de recherche sans augmenter significativement la charge de travail des personnes impliquées dans ce processus.

La volonté est de se placer dans un contexte plus large que celui de la temporalité de la thèse et au-delà du modèle biologique des crustacés d'eau douce ou de la Nouvelle-Calédonie-Kanaky. En effet, les connaissances sur les espèces, les milieux, les méthodes, les données, les analyses et la gestion du projet acquises pendant un projet de recherche peuvent se transmettre à une grande diversité d'acteurs à condition d'être pérenne et réutilisable.

Avant d'envisager la manière de pérenniser les connaissances acquises, il est crucial de s'interroger sur les publics qui pourraient avoir besoin de ces connaissances et à quelles fins. Entre autres, la recherche demande à vérifier les résultats et conclusions produites, à refaire la collecte ou l'analyse des données ou encore à réutiliser les protocoles, les données et les analyses. L'expertise cherche à référencer son travail, à appuyer les résultats et recommandations produites ou encore à appliquer les protocoles et analyses développées afin de répondre à des cadres légaux. Enfin, les acteurs du territoire (*i.e.*, citoyens, décideurs, entreprises) peuvent souhaiter se renseigner sur leur environnement, s'emparer des données et résultats qui y sont produits mais aussi s'approprier les connaissances générées pour actionner des démarches en faveur du territoire.

Ce chapitre est écrit en suivant la structure du cycle de vie de la recherche. Cependant, il a semblé important de distinguer le travail en deux sections. La

première se concentre sur le cadre de travail et les moyens utilisés pendant la thèse ainsi que leurs apports à la pérennité du projet. L'objectif étant de proposer un cadre de travail clair et accessible basé sur des solutions existantes en privilégiant la sobriété, les éléments présentés dans cette première section pourront sembler simples. La seconde section porte sur les parties « Nettoyer » et « Analyser » du cycle de vie de la recherche qui ont fait l'objet d'une attention particulière dans cet axe de la thèse. En effet, bien que des solutions techniques innovantes pour pérenniser et ré-utiliser les productions de ces parties soient existantes, elles n'ont commencé à être développées pour la recherche sur la biodiversité que récemment (Hardisty *et al.*, 2016). La mise en œuvre de ces solutions peut s'avérer complexe et mérite d'être accompagnée en proposant des protocoles dédiés aux communautés étudiant la biodiversité (Hampton *et al.*, 2017; Emery *et al.*, 2021). De fait, un cadre à appliquer pendant le développement des processus de nettoyage et d'analyse des données a été formalisé pendant cette thèse afin d'optimiser l'application des bonnes pratiques de recherche. Ces bonnes pratiques permettent de décupler les applications potentielles et les perspectives d'un projet de recherche. Il a fait l'objet d'une publication qui présente ce cadre ainsi que la plateforme web Galaxy-Ecology (voir partie III de ce chapitre).

II. Cadre de travail transparent, reproductible et (ré)utilisable

II. 1. Planifier

II. 1. i. Mise en œuvre

La planification est une étape particulièrement décisive pour le projet de recherche. L'ensemble des étapes du cycle de vie de la recherche y sont réfléchies, notamment via la mise en place des collaborations (*e.g.*, avec des acteurs locaux, des laboratoires) et des partenariats (*e.g.*, avec des financeurs, des infrastructures). Le Plan de Gestion de Données (PGD) initial de la thèse a été rédigé dès le début de la thèse et est disponible annexe ES1 et en ligne (<https://doi.org/10.48579/PRO/DGVWPW>; fichier READ ME).

Le design expérimental est défini à ce stade du projet, il permet de déterminer les protocoles qui seront utilisés pour collecter la donnée et assurer qu'elles pourront répondre aux questions posées. Le projet de recherche de cette thèse se base sur des échantillons récoltés majoritairement en amont avec un protocole qui a été défini dans le cadre d'un projet différent. Cependant, les protocoles utilisés et les quantités d'échantillons et d'individus à traiter pour les

tris ou la biologie moléculaire ont pu être définis en amont des phases de laboratoire. Ceci a permis d'anticiper la demande de financement pour les manipulations ou encore les gratifications de stages.

II. 1. ii. Apports pour la pérennisation

Premièrement, la planification pose les fondations sur lesquelles la transparence, la reproductibilité, la ré-utilisabilité et la pérennité du projet de recherche vont se baser. Par l'écriture et le partage d'une première version de plan de gestion de données, le cadre de travail dans lequel s'inscrit le projet de recherche est, par définition, plus transparent et pourra être réutilisable pour la mise en place d'autres projets.

De plus, il est un support utile pour dialoguer avec les collaborateurs et partenaires sur le projet de recherche dans son ensemble. Des accords peuvent être passés et notés en amont du projet concernant la gestion des objets numériques, leurs modalités de réutilisation et de partage. Notamment, l'utilisation des principes FAIR (Wilkinson *et al.*, 2016) et CARE (Carroll *et al.*, 2020) permet d'assurer un équilibre entre les principes de sciences ouvertes et la protection des publics potentiellement impactés par les travaux effectués ou mentionnés dans les données. Par la transparence des données et de leurs modes d'acquisition ainsi que la définition claire des modalités de réutilisation, la réutilisation des données par tous les publics est favorisée.

La transparence sur les données qui ne seront pas partagées publiquement permet aux personnes participant au projet mais aussi aux personnes externes d'avoir connaissance de l'entièreté des données produites. De plus, les motifs avancés pour le non-partage des données permettent aux personnes souhaitant reproduire un projet de recherche similaire d'anticiper le type d'information à ne pas partager.

II. 2. Collecter

II. 2. i. Mise en œuvre

La collecte des données rassemble des étapes telles que l'échantillonnage, les phases de laboratoire ou encore l'investigation de sources de données existantes. Les protocoles représentent le cadre dans lequel la récolte de données a été réalisée tandis que les méthodes détaillent l'application des protocoles dans une temporalité et un projet précis (<https://eml.ecoinformatics.org/eml-schema> ;

section 5.1.6). Ici, en ce qui concerne la biologie moléculaire, les protocoles ont été saisis sur la plateforme protocols.io et les méthodes détaillées inscrites dans le cahier de laboratoire (tab. 3.1).

Tableau 3.1 – Types et liens vers les productions liées à la collecte de données.

Type de production	Liens
Protocole d'extraction ADN	https://dx.doi.org/10.17504/protocols.io.36wgqnm9yvk5/v1 https://dx.doi.org/10.17504/protocols.io.j8nlk4x7wg5r/v2
Protocole de PCR	https://dx.doi.org/10.17504/protocols.io.261ge4y4jv47/v3 https://dx.doi.org/10.17504/protocols.io.n92ld8q57v5b/v1
Cahier de laboratoire	Dépôt HAL en préparation

II. 2. ii. Apports pour la pérennisation

Dans une publication scientifique, ce sont les méthodes et non les protocoles qui sont décrits en priorité. En effet, les protocoles peuvent être pensés et décrits afin d'être utilisés dans d'autres contextes et par d'autres personnes. Cette manière de décrire les protocoles peut notamment permettre de rendre les méthodes plus reproductibles grâce à cette base commune que représente le protocole. Les méthodes décrites dans les articles scientifiques permettent en priorité de citer les protocoles utilisés (*e.g.*, identification d'espèces sur transects, utilisation d'un protocole de PCR) et les modalités de ce protocole qui sont propres aux résultats décrits (*e.g.*, dates de réalisation des transects, amorces utilisées pour la PCR) afin de permettre aux pairs d'évaluer la validité et la qualité des résultats.

Le fait de partager les protocoles à travers une plateforme dédiée les rends plus facile à trouver. En effet, rechercher un protocole à travers des articles scientifiques est souvent laborieux. Une telle plateforme permet de trouver des protocoles directement par leur nom, par le matériel utilisé ou encore par le modèle biologique utilisé.

II. 3. Organiser

II. 3. i. Mise en œuvre

Les individus des populations du copépode *Boeckella* publiées dans le premier article de cette thèse (Royaux *et al.*, 2024) ont été préparés pour être déposés en collection du Muséum National d'Histoire Naturelle. A terme, lorsqu'ils seront tous traités, tous les échantillons récoltés y seront déposés (tab. 3.2).

Les données ont été entreposées en continu sur le cloud du Centre e-Science du Grand Ouest (CeSGO), un espace de stockage et de partage de données dédié à la recherche et associé à un centre de calcul scientifique. Un logiciel client de ce cloud a été installé sur deux ordinateurs pour que les données puissent être actualisées en parallèle d'une part par moi-même et, d'autre part, par les étudiantes et étudiants ayant collaboré au tri des échantillons. L'organisation des données a été pensée pour qu'elles puissent être renseignées en collaboration, soit en cherchant un équilibre entre l'aisance de saisie, la rationalisation du volume nécessaire au stockage et la qualité des données. L'organisation des données a évolué avec l'usage. Les données de génétique y sont également stockées. Les données environnementales utilisées pour les modélisations dans le chapitre II sont toutes disponibles librement en ligne. Elles ont été téléchargées et stockées localement et, afin d'assurer une bonne transparence, tous les nettoyages effectués ont été réalisés à travers des scripts partagés via la plateforme GitHub (voir partie « Ecrire et publier »).

En s'inspirant de la manière de rédiger un « data paper » (*i.e.* publication décrivant un jeu de données), les métadonnées détaillées sur chacun des champs (*i.e.*, colonnes des tableaux notamment) ont été écrites en français dans un document texte Word (annexe ES2) afin d'être lisible et utilisable par les personnes collaborant sur le projet.

Le plan de gestion de données est un document évolutif et ses mises à jour régulières sont importantes afin de conserver un suivi de l'évolution de l'organisation du projet (tab. 3.2).

Tableau 3.2 – Types et liens vers les productions liées à la collecte de données.

Type de production	Liens - identifiants
Identifiants des dépôts en collection	De MNHN-UI-2017-2297 à MNHN-UI-2017-2316
Versions du PGD	Dépôts sur Data.Indores en préparation

II. 3. ii. Apports pour la pérennisation

Le dépôt en collection permet de conserver les échantillons récoltés à long terme. En effet, les plus vieux spécimens de crustacé répertoriés dans les collections du MNHN sont datés de la fin du 18^e siècle. Bien que des démarches administratives parfois lourdes soient nécessaire pour étudier les échantillons déposés en collection,

ceux-ci sont rendus faciles à trouver, accessibles et ré-utilisables par ce biais. De plus, dans le cadre des principes CARE, les échantillons prélevés dans cette thèse pourront éventuellement être restitués au territoire de la NCK si les laboratoires locaux le souhaitent et ont les infrastructures nécessaires.

Bien qu'il soit important d'avoir un format de données cadré pour les pérenniser, la réalité d'un projet de recherche nécessite de conserver une forme de flexibilité afin de s'adapter aux usages. En effet, les formats trop contraints peuvent entraîner une perte d'information importante. Il s'agit d'un équilibre à trouver pour conserver une forme contrainte en laissant la possibilité d'ajouter des informations cruciales qui n'avaient pas été formalisées à l'origine. Bien entendu, tout ceci est à réfléchir en gardant en tête les problématiques liées au « big data ». De fait, une quantité importante de stockage peut être épargnée via une organisation réfléchie des jeux de données, notamment en se référant aux 5 V du « big data » (Vélocité, Volume, Valeur, Variété, Véracité).

II. 4. Nettoyer et Analyser

Le nettoyage et l'analyse de données est un point particulièrement complexe quand il s'agit de les pérenniser par la transparence, la reproductibilité et la réutilisation. C'est pourquoi elle a fait l'objet d'un développement plus approfondi qui est présenté dans l'article en partie III de ce chapitre. Les développements effectués dans ce cadre sont détaillés dans l'encadré C.

Bien que la majeure partie des nettoyages et analyses soient rendues disponibles, un script de nettoyage des données de cette thèse ne sera pas partagé. En effet, c'est un script de « pré-nettoyage », il vérifie chaque champ des tableaux de données pour détecter de potentielles fautes de frappe (*e.g.*, caractères spéciaux, dénomination de taxons d'après le référentiel du NCBI ; National Center for Biotechnology Information), données manquantes (*e.g.*, dates non renseignées ou incomplètes, identifiant de localité inexistant) ou encore erronée (*e.g.*, coordonnées de localité incohérentes, information dans la mauvaise colonne). Ce type de nettoyage est très spécifique aux données produites et est difficilement réutilisable. Cependant, ce processus pourrait être amélioré par l'utilisation d'outils pour le nettoyage de données et le repérage de valeurs aberrantes (outliers) tel que OpenRefine.

II. 5. Ecrire et publier

II. 5. i. Mise en œuvre

Au total, 86 séquences génétiques ont été partagées et 88 sont en cours de traitement sur l'entrepôt international GenBank (NCBI). Les 348 fichiers bruts de séquençage, les données sur les individus, les échantillons et les localités ont été ou seront partagées sur l'entrepôt national data.InDoRES (Inventaire des Données de la Recherche en Environnement et Sociétés). Les métadonnées écrites en français sont associées au Plan de Gestion de Données dans le fichier « ReadMe » du dépôt data.InDoRES, ceux-ci ont été adaptés en langage Markdown (lisible par les ordinateurs et exportable en divers formats, *e.g.*, HTML, PDF). Celles-ci ont ensuite été adaptées au standard Ecological Metadata Language qui a permis la production automatisée d'un rapport sur les métadonnées et d'une ébauche de « data paper » en anglais (annexe ES3). Les 26 scripts R de nettoyage et analyses de données ont été et seront mis à disposition sur GitHub (tab. 3.3).

Le premier article publié dans le cadre de cette thèse (présenté dans le chapitre I) a été l'objet d'un processus de révision fermée et est disponible sur HAL (tab. 3.3). L'article présenté dans ce chapitre a été proposé à travers le système de révision ouverte par les pairs « Peer Community In Ecology ». Cet article a été révisé et recommandé par une personne anonyme et deux chercheurs internationaux, Nick Isaak et Timothée Poisot (tab. 3.3).

A travers les collaborations avec des acteurs locaux, un besoin particulier a été identifié : la production d'une clé d'identification illustrée et didactique. Ce travail est en cours de construction. En outre, une communication dans un ouvrage collectif de littérature dite « grise », a été écrite sur l'endémisme observé chez les espèces du genre *Boeckella*. Cet ouvrage intitulé « Histoire des crustacés » sera publié, en avril, par les éditions EDP Sciences et a pour cible le grand public à partir du niveau collège (annexe ES4).

Durant la thèse, certaines approches méthodologiques développées en équipe autour du traitement des échantillons ont amené à proposer une déclaration d'invention à la cellule de valorisation du Muséum national d'Histoire naturelle. Dans un premier temps, le contenu de la déclaration d'invention doit rester strictement confidentiel, et donc aller contre les principes de science ouverte. Cependant, à terme, ce processus de valorisation permet de diffuser les

applications d'un protocole de recherche à l'expertise et d'augmenter son potentiel de réutilisation.

Tableau 3.3 – Types et liens vers les productions liées à la publication et à l'écriture.

Type de production	Liens - identifiants
Identifiants des dépôts NCBI	De MNHN-UI-2017-2297 à MNHN-UI-2017-2316
Données partagées	https://doi.org/10.48579/PRO/DGVWPW
Révision ouverte PCI	Pre-print et manuscrit final : https://ecoevorxiv.org/repository/view/6965/ Révision : en cours de mise à disposition par PCI
Version HAL premier article	https://hal.science/hal-04681440
Dépôts GitHub	https://github.com/ColineRoyaux/population_genetics https://github.com/ColineRoyaux/cleanSIG https://github.com/ColineRoyaux/DM

II. 5. ii. Apports pour la pérennisation

L'ouverture des données et processus analytiques liés aux publications scientifiques permet aux comités de révision et aux lecteurs des articles de vérifier la validité des résultats présentés. De plus, par leur partage sur des plateformes dédiées (et non uniquement en annexe des articles), les données et processus analytiques sont plus faciles à trouver pour tous les publics.

L'écriture et la mise à disposition des métadonnées dans un format lisible pour l'humain en plusieurs langues permet d'étendre l'accessibilité et la transparence sur les données à un public plus large (fichiers Word et Markdown). De plus, l'utilisation d'un standard de métadonnées (EML) permet de rendre les données lisibles par les ordinateurs et plus facile à utiliser en interaction avec d'autres données (interopérable).

La construction de collaborations durables avec des acteurs des territoires considérés est cruciale pour pérenniser les connaissances produites pour un public plus large que celui de la recherche. En effet, le format et le type d'information nécessaire est propre à chaque public et nécessite donc des échanges récurrents.

Le niveau de réutilisabilité des processus analytiques publié dépend de la façon dont ils ont été développés et partagés, ceci est détaillé dans la partie III de ce chapitre.

III. Cadre pour les bonnes pratiques de développement des processus analytiques

Guidance framework to apply best practices in ecological data analysis: Lessons learned from building Galaxy-Ecology

Royaux Coline | Mihoub Jean-Baptiste | Jossé Marie | Pelletier Dominique | Norvez Olivier | Reecht Yves | Fouilloux Anne | Rasche Helena | Hiltemann Saskia | Batut Bérénice | Eléaume Marc | Segumineau Pauline | Massé Guillaume | Amossé Alan | Bissery Claire | Lorrilliere Romain | Martin Alexis | Bas Yves | Virgoulay Thimothée | Chambon Valentin | Arnaud Elie | Michon Elisa | Urfer Clara | Trigodet Eloïse | Delannoy Marie | Loïs Gregoire | Julliard Romain | Grüning Björn | The Galaxy-E community | Le Bras Yvan

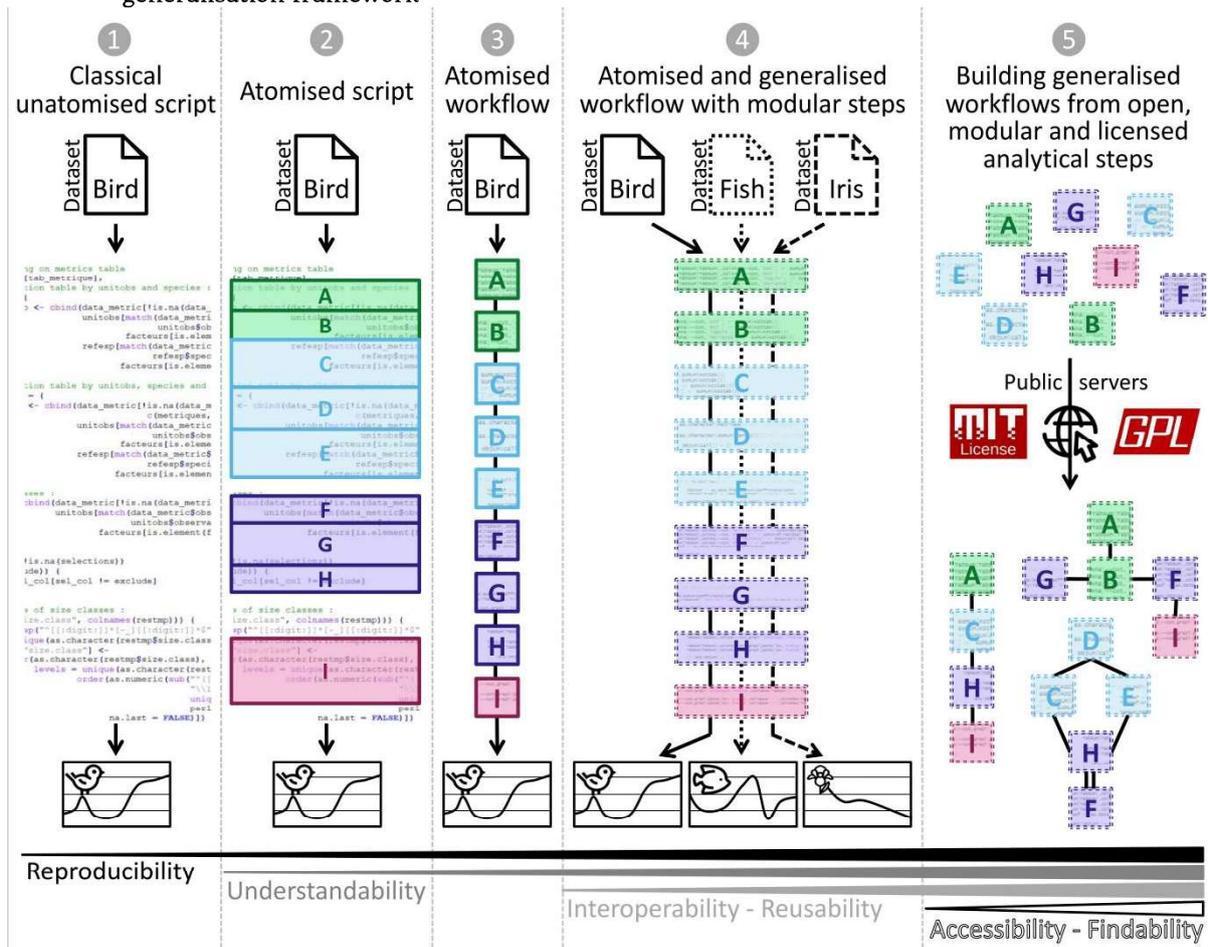
Accepted, Peer Community In Ecology

<https://doi.org/10.32942/X2G033>

Abstract

Numerous conceptual frameworks exist for best practices in research data and analysis (*e.g.* Open Science and FAIR principles). In practice, there is a need for further progress to improve transparency, reproducibility, and confidence in ecology. Here, we propose a practical and operational framework for researchers and experts in ecology to achieve best practices for building analytical procedures from individual research projects to production-level analytical pipelines. We introduce the concept of atomisation to identify analytical steps which support generalisation by allowing us to go beyond single analyses. The term atomisation is employed to convey the idea of single analytical steps as “atoms” composing an analytical procedure. When generalised, “atoms” can be used in more than a single case analysis. These guidelines were established during the development of the Galaxy-Ecology initiative, a web platform dedicated to data analysis in ecology. Galaxy-Ecology allows us to demonstrate a way to reach higher levels of reproducibility in ecological sciences by increasing the accessibility and reusability of analytical workflows once atomised and generalised.

Graphical abstract – Levels of attainable best practices through the atomisation – generalisation framework



Keywords

Biodiversity, Reproducible analyses, Galaxy, Best practices, Atomisation, Generalisation, Workflows, Ecoinformatics, Conda, Container, Common Workflow Language, RO-CRATE

Introduction

Ecology's Reproducibility Crisis

Research in ecology is increasingly shaped by the availability of novel analytical solutions and statistical tools. Given the ever-growing amount of data available, much attention is often given to the thought process behind statistical analyses to handle different data distributions, pseudo-replication, and sampling biases for instance (NERC 2010, 2012; Hampton *et al.*, 2017; Emery *et al.*, 2021). Despite the high-quality standards required by the scientific community from data access to analysis, the level of complexity of ecological systems makes results difficult to reproduce. The ongoing “reproducibility crisis” has also led researchers to pay

closer attention to the quality of analyses to increase confidence in their studies and conclusions (Fanelli, 2018; Ioannidis, 2022). Reproducibility (*i.e.* different teams and experimental setups obtaining similar results; Plesser, 2018) is one of the main criteria for evaluating robust science and reliable conclusions. The term “reproducibility” is a relative concept and has known various definitions depending on field and context. Reproducibility of analyses (“computational reproducibility”) is defined by Cohen-Boulakia *et al.* (2017) as the ability of distinct analyses to reach to the same conclusion.

In the current context of the global biodiversity crisis, the scientific community needs to use all available data and provide as robust as possible evidence regarding the state and dynamic of ecological systems, from genetic to ecosystem. At the same time, using analytical tools to provide robust evidence can be complex and may require advanced skills that are not widely available across the scientific community (Hampton *et al.*, 2017). Therefore, operational solutions and methodological guidelines can allow analytical workflows to be more accessible without degrading the scientific quality of analyses, and thus, promote efficient and broad deployment of best practices.

Is the ecology community failing to meet best practices?

The first step towards reproducibility is knowing current best practices and recommendations. Among them, the FAIR principles (Wilkinson *et al.*, 2016), for which the availability of the data and the code used for each published result is an essential criterion, may be key for appropriate management through the data life cycle (Michener, 2015). The FAIR principles (see also CARE principles by Carroll *et al.*, 2020) are considered as a founding framework to share data along four important elements: "Findable" for humans and machines; "Accessible" with a detailed access procedure; "Interoperable" for interaction with other data or applications; "Reusable" in an identical or different context. In addition to these principles, propositions have been delimited within several thematic communities in ecology to evaluate and enhance best practices application, notably the Species Distribution Modelling communities (Araújo *et al.*, 2019; Zurell *et al.*, 2020).

Although data accessibility has been substantially improved in ecology during the past decade, sharing analytical scripts and codes remain largely marginal (Archmiller *et al.*, 2020; Culina *et al.*, 2020; Minocher *et al.*, 2021; Ivimey-Cook *et al.*, 2023). However, even if sharing code is necessary to achieve good

computational reproducibility, it is insufficient. Therefore, the utilisation of computational workflows has been suggested as a solution for improving computational reproducibility (Cohen-Boulakia *et al.*, 2017; Grüning *et al.*, 2018) through software such as Snakemake (Köster & Rahmann, 2012), Nextflow (Di Tommaso *et al.*, 2017), or Galaxy (The Galaxy Community, 2022). A workflow is generally defined as a sequence of distinct computational tasks for a particular objective (Goble *et al.*, 2020). As such, a workflow represents the backbone of a single specific analysis. Throughout the analytical procedure, a typical workflow starts with raw data, which can be extracted from several databases or data files and processed through a series of analytical steps. The products resulting from these analytical steps (*i.e.* the outputs of the computational workflow) can be data files, graphic representations and any associated metrics.

When properly designed, a certain level of reproducibility can be easily achieved since workflow languages naturally capture the following four key elements (Cohen-Boulakia *et al.*, 2017):

- the specificities of the workflow, the analysis steps and associated tools;
- the workflow entries, datasets and parameters;
- the environment and context of the use of the workflow;
- the results obtained and the outputs of the workflow.

In the original publication of Wilkinson *et al.* (2016), the focus of FAIR principles was mainly on observational data. However, the principles can be applied to software and computational workflows (Lamprecht *et al.*, 2019; Goble *et al.*, 2020). For instance, a code shared as supplementary material of a non-open access publication could be considered as "Interoperable" but is not easily "Findable", "Accessible", or "Reusable". In contrast, a large block of code consisting of several hundred lines, from data pre-processing to final results and graphics as pictured in the Graphical abstract ①, may require efforts to understand and adapt to other kinds of data ("non-reusable"), mainly if annotations or comments are limited. Similarly, an analytical procedure shared without indicating the versions of hardware, software, and packages has a low chance of producing identical outputs, making it less reproducible. These issues may harm the scientific community by preventing fully transparent communication among users about knowledge production and practice comparison. They can also be detrimental to individual authors, when they need to update or run new analyses.

Impact on Ecology Research

The efficiency of the scientific process is greatly affected by the lack of computational reproducibility and FAIRness of analytical procedures. The adoption of FAIR practices was estimated to save 10.2 billion € per year in Europe (Munafò *et al.*, 2017; European Commission, 2018; Gomes *et al.*, 2022). Moreover, consistent application of reproducibility and FAIR principles will improve trust in research studies and scientific reports (Powers & Hampton, 2019; Lortie, 2021; Jenkins *et al.*, 2023).

The widespread use of computational languages to process large-scale data and analyse complex systems has been a major advance in studying the ecosphere at any spatio-temporal scale (Michener & Jones, 2012; Farley *et al.*, 2018). However, the ever-growing technical and programming skills required to take advantage of such computational solutions by the scientific community raise new challenges (Jetz *et al.*, 2019; Boyd *et al.*, 2023; Leroy, 2023). The use of increasingly complex analytical solutions, paired with different approaches or programming languages, creates barriers to uptake and challenges for peer-review. Indeed, many ecologists have acquired their programming skills through self-study or through courses that combine instruction in statistics and ecological principles with an introduction to programming. This learning process does not inherently compromise the quality of the analyses and results; however, it may lead to inappropriate coding habits. As a response to this situation, adequate training was identified by life science researchers (*Community Survey Report*, 2013; Williams & Teal, 2017; Larcombe *et al.*, 2017), as it would help involve more people in the understanding of current analytical solutions and benefit to scientific cooperation (Touchon & McCoy, 2016; Gownaris *et al.*, 2022). Research is typically structured through a highly competitive organisation, with a potentially detrimental effect on scientific knowledge (Fang & Casadevall, 2015). Instead, fostering collaboration and collective intelligence by promoting transparent sharing of analytical procedures, would offer more persistent and robust ways to achieve actionable science (Ellemers, 2021). Such efforts would be of paramount importance in environmental sciences and the conservation of biodiversity by providing governance and guiding actions with increasingly robust evidence (Keenan *et al.*, 2012).

Are there simple and ready-to-use solutions?

In this note, we aim to promote the reuse of existing concepts and solutions as pillars toward better practices for ecological analyses by providing a streamlined framework. We believe the atomisation-generalisation framework presented in the second part of this note represents an operational and actionable path for researchers and experts to attain levels of best practices (*e.g.* reproducibility, FAIR, open science, R compendium; Casajus, 2023) with no more investment than they are able or willing to provide (Field *et al.*, 2014). Atomisation is used to refer to the identification of distinct analytical steps each constituting an analytical procedure. It is a non-standard term introduced in this note to convey the idea of analytical “atoms”. As for atom particles that etymologically correspond to “indivisible” but are composed of subatomic particles, an analytical atom represents a single analytical step composed of several functions. Generalisation involves the alteration of an analytical step to enlarge its applicability in diverse contexts and for diverse purposes. Therefore, generalisation cannot be efficiently achieved without prior atomisation.

Atomisation and Generalisation are central organising principles in the design of the Galaxy-Ecology (Galaxy-E) initiative (see section III). Galaxy-E is a demonstration platform for applying best practices such as the FAIR principles and computational reproducibility for analytical procedures in ecology. Hence, this technical note is partly Galaxy-oriented, not to present the platform as a prescriptive solution but to give an operational example of the best practices it helps to achieve.

Guidelines for best practices

Atomisation: what is it and why?

Atomisation refers to dividing an analytical procedure into several specific steps (“atoms”; Graphical abstract ②) generating a suite of elementary analytical steps as pictured in the Graphical abstract ③. For instance, in a maximally-atomised workflow, each small step would be conducted by its own bespoke function. Breaking down the analytical process into atoms functioning as building blocks allows for better understanding, modularity, and visibility of the analytical flow. It permits making it more accessible to a broader audience or facilitating the peer-review process. Indeed, an extended one-block code that imports raw data, makes pre-processing steps (*e.g.* filter, formatting), conducts analyses (*e.g.* distribution study, modelling), and performs final representations of results (*e.g.* maps, plots) can be challenging to understand and reuse by others or even the same person after some time.

McIntire *et al.* (2022) described the PERFICT approach (Prediction, Evaluation, Reusability, Free access, Interoperability, Continuous workflows, and routine Tests) to set a new foundation for models in predictive ecology. This can be applied more generally to the analytical procedure in ecology and biodiversity. In their article, McIntire and collaborators make an analogy between code development and Lego® construction, similar to our definition of atomisation. Functions are a workflow’s most fundamental analytical steps and can be seen as modular pieces, alike single pieces of Lego®. Modules can be created from a single or series of successive functions comparably as in Lego® structures made of several pieces (*e.g.* meant to build cars, houses, or road). These modules (or atoms, tools) can be used as standalone or combined to make simple to complex analytical workflows (*e.g.* data formatting or curation, running statistical models, or generating graphical elements for visualisation). Doing so, the atomisation approach may facilitate sharing or teaching analytical practices since beginners can easily understand the general organisation of the analytical procedure by simply reading the list of steps in the analysis with a limited degree of complexity. Decoupling programming skills from analytical skills can make data processing more accessible to a wider audience. Indeed, once each elementary step is clearly identified and delimited along the atomisation process, it is easier to grasp the whole analytical procedure and focus on the review of each step at a time or (re)use it. New workflows can further be generated by recombining existing,

validated or peer-reviewed elementary steps in innovative ways. This process can save time, increase confidence, and avoid potential programming mistakes, allowing greater focus on understanding the analytical workflow.

Generalisation: what is it and why?

