



HAL
open science

La phylogéographie comparée d'espèces hydrothermales du Pacifique Ouest à l'heure de la génomique des populations

Adrien Tran Lu Y

► **To cite this version:**

Adrien Tran Lu Y. La phylogéographie comparée d'espèces hydrothermales du Pacifique Ouest à l'heure de la génomique des populations. Autre [q-bio.OT]. Université de Montpellier, 2022. Français. NNT : 2022UMONG017 . tel-03848079

HAL Id: tel-03848079

<https://theses.hal.science/tel-03848079>

Submitted on 10 Nov 2022

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.

**THÈSE POUR OBTENIR LE GRADE DE DOCTEUR
DE L'UNIVERSITÉ DE MONTPELLIER**

En génétique et génomique

École doctorale GAIA

Unité de recherche ISEM

**La phylogéographie comparée d'espèces
hydrothermales du Pacifique Ouest à l'heure de la
génomique des populations**

**Présentée par Adrien Tran Lu Y
Le 11 Mai 2022**

**Sous la direction de François Bonhomme
et Stéphane Hourdez**

Devant le jury composé de

MATHIEU GAUTIER, Directeur de recherche, INRAE

PIERRE CHEVALDONNE, Directeur de recherche, IMBE

FRANÇOIS BONHOMME, Directeur de recherche émérite, CNRS

STEPHANE HOURDEZ, Chargé de recherche, CNRS

CYNTHIA RIGINOS, Professeur, University of Queensland

SARAH SAMADI, Professeur, MNHN

Président du jury

Examineur

Directeur de thèse

Co-directeur de thèse

Rapportrice

Rapportrice



**UNIVERSITÉ
DE MONTPELLIER**

Résumé :

Cette thèse a pour objectif principal d'explorer les patrons de différenciation phylogéographique de plusieurs espèces inféodées aux sources hydrothermales profondes des bassins d'arrière-arc du Pacifique Ouest. Ces sources constituent un habitat discontinu car entrecoupé par des arcs volcaniques, des fosses de subduction et des plaines abyssales. Les écosystèmes associés à ces sources sont composés d'une faune abondante d'espèces dépendant entièrement de la chimiosynthèse bactérienne. Dans ce contexte très fragmenté se pose la question du degré de connectivité des peuplements des différentes espèces hydrothermales ainsi que leur histoire phylogéographique. Peu étudiées jusqu'à présent, nous abordons ici ces questions avec les méthodologies de la génomique des populations qui permettent de s'affranchir des principales limites liées à l'utilisation d'un faible nombre de marqueurs et apportent une meilleure résolution des processus démo-génétiques.

Le premier chapitre s'intéresse à un gastéropode emblématique de ces écosystèmes, *Ifremeria nautiliei* (Provannidae) échantillonné sur l'ensemble de son aire de répartition. L'obtention de 10 570 marqueurs génomiques (ddRad) indépendants a permis d'identifier une structure en deux métapopulations séparées par les Iles Salomon et Vanuatu et quasi-homogènes génétiquement de part et d'autre. Les reconstructions démo-génétiques indiquent que ces métapopulations sont actuellement connectées par un flux génique faible, bi-directionnel et légèrement asymétrique et qu'elles auraient divergé il y a environ 70 000 générations. Cependant une analyse détaillée des locus ne se conformant pas à la distribution neutre attendue (« outliers ») montre une limite de connectivité démographique à l'échelle des bassins au sein de chaque métapopulation. Ceci peut avoir des conséquences pour la recolonisation des sites en cas de destruction massive.

Le chapitre 2 compare les patrons phylogéographiques de quatre espèces (*I. nautiliei*, *Alviniconcha kojimai*, *Shinkailepas tollmanni* et *Eochionelasmus ohtai*) avec la même méthodologie que pour le chapitre 1. Comme pour *I. nautiliei*, les trois autres espèces montrent une structure en deux métapopulations avec un faible flux génique bidirectionnel légèrement asymétrique vers l'ouest. Cette bi-directionnalité pourrait suggérer que les stratégies de dispersion larvaire ne sont pas unimodales et donc plus complexes que celles proposées jusqu'à présent. Les reconstructions démo-génétiques montrent que la plupart des espèces auraient divergé sous un modèle de contact secondaire avec une séparation entre métapopulations estimée à 70 000 - 100 000 générations à l'exception du gastéropode *A. kojimai* dont la divergence est très récente. Les temporalités de remise en contact secondaire et de niveau de flux génique varient cependant d'une espèce à l'autre, des différences probablement attribuables à l'interaction entre l'histoire géologique de la région et leurs traits d'histoires de vie qui eux-mêmes conditionnent leur distribution biogéographique.

Mots Clés : Espèces hydrothermales - Génomique des populations - Phylogéographie comparative - ddRAD - Connectivité - Pacifique occidental - Bassin arrière-arc.

Abstract:

The main objective of this thesis is to explore the phylogeographic differentiation patterns of several species that inhabit deep hydrothermal vents in the back-arc basins of the western Pacific. These vents constitute a discontinuous habitat as they are interspersed with volcanic arcs, subduction trenches and abyssal plains. The ecosystems associated with these sources are composed of an abundant fauna of species entirely dependent on bacterial chemosynthesis. In this highly fragmented context, the question arises as to the degree of connectivity of the populations of the different hydrothermal species as well as their phylogeographic history. Little studied until now, we approach these questions here with the methodologies of population genomics which allow us to get rid of the main limitations linked to the use of a small number of markers and bring a better resolution of the demo-genetic processes.

The first chapter focuses on a gastropod emblematic of these ecosystems, *Ifremeria nautiliei* (Provannidae) sampled throughout its range. By obtaining 10,570 independent genomic markers (ddRad), we identified a structure in two metapopulations separated by the Solomon Islands and Vanuatu and almost genetically homogenous on both sides. Demo-genetic reconstructions indicate that these metapopulations are currently connected by a weak, bi-directional and slightly asymmetric gene flow and that they would have diverged about 70,000 generations ago. However, a detailed analysis of loci that do not conform to the expected neutral distribution ("outliers") shows a limit of demographic connectivity at the basin scale within each metapopulation. This may have implications for site recolonization in the event of mass destruction.

Second chapter compares the phylogeographic patterns of four species (*I. nautiliei*, *Alviniconcha kojimai*, *Shinkailepas tollmanni*, and *Eochionelasmus ohtai*) using the same methodology as Chapter one. As with *I. nautiliei*, the other three species show a two-metapopulation structure with weak bi-directional gene flow slightly asymmetrically to the west. This bi-directionality might suggest that larval dispersal strategies are not unimodal and thus more complex than those proposed so far. Demo-genetic reconstructions show that most species would have diverged under a secondary contact pattern with a separation between metapopulations estimated at 70 000 - 100 000 generations with the exception of the gastropod *A. kojimai* whose divergence is very recent. However, the timing of secondary contact and gene flow levels vary from species to species, differences that are probably attributable to the interaction between the geological history of the region and their life history traits, which themselves condition their biogeographic distribution.

Keywords: Hydrothermal species - Population genomics - Comparative phylogeography - ddRAD - Connectivity - Western Pacific - Back-arc basin.

TABLE DES MATIERES

Introduction générale.....	1
I. De la biogéographie à la phylogéographie	2
1. Introduction	2
2. Patrons phylogéographiques en milieu terrestre.....	4
3. Patrons phylogéographiques en milieu marin.....	9
II. De la génomique des populations à la phylogéographie	11
1. Introduction	11
2. Les modèles démo-génétiques et la reconstitution des événements passés de divergence.....	12
Les questions posées	15
III. Les modèles d'étude de ce travail	16
1. Milieux océaniques profonds et monde hydrothermal.....	16
2. Modèles d'étude et questions abordées	22
 Chapitre I : Structure génétique et connectivité chez le gasteropode hydrothermal <i>Ifremeria nautilei</i>	 27
I. Introduction	28
II. Résumé (français)	28
III. Article (Molecular Ecology)	29
IV. Conclusions et perspectives	66
 Chapitre II : Phylogéographie comparée des espèces hydrothermales du Pacifique ouest	 70
I. Introduction	71
II. Matériel et Méthodes	73

III. Résultats	76
1. Structuration des populations	76
Analyses inter-métapopulations	76
Analyses intra-métapopulations.....	78
2. Introgression	79
3. Diversité génétique	81
4. Inférence de l'histoire de divergence	82
Modèles démo-génétiques.....	83
Temporalité de la divergence des métapopulations.....	85
Flux génique.....	86
IV. Discussion	88
1. Comparaison entre espèces par approche génomique (ddRAD)	89
Existe-t-il des déterminants géomorphologiques concordants pour les différentes espèces ?.....	90
(i) Différentiation Est/Ouest.....	90
(ii) Directionnalité des flux géniques	94
Existe-t-il une temporalité de la divergence commune à toutes les espèces ?	100
(iii) Temporalité de la divergence des métapopulations.....	100
2. Comparaison des résultats avec la littérature existante	105
(iv) Structuration et flux génique des génomes nucléaires et mitochondriaux	105
(v) Comparaison avec d'autres études phylogéographiques du monde hydrothermal.....	108
Et la spéciation dans tout ça ?	110
Discussion générale	114
Comparaison avec les modèles phylogéographiques Atlantique/Méditerranée ...	115
Biologie de la conservation, connectivité génétique, démographique et résilience	117
Conclusion générale et perspectives	122

Bilan des résultats	123
Outils pour aborder les questions de résilience et perspectives liées à l'obtention de génomes complets	125
Références	128
Annexes	154
Annexe 1	155
Tables.....	159
Figures.....	170
Annexe 2	193
1. Assemblage de novo pour <i>A. kojimai</i>	193
2. Assemblage <i>de novo</i> pour <i>S. tollmanni</i>	200
3. Assemblage <i>de novo</i> pour <i>E. ohtai</i>	208
4. Analyses	216
Tableaux.....	216
Figures.....	222
Remerciements	232

INTRODUCTION GENERALE

I. DE LA BIOGEOGRAPHIE A LA PHYLOGEOGRAPHIE

1. Introduction

Comprendre la distribution actuelle des organismes vivants sur Terre reste une question majeure en biologie, à l'intersection de plusieurs domaines allant de la paléontologie à la taxonomie en passant par l'écologie, mais aussi la géologie et la climatologie.