Generalisation refers to the modification of an analytical procedure to make it applicable to many settings, by removing specificities related to a particular data file or data format. This means trying to avoid hard-coding anything that is specific to the structure of the original dataset (e.g. number of years). Generalisation aims to optimise the reusability at different times (e.g. regular result update), enlarge the application of a given analysis to different input data files while keeping the initial analytical procedure fully reproducible as pictured in the Graphical abstract ④. Generalising an analytical step requires identifying key elements and invariant parameters from those that must be adaptable to allow for the analysis to be applied to specific characteristics of various datasets. These parameters must be implemented to be easily modified if needed. Generalisation can be tricky because the higher the flexibility of an analytical step, the greater the risk of errors in its use. This is why generalisation should be complemented by clear statement and an implementation of red flags and warnings to prevent such events. As with atomisation, generalisation is primarily a conceptual way to build analytical procedures. It requires minor change of practices to reach certain degree of generalisation, avoiding additional effort later for reusability, reproducibility, and share.

Practical steps towards atomised and generalised coding

Breaking down codes into elementary steps to achieve atomisation is not an intuitive task at first as it may target a single function or a more intricate set of several functions. There could be different degrees of atomisation, depending on the grain required to decompose the analytical process (fig. 3.1.1; tab. 3.1.1). The application of general guidelines and best practices implies finding a balance between the most appropriate degree of atomisation and generalisation. This depends on the type of analytical procedure or the targeted audience (e.g. with different interests and programming skills). Attention to this balance is critical to ensure that the analytical procedures could be reused. For instance, a workflow in which each function would be considered as a unique elementary step would optimise the flexibility but may likely add unnecessary complexity. At the other

extreme, considering a whole analytical workflow as an elementary step may make it ready-to-use and simplify its application, but would be too coarse and therefore limit flexibility by violating the principle of atomisation.

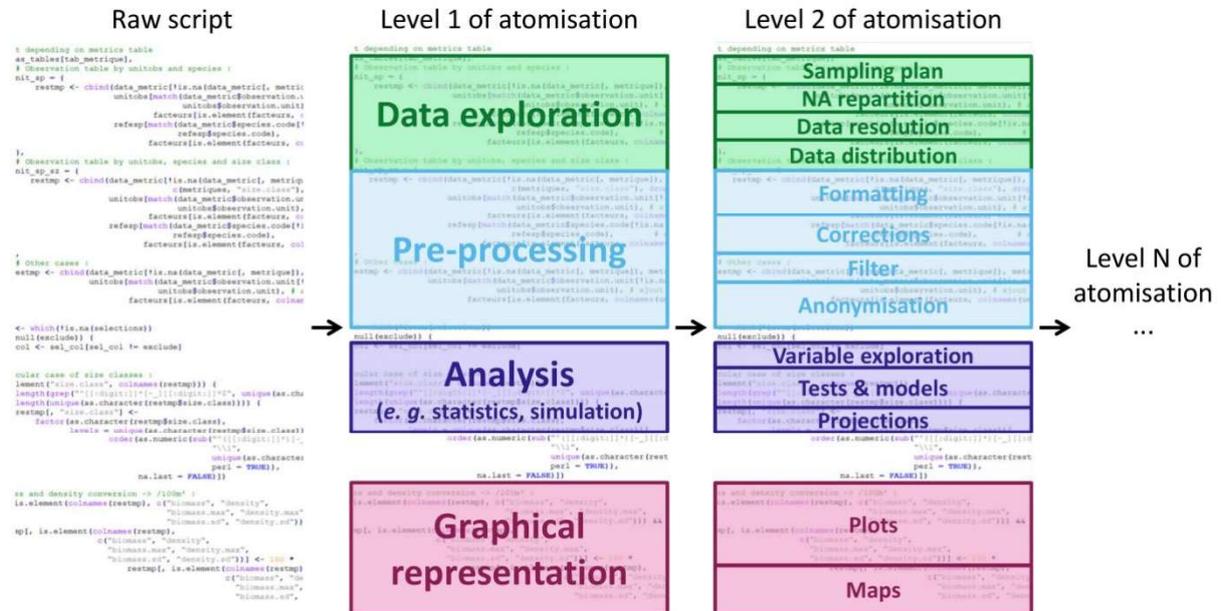


Figure 3.1.1 - Illustration of the atomisation of an existing code. The first level of atomisation is delimitating the large sections of an analytical procedure that exist in almost all procedures. This first level is conveyed using same colours to the second level of atomisation where more detailed and specific analytical steps are illustrated in each section. The process of atomisation can continue through a multitude of levels, ultimately leading to the maximally atomised procedure, which is comprised of a single function.

Table 3.1.1 - Example of atomisation levels

Level 1 - big shape	Level 2	Level 3
Data exploration	Sampling plan	Complete Balanced
	Missing values	Proportion Distribution
	Data granularity	Geographic resolution Temporal resolution Measure resolution
	Data distribution	Geographic coverage Temporal coverage Measures ranges Summaries

Pre-processing	Formatting	Change file format Change general format
	Corrections	Remove special characters Remove low trust observations Correct measures
	Filtering	Remove unwanted observations
	Anonymisation	Anonymise names Anonymise localities Anonymise species

Analysis	Variable exploration	PCA Collinearity Correlation
	Unimodal tests	Linear Models χ^2 Student
	Statistical models	Generalised Linear Models Generalised Additive Models Random Forest
	Models Evaluation	Evaluation metrics (<i>e.g.</i> AIC, Jaccard) Validation methods
	Projections	Geographical projections Temporal projections

Representation	Plot	Raw variables Modelled results
	Map	Observations Projections

A few changes in code-writing habits can enhance the reusability of the analytical procedure by generating easy-to-understand analytical procedure without investing much time. It is best to develop each elementary step directly in separate code files and to give details of the order in which elementary steps are used for each analytical workflow. To ensure reproducibility and traceability of the results, each computation of the analytical workflow should be associated with the details of the parameters settings and datasets used. From a practical point of view, a couple of recommendations could be made for coding elementary steps to facilitate generalisation and ease the reuse. Once each elementary step is defined,

we recommend all dependencies (*e.g.* software version, packages, libraries and their versions) to be set at the same place, at the start of the code, followed by modular parameters (*e.g.* input file location and name, column selection, modelling parameters, data specificities, output saving location). When the script of the elementary step is completed, modular parameters should be the only part of the code that may be modified in future reuse. Dependencies and subsequent computational tasks should be left untouched to ensure the integrity of the analysis and then, reproducibility. In the end, it is best to add an open-source license to any analytical procedure shared publicly (*e.g.* MIT, GPL). It permits to clearly state the terms and conditions of diffusion, share and reuse.

As such, atomisation and generalisation may overcome social or psychological barriers related to transparent sharing, either related to securing ownership (*e.g.* DOI) and to embarrassment or fear during a peer-review process (Gomes *et al.*, 2022). Indeed, as atomisation and generalisation notably permit higher readability of codes, it would be more straightforward for the writer or even trusted peers to verify and review the steps before submission.

Atomisation and generalisation are related and complementary concepts that may be applied from the earliest stages of the programming development. Indeed, atomisation into adequate elementary steps is necessary to properly generalise an analytical procedure as it permits to enhance the modularity of the procedure and its capacity to be tailored to different data types.

Entering a new dimension: the Galaxy-E initiative example

Developing open and properly atomised and generalised analytical procedures can already represent a significant step forward in terms of best practice. Galaxy is a good illustration of atomisation and generalisation with easier management of analytical workflows. The platform proposes many analytical tools that represent generalised and atomised elementary steps. These tools are modular and openly licensed, which permits to build generalised workflows as pictured in the Graphical abstract 5.

Galaxy (The Galaxy Community, 2022) is a workflow-oriented web platform for analysing data and sharing outputs. It allows scientists to share, develop, and use various datasets and data processing tools (*e.g.* data formatting, statistical tests, graphic representations).

Galaxy enables good reproducibility for data exploration and analyses, helps compute intricate analyses on big data files, enables collaboration, and can support the teaching process. Galaxy-E is a Galaxy server dedicated to ecological analyses maintained by the European Galaxy team (supported by the German Federal Ministry of Education and Research and the German Network for Bioinformatics Infrastructure), and is available at <https://ecology.usegalaxy.eu>.

Galaxy-E is mostly aimed at scientists that process biodiversity data and already understand the general functioning of the analytical procedures they want to produce. The rationale for a user would be to create or reuse analytical workflows with high FAIRness in a collaborative and open source platform. It can be used for individual analyses as well as for collaborative projects. In some cases, if the analytical procedure is already clearly defined, it can be used by citizens or for teaching.

There are different Galaxy servers, at global, continental, and national levels (European and French levels for example), but also according to the fields (*e.g.*, biomedical, ecology, climate). The Galaxy-E initiative is hosted by European (<https://ecology.usegalaxy.eu>) and French (<https://ecology.usegalaxy.fr>) servers.

Datasets can be uploaded on a Galaxy server from a local device, an online server, or a database. Users can then access every available tool (fig. 3.1.2, left panel) to modify, explore, and analyse their data. All tools used, parameters, and data (inputs and outputs) of the analysis are saved in a private “Galaxy history” (fig. 3.1.2, right panel), documenting every step of the analytical procedure and recording the provenance of each output. From any history, the user can extract a workflow (fig. 3.1.3) or directly share or publish the history itself. Workflows are reusable through WorkflowHub (<https://workflowhub.eu>) or Dockstore (<https://dockstore.org>) and exportable in CWL and RO-CRATE standards.

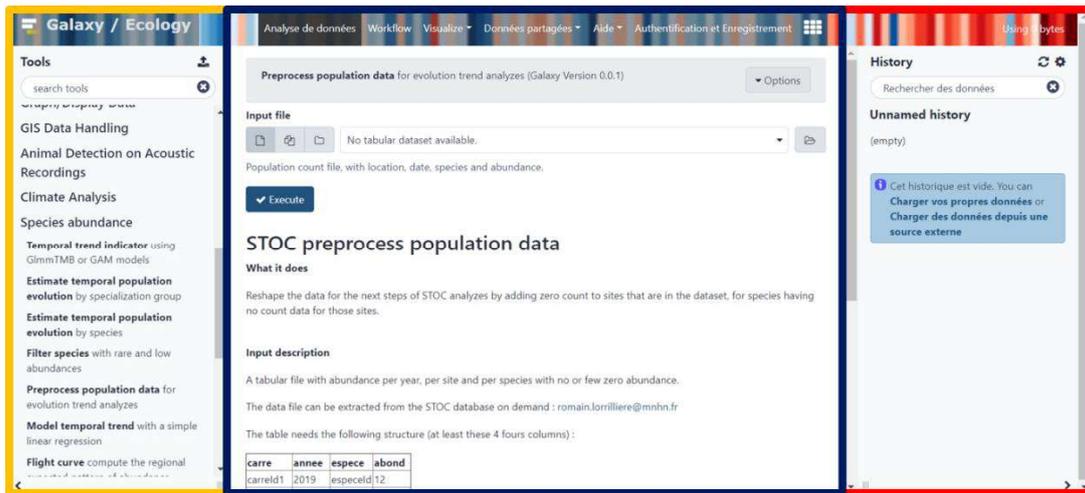


Figure 3.1.2 - Galaxy-Ecology users' interface <https://ecology.usegalaxy.eu>. Yellow panel on the left: analysis tool list; blue panel in the middle: current tool interface; red panel on the right: Galaxy analysis history

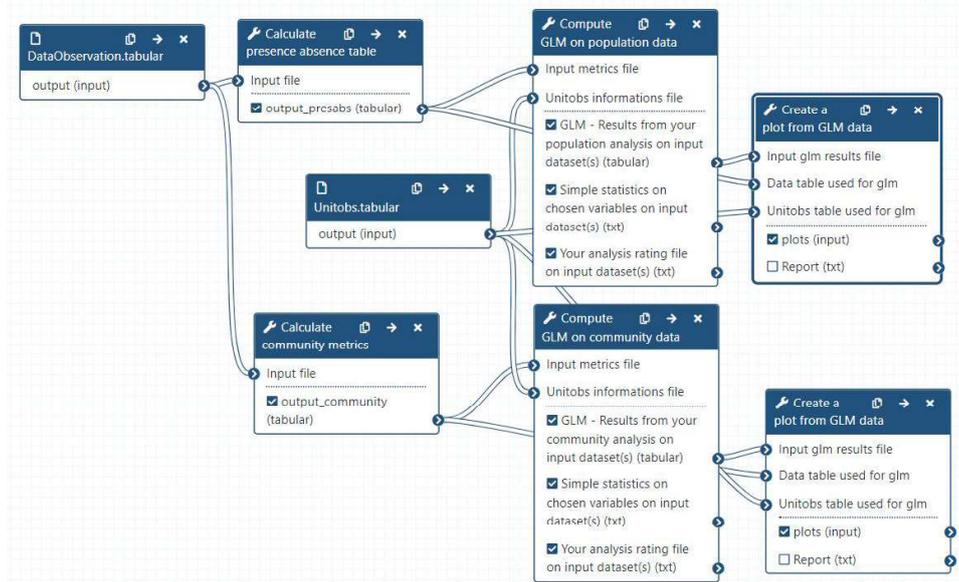


Figure 3.1.3 - Representation of a Galaxy workflow in the editing interface of a Galaxy server. Each box represents an analysis tool, and the lines represent the flow of data through the tools. In relation with the atomisation-generalisation framework, each box (tool) corresponds to an atomised and generalised step with editable parameters, inputs and outputs.

Any analytical procedure can be adapted on the platform and Galaxy can be used through the whole data life cycle (https://rdmkit.elixir-europe.org/galaxy_assembly). One can use off-the-shelf tools, workflows, and tutorials to design an analytical procedure, or suggest, develop, and share new workflows and tutorials, two aspects that do not require coding skills.

As each Galaxy tools are atomised and generalised elementary steps that can be articulated in a workflow, the Galaxy platform benefits from the same advantages as atomisation and generalisation and can help enhancing best practice application (tab. 3.1.2).

Table 3.1.2 - Illustration of how the atomisation-generalisation framework and Galaxy implements and conforms to best practice.

	Atomised-generalised code	Galaxy
Reproducibility and transparency	<p>Environment, software and package versions</p> <p>Can be indicated but possibly hard to manage</p> <p>Can also be set as an output of the analysis (<i>e.g.</i> session info)</p> <p>Packages written in each coded elementary step or using a versioning system such as Conda</p>	<p>Entirely packaged with Conda package manager and BioContainers</p> <p>Possibility to store analytical procedures as containers for persistent execution</p>
Inputs and parameters	<p>One must keep track of different parametrisation and input settings at each computation</p>	<p>Automatically tracked and shareable with the “Galaxy history”</p>
Peer-review	<p>Organisation of the analytical procedure reviewable by non-code developers</p> <p>Code developers might be able to detect errors as it is easier in shorter scripts</p> <p>Transparency over the development process achievable through Git</p>	<p>Reviewable “Galaxy history” and re-executable workflow</p> <p>Continuous peer-reviewed of tools with open-source code</p> <p>Transparency over the development process through Git</p> <p>The workflows can be reviewed by the Intergalactic Workflow Commission (IWC) for best practices</p>
Output provenance	<p>Can be tracked and reproduced in some cases</p>	<p>Tracked with the “Galaxy history” and reproducible with workflow</p>
FAIR principles	<p>Findable</p> <p>If properly shared</p>	<p>Web-based solution</p> <p>Unified system for data and software citation and attribution</p> <p>Tools can be made available on several servers</p> <p>Tools can be linked to tools registries and annotated with different ontologies</p> <p>Annotated workflows findable on WorkflowHub (https://workflowhub.eu) and Dockstore (https://dockstore.org)</p> <p>Free distribution of tools via the Galaxy ToolShed and workflows via WorkflowHub and Dockstore under an open-source licence</p>
Accessible	<p>If properly shared</p>	<p>Use different software, computational language and library versions on a single platform with the Conda package management system</p> <p>Workflows exportable in JSON and shareable through several standards (<i>e.g.</i> Common Workflow Language; Crusoe <i>et al.</i>, 2021 and Research Object Crate; Soiland-Reyes <i>et al.</i>, 2022)</p>
Interoperable	<p>When properly generalised, different elementary steps should be useable in interaction with each other</p>	<p>Tools, histories and workflows are re-executable, reusable and adaptable with different analytical procedure, parametrisation and/or inputs. Open-source code can be used outside of a Galaxy server</p>
Reusable	<p>Generalised elementary steps are reusable and adaptable with different analytical procedure, parametrisation and/or inputs</p> <p>The analytical procedure is clearer when properly atomised</p>	<p>Tools interface, workflow annotations, help sections and tutorials are a valuable help</p>
Technical and knowledge gaps	<p>Learning the analytical procedure design separately from computing languages, giving structure to trainees</p> <p>Reusability of elementary steps for trainees</p> <p>Need for a computation cluster if large data or demanding algorithm</p> <p>Achievable through collaborative code-editing applications</p>	<p>Experimenting with intricate analyses without computer code first</p> <p>Tutorials and videos from Galaxy Training Network (https://training.galaxyproject.org) - Galaxy community</p> <p>HPC (High Performance Computing) through an interface</p> <p>Bulk (meta)data manipulation</p> <p>With anyone through a Galaxy server</p>
Collaboration and attribution	<p>Analysis design and development</p> <p>Citation</p> <p>Easy reuse of openly shared elementary steps could lead to higher citation rates</p>	<p>Each tool, workflow, and tutorial are provided with a unique identifier for proper attribution and citation</p>

The Galaxy platform emphasises (i) accessibility of tools and data even without programming experience, (ii) reproducibility through the easy creation and reuse of analysis workflows, (iii) transparency through the open-source distribution of underlying codes; and (iv) community support.

For scientists, from a user's point of view, it offers extensive computing power and a graphical interface to use analysis workflows, even without experience in software development. Web-based access allows easy sharing of analytical workflows between collaborators and with a broader audience. Galaxy supports tools in almost any computational language, including R and Python, two of the most used languages in ecology, with many packages dedicated to ecological and biodiversity-oriented analyses incorporated (Lai *et al.*, 2019).

Anyone can use the tools on Galaxy and/or develop new tools and workflows to make them available to all by publishing them in the shared Galaxy ToolShed (<https://toolshed.g2.bx.psu.edu/>) which ensures that the tools and dependencies can be installed on any Galaxy servers. Any analytical procedure or workflow can be shared and enriched in parallel by several users, facilitating teamwork.

The platform is community-driven which permits continuous peer review of the platform and of the tools, workflows and tutorials provided. Many tutorials are available on the Galaxy Training Network (GTN; <https://training.galaxyproject.org/>) which is a valuable asset to the accessibility and reusability of tools and workflows (Batut *et al.*, 2018; Hiltemann *et al.*, 2023).

If enough researchers and experts start using and contributing to the platform, the number and content of available analytical procedures could expand at the same pace as latest analytical methodologies are integrated to research processes. If a different platform fits best and is more widely used by ecological and biodiversity scientific communities in the end, the work done on Galaxy will not be lost as tools are easily transposable to other interfaces (*e.g.* scripts directly usable with R, Python, etc., translation of workflows to other workflow engines).

Galaxy is ready to use and has proved its efficiency and suitability in other research fields, including genomics and climate science (Knijn *et al.*, 2020; Serrano-Solano *et al.*, 2022). Galaxy-Ecology has implemented workflows for biodiversity data exploration, eDNA processing, general population and community metrics and models, ecoregionalisation, NDVI (Normalised difference vegetation

index) computation with Sentinel-2 data among others (see some examples: <https://workflowhub.eu/workflows/657>) and tutorials for several of them are available on the GTN platform (see <https://training.galaxyproject.org/training-material/topics/ecology>).

In addition to using existing tools, users may develop and upload entirely new tools and workflows to the Galaxy server in any computational language to make them accessible to all other users.

Galaxy is a participative platform and several ways to participate to Galaxy exist depending on one's skills, available time, and needs. Anyone can participate to the Galaxy-Ecology initiative by:

- Sharing datasets, histories and workflows;
- Giving feedback on servers, tools, and workflows;
- Sharing tools and workflows ideas (eventually with code) through Git issues;
- Asking for tool modifications through issues;
- Modifying existing tools or proposing new tools through GitHub or GitLab;
- Writing or contributing to a GTN tutorial on a specific functionality or a workflow on the Galaxy Training Network platform;
- Create learning pathways, a set of tutorials curated by community experts to form a coherent set of lessons around a topic, building up knowledge (<https://training.galaxyproject.org/training-material/learning-pathways>);
- Propose training events and help users in the utilisation of a workflow and tutorial.

Analyses are rarely computed only once. Any analysis with a generalisation potential is a suitable candidate to be Galaxy-fied. A methodological framework is presented in online supplementary material (https://github.com/ColineRoyaux/Galaxy_Templates/blob/main/Methods/Methods%20-%20How%20to%20Galaxy-fy%20your%20analytical%20procedure_.md) at three levels depending on potential interests, computing language skills, and willingness to invest more or less time in the process: (i) 'user' relying on existing Galaxy tools and workflows to analyse data (lower time investment), (ii) 'developer' relying on existing and validated analytical procedure to develop Galaxy tools and

workflows (highest time investment), and (iii) ‘trainer’ relying on existing Galaxy tools to share workflows and create training material (variable time investment).

Discussion and limitations

There are many best practices and recommendations existing for analytical procedures, data management, and computational code development. The levels of application of these best practices fall within a continuum offering a range of possibilities from the sole sharing of processed and interpreted results with a brief description of methods to an executable paper published within a container and emulated virtual machine (Strijkers *et al.*, 2011; Grüning *et al.*, 2018). Situated somewhere in between the aforementioned extremes, the atomisation – generalisation framework and the utilisation of the Galaxy platform might represent viable solutions offering a satisfactory level of best practices.

Atomisation and generalisation of computer codes can represent a relatively low investment strategy to attain certain levels of best practices such as transparency and reusability. It also carries advantages such as easier peer review, modularity of analytical procedures and, consequently, time savings. Indeed, applying the framework is not sufficient to attain the highest levels of best practices. For reproducibility and transparency, the management of the environment, software and package versions can be hard to maintain and record. For example, on a local computer a comprehensive tracking of input, outputs and codes requires meticulous management of folder structure in the environment. Additionally, non-code developers will be able to partially review the analytical procedure only if the workflow is clearly outlined in an adapted format (*e.g.* table, graphical representation). Accessibility and findability of the atomised and generalised analytical procedure is dependent of its proper sharing (*e.g.* persistent link, open repository).

Galaxy can represent an easier gateway towards higher levels of best practice as sharing a complete, detailed and (re-)executable analytical procedure is facilitated through provenance tracking and automatic metadata enrichment. In comparison, many scientific workflow management systems, such as Snakemake, Nextflow or the R package Targets, operate from the command line. In ecology, numerous initiatives have tried to introduce such systems, starting with more user-friendly solutions. For example, the KNIME and Kepler systems with the CoESRA initiative (Collaborative Environment for Scholarly Research and Analysis) in

Australia; Taverna with the BioVeL initiative (Biodiversity Virtual e-Laboratory) in Europe; or very recently, the BON in a Box pipeline engine. These systems are more accessible to new users by offering a graphical interface while achieving high specificity (Berthold *et al.*, 2007; Hardisty *et al.*, 2016; <https://boninabox.geobon.org/>). However, good computer programming or scientific workflow management knowledge is still necessary to use these applications appropriately.

In comparison to the atomisation-generalisation framework, Galaxy can be rightfully seen as necessitating more time investment for scientists with programming experience as it requires to learn to use a new platform. Additionally, more effort may be required on Galaxy when an additional analytical step needs to be developed, but the Galaxy community can be an efficient crutch on which hard-pressed scientists can rely. Indeed, one can ask for help on the implementation of tools whether one knows computing languages and can share their code or not.

This note showcases a simple proposition to achieve best practices in analytical procedures with two plain guidelines: atomisation and generalisation. This straightforward framework represents a different manner to think and build analytical procedures; it doesn't require using a new technology or learning to use a new software. In terms of attaining higher levels of best practice, whether it is through the atomisation-generalisation framework, Galaxy, a combination of the two or otherwise, the optimal approach is to be determined by individuals depending on their interests, projects, and available resources. Relying on existing solutions as much as possible is, in our perspective, an efficient way to achieve a better understanding of best practices and their implications. Given the current environmental crisis, science has the major political and social responsibility to maintain good levels of transparency, reproducibility and efficiency.

Acknowledgements

Authors want to thank Sandrine Pavoine for its highly relevant and helpful advice and reviews on both the content and the form of the article. Authors are thankful to Thimothée Poisot (recommender), Nick Isaac (reviewer) and one anonymous reviewer for their advice during the Peer Community In review. Their help and suggestions on the structure and the content of the manuscript really helped to get the message of the article across in a more accessible manner.

Funding were provided by the European Union through the Erasmus+ Gallantries project; the Agence Nationale de la Recherche through the 65 Million d'Observateurs and the IA-Biodiv projects; the French National Fund for Open Science through the OpenMetaPaper project; the European commission through the H2020 EOSC-Pillar, GAPARS projects, and Horizon Europe FAIRE EASE project; the GO FAIR initiative through the BiodiFAIRse Implementation Network; the Blue Nature Alliance; and the Antarctic and Southern Ocean Coalition. Finally, funding by the French Ministry of Higher Education and Research were provided for the "Pôle national de données de biodiversité" e-infrastructure.

Encadré C: Les analyses de la thèse disponibles sur Galaxy ou en atomisé-généralisé

Vingt-six scripts R ont été développés avec le cadre de travail atomisé – généralisé, ils sont ou seront disponibles en ligne rapidement et seront adaptés en outils Galaxy à terme :

- https://github.com/ColineRoyaux/population_genetics
- <https://github.com/ColineRoyaux/cleanSIG>
- <https://github.com/ColineRoyaux/DM>

Trois outils Galaxy ont été développés dans le cadre de la thèse dont un est en ligne (https://usegalaxy.eu/root?tool_id=toolshed.g2.bx.psu.edu/repos/ecology/ab1_fastq_converter/ab1_fastq_converter/1.20.0), les autres seront partagés dans un temps court.

Deux workflows Galaxy réutilisables ont été développés dans le cadre de la thèse et au moins trois autres seront mis à disposition dans un temps court afin de réaliser des analyses de génétique des populations, du nettoyage de données SIG et de la modélisation statistique :

- https://usegalaxy.fr/u/coline_royaux/w/sanger1--from-ab1-to-aligned-fasta-imported-from-uploaded-file
- https://usegalaxy.eu/u/coline_royaux/w/sanger-5--from-hand-corrected-haplotypes-table-to-population-haplotypes-map-and-euclidean-distance-matrix

Un tutoriel Galaxy a été rédigé pour le nettoyage de données de séquençage (https://training.galaxyproject.org/training-material/topics/sequence-analysis/tutorials/Manage_AB1_Sanger/tutorial.html). D'autres tutoriels sont prévus pour les autres workflows Galaxy.

Discussion générale, conclusion et perspectives

I. Principaux résultats

Les systèmes insulaires tels que les îles et les mares sont depuis longtemps considérés comme des laboratoires de biodiversité (Whittaker *et al.*, 2023). Chaque élément des systèmes insulaires (une île ou une mare) a des caractéristiques propres et uniques qui en fait des environnements fascinants. Du fait de leur spécificité et de leur isolation, la biodiversité de ces milieux est particulièrement vulnérable. Pourtant, l'importance écologique des milieux ponctuels d'eau douce et des îles a été démontrée, notamment comme refuges pour certains groupes taxonomiques à des moments de grands bouleversements (Veron *et al.*, 2019; Heads, 2023). De fait, une meilleure connaissance de ces systèmes et de leur biodiversité est primordiale pour encourager leur protection de façon rationnelle et efficace.

Dans ce travail de thèse, plus de la moitié des espèces de crustacés (Branchiopodes, Calanoïdes) identifiées dans les mares de Nouvelle-Calédonie-Kanaky n'étaient pas signalées auparavant sur l'archipel (Chapitre I). Cette proportion démontre à elle seule le déficit de connaissances qui était existant sur ces groupes. Ce déficit existe toujours et de nombreuses espèces n'ont pas été détectées (Chapitre I et II).

L'aspect principal ressortant des résultats des deux premiers chapitres est la structuration de la biodiversité des crustacés d'eau douce en fonction des massifs ultramafiques observable des gènes aux assemblages d'espèces. Ce constat n'est pas surprenant, il est observé chez de nombreux groupes (Caesar *et al.*, 2017).

La diversité génétique semble plus élevée et géographiquement structurée sur les massifs ultramafiques en comparaison avec le reste de la NCK (Chapitre I). De plus, les mares présentes sur les sols ultramafiques semblent accueillir une plus grande proportion d'espèces endémiques voire micro-endémiques (Chapitre I et II). Ces constats pourraient résulter de la spécificité des sols ultramafiques qui entraîne une dispersion réduite chez les espèces de crustacés de ces milieux. Hors des massifs ultramafiques, l'endémisme existe mais est largement inférieur en proportion. Cependant, la diversité spécifique y est plus élevée que sur les sols ultramafiques (Chapitre II). En effet, la grande diversité de paysages non ultramafiques sur l'archipel ainsi que la connectivité qui a été démontrée avec l'Australie pourraient en partie expliquer ce constat. La grande distance séparant l'Australie de la NCK suggère que l'archipel néo-calédonien pourrait constituer un

système fortement isolé. Cependant, il semble que la région non-ultramafique de l'archipel constitue un système insulaire peu isolé de l'Australie tandis que la région ultramafique est un système insulaire isolé de l'Australie mais aussi partiellement de la région non-ultramafique. Dans ce contexte, la biodiversité des mares de NCK pourrait être qualifiée de système insulaire cloisonné. Au-delà du type de massif, le niveau de précipitations semble avoir un effet sur les assemblages d'espèces retrouvés dans les mares. Cependant, il est probable que les caractéristiques d'assèchement du milieu aient un effet plus direct sur les assemblages et la diversité. De plus, il semble que les assemblages d'espèces de l'archipel n'aient pas tous été échantillonnés du fait de la faible probabilité de présence des deux assemblages majoritaires sur les zones de frontière entre les types de massifs (Chapitre II).

Enfin, dans le dernier chapitre, les leviers utilisés tout au long du projet de recherche afin d'augmenter le niveau de transparence, d'accessibilité et de réutilisabilité des connaissances acquises sur les milieux, les espèces et les méthodes utilisées ont été détaillés. Par ce biais, il a été démontré qu'il était réalisable d'augmenter le niveau de bonnes pratiques de recherche avec des solutions nécessitant peu d'investissement temporel ou monétaire. De plus, les apports de chaque approche ont pu être mis en avant et peuvent bénéficier à de nombreux acteurs. Sur le long terme, la mise en œuvre de ces bonnes pratiques sont bénéfiques pour le projet et les personnes investies, et permettent de faciliter la mise en place d'un nouveau cycle de recherche.

II. Éléments de conservation

II. 1. Les milieux d'eau douce

Bien que l'eau douce soit une ressource vitale notamment aux populations humaines, la mise en place d'aires protégées (en dehors des Aires Marines Protégées) est davantage basée sur les écosystèmes terrestres (Lynch *et al.*, 2023). Cependant, les zones prioritaires pour les environnements aquatiques sont différentes (Herbert *et al.*, 2010). Lorsque les écosystèmes aquatiques sont pris en compte, les poissons et décapodes des rivières sont préférentiellement étudiés (Faghihinia *et al.*, 2021). La diversité des autres crustacés dans les eaux douces étant fortement influencée (souvent négativement) par la présence de poissons (Kuczyńska-Kippen & Pronin, 2018), la protection prioritaire des poissons est potentiellement délétère pour la majorité des crustacés dont ils sont les prédateurs. De plus, les mares sont des environnements d'eau douce moins étudiés que les lacs

ou les rivières (Faghihinia *et al.*, 2021; Hill *et al.*, 2021). Dans ces milieux fonctionnant en réseau, la prise en compte de la connectivité dans les décisions de conservation est primordiale pour la mise en place de mesures cohérentes et efficaces dans des régions stratégiques (*e.g.*, trame verte et bleue). De fait, la génétique des populations est un outil efficace pour évaluer la connectivité et identifier les barrières potentielles entre milieux ou les réseaux dégradés (Manel *et al.*, 2020). Le mode de dispersion des groupes de crustacés étudiés dans cette thèse en fait un modèle particulièrement intéressant pour estimer la connectivité entre milieux aquatiques ponctuels (Whittaker *et al.*, 2023).