C'est lors des grandes découvertes et des premières descriptions de la diversité du vivant par les explorateurs naturalistes sous différentes latitudes que la biogéographie débuta réellement comme science. Dans un premier temps, celle-ci a consisté principalement en la description de la distribution des espèces à travers les différentes zones climatiques au sein des différents continents et ainsi à définir des grandes provinces biogéographiques. Cependant, très vite, ces premières descriptions et études ont débouché sur un ensemble de questions concernant les mécanismes à l'origine des patrons biogéographiques observés : A zone climatique équivalente, pourquoi certaines régions ont-elles une plus grande richesse spécifique que d'autres ? Comment explique-t-on les phénomènes de vicariance ? Pourquoi la distribution des espèces va-t-elle de l'endémisme au cosmopolitisme ? Quel est le rôle de la compétition interspécifique (et inversement de la facilitation) dans les assemblages d'espèces observés ? Quelles sont les profondeurs historiques de ces patrons de distribution et comment se mettent-ils en place au cours du temps ?

C'est principalement à partir du moment où la théorie de l'évolution a été largement acceptée, que le lien entre biogéographie et mécanismes de formation d'espèces nouvelles (que l'on appelle maintenant vicariance ou spéciation) est devenu une thématique centrale de recherche et a conduit à de nombreuses études sur les histoires naturelles des espèces ayant mené à leurs distributions actuelles. C'est justement à partir de ses observations sur la diversité des formes vivantes observées dans différents lieux lors de sa circumnavigation à bord du HMS Beagle, que Charles Darwin élaborait sa théorie sur l'origine des espèces par sélection naturelle. Le cas emblématique des pinsons des Galapagos qu'il a étudié illustre par exemple le lien entre endémisme et spéciation (Brüniche-Olsen et al., 2019; Grant, 1999; Grant & Grant, 2009; Sulloway, 1982).

La théorie de l'évolution a donc été le point de départ des nombreuses études concernant les origines de la biodiversité (même si le terme n'existait pas encore) et de sa répartition sur notre planète. Pour ce faire l'espèce a été choisie comme unité d'observation et de quantification de cette biodiversité, bien que sa définition et sa délimitation reste un défi en biologie (De Queiroz, 2007). Cependant, jusqu'à relativement récemment, la temporalité des mécanismes liés à la formation des espèces n'était pas ou très peu documentée car difficilement accessible. Les choses ont commencé à évoluer dans les années 1960 –70 en couplant les progrès des analyses phylogénétiques, ceux de la biochimie des protéines et de l'ADN, conjointement avec ceux sur l'histoire de la Terre et de ses continents, elle-même permise par la théorisation de la tectonique des plaques, apportant ainsi un cadre géographique à l'évolution des espèces. Dans les années 1980-90, la généralisation des analyses moléculaires et les progrès de la génétique des populations ont permis de mettre en évidence les liens de proximité génétique qui existent au sein des populations d'une espèce mais aussi entre espèces proches. John Avise (2010) a ainsi introduit le terme « phylogéographie » et théorisé avec Templeton (1998) cette discipline qui cherche à établir des possibles corrélations entre les niveaux de structuration génétique d'une espèce, sa distribution géographique et les événements spatio-temporels géologiques ou climatiques. Avec l'avènement de ces études de phylogénie moléculaire, les estimations des temps de divergence ont pu être réalisées de manière de plus en plus fiable par le développement et l'utilisation de modèles d'évolution, la calibration d'horloges moléculaires expérimentalement ou par datation précise d'évènements géologiques majeurs, mais aussi par leurs étroits liens avec différents évènements spatio-temporels. Un des exemples les plus emblématiques de ces études phylogéographiques est celui qui concerne notre espèce, *Homo sapiens*. Les premières études qui lui ont été consacrées ont confirmé à partir de données moléculaires sur l'ADN mitochondrial l'hypothèse de l'origine africaine de l'Homme moderne qui n'était pas fermement établie d'après les fossiles (Cann et al., 1987; Vigilant et al., 1991; Wilson et al., 1985)

Bien que ces études n'aient été dans un premier temps qu'à l'échelle d'une ou de quelques espèces, l'évolution de ce pan de la recherche a rapidement abouti à des études comparatives entre espèces.

Cette comparaison plurispécifique ou phylogéographie comparative intègre ainsi sur une base intra- et inter-spécifique, la biogéographie et la génétique des populations pour étudier les déterminants ou facteurs responsables des différents patrons de distribution des espèces observés et leur congruence autour de certains événements temporels.

2. Patrons phylogéographiques en milieu terrestre

Les grandes questions de biogéographie et de phylogéographie, concernant l'impact et le rôle des variations climatiques, de la tectonique ou autres barrières sur les patrons de distributions des espèces, ont commencé à être étudiés à différentes échelles notamment celle de l'Europe (domaine paléarctique), ou de l'Amérique du Nord (néarctique) principalement sur des espèces terrestres.

Ces études ont mis en lumière que de nombreuses espèces ont subi une succession de réductions et d'expansions de leurs aires de répartition au cours des cycles glaciaires et interglaciaires du Quaternaire. Ainsi, pendant ces périodes glaciaires, un nombre important d'espèces et de populations se sont retrouvées isolées au sein de différents refuges glaciaires, formant en conséquence des populations génétiquement différenciées (Hewitt, 1996, 1999, 2000). Bien que toutes ces espèces aient subi les mêmes cycles climatiques, elles ne partagent pas toutes nécessairement des patrons similaires de distribution. Par exemple, pour le Paléarctique (pour sa partie européenne), trois grands types de patrons ou tendances ont été décrits dans la littérature (Schmitt, 2007).

Le premier décrit est celui des espèces dites « méditerranéennes ». Ces dernières auraient subi un isolement génétique lors de leur cantonnement dans les refuges des péninsules méditerranéennes (ibérique, italienne, grecque) lors des périodes glaciaires. A la suite de ces dernières, les recolonisations furent impactées différenciellement par plusieurs barrières physiques à la dispersion tels que les Pyrénées, les Alpes ou les Balkans, générant ainsi des réponses variables en termes d'aires de distribution selon les traits d'histoire de vie des espèces, décrivant ainsi une multitude de patrons de distribution.

Le deuxième patron identifié est celui des espèces décrites comme étant « continentales/sibériennes ». Les origines des patrons de ces dernières sont assez complexes. D'une part, ils proviendraient pour certaines espèces d'évènements de recolonisation post-

glaciaire similairement aux espèces dites « méditerranéennes » mais provenant de refuges glaciaires paléarctiques « extra-méditerranéens » se trouvant en Asie, n'ayant ainsi pas été affectés par les mêmes barrières. D'autre part, en fonction des espèces, plusieurs autres refuges glaciaires « extra-méditerranéens » ont été mis en évidence en Europe. Cette hétérogénéité due à ces différents refuges glaciaires se traduit ainsi par des histoires complexes de colonisation (Tzedakis et al., 2013).

Enfin, le dernier grand type de patron ou tendance est celui des espèces dites « arctiques/alpines », affichant un certain nombre de patrons variés. Ceux-ci caractérisent des espèces ayant eu une large distribution lors des périodes glaciaires et qui ont subi une rétractation de leur aire de répartition vers les hautes latitudes ou en altitude dans les chaînes de montagnes européennes pendant les périodes interglaciaires. Ces patrons peuvent aussi inclure des espèces ayant eu des distributions disjointes durant les périodes glaciaires, probablement en lien avec leurs niches écologiques, les ayant ainsi restreintes aux habitats humides proches des montagnes par exemple. Bien que souvent perçu comme un patron simple et facilement identifiable, celui-ci affiche en réalité une grande diversité et complexité.

S'il est évident que ces études de phylogéographie comparée ont mis en lumière ces grandes tendances pour expliquer les patrons biogéographiques, la multitude de différences notables dans ces dernières ont montré que les réponses spécifiques peuvent varier bien que les différentes espèces aient subi de manière simultanée les événements climatiques. Cette variabilité serait d'une part probablement due aux traits d'histoire de vie intrinsèques aux espèces, tels que les capacités de dispersion et sensibilité aux barrières physiques (chaînes de montagne...), mais elle pourrait aussi être liée aux localisations des refuges glaciaires ou interglaciaires (Hewitt, 2011). Ces différents facteurs ont ainsi engendré des histoires complexes de colonisation post-glaciaire, générant en conséquence les patrons actuels des espèces (Figure 1) (Hewitt, 2011).