Dans un temps court, la production d'une clé d'identification des espèces de crustacés des mares pourra être utile à l'expertise en conservation biologique sur l'archipel. Cette clé d'identification pourrait, à terme, être mise à disposition via l'application XPER3 sous forme de clé interactive plus accessible (Ung *et al.*, 2010). Ces groupes sont échantillonnables en grand nombre simplement à l'aide d'un filet à plancton ou d'une pelle dans le cas de l'étude des sédiments. Grâce aux œufs de résistance, l'étude des milieux temporaires n'est pas forcément conditionnée par la présence d'eau. Cependant, dans notre cas, l'étude de la connectivité nécessite d'obtenir des données sur les séquences ADN. La biologie moléculaire est encore peu utilisée notamment dans l'application des mesures de conservation par manque de moyens financiers et d'expertise (Haig *et al.*, 2016; Taylor *et al.*, 2017). Ces problématiques sont de moins en moins importantes du fait de l'amélioration significative des méthodes de reconstitution des séquences et du coût du séquençage qui diminue d'année en année. Notamment, l'utilisation de l'ADN environnemental devient de plus en plus répandue (Reid *et al.*, 2019; Beng & Corlett, 2020).

Les avantages de l'ADN environnemental sont indéniables (*e.g.*, rapide, non-destructive, non-invasive, large spectre de taxons, haut niveau de détection) mais cette méthode nécessite de s'appuyer sur des séquences de références d'individus formellement identifiés. Une alternative à l'ADN environnemental est l'ADN massal qui se base sur une extraction de plusieurs individus d'espèces différentes sans qu'un tri ne soit nécessaire. Toutefois, il sera impossible de détecter une espèce qui n'a jamais été séquencée auparavant. Il sera toujours nécessaire d'avoir une bonne connaissance des espèces du milieu pour l'étudier de cette façon et donc construire des bases de référence appuyées par des taxonomistes. Notamment, les résultats obtenus dans le Chapitre I ont permis d'alimenter ces bases de références à travers un travail associant morphologie et génétique.

II. 2. Les milieux ultramafiques

Le lien fort existant entre la présence de massifs ultramafique et la structuration de la biodiversité en NCK pose une problématique palpable vis-à-vis de l'industrie minière sur l'archipel. Ce phénomène d'endémisme n'est pas confiné à la NCK et est retrouvé dans les mares sur d'autres terrains riches en métaux lourds exploités par des mines. Notamment, au Brésil dans l'état de Minas Gerais (« Mines communes » en Portugais) et du Pará, trois espèces de Branchiopodes sont endémiques des mares ultramafiques, *Branchinecta ferrolimneta* à Nova Lima (Rogers & Ferreira, 2007), et *Dendrocephalus carajaensis*, et *D. xikrini* de la mine des Carajás (Rogers & Gomes, 2012; Rabet *et al.*, 2018). Il serait intéressant de comparer les deux situations pour mieux comprendre l'endémisme lié à la présence de métaux lourds dans les sols ainsi que les problèmes de conservation liés à l'exploitation minière. Sur le territoire de la NCK, la grande concentration de mares retrouvée sur des massifs ultramafiques au sud de Grande Terre à proximité de la mine de Goro (Jeanpert *et al.*, 2016) en fait un milieu particulièrement sujet aux perturbations. En effet, les mines ont de nombreux effets impactant la qualité des eaux ou encore la viabilité des milieux (*e.g.*, mares détruites par l'extraction ou remblayées par la fine poussière produite ; Royaux *et al.*, 2024). En effet, parmi les 11 espèces évaluées par l'IUCN (International Union for Conservation of Nature and Natural Resources) comme éteintes du fait de l'industrie minière, 4 étaient présentes sur des îles du Pacifique dont une en NCK (la plante *Pycnandra micrantha*). De plus, la pauvreté des substrats ultramafiques rend la restauration des environnements de ces régions beaucoup plus longue qu'ailleurs (Ung *et al.*, 2010).

Bien que les menaces pour la biodiversité semblent plus concentrées sur les massifs ultramafiques avec la présence de l'industrie minière, le reste de l'archipel n'est pas épargné par les perturbations anthropiques. Même si les zones urbaines avec des sols imperméabilisés sont peu répandues sur l'archipel, près de 40% du territoire est recouvert d'un couvert végétal dit « modifié » avec plus de 350 espèces introduites naturalisées. Cette végétation modifiée est majoritairement due à une grande quantité de feux (volontaires ou accidentels) et au développement de terres agricoles (Isnard & Jaffré, 2024). Il est donc important de mettre en place des mesures qui prennent en compte tout l'archipel en dissociant bien la situation en terrain ultramafique et non-ultramafique.

II. 3. Les méthodes d'analyses

Une technique de modélisation similaire à celle utilisée dans le chapitre II, a déjà été appliquée sur des espèces végétales micro-endémiques à l'échelle du mont Kaala sur Grande Terre afin de définir des zones d'intérêt pour la conservation (Lannuzel *et al.*, 2021). Les crustacés des mares pourraient être également utilisés pour ce type de modèles qui aide à la gestion de la conservation à une échelle plus fine. De plus, comme évoqué dans le chapitre II, la prise en compte de l'hydropériode est très complexe mais pourrait être réalisable dans un premier temps à petite échelle. En effet, il serait possible d'opérer des détections du niveau d'eau dans les mares avec la technologie LiDAR à plusieurs moments de l'année en parallèle de campagnes d'échantillonnage (Jeanpert *et al.*, 2016). A partir de ces données, une cartographie de l'hydropériode pourrait être inférée et utilisée afin de caractériser les liens entre les caractéristiques temporaires de mares et les espèces qui y sont identifiées.

Au-delà des crustacés des mares et même de la NCK, comme expliqué dans le chapitre III, les scripts analytiques produits ont été pensés pour être utilisables dans d'autres contextes géographiques ou temporels. A leur mise à disposition sur Galaxy avec un tutoriel adapté, les processus analytiques utilisés dans cette thèse pourront être appliqués par un plus large public (Batut *et al.*, 2018; Hiltemann *et al.*, 2023). Notamment, la méthode des modèles d'ensembles utilisée dans le chapitre II est depuis presque 30 ans recommandée comme outil pour le rapportage et la prise de décision de conservation (*e.g.*, évaluation des risques d'invasion biologique, identification de zones à haut potentiel, support à la mise en place de réserves ; Beerling *et al.*, 1995; Engler *et al.*, 2004; Araújo *et al.*, 2004; Guisan & Thuiller, 2005; Guisan *et al.*, 2013). Or, cette méthode est complexe, elle est très sensible aux biais et doit être interprétée avec prudence (Guillera-Arroita *et al.*, 2015). De ce fait, de nombreux cadres et outils ont été développés pour faciliter son utilisation (Thuiller *et al.*, 2009; Feng *et al.*, 2019b; Valavi *et al.*, 2021; Boyd *et al.*, 2023) mais de nombreux choix méthodologiques restent peu adressés par la littérature (Leroy, 2023). La mise en place d'un outil dédié à l'application de cette méthode pour différents publics avec plusieurs niveaux de complexité pourrait être un moyen d'élargir son utilisation. Ce type d'outil a été développé pour d'autres méthodes d'analyses notamment pour aider à la création d'Aires Marines Protégées (AMP) et appuyer le travail des gestionnaires d'AMP via l'application PAMPA (Pelletier *et al.*, 2014) utilisée notamment en NCK. Une partie du processus analytique de cette application a été adaptée sur la plateforme Galaxy par mes soins (Pelletier *et al.*,

2021). De plus, le développement d'un outil à plusieurs niveaux de complexité a déjà été testé à travers la plateforme Galaxy via le projet Vigie-nature école. Vigie-nature est un ensemble d'observatoires de science participative porté par le Muséum national d'Histoire naturelle (<https://www.vigienature.fr>). Ces observatoires portent sur de nombreux groupes tels que les oiseaux (Suivi Temporel des Oiseaux Communs ; STOC), la flore urbaine (STREETS) ou encore les chiroptères (Vigie-chiro) et visent des publics diversifiés (*e.g.*, naturalistes ou non, citadins ou non). Vigie-nature école est un projet qui vise plus particulièrement les publics scolaires avec des protocoles adaptés. Dans le cadre de ce projet, des outils ont été développés sur les serveurs de Galaxy pour proposer une plateforme d'analyse de données aux collégiens et lycéens participant aux observatoires. Un processus analytique bien défini et une interface adaptée a été pensée pour ces publics à travers l'outil Galaxy Bricks (<https://www.vigienature-ecole.fr/bricks>).

III. Perspectives

III. 1. Partage des connaissances et publications

De nombreuses perspectives ont émergé des travaux effectués pendant cette thèse. Tout d'abord, sur un temps court, huit espèces sont à décrire formellement et une liste actualisée des espèces de crustacés d'eau douce présentes sur l'archipel devra être publiée. De plus, les analyses de génétiques de chaque groupe présenté dans le Chapitre I seront également à soumettre sous forme d'article. Comme évoqué dans l'encadré B, les modèles de distribution des espèces basées sur la présence des œufs de résistance sont à perfectionner.

En parallèle, la finalisation des outils et flux automatisés (workflows) Galaxy amorcés pour les analyses réalisées dans les trois publications des Chapitres I et II ainsi que l'écriture de plusieurs tutoriels pour leur utilisation seront réalisées rapidement. Ces productions permettront d'augmenter significativement l'accessibilité et la reproductibilité des travaux effectués. Elles pourront également aider à augmenter la réutilisation des processus analytiques pour d'autres contextes de recherche mais aussi par d'autres publics.

Etant donné la multitude d'espèces non répertoriées sur l'archipel qui ont été détectées dans cette thèse, il serait important de faire suivre les informations sur la plateforme du GBIF (Global Biodiversity Information Facility) qui répertorie les occurrences d'espèces à l'échelle mondiale. De plus, au niveau français, l'INPN (Inventaire National du Patrimoine Naturel) produit un référentiel taxonomique

régulièrement mis à jour qui devrait intégrer les espèces étudiées ici. Cependant, des doublons d'occurrence peuvent arriver entre les échantillons déposés en collections, les données mises sur un entrepôt de données comme Data.InDoRES, celles sur la plateforme GenBank et celles sur l'INPN et le GBIF. De ce fait, des échanges ont été amorcés avec ces différents systèmes d'information afin d'avoir une gestion le plus simple possible de ces interdépendances (Norvez *et al.*, 2023).

III. 2. Études complémentaires

L'identification formelles des espèces d'Ostracodes, d'Harpacticoides et de Cyclopoïdes pourraient être réalisées à travers des collaborations avec des chercheurs et chercheuses spécialisées sur ces groupes telles que Maria Hołyńska en Pologne, Koen Martens en Belgique ou encore Geoff Boxshall en Angleterre. De plus, une révision taxonomique des Chydoridae est à envisager en collaboration notamment avec Brian Timms en Australie.

Le scénario idéal pour évaluer la validité du modèle de probabilité de présence de chaque assemblage d'espèce serait d'aller vérifier avec des échantillonnages indépendants la distribution de ces assemblages. En complément des milieux étudiés sur l'archipel dans cette thèse, il serait intéressant d'échantillonner davantage dans les régions où les modèles prédisent de basses probabilités de présence des deux assemblages. Ceci pourrait permettre de vérifier si ces régions consistent en un assemblage distinct ou des zones de transition entre les deux types d'assemblages. Le gradient environnemental pourrait être caractérisé dans les mares avec des dosages de la matière organique en suspension ou encore la concentration en métaux lourds. La théorie d'un assemblage distinct pourrait être appuyée par la distribution de *B. sibletii* qui a été observé exclusivement dans des mares ultramafiques à forte concentration en matière organique et à l'observation d'un œuf d'une potentielle nouvelle espèce d'*Australimnadia* dans ce type de milieu. En ajoutant des récoltes, il serait aussi possible de statuer sur la nature des assemblages mineurs (Canala et zone de vie de la mine de Goro au sud-ouest de Grande Terre) déconnectés des deux assemblages majeurs. De plus, certaines zones ont été insuffisamment prospectées comme les zones ultramafiques du nord ou d'altitude (Maquis de haute altitude).

Les modèles calculés et présentés dans le Chapitre II sont régulièrement utilisés pour projeter l'aire de répartition supposée des espèces ou assemblages dans le futur avec différents scénarios de changements climatiques (Buisson & Lek,

2010; Zurell *et al.*, 2020). Il serait effectivement intéressant de pouvoir estimer les potentielles modifications des aires environnementales favorables aux assemblages d'espèces dans le futur en fonction de différents scénarios climatiques. Lorsqu'elles sont de bonne qualité, ce type de projection permet d'anticiper les changements climatiques pour la création d'aires protégées qui seront efficaces sur le long terme (Guisan *et al.*, 2013; Pollock *et al.*, 2020; Verniest, 2023).

L'évaluation des caractéristiques d'assèchement et de l'hydropériode demandent un investissement lourd en moyens humains, financiers et temporels. Or, ces caractéristiques conditionnent les espèces et assemblages de crustacés qui sont observés dans les milieux (Tavernini, 2008; Seminara *et al.*, 2015; Kuczyńska-Kippen & Pronin, 2018; Coccia *et al.*, 2024). L'étude approfondie du lien entre ces caractéristiques et la présence de certaines espèces ou assemblages pourrait permettre de faciliter l'étude de l'assèchement et de la mise en eau d'un milieu. D'après Hill *et al.* (2021), l'étude des successions est un sujet de recherche lacunaire et prioritaire en écologie des mares. En effet, la présence d'un assemblage ou d'une espèce donnée pourrait suffire à estimer l'hydropériode et les caractéristiques d'assèchement d'une localité.

Une étude approfondie des temps de diapause des différentes espèces en fonction de plusieurs facteurs tels que le temps d'assèchement, le temps en eau, la température ou encore la luminosité pourrait constituer un avancement majeur pour comprendre l'écologie des organismes étudiés et leur futur. En effet, une connaissance fine des facteurs activateurs ou inhibiteurs des différentes étapes de la diapause associée à des modèles de distribution présents et futurs pourrait permettre de prévenir de potentielles extinctions ou invasions. Par exemple, dans l'hexagone, les œufs de résistance de l'espèce *Eubbranchipus grubii* nécessitent une période de gel pour éclore (analogue à la vernalisation chez certaines plantes ; Mossin, 1986). Sachant cela, cette espèce risque de disparaître dans les années à venir du fait du réchauffement des températures pendant les périodes hivernales causant un raccourcissement voire une disparition des périodes de gel dans certaines régions. Cette caractéristique des œufs est essentielle à prendre en compte si l'on souhaite préserver l'espèce.

En complément, l'étude des caractéristiques des structures de résistances (œufs, ephippies, adultes enkystés) pourrait permettre d'évaluer les capacités de dispersion des espèces étudiées et leur potentielle résilience face aux changements climatiques. En effet, Timms *et al.* (2009), ont mis en avant les caractéristiques des

œufs d'espèces de *Parartemia* à s'accoler au sédiment, limitant leurs capacités de dispersion et pouvant expliquer l'endémisme récurrent retrouvé chez ce genre. En NCK, ce type d'étude pourrait donner davantage d'éléments pour expliquer les différences observées entre la répartition de la diversité génétique chez les espèces présentes sur les massifs ultramafiques, hors de ceux-ci ou sur les deux.

A d'autres fins, il est également possible de projeter les aires de répartition supposées dans le passé. Par exemple, Poncet *et al.*, (2013) ont pu estimer l'aire favorable à l'espèce de plante endémique de NCK *Amborella trichopoda* au dernier maximum glaciaire (-21 000 ans) à l'aide de cette méthodologie. Entre autres perspectives, cela pourrait permettre de répondre partiellement aux questionnements évoqués dans le second article du Chapitre I quant à la distribution passée du genre *Boeckella*. Cependant, il sera important de prendre en compte que l'hypothèse selon laquelle les variables environnementales préférées actuellement par un taxon étaient les mêmes dans un passé lointain est partiellement fausse.

III. 3. Au-delà des modèles d'études

En se basant sur des modèles bien définis tels que les crustacés d'eau douce, les mares, ou encore la Nouvelle-Calédonie-Kanaky, il a été possible de dessiner des cadres de travail qui peuvent s'étendre à d'autres contextes taxonomiques ou géographiques. Notamment, le chapitre III de cette thèse a permis de mettre en place des cadres pour la réutilisation des connaissances dans un contexte d'étude de la biodiversité.

Premièrement, la distribution presque ubiquiste des mares d'eau douce sur la surface terrestre en fait un milieu étudiable dans la majorité des biotopes et à toutes les échelles. L'ensemble des méthodes et analyses utilisées dans cette thèse pourraient s'appliquer ailleurs dans le monde. En effet, les micro-crustacés d'eau douce restent des groupes méconnus à l'échelle mondiale et le nombre d'espèces décrites reste à ce jour trop limité pour nous permettre d'atteindre, d'ici 2100, un nombre d'espèces connues proche de la réalité (Macedo *et al.*, 2024). L'hexagone ne fait pas exception et les groupes étudiés pourraient être encore moins connus qu'en NCK. En effet, si on considère l'exemple des collections du MNHN, seuls 122 occurrences sont répertoriées dans les bases de données en France hexagonale contre 361 en NCK.

Au-delà d'assemblages d'espèces séparés géographiquement, les successions annuelles ou pluri-annuelles retrouvées dans les milieux temporaires et permanents permettent l'étude des assemblages d'espèces séparés par le temps dans une même localité géographique (Montaña *et al.*, 2022). Des communautés distinctes en été et en hiver ont été retrouvées notamment en Camargue et en Angleterre (Nourisson & Aguesse, 1961; Hill *et al.*, 2016). De la même manière que les bordures géographiques entre deux habitats montrent régulièrement une richesse spécifique plus élevée (Willmer *et al.*, 2022), il a été observé une tendance similaire dans ces mares à travers le temps avec une richesse plus élevée à l'automne. Cette dimension temporelle de la composition des assemblages complexifie l'échantillonnage, la détection des espèces et les analyses qui en découlent.

Au-delà des assemblages actuels, les micro-crustacés des mares sont des modèles applicables pour l'étude des environnements passé. En effet, des fragments d'individus et des œufs de résistance de Branchiopodes ont déjà été identifiés dans des paléo-sédiments ou encore dans les poils de grands mammifères du Pléistocène (Kotov *et al.*, 2019; Rogers *et al.*, 2021). Les œufs de résistance permettent de conserver l'ADN pendant une très longue période pouvant, sous les bonnes conditions, aller jusqu'à plusieurs milliers d'années (*e.g.*, séquences de *Boeckella poppei* extraites d'œufs datés à -9 000 ans ; Bissett *et al.*, 2005). Sans aller aussi loin dans le temps, la recherche d'œufs de résistance dans des sédiments plus profonds que ce qui a été étudié ici pourrait permettre de reconstituer la diversité génétique et les assemblages d'espèces présentes 30 ans en arrière, avant la révolution industrielle ou même avant l'arrivée du capitaine Cook en NCK. Ces informations représentent des éléments cruciaux pour ces groupes dont les traces historiques ne sont que très parcellaires. De plus, chez certains groupes et selon les conditions du milieu, les œufs de résistances peuvent éclore après plusieurs décennies, voire plusieurs siècles (*e.g.*, éclosion d'œufs de copépode datés à 330 ans ; Hairston *et al.*, 1995). En France, des traces de *Lepidurus articus* retrouvée et les vestiges des animaux dans les paléo-sédiments ont permis d'affiner les connaissances sur le climat local entre -30 000 et -20 000 ans (Ponel, 1994).

Références

A

- Adamowicz S.J., Menu-Marque S., Halse S.A., Topan J.C., Zemplak T.S., Hebert P.D.N., Witt J.D.S. The evolutionary diversification of the Centropagidae (Crustacea, Calanoida): A history of habitat shifts. *Molecular Phylogenetics and Evolution* 2010; **55**: 418–430.
- Affholder O., Antoine P.O., Beck R.M.D. The “Diahot Tooth” is a Miocene rhinocerotid fossil brought by humans to New Caledonia. *Journal of Mammalian Evolution* 2024; **31**: 1–13.
- Alekseev V.R. Revision of the genus Eucyclops (Claus, 1893) and subfamily Eucyclopinae of the world fauna. *Arthropoda Selecta* 2019; **28**: 490–514.
- Alekseev V.R. Revision Of The Genus Mesocyclops Sars 1914 (Copepoda, Cyclopidae) Of The World Fauna. *Zoologicheskii Zhurnal* 2020; **99**: 1323–1344.
- Alfonso G., Belmonte G. Expanding distribution of *Boeckella triarticulata* (Thomson, 1883) (Copepoda: Calanoida: Centropagidae) in Southern Italy. *Aquatic Invasions* 2008; **3**: 247–251.
- Allouche O., Tsoar A., Kadmon R. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* 2006; **43**: 1223–1232.
- Amir H., Cavaloc Y., Crossay T., Bourles A., Gensous S., Lagrange A., Burtet-Sarramegna V., Guentas L. Importance and roles of arbuscular mycorrhizal fungi in New Caledonian ultramafic soils. *Botany Letters* 2023; **170**: 449–458.
- Araújo M.B., Anderson R.P., Barbosa A.M., Beale C.M., Dormann C.F., Early R., Garcia R.A., Guisan A., Maiorano L., Naimi B., O’Hara R.B., Zimmermann N.E., Rahbek C. Standards for distribution models in biodiversity assessments. *Science Advances* 2019; **5**: 1–12.

Araújo M.B., Cabeza M., Thuiller W., Hannah L., Williams P.H. Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. *Global Change Biology* 2004; **10**: 1618–1626.

Archmiller A.A., Johnson A.D., Nolan J., Edwards M., Elliott L.H., Ferguson J.M., Iannarilli F., Vélez J., Vitense K., Johnson D.H., Fieberg J. Computational Reproducibility in The Wildlife Society’s Flagship Journals. *Journal of Wildlife Management* 2020; **84**: 1012–1017.

B

Bagella S., Gascón S., Caria M.C., Sala J., Mariani M.A., Boix D. Identifying key environmental factors related to plant and crustacean assemblages in Mediterranean temporary ponds. *Biodiversity and Conservation* 2010; **19**: 1749–1768.

Bandelt H.J., Forster P., Rohl A. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 1999; **16**: 37–48.

Barber R.A., Ball S.G., Morris R.K.A., Gilbert F. Target-group backgrounds prove effective at correcting sampling bias in Maxent models. *Diversity and Distributions* 2022; **28**: 128–141.

Bargier N., Dominique Y., Humbert J.F., Parpet J.F., Lacroix G., Jeanpert J., Genthon P., Fare A., Vendé-Leclerc M. *Diagnose des dolines de Nouvelle-Calédonie*.

Bastian M., Heymann S., Jacomy M. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media, ICWSM 2009* 2009;: 361–362.

Batterbury S.P.J., Kowasch M. Introduction: Geographical Understanding and “Listening” in New Caledonia-Kanaky. *Geographies of New Caledonia-Kanaky* 2024; 1–9.

- Batut B., Hiltemann S., Bagnacani A., Baker D., Bhardwaj V., Blank C., Bretaudeau A., Brillet-Guéguen L., Čech M., Chilton J., Clements D., Doppelt-Azeroual O., Erxleben A., Freeberg M.A., Gladman S., Hoogstrate Y., Hotz H.R., Houwaart T., Jagtap P., Larivière D., Le Corguillé G., Manke T., Mareuil F., Ramírez F., Ryan D., Sigloch F.C., Soranzo N., Wolff J., Videm P., Wolfien M., Wubuli A., Yusuf D., Taylor J., Backofen R., Nekrutenko A., Grüning B. Community-Driven Data Analysis Training for Biology. *Cell Systems* 2018; **6**: 752-758.e1.
- Bayly I.A.E. A Revision Of The Australasian Species Of The Freshwater Genera Boeckella And Hemiboeckella (Copepoda: Calanoida). 1964
- Bayly I.A.E. A new species of Boeckella (Copepoda: Calanoida) and additional comments on some other species of the genus. *Papers and Proceedings of the Royal Society of Tasmania* 1967; **101**: 97–102.
- Bayly I.A.E. A new species of Boeckella (Copepoda: Calanoida) from Western Australia, and comments on two other congeners. *Records of the West Australian Museum* 1985; **12**: 79–84.
- Bayly I.A.E. *The non-marine Centropagidae (Copepoda, Calanoida) of the world*. The Hague, Netherlands: SPB Academic Publishing.
- Bayly I. Fusion of the genera Boeckella and Pseudoboeckella (Copepoda) and revision of their species from South America and sub-Antarctic islands. *Revista Chilena de Historia Natural* 1992b; **65**: 17–63.
- Bayly I.A.E., Morton D.W. Aspects of the zoogeography of Australian microcrustaceans. *SIL Proceedings* 1978; **20**: 2537–2540.
- Beck R.M.D. A dated phylogeny of marsupials using a molecular supermatrix and multiple fossil constraints. *Journal of Mammalogy* 2008; **89**: 175–189.
- Beerling D.J., Huntley B., Bailey J.P. Climate and the distribution of *Fallopia japonica*: use of an introduced species to test the predictive capacity of response surfaces. *Journal of Vegetation Science* 1995; **6**: 269–282.
- Bellec L., Rabet N. Dating of the Limnadiidae family suggests an American origin of Eulimnadia. *Hydrobiologia* 2016; **773**: 149–161.
- Beng K.C., Corlett R.T. *Applications of environmental DNA (eDNA) in ecology and conservation: opportunities, challenges and prospects*. Springer Netherlands.

- Berman M., Andersen A.N. New Caledonia has a depauperate subterranean ant fauna, despite spectacular radiations above ground. *Biodiversity and Conservation* 2012; **21**: 2489–2497.
- Berry R.J. Diversity and Differentiation: The Importance of Island Biology for General Theory. *Oikos* 1983; **41**: 523.
- Berthold M.R., Cebron N., Dill F., Gabriel T.R., Kötter T., Meinel T., Ohl P., Sieb C., Thiel K., Wiswedel B. KNIME: The Konstanz Information Miner. *Studies in Classification, Data Analysis, and Knowledge Organization* 2007; 319–326.
- Biggs J., Williams P., Whitfield M., Nicolet P., Weatherby A. 15 Years of pond assessment in Britain: Results and lessons learned from the work of Pond Conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems* 2005; **15**: 693–714.
- Bissett A., Gibson J.A.E., Jarman S.N., Swadling K.M., Cromer L. Isolation, amplification and identification of ancient copepod DNA from lake sediments. *Limnology and Oceanography* 2005; 533–542.
- Bloomfield N.J., Knerr N., Encinas-Viso F. A comparison of network and clustering methods to detect biogeographical regions. *Ecography* 2018; **41**: 1–10.
- Boix D., Sala J., Moreno-Amich R. The faunal composition of Espolla pond (NE Iberian Peninsula): The neglected biodiversity of temporary waters. *Wetlands* 2001; **21**: 577–592.
- Borgman C.L. *Qu'est-ce que le travail scientifique des données ?* OpenEdition Press.
- Boula A., Laporte-Magoni C., Gunkel-Grillon P., Bour O., Selmaoui-Folcher N. Potential contamination of stream waters by ultramafic mining sediments: Identification of geochemical markers (New Caledonia). *Journal of Geochemical Exploration* 2022; **232**: 106879.
- Boyd R.J., August T.A., Cooke R., Logie M., Mancini F., Powney G.D., Roy D.B., Turvey K., Isaac N.J.B. An operational workflow for producing periodic estimates of species occupancy at national scales. *Biological Reviews* 2023; **98**: 1492–1508.
- Brady, G.S. 1875. Note on Entomostraca from Kerguelen's Land and the south Indian Ocean. *Annals and Magazine of Natural History*, **16**: 162.
- Brady, G.S. 1906. On the Entomostracan Fauna of the New Zealand Lakes. *Proceedings of the Zoological Society of London*, 1906: 696–697.
- Breiman L. Random forests. *Machine Learning* 2001; **45**: 5–32.

- Brendonck L., Pinceel T., Ortells R. Dormancy and dispersal as mediators of zooplankton population and community dynamics along a hydrological disturbance gradient in inland temporary pools. *Hydrobiologia* 2017; **796**: 201–222.
- Brock M.A., Nielsen D.L., Shiel R.J., Green J.D., Langley J.D. Drought and aquatic community resilience: the role of eggs and seeds in sediments of temporary wetlands. *Freshwater Biology* 2003; **48**: 1207–1218.
- Brown J.H. Why are there so many species in the tropics? *Journal of Biogeography* 2014; **41**: 8–22.
- Bryant D.M. Effects of prey density and site character on estuary usage by overwintering waders (Charadrii). *Estuarine and Coastal Marine Science* 1979; **9**: 369–384.
- Brysiewicz A., Sługocki, Wesołowski P., Czerniawski R. Zooplankton community structure in small ponds in relation to fish community and environmental factors. *Applied Ecology and Environmental Research* 2017; **15**: 929–941.
- Buisson L., Grenouillet G., Casajus N., Lek S. Predicting the Potential Impacts of Climate Change on Stream Fish Assemblages. *American Fisheries Society Symposium* 2010; **73**: 327–346.
- Caesar M., Grandcolas P., Pellens R. Outstanding micro-endemism in New Caledonia: More than one out of ten animal species have a very restricted distribution range. *Plos One* 2017; **12**: e0181437.

C

- Carroll S.R., Garba I., Figueroa-Rodríguez O.L., Holbrook J., Lovett R., Materechera S., Parsons M., Raseroka K., Rodriguez-Lonebear D., Rowe R., Sara R., Walker J.D., Anderson J., Hudson M. The CARE principles for indigenous data governance. *Data Science Journal* 2020; **19**: 1–12.
- Carroll S.R., Herczog E., Hudson M., Russell K., Stall S. Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Scientific Data* 2021 *8:1* 2021; **8**: 1–6.
- Casajus N. {rcompendium} {An} {R} package to create a package or research compendium structure. 2023.

- Cervantes-Martínez A., Gutiérrez-Aguirre M.A., Suárez-Morales E., Jaime S. Phenetic and Genetic Variability of Continental and Island Populations of the Freshwater Copepod *Mastigodiatomus ha Cervantes*, 2020 (Copepoda): A Case of Dispersal? *Diversity 2021, Vol. 13, Page 279* 2021; **13**: 279.
- Chao A., Hu K.H., Chao A., Henderson P.A., Chiu C.H., Hu F. A Quick Introduction to iNEXT.3D via Examples HOW TO RUN iNEXT.3D. 2024.
- Chao A., Kubota Y., Zelený D., Chiu C.H., Li C.F., Kusumoto B., Yasuhara M., Thorn S., Wei C.L., Costello M.J., Colwell R.K. Quantifying sample completeness and comparing diversities among assemblages. *Ecological Research* 2020; **35**: 292–314.
- Chazeau J. Research on New Caledonian Terrestrial Fauna: Achievements and Prospects. *Biodiversity Letters* 1993; **1**: 123.
- Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 2016; **13-17-August-2016**: 785–794.
- Chen X.Y., He F. Speciation and endemism under the model of island biogeography. *Ecology* 2009; **90**: 39–45.
- Chevillotte V., Chardon D., Beauvais A., Maurizot P., Colin F. Long-term tropical morphogenesis of New Caledonia (Southwest Pacific): Importance of positive epeirogeny and climate change. *Geomorphology* 2006; **81**: 361–375.
- CITES. Convention sur le commerce international des espèces de faune et de flore sauvages menacées d'extinction. 1973.
- Coccia C., Almeida B.A., Badosa A., Diniz L.P., Brendonck L., Frisch D., Green A.J. Hydroperiod length, not pond age, determines zooplankton taxonomic and functional diversity in temporary ponds. *Ecological Indicators* 2024; **159**: 111632.
- Cohen-Boulakia S., Belhajjame K., Collin O., Chopard J., Froidevaux C., Gaignard A., Hinsén K., Larmande P., Bras Y. Le, Lemoine F., Mareuil F., Ménager H., Pradal C., Blanchet C. Scientific workflows for computational reproducibility in the life sciences: Status, challenges and opportunities. *Future Generation Computer Systems* 2017; **75**: 284–298.
- Condamine F.L., Leslie A.B., Antonelli A. Ancient islands acted as refugia and pumps for conifer diversity. *Cladistics* 2017; **33**: 69–92.
- Convention on Biological Diversity - Global Biodiversity Outlook 5*. 2020.

Crusoe M.R., Abeln S., Iosup A., Amstutz P., Chilton J., Tijanić N., Ménager H., Soiland-Reyes S., Goble C. Methods Included: Standardizing Computational Reuse and Portability with the Common Workflow Language. *Communications of the ACM* 2021; **65**: 54–63.

Culina A., van den Berg I., Evans S., Sánchez-Tójar A. Low availability of code in ecology: A call for urgent action. *PLOS Biology* 2020; **18**: e3000763.

D

Daday, E 1901. Diagnoses praecursoriae Copepodorum novorum e Patagonia. *Természetrájsz Füzetek*, **24**: 345–350.

Darwin C. *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*. London: John Murray, 1859.

De Meester L. Local genetic differentiation and adaptation in freshwater zooplankton populations: Patterns and processes. *Écoscience* 1996; **3**: 385–399.

De Meester L., Declerck S.A.J., Ger K.A. Beyond Daphnia: a plea for a more inclusive and unifying approach to freshwater zooplankton ecology. *Hydrobiologia* 2023; **850**: 4693–4703.

Defaye D. Description of the First Boeckella (Copepoda , Calanoida , Centropagidae) from New Caledonia. *Crustacean* 1998; **71**: 686–699.

De Guerne, J. & Richard, J. 1889. Révision des Calanides d'eau douce. *Mémoires de la Société Zoologique de France*. **2**: 1–129.

Dereeper A., Guignon V., Blanc G., Audic S., Buffet S., Chevenet F., Dufayard J.F., Guindon S., Lefort V., Lescot M., Claverie J.M., Gascuel O. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research* 2008; **36**: W465–W469.

Desoutter L., Bertaud A. *Bilan technique 2018 : Synthèse annuelle des résultats des suivis environnementaux du Grand Sud pour l'année 2018 / Oeil Nouvelle-Calédonie*. Nouméa.

Di Tommaso P., Chatzou M., Floden E.W., Barja P., Palumbo E., Notredame C. Nextflow enables reproducible computational workflows. *Nature Biotechnology* 2017; **35**: 316–319.

- Doerr V.A.J., Barrett T., Doerr E.D. Connectivity, dispersal behaviour and conservation under climate change: a response to Hodgson et al. *Journal of Applied Ecology* 2011; **48**: 143–147.
- Downing J.A., Prairie Y.T., Cole J.J., Duarte C.M., Tranvik L.J., Striegl R.G., McDowell W.H., Kortelainen P., Caraco N.F., Melack J.M., Middelburg J.J. The global abundance and size distribution of lakes, ponds, and impoundments. *Limnology and Oceanography* 2006; **51**: 2388–2397.
- Dray S., Dufour A.B. The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software* 2007; **22**: 1–20.
- Drew L.W. Are We Losing the Science of Taxonomy? As need grows, numbers and training are failing to keep up. *BioScience* 2011; **61**: 942–946.
- Duan B., Kang T., Wan H., Liu W., Zhang F., Mu S., Guan Y., Li Z., Tian Y., Kang X. Microsatellite markers reveal genetic diversity and population structure of *Portunus trituberculatus* in the Bohai Sea, China. *Scientific Reports* 2023 *13:1* 2023; **13**: 1–12.
- Dubois J., Launay J., Recy J. Uplift movements in New Caledonia-Loyalty Islands area and their plate tectonics interpretation. *Tectonophysics* 1974; **24**: 133–150.
- Duchêne S., Archer F.I., Vilstrup J., Caballero S., Morin P.A. Mitogenome Phylogenetics: The Impact of Using Single Regions and Partitioning Schemes on Topology, Substitution Rate and Divergence Time Estimation. *PLOS ONE* 2011; **6**: e27138.
- Dussart B.H. *Parathalassius fagesi* gen. et sp. nov. (Centropagidae), copépode nouveau de Nouvelle-Calédonie. *Cahiers de Biologie Marine* 1986; **27**: 63–68.

E

- El-moor Loureiro L.M.A. Manual de Identificação de Cladóceros Límnicos do Brasil. 1997.
- Ellemers N. Science as collaborative knowledge generation. *British Journal of Social Psychology* 2021; **60**: 1–28.
- EMBL Australia Bioinformatics Resource. *Community Survey Report* 2013 <https://www.embl-abr.org.au/news/braembl-community-survey-report-2013/>.

- Emery N.C., Crispo E., Supp S.R., Farrell K.J., Kerkhoff A.J., Bledsoe E.K., O'Donnell K.L., McCall A.C., Aiello-Lammens M.E. Data Science in Undergraduate Life Science Education: A Need for Instructor Skills Training. *BioScience* 2021; **71**: 1274–1287.
- van den Ende C., White L.T., van Welzen P.C. The existence and break-up of the Antarctic land bridge as indicated by both amphi-Pacific distributions and tectonics. *Gondwana Research* 2017; **44**: 219–227.
- Engler R., Guisan A., Rechsteiner L. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology* 2004; **41**: 263–274.
- Enright N.J., Rigg L., Jaffré T. Environmental controls on species composition along a (maquis) shrubland to forest gradient on ultramafics at Mont Do, New Caledonia. *South African Journal of Science* 2001; **97**: 573–580.
- European Commission, Directorate-General for Research and Innovation. Cost-benefit analysis for FAIR research data: cost of not having FAIR research data. *Publications Office* 2018

F

- Faghihinia M., Xu Y., Liu D., Wu N. Freshwater biodiversity at different habitats: Research hotspots with persistent and emerging themes. *Ecological Indicators* 2021; **129**: 107926.
- Fairbridge W.S. Three new species of *Boeckella* with an account of the developmental stages of *B. opaqua* n. sp. *Journal of The Royal Society of Western Australia* 1945
- Fanelli D. Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences of the United States of America* 2018; **115**: 2628–2631.
- Fang F.C., Casadevall A. Competitive Science: Is Competition Ruining Science? *Infection and Immunity* 2015; **83**: 1229–1233.
- Farley S.S., Dawson A., Goring S.J., Williams J.W. Situating Ecology as a Big-Data Science: Current Advances, Challenges, and Solutions. *BioScience* 2018; **68**: 563–576.
- Feld C.K., Birk S., Eme D., Gerisch M., Hering D., Kernan M., Maileht K., Mischke U., Ott I., Pletterbauer F., Poikane S., Salgado J., Sayer C.D., Van Wichelen J.,

- Malard F. Disentangling the effects of land use and geo-climatic factors on diversity in European freshwater ecosystems. *Ecological Indicators* 2016; **60**: 71–83.
- Feng X., Park D.S., Walker C., Peterson A.T., Merow C., Papeş M. ecological niche models. *Nature Ecology & Evolution* 2019a; **3**.
- Feng X., Park D.S., Walker C., Peterson A.T., Merow C., Papeş M. A checklist for maximizing reproducibility of ecological niche models. *Nature Ecology & Evolution* 2019 3:10 2019b; **3**: 1382–1395.
- Fernández-Palacios J.M., Kreft H., Irl S.D.H., Norder S., Ah-Peng C., Borges P.A.V., Burns K.C., de Nascimento L., Meyer J.Y., Montes E., Drake D.R. Scientists' warning – The outstanding biodiversity of islands is in peril. *Global Ecology and Conservation* 2021; **31**: e01847.
- Field B., Booth A., Ilott I., Gerrish K. Using the Knowledge to Action Framework in practice: a citation analysis and systematic review. *Implementation science: IS* 2014; **9**: 172.
- Folmer O., Black M., Hoeh W., Lutz R., Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular marine biology and biotechnology* 1994; **3**: 294–299.
- Fontaneto D. Long-distance passive dispersal in microscopic aquatic animals. *Movement Ecology* 2019; **7**: 10.
- Franklin J., Miller J.A. *Mapping Species Distributions: Spatial Inference and Prediction* 2010; 1–320.
- Fryer G. Crustacean diversity in relation to the size of water bodies: some facts and problems. *Freshwater Biology* 1985; **15**: 347–361.

G

- Gaiji S., Chavan V., Ariño A.H., Otegui J., Hobern D., Sood R., Robles E. Content assessment of the primary biodiversity data published through GBIF network: Status, challenges and potentials. *Biodiversity Informatics* 2013; **8**: 94–172.
- Gajbe S.B., Tiwari A., Gopalji, Singh R.K. Evaluation and analysis of Data Management Plan tools: A parametric approach. *Information Processing & Management* 2021; **58**: 102480.

- Gardner R.C., Finlayson C. State of the World's Wetlands and Their Services to People By. *Ramsar Convention Secretariat* 2020; 2020–2025.
- Garrigue C., Bonneville C.D., Cleguer C., Oremus M. Extremely Low mtDNA Diversity and High Genetic Differentiation Reveal the Precarious Genetic Status of Dugongs in New Caledonia, South Pacific. *Journal of Heredity* 2022; **113**: 516–524.
- Gaudeul M., Gardner M.F., Thomas P., Ennos R.A., Hollingsworth P.M. Evolutionary dynamics of emblematic *Araucaria* species (Araucariaceae) in New Caledonia: Nuclear and chloroplast markers suggest recent diversification, introgression, and a tight link between genetics and geography within species. *BMC Evolutionary Biology* 2014; **14**: 1–21.
- Germande O., Gunkel-Grillon P., Dominique Y., Feurtet-Mazel A., Bierque E., Dassié E., Daffe G., Pierron F., Baudrimont I., Baudrimont M. Impact of nickel mining in New Caledonia on marbled eels *Anguilla marmorata*. *Journal of Hazardous Materials* 2022; **436**: 129285.
- Goble C., Cohen-Boulakia S., Soiland-Reyes S., Garijo D., Gil Y., Crusoe M.R., Peters K., Schober D. FAIR Computational Workflows. *Data Intelligence* 2020; **2**: 108–121.
- Gomes D.G.E., Pottier P., Crystal-Ornelas R., Hudgins E.J., Foroughirad V., Sánchez-Reyes L.L., Turba R., Martinez P.A., Moreau D., Bertram M.G., Smout C.A., Gaynor K.M. Why don't we share data and code? Perceived barriers and benefits to public archiving practices. *Proceedings of the Royal Society B: Biological Sciences* 2022; **289**.
- Gorman M. *Island Ecology*. Springer Netherlands. 2012.
- Gourmelon V., Maggia L., Powell J.R., Gigante S., Hortal S., Gueunier C., Letellier K., Carriconde F. Environmental and Geographical Factors Structure Soil Microbial Diversity in New Caledonian Ultramafic Substrates: A Metagenomic Approach. *PLOS ONE* 2016; **11**: e0167405.
- Gownaris N.J., Vermeir K., Bittner M.I., Gunawardena L., Kaur-Ghumaan S., Lepenies R., Ntsefong G.N., Zakari I.S. Barriers to Full Participation in the Open Science Life Cycle among Early Career Researchers. *Data Science Journal* 2022; **21**.
- Graham I.D., Logan J., Harrison M.B., Straus S.E., Tetroe J., Caswell W., Robinson N. Lost in Knowledge Translation: Time for a Map? 2006; **26**: 13–24.
- Grandcolas P., Murienne J., Robillard T., Desutter-Grandcolas L., Jourdan H.,

- Guilbert E., Deharveng L. Review. New Caledonia: A very old Darwinian island? *Philosophical Transactions of the Royal Society B: Biological Sciences* 2008; **363**: 3309–3317.
- Grüning B., Chilton J., Köster J., Dale R., Soranzo N., van den Beek M., Goecks J., Backofen R., Nekrutenko A., Taylor J. Practical Computational Reproducibility in the Life Sciences. *Cell Systems* 2018; **6**: 631–635.
- Gu Y.L., Sun C.H., Liu P., Zhang X., Sinev A.Y., Dumont H.J., Han B.P. Complete mitochondrial genome of *Ovalona pulchella* (Branchiopoda, Cladocera) as the first representative in the family Chydoridae: Gene rearrangements and phylogenetic analysis of Cladocera. *Gene* 2022; **818**: 146230.
- De Guerne J., Richard J. Révision des Calanides d'eau douce. *Société Zoologique de France* 1889; 53–144.
- Guillera-Arroita G., Lahoz-Monfort J.J., Elith J., Gordon A., Kujala H., Lentini P.E., McCarthy M.A., Tingley R., Wintle B.A. Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography* 2015; **24**: 276–292.
- Guisan A., Rahbek C. SESAM – a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography* 2011; **38**: 1433–1444.
- Guisan A., Thuiller W. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 2005; **8**: 993–1009.
- Guisan A., Thuiller W., Zimmermann N.E. Habitat Suitability and Distribution Models With Applications in R. *Habitat Suitability and Distribution Models: With Applications in R* 2017.
- Guisan A., Tingley R., Baumgartner J.B., Naujokaitis-Lewis I., Sutcliffe P.R., Tulloch A.I.T., Regan T.J., Brotons L., McDonald-Madden E., Mantyka-Pringle C., Martin T.G., Rhodes J.R., Maggini R., Setterfield S.A., Elith J., Schwartz M.W., Wintle B.A., Broennimann O., Austin M., Ferrier S., Kearney M.R., Possingham H.P., Buckley Y.M. Predicting species distributions for conservation decisions. *Ecology Letters* 2013; **16**: 1424–1435.

H

- Haase M., Bouchet P. Radiation of crenobiontic gastropods on an ancient continental island: The Hemistomia-clade in New Caledonia (Gastropoda: Hydrobiidae). *Hydrobiologia* 1998; **367**: 43–129.
- Haelewaters D., Hofmann T.A., Romero-Olivares A.L. Ten simple rules for Global North researchers to stop perpetuating helicopter research in the Global South. *PLoS Computational Biology* 2021; **17**: 1–8.
- Haig S.M., Miller M.P., Bellinger R., Draheim H.M., Mercer D.M., Mullins T.D. The conservation genetics juggling act: integrating genetics and ecology, science and policy. *Evolutionary Applications* 2016; **9**: 181–195.
- Hairston N.G., Van Brunt R.A., Kearns C.M., Engstrom D.R. Age and Survivorship of Diapausing Eggs in a Sediment Egg Bank. *Ecology* 1995; **76**: 1706–1711.
- Hammill E., Rogers A., Beckerman A.P. Costs, benefits and the evolution of inducible defences: a case study with *Daphnia pulex*. *Journal of Evolutionary Biology* 2008; **21**: 705–715.
- Hamond R. Non-marine Harpacticoid Copepods of Australia. I. Canthocamptidae of the Genus *Canthocamptus* Westwood s.lat. and *Fibulacamptus*, gen. nov., and Including the Description of a Related New Species of *Canthocamptus* from New Caledonia. *Invertebr. Taxon* 1987; **1**: 1023–247.
- Hampton S.E., Jones M.B., Wasser L.A., Schildhauer M.P., Supp S.R., Brun J., Hernandez R.R., Boettiger C., Collins S.L., Gross L.J., Fernández D.S., Budden A., White E.P., Teal T.K., Labou S.G., Aukema J.E. Skills and Knowledge for Data-Intensive Environmental Research. *BioScience* 2017; **67**: 546–557.
- Hanley J.A., McNeil B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; **143**: 29–36.
- Harada I. A new Copepod species parasitic on Formosan fresh-water fishes. *Journal of the Society of Tropical Agriculture Taiwan* 1930; **2**: 71–76.

- Hardisty A.R., Bacall F., Beard N., Balcázar-Vargas M.P., Balech B., Barcza Z., Bourlat S.J., Giovanni R., Jong Y., Leo F., Dobor L., Donvito G., Fellows D., Guerra A.F., Ferreira N., Fetyukova Y., Fosso B., Giddy J., Goble C., Güntsch A., Haines R., Ernst V.H., Hettling H., Hidy D., Horváth F., Ittzés D., Ittzés P., Jones A., Kottmann R., Kulawik R., Leidenberger S., Lyytikäinen-Saarenmaa P., Mathew C., Morrison N., Nenadic A., Hidalgo A.N., Obst M., Oostermeijer G., Paymal E., Pesole G., Pinto S., Poigné A., Fernandez F.Q., Santamaria M., Saarenmaa H., Sipos G., Sylla K.H., Tähtinen M., Vicario S., Vos R.A., Williams A.R., Yilmaz P. BioVeL: A virtual laboratory for data analysis and modelling in biodiversity science and ecology. *BMC Ecology* 2016; **16**.
- Haÿ V., Mennesson M.I., Carpentier C., Dahruddin H., Sauri S., Limmon G., Wowor D., Hubert N., Keith P., Lord C. Phylogeography of *Microphis retzii* (Bleeker, 1856) and *Microphis brachyurus* (Bleeker, 1854) in the Pacific. *Journal of Fish Biology* 2024; 1–19.
- Heads M. Methods in molecular biogeography: The case of New Caledonia. *Journal of Biogeography* 2023; **50**: 1437–1453.
- Herbert M.E., McIntyre P.B., Doran P.J., Allan J.D., Abell R. Terrestrial Reserve Networks Do Not Adequately Represent Aquatic Ecosystems Las Redes de Reservas Terrestres no Representan a los Ecosistemas Acuáticos Adecuadamente. *Conservation Biology* 2010; **24**: 1002–1011.
- Herteux C.E., Gawlik D.E., Smith L.L. Habitat Characteristics Affecting Wading Bird Use of Geographically Isolated Wetlands in the U.S. Southeastern Coastal Plain. *Wetlands* 2019; **40**: 1149–1159.
- Hill M.O. Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology* 1973; **54**: 427–432.
- Hill M.J., Greaves H.M., Sayer C.D., Hassall C., Milin M., Milner V.S., Marazzi L., Hall R., Harper L.R., Thornhill I., Walton R., Biggs J., Ewald N., Law A., Willby N., White J.C., Briers R.A., Mathers K.L., Jeffries M.J., Wood P.J. Pond ecology and conservation: research priorities and knowledge gaps. *Ecosphere* 2021; **12**: e03853.
- Hill M.F., Hastings A., Botsford L.W. The Effects of Small Dispersal Rates on Extinction Times in Structured Metapopulation Models. <https://doi.org/10.1086/341526> 2002; **160**: 389–402.

- Hill M.J., Sayer C.D., Wood P.J. When is the best time to sample aquatic macroinvertebrates in ponds for biodiversity assessment? *Environmental Monitoring and Assessment* 2016; **188**: 1–11.
- Hiltemann S., Rasche H., Gladman S., Hotz H.R., Larivière D., Blankenberg D., Jagtap P.D., Wollmann T., Bretaudeau A., Goué N., Griffin T.J., Royaux C., Bras Y. Le, Mehta S., Syme A., Coppens F., Droesbeke B., Soranzo N., Bacon W., Psomopoulos F., Gallardo-Alba C., Davis J., Föll M.C., Fahrner M., Doyle M.A., Serrano-Solano B., Fouilloux A.C., van Heusden P., Maier W., Clements D., Heyl F., Grüning B., Batut B. Galaxy Training: A powerful framework for teaching! *PLOS Computational Biology* 2023; **19**: e1010752.
- Hołyńska M., Śługocki Ł. Freshwater microcrustaceans (Copepoda: Cyclopidae) on islands: a review. *Hydrobiologia* 2023; **850**: 183–201.
- Hortal J., De Bello F., Diniz-Filho J.A.F., Lewinsohn T.M., Lobo J.M., Ladle R.J. Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics* 2015; **46**: 523–549.
- Hrivniak Ľ., Sartori M., Sroka P., Bojková J. Big diversity in a small hotspot: two new species of Leptophlebiidae (Insecta, Ephemeroptera) from New Caledonia. *ZooKeys* 2023; **1143**: 71–88.
- Huang Q., Xu L., Xie L., Liu P., Rizo E.Z.C., Han B.P. Spatial and temporal variation of genetic diversity and genetic differentiation in *Daphnia galeata* populations in four large reservoirs in southern China. *Frontiers in Microbiology* 2022; **13**: 1041011.
- Hudel L., Georg L., Henkel L. New distribution records: Four species of giant geckos (Gekkota: Diplodactylidae) occur in syntopy on Île des Pins, New Caledonia. *Herpetology Notes* 2020; **13**: 261–265.
- Hughes J.B., Daily G.C., Ehrlich P.R. Population diversity: its extent and extinction. *Science (New York, N.Y.)* 1997; **278**: 689–692.
- Hughes A.C., Orr M.C., Ma K., Costello M.J., Waller J., Provoost P., Yang Q., Zhu C., Qiao H. Sampling biases shape our view of the natural world. *Ecography* 2021; **44**: 1259–1269.

I
Incagnone G., Marrone F., Barone R., Robba L., Naselli-Flores L. How do freshwater organisms cross the “dry ocean”? A review on passive dispersal and colonization processes with a special focus on temporary ponds. *Hydrobiologia* 2015; **750**: 103–123.

Inventaire National du Patrimoine Naturel. 2023. Inventaire des collectivités d'outre-mer et de la métropole (icom_16) issu des données biogéographiques de TAXREF V16 [<https://inpn.mnhn.fr/espece/jeudonnees/72419>].

Ioannidis J.P.A. Correction: Why Most Published Research Findings Are False. *PLOS Medicine* 2022; **19**: e1004085.

Isnard S., Jaffré T. What Makes New Caledonia’s Flora So Outstanding? *Geographies of New Caledonia-Kanaky* 2024; 21–32.

Isnard S., L’huillier L., Rigault F., Jaffré T. How did the ultramafic soils shape the flora of the New Caledonian hotspot? *Plant and Soil* 2016; **403**: 53–76.

Ivimey-Cook E.R., Pick J.L., Bairos-Novak K.R., Culina A., Gould E., Grainger M., Marshall B.M., Moreau D., Paquet M., Royauté R., Sánchez-Tójar A., Silva I., Windecker S.M. Implementing code review in the scientific workflow: Insights from ecology and evolutionary biology. *Journal of Evolutionary Biology* 2023; **36**: 1347–1356.

J

Jaccard P. Distribution de la Flore Alpine dans le Bassin des Dranses et dans quelques régions voisines. *Bulletin de la Societe Vaudoise des Sciences Naturelles* 1901; **37**: 241–272.

Jaffre T. Comparative study of plant formations and flora of ultramafic rocks in New Caledonia and other tropical regions around the world (Guillaumet, JL and Belin, M and Puig, H, Ed.). *Tropical Phytogeography: Realities And Perspectives* 1996; 137.

Jaffré T., Morat P., Veillon J.M. La Flore: Caractéristiques et composition floristique des principales formations végétales. *Bois et forêts des tropiques* 1994; **242**.

- Jaffré T., Munzinger J., Lowry P.P. Threats to the conifer species found on New Caledonia's ultramafic massifs and proposals for urgently needed measures to improve their protection. *Biodiversity and Conservation* 2010; **19**: 1485–1502.
- Jeanpert J., Genthon P., Maurizot P., Folio J.L., Vendé-Leclerc M., Sérino J., Join J.L., Iseppi M. Morphology and distribution of dolines on ultramafic rocks from airborne LiDAR data: the case of southern Grande Terre in New Caledonia (SW Pacific). *Earth Surface Processes and Landforms* 2016; **41**: 1854–1868.
- Jenkins G.B., Beckerman A.P., Bellard C., Benítez-López A., Ellison A.M., Foote C.G., Hufton A.L., Lashley M.A., Lortie C.J., Ma Z., Moore A.J., Narum S.R., Nilsson J., O'Boyle B., Provete D.B., Razgour O., Rieseberg L., Riginos C., Santini L., Sibbett B., Peres-Neto P.R. Reproducibility in ecology and evolution: Minimum standards for data and code. *Ecology and Evolution* 2023; **13**: e9961.
- Jetz W., McGeoch M.A., Guralnick R., Ferrier S., Beck J., Costello M.J., Fernandez M., Geller G.N., Keil P., Merow C., Meyer C., Muller-Karger F.E., Pereira H.M., Regan E.C., Schmeller D.S., Turak E. Essential biodiversity variables for mapping and monitoring species populations. *Nature Ecology and Evolution* 2019; **3**: 539–551.
- Johanson K.A., Wells A. New Caledonia 's Trichoptera — present status of knowledge. *Zoosymposia* 2019; **102**: 87–102.
- Junk W.J. Long-term environmental trends and the future of tropical wetlands. *Environmental Conservation* 2002; **29**: 414–435.

K

- Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., Von Haeseler A., Jermini L.S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 2017 **14**:6 2017; **14**: 587–589.
- Katoh, K., Misawa, K., Kuma, K. & Miyata, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, **30**: 3059–3066.
- Keenan M., Cutler P., Marks J., Meylan R., Smith C., Koivisto E. Orienting international science cooperation to meet global 'grand challenges'. *Science*

and Public Policy 2012; **39**: 166–177.

- Keith P. Threatened fishes of the world: *Galaxias neocaledonicus* Weber & Beaufort, 1913 (Galaxiidae). *Environmental Biology of Fishes* 2002; **63**: 26.
- Kier G., Kreft H., Lee T.M., Jetz W., Ibisch P.L., Nowicki C., Mutke J., Barthlott W. A global assessment of endemism and species richness across island and mainland regions. *Proceedings of the National Academy of Sciences of the United States of America* 2009; **106**: 9322–7.
- Kierczak J., Pietranik A., Pędziwiatr A. Ultramafic geoecosystems as a natural source of Ni, Cr, and Co to the environment: A review. *Science of The Total Environment* 2021; **755**: 142620.
- Kisseih A., Higuti J., Martens K. Re-description of *strandesia sanoamuangae savatenalinton & martens*, 2010 and description of a new species of *strandesia* (Crustacea, ostracoda) from Grande Terre, New Caledonia. *European Journal of Taxonomy* 2020; **2020**: 1–28.
- Knight A.T., Cowling R.M., Rouget M., Balmford A., Lombard A.T., Campbell B.M. Knowing but Not Doing: Selecting Priority Conservation Areas and the Research-Implementation Gap. *Biology* 2008; **22**: 610–617.
- Knijn A., Michelacci V., Orsini M., Morabito S. Advanced research infrastructure for experimentation in genomics (ARIES): A lustrum of galaxy experience. *bioRxiv* 2020; 2020.05.14.095901.
- Korn M., Rabet N., Ghatte H. V., Marrone F., Hundsdoerfer A.K. Molecular phylogeny of the Notostraca. *Molecular phylogenetics and evolution* 2013; **69**: 1159–1171.
- Korovchinsky N.M. Review of Sididae (Crustacea: Cladocera: Ctenopoda) of the Pacific Ocean Islands, with description of a new species of *Diaphanosoma* from West Samoa. *Hydrobiologia* 2001.
- Köster J., Rahmann S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* 2012; **28**: 2520–2522.
- Kotov A.A., Elias-Gutierrez M. A phylogenetic analysis of *Ilyocryptus sars*, 1862 (cladocera: Ilyocryptidae). *International Review of Hydrobiology* 2009; **94**: 208–225.
- Kotov A.A., Zharov A.A., Chernova O.F., Neretina A.N., Gololobova M.A., Trofimova S.S., Zinovyev E. V., Izyumova E.I., Zanina O.G., Kirillova I. V., Shidlovskiy F.K. Crustacea (Branchiopoda) among Organic Remains from Mammoth

Hair. *Biology Bulletin* 2019; **46**: 850–863.

Kreft H., Jetz W. A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography* 2010; **37**: 2029–2053.

Kuczyńska-Kippen N., Pronin M. Diversity and zooplankton species associated with certain hydroperiods and fish state in field ponds. *Ecological Indicators* 2018; **90**: 171–178.

Kueffer C., Kinney K. What is the importance of islands to environmental conservation? *Environmental Conservation* 2017; **44**: 311–322.

Kurata K., Jaffré T., Setoguchi H. Genetic diversity and geographical structure of the pitcher plant *Nepenthes vieillardii* in New Caledonia: A chloroplast DNA haplotype analysis. *American Journal of Botany* 2008; **95**: 1632–1644.

L

L'Huillier L., Jaffré T. *L'exploitation des minerais de nickel en Nouvelle-Calédonie*.

L'huillier L., Jaffré T., Wulff A. Mines et Environnement en Nouvelle-Calédonie : les milieux sur substrats ultramafiques et leur restauration Synthèses Etudes. 2010.

Lai J., Lortie C.J., Muenchen R.A., Yang J., Ma K. Evaluating the popularity of R in ecology. *Ecosphere* 2019; **10**: e02567.

Lamprecht A.L., Garcia L., Kuzak M., Martinez C., Arcila R., Martin Del Pico E., Dominguez Del Angel V., van de Sandt S., Ison J., Martinez P.A., McQuilton P., Valencia A., Harrow J., Psomopoulos F., Gelpi J.L., Chue Hong N., Goble C., Capella-Gutierrez S. Towards FAIR principles for research software. *Data Science* 2019; **3**: 37–59.

Lanfear R., Frandsen P.B., Wright A.M., Senfeld T., Calcott B. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Molecular Biology and Evolution* 2016; **34**: 772–773.

Lannuzel G., Balmot J., Dubos N., Thibault M., Fogliani B. High-resolution topographic variables accurately predict the distribution of rare plant species for conservation area selection in a narrow-endemism hotspot in New Caledonia. *Biodiversity and Conservation* 2021; **30**: 963–990.

- Larcombe L., Hendricusdottir R., Attwood T., Bacall F., Beard N., Bellis L., Dunn W., Hancock J., Nenadic A., Orengo C., Overduin B., Sansone S., Thurston M., Viant M., Winder C., Goble C., Ponting C., Rustici G. ELIXIR-UK role in bioinformatics training at the national level and across ELIXIR. *F1000Research* 2017; **6**: 952.
- Lauber T.B., Stedman R.C., Decker D.J., Knuth B.A. Linking Knowledge to Action in Collaborative Conservation. *Conservation Biology* 2011; **25**: 1186–1194.
- Leonardo G.A., Valentina B., Annagrazia P., Genuario Belmonte F., Alfonso G., Belmonte Á.G., Belmonte G., Beccarisi L., Pieri V., Frassanito A., Nazionale Dell' P., Murgia A. Using crustaceans to identify different pond types. A case study from the Alta Murgia National Park, Apulia (South-eastern Italy). *Hydrobiologia* 2016; **782**: 53–69.
- Leroy B. Choosing presence-only species distribution models. *Journal of Biogeography* 2023; **50**: 247–250.
- Leroy B., Dias M.S., Giraud E., Hugueny B., Jézéquel C., Leprieur F., Oberdorff T., Tedesco P.A. Global biogeographical regions of freshwater fish species. *Journal of Biogeography* 2019; **46**: 2407–2419.
- Longhurst A.R. A review of the Notostraca. *Bulletin of the British Museum (Natural History) Zoology* 1955; **3**: 1–57.
- Lortie C.J. The early bird gets the return: The benefits of publishing your data sooner. *Ecology and Evolution* 2021; **11**: 10736–10740.
- Lynch A.J., Cooke S.J., Arthington A.H., Baigun C., Bossenbroek L., Dickens C., Harrison I., Kimirei I., Langhans S.D., Murchie K.J., Olden J.D., Ormerod S.J., Owuor M., Raghavan R., Samways M.J., Schinegger R., Sharma S., Tachamo-Shah R.D., Tickner D., Tweddle D., Young N., Jähnig S.C. People need freshwater biodiversity. *Wiley Interdisciplinary Reviews: Water* 2023; **10**: e1633.

M

- MacArthur R.H., Wilson E.O. An Equilibrium Theory Of Insular Zoogeography. *Evolution* 1963; **17**: 373–387.
- Macedo R.L., Toutain M., Reid J., Soto I., Haubrock P.J., Cuthbert R.N., Rocha O. Substantial unrealised global biodiversity of continental microcrustaceans. *Journal of Plankton Research* 2024; **46**: 338–347.