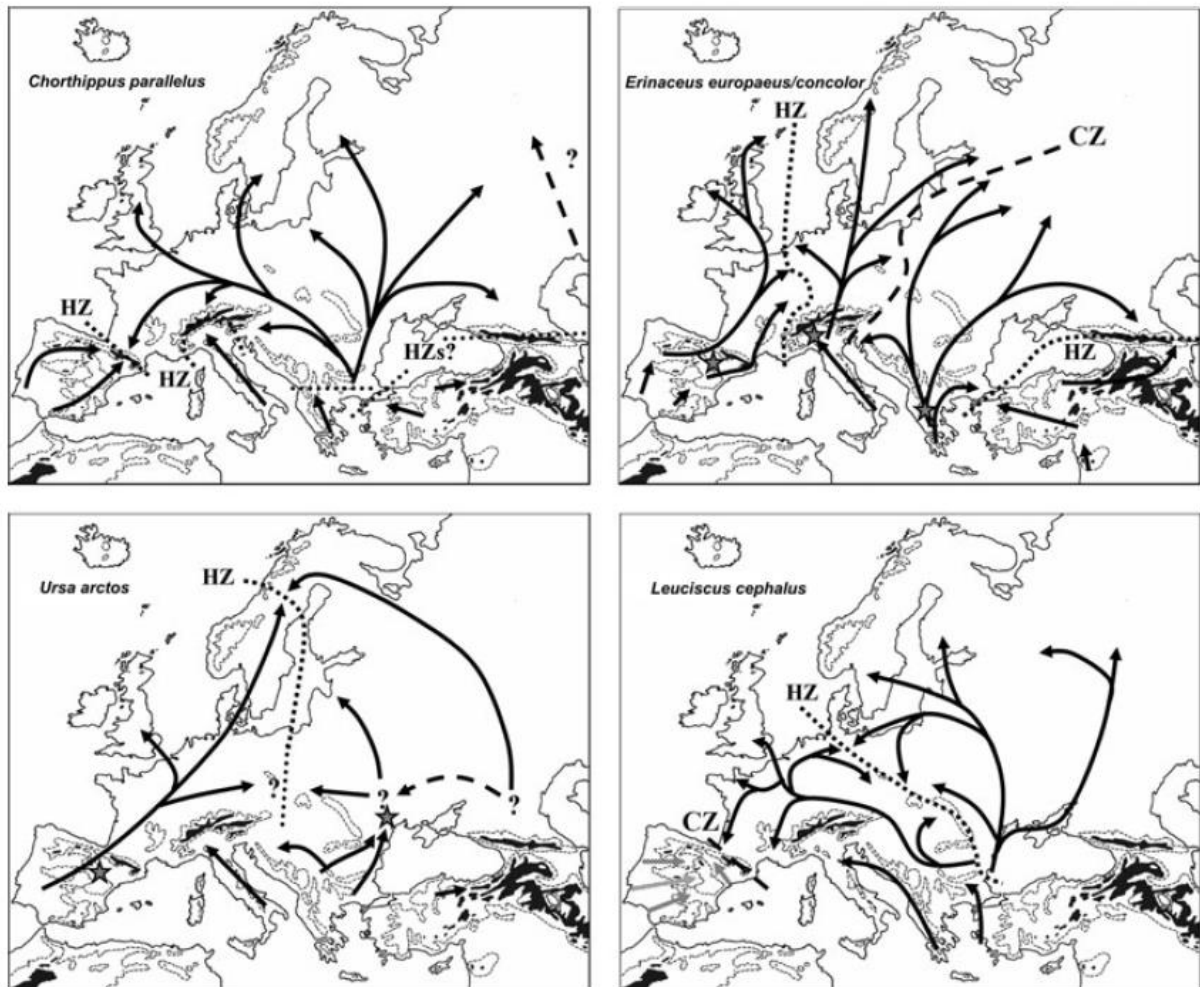


Figure 1 : Possibles routes de colonisation postglaciaire en Europe, depuis des refuges supposés et estimés a partir des données génétiques, climatiques et fossiles pour quatres espèces. Le criquet des pâtures (*Chorthippus parallelus*), le hérisson européen (*Erinaceus europaeus/concolor*), L'ours brun (*Ursus arctos*) et la Chevesne (*Leuciscus cephalus*). HZ : Hybride zone. Les flèches indiquent les possibles routes de colonisation et les traits en pointillé les zones de contacts entre populations. Tiré d'Hewitt 2011, Figure 2

Si ces études ont montré que les cycles glaciaires ont grandement influencé les patrons biogéographiques des régions tempérées terrestres, ces influences ne sont pas a priori aussi évidentes pour les régions inter-tropicales où l'effet des glaciations n'est pas aussi marqué. Cependant, il a été montré que les patrons phylogéographiques de ces régions ont subi l'effet d'une part des variations drastiques dans la saisonnalité des précipitations et du niveau des océans (Colinvaux et al., 1996; Gasse, 2000; O. J. Wilson et al., 2021), mais aussi d'autre part l'effet de barrières des chaînes de montagnes, des systèmes hydrographiques (rivières) ou

encore d'autres facteurs biologiques et environnementaux (Bonatelli et al., 2022; Pellegrino et al., 2005; Thomé et al., 2014)

Plusieurs études s'intéressant à comprendre l'histoire et l'origine de la diversification de différents amphibiens en région tropicale par exemple, ont montré que les déterminants à l'origine des patrons de distribution sont très variés, même au sein d'un genre et d'une région biogéographique (Arteaga et al., 2016; Crawford et al., 2007; d'Horta et al., 2013; Fouquet et al., 2015; Vacher et al., 2020; Zeisset & Beebee, 2008). Ainsi, certains facteurs influençant ces patrons ont une origine paléogéographique par la mise en place une barrière géologique ou climatique (thermique) pour certaines espèces, alors que d'autres (hydrologie) ont agi en créant des refuges lors des périodes de sécheresse, avec des cycles climatiques et oscillations des pluies associés, générant une fragmentation de l'habitat comme des îlots forestiers.

Les variations climatiques allant de changement drastique tel que les glaciations ou des variations dans les saisonnalités ont eu une contribution non négligeable sur l'origine des patrons biogéographiques des espèces terrestres. Des événements tectoniques majeurs ont aussi eu une contribution importante dans la formation de ces dernières, de par leur effet de barrière à la dispersion principalement mais aussi dans certains cas facilitant le contact des populations ou espèces.

Un des exemples emblématiques est celui de la fermeture de l'isthme de Panama, qui a résulté de la collision des plaques continentales de l'Amérique du Nord et du Sud, fermant ainsi la jonction entre les océans Pacifique et Atlantique, il y a environ 3 – 3,5 millions d'années (Keigwin, 1978; O'Dea et al., 2016). Une des conséquences principales et directes de cette collision entre les deux continents est la formation d'un pont terrestre, permettant ainsi aux faunes respectives de s'échanger et de coloniser de nouveaux habitats, tout en stoppant les échanges entre faunes marines des deux océans (Knowlton & Weigt, 1998). De nombreuses études se sont ainsi intéressées à comprendre les conséquences de ce contact sur les patrons biogéographiques terrestres (Cody et al., 2010; Pinto-Sánchez et al., 2012). Elles ont montré que cela a permis aux faunes respectives de se disperser entre les deux continents formant ainsi le « Grand échange biotique américain ». Néanmoins, il a aussi été démontré que l'échange des faunes s'est fait de manière progressive avec une arrivée d'espèces plus dispersantes avant la fermeture de l'isthme de Panama, montrant ainsi que les échanges

faunistiques et floristiques, liés au rapprochement des deux continents (et les nombreuses extinctions qui en ont résulté), ne se seraient donc pas réalisés de manière simultanée mais à différentes périodes précédant la fermeture complète de l'isthme (Bacon et al., 2015; Carranza, 2004).

Un autre cas emblématique est celui de la séparation biogéographique, souvent mis en avant est ce qui correspond à ce qui est appelé de nos jours la « Ligne Wallace ». Lors de ses explorations dans l'Indopacifique, l'explorateur naturaliste Alfred Wallace a identifié et décrit une zone de démarcation nette dans les patrons de composition d'espèces en Asie du Sud-Est (Wallace, 1860). Pour expliquer cette séparation, il a été proposé par la suite que lors de la baisse du niveau des océans au cours des glaciations, les régions de part et d'autre de cette ligne correspondaient à des zones terrestres en continuité avec leur continent respectif, mais séparées par une zone d'eaux profondes entre les deux plaques océano-continetales eurasiennne et australienne. Cette zone aurait dès lors constitué une barrière géographique et physique limitant la dispersion d'espèces terrestres. Bien que cette première hypothèse de vicariance ait été posée pour cette région, plusieurs études ont montré par la suite que les patrons observés ne correspondaient pas nécessairement à ce cas de figure, mais affichaient plutôt des distributions biogéographiques plus complexes et qui sont possiblement nuancé par les traits d'histoire de vie des espèces (Esselstyn et al., 2010; Jones & Kennedy, 2008; Mitchell et al., 2014; Tänzler et al., 2014).

Ces différents exemples de phylogéographie comparée en milieu terrestre ont permis de mettre en lumière les contributions respectives de plusieurs facteurs extrinsèques à la formation et aux fluctuations de ces distributions d'espèces terrestres, tels que les cycles climatiques, la tectonique, les diverses barrières physiques et géographiques, ayant ainsi conduit à des rétractions et expansions de ces aires de distribution, qu'elles soient tropicales ou tempérées. Bien que ces facteurs extrinsèques aient exercé une influence importante dans l'origine des répartitions actuelles des espèces dans une région donnée, les facteurs intrinsèques aux espèces (traits d'histoires de vie) ont aussi montré leur influence non négligeable et qui tempère celle des facteurs extrinsèques, notamment à travers les capacités de dispersion mais aussi par d'autres facteurs biologiques. En résumé, ces études ont ainsi mis en avant la complexité des origines de ces distributions, pouvant afficher une forte variabilité, même au sein d'une région biogéographique.

3. Patrons phylogéographiques en milieu marin

Comme mentionné précédemment, la fermeture de l'isthme de Panama a grandement influencé les faunes terrestres américaines, mais celle-ci a aussi eu des conséquences drastiques et majeures sur la circulation océanique atlantique, le climat et les espèces marines situées de part et d'autre de l'Isthme (Haug & Tiedemann, 1998). La première conséquence de la fermeture progressive de cet isthme a été un effet de vicariance stricte des populations marines de part et d'autre de la barrière formée. D'autre part, la fermeture a aussi eu pour conséquence la formation du fameux courant « Gulf Stream » (ou dérive nord-atlantique). Plusieurs études se sont penchées sur la datation de la divergence en lien avec la fermeture de l'isthme, supposant que celle-ci soit commune pour toutes les espèces marines de part et d'autre. Elles ont montré que pour certaines espèces, le début de la divergence concorde avec la mise en place de l'isthme, mais que pour d'autres, la divergence serait plus ancienne et aurait été provoquée par des changements environnementaux tels que la salinité (Knowlton & Weigt, 1998).

Un autre exemple notable en milieu marin est l'ouverture du détroit de Béring, il y a environ 5 millions d'années (Gladenkov et al., 2002; Marinovich & Gladenkov, 1999). Cette ouverture entre l'océan Pacifique et l'Atlantique a eu des conséquences similaires au « grand échange biotique » de la faune terrestre américaine lors que la fermeture de l'isthme de Panama, permettant ainsi aux faunes marines restreintes de chaque océan de se disperser, générant ainsi « l'échange trans-arctique » pour les espèces des bassins du Nord Pacifique et de l'Atlantique (Laakkonen et al., 2013, 2021; Luttikhuizen et al., 2018). Cependant, il a été montré que non seulement l'ouverture physique de ce détroit a eu des conséquences sur les patrons de distribution des espèces, mais cet effet a été combiné à celui des cycles glaciaires, générant ainsi des échanges interocéaniques répétés pour certains taxons (Laakkonen et al., 2021).

Bien que la séparation des océans par les continents soit évidente comme dans le cas précédemment cité de l'isthme de Panama ou du détroit de Béring, les interconnexions des océans rendent difficiles la compréhension des déterminants qui sont à l'origine des patrons de biodiversité actuels pour les espèces marines. Cela tient d'une part à la quasi-impossibilité d'identifier des barrières physiques nettes, telles qu'un front océanique ou des courants

divergents, mais aussi d'autre part au fait que la dispersion d'une grande majorité d'espèces marines, qu'elles soient mobiles ou fixées, est impactée de manière plus ou moins importante par les courants selon leurs traits d'histoire de vie (période de reproduction, durée de vie larvaire, etc...). Malgré cela, un certain nombre d'études se sont attelées à identifier les contributions relatives de ces déterminants sur les patrons de distribution des espèces et de leurs populations.

Kelly & Palumbi (2010) ont comparé la variabilité génétique de 50 espèces d'invertébrés marins côtiers du Pacifique Nord. Ils ont mis en évidence que les espèces n'arborant pas de phase larvaire pélagique affichent les plus forts degrés de différenciation génétique. A contrario, dans cette étude, les espèces ayant un développement pélagique présentent une forte disparité d'un point de vue de la différenciation génétique, allant d'espèces dont les populations sont très fortement différenciées à une quasi-homogénéité génétique. Malgré cette forte hétérogénéité dans les patrons de différenciation, deux grandes tendances ont été observées. La première correspond aux espèces qui affichent une structuration génétique semblable, due à une zone de transition commune, dont l'origine de cette dernière serait potentiellement due aux courants. Le deuxième patron est celui des espèces qui présentent une différenciation locale le long de la zone de balancement des marées (écotypes) et dont l'explication pourrait être due aux traits d'histoire de vie des espèces et à l'adaptation (Johannesson, 2003; Johannesson et al., 2010).

D'autres études se sont intéressées à comprendre les patrons de phylogéographie d'espèces marines dans d'autres régions du monde. Un des cas relativement bien étudiés est celui des espèces atlantico-méditerranéennes. Patarnello et al. (2007) ont publié une synthèse des données de la littérature sur plusieurs espèces atlantico-méditerranéennes qui a montré que la zone du détroit de Gibraltar représentait une fracture phylogéographique majeure pour beaucoup d'espèces. De façon analogue au cas précédemment décrit sur le Nord-Pacifique, les différentes espèces montrent une forte diversité du point de vue de la différenciation génétique entre populations de part et d'autre de la zone de transition entre la mer d'Alboran et le détroit de Gibraltar. Cette dernière serait ainsi une zone de moindre échange ayant fonctionné comme barrière à chaque baisse du niveau de la mer durant les dernières périodes glaciaires, formant ainsi par allopatrie une lignée atlantique et une lignée méditerranéenne génétiquement distinctes et remises en contact après chaque période glaciaire. Plus

récemment une étude, s'intéressant à l'hétérogénéité de cette différenciation entre les lignées atlantique et méditerranéenne, en lien avec les traits d'histoire de vie chez des poissons téléostéens a montré que ces traits ont une influence importante sur la diversité génétique des espèces lorsqu'ils sont corrélés avec la dispersion (Barry et al., 2022). Ces travaux permettent de nuancer le rôle des différents processus extrinsèques à la biologie des espèces comme l'histoire du climat et la barrière physique à la dispersion qu'a constitué la baisse du niveau des océans associée aux fluctuations climatiques.

Bien que moins évidents que pour la faune terrestre, ces variations climatiques et du niveau de la mer ont donc aussi induit la formation de refuges glaciaires en milieu marin (Jolly et al., 2006; Maggs et al., 2008).

Malgré les difficultés à observer les barrières ou processus hydrologiques qui ont influencé la distribution des espèces marines, les études utilisant des approches de phylogéographie comparative sont nécessaires pour identifier et comprendre les processus qui ont forgé ces patrons de distribution, ainsi que l'effet des traits d'histoire de vie des espèces qui peuvent atténuer ou exacerber ces processus, notamment quand ils affectent la dispersion et la taille des populations.

II. DE LA GENOMIQUE DES POPULATIONS A LA PHYLOGEOGRAPHIE

1. Introduction

La phylogéographie, discipline introduite par John Avise (2000) & Templeton (1998), vise à comprendre les liens entre les niveaux de structuration génétique d'une espèce, sa distribution géographique et les événements géologiques ou climatiques, en replaçant les différences génétiques des populations d'une espèce dans un cadre historique et géographique. Les premières études à grande échelle se sont largement basées sur la divergence de l'ADN mitochondrial (mtDNA), qui reste encore très employé dans le domaine et a été à l'origine de sa notoriété (Avise, 2009). L'ADN mitochondrial, bien que très souvent assez informatif et résolutif concernant l'évolution des lignées maternelles ou la délimitation d'espèces, est néanmoins sujet à différents problèmes potentiels. Il existe en effet assez souvent des discordances entre les arbres phylogénétiques obtenus à partir de l'ADN

mitochondrial et de l'ADN nucléaire. En revisitant ces discordances mito-nucléaires, Toews & Brelsford (2012) ont pu mettre en lumière que dans la majorité des cas, elles seraient dues à des mécanismes tels que des introgressions adaptatives (capture mitochondriale), des disparités démographiques (sex-ratio déséquilibré) ou à des biais asymétriques liés au sexe. Quoiqu'il en soit, le génome mitochondrial ségrége comme un locus unique lié au sexe sans recombinaison, et ne peut donc pas à lui seul garder la trace de toute l'histoire démographique d'une espèce donnée.

L'ADN qui compose le génome nucléaire est beaucoup plus riche de ce point de vue-là car c'est l'évolution de dizaine de milliers de gènes indépendants qu'il enregistre. De nos jours, les marqueurs nucléaires les plus couramment utilisés sont les mutations nucléotidiques ponctuelles ou SNPs (*Single Nucleotide Polymorphism*), qui correspondent simplement au remplacement d'un nucléotide par un autre dans une séquence d'ADN. L'avènement des nouvelles méthodes de séquençage a rendu possible l'accès à des milliers de marqueurs génomiques SNPs, associés avec le développement de nouveaux outils d'analyses bio-informatiques. Cependant, ce n'est pas le seul aspect technique qui prime, c'est aussi le cadre théorique et conceptuel qui a fait de véritables bonds en avant, avec des méthodes de calcul basées sur la théorie de la coalescence ou les équations de diffusion dans un cadre interprétatif dépendant des modèles évolutionnistes qui permettent de tester des hypothèses de plus en plus réalistes et complexes, permettant notamment d'expliquer tel ou tel des patrons de distribution observés (Edwards et al., 2015, 2022). Ce sont ainsi de nouveaux champs disciplinaires des sciences de l'évolution qui ont vu le jour, notamment génomique de la spéciation et la génomique du paysage.

2. Les modèles démo-génétiques et la reconstitution des événements passés de divergence

L'accès à un très grand nombre de marqueurs pour un nombre important d'individus permet désormais d'analyser les informations contenues au sein du polymorphisme génomique. Le paragraphe qui suit ne concerne que les données qui peuvent être considérées comme de première génération, c'est à dire non-assemblées sous forme de génomes complets, ce qui est encore le cas pour de nombreuses espèces non-modèle. Pour rendre compte des patrons de diversification spatiale observés - ce qui est le cœur même de la phylogéographie - il est

possible désormais de simuler des modèles démo-génétiques variés et complexes, sous différentes contraintes (migration ralentie à cause d'incompatibilités génomiques, effet de la sélection en liaison, recombinaison variable le long du génome, modèle de mutation, etc...). Ceci permet de reconstituer des scénarios de divergence complexes tels qu'estimer des temps de divergence stricte, de contacts secondaires éventuels, d'introgessions passées, ou bien encore d'estimer la directionnalité et l'intensité du flux génique entre populations.

Ainsi, si l'on se place dans le cadre le plus simple, c'est-à-dire celui de l'histoire évolutive de deux lignées sœurs avec des locus tous neutres et indépendants, plusieurs modèles peuvent être introduits permettant de distinguer des scénarios de divergence avec ou sans flux génique. Nielsen & Wakeley (2001) ont ainsi introduit une approche par vraisemblance pour distinguer le cas de divergence stricte « *Strict Isolation* » (SI) avec celui de divergence avec migration « *Isolation with Migration* » (IM), permettant ainsi de différencier clairement l'effet du tri incomplet de polymorphisme ancestral de celui de la migration. Néanmoins, ce dernier modèle ne permet pas de déterminer si les flux géniques résiduels sont constants au cours du temps ou bien résultent d'une ancienne migration ("*Ancient Migration*" AM) ou d'un contact secondaire ("*Secondary Contact*" SC).

Cependant ces modélisations « simples » qui capturent un certain nombre d'effets démographiques (taille des populations, migration, temps de divergence), ne prennent pas en compte différents processus évolutifs affectant le génome qui peuvent mener à des interprétations démographiques biaisées (Ewing & Jensen, 2016; Schrider et al., 2016). Il est connu que l'effet des différentes formes de sélection est en général très variable selon le taux de recombinaison et varie en fonction des parties du génome considérées, générant ainsi une hétérogénéité de polymorphisme au sein du génome. (Barton & Bengtsson, 1986; Harrison & Larson, 2016; Roze & Barton, 2006; Wolf & Ellegren, 2017). En effet, la sélection de fond (*background selection*), et les balayages sélectifs (*selective sweeps*) génèrent une réduction locale de diversité nucléotidique d'autant plus forte que le taux de recombinaison est faible (Burri, 2017; Charlesworth et al., 1997; Hill & Robertson, 1966). Par ailleurs, du fait d'adaptations locales différentielles ou d'incompatibilités génomiques, certains locus peuvent être amenés à être contre-sélectionnés quand il y a du flux génique entre les deux populations. Cela se traduit par une réduction locale de migration effective (m_e), alors que l'effet de la

sélection en liaison sur les zones à faible taux de recombinaison peut être modélisé par une réduction locale de la taille efficace des populations (N_e).

Ces facteurs d'hétérogénéité génomique peuvent désormais être pris en compte dans les reconstructions démo-génétiques que ce soit par simulation dans un cadre d'approximation bayésienne (*Approximate Bayesian Computation ABC*) (Beaumont et al., 2002) ou par des approches de vraisemblance avec des équations de diffusion (Gutenkunst et al., 2009; Rougeux et al., 2017). Ces différentes approches ont montré l'importance de capturer cette hétérogénéité qu'elle soit générée par des locus à migration effective réduite encore dits « barrières » (Moan et al., 2016; Tine et al., 2014) ou/et à l'effet de la sélection en liaison (Rougemont et al., 2017; Sousa et al., 2013). Ces évolutions théoriques, techniques et analytiques, ont permis l'essor de la génomique des populations et de tester des scénarios démo-génétiques complexes tout en étant de plus en plus réalistes pour refléter au mieux l'histoire évolutive des espèces (acquisition de l'isolement reproductif) et les différents mécanismes évolutifs par lesquels elle se traduit au sein des génomes, ainsi que leurs contributions à l'origine des patrons de distribution observés. Néanmoins, bien que permettant de reconstruire l'histoire démo-génétique, la majorité de ces méthodes n'utilisent que l'information génomique ponctuelles (SNPs) et les statistiques associées à l'exclusion de celle concernant de la liaison génétique entre région et architecture génomique, tels que celle des haplotypes et cartes de recombinaison, qui peuvent être utilisées pour inférer l'histoire évolutive (Browning et al., 2018; Browning & Browning, 2011; Ni et al., 2019).