- Mandal S., Abbott Wilkins R., Shurin J.B. Compensatory grazing by *Daphnia* generates a trade-off between top-down and bottom-up effects across phytoplankton taxa. *Ecosphere* 2018; **9**: e02537.
- Manel S., Guerin P.E., Mouillot D., Blanchet S., Velez L., Albouy C., Pellissier L. Global determinants of freshwater and marine fish genetic diversity. *Nature Communications* 2020 *11:1* 2020; **11**: 1–9.
- Manuel RAMSAR. Manuels Ramsar pour l'utilisation rationnelle des zones humides. *Manuel 17: Inscription de Sites Ramsar* 2010; 132.
- Marques J.M., Carreira P.M., Carvalho M.R., Matias M.J., Goff F.E., Basto M.J., Graça R.C., Aires-Barros L., Rocha L. Origins of high pH mineral waters from ultramafic rocks, Central Portugal. *Applied Geochemistry* 2008; **23**: 3278–3289.
- Martens K., Yavuzatmaca M., Higuti J. On a new species of the genus *Cyprinotus* (Crustacea, ostracoda) from a temporary wetland in New Caledonia (Pacific ocean), with a reappraisal of the genus. *European Journal of Taxonomy* 2019; **2019**: 1–22.
- Martin G.K., Beisner B.E., Chain F.J.J., Cristescu M.E., del Giorgio P.A., Derry A.M. Freshwater zooplankton metapopulations and metacommunities respond differently to environmental and spatial variation. *Ecology* 2021; **102**: 0–1.
- Maturana C.S., Rosenfeld S., Biersma E.M., Segovia N.I., González-Wevar C.A., Díaz A., Naretto J., Duggan I.C., Hogg I.D., Poulin E., Convey P., Jackson J.A. Historical biogeography of the Gondwanan freshwater genus *Boeckella* (Crustacea): Timing and modes of speciation in the Southern Hemisphere. *Diversity and Distributions* 2021; **27**: 2330–2343.
- Maturana C.S., Rosenfeld S., Naretto J., Convey P., Poulin E., Lopez Z., Gerard K., Contador T., Rendoll J., Jackson J.A., Hogg I.D., Duggan I.C., Bahamonde F. Occurrences of *Boeckella* at high latitudes in the southern hemisphere. *GBIF* 2019.
- Maturana C.S., Segovia N.I., González-Wevar C.A., Díaz A., Rosenfeld S., Poulin E., Jackson J.A., Convey P. Evidence of strong small-scale population structure in the Antarctic freshwater copepod *Boeckella poppei* in lakes on Signy Island, South Orkney Islands. *Limnology and Oceanography* 2020; **65**: 2024–2040.
- Maurizot P., Campbell H.J. Palaeobiogeography of New Caledonia. *Geological Society Memoir* 2020; **51**: 189–213.

- McIntire E.J.B., Chubaty A.M., Cumming S.G., Andison D., Barros C., Boisvenue C., Haché S., Luo Y., Micheletti T., Stewart F.E.C. PERFICT: A Re-imagined foundation for predictive ecology. *Ecology Letters* 2022; **25**: 1345–1351.
- McKinney M.L. Extinction vulnerability and selectivity: Combining ecological and paleontological views. *Annual Review of Ecology and Systematics* 1997; **28**: 495–516.
- Menesson M.I., Bonillo C., Feunteun E., Keith P. Phylogeography of *Eleotris fusca* (Teleostei: Gobioidae: Eleotridae) in the Indo-Pacific area reveals a cryptic species in the Indian Ocean. *Conservation Genetics* 2018; **19**: 1025–1038.
- Meusel F., Schwentner M. Molecular and morphological delimitation of Australian *Triops* species (Crustacea: Branchiopoda: Notostraca)—large diversity and little morphological differentiation. *Organisms Diversity and Evolution* 2017; **17**: 137–156.
- Michener W.K. Ten Simple Rules for Creating a Good Data Management Plan. *PLOS Computational Biology* 2015; **11**: e1004525.
- Michener W.K., Brunt J.W., Helly J.J., Kirchner T.B., Stafford S.G. Nongeospatial metadata for the ecological sciences. *Ecological Applications* 1997; **7**: 330–342.
- Michener W.K., Jones M.B. Ecoinformatics: Supporting ecology as a data-intensive science. *Trends in Ecology and Evolution* 2012; **27**: 85–93.
- Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., Von Haeseler A., Lanfear R., Teeling E. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular biology and evolution* 2020; **37**: 1530–1534.
- Minocher R., Atmaca S., Bavero C., McElreath R., Beheim B. Estimating the reproducibility of social learning research published between 1955 and 2018. *Royal Society Open Science* 2021; **8**.
- Montaña C.G., Keppeler F.W., Laughrey C.P., Schalk C.M. Community assembly within ponds: the roles of space, time, and environmental gradients. *Aquatic Ecology* 2022; **56**: 123–142.
- Mossin J. Physicochemical Factors Inducing Embryonic Development and Spring Hatching of the European Fairy Shrimp *Siphonophanes grubei* (Dybowski) (Crustacea: Anostraca). 1986; **6**: 693–704.

Mrázek, A. 1901. Süßwasser-Copepoden. *Ergebnisse der Hamburger Magalhaensischen Sammelreise 1892/93, Hamburg Natural History Museum*, **2**: 1–29.

Munafò M.R., Nosek B.A., Bishop D.V.M., Button K.S., Chambers C.D., Percie Du Sert N., Simonsohn U., Wagenmakers E.J., Ware J.J., Ioannidis J.P.A. A manifesto for reproducible science. *Nature Human Behaviour* 2017 **1**:1 2017; **1**: 1–9.

N

Naganawa H. First record of *Triops strenuus* Wolf, 1911 (Branchiopoda, Notostraca), a tadpole shrimp of Australian origin, from Japan. *Crustaceana* 2018; **91**: 425–438.

Nattier R., Grandcolas P., Pellens R., Jourdan H., Couloux A., Poulain S., Robillard T. Climate and Soil Type Together Explain the Distribution of Microendemic Species in a Biodiversity Hotspot (V Ketmaier, Ed.). *PLoS ONE* 2013; **8**: e80811.

Nattier R., Pellens R., Robillard T., Jourdan H., Legendre F., Caesar M., Nel A., Grandcolas P. Updating the Phylogenetic Dating of New Caledonian Biodiversity with a Meta-analysis of the Available Evidence. *Scientific Reports* 2017; **7**: 3705.

Natural Environment Research Council (2010, 2012) Most Wanted: Postgraduate Skills Needs in the Environment Sector.

Norvez O., Milon T., Pamerlon S., Archambeau A., Bouix T., Cheminée O., Bras Y. Le, Robert S., Vinet C. Comprendre, partager, ré-utiliser les données de biodiversité. 2023.

Nourisson M., Aguesse P. Cycle annuel des phyllopoïdes d'une mare temporaire de Camargue. *Bulletin de la Société Zoologique de France* 1961; **86**: 754–762.

O

Olesen J., Pöllabauer C., Sigvardt Z.M.S., Rogers D.C. A new species of *Lynceus Müller*, 1776 from New Caledonia (Crustacea: Branchiopoda: Laevicaudata) from dolines, with remarks on zoogeography. *European Journal of Taxonomy* 2016; **224**: 1–18.

Olmo C., Armengol X., Antón-Pardo M., Ortells R. The environmental and zooplankton community changes in restored ponds over 4 years. *Journal of Plankton Research* 2016; **38**: 490–501.

P

Padial J.M., Miralles A., De la Riva I., Vences M. The integrative future of taxonomy. *Frontiers in Zoology* 2010; **7**: 1–14.

Paradis E. Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics* 2010; **26**: 419–420.

Parmar T.K., Rawtani D., Agrawal Y.K. Bioindicators: the natural indicator of environmental pollution. *Frontiers in Life Science* 2016; **9**: 110–118.

Parry V., Kiemel K., Pawlak J., Eccard J., Tiedemann R., Weithoff G. Drivers of zooplankton dispersal in a pond metacommunity. *Hydrobiologia* 2024; **851**: 2875–2893.

Pascal M., Richer de Forges B., Le Guyader H., Simberloff D. Mining and Other Threats to the New Caledonia Biodiversity Hotspot. *Conservation Biology* 2008; **22**: 498–499.

Pattanayak S., Das S., Navyasri K. Bioindicator Emerged as a Potential Environmental Marker. *International Journal of Agriculture Environment and Biotechnology* 2020; **13**: 339–344.

Pekel J.F., Cottam A., Gorelick N., Belward A.S. High-resolution mapping of global surface water and its long-term changes. *Nature* 2016; **540**: 418–422.

Pelletier B. Geology of the New Caledonia region and its implications for the study of the New Caledonian biodiversity. *Compendium of marine species of New Caledonia* 2007; 19–32.

Pelletier D., Bissery C., Gonson C. PAMPA: Indicateurs de la Performance d'Aires Marines Protégées pour la gestion des écosystèmes côtiers, des ressources et de leurs usages. 2014.

Pelletier D., Roos D., Bouchouca M., Schohn T., Roman W., Gonson C., Bockel T., Carpentier L., Preuss B., Powell A., Garcia J., Gaboriau M., Cadé F., Royaux C., Le Bras Y., Reece Y. A Standardized Workflow Based on the STAVIRO Unbaited Underwater Video System for Monitoring Fish and Habitat Essential Biodiversity Variables in Coastal Areas. *Frontiers in Marine Science* 2021; **8**: 689280.

- Pentinsaari M., Ratnasingham S., Miller S.E., Hebert P.D.N. BOLD and GenBank revisited – Do identification errors arise in the lab or in the sequence libraries? *PLOS ONE* 2020; **15**: e0231814.
- Pillon Y., Hopkins H.C.F., Munzinger J., Amir H., Chase M.W. Cryptic species, gene recombination and hybridization in the genus *Spiraeanthemum* (Cunoniaceae) from New Caledonia. *Botanical Journal of the Linnean Society* 2009; **161**: 137–152.
- Pillon Y., Jaffré T., Birnbaum P., Bruy D., Cluzel D., Ducousso M., Fogliani B., Ibanez T., Jourdan H., Lagarde L., Léopold A., Munzinger J., Pouteau R., Read J., Isnard S. Infertile landscapes on an old oceanic island: the biodiversity hotspot of New Caledonia. *Biological Journal of the Linnean Society* 2021; **133**: 317–341.
- Pitoiset A. *Le nickel en Nouvelle Calédonie*. 2016.
- Plesser H.E. Reproducibility vs. Replicability: A brief history of a confused terminology. *Frontiers in Neuroinformatics* 2018; **11**: 76.
- Pollock L.J., Connor L.M.J.O., Mokany K., Rosauer D.F., Talluto M. V, Thuiller W. Protecting Biodiversity (in All Its Complexity): New Models and Methods. *Trends in Ecology & Evolution* 2020; **35**: 1119–1128.
- Poncet V., Munoz F., Munzinger J., Pillon Y., Gomez C., Couderc M., Tranchant-Dubreuil C., Hamon S., De Kochko A. Phylogeography and niche modelling of the relict plant *Amborella trichopoda* (Amborellaceae) reveal multiple Pleistocene refugia in New Caledonia. *Molecular Ecology* 2013; **22**: 6163–6178.
- Ponel P. Les fluctuations climatiques au Pleniglaciaire wurmien deduites des assemblages d'arthropodes fossiles a La Grande Pile (Haute-Saone, France) [Climatic variations during the Wurmian Pleniglacial inferred from fossil arthropod assemblages at La Grande Pile. *Comptes Rendus de l'Academie des Sciences Serie II A Sciences de la Terre et des Planetes* 1994; **319**: 845–852.
- Powers S.M., Hampton S.E. Open science, reproducibility, and transparency in ecology. *Ecological Applications* 2019; **29**.

Q

Quinlan K., Bayly I.A.E. A new species of *Boeckella* (Copepoda: Calanoida) from arid Western Australia, an updated key, and aspects of claypan ecology. *Records of the Western Australian Museum* 2017; **32**: 191.

R

Rabet N., Lacau S., Bozelli R.L. Richness of *Dendrocephalus* (Branchiopoda, Anostraca) in Brazil with the description of two new species. *European Journal of Taxonomy* 2018; **2018**.

Rachel Carson. *Silent Spring*. 1962.

Ramiro-Sánchez B., Martin A., Leroy B. The epitome of data paucity: Deep-sea habitats of the Southern Indian Ocean. 2023.

Reid A.J., Carlson A.K., Creed I.F., Eliason E.J., Gell P.A., Johnson P.T.J., Kidd K.A., MacCormack T.J., Olden J.D., Ormerod S.J., Smol J.P., Taylor W.W., Tockner K., Vermaire J.C., Dudgeon D., Cooke S.J. Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biological Reviews* 2019; **94**: 849–873.

Rhazi L., Grillas P., Saber E.R., Rhazi M., Brendonck L., Waterkeyn A. Vegetation of Mediterranean temporary pools: A fading jewel? *Hydrobiologia* 2012; **689**: 23–36.

Richardson D.C., Holgerson M.A., Farragher M.J., Hoffman K.K., King K.B.S., Alfonso M.B., Andersen M.R., Cheruveil K.S., Coleman K.A., Farruggia M.J., Fernandez R.L., Hondula K.L., López Moreira Mazacotte G.A., Paul K., Peierls B.L., Rabaey J.S., Sadro S., Sánchez M.L., Smyth R.L., Sweetman J.N. A functional definition to distinguish ponds from lakes and wetlands. *Scientific Reports* 2022 12:1 2022; **12**: 1–13.

Richer de Forges B., Pascal M. La Nouvelle-Calédonie, un « point chaud » de la biodiversité mondiale gravement menacé par l'exploitation minière. *Journal de la société des océanistes* 2008; 95–112.

Ridgeway G. The state of boosting. *Computing Science and Statistics* 1999; **31**: 172–181.

- Roberts D.R., Bahn V., Ciuti S., Boyce M.S., Elith J., Guillera-Arroita G., Hauenstein S., Lahoz-Monfort J.J., Schröder B., Thuiller W., Warton D.I., Wintle B.A., Hartig F., Dormann C.F. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 2017; **40**: 913–929.
- Rodary E. Which Environmental Policies for New Caledonia? *Geographies of New Caledonia-Kanaky* 2024; 13–20.
- Rodrigues A.V., Nakamura G., Staggemeier V.G., Duarte L. Species misidentification affects biodiversity metrics: Dealing with this issue using the new R package *naturaList*. *Ecological Informatics* 2022; **69**: 101625.
- Rogers D.C., Ferreira A. A new species of Branchinecta (crustacea: Anostraca) from Brasil. *Zootaxa* 2007; 27–34.
- Rogers D.C., Gomes J. A new species of Dendrocephalus (Crustacea, Anostraca) from Serra dos Carajás (Pará State, Brasil). 2012.
- Rogers D.C., Zharov A.A., Neretina A.N., Kuzmina S.A., Kotov A.A. A Review of Recently Discovered Remains of the Pleistocene Branchiopods (Anostraca, Notostraca) from NE Siberia and Arctic Canada. *Water* 2021, Vol. 13, Page 280 2021; **13**: 280.
- Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology* 2012; **61**: 539–542.
- Rosvall M., Bergstrom C. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences* 2008; **105**: 1118–1123.
- Royaux C., Charpin N., Rabet N. Genetic variability of New Caledonian Boeckella de Guerne & Richard, 1889 (Copepoda: Calanoida), with the description of a new species. *Journal of Crustacean Biology* 2024; **44**: 1–17.
- Ruhfel B.R., Bove C.P., Philbrick C.T., Davis C.C. Dispersal largely explains the gondwanan distribution of the ancient tropical clusioid plant clade. *American Journal of Botany* 2016; **103**: 1117–1128.
- Ruhsam M., Clark A., Finger A., Wulff A.S., Mill R.R., Thomas P.I., Gardner M.F., Gaudeul M., Ennos R.A., Hollingsworth P.M. Hidden in plain view: Cryptic diversity in the emblematic Araucaria of New Caledonia. *American Journal of Botany* 2016; **103**: 888–898.

Rundle H.D., Nosil P. Ecological speciation. *Ecology Letters* 2005; **8**: 336–352.

S

Sandel B., Arge L., Dalsgaard B., Davies R.G., Gaston K.J., Sutherland W.J., Svenning J.C. The influence of late quaternary climate-change velocity on species endemism. *Science* 2011; **334**: 660–664.

Sanger F., Nicklen S., Coulson A.R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 1977; **74**: 5463–7.

Scheihing R., Cardenas L., Nespolo R.F., Krall P., Walz K., Kohshima S., Labarca P. Morphological and molecular analysis of centropagids from the high Andean plateau (Copepoda: Calanoidea). *Hydrobiologia* 2010; **637**: 45–52.

Schluter D. Ecology and the origin of species. *Trends in Ecology & Evolution* 2001; **16**: 372–380.

Schneider D.W., Frost T.M. Habitat duration and community structure in temporary ponds. *Journal of the North American Benthological Society* 1996; **15**: 64–86.

Schneider K., Le Mestre M., Desrioux I., Gunkel-Grillon P. Occurrence of arsenic in ultramafic rocks' alterites from nickel mines in New Caledonia: implications for the contamination of surface waters. *Environmental Chemistry Letters* 2020; **18**: 1397–1402.

Schwentner M., Timms B. V., Richter S. Spinicaudata (Branchiopoda: Diplostraca) in Australia's Arid Zone: Unparalleled Diversity at Regional Scales and within Water Bodies. *Journal of Crustacean Biology* 2015; **35**: 366–378.

Scott, T. 1914. Remarks on some Copepoda from the Falkland Islands collected by Mr. Rupert Vallentin, F. L.S. *Annals and Magazine of Natural History*, **13**: 1–11.

Searle, J. 1911. Some Victorian Copepoda new to science. *Victorian Naturalist*, **27**: 174–178.

Seminara M., Vagaggini D., Stoch F. Long-term monitoring of astatic water bodies: microcrustaceans as indicators of hydroperiod length in ponds and pools. *Rendiconti Lincei* 2015; **26**: 345–352.

Serrano-Solano B., Fouilloux A., Eguinoa I., Kalaš M., Grüning B., Coppens F. Galaxy:

- A Decade of Realising CWFR Concepts. *Data Intelligence* 2022; **4**: 358–371.
- Sichlau M.H., Hansen J.L.S., Andersen T.J., Hansen B.W. Distribution and mortality of diapause eggs from calanoid copepods in relation to sedimentation regimes. *Marine Biology* 2011; **158**: 665–676.
- Sigvardt Z.M., Olesen J., Christopher Rogers D., Timms B., Mlambo M., Rabet N., Palero F., Jørgen Olesen C. Multilocus phylogenetics of smooth clam shrimps (Branchiopoda, Laevicaudata). *Zoologica Scripta* 2021; **50**: 795–811.
- Sinev A.Y. Armatalona gen. n. – a new genus of subfamily Aloninae (Anomopoda, Chydoridae), separated from genus Alona Baird, 1840. *Hydrobiologia* 2004 *520:1* 2004; **520**: 29–47.
- Sinev A.Y., Kobayashi T. Redescription of the endemic Australian cladoceran Alona willisi (Smirnov, 1989) and its assignment to Acanthalona gen. nov.(Cladocera: Anomopoda: Chydoridae). *Zootaxa* 2012; 43–55.
- Smirnov N.N., Timms B.V. A revision of the Australian Cladocera. 1983.
- Soiland-Reyes S., Sefton P., Crosas M., Castro L.J., Coppens F., Fernández J.M., Garijo D., Grüning B., La Rosa M., Leo S., Ó Carragáin E., Portier M., Trisovic A., Community R.C., Groth P., Goble C. Packaging research artefacts with RO-Crate. *Data Science* 2022; **5**: 97–138.
- Soltis D.E., Smith S.A., Cellinese N., Wurdack K.J., Tank D.C., Brockington S.F., Refulio-Rodriguez N.F., Walker J.B., Moore M.J., Carlsward B.S., Bell C.D., Latvis M., Crawley S., Black C., Diouf D., Xi Z., Rushworth C.A., Gitzendanner M.A., Sytsma K.J., Qiu Y.L., Hilu K.W., Davis C.C., Sanderson M.J., Beaman R.S., Olmstead R.G., Judd W.S., Donoghue M.J., Soltis P.S. Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany* 2011; **98**: 704–730.
- Stefanoudis P. V., Sayer C.D., Greaves H.M., Davidson T.A., Robson H., Almeida D., Smith E. Consequences of fish for cladoceran, water beetle and macrophyte communities in a farmland pond landscape: Implications for conservation. *Fundamental and Applied Limnology* 2017; **190**: 141–156.
- Stingelin T. Cladoceren von Neu-Caledonien. *Nova Caledonia* 1915; **2**: 197–208.
- Strachan S.R., Chester E.T., Robson B.J. Freshwater Invertebrate Life History Strategies for Surviving Desiccation. *Springer Science Reviews* 2015; **3**: 57–75.
- Strayer D.L., Dudgeon D. Freshwater biodiversity conservation: Recent progress and

future challenges. *Journal of the North American Benthological Society* 2010; **29**: 344–358.

Strijkers R., Cushing R., Vasyunin D., De Laat C., Belloum A.S.Z., Meijer R. Toward executable scientific publications. *Procedia Computer Science* 2011; **4**: 707–715.

Stroud J.T., Bush M.R., Ladd M.C., Nowicki R.J., Shantz A.A., Sweatman J. Is a community still a community? Reviewing definitions of key terms in community ecology. *Ecology and Evolution* 2015; **5**: 4757–4765.

T

Tavernini S. Seasonal and inter-annual zooplankton dynamics in temporary pools with different hydroperiods. *Limnologica* 2008; **38**: 63–75.

Taylor H., Dussex N., van Heezik Y. Bridging the conservation genetics gap by identifying barriers to implementation for conservation practitioners. *Global Ecology and Conservation* 2017; **10**: 231–242.

Templeton A.R., Crandall K.A., Sing C.F. A cladistic analysis of phenotypic association with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 1992; **132**: 619–635.

Teske P.R., Golla T.R., Sandoval-Castillo J., Emami-Khoyi A., van der Lingen C.D., von der Heyden S., Chiazari B., Jansen van Vuuren B., Beheregaray L.B. Mitochondrial DNA is unsuitable to test for isolation by distance. *Scientific Reports 2018 8:1* 2018; **8**: 1–9.

The Galaxy Community. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Research* 2022; **50**: W345–W351.

Thomson, G.M. 1883. On the New Zealand Copepoda. *Transactions and Proceedings of the New Zealand Institute*, **15**: 93–116.

Thuiller W., Lafourcade B., Engler R., Araújo M.B. BIOMOD - A platform for ensemble forecasting of species distributions. *Ecography* 2009; **32**: 369–373.

Tickner D., Opperman J.J., Abell R., Acreman M., Arthington A.H., Bunn S.E., Cooke S.J., Dalton J., Darwall W., Edwards G., Harrison I., Hughes K., Jones T., Leclère D., Lynch A.J., Leonard P., McClain M.E., Muruven D., Olden J.D.,

Ormerod S.J., Robinson J., Tharme R.E., Thieme M., Tockner K., Wright M., Young L. Bending the Curve of Global Freshwater Biodiversity Loss: An Emergency Recovery Plan. *BioScience* 2020; **70**: 330–342.

Timms B. V. The Cladocera (Crustacea) of New Caledonia. *Proceedings of the Linnean Society N.S.W* 1985; **108**: 47–57.

Timms B. V. A revised identification guide to the fairy shrimps (Crustacea: Anostraca: Anostracina) of Australia. 2015.

Timms B. V. A partial revision of the Australian Eulimnadia Packard, 1874 (Branchiopoda: Spinicaudata: Limnadiidae). *Zootaxa* 2016; **4066**: 351–389.

Timms B. V., Pinder A.M., Campagna V.S. The biogeography and conservation status of the Australian endemic brine shrimp *Parartemia* (Crustacea, Anostraca, Parartemiidae). *Conservation Science Western Australia* 2009; **7**: 413–427.

Touchon J.C., McCoy M.W. The mismatch between current statistical practice and doctoral training in ecology. *Ecosphere* 2016; **7**: e01394.

Toussaint M. L'épreuve du feu. Politiques de la nature, savoirs, feux de brousse et décolonisation en Nouvelle-Calédonie. 2018.

Tsuji S., Shibata N., Sawada H., Ushio M. Quantitative evaluation of intraspecific genetic diversity in a natural fish population using environmental DNA analysis. *Molecular Ecology Resources* 2020; **20**: 1323–1332.

U

Ung V., Dubus G., Zaragüeta-Bagils R., Vignes-Lebbe R. Xper2: introducing e-taxonomy. *Bioinformatics* 2010; **26**: 703–704.

V

Valavi R., Guillera-Arroita G., Lahoz-Monfort J.J., Elith J. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecological Monographs* 2021; e01486.

Vamosi J.C., Vamosi S.M. Extinction Risk Escalates in the Tropics. *PLOS ONE* 2008; **3**: e3886.

Van Damme K., Sinev A. Tropical Amphi-Pacific disjunctions in the Cladocera (Crustacea: Branchiopoda). *Article in Journal of Limnology* 2013; **72**: 209–

244.

- Van Damme K., Dumont H.J. The 'true' genus *Alona* Baird, 1843 (Crustacea: Cladocera: Anomopoda): position of the *A. quadrangularis*-group and description of a new species from the Democratic Republic of Congo. *Zootaxa* 2008; **1843**: 1–25.
- Varela S., Anderson R.P., García-Valdés R., Fernández-González F. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography* 2014; **37**: 1084–1091.
- Verniest F. Adaptation du réseau d'aires protégées aux changements globaux pour mieux conserver: cas des communautés d'oiseaux d'eau hivernants des zones humides. Paris, Muséum national d'histoire naturelle.
- Veron S., Haevermans T., Govaerts R., Mouchet M., Pellens R. Distribution and relative age of endemism across islands worldwide. *Scientific Reports* 2019 *9:1* 2019; **9**: 1–12.
- Vilgrain L., Maps F., Basedow S., Trudnowska E., Madoui M., Niehoff B., Ayata S. Copepods' true colors: astaxanthin pigmentation as an indicator of fitness. *Ecosphere* 2023; **14**: 1–25.
- Vilhena D.A., Antonelli A. A network approach for identifying and delimiting biogeographical regions. *Nature Communications* 2015 *6:1* 2015; **6**: 1–9.

W

- Walter, T.C. & Boxshall, G. 2021. World of Copepods database. Boeckella Guerne & Richard, 1889 [https://www.marinespecies.org/copepoda/aphia.php?p=taxdetails&id=346853].
- Wang C.C., Liu J.Y., Chou L.S. Egg Bank Spatial Structure and Functional Size of Three Sympatric Branchiopods (Branchiopoda) in Siangtian Pond, Taiwan. *Journal of Crustacean Biology* 2014; **34**: 412–421.
- Warton D.I., Blanchet F.G., O'Hara R.B., Ovaskainen O., Taskinen S., Walker S.C., Hui F.K.C. So Many Variables: Joint Modeling in Community Ecology. *Trends in Ecology and Evolution* 2015; **30**: 766–779.
- Weeks S.C., Chapman E.G., Rogers D.C., Senyo D.M., Hoeh W.R. Evolutionary transitions among dioecy, androdioecy and hermaphroditism in limnadiid clam shrimp (Branchiopoda: Spinicaudata). *Journal of Evolutionary Biology* 2009; **22**: 1781–1799.

- White L.T., Gibson G.M., Lister G.S. A reassessment of paleogeographic reconstructions of eastern Gondwana: Bringing geology back into the equation. *Gondwana Research* 2013; **24**: 984–998.
- Whittaker R.J., Fernández-Palacios J.M., Matthews T.J., Borregaard M.K., Triantis K.A. Island biogeography: Taking the long view of nature's laboratories. *Science* 2017; **357**.
- Whittaker R.J., Fernández-Palacios J.M., Matthews T.J. (Thomas J., Whittaker R.J.) *Island biogeography: geo-environmental dynamics, ecology, evolution, human impact, and conservation*. Oxford University Press, 2023.
- Wilkinson M.D., Dumontier M., Aalbersberg I.J., Appleton G., Axton M., Baak A., Blomberg N., Boiten J.W., da Silva Santos L.B., Bourne P.E., Bouwman J., Brookes A.J., Clark T., Crosas M., Dillo I., Dumon O., Edmunds S., Evelo C.T., Finkers R., Gonzalez-Beltran A., Gray A.J.G., Groth P., Goble C., Grethe J.S., Heringa J., Hoen P.A.C., Hooft R., Kuhn T., Kok R., Kok J., Lusher S.J., Martone M.E., Mons A., Packer A.L., Persson B., Rocca-Serra P., Roos M., van Schaik R., Sansone S.A., Schultes E., Sengstag T., Slater T., Strawn G., Swertz M.A., Thompson M., Van Der Lei J., Van Mulligen E., Velterop J., Waagmeester A., Wittenburg P., Wolstencroft K., Zhao J., Mons B. Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 2016; **3**: 1–9.
- Williams D.D. *The Biology of Temporary Waters*. Oxford University Press. 2007.
- Williams J.J., Teal T.K. A vision for collaborative training infrastructure for bioinformatics. *Annals of the New York Academy of Sciences* 2017; **1387**: 54–60.
- Willmer J.N.G., Püttker T., Prevedello J.A. Global impacts of edge effects on species richness. *Biological Conservation* 2022; **272**: 109654.
- Winter D.J. mmod: an R library for the calculation of population differentiation statistics. *Molecular Ecology Resources* 2012; **12**: 1158–1160.
- Woolley S.N.C., Foster S.D., Bax N.J., Currie J.C., Dunn D.C., Hansen C., Hill N., O'Hara T.D., Ovaskainen O., Sayre R., Vanhatalo J.P., Dunstan P.K. Bioregions in Marine Environments: Combining Biological and Environmental Data for Management and Scientific Understanding. *BioScience* 2020; **70**: 48–59.

X

Xiang C.Y., Gao F., Jakovlić I., Lei H.P., Hu Y., Zhang H., Zou H., Wang G.T., Zhang D. Using PhyloSuite for molecular phylogeny and tree-based analyses. *iMeta* 2023; **2**: e87.

Z

Zakardjian M., Mahé P., Geslin B., Jourdan H. Plant–pollinator interactions in ultramafic and non-ultramafic environments in New Caledonia. *Botany Letters* 2023; **170**: 459–467.

Zhang D., Gao F., Jakovlić I., Zou H., Zhang J., Li W.X., Wang G.T. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Molecular ecology resources* 2020; **20**: 348–355.

Zurell D., Franklin J., König C., Bouchet P.J., Dormann C.F., Elith J., Fandos G., Feng X., Guillera-Arroita G., Guisan A., Lahoz-Monfort J.J., Leitão P.J., Park D.S., Peterson A.T., Rapacciuolo G., Schmatz D.R., Schröder B., Serra-Diaz J.M., Thuiller W., Yates K.L., Zimmermann N.E., Merow C. A standard protocol for reporting species distribution models. *Ecography* 2020; **43**: 1261–1277.

Annexe A

Chapitre I – Partie II – Article I

AS1 Figure. Table of present *Boeckella* species and associated regions.

Boeckella species	Region
<i>Boeckella antiqua</i> Menu-Marque & Balseiro, 2000	South America
<i>Boeckella bergi</i> Richard, 1897	South America
<i>Boeckella brevicaudata</i> (Brady, 1875)	Antarctica & Sub-antarctic Islands
<i>Boeckella calcaris</i> (Harding, 1955)	South America
<i>Boeckella diamantina</i> Menu-Marque & Zúñiga, 1994	South America
<i>Boeckella gibbosa</i> (Brehm, 1935)	South America
<i>Boeckella gracilipes</i> Daday, 1901	South America
<i>Boeckella gracilis</i> (Daday, 1902)	South America
<i>Boeckella longicauda</i> Daday, 1901	South America
<i>Boeckella meteoris</i> Kiefer, 1928	South America
<i>Boeckella michaelsoni</i> (Mrázek, 1901)	South America
<i>Boeckella occidentalis</i> Marsh, 1906	South America
<i>Boeckella palustris</i> (Harding, 1955)	South America
<i>Boeckella poopuensis</i> Marsh, 1906	South America
<i>Boeckella poppei</i> (Mrázek, 1901)	South America
<i>Boeckella silvestri</i> (Daday, 1901)	South America
<i>Boeckella titicacae</i> Harding, 1955	South America
<i>Boeckella vallentini</i> (Scott T., 1914)	South America
<i>Boeckella bispinosa</i> Bayly, 1967	Australia
<i>Boeckella delicata</i> Percival, 1937	Australia & New Zealand
<i>Boeckella dilatata</i> Sars G.O., 1904	New Zealand
<i>Boeckella fluvialis</i> Henry, 1922	Australia
<i>Boeckella geniculata</i> Bayly, 1964	Australia
<i>Boeckella hamata</i> Brehm, 1928	New Zealand
<i>Boeckella major</i> Searle, 1938	Australia
<i>Boeckella minuta</i> Sars G.O., 1896	Australia
<i>Boeckella montana</i> Bayly, 1964	Australia
<i>Boeckella nyoraensis</i> Searle, 1912	Australia
<i>Boeckella opaqua</i> Fairbridge, 1945	Australia
<i>Boeckella pilkililli</i> Quinlan & Bayly, 2017	Australia

Boeckella propinqua Sars G.O., 1904	Australia & New Zealand
Boeckella pseudochelae Searle, 1912	Australia
Boeckella robusta Sars G.O., 1896	Australia
Boeckella rubra Smith G.W., 1909	Australia
Boeckella saycei Sars G.O., 1908	Australia
Boeckella shieli Bayly, 1985	Australia
Boeckella spinogibba Defaye, 1998	New Caledonia
Boeckella symmetrica Sars G.O., 1908	Australia
Boeckella tanea Chapman, 1973	New Zealand
Boeckella timmsi Bayly, 1998	Australia
Boeckella triarticulata (Thomson G.M., 1883)	Australia & New Zealand (Introduced in Mongolia and Italy)
Boeckella sibleti sp. nov.	New Caledonia

AS2 Figure. Distribution of amplified *Boeckella sibleti* sp. nov. and *B. spinogibba*.