Plusieurs exemples de ces utilisations de marqueurs génomiques ont permis de retracer ou de reconstruire finement l'histoire évolutive des populations. Tine et al. (2014) et Duranton et al. (2018) s'intéressant à l'étude de la divergence des deux lignées génétiques du bar européen (atlantique et méditerranéenne) ont mis en lumière que ces dernières ont commencé à diverger, il y a environ 270 000 ans et affiche un contact secondaire récent, à partir du début de l'Holocène (environ 12 000 ans), tout en mettant en évidence l'importance de l'hétérogénéité des flux géniques au sein du génome.

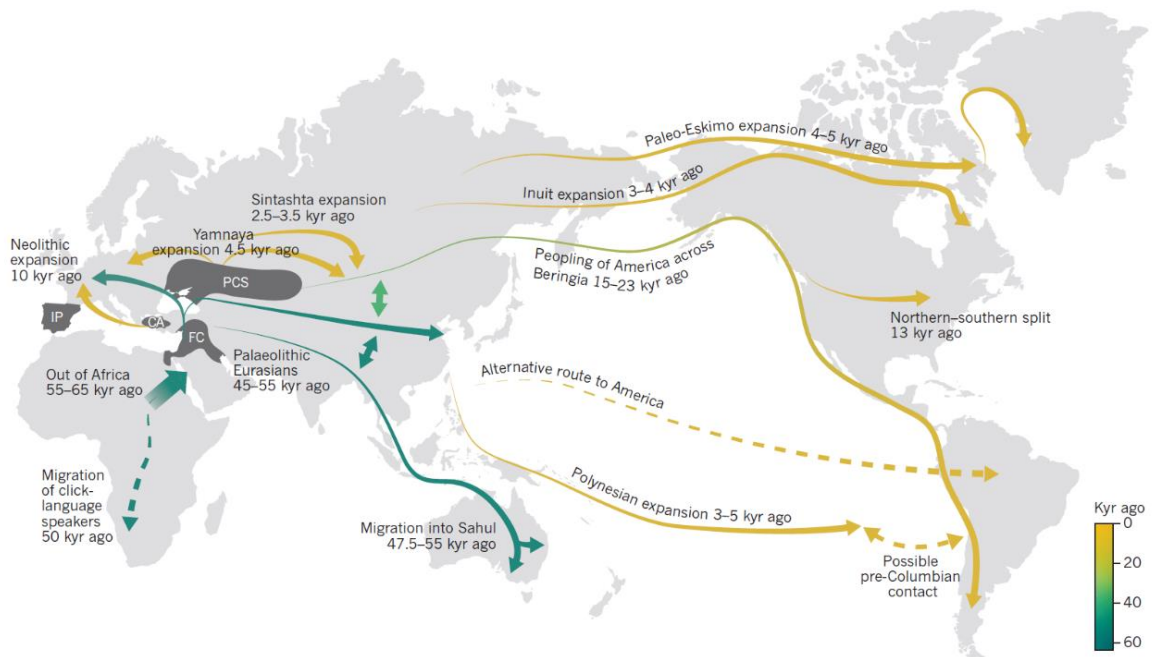


Figure 2 : Routes majeurs de migration de l'Homme moderne inféré à partir de données génomiques. en pointillé les routes encore débattues. CA, Central Anatolia; FC, Fertile Crescent; IP, Iberian Peninsula; PCS, Pontic–Caspian steppe. De (Nielsen et al., 2017), Figure 3. Kyr : (Kilo) Thousand year ago

Un autre exemple assez emblématique est celui de la reconstruction de l'histoire évolutive de l'Homme moderne « *Homo sapiens* ». En effet, comme cité précédemment, plusieurs études ont montré que l'origine géographique de *Homo sapiens* se situerait en Afrique. Cependant la génomique a permis de mettre en évidence que le génome de *sapiens* avait été introgressé par différents hominidés (hominidés africains archaïques, Néandertaliens et Dénisoviens) lors de sa dispersion en dehors de l'Afrique (Figure 2) (Gokcumen, 2020; Mondal et al., 2019; Nielsen et al., 2017; Teixeira & Cooper, 2019).

Les questions posées

A travers le lien entre génomique des populations et phylogéographie, la biodiversité peut être ainsi étudiée à différentes échelles spatiales et/ou temporelles, allant des biomes, aux séquences d'ADN et à leurs mutations, tout en passant de l'échelle du cycle circadien à celle des temps géologiques. Ainsi, l'intersection des méthodes actuelles de la génomique des populations et de la phylogéographie permettent de commencer à élucider certaines questions majeures en biologie de l'évolution telles que : Quel est le rôle de la géographie dans les mécanismes de divergence et d'acquisition de l'isolement reproductif ? Ces

mécanismes d'isolement résultent-ils principalement de barrières au flux génique liées à des événements climatiques ou géologiques ou/et bien dépendent-ils de phénomènes d'adaptation locale ? Quelle est leur sensibilité aux traits d'histoire de vie des espèces ? Avec quelle temporalité et par quels mécanismes précis les incompatibilités génétiques à la base de l'isolement reproductif se mettent-elles en place et sont-elles associées à des effets sélectifs ? Répondre à ces questions permet d'avoir ainsi accès aux mécanismes génomique des phénomènes de divergence en plus des facteurs extrinsèques aux espèces, ayant conduit *in fine* à la spéciation.

Ces questions sont d'un intérêt fondamental dans la théorie de l'évolution pour la compréhension des mécanismes de spéciation. Mais leur portée dépasse largement ce cadre théorique car l'impact de l'Homme sur la biosphère est tel qu'il engendre une accélération du taux d'extinction des espèces à l'échelle planétaire (Barnosky et al., 2011). La question des mécanismes susceptibles de créer de la biodiversité par la formation de nouvelles espèces et de la temporalité à laquelle ces processus se font est donc particulièrement légitime et d'actualité. Ainsi reconstituer les histoires évolutives des espèces à partir de données génomiques et voir leurs liens possibles avec les différents facteurs qui auraient pu les influencer, devient un point clé pour la compréhension des origines de la biodiversité à différentes échelles.

III. LES MODELES D'ETUDE DE CE TRAVAIL

1. Milieux océaniques profonds et monde hydrothermal

L'environnement marin profond occupe 95% des océans et bien qu'il constitue le biome le plus vaste de la planète, il reste le moins étudié. Au sein de celui-ci, les écosystèmes hydrothermaux forment un ensemble à part. Bien que découverts très récemment (fin des années 1970), ce type d'écosystème est présent dans tous les océans, sous toutes les latitudes, majoritairement au niveau des zones d'accrétion tectoniques expansives comme les dorsales océaniques mais également au niveau de points chauds, d'arcs volcaniques et de bassins arrière-arc (Figure 3). Quels que soient ces différents ensembles géologiques, les

milieux hydrothermaux associés forment toujours des habitats fragmentés présentant des fortes dynamiques temporelles calquées sur l’histoire géotectonique des régions concernées (Matabos & Jollivet, 2019; Tunnicliffe & Bone, 1988; Van Dover, 1995; Van Dover et al., 2001, 2002).

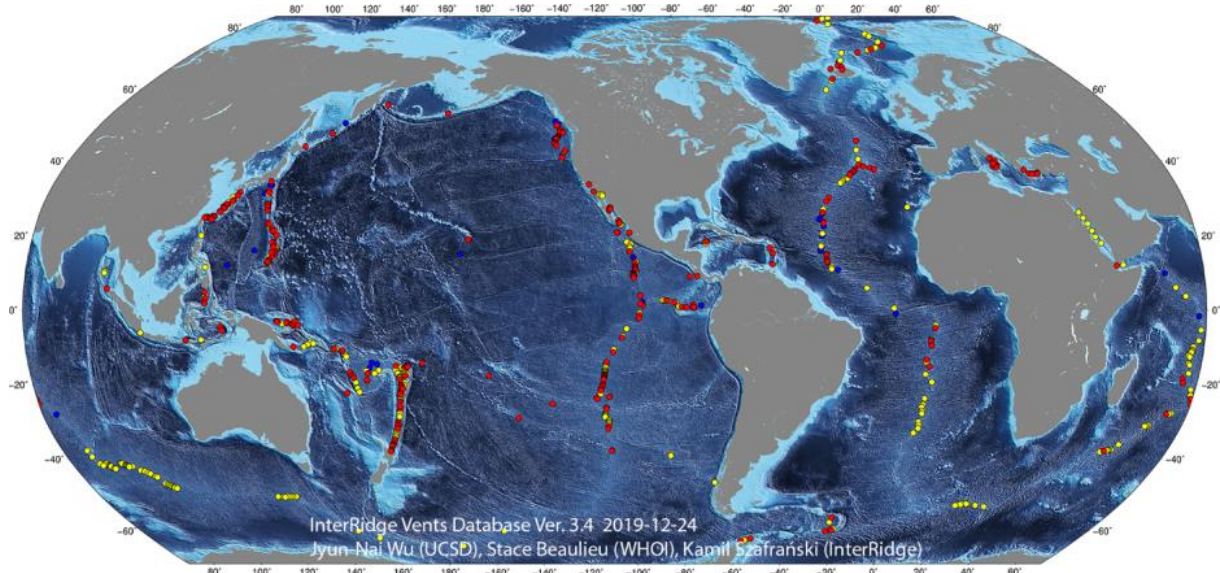


Figure 3 : Carte des sources hydrothermales autour du monde. Les points rouges représentent les sources connues, les jaunes celles inferées et bleus supposées) (issue de la base de donnée InterRidge Vents Database Ver3.4 2019-12-24)

Ces écosystèmes profonds arborent une riche faune « endémique » spécialisé, qui est entièrement dépendante de la production microbienne locale par chimiosynthèse, ne dépendant ainsi pas de producteurs primaires photosynthétiques, au contraire de la majorité des autres écosystèmes, même profonds. Dans ce type d'environnement où l’impact des cycles climatiques ou des évènements géologiques n’est en aucun cas évident, la question des processus qui ont façonné les patrons de distribution des espèces hydrothermales reste encore largement à explorer. Tunnicliffe (1992) a suggéré que ces faunes hydrothermales pourraient provenir d’une origine commune puis auraient divergé pour éventuellement former différentes provinces biogéographiques.

Suite à cette première étude, les suivantes se sont attelées à définir et à décrire les grandes régions biogéographiques pour l’environnement hydrothermal profond. Moalic et al. (2012) ont montré par une approche de réseau que les sources hydrothermales, se découpaient en grandes provinces biogéographiques relativement bien soutenues. Outre ces inférences, la

topologie de leur réseau a de plus permis de suggérer des voies d'échanges préférentiels entre provinces. Ils ont ainsi proposé que la région biogéographique du Pacifique Occidental ait joué un rôle important pour les communautés hydrothermales à différentes échelles, de par sa position centrale au sein du réseau (Desbruyères et al., 2006; Moalic et al., 2012; Tunnicliffe et al., 1998; Van Dover et al., 2002). Cette région arborant un fort taux d'endémisme d'espèce hydrothermales, pourrait avoir joué un rôle majeur dans l'origine de la faune hydrothermale actuelle, ce que suggéraient déjà Little & Vrijenhoek (2003) dans leur étude combinant les données paléontologiques et génétiques. Les datations d'origine de ces faunes pointent généralement vers des âges parfois divergents (fossiles vs génétique) mais l'origine des lignées hydrothermales se situerait à environ 100 Ma (Millions d'années), bien que les crevettes et les moules bathymodilinae soient plus récentes, datant du Miocène (Little & Vrijenhoek, 2003; Vrijenhoek, 2010).

Par leur nature souvent éphémère et instable, les sources hydrothermales ont conduit à une perpétuelle dynamique de recolonisation de l'habitat. Ainsi, l'arrêt du fluide hydrothermal et donc de la source se traduit par la mort (ou la survie provisoire avec arrêt de la reproduction) d'une grande partie des individus qui lui sont associés. Cette dynamique, avec des taux d'extinction élevés sur les dorsales rapides, conduit dans certains cas à des stratégies dispersives pouvant favoriser soit la dispersion de proche en proche, soit une dispersion 'longue durée' et donc une stratégie dispersive à longue distance (Vrijenhoek, 1997). Ainsi lorsque l'activité hydrothermale se déplace le long des dorsales océaniques cela donne lieu de manière épisodique soit à des contacts secondaires reliant ainsi des populations précédemment isolées, soit à des phases d'extinction et de recolonisation qui affectent souvent la quasi-totalité d'un champ hydrothermal (Tunnicliffe et al., 1998).

Des études moléculaires se sont ainsi intéressées à comprendre la distribution des espèces hydrothermales à l'échelle de ces grandes provinces biogéographiques. Une grande partie de ces études se sont principalement intéressées aux distributions des espèces hydrothermales associées aux ensembles géologiques quasi-continus, tels que les dorsales océaniques rapides (Dorsale Est-Pacifique) ou lentes (Dorsale Médio-Atlantique). Ces études ont montré que la plupart des espèces hydrothermales sont généralement capables de maintenir une connectivité sur des milliers de kilomètres (Craddock et al., 1995; Hurtado et al., 2004; Jollivet

et al., 1995; S. Teixeira et al., 2011, 2012; Won et al., 2003) avec une forte capacité de colonisation des sites nouvellement formés (Mullineaux et al., 2010), en suivant parfois un modèle dit par « pas japonais » (*stepping-stone*) (Audzijonyte & Vrijenhoek, 2010; Breusing et al., 2016; Vrijenhoek, 2010). Cependant, ces études et d'autres ont mis en lumière qu'il existait aussi des facteurs extrinsèques qui influencent les patrons de distribution des espèces et populations (Breusing et al., 2016; Johnson et al., 2006; Plouviez et al., 2009). D'autre part, les contributions de ces facteurs ne sont pas équivalentes pour toutes les espèces. Ces différents facteurs extrinsèques peuvent être assez variés, allant d'événements géologiques agissant comme une barrière à la dispersion tels que la formation de microplaques ou de failles transformantes, mais aussi à des limites liées aux traits d'histoire de vie (dispersion) ou écologique, tels que les courants, la profondeur (pression) ou encore les compositions des fluides.

A l'opposé de ce que l'on retrouve dans les systèmes de dorsales océaniques, le Pacifique Ouest est formé d'un ensemble complexe et discontinu de bassins arrière-arc et d'arcs volcaniques (Figure 4). Les bassins arrière-arc sont la résultante du mécanisme de subduction. Lors de cette subduction, la plaque sus-jacente est étirée et s'amincit, provoquant une remontée et fonte partielle du manteau, occasionnant ainsi une activité volcanique par la formation d'une zone de rift analogue à une dorsale océanique et menant à une activité hydrothermale par le biais de la fracturation du plancher océanique. L'histoire géologique des bassins arrière-arc du Pacifique Ouest est assez récente à l'échelle des temps géologiques (1 à 7 millions d'années). La faune hydrothermale associée est endémique à cette région, et se retrouve partagée entre ces bassins, allant du bassin des Mariannes (pour certaines espèces) à la Nouvelle Zélande (arc de Kermadec) en passant par le bassin de Lau (Iles Tonga), le bassin de Manus (Papouasie-Nouvelle Guinée, PNG), le bassin de Woodlark (PNG et Iles Salomon), le bassin de Nord-Fidji (Fidji) et l'arc volcanique de Futuna (Futuna).

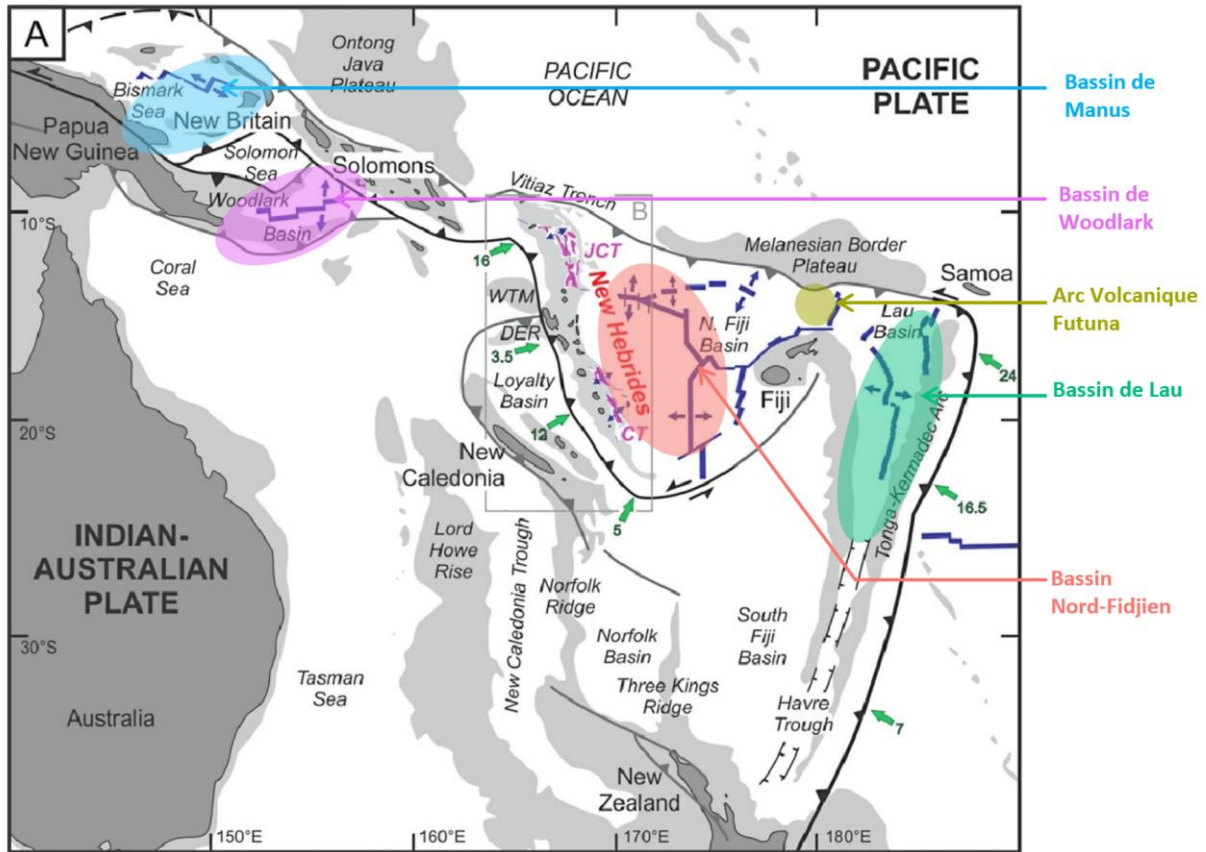


Figure 4 : Représentation des ensembles géologiques du Pacifique Occidental. Les cercles de couleurs représentent les bassins d'arrière-arc et les zones en bleu foncées représentent les zones d'accrétions et d'expansion de la croûte océanique associés avec leurs orientations. (adapté de Keith et al., 2021, Ore geological reviews, Volume 135 Figure 1)

Un certain nombre d'approches génétiques se sont intéressées à tout d'abord décrire et comprendre les structures des populations des espèces hydrothermales du Pacifique Ouest. Ces études ont montré que les quelques espèces étudiées affichaient soit une absence de structuration, soit une très nette différenciation génétique entre le bassin de Manus et celui de Lau, et dans quelques de cas, ont estimé un faible flux génique asymétrique orienté de l'ouest vers l'est (Breusing et al., 2021; Plouviez et al., 2019; Thaler et al., 2011, 2014; Yahagi et al., 2020).

Bien que ces dernières études se soient intéressées à comparer leurs résultats aux précédentes, elles ne disposaient pas d'un échantillonnage exhaustif de tous les bassins arrière-arc pour essayer d'identifier les processus qui ont façonné les patrons actuels de dispersion, et par ailleurs, elles n'ont utilisé qu'un faible nombre de marqueurs génétiques. Il

en résulte une limitation des analyses et reconstructions démo-génétiques de l'histoire évolutive des espèces et d'autre part ces études ne reflètent que l'histoire évolutive associée à ces marqueurs. Malgré ces études pionnières sur la faune hydrothermale dans cette région du monde, les connaissances actuelles sur les processus qui ont influencé les patrons biogéographiques de ces faunes emblématiques restent limitées. Leur compréhension a donc un double intérêt que cela soit pour la recherche fondamentale et la compréhension du vivant, mais aussi pour appréhender les influences et l'impact par des activités anthropiques telles que l'exploitation minière en milieu profond.