Locality	Sampling date	Total number	female	male	<i>Boeckella sibleti</i> (N)	<i>Boeckella spinogibba</i> Abundances of each haplotype													Haplotype richness	Catalog number
						B_spi_COI01	B_spi_COI02	B_spi_COI03	B_spi_COI04	B_spi_COI05	B_spi_COI06	B_spi_COI07	B_spi_COI08	B_spi_COI09	B_spi_COI10	B_spi_COI11	B_spi_COI12	B_spi_COI13		
HYNC920	5 June 2018	5	3	2	0	2	1	0	0	0	2	0	0	0	0	0	0	0	3	MNHN-UI-2017-2303
HYNC925	16 June 2018	5	3	2	0	5	0	0	0	0	0	0	0	0	0	0	0	0	1	MNHN-UI-2017-2304
HYNC918	16 June 2018	10	6	4	0	0	0	10	0	0	0	0	0	0	0	0	0	0	1	MNHN-UI-2017-2305
HYNC4102	16 June 2018	8	6	2	0	4	0	2	0	0	0	0	0	0	1	1	0	0	4	MNHN-UI-2017-2306 MNHN-UI-2017-2307
HYNC4103	16 June 2018	10	6	4	0	0	0	0	9	0	0	0	1	0	0	0	0	0	2	MNHN-UI-2017-2308
HYNC4104	16 June 2018	9	5	4	0	0	0	0	0	8	0	1	0	0	0	0	0	0	2	MNHN-UI-2017-2309
HYNC931	16 June 2018	9	6	3	0	9	0	0	0	0	0	0	0	0	0	0	0	0	1	MNHN-UI-2017-2310 MNHN-UI-2017-2311
HYNC927	16 June 2018	6	3	3	0	0	6	0	0	0	0	0	0	0	0	0	0	0	1	MNHN-UI-2017-2312
HYNC4109	16 June 2018	4	3	1	0	0	4	0	0	0	0	0	0	0	0	0	0	0	1	MNHN-UI-2017-2313
HYNC4111	16 June 2018	5	2	3	0	0	4	0	0	0	0	0	0	0	0	0	0	1	2	MNHN-UI-2017-2314
HYNC928	5 June 2018	10	7	3	10	0	0	0	0	0	0	0	0	0	0	0	0	0	1	See 'Description' in text
HYNC4110	16 June 2018	3	3	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	3	MNHN-UI-2017-2315
Total		84	53	31	10	20	16	12	9	8	2	1								

AS3 Figure. Accession numbers for sequences obtained for *Boeckella sibleti* sp. nov. and *B. spinogibba*.

Species	Haplotype	NCBI accession numbers
<i>Boeckella spinogibba</i>	B_spi_COI01	OQ117983.1; OQ117984.1; OQ117987.1; OQ117992.1; OQ117993.1; OQ118000.1; OQ118001.1; OQ118003.1; OQ118010.1; OQ118011.1; OQ118018.1; OQ118019.1; OQ118021.1; OQ118024.1; OQ118025.1; OQ118029.1; OQ118030.1; OQ118032.1; OQ118035.1; OQ118040.1
	B_spi_COI02	OQ118002.1; OQ118014.1; OQ118022.1; OQ118033.1; OQ118042.1; OQ118043.1; OQ118047.1; OQ118048.1; OQ118049.1; OQ118052.1; OQ118054.1; OQ118057.1; OQ118059.1; OQ118061.1; OQ118062.1; OQ118064.1
	B_spi_COI03	OQ117985.1; OQ117986.1; OQ117994.1; OQ117995.1; OQ118004.1; OQ118005.1; OQ118012.1; OQ118015.1; OQ118020.1; OQ118023.1; OQ118031.1; OQ118034.1
	B_spi_COI04	OQ117988.1; OQ117997.1; OQ117998.1; OQ118007.1; OQ118008.1; OQ118016.1; OQ118026.1; OQ118036.1; OQ118037.1
	B_spi_COI05	OQ117990.1; OQ117991.1; OQ117999.1; OQ118009.1; OQ118017.1; OQ118027.1; OQ118038.1; OQ118039.1
	B_spi_COI06	OQ118013.1; OQ118041.1
	B_spi_COI07	OQ118028.1
	B_spi_COI08	OQ117989.1
	B_spi_COI09	OQ118058.1
	B_spi_COI10	OQ118006.1
	B_spi_COI11	OQ117996.1
	B_spi_COI12	OQ118053.1
	B_spi_COI13	OQ118044.1
	B_spi_28S01	OR612785.1
<i>Boeckella sibleti</i> sp. nov.	B_sib_COI01	OQ118045.1; OQ118046.1; OQ118050.1; OQ118051.1; OQ118055.1; OQ118056.1; OQ118060.1; OQ118063.1; OQ118065.1; OQ118066.1
	B_sib_28S01	OR612784.1

AS4 Figure. Accession numbers for sequences used in producing the 28S+COI phylogeny after Maturana *et al.* (2021).

Species	Haplotype after Maturana <i>et al.</i> 2021	Locality	NCBI accessions	
			COI	28S
<i>Boeckella spinogibba</i>	This study M156	New Caledonia	OQ117983.1	OR612785.1
<i>Boeckella sibleti</i> nov. sp.	This study M223	New Caledonia	OQ118045.1	OR612784.1
<i>Boeckella brasiliensis</i>	Hap_106	Patagonia	MZ757054.2	MZ773003.1
	Hap_90	Patagonia	MZ757055.2	MZ772978.1
<i>Boeckella vallentini</i>	Hap_107	Crozet Islands	MZ757051.1	MZ772977.1
	Hap_108	Kerguelen Islands	MZ757026.3	MZ772986.1
	Hap_109	Kerguelen Islands	MZ757027.3	MZ772986.1
<i>Boeckella poppei</i> (clade i)	Hap_111	South Shetland Islands	OP547788.1	MN088122.1
	Hap_112	Signy Island	OP547854.1	MN088122.1
	Hap_117	South Georgia	OP547806.1	MN088122.1
	Hap_116	South Georgia	OP547805.1	MN088122.1
	Hap_115	Antarctic Peninsula	OP547791.1	MN088122.1
	Hap_113	Patagonia	OP547831.1	OP547503.1
<i>Boeckella poppei</i> (clade ii)	Hap_120	Falkland	MZ757074.1	MZ772979.1
	Hap_123	Patagonia	MZ757076.1	MN088122.1
	Hap_126	Patagonia	MZ757028.2	MZ772982.2
<i>Boeckella poppei</i> (clade iii)	HpIH02	Hornos Island	MZ757030.1	MZ772983.1
	HpIH04	Hornos Island	MZ757031.1	MZ772981.1
<i>Boeckella pooensis</i>	Hap_16	South Argentina	MZ757028.1	GU049746.1
<i>Boeckella bergi</i>	Hap_18	South Argentina	DQ356543.1	GU049735.1
<i>Boeckella antiqua</i>	Hap_129	Patagonia	MZ757053.1	GU049733.1
<i>Boeckella meteoris</i>	Hap_38	Patagonia	MZ757063.2	GU049738.1
	Hap_41	Patagonia	MZ757066.2	MZ772998.1
	Hap_47	Patagonia	MZ757035.1	GU049738.1
	Hap_52	Magallanes	MZ757077.2	GU049738.1
<i>Boeckella michaelsoni</i>	Hap_6	Tierra del Fuego	MZ757026.1	GU049742.1
<i>Boeckella brevicaudata</i>	Hap_22	Navarino Island	MZ757047.2	MZ773000.1
	Hap_26	Navarino Island	MZ757051.2	MZ772999.1
	Hap_19	Kerguelen Islands	MZ757027.3	MZ772985.1
<i>Boeckella pseudochelae</i>	Hap_68	Australia	MZ757046.1	MZ772992.1
<i>Boeckella symmetrica</i>	Hap_70	New Zealand	MZ757081.1	MZ772995.1
<i>Boeckella triarticulata</i>	Hap_71	Australia	MZ757082.1	MZ772994.1
<i>Boeckella triarticulata</i>	Hap_74	Australia	MZ757083.1	GU049755.1

<i>Boeckella triarticulata</i>	Hap_72	Australia	MZ757084.1	GU049756.1
<i>Boeckella fluvialis</i>	Hap_75	Australia	MZ757086.1	GU049753.1
<i>Boeckella fluvialis</i>	Hap_76	Australia	MZ757087.1	MZ772989.1
<i>Boeckella propinqua</i>	Hap_80	New Zealand	MZ757091.1	MZ772993.1
<i>Boeckella tanea</i>	Hap_84	New Zealand	MZ757095.1	MZ772996.1
<i>Boeckella hamata</i>	Hap_78	New Zealand	MZ757089.1	MZ772990.1
<i>Parabroteas sarsi</i>	Hap_1	Argentina	DQ356574.1	GU049767.1
<i>Calamoecia ampulla</i>	Hap_5	Australia	JX436336.1	GU049758.1

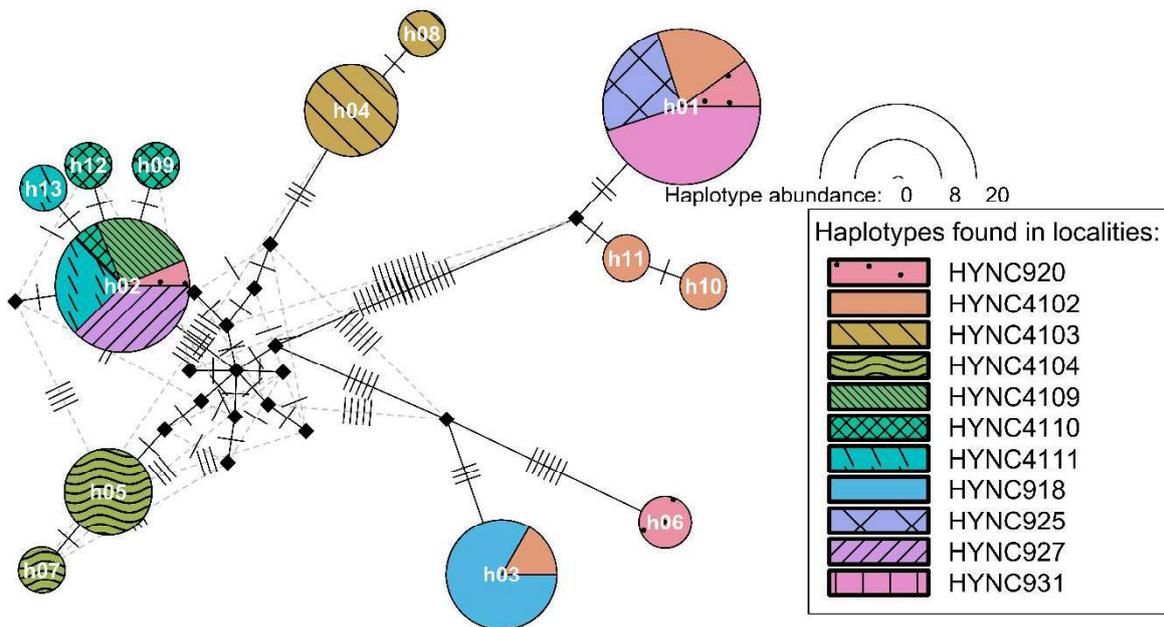
AS6 Figure. Raw results from bPTP analyses.

bPTP analysis on 28S+COI phylogeny (Fig. 8): https://species.h-its.org/ptp/result/?job_id=124210&email=coline.royaux@mnhn.fr

bPTP analysis on COI phylogeny (Fig. AS5): https://species.h-its.org/ptp/result/?job_id=124367&email=coline.royaux@mnhn.fr

AS7 Figure. Haplotype network with all alternative links and median vectors of 74 individuals of *B. spinogibba*.

S7 Figure. Haplotype network with all alternative links (dotted grey lines) and median vectors of 74 individuals of *B. spinogibba* from 12 ponds in the Plaine des Lacs. Haplotype names, originally of the form B_spi_COIXX, are abbreviated to hXX for the purpose of this representation.



Annexe B

Chapitre I – Partie II – Article II

Table BS1 - Haplotypes and corresponding National Center for Biotechnology Information (NCBI) accession numbers published in the present article

Species	Haplotype	NCBI accession numbers
<i>Boeckella spinogibba</i>	B_spi_COI01	PQ181012.1 ; PQ181013.1 ; PQ181035.1 ; PQ181076.1 ; PQ181077.1 ; PQ181078.1 ; PQ181079.1
	B_spi_COI02	PQ181040.1 ; PQ181041.1 ; PQ181042.1 ; PQ181043.1 ; PQ181044.1 ; PQ181045.1 ; PQ181046.1 ; PQ181047.1 ; PQ181048.1 ; PQ181082.1 ; PQ181083.1 ; PQ181084.1 ; PQ181085.1
	B_spi_COI06	PQ181051.1 ; PQ181052.1 ; PQ181053.1 ; PQ181054.1 ; PQ181055.1 ; PQ181056.1 ; PQ181057.1 ; PQ181058.1
	B_spi_COI09	PQ181080.1
	B_spi_COI02	
	uncomplete seq.	-
	B_sib_COI14	PQ181037.1
	B_spi_COI14	PQ181038.1
	B_spi_COI15	PQ181049.1
	B_spi_COI16	PQ181036.1 ; PQ181081.1
B_spi_COI17	PQ181038.1	
B_spi_COI18	PQ181050.1	
<i>Boeckella sibleti</i>	B_sib_COI01	PQ181017.1
	B_sib_COI02	PQ181031.1 ; PQ181032.1 ; PQ181033.1 ; PQ181034.1
	B_sib_COI03	PQ181028.1
	B_sib_COI04	PQ181030.1
	B_sib_COI05	PQ181029.1 ; PQ181092.1
	B_sib_COI06	PQ181018.1 ; PQ181019.1 ; PQ181020.1 ; PQ181021.1 ; PQ181022.1 ; PQ181023.1 ; PQ181090.1
	B_sib_COI06	
	uncomplete seq.	-
	B_sib_COI07	PQ181027.1
	B_sib_COI07	PQ181024.1 ; PQ181025.1 ; PQ181026.1 ; PQ181091.1
B_sib_COI08	PQ181093.1	
B_sib_COI09	PQ181089.1	
<i>Boeckella</i> sp. A	B_nsa_COI01	PQ181014.1 ; PQ181069.1 ; PQ181070.1 ; PQ181071.1 ; PQ181071.1 ; PQ181073.1 ; PQ181074.1 ; PQ181086.1 ;

		PQ181087.1
	B_nsa_28S01	PQ186733.1
<i>Boeckella sp. B</i>	B_nsb_COI01	PQ181059.1 ; PQ181060.1 ; PQ181061.1 ; PQ181062.1 ; PQ181063.1 ; PQ181064.1 ; PQ181065.1 ; PQ181066.1 ; PQ181067.1 ; PQ181068.1 ; PQ181088.1
	B_nsb_28S01	PQ186731.1
<i>Boeckella sp. C</i>	B_nsc_COI01	PQ181011.1
	B_nsc_28S01	PQ186734.1
<i>Boeckella sp. D</i>	B_nsd_COI01	PQ181015.1 ; PQ181016.1 ; PQ181075.1
	B_nsd_28S01	PQ186732.1

Table BS2 - Accession number table for 28S+COI phylogeny (Fig. 2), after Maturana *et al.* (2021) and Royaux *et al.* (2024). Concatenations from this study are in bold at the beginning of the table.

Species	Name	Locality	NCBI accessions	
			COI	28S
<i>Boeckella sp. A</i>	M229	Grande Terre	PQ181014.1	PQ186733.1
<i>Boeckella sp. B</i>	M282	Grande Terre	PQ181062.1	PQ186731.1
<i>Boeckella sp. C</i>	M209	Grande Terre	PQ181011.1	PQ186734.1
<i>Boeckella sp. D</i>	M261	Belep	PQ181015.1	PQ186732.1
<i>Boeckella spinogibba</i>	M156	Grande Terre	OQ117983.1	OR612785.1
<i>Boeckella sibleti</i>	M223	Grande Terre	OQ118045.1	OR612784.1
<i>Boeckella brasiliensis</i>	Hap_106	Patagonia	MZ757054.2	MZ773003.1
	Hap_90	Patagonia	MZ757055.2	MZ772978.1
<i>Boeckella vallentini</i>	Hap_107	Crozet Islands	MZ757051.1	MZ772977.1
	Hap_108	Kerguelen Islands	MZ757026.3	MZ772986.1
	Hap_109	Kerguelen Islands	MZ757027.3	MZ772986.1
<i>Boeckella poppei</i> (clade i)	Hap_111	South Shetland Islands	OP547788.1	MN088122.1
	Hap_112	Signy Island	OP547854.1	MN088122.1
	Hap_117	South Georgia	OP547806.1	MN088122.1
	Hap_116	South Georgia	OP547805.1	MN088122.1
	Hap_115	Antarctic Peninsula	OP547791.1	MN088122.1
	Hap_113	Patagonia	OP547831.1	OP547503.1
<i>Boeckella poppei</i> (clade ii)	Hap_120	Falkland	MZ757074.1	MZ772979.1
	Hap_123	Patagonia	MZ757076.1	MN088122.1
	Hap_126	Patagonia	MZ757028.2	MZ772982.2
<i>Boeckella poppei</i> (clade iii)	HpIH02	Hornos Island	MZ757030.1	MZ772983.1
	HpIH04	Hornos Island	MZ757031.1	MZ772981.1
<i>Boeckella poopeensis</i>	Hap_16	South Argentina	MZ757028.1	GU049746.1
<i>Boeckella bergi</i>	Hap_18	South Argentina	DQ356543.1	GU049735.1
<i>Boeckella antiqua</i>	Hap_129	Patagonia	MZ757053.1	GU049733.1
<i>Boeckella meteoris</i>	Hap_38	Patagonia	MZ757063.2	GU049738.1
	Hap_41	Patagonia	MZ757066.2	MZ772998.1
	Hap_47	Patagonia	MZ757035.1	GU049738.1
	Hap_52	Magallanes	MZ757077.2	GU049738.1
<i>Boeckella michaelseni</i>	Hap_6	Tierra del Fuego	MZ757026.1	GU049742.1
<i>Boeckella brevicaudata</i>	Hap_22	Navarino Island	MZ757047.2	MZ773000.1
	Hap_26	Navarino Island	MZ757051.2	MZ772999.1
<i>Boeckella pseudochelae</i>	Hap_68	Australia	MZ757046.1	MZ772992.1
<i>Boeckella symmetrica</i>	Hap_70	New Zealand	MZ757081.1	MZ772995.1
<i>Boeckella triarticulata</i>	Hap_71	Australia	MZ757082.1	MZ772994.1
	Hap_74	Australia	MZ757083.1	GU049755.1
	Hap_72	Australia	MZ757084.1	GU049756.1
<i>Boeckella fluvialis</i>	Hap_75	Australia	MZ757086.1	GU049753.1
	Hap_76	Australia	MZ757087.1	MZ772989.1
<i>Boeckella propinqua</i>	Hap_80	New Zealand	MZ757091.1	MZ772993.1
<i>Boeckella tanea</i>	Hap_84	New Zealand	MZ757095.1	MZ772996.1
<i>Boeckella hamata</i>	Hap_78	New Zealand	MZ757089.1	MZ772990.1
<i>Parabroteas sarsi</i>	Hap_1	Argentina	DQ356574.1	GU049767.1
<i>Calamoecia ampulla</i>	Hap_5	Australia	JX436336.1	GU049758.1

Table BS3 - Haplotypes and corresponding accession numbers retrieved from the National Center for Biotechnology Information (NCBI) for population structure analyses.

Species	Haplotype	NCBI accession numbers
<i>Boeckella spinogibba</i>	B_spi_COI01	OQ117983.1; OQ117984.1; OQ117987.1; OQ117992.1; OQ117993.1; OQ118000.1; OQ118001.1; OQ118003.1; OQ118010.1; OQ118011.1; OQ118018.1; OQ118019.1; OQ118021.1; OQ118024.1; OQ118025.1; OQ118029.1; OQ118030.1; OQ118032.1; OQ118035.1; OQ118040.1
	B_spi_COI02	OQ118002.1; OQ118014.1; OQ118022.1; OQ118033.1; OQ118042.1; OQ118043.1; OQ118047.1; OQ118048.1; OQ118049.1; OQ118052.1; OQ118054.1; OQ118057.1; OQ118059.1; OQ118061.1; OQ118062.1; OQ118064.1
	B_spi_COI03	OQ117985.1; OQ117986.1; OQ117994.1; OQ117995.1; OQ118004.1; OQ118005.1; OQ118012.1; OQ118015.1; OQ118020.1; OQ118023.1; OQ118031.1; OQ118034.1
	B_spi_COI04	OQ117988.1; OQ117997.1; OQ117998.1; OQ118007.1; OQ118008.1; OQ118016.1; OQ118026.1; OQ118036.1; OQ118037.1
	B_spi_COI05	OQ117990.1; OQ117991.1; OQ117999.1; OQ118009.1; OQ118017.1; OQ118027.1; OQ118038.1; OQ118039.1
	B_spi_COI06	OQ118013.1; OQ118041.1
	B_spi_COI07	OQ118028.1
	B_spi_COI08	OQ117989.1
	B_spi_COI09	OQ118058.1
	B_spi_COI10	OQ118006.1
	B_spi_COI11	OQ117996.1
	B_spi_COI12	OQ118053.1
	B_spi_COI13	OQ118044.1
<i>Boeckella sibletii</i>	B_sib_COI01	OQ118045.1; OQ118046.1; OQ118050.1; OQ118051.1; OQ118055.1; OQ118056.1; OQ118060.1; OQ118063.1; OQ118065.1; OQ118066.1

Table BS4 - NGS reads used for each retrieved sequence.

Isolate name	Number of individuals used for DNA extraction	N reads	NCBI accessions
M156	1	985208	PQ181012.1
M209	1	612554	PQ181011.1 ; PQ186734.1
M223	50	1749992	PQ181017.1
M229	1	1065526	PQ181014 ; PQ186733.1
M263	1	1852528	PQ181016.1
M261	10	1219094	PQ181015.1 ; PQ186732.1
M282	1	1139698	PQ186731.1

Figure BS5 - Raw results from bPTP analyses. Monophyletic groups forming species according to the analyses are represented in red on the trees.

https://species.h-its.org/ptp/result/?job_id=282&email=coline.royaux@mnhn.fr

Table B56 - Uncorrected genetic distance matrix between each haplotype.

		<i>Boeckella spinogibba</i>																		<i>Boeckella sibirici</i>								<i>Boeckella sp. nov. A</i>	<i>Boeckella sp. nov. B</i>	<i>Boeckella sp. nov. C</i>				
		B_spi_COI01	B_spi_COI02	B_spi_COI03	B_spi_COI04	B_spi_COI05	B_spi_COI06	B_spi_COI07	B_spi_COI08	B_spi_COI09	B_spi_COI10	B_spi_COI11	B_spi_COI12	B_spi_COI13	B_spi_COI14	B_spi_COI15	B_spi_COI16	B_spi_COI17	B_spi_COI18	B_sib_COI01	B_sib_COI02	B_sib_COI03	B_sib_COI04	B_sib_COI05	B_sib_COI06	B_sib_COI07	B_sib_COI08	B_sib_COI09						
<i>Boeckella spinogibba</i>	B_spi_COI02	2.13																																
	B_spi_COI03	2.89	1.67																															
	B_spi_COI04	2.43	1.22	1.98																														
	B_spi_COI05	2.28	0.76	1.82	1.06																													
	B_spi_COI06	2.74	2.13	1.37	2.43	2.28																												
	B_spi_COI07	2.43	0.91	1.98	1.22	2.43	2.43	2.43	1.52																									
	B_spi_COI08	2.58	1.37	2.13	0.15	2.22	2.58	1.37																										
	B_spi_COI09	1.98	0.15	1.82	1.37	0.91	2.28	2.43	2.58	1.98																								
	B_spi_COI10	0.46	1.98	2.74	2.28	2.13	2.58	2.28	2.43	1.82	0.15																							
	B_spi_COI12	2.28	0.15	1.82	1.37	0.76	2.28	0.91	1.52	0.30	2.28	2.13																						
	B_spi_COI13	2.28	0.15	1.82	1.37	0.91	2.28	1.06	1.52	0.30	2.28	2.13	0.30																					
	B_spi_COI14	2.28	0.15	1.82	1.37	0.91	2.28	1.06	1.52	0.30	2.28	2.13	0.30	0.30																				
	B_spi_COI15	2.89	1.98	1.22	2.28	2.13	2.58	2.74	2.13	2.13	2.13																							
	B_spi_COI16	2.28	0.15	1.82	1.37	0.91	2.28	1.06	1.52	0.30	2.28	2.13	0.30	0.30	0.30	2.13																		
	B_spi_COI17	2.28	0.15	1.82	1.37	0.91	2.28	1.06	1.52	0.30	2.28	2.13	0.30	0.30	0.30	2.13	0.30																	
	B_spi_COI18	2.89	2.28	1.52	2.58	2.43	0.15	2.58	2.74	2.43	2.43	2.58	2.74	2.43	2.43	2.43	2.43	2.43	2.43															
	<i>Boeckella sibirici</i>	B_sib_COI01	8.51	8.97	8.66	9.57	9.12	8.66	8.97	9.73	9.12	8.51	8.66	9.12	8.81	9.12	8.81	8.81																
		B_sib_COI02	9.57	9.57	9.57	10.2	10	9.73	10.2	10	9.73	9.57	9.73	9.73	9.42	9.73	9.42	9.88	9.88	3.95														
B_sib_COI03		9.73	9.73	9.73	10.3	10.2	9.88	10.3	10.2	9.88	9.73	9.88	9.88	9.57	9.88	9.57	9.88	9.57	10	4.10	0.15													
B_sib_COI04		9.73	9.73	9.73	10.3	10.2	9.88	10.3	10.2	9.88	9.73	9.88	9.88	9.57	9.88	9.57	9.88	9.57	10	4.10	0.15	0.30												
B_sib_COI05		9.42	9.42	9.42	10.0	9.88	9.57	10.3	9.88	9.57	9.42	9.57	9.57	9.27	9.57	9.27	9.57	9.27	9.73	3.80	0.15	0.30	0.30											
B_sib_COI06		9.73	9.73	9.73	10.3	10.2	9.57	10.3	10.2	9.88	9.73	9.88	9.88	9.57	9.88	9.57	9.88	9.57	9.73	4.41	1.37	1.52	1.52	1.52										
B_sib_COI07		9.57	9.57	9.57	10.2	10	9.42	10.2	10	9.73	9.57	9.73	9.73	9.42	9.73	9.42	9.73	9.42	9.57	4.26	1.22	1.37	1.37	1.37	0.15									
B_sib_COI08		9.57	9.57	9.57	10.2	10	9.73	10.2	10	9.73	9.57	9.73	9.73	9.42	9.73	9.42	9.73	9.42	9.88	3.95	0.30	0.46	0.46	0.15	1.67	1.52								
B_sib_COI09		8.81	9.88	9.27	10.2	10	8.97	10.2	10	10	9.12	9.27	10	9.73	10	9.12	10	9.73	9.12	3.34	2.74	2.89	2.89	2.89	2.58	2.43	3.04							
<i>Boeckella sp. nov. A</i>		10.2	10.2	9.88	10.8	10.3	9.73	10.2	10.9	10.3	10.2	10.3	10.3	10	10.3	9.73	10	10.3	9.88	8.51	7.75	7.90	7.90	7.90	8.21	8.36	8.05	8.05						
<i>Boeckella sp. nov. B</i>	9.27	9.42	9.27	9.73	9.57	9.42	9.42	9.57	9.57	9.27	9.42	9.57	9.42	9.27	9.57	9.27	9.27	9.27	7.60	6.08	6.23	6.23	6.23	6.84	6.69	6.38	6.53	6.53						
<i>Boeckella sp. nov. C</i>	10.8	11.1	11.5	11.7	11.1	12	10.9	11.5	11.1	11.4	11.2	10.9	10.9	11.2	11.8	11.2	10.9	11.8	10.3	11.1	11.2	11.2	11.2	10.9	11.4	11.2	10.9	12.2	10.5					
<i>Boeckella sp. nov. D</i>	9.12	9.57	9.27	9.88	9.57	9.57	9.42	9.73	9.73	9.12	9.27	9.42	9.42	9.42	9.73	9.42	9.73	9.42	5.17	5.02	5.17	5.17	5.17	4.86	5.47	5.32	5.02	5.47	8.21	6.23	6.23	10.3		

Figure BS7 - Median-Joining haplotype network of *B. spinogibba* with alternative links, black squares are median vectors; haplotype names, originally of the form B_spi_COIXX, are abbreviated to hXX for the purpose of this representation.

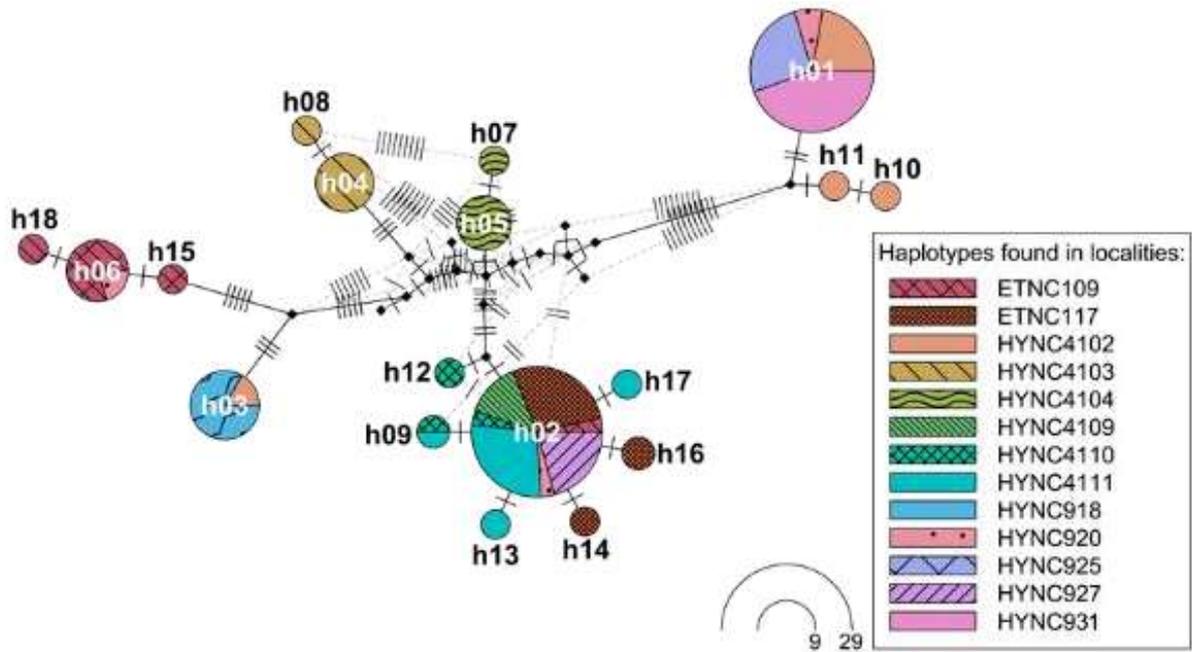


Figure BS8 - Parsimony haplotype network of *B. spinogibba*; haplotype names, originally of the form B_spi_COIXX, are abbreviated to hXX for the purpose of this representation.