Deux grandes hypothèses peuvent être émises concernant l'origine de ces patrons biogéographiques : une première propose que les patrons de distribution des espèces hydrothermales du Pacifique Ouest soient issus d'une fragmentation d'un ensemble passé et continu de dorsales en plusieurs bassins arrière-arc, séparant ainsi au cours du temps les populations. La seconde hypothèse proposerait que l'origine des patrons actuels proviendrait en partie de l'effet d'une dispersion dérivée selon le régime hydrologique de la région avec des sens de migration dépendant de la position des larves dans la colonne d'eau associée à une colonisation des habitats hydrothermaux nouvellement formés au sein de chaque bassin. Ainsi, se posent les questions sur les processus qui ont contribué à la distribution des espèces hydrothermales du Pacifique Ouest. Celle-ci serait-elle majoritairement d'origine géographique et géologique car découlant de la mise en place des arcs volcaniques et zones d'accrétions des bassins ? Ou bien serait-elle plutôt due à l'effet des principaux courants océaniques actuels et/ou liés à l'hydrodynamisme associé aux dernières glaciations, cet effet étant lui-même nuancé par les traits d'histoire de vie des espèces qui conditionnent la dispersion des espèces ?

Les attendus pour la première hypothèse seraient donc que l'on ait une structuration des populations par bassin pour les espèces hydrothermales de cette région, tout en affichant des nuances dans les degrés de connectivité modulés par les traits d'histoire de vie.

Pour la deuxième, les attendus peuvent varier d'une histoire de colonisation en lien avec la formation des bassins et des zones d'accrétions associées en provenance d'une zone ancestrale, à une histoire récente de colonisation/recolonisation ou de remise en contact des

populations, principalement due aux changements hydrologiques que les glaciations peuvent engendrer.

2. Modèles d'étude et questions abordées

Pour répondre à ces questions et effectuer une approche de phylogéographie comparée entre espèces hydrothermales du Pacifique Ouest, nous avons choisi d'utiliser des espèces arborant des traits d'histoire de vie différents : les gastéropodes Provannidae *Ifremeria nautili* et *Alviniconcha kojimai*, le gastéropode patelliforme *Shinkailepas tollmanni*, et le cirripède *Eochionelasmus ohtai* (Figure 5). Le choix et l'échantillonnage de ces espèces a été réalisé de telle sorte que les espèces choisies partagent et couvrent la même aire de répartition sur l'ensemble des bassins d'arrières-arc Pacifique Ouest (Figure 4).

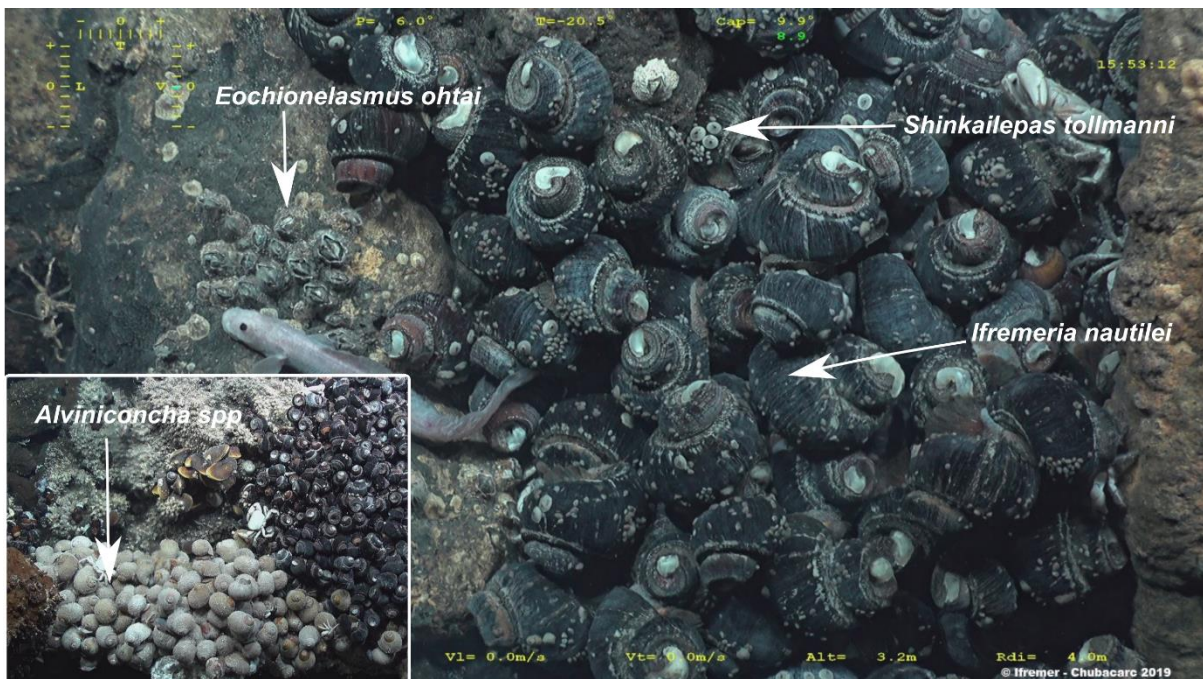


Figure 5 : Photo des animaux in situ prise pendant la campagne océanographique ChubacArc 2019.

Les écosystèmes hydrothermaux du Pacifique Ouest sont assez originaux par rapport ceux des autres régions du monde, car ils sont dominés par des assemblages de grands gastéropodes symbiotiques (*I. nautili* et *Alviniconcha spp.*) et des moulières correspondant à plusieurs espèces des genres *Bathymodiolus* et *Gigantidas* qui sont aussi impliquées dans des symbioses thioautotrophes et méthylo-trophes. Ces espèces s'agrègent dans les zones de diffusion et à la base des cheminées hydrothermales où elles constituent la base des communautés

hydrothermales (espèces architectes). Dans les autres régions du globe, les communautés sont dominées soit par des bivalves (moules hydrothermales, sous-famille des Bathymodiolinae), des vers tubicoles (vestmentifères de la famille des Siboglinidae) ou des crustacés hydrothermaux (principalement des crevettes Alvinocarididae ou des galathées) (Mullineaux, 2013; Wolff, 2005). Sur les parois des cheminées avec des températures généralement plus élevées, d'autres communautés moins diverses sont aussi souvent trouvées. Les vers de la famille des Alvinellidae, parfois très abondants, sont les principaux métazoaires de ces communautés et ils sont accompagnés de quelques autres espèces en moindre abondance (Desbruyeres et al., 2006).

Ifremeria nautilei et *Alviniconcha kojimai* sont des grands gastéropodes appartenant à la même famille (Provannidae) et forment des clades distincts dont le temps de divergence est estimé entre 15-35 millions d'année (Johnson et al., 2010). Ils arborent une relation symbiotique avec des bactéries chimiosynthétiques permettant ainsi d'accéder à une source d'énergie fournie par les fluides hydrothermaux qui contiennent des composés réduits tels que de l'hydrogène et son sulfure, qui sont utilisés par les bactéries symbiotiques (Beinart et al., 2014; Breusing et al., 2020; Seston et al., 2016). A l'état adulte, ils forment des agrégats proches des sorties des fluides hydrothermaux. Ces espèces sont très peu mobiles mais peuvent survivre en l'absence de fluide aux températures de l'eau du fond et se déplacer pour retrouver une nouvelle source active, a priori sur de très courtes distances. *I. nautilei* est une espèce gonochorique avec une fécondation interne. Les femelles possèdent une poche incubatrice où les larves de type lécithotrophe se développent jusqu'au stade de larves véligères (larves coquillées particulières appelées larves de Warèn). Les larves sont ensuite émises dans la colonne d'eau où elles sont capables de nager et peuvent se disperser sans que l'on connaisse vraiment leur durée de vie larvaire pélagique (Reynolds et al., 2010; Warèn & Bouchet, 1993). *A. kojimai* est quant à elle hermaphrodite avec un maintien d'individus mâles et femelles/hermaphrodites dans une proportion de 1 :1 (F. Pradillon, comm. pers.). Les femelles ne possèdent pas de poche incubatrice et émettent leurs gamètes directement dans la colonne d'eau. Les larves sont de petite taille (90 μm) et probablement planctonotrophes.

Shinkailepas tollmanni est une espèce de patelle de la famille Phenacolepadidae, qui se nourrit par broutage (« grazing ») à partir de biomasse bactérienne filamenteuse attachée aux roches et coquilles d'*Ifremeria* dans les zones de diffusion. Cette espèce est retrouvée plus particulièrement sur les coquilles d'*I. nautiliei* et possède un cycle de vie particulier où les œufs fécondés à l'intérieur de la femelle sont pondus à l'intérieur de capsules (quelques dizaines d'individus par capsule) qui sont ensuite déposées exclusivement sur les coquilles d'*Ifremeria*. Les larves se développent jusqu'au stade véligère dans ces capsules. Les larves sont ensuite relâchées dans la colonne après la rupture de la paroi de la capsule. Les larves peuvent alors disperser sur une durée qui reste encore inconnue. Chez une autre espèce de *Shinkailepas*, il a été montré que la larve est capable de nager verticalement vers la surface et pourrait être planctonotrophe (Yahagi et al., 2017).

Eochionelasmus ohtai est une espèce de cirripède de la famille des Chionelasmatidae qui vit relativement proche des fluides hydrothermaux diffus. Elle est considérée comme un « fossile vivant » intermédiaire entre les formes sessiles et pédonculées des cirripèdes (Newman, 1985). Très peu d'informations sont malheureusement disponibles sur sa biologie et ses traits d'histoire de vie. Des observations récentes indiquent cependant que l'espèce serait un hermaphrodite simultané possédant de très gros ovocytes vitellogéniques (300 µm en moyenne) et donc possédant possiblement des larves à développement lécithotrophe allant jusqu'au stade cypris (Yamaguchi & Newman, 1997). L'espèce est probablement incubatrice avec une fécondité par hermaphrodite estimée à environ 150 œufs par ponte sans que l'on sache exactement si la production d'œufs est continue sur l'année.

Nos espèces présentent donc des fécondités très contrastées avec un nombre d'œufs par femelle très important (plusieurs milliers) chez les Provannidae et relativement réduit chez *Eochionelasmus ohtai* et *Shinkailepas tollmanni*) où le nombre d'œufs peut varier de plusieurs dizaines à quelques centaines selon le nombre de capsules pondus ou d'œufs en développement par individu.