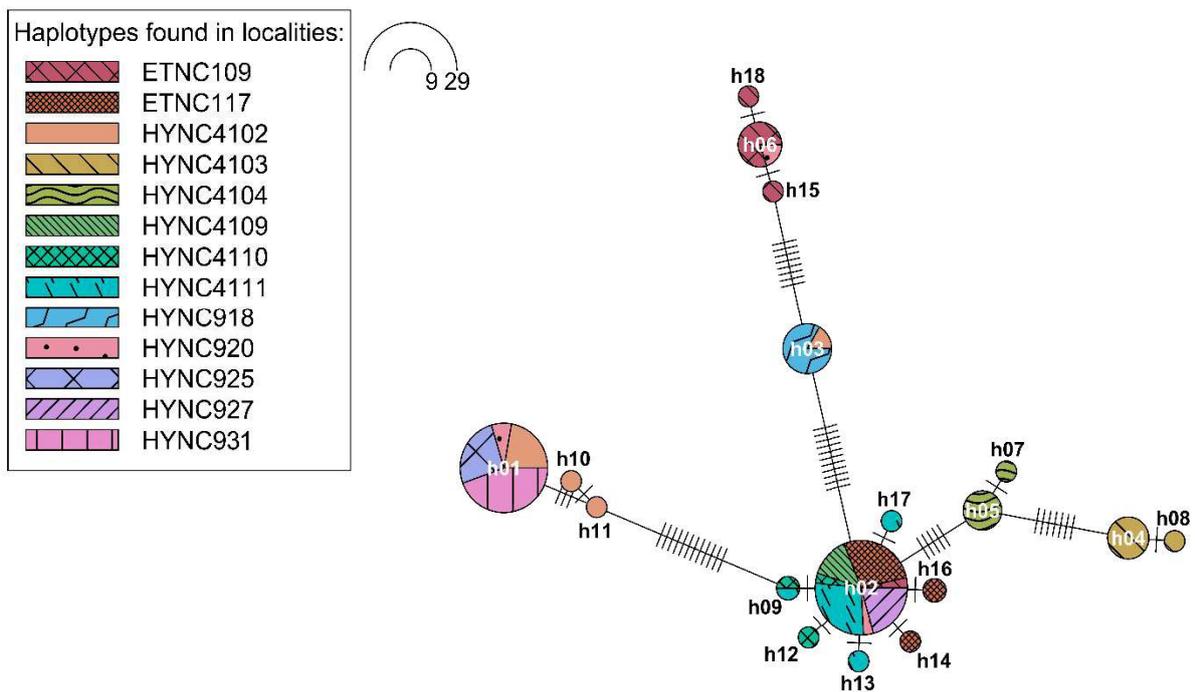


Figure BS9 - Median-Joining haplotype network of *B. sibleti* with alternative links, black squares are median vectors; haplotype names, originally of the form B_sib_COIXX, are abbreviated to hXX for the purpose of this representation.

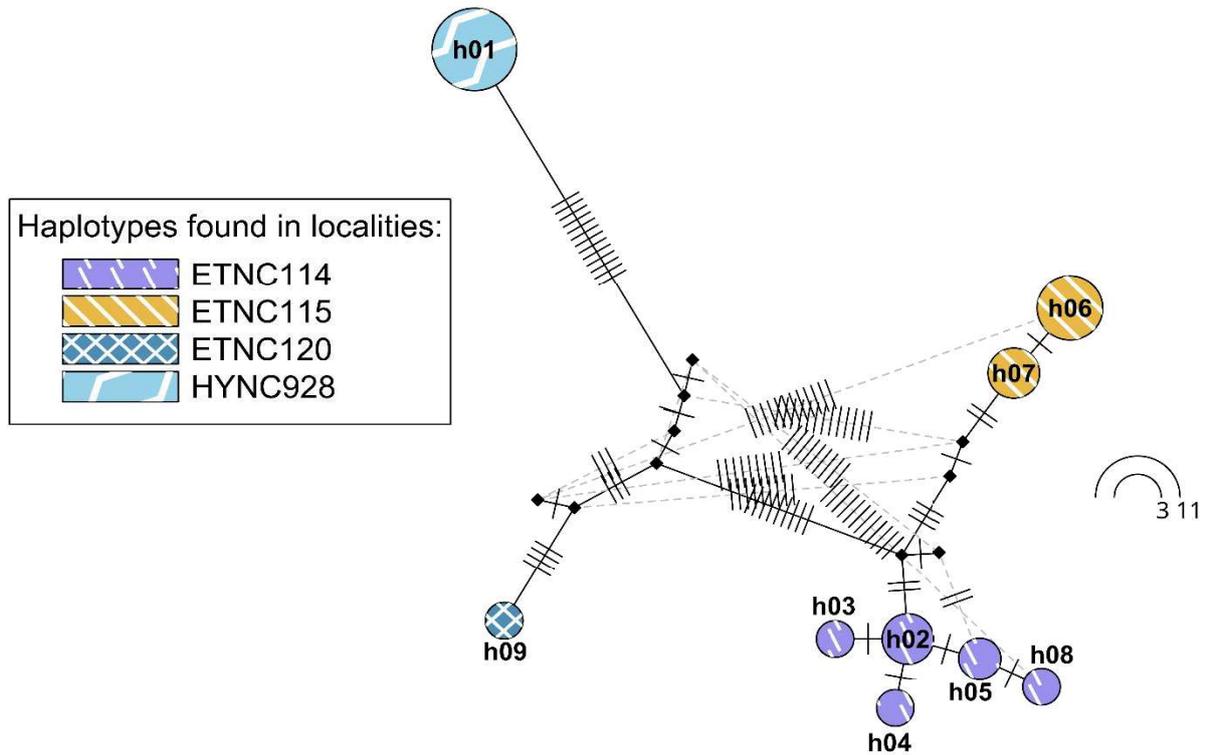
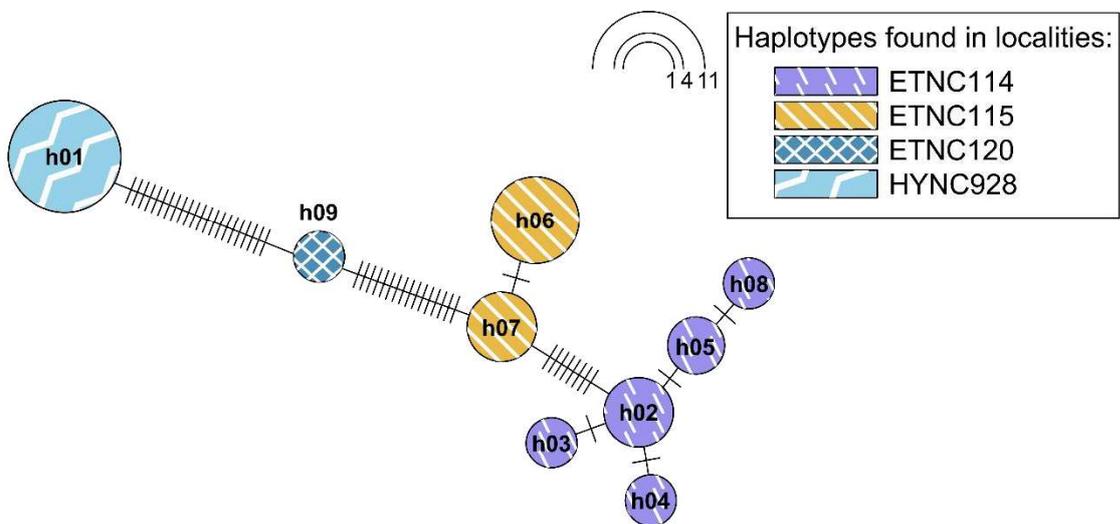


Figure BS10 - Parsimony haplotype network of *B. sibleti*; haplotype names, originally of the form B_sib_COIXX, are abbreviated to hXX for the purpose of this representation.



Annexe C

Chapitre I – Partie IV

Tableau CS1 – Numéros d’accessions NCBI pour la phylogénie COI du genre *Eulimndia* (fig. 1.3).

Espèce	Pays	Haplotype	Accessions NCBI
<i>Eulimnadia sp. A</i> (cette thèse)	NCK	E_nsa_COI01	
		E_nsa_COI07	
		E_nsa_COI04	
		E_nsa_COI06	
		E_nsa_COI02	
		E_nsa_COI03	
<i>Eulimnadia sp. C</i> (Schwentner <i>et al.</i> , 2015)	Australie		KP754149.1
			KP754151.1
			KP754152.1
			KP754153.1
<i>Eulimnadia sp. D</i> (Schwentner <i>et al.</i> , 2015)	Australie		KP754154.1
			KP754155.1 ; KP754156.1
<i>Eulimnadia sp. E</i> (Schwentner <i>et al.</i> , 2015)	Australie		KP754157.1 ; KP754158.1 ; KP754159.1
<i>Eulimnadia sp. F</i> (Schwentner <i>et al.</i> , 2015)	Australie		KP754163.1 ; KP754164.1 ; KP754165.1
<i>Eulimnadia sp. K</i> (Schwentner <i>et al.</i> , 2015)	Australie		KP754194.1 ; KP754195.1
			KP754198.1
			KP754201.1
<i>Eulimnadia sp. I</i> (Schwentner <i>et al.</i> , 2015)	Australie		KP754202.1
<i>Eulimnadia sp. G</i> (Schwentner <i>et al.</i> , 2015)	Australie		KP754167.1 ; KP754169.1 ; KP754172.1
			KP754171.1
			KP754168.1
<i>Eulimnadia sp. H</i> (Schwentner <i>et al.</i> , 2015)	Australie		KP754174.1
			KP754175.1 ; KP754176.1 ; KP754177.1 ; KP754179.1
			KP754178.1
			KP754184.1

		KP754182.1
		KP754183.1
		KP754180.1
		KP754181.1
<i>Eulimnadia sp. N</i> (Schwentner <i>et al.</i> , 2015)	Australie	KP754188.1
		KP754189.1
<i>Eulimnadia sp. O</i> (Schwentner <i>et al.</i> , 2015)	Australie	KP754190.1 ; KP754192.1
<i>Eulimnadia sp. M</i> (Schwentner <i>et al.</i> , 2015)	Australie	KP754203.1
<i>Eulimnadia sp. L</i> (Schwentner <i>et al.</i> , 2015)	Australie	KP754204.1
<i>Eulimnadia sp. 2</i> (Weeks <i>et al.</i> , 2009)	Australie	FJ499208.1
	Australie	FJ499209.1
	Australie	FJ499217.1
<i>Eulimnadia sp. B</i> (Schwentner <i>et al.</i> , 2015)	Australie	KP754185.1
	Australie	KP754186.1 ; KP754187.1
<i>Eulimnadia africana</i>	Botswana	FJ499195.1
	Afrique du Sud	FJ499202.1
	Botswana	FJ499220.1
<i>Eulimnadia braueriana</i>	Japon	EF189667.1
<i>Eulimnadia cylindrova</i>	Japon	FJ499139.1

Tableau CS2 – Numéros d’accessions NCBI pour la phylogénie COI du genre *Triops* (fig. 1.7).

Espèce	Pays	Haplotype	Accessions NCBI
<i>Triops longicaudatus</i>	USA		KM516710.1
			HQ908551.1
			AY639934.1
<i>Triops newberryi</i>	USA		HQ908516.1
			HQ908497.1
			HQ908514.1
			HQ908515.1
			KM516712.1
			HQ908510.1
			HQ908511.1
			HQ908498.1
<i>Triops longicaudatus</i>	Corée		JQ609205.1
			JQ609245.1
	Mexique		GQ144444.1
	USA		HM883938.1
			HM883939.1
	Corée		JQ609192.1
			JQ609189.1
			GU475465.1
			JQ609244.1
	USA		KM516711.1
			HQ908543.1
	<i>Triops sp.</i> (Korn <i>et al.</i> , 2013)	Bolivie	
<i>Triops sp.</i> (Naganawa, 2018)	Australie		LC222465.1
<i>Triops sp. P</i>	Australie	h96	KX162367.1
<i>Triops sp. H</i>	Australie	h98	KX162409.1
		h97	KX162406.1
		h101	KX162446.1
		h99	KX162411.1
		h100	KX162443.1
<i>Triops sp. M</i>	Australie	h95	KX162341.1
<i>Triops sp. F (1)</i>	Australie	h105	KX162501.1
		h107	KX162500.1
		h106	KX162499.1
(F2)	Australie	h102	KX162466.1
		h103	KX162416.1
		h103	KX162467.1
		h104	KX162470.1
<i>Triops sp. I</i>	Australie	h17	KX162426.1
		h13	KX162325.1
		h15	KX162330.1
		h16	KX162389.1
<i>Triops sp. N</i>	Australie	h12	KX162431.1

<i>Triops sp. O</i>	Australie	h6	KX162392.1
		h7	KX162429.1
		h8	KX162368.1
		h5	KX162430.1
		h9	KX162390.1
		h3 ; h4	DQ310624.1
		h11	KX162324.1
<i>Triops sp. C (1)</i>	Australie	h78	KX162450.1
		h81	KX162492.1
		h79	KX162452.1
		h80	KX162457.1
(C2)	Australie	h76	KX162454.1
		h77	KX162489.1
		h75	EF189677.1
		h76	KX162459.1
<i>Triops sp. T (1)</i>	Australie	h89	JN190396.1
		h90	JN190398.1
(T2)	Australie	h85	JN175239.1
		h86	JN175240.1
		h82	JN175230.1
		h83	JN175231.1
		h83	JN175232.1
		h87	JN175238.1
		h88	JN190397.1
		LC222466.1	
<i>Triops sp. S</i>	Australie	h91	JN175233.1
		h93	JN175244.1
		h92	JN175237.1
<i>Triops sp. Q</i>	Australie	h94	JN175245.1
<i>Triops sp. E</i>	Australie	h45	KX162419.1
		h46	KX162471.1
		h48	KX162493.1
		h47	KX162473.1
		h49	KX162472.1
<i>Triops sp. L</i>	Australie		LC222464.1
		h40	KX162420.1
		h39	JN175242.1
		h41	KX162455.1
		h44	KX162509.1
		h42	KX162491.1
		h43	KX162502.1
<i>Triops australiensis</i>	Australie	h35	KX162398.1
		h35	KX162400.1
		h36	KX162343.1
		h37	KX162508.1
		h33	KX162511.1
		h34	KX162484.1
		h30	KX162423.1
		h29	KX162355.1
		h31	KX162424.1
		h32	KX162510.1
		h38	KX162465.1
		h21	KX162449.1
		h19	KX162334.1

		h18	DQ310625.1
		h20	KX162335.1
		h22	KX162494.1
		h25	KX162444.1
		h24	KX162344.1
		h28	KX162365.1
<i>Triops sp. K</i>	Australie	h67	KX162361.1
		h66	KX162364.1
		h62	KX162326.1
		h63	KX162360.1
		h65	KX162363.1
		h64	KX162327.1
<i>Triops intermedius</i>	NCK	T_int_COI02 T_int_COI01	
(A2)	Australie	h50	KX162483.1
		h51	KX162482.1
(A1)	Australie	h60	KX162401.1
		h58	KX162339.1
		h59	KX162340.1
		h61	KX162347.1
		h54	KX162346.1
		h52	DQ343234.1
		h53	KX162378.1
		h56	KX162481.1
		h55	KX162433.1
		h57	KX162505.1
<i>Triops strenuus</i> (B1)	Australie	h73	KX162370.1
(B3)	Australie	h72	KX162369.1
<i>Triops strenuus</i>	Japon		LC222463.1
(B2)	Australie	h69	KX162381.1
		h68	DQ343235.1
		h70	KX162382.1
		h71b	KX162438.1
		h71	KX162440.1
(B4)		h74	KX162373.1
<i>Triops gracilis</i>	Australie	h108	JN175235.1
		h109	JN175236.1
<i>Triops sp. G</i>	Australie	h113	KX162462.1
		h114	KX162464.1
		h110	KX162460.1
		h112	KX162399.1
		h111	KX162410.1
<i>Triops sp. J</i>	Australie	h1	KX162359.1
		h1	KX162356.1
		h2	KX162357.1
<i>Lepidurus apus viridis</i>	Australie		HF911398.1

Tableau CS3 - Numéros d'accessions NCBI pour la phylogénie multilocus 18S+COI+16S du genre *Lynceus* (fig. 1.12).

Espèce	Pays	Indiv	Accessions NCBI		
			18S	COI	16S
<i>Lynceus planifascius</i>	Thailand	ZS004	MZ539580.1	MZ539940.1	MZ539593.1
<i>Lynceus grossipedia</i>	Mongolia	ZS056	MZ539575.1	MN515429.1	MZ539590.1
<i>Lynceus gracilicornis</i>	USA	ZS005	AF144215.1	MZ539937.1	MZ539589.1
<i>Lynceus biformis</i>	Japon	ZS017	MZ539571.1	MN515425.1	MZ539586.1
<i>Lynceus sp. MM</i>	Afrique du Sud	M176			
<i>Lynceus brachyurus</i>	Denmark	ZS031	MZ539573.1	MN515426.1	MK356563.1
	USA	ZS006	MZ539572.1	MN515427.1	MZ539588.1
<i>Lynceus huentelauquensis</i>	Chili	ZS021	MZ539576.1	MZ539938.1	MK356565.1
<i>Lynceus sp. B</i> (cette thèse)	Canala - NCK	M158			
<i>Lynceus insularis</i>	Goro - NCK	M287			
		ZS060	MZ539577.1	MZ539939.1	MZ539591.1
<i>Lynceus sp. A</i> (cette thèse)	Nouméa - NCK	M061			
	Mont Dore - NCK	M125			
	Koumac - NCK	M110			
	Australie	M106			
<i>Lynceus baylyi</i>	Australie	ZS002	MZ539570.1	MN515424.1	MZ539585.1
<i>Lynceus macleayanus</i>	Australie	ZS008	MZ539578.1	MN515430.1	MK356566.1
<i>Lynceus tatei</i>	Australie	ZS010	MZ539582.1	MN515433.1	MK356568.1
<i>Lynceus magdaleanae</i>	Australie	CS90	MZ539579.1	MN515431.1	MK356567.1
<i>Lynceus susanneae</i>	Australie	ZS003	MZ539581.1	MN515432.1	MZ539595.1
<i>Cyclestheria hislopi</i>	Thaïlande		MZ539569.1	MZ539935.1	MZ539584.1
<i>Triops cancriformis</i>	Japon		EF189638.1	AB084514.1	AB084514.1

Annexe D

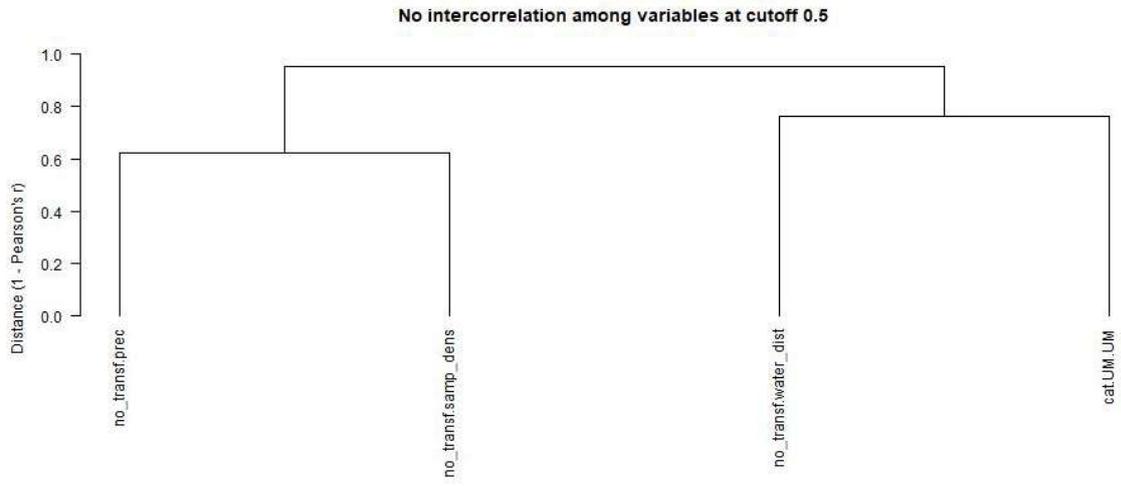
Chapitre II – Partie II

Table DS1 - Sampled locations sorted in full.

Sample single identifier	Longitude (x)	Latitude (y)	Sampling date(s)	Other names given to the location
HYNC4114	166.42577	-22.16381	17-06-2018	
HYNC4014	165.49944	-21.15934	07-06-2018	
HYNC4000	166.22052	-21.97417	04-06-2018	
HYNC918	166.95902	-22.26997	16-06-2018	HYNC4101
ETNC006	164.95308	-21.3421	15-04-2021	Pindai mare temporaire
HYNC920	166.9521	-22.26282	05-06-2018	HYNC4005
HYNC4111	166.95397	-22.26296	16-06-2018 12-06-2020	Doline croissante
HYNC4066	165.17555	-21.45432	13-06-2018	
HYNC4102	166.94141	-22.25926	16-06-2018 17-05-2022	
HYNC4109	166.95335	-22.26197	16-06-2018	
HYNC2608	165.86094	-21.48206	09-04-2018	
HYNC909	164.659	-20.628	17-11-2016	
ETNC110	166.99156	-22.28856	03-06-2020	DT41; DP41
ETNC109	166.95919	-22.25989	08-06-2020	DP46
HYNC2604	165.867	-21.48756	09-04-2018	
HYNC925	166.95259	-22.26107	16-06-2018 12-06-2020	HYNC4107; Lac en long
HYNC4110	166.95346	-22.26281	16-06-2018	
ETNC117	166.952	-22.26239	05-06-2020	
HYNC928	166.72775	-22.21552	25-03-2017 05-06-2018	HYNC4004
ETNC002	166.04828	-22.00455	14-01-2021	Ducos mare 1
ETNC107	166.96889	-22.27547	09-06-2020	DP13
ETNC120	166.62531	-22.07308	23-06-2020	Pourina; PPRB
HYNC4104	166.94773	-22.26835	16-06-2018	
ETNC116	166.90658	-22.32086	22-06-2020	DOL_11
ETNC113	166.89847	-22.30944	05-06-2020	DT70
HYNC4112	166.36215	-22.16813	17-06-2018	

HYNC2741	165.94847	-21.90796	08-12-2017	
HYNC2703	166.00233	-21.92528	07-12-2017	
ETNC004	166.05052	-22.00462	14-01-2021	Ducos mare 3
HYNC4098	166.82705	-22.27972	16-06-2018	
HYNC927	166.95324	-22.2621	16-06-2018	HYNC4108
HYNC4105	166.94776	-22.26257	16-06-2018	
ETNC008	164.94987	-21.3238	15-05-2021	Pindai mare temporaire 2
HYNC931	166.94835	-22.25917	16-06-2018 05-06-2020	HYNC4106; DT53
ETNC001	166.03328	-22.01154	14-01-2021	Mare aux canards
HYNC4003	166.22512	-22.03785	04-06-2018	
HYNC4060	165.19906	-21.33434	13-06-2018	
HYNC4079	164.45752	-20.72473	14-06-2018	
HYNC4012	165.50048	-21.16028	07-06-2018	
HYNC4008	166.39332	-22.26009	06-06-2018	
HYNC4075	164.46567	-20.7248	14-06-2018	
ETNC112	166.96989	-22.27567	09-06-2020	DT45; DP45
ETNC124	166.97336	-22.33189	26-06-2020	DP60
HYNC924	166.95164	-22.26234	05-06-2020	DT20
HYNC4009	166.40684	-22.27068	06-06-2018	
HYNC4055	164.66218	-20.96972	12-06-2018	
HYNC2747	166.00635	-21.93062	08-12-2017	
HYNC4084	164.46736	-20.8234	14-06-2018	
HYNC2802	165.46647	-21.51258	22-11-2017	HYNC1802
HYNC2672	165.81242	-21.86069	04-12-2017	
HYNC4103	166.94166	-22.26875	16-06-2018	
ETNC302	166.8873	-22.09387	19-05-2022	Unia02
ETNC310	166.66102	-22.20532	18-05-202	Doline mare Valentin
HYNC4019	165.34887	-21.03231	08-06-2018	
HYNC4002	166.23357	-21.97767	04-06-2018	
ETNC209	163.69716	-19.83093	24-04-2022	Belep09
HYNC4006	166.39224	-22.26134	06-06-2018	
HYNC4020	165.37911	-21.02222	08-06-2018	
ETNC401	166.15042	-21.74806	23-04-2023	Pic ningua
ETNC016	165.97887	-21.72764	22-02-2021	Mont do
HYNC2699	166.00122	-21.93533	05-12-2017	
HYNC4074	164.46548	-20.72657	14-06-2018	
HYNC4024	165.24526	-20.87638	08-06-2018	
ETNC404	166.62423	-22.05538	22-04-2023	Haute pourina nord
ETNC406	163.93282	-20.16094	01-04-2023	ABC_002

Figure DS2 – Dendrogramme de corrélation de Pearson



Annexe E

Chapitre III – Partie II

Annexe ES1 – Plan de gestion de données de la thèse

Plan de Gestion des Données sur le zooplancton et les mares de Nouvelle-Calédonie

Mots clés : Nouvelle-Calédonie, zooplancton, mares, eau douce, génétique des populations, taxonomie

Particulièrement dans le cadre de la thèse de Coline Royaux encadrée par Nicolas Rabet (HDR) et Yvan Le Bras (PhD).

Description du projet de recherche :

Le micro-endémisme en Nouvelle-Calédonie : Caractérisation à partir du zooplancton dulçaquicole, recherche des facteurs explicatifs majeurs et proposition d'un modèle d'écologie fonctionnelle

Le contexte biologique, géologique et climatique de la Nouvelle-Calédonie est particulièrement remarquable de par l'association de facteurs et de conditions à caractères exceptionnels. Bien qu'elle ait une superficie assez réduite (approximativement la même que les Pyrénées), elle est un des hotspots de biodiversité majeurs dans le monde et possède la plus grande proportion d'espèces endémiques décrites sur tous les territoires français (56% selon le Livret INPN, 2021). Il a été longtemps pensé que cette biodiversité remarquable était due à sa dissociation de l'Australie il y a environ 80 Ma ayant entraîné des spéciations indépendantes en Nouvelle-Calédonie mais il a été démontré que l'île était immergée jusqu'à il y a 37 Ma et aurait donc connu une re-colonisation postérieure par le vivant. L'évènement tectonique qui a engendré l'émersion de la Nouvelle-Calédonie a également provoqué un charriage du manteau terrestre en surface qui a formé, après altération, des massifs riches en métaux lourds connaissant aujourd'hui une forte pression d'exploitation par l'industrie minière.

Cette thèse comporte 3 volets majeurs, les deux premiers portent particulièrement sur la documentation et l'explication de l'endémisme et du micro-endémisme chez les communautés de zooplancton des mares de Nouvelle-Calédonie. En effet, les mares de Nouvelle-Calédonie restent des milieux méconnus et un travail préliminaire effectué à l'occasion de mon stage de Master 1, également au sein de l'équipe BIOPAC, suggère la présence de nombreuses espèces encore à découvrir et à décrire. La documentation de l'endémisme et du micro-endémisme au sein de ces communautés passera par la combinaison d'approches de génétique et de morphologie. A partir de cette documentation, un travail de modélisation (modèles de niche écologique, modèles de répartition prédictive,

extrapolation en zones à fort potentiel) permettra de caractériser les facteurs explicatifs majeurs de ce phénomène chez les communautés de zooplancton dulçaquicole.

Ce PGD est essentiellement destiné à répondre à ces deux premiers volets et constitue une partie du dernier axe qui a pour vocation de rendre les découvertes et connaissances produites par cette thèse directement actionnables pour la protection des écosystèmes. Cette partie de la thèse est particulièrement rendue possible par la collaboration avec l'UMS Patrimoine Naturel à travers le Pôle National de Données de Biodiversité et elle est née de plusieurs constats faits en écologie et autour des sciences de la conservation de la biodiversité. De fait, de grandes avancées ont été accomplies depuis les premières alertes majeures sur la perte de biodiversité dans les années 1960 : nous avons une recherche et une expertise mondiale de grande qualité dans ces domaines, des concepts fonctionnels, des méthodes très variées, adaptables et pratiques. Or, tout ceci est grandement sous-exploité et gagnerait à être plus transparent, reproductible et accessible.

Description du financement associé au projet :

Ecole Doctorale 227 : Co-financement du projet de thèse. Ecole Doctorale « Sciences de la Nature et de l'Homme : évolution et écologie ».

ERASMUS+ : Co-financement du projet de thèse. Projet centré sur l'analyse de données dans le cadre de la recherche pour la production de connaissances. Il se concentre essentiellement sur l'identification des compétences requises pour l'acquisition de savoirs à partir d'études génétiques à large échelle appliquées à l'écologie et la biodiversité.

Appel à Projets SUD de BOREA : Barcoding calanoïdes.

Action Transversale du Muséum : Acronyme planctoncal.

Le projet consiste à étudier le plancton d'eau douce de Nouvelle-Calédonie, qui a été jusque-là négligé, pour identifier et quantifier le micro-endémisme dans ce type d'habitat. Nous nous baserons sur les Branchiopodes et les Copépodes Calanoïdes (éléments dominants du plancton). Notre objectif, outre de combler les lacunes de connaissances, est de mieux localiser l'endémisme calédonien et les problèmes de conservation qui en découlent (activité minière, aménagement urbain ou touristique). Nous pensons être capables d'estimer si l'endémisme est intrinsèque à certains genres de Branchiopodes ou Copépodes et/ou plus à des milieux (mares de l'ouest côtier par opposition aux mares des milieux riches en nickel (ultramafique)) et identifier les taxons clés pouvant servir d'espèce(s) indicatrice(s) à la gestion de la conservation.

Contacts :

Coline Royaux – coline.royaux@mnhn.fr

Yvan Le Bras – yvan.le-bras@mnhn.fr

Nicolas Rabet – nicolas.rabet@mnhn.fr

Contributeurs :

Coline Royaux – coline.royaux@mnhn.fr

Yvan Le Bras – yvan.le-bras@mnhn.fr

Nicolas Rabet – nicolas.rabet@mnhn.fr
Frederic Busson
Felix Navarro
Lucie Lenglet
Gabrielle Ourous

Coûts associés à la gestion de données :

Environ 5% du temps travaillé de Coline Royaux.

Infrastructure : voir avec CESGO selon la quantité de données stockées et pour une idée du coût CO2

Description des jeux de données :

Les jeux de données produits dans le cadre de cette thèse se basent sur des séquençages d'individus de crustacés planctoniques de mares temporaires de Nouvelle-Calédonie qui ont été collectés en majeure partie à l'occasion des campagnes La Planète Revisitée entre 2016 et 2018. Une partie ont été et seront collectés par Nathalie Mary du bureau d'études ETHYCO et Nicolas Charpin de l'association vies d'Ô douce. Il est également possible qu'une phase de terrain s'organise en Nouvelle-Calédonie pendant la thèse pour enrichir davantage les jeux de données, à condition d'obtenir des financements suffisants.

Dans cette première version, les données sont composées de fichiers tabulés, images pour les localités et organismes étudiés et texte pour les séquences. Le fichier tabulé est composé sept feuilles :

- « Individuals » : Rassemble les données récoltées sur les individus séquencés ou observés au Microscope Electronique à Balayage (MEB), au microscope optique ou à la loupe binoculaire. Il renseigne sur le genre et l'espèce de l'individu, son identifiant de collection au Muséum National d'Histoire Naturelle (MNHN) et la ou les personne(s) ayant réalisé les manipulations en laboratoire, ayant saisi et nettoyé les données. S'il a été extrait, sa localisation parmi les puits des plaques d'extraction et de séquençage, les différentes amorces testées et les haplotypes retrouvés.
- « Samples » : Rassemble les données sur les différents échantillons en eau, de terre ou issus d'élevages récoltés. Il renseigne entre autre la localité d'origine, la date d'échantillonnage, le type d'échantillon et d'autres informations complémentaires.
- « Observation » : Rassemble les données d'observation de terrain qui ont pu être faites sur le terrain en direct, par des mesures (e.g. pH, température) et/ou à travers les photos prises sur place, des images satellite et des photographies aériennes. Il renseigne tout type d'informations pouvant être issues de ces différentes observations comme par exemple si la mare est en eau ou sèche.
- « Sorting » : Rassemble les données issues des tris des échantillons. Il renseigne l'échantillon utilisé, le genre des individus comptés, l'espèce si c'est possible et le comptage et/ou la présence-absence du taxon.

- « Localities » : Rassemble les données sur les localités observées et/ou échantillonnées. Il renseigne la latitude et longitude de la localité, si elle a déjà été échantillonnées ou non, si elle a pu être mal étiquetée sur le terrain, quand et sous quel nom et si elle est considérée comme étant sur un substrat ultramafique ou non.
- « Extraction_plates » : Rassemble les données sur les différentes plaques d'extraction ADN. Il renseigne le nombre d'échantillons de Nouvelle-Calédonie contenues dans la plaque, la date d'extraction et des commentaires si nécessaire.
- « Sequencing_plates » : Rassemble les données sur les différentes plaques de séquençage. Il renseigne le nombre d'échantillons de Nouvelle-Calédonie contenues dans la plaque, la date de séquençage et des commentaires si nécessaire.

Les photographies des localités prises sur le terrain sont rassemblées dans des noms de dossiers correspondant à leurs identifiants.

Les images des organismes observés au MEB, au microscope optique et à la loupe binoculaire sont rassemblées par genre et par technique d'observation. Les fichiers de séquences issues de la technique Sanger sont disponibles en version brutes, comme elles ont été communiquées par les laboratoires, dans des archives .zip contenant un fichier .fasta rassemblant toutes les séquences obtenues et un fichier .xls avec un rapport sur la longueur et la qualité des séquences. Enfin, cinq fichiers pour chaque brin d'ADN amplifié (H et L) de chaque individu séquencé (soit 10 fichiers pour un individu séquencé) :

- Un fichier .ab1 et un fichier .scf, contenant la séquence et le chromatogramme de séquençage éditable ;
- Un fichier .pdf, contenant la séquence et le chromatogramme de séquençage non-éditable ;
- Un fichier .seq, contenant la séquence entière ;
- Un fichier .clipped, contenant la séquence avec les extrémités de mauvaises qualités coupées.

Les séquences sont également rendues disponibles sous formes d'haplotypes nettoyés et alignés dans des fichiers .fasta pour chaque genre étudié.

Les séquences des amorces utilisées pour les amplifications et séquençages d'ADN sont également disponibles dans des fichiers .fasta.

Les métadonnées avec le détail du contenu de chaque type de fichier et champ de données seront renseignées en Ecological Metadata Language (EML) à l'aide de l'application MetaShARK.

Distribution des jeux de données :

Espace de travail collaboratif sur CESCO et elles seront rendues disponibles pour l'utilisation au territoire de la Nouvelle Calédonie et au public via un entrepôt de données national et/ou européen.

Les données issues des séquençages ADN seront publiées sur le National Center for Biotechnology Information (NCBI).

Licence des données et métadonnées :

CC BY 4.0.

Hébergeur des données :

Les données seront hébergées sur CESCO pendant la modification et à la fin de la thèse elles seront publiées sur un entrepôt de données national (CNRS-INEE) et/ou européen (ZENODO).

On prévoit également la publication des différents haplotypes sur le NCBI et la mise en collection des échantillons au MNHN.

Sécurité et sensibilité des données :

Certaines localités sont associées à des noms de propriétaires terriens à ne pas divulguer dans la version ouverte des données.

Ressources techniques nécessaires à la production, stockage et gestion des données :

Production des données : La production des données demande des ressources techniques au moment de l'échantillonnage (e.g. voiture pour se rendre sur les sites, filet à plancton) et des manipulations en laboratoire (e.g. loupe binoculaire, micropipette) dont les protocoles seront rendus disponibles à travers la plateforme www.protocols.io.

Stockage des données : La saisie des données pouvant être faite par le biais de plusieurs utilisateurs dans des lieux différents, il est nécessaire d'avoir accès à un serveur pour effectuer un travail en commun mis à jour automatiquement. Les données sont donc partagées entre les collaborateurs sur la plateforme CESCO et seront mises en synchronisation automatique sur autant de PC que nécessaires. Pour l'instant, le PC personnel de Coline Royaux et le PC professionnel de Nicolas Rabet.

Gestion des données : La gestion des données se fera avec les PC synchronisés, mais plus particulièrement celui de Coline Royaux. La gestion des métadonnées se fera sur MetaShARK.

Pour la gestion des données, le logiciel utilisé pour la saisie est Excel, ce choix n'est pas idéal dans le sens où Excel est un logiciel de calcul qui peut interpréter les entrées de manière indésirable (e.g. si un identifiant est nommé 1E3 il sera interprété comme 1000). L'idéal serait d'utiliser un simple logiciel de traitement de texte tabulé qui permet de faire des entrées de texte sans interprétation dans des cases délimitées pour ensuite l'exporter en n'importe quel format. Or, un tel logiciel n'a pas été trouvé et Excel est le plus utilisé, c'est pourquoi le choix se porte sur ce logiciel.

Standard(s) de métadonnées utilisé(s) pour la description des données :

Ecological Metadata Language (EML).

Description fine des champs et règles de saisie

Règles générales :

- TOUS LES CARACTERES SPECIAUX SONT A EVITER : é è ^ \$ £ # ' (| \ » ; , ? / : € à @ etc ...
- On utilise le tiret du 8 pour séparer à la place des espaces

Ta b	champ	contient	règles	REGEX	exemple
Individuels	ID_ind	Identifiant unique des individus séquencés ou observés	<ul style="list-style-type: none"> - Partie rendant l'identifiant unique, n'importe quelle chaîne de caractères d'au moins 1 caractère (chiffres, lettres majuscules et/ou minuscules). Dans le cas d'un individu dont l'ADN a été extrait, on prend l'identifiant de plaque et du puit d'extraction. - Lettres identifiant le genre de l'individu (lettres minuscules, 3 caractères au minimum) - Lettre majuscule marquant le début d'identifiant de localité - Identifiant de localité composée de 3 chiffres minimum (lettres minuscules acceptées) 	<code>[0-9A-Za-z]{5}_[A-Z][0-9]{2}[0-9a-z]{+}</code>	P8A1_boeck _H2608
	ID_sample	Identifiant échantillons desquels les individus ont été extraits	Voir règles dans la partie tableau « Samples »		
	genus	Genre identifié de l'individu	Nom latin entier du genre en minuscules Lettres minuscules seulement	<code>[a-z]{2}[a-z]{+}</code>	boeckella
	species	Espèce identifiée de l'individu	Nom latin entier de l'espèce en minuscules, séparations avec un tiret du 8 « _ », si nouvelle espèce : nov_sp_X (Attention de ne pas donner la même lettre X à deux espèces distinctes). Minuscules et « _ » seulement, 1 majuscule possible, valeurs NA autorisées	<code>[a-z]{+}_[a-z]{+}[a-z]{1}*[A-Z]?</code>	boeckella_no v_sp_A
	sex	Sexe de l'individu	Lettres minuscules seulement, valeurs NA autorisées	<code>(female male hermap hrodite juvenile putat ive_hermaphrodite)</code>	female
	obs_technique	Méthode d'identification et de	Microscope électronique à balayage (meb), loupe binoculaire (binoc_microscope), microscope optique (optic_microscope), ...	<code>[a-z]{+}</code>	binoc_micro scope

		sexage de l'individu	Lettres minuscules et tiret du 8 « _ » seulement		
breed_T_F	Est-ce un individu extrait d'un élevage ?		booléen, seuls TRUE ou FALSE sont autorisés Valeurs NA autorisées	(TRUE FALSE)	TRUE
extracted_T_F	A-t-on extrait l'ADN de cet individu ?		booléen, seuls TRUE ou FALSE sont autorisés	(TRUE FALSE)	FALSE
ID_ext_plate	Identifiant de la plaque d'extraction où a été extrait l'individu		Voir règles dans la partie tableau « Extraction_plates » Valeurs NA autorisées si l'individu n'a pas été extrait		
ext_plate_loc	Localisation de l'individu dans la plaque d'extraction		- Les 12 colonnes de la plaque sont représentées par les nombres de 1 à 12 - Les 8 lignes sont représentées par des lettres majuscules de A à H Valeurs NA autorisées si l'individu n'a pas été extrait Ici cas d'une plaque de 96 puits mais il y a d'autres possibilités	[A-H]([1-9])1[0-2]	B12
n_tested_y_primers_PCR	Combien de couples de primers ont été utilisés dans les PCR pour amplifier la région y de l'individu ?		Valeur chiffrée Valeurs NA autorisées si l'individu n'a pas été mis en PCR	[0-9]+	4
ID_y_primers_PCRx	Identifiant du couple de primers utilisé pour amplifier la région y de l'individu dans la tentative x		Identifiant pour le brin lourd (H) en premier puis le brin léger (L) Au sein des identifiants, lettres majuscules, minuscules et chiffres autorisés seulement Valeurs NA autorisées si l'individu n'a pas été mis en PCR	H[0-9A-Za-z]+_L[0-9A-Za-z]+	Hlync2_Lfolmer
positive_y_PCRx_T_F	La tentative x de PCR pour amplifier la région y de l'individu a fonctionné ?		booléen, seuls TRUE ou FALSE sont autorisés Valeurs NA autorisées si l'individu n'a pas été mis en PCR	(TRUE FALSE)	NA
ID_y_seqz_plate	Identifiant de la plaque séquençage de la tentative z de séquençage de la région y de l'individu		Voir règles dans la partie tableau « Sequencing_plates » Valeurs NA autorisées si l'individu n'a pas été séquençé		

y_seqz_plate_lo c	Localisation de l'individu dans la plaque séquençage de la tentative z de séquençage de sa région y	- Les 12 colonnes de la plaque sont représentées par les nombres de 1 à 12 - Les 8 lignes sont représentées par des lettres majuscules de A à H Valeurs NA autorisées si l'individu n'a pas été séquençé	[A-H]([1-9] 1[0-2])	A2
ID_y_primers_s eqz	Identifiant du couple de primers utilisé pour séquençer la région y de l'individu dans la tentative z	Identifiant pour le brin lourd (H) en premier puis le brin léger (L) Au sein des identifiants, lettres majuscules, minuscules et chiffres autorisés seulement Valeurs NA autorisées si l'individu n'a pas été séquençé	H[0-9A-Za-z]+_L[0-9A-Za-z]+	Hlync2_Llyn c2
positive_y_seqz _T_F	La tentative z de séquençage de la région y de l'individu a-t-elle fonctionné ?	booléen, seuls TRUE ou FALSE sont autorisés Valeurs NA autorisées si l'individu n'a pas été séquençé	(TRUE FALSE)	TRUE
y_haplotype	De quel haplotype de la région y est l'individu ? Identifiant unique de l'haplotype trouvé chez l'individu décrit	- Première lettre du genre de l'individu en majuscule - 3 premières lettres de l'espèce de l'individu en minuscules - Acronyme de la région ADN concernée en majuscules, minuscules ou chiffres - Numéro aloué à l'haplotype en 2 chiffres Valeurs NA autorisées si l'individu n'a pas été séquençé ?????	[A-Z]_[a-z]{3}_[A-Za-z0-9]+[0-9]{2}	B_nsa_CO10 9
ID_collection_ MNHN	?????	?????	?????	?????
input_operator	Personne qui a entré les données	nom_prenom minuscules seulement, si prénom ou nom composé on colle sans espace ni tiret	[a-z]+_[a-z]+	annemarie_d upont
lab_operator	Personne qui a fait les manip en labo	nom_prenom minuscules seulement, si prénom ou nom composé on colle sans espace ni tiret	[a-z]+_[a-z]+	coline_royau x
ID_sample	Identifiant unique des échantillons	- Partie rendant l'identifiant unique, n'importe quelle chaîne de caractères d'au moins 1 caractère (chiffres, lettres majuscules et/ou minuscules). Dans le cas d'un échantillon d'élevage, on met le nombre de jour en élevage au moment du prélèvement suivi de l'identifiant d'élevage.	[0-9A-Za-z]+_[WSB]_[A-Z][0-9]{2}[0-9a-z]+_[0-9]{4}	37MA233_B _H2608_201 8
Samples				

	échantillon ?	colle sans espace ni tiret		upont
input_operator	Personne qui a entré les données	nom_prenom minuscules seulement, si prénom ou nom composé on colle sans espace ni tiret	[a-z]+_[a-z]+	annemarie_d upont
ID_locality	Identifiant unique de la localité	- Suite de lettres majuscules. Par exemple, HYNC est l'identifiant de La Planète Revisitée en Nouvelle-Calédonie. - Suite de 3 chiffres minimum	[A-Z]+[0-9]{2}[0-9]+[a-z]?	HYNC187
y_latitude	Latitude	Nombre réel en degré décimal (décimale notée avec un point)	-90 ≤ R ≤ 90	-21.699833
x_longitude	Longitude	Nombre réel en degré décimal (décimale notée avec un point)	-180 ≤ R ≤ 180	165.558639
sampled_T_F	La localité a-t-elle déjà été échantillonnée ?	booléen, seuls TRUE ou FALSE sont autorisés	(TRUE FALSE)	TRUE
anterior_locality_ID	Si erreur quand on a nommé la localité sur place précédemment, le « faux » nom	- Année où la localité a été nommée de la mauvaise façon - Faux nom qui a été donné : Suite de lettres majuscules + suite de 3 chiffres minimum Valeurs NA autorisées si aucune erreur n'a été faite	[0-9]{4}_[A-Z]+[0-9]{2}[0-9]+	2016_HYNC 1875
posterior_locality_ID	Si erreur quand on a nommé la localité sur place les fois d'après, le « faux » nom	Pareil Valeurs NA autorisées si aucune erreur n'a été faite	[0-9]{4}_[A-Z]+[0-9]{2}[0-9]+	2018_HYNC 4098
NM_locality_ID	Identifiant de localité donné par Nathalie Mary	Souvent « Doline ... », souvent contient des caractères spéciaux etc ... donc a mettre entre guillemets " ... "	".+"	"Doline isolée"
ultramafic_T_F	La mare a-t-elle semblé être sur substrat ultramafique ?	booléen, seuls TRUE ou FALSE sont autorisés Champ à croiser avec couches des sols.	(TRUE FALSE)	TRUE
ID_obs	Identifiant unique des observations	- Partie rendant l'identifiant unique, n'importe quelle chaîne de caractères d'au moins 1 caractère (chiffres, lettres majuscules et/ou minuscules). - Type d'observation (P : photo directement prise sur place, L : observation faite sur place, M : mesure faite sur place, V : plusieurs types d'observations sur place (P,L et/ou M à la fois), A : photo aérienne, I : Image satellite)	[0-9A-Za-z]+_[PLMVAI]_[AZ][0-9]{2}[0-9a-z]+_[0-9]{4}	1_V_H2608 _2018

			<p>- Localité d'observation - Année d'observation</p>		
ID_locality	ID localité observée		Voir règles dans la partie tableau « Localities »		
obs_type	Type d'observation effectuée		photo : photo directement prise sur place, live : observation faite sur place, measure : mesure faite sur place, various : plusieurs types d'observations sur place, aerial_pic : photo aérienne, img_sat : Image satellite	(photo live measure various aerial_pic img_sat)	live
URL	Lien pointant vers l'origine de l'observation		Ici, seul champ où les caractères spéciaux sont autorisés. NA autorisés. L'URL doit être mis entre guillemets "...", peut être un lien internet ou pointer vers un objet contenu dans le jeu de données	","	"localities_photos/HYNC4104"
obs_date_DD_MM_YYYY	Date de l'observation		Sous la forme JJ_MM_AAAA Chiffres seulement	(0[1-9] [12][0-9] [3][01])_(0[1-9] [10-2])_[0-9]{4}	23_08_2017
obs_year	Année d'observation		Sous la forme AAAA, chiffres seulement	[0-9]{4}	2017
wet_T_F	Mare en eau ?		booléen, seuls TRUE ou FALSE sont autorisés	(TRUE FALSE)	TRUE
sample_T_F	Echantillon pris ?		booléen, seuls TRUE ou FALSE sont autorisés	(TRUE FALSE)	TRUE
ID_sample	ID de l'échantillon pris		Voir règles dans la partie tableau « Samples »		
pH	Mesure de pH		Nombre réel positif, décimale notée avec un point. Valeurs NA autorisées	0 ≤ R ≤ 15	7.8
conductivity_microS	Mesure de conductance		Nombre réel positif en microSiemens, décimale notée avec un point Valeurs NA autorisées	+R	197
temperature_celsius	Mesure de température		Nombre réel en degrés celsius, décimale notée avec un point Valeurs NA autorisées	R	20.2
pool_type	Type de mare		Minuscules et tiret du 8 « _ » seulement. Garder les mêmes façons d'écrire des milieux similaires, on reste dans la liste suivante dans la mesure du possible : - swamp : marécage, - puddle : flaque, - pond : mare, - rock_pool : bassin de roche, - hill_reservoir : réservoir collinaire, - hole : trou d'eau, - sinkhole : doline, - cutoff : bras mort, - mangrove	[a-z_]+	puddle
obs_superficy	Superficie de la mare		A mesurer en priorité au plus grand sur les images satellites mais possibilité de noter les mesures à d'autres moments. En m².		
div_infos	Informations qui ne		Minucules et tirets du 8 « _ » seulement, NA autorisés. Réduire les mots mais on en met assez pour ça		

	restent pas dans les champs renseignés	reste compréhensible, on essaye de mettre des tirets du 8 « _ » au lieu des espaces, pas de virgule, d'accent, d'apostrophe, ... Si on voit que certaines informations sont récurrentes (taille de la mare, ...) mais n'est pas encore dans les champs on peut créer un nouveau champ.
input_operator	Personne qui a entré les données	nom_prenom minuscules seulement, si prénom ou nom composé on colle sans espace ni tiret
observer	Personne qui a réalisé l'observation	nom_prenom minuscules seulement, si prénom ou nom composé on colle sans espace ni tiret
input_origin	Origine de l'observation	Origine des images satellites, etc ... Que minuscules et « _ », NA autorisés
ID_sample	ID de l'échantillon	Voir règles dans la partie tableau « Samples »
genus	Genre des individus	Nom latin entier du genre en minuscules Lettres minuscules seulement
species	Espèce identifiée des individus	Nom latin entier de l'espèce en minuscules, séparations avec un tiret du 8 « _ », si nouvelle espèce : nov_sp_X (Attention de ne pas donner la même lettre X à deux espèces distinctes). Minuscules et « _ » seulement, 1 majuscule possible, valeurs NA autorisées
certainty	L'identification est sûre à quel point ?	Note donnée par l'observateur : A : Sûre, B : Léger doute, C : Doute raisonnable, D : Gros doute, E : Incertain
presence_T_F	Est-elle présente ?	booléen, seuls TRUE ou FALSE sont autorisés. Vraisemblablement que des TRUE ici pendant la saisie, les FALSE seront ajoutés ensuite
stage	Quel(s) stade(s) de développement vu(s) ?	(eggs? larvae remain s? adults? juv)
sex_ratio_M	Proportion de mâles	0 ≤ ℝ ≤ 1
sex_ratio_F	Proportion de femelles	0 ≤ ℝ ≤ 1
sex_ratio_H	Proportion d'hermaphrodites	0 ≤ ℝ ≤ 1
juvenile_ratio	Proportion de juvéniles	0 ≤ ℝ ≤ 1
comment	Commentaire	Minuscules et tirets du 8 « _ » seulement, NA autorisés. Mêmes règles que « div_infos » dans « Observation ».
observer	Personne qui a trié	Mêmes règles que tous les champs où on entre des prénom nom [a-z]+ [a-z]+ annie_porti

Sortine

Extraction_plates		ID_ext_plate	Identifiant unique de la plaque d'extraction	P ou T majuscule suivi d'un identifiant d'un chiffre minimum P : plaque de 96 puits ensembles. T : tubes eppendorf séparés.	[PT][0-9]+	P8
n_samp_NC	Nombre d'échantillons d'intérêt dans la plaque			Ici, échantillons de plancton de Nouvelle-Calédonie seulement Nombre entier seulement	[0-9]+	67
extraction_date_DD_MM_YYYY	Date de l'extraction ADN			Sous la forme JJ_MM_AAAA Chiffres seulement	(0[1-9] [12][0-9] [301])_(0[1-9] [10-2])_[0-9]{4}	14_12_2018
extraction_year	Année d'extraction			Sous la forme AAAA, chiffres seulement	[0-9]{4}	2018
comment	Commentaire			Sur la plaque s'il y a un incident par exemple, même règle que « div_infos » dans « Observation », NA autorisés		
Sequencing_plates		ID_seq_plate	Identifiant unique de la plaque d'extraction	Une lettre majuscule (E pour Eurofins par exemple) suivie de l'année de séquençage et d'un identifiant d'un chiffre minimum	[A-Z][0-9]{4}_[0-9]+	E2019_19
n_ech_NC	Nombre d'échantillons d'intérêt dans la plaque			Ici, échantillons de plancton de Nouvelle-Calédonie seulement Nombre entier seulement	[0-9]+	67
seq_date_DD_MM_YYYY	Date du séquençage			Sous la forme JJ_MM_AAAA Chiffres seulement	(0[1-9] [12][0-9] [301])_(0[1-9] [10-2])_[0-9]{4}	16_12_2018
seq_year	Année de séquençage			Sous la forme AAAA, chiffres seulement	[0-9]{4}	2018
comment	Commentaire			Sur la plaque s'il y a un incident par exemple, même règle que « div_infos » dans « Observation », NA autorisés		

Occurrences of the Boeckella genus in New Caledonia

2024-02-12

Abstract

This dataset contains the occurrences of the copepod genus *Boeckella* in New Caledonia gathered during an expedition in 2018 of the program "La Planète Revisitée". It is the second time this genus is recorded in New Caledonia and a new species has been found *Boeckella sibleti* sp. nov. This dataset is used in the article by Royaux et al., 2023. It is a subset of Coline Royaux's PhD thesis dataset which will be updated through the research and validation process. Partial COI (Cytochrome Oxidase I) gene of 74 individuals of *Boeckella spinogibba* and 10 individuals of *Boeckella sibleti* has been sequenced with Sanger method. Original AB1 data files for these sequenced are shared along with tabular data files on individuals (containing e.g. sample ID, species and sex identified, NCBI GenBank accession for cleaned sequences and used primers for amplifications), occurrences (containing e.g. sample ID, species identified, French Natural History Museum (MNHN) collection ID), samples (containing e.g. sample type, location ID, sampling date, sampler name), locations (containing e.g. other names of the location, coordinates in decimal degrees WGS84) and DNA extractions (containing e.g. extraction date, laboratory operator) and fasta files with all found haplotypes sequences and primer used. DNA extraction (<https://dx.doi.org/10.17504/protocols.io.j8nlk4x7wg5r/v1>) and PCR (dx.doi.org/10.17504/protocols.io.261ge4y4jv47/v2) protocols are available on protocols.io. Sequence cleaning workflow is available on Galaxy at https://usegalaxy.fr/u/coline_royaux/w/sanger1--from-ab1-to-aligned-fasta-imported-from-uploaded-file and history retracing data cleaning steps at https://usegalaxy.fr/u/coline_royaux/h/automatic-boeckella-nc-data-cleaning-1.

Metadata Provider

[Author list](#)

Creators

- Coline Royaux coline.royaux@mnhn.fr
- Nicolas Charpin nicolas.charpin@vod.fr
- Nicolas Rabet nicolas.rabet@mnhn.fr

Associated parties

Keywords

- Biodiversity
- Crustacea
- Zooplankton
- Endemic animals
- Population genetics
- Phylogeny
- Aquatic biodiversity
- Freshwater biodiversity
- Pond animals

Annotations

Temporal coverage

Start date	End date
2018-06-01	2018-06-30

Geographic coverage

New caledonia

West:	161
East:	169
North:	-17
South: -24	

Taxonomic coverage

Rank	Value
------	-------

Data tables

extractions_boeck1_2023-02-01.tsv

Physical: extractions_boeck1_2023-02-01.tsv

attribute name	definition	format, unit or codes
ID_ext_plate	Description for: ID_ext_plate	
n_samp_NC	Description for: n_samp_NC	dimensionless
extraction_date_DD_MM_YYYY	Description for: extraction_date_DD_MM_YYYY	DD_MM_YYYY

attribute name	definition	format, unit or codes
extraction_year	Description for: extraction_year	YYYY
comment	Description for: comment	96_wells_plate = Value: 96_wells_plate for attribute: comment

individuals_boeck1_2023-07-14.tsv

Physical: individuals_boeck1_2023-07-14.tsv

attribute name	definition	format, unit or codes
ID_ind	Description for: ID_ind	
ID_sample	Description for: ID_sample	
genus	Description for: genus	boeckella = Value: boeckella for attribute: genus
species	Description for: species	boeckella_spinogibba = Value: boeckella_spinogibba for attribute: species boeckella_sibletti = Value: boeckella_sibletti for attribute: species
sex	Description for: sex	female = Value: female for attribute: sex male = Value: male for attribute: sex
obs_technique	Description for: obs_technique	binoc_microscope = Value: binoc_microscope for attribute: obs_technique
extracted_T_F	Description for: extracted_T_F	TRUE = Value: TRUE for attribute: extracted_T_F
ID_ext_plate	Description for: ID_ext_plate	P8 = Value: P8 for attribute: ID_ext_plate P9 = Value: P9 for attribute: ID_ext_plate
ext_plate_loc	Description for: ext_plate_loc	
n_tested_COI_primers_PCR	Description for: n_tested_COI_primers_PCR	dimensionless
ID_COI_primers_PCR1	Description for: ID_COI_primers_PCR1	Hfolmer_Lfolmer = Value: Hfolmer_Lfolmer for attribute: ID_COI_primers_PCR1
positive_COI_PCR1_T_F	Description for: positive_COI_PCR1_T_F	TRUE = Value: TRUE for attribute: positive_COI_PCR1_T_F

attribute name	definition	format, unit or codes
ID_COI_primers_PCR2	Description for: ID_COI_primers_PCR2	
positive_COI_PCR2_T_F	Description for: positive_COI_PCR2_T_F	
ID_COI_seq1_plate	Description for: ID_COI_seq1_plate	E2019_19 = Value: E2019_19 for attribute: ID_COI_seq1_plate
COI_seq1_plate_loc	Description for: COI_seq1_plate_loc	
ID_COI_primers_seq1	Description for: ID_COI_primers_seq1	Hfolmer_Lfolmer = Value: Hfolmer_Lfolmer for attribute: ID_COI_primers_seq1
positive_COI_seq1_T_F	Description for: positive_COI_seq1_T_F	TRUE = Value: TRUE for attribute: positive_COI_seq1_T_F
COI_haplotype	Description for: COI_haplotype	B_spi_COI03 = Value: B_spi_COI03 for attribute: COI_haplotype B_spi_COI01 = Valu e: B_spi_COI01 for attribute: COI_haplotype B_spi_COI11 = Valu e: B_spi_COI11 for attribute: COI_haplotype B_spi_COI10 = Valu e: B_spi_COI10 for attribute: COI_haplotype B_spi_COI04 = Valu e: B_spi_COI04 for attribute: COI_haplotype B_spi_COI08 = Valu e: B_spi_COI08 for attribute: COI_haplotype B_spi_COI05 = Valu e: B_spi_COI05 for attribute: COI_haplotype B_spi_COI07 = Valu e: B_spi_COI07 for attribute: COI_haplotype B_spi_COI06 = Valu e: B_spi_COI06 for attribute: COI_haplotype B_spi_COI02 = Valu e: B_spi_COI02 for attribute: COI_haplotype B_spi_COI12 = Valu e: B_spi_COI12 for attribute: COI_haplotype B_spi_COI09 = Valu e: B_spi_COI09 for attribute: COI_haplotype B_spi_COI13 = Valu e: B_spi_COI13 for attribute: COI_haplotype B_sib_COI01 = Valu e: B_sib_COI01 for attribute: COI_haplotype
COI_NCBI_accession	Description for: COI_NCBI_accession	

attribute name	definition	format, unit or codes
input_operator	Description for: input_operator	coline_royaux = Value: coline_royaux for attribute: input_operator
lab_operator	Description for: lab_operator	coline_royaux = Value: coline_royaux for attribute: lab_operator

locations_boeck1_2023-07-14.tsv

Physical: locations_boeck1_2023-07-14.tsv

attribute name	definition	format, unit or codes
ID_locality	Description for: ID_locality	
y_latitude	Description for: y_latitude	dimensionless
x_longitude	Description for: x_longitude	dimensionless
sampled_T_F	Description for: sampled_T_F	TRUE = Value: TRUE for attribute: sampled_T_F
posterior_locality_ID	Description for: posterior_locality_ID	
ET_locality_ID	Description for: ET_locality_ID	
ultramafic_T_F	Description for: ultramafic_T_F	TRUE = Value: TRUE for attribute: ultramafic_T_F

samples_boeck1_2023-07-14.tsv

Physical: samples_boeck1_2023-07-14.tsv

attribute name	definition	format, unit or codes
ID_sample	Description for: ID_sample	
ID_locality	Description for: ID_locality	
sampling_date_DD_MM_YY Y	Description for: sampling_date_DD_MM_YY Y	MM_DD_YYYY
sampling_year	Description for: sampling_year	YYYY
sample_type	Description for: sample_type	wild = Value: wild for attribute: sample_type
landowner_at_sampling_date	Description for: landowner_at_sampling_date	YYYY-MM-DD

attribute name	definition	format, unit or codes
living_animals_eggs_found_T_F	Description for: living_animals_eggs_found_T_F	TRUE = Value: TRUE for attribute: living_animals_eggs_found_T_F
sorted_T_F	Description for: sorted_T_F	TRUE = Value: TRUE for attribute: sorted_T_F
sampler	Description for: sampler	nicolas_rabet = Value: nicolas_rabet for attribute: sampler
input_operator	Description for: input_operator	coline_royaux = Value: coline_royaux for attribute: input_operator

Spatial Rasters

Spatial Vectors

Custom units

unit name	parent SI unit	unit type	description
-----------	----------------	-----------	-------------

Other entities

AB1_Boeckella.zip

haplotypes_boeck1_2023-02-01.fasta

Intellectual Rights

This information is released under the Creative Commons license - Attribution - CC BY (<https://creativecommons.org/licenses/by/4.0/>). The consumer of these data ("Data User" herein) is required to cite it appropriately in any publication that results from its use. The Data User should realize that these data may be actively used by others for ongoing research and that coordination may be necessary to prevent duplicate publication. The Data User is urged to contact the authors of these data if any questions about methodology or results occur. Where appropriate, the Data User is encouraged to consider collaboration or co-authorship with the authors. The Data User should realize that misinterpretation of data may occur if used out of context of the original study. While substantial efforts are made to ensure the accuracy of data and associated documentation, complete accuracy of data sets cannot be guaranteed. All data are made available "as is." The Data User should be aware, however, that data are updated periodically and it is the responsibility of the Data User to check for new versions of the data. The data authors and the repository where these data

were obtained shall not be liable for damages resulting from any use or misinterpretation of the data. Thank you.

Development supported by 
created with [emldown](#)

Ouvrage collectif – Histoire des crustacés

Titre

Les copépodes de Nouvelle-Calédonie : de l'endémisme à petite échelle ?

Prévoir un titre simple, informatif ou énigmatique, un titre jeu de mots, un titre question...etc.

Auteurs

Coline Royaux, Frédéric Busson, Nicolas Charpin, Nicolas Rabet

Lister les auteurs (prénom, nom), 5 auteurs maximum

Chapeau

En Nouvelle-Calédonie, les copépodes du genre *Boeckella* sont isolés dans des réseaux de mares limitées aux zones riches en métaux et ont évolué d'une manière originale par rapport au reste du monde.

Texte (page droite)

Le plancton des mares est composé d'une grande variété d'organismes. En eau douce comme en eau de mer, les crustacés et notamment les copépodes, représentent une part importante du plancton. Parmi eux, le genre *Boeckella* est emblématique des mares continentales en Amérique du Sud, en Antarctique, dans les îles subantarctiques, en Australie et en Nouvelle-Calédonie.

Jusqu'à là négligée, l'étude des mares de Nouvelle-Calédonie montre que les espèces de *Boeckella* ont l'originalité d'y être exclusivement localisées dans les zones minières où les taux de métaux lourds sont naturellement très élevés. La diversité génétique de ces micro-crustacés suggère qu'il existe un fort isolement des populations, avec une histoire locale ancienne. La diversité morphologique est faible et tranche avec ce qui est observé ailleurs dans le monde. En Nouvelle-Calédonie, ces organismes sont cantonnés à de petits territoires et sont également isolés aux niveaux génétique et morphologique interprétable comme une forme d'endémisme. Sur ce territoire, leur aire de distribution est réduite à quelques dizaines de km², contre des milliers en Amérique du Sud. Dans ce contexte, nous utilisons le concept de micro-endémisme avec la démonstration que, la Nouvelle-Calédonie se révèle pleine de surprises. Malheureusement, nos curieux *Boeckella* habitent exclusivement les territoires qui intéressent l'homme pour l'exploitation du nickel, générant un conflit entre conservation et exploitation minière.