L'évolution des espèces abyssales vers un mode de développement larvaire lécithotrophe ou planctonotrophe est intéressante car ce mode pourrait alors conditionner la dispersion à différents niveaux de la colonne d'eau, les larves planctonotrophes dispersant plutôt en surface et les lécithotrophes sur le fond. Dans ce cas, les deux modes pourraient être

hautement dispersifs, le développement des larves lécithotrophes pouvant être une adaptation aux eaux froides et oligotrophes où le développement embryonnaire est fortement ralenti et la métamorphose extrêmement différée et potentiellement initiée lorsque les conditions sont localement favorables à l'installation de la larve (Pradillon et al., 2001).

Le présent travail de thèse vise donc à identifier et à comprendre les patrons de structure et de connectivité des populations des espèces hydrothermales du Pacifique Ouest. Dans ce contexte très particulier des bassins d'arrière-arc, les histoires évolutive et démographique des espèces hydrothermales sont-elles ainsi (1) principalement influencée par la formation des bassins arrière-arc et des zones d'accrétions associées, générant ainsi une différenciation des populations par bassin ou bien (2) sont-elles principalement influencées par les traits d'histoire de vies des espèces et/ou des événements climatiques ? Cette thèse a pour principale ambition d'examiner les patrons de structuration génétique de différentes espèces hydrothermales endémiques et emblématiques du Pacifique Ouest, en comparant des espèces aux traits d'histoire de vie contrastés.

La première partie (chapitre 1) de ce travail de thèse s'est focalisée sur la compréhension de la structuration génétique des populations d'*I. nautiliei* à différentes échelles, par des approches de génomique des populations.

Dans la seconde partie, en utilisant les mêmes approches méthodologiques, (Chapitre 2), nous avons cherché par une approche comparative plurispécifique à comprendre d'une part si les différentes espèces étudiées montrent des patrons de structuration similaires et si les nuances affichées sont en lien avec différents facteurs intrinsèques ou extrinsèques aux espèces. D'autre part, similairement à ce qui a été réalisé sur le premier chapitre pour *I. nautiliei*, des modèles de reconstruction démo-génétique ont été utilisés sur les différentes espèces, de manière indépendante et à partir de données moléculaires génomiques. Ceci permet de reconstruire l'histoire de la divergence des populations pour chaque espèce et d'estimer sa temporalité.

**CHAPITRE I : STRUCTURE GENETIQUE
ET CONNECTIVITE CHEZ LE
GASTEROPODE HYDROTHERMAL
*IFREMERIA NAUTILEI***

I. INTRODUCTION

Les résultats de ce chapitre sont exposés dans l'article joint, sous-pressé dans le journal « *Molecular Ecology* ».

Dans ce premier chapitre, nous nous sommes penchés par des approches de génomique des populations sur une des espèces architectes emblématique des communautés hydrothermales profondes du Pacifique Ouest. Ce chapitre se découpe en trois parties, la première visant à affiner les connaissances concernant la structuration génétique et géographique. La deuxième est une reconstruction de la divergence des populations par des approches démo-génétiques avec une approche par équation de diffusion sous le programme d'inférences *dadi* avec la prise en compte de différents paramètres qui prennent en compte l'hétérogénéité génomique (effet de sélection en liaison et locus barrière). Enfin, la troisième et dernière partie a consisté en l'identification et l'utilisation de marqueurs dits aberrants dits « outliers » pour détecter des subtiles différences concernant la connectivité aux échelles intra- et inter-populations.

II. RESUME (FRANÇAIS)

Les cheminées hydrothermales profondes forment des grappes d'habitats éphémères qui soulèvent des questions intéressantes sur l'évolution et la dynamique de la faune endémique associée, constamment soumise à des processus d'extinction-recolonisation. Ces environnements riches en métaux sont aussi convoités pour les ressources minérales qu'ils abritent, soulevant ainsi de récentes préoccupations en matière de conservation. Le destin évolutif et la résilience démographique des espèces hydrothermales dépendent fortement du degré de connectivité entre et au sein de leurs métapopulations fragmentées. Cependant dans l'environnement profond l'évaluation de la connectivité est difficile et nécessite généralement des approches génétiques indirectes. Grâce aux avancées théoriques et techniques, il est désormais possible d'améliorer l'évaluation de la connectivité génétique à fine échelle grâce à une analyse minutieuse de la différenciation génétique à l'échelle du génome.

Nous avons exploré la connectivité des populations du gastéropode hydrothermal *Ifremeria nautiliei* dans son aire de répartition, qui englobe cinq bassins arrière-arc distincts dans le

pacifique Ouest. L'analyse globale, basée sur 10 570 marqueurs nucléaires (SNP) indépendants et dérivés du séquençage de l'ADN associés au site de restriction à double digestion (ddRAD-seq), a mis en évidence deux groupes génétiques semi-isolés et homogènes, confirmant d'une part les précédents résultats (Thaler et al., 2011). Cependant nos résultats vont plus loin, étendant ces derniers à une échelle géographique plus importante grâce à l'échantillonnage hiérarchique réalisé. Les modélisations démo-génétiques suggèrent que ces deux ensembles génétiques ont commencé à diverger il y a environ 70 000 générations en considérant un taux de mutation donnée (10^{-8}), tout en continuant à présenter un faible flux génétique et légèrement asymétrique. Ces modélisations ont aussi mis en évidence un effet non négligeable de la sélection en liaison ou/et l'existant de locus ayant une migration réduite montrant que l'hétérogénéité du génome est aussi prendre en compte dans les modélisations.

En outre, une analyse minutieuse des locus dits aberrants dits « outliers » a mis en lumière de subtiles limitations de la connectivité entre bassins voisins au sein des deux ensembles génétiques, tout en confirmant le signal de connectivité entre ces derniers. Cette découverte indique que la migration n'est probablement pas assez forte pour contrebalancer totalement la dérive ou la sélection locale, remettant ainsi en question le potentiel de résilience démographique à cette dernière échelle géographique.

Ces résultats illustrent le potentiel des grands ensembles de données génomiques pour comprendre les modèles de connectivité à fine échelle pour les communautés des sources hydrothermales profondes et les milieux abyssaux de façon générale.

III. ARTICLE (MOLECULAR ECOLOGY)

Doi de l'article : <https://doi.org/10.1111/mec.16430>

Le matériel supplémentaire cité dans l'article fait référence aux tableaux et figures dans l'annexe 1.

Subtle limits to connectivity revealed by outlier loci within two divergent metapopulations of the deep-sea hydrothermal gastropod *Ifremeria nautili*

Adrien Tran Lu Y^{1*}, Stéphanie Ruault², Claire Daguin-Thiébaud², Jade Castel², Nicolas Bierne¹, Thomas Broquet², Patrick Wincker⁵, Aude Perdereau⁵, Sophie Arnaud-Haond³, Pierre-Alexandre Gagnaire¹, Didier Jollivet², Stéphane Hourdez⁴, François Bonhomme¹

1:ISEM, Institut des Sciences de l'Evolution, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France,

2: Sorbonne Université, CNRS, UMR 7144, 'Dynamique de la Diversité Marine' (DyDiv) Lab, Station biologique de Roscoff, Place G. Teissier, 29680 Roscoff, France

3: MARBEC, Marine Biodiversity Exploitation and Conservation, Univ Montpellier, CNRS, IFREMER, IRD, Sète, France

4: Sorbonne Université, CNRS, UMR 8222, Laboratoire d'Ecogéochimie des Environnements Benthiques, Observatoire Océanologique de Banyuls, Avenue Pierre Fabre, 66650, Banyuls-sur-Mer, France

5: Génomique Métabolique, Génoscope, Institut de Biologie François Jacob, CEA, CNRS, Université Évry, Université Paris-Saclay, Évry, France

*Corresponding author: adrien.tran-lu-y@umontpellier.fr

Keywords:

Genetic connectivity, Demographic inference, ddRAD-seq, Hydrothermal vents, Western Pacific, Outlier detection

Abstract. Hydrothermal vents form archipelagos of ephemeral deep-sea habitats that raise interesting questions about the evolution and dynamics of the associated endemic fauna, constantly subject to extinction-recolonization processes. These metal-rich environments are coveted for the mineral resources they harbor, thus raising recent conservation concerns. The evolutionary fate and demographic resilience of hydrothermal species strongly depend on the degree of connectivity among and within their fragmented metapopulations. In the deep sea, however, assessing connectivity is difficult and usually requires indirect genetic approaches. Improved detection of fine-scale genetic connectivity is now possible based on genome-wide screening for genetic differentiation.

Here, we explored population connectivity in the hydrothermal vent snail *Ifremeria nautiliei* across its species range encompassing five distinct back-arc basins in the Southwest Pacific. The global analysis, based on 10 570 single nucleotide polymorphism (SNP) markers derived from double digest restriction-site associated DNA sequencing (ddRAD-seq), depicted two semi-isolated and homogeneous genetic clusters. Demo-genetic modeling suggests that these two groups began to diverge about 70 000 generations ago, but continue to exhibit weak and slightly asymmetrical gene flow. Furthermore, a careful analysis of outlier loci showed subtle limitations to connectivity between neighboring basins within both groups. This finding indicates that migration is not strong enough to totally counterbalance drift or local selection, hence questioning the potential for demographic resilience at this latter geographical scale. These results illustrate the potential of large genomic datasets to understand fine-scale connectivity patterns in hydrothermal vents and the deep sea.

Introduction

Understanding the connectivity of populations is a central issue for evolutionary ecology, conservation and management (Cayuela et al., 2018). Direct approaches such as population monitoring or mark-recapture experiments are rarely applicable in marine environments, because many marine species have large population sizes and high dispersal capabilities due to their minute pelagic propagules. These characteristics are likely to reduce the ability of population genetics to assess population connectivity at local and regional scales, except in situations where there is sufficient genetic differentiation or where a large fraction of the population can be sampled (Jones et al., 2005; Pinsky et al., 2010). Deep-sea hydrothermal ecosystems have attracted much attention since their discovery in the late 1970s (Van Dover et al., 2001; Tunnicliffe et al., 1998) due to the oasis-like distribution of these unique and chaotic habitats harboring rich and endemic fauna. Hydrothermal environments are mostly found in tectonically active areas, such as mid-oceanic ridges, where neighboring vents are often separated by tens of meters to hundreds of kilometers, resulting in an almost linear, but fragmented and unstable distribution of vent communities (Chevaldonné et al., 1997; Hannington et al., 2011; Jollivet et al., 1999; Vrijenhoek, 2010).

Studies of slow (*e.g.* Mid-Atlantic Ridge) and fast (*e.g.* East Pacific Rise) spreading ridges have shown that number of species are able to maintain gene flow over thousands of kilometers (Breusing et al., 2016; Craddock et al., 1995; Hurtado et al., 2004; Jollivet et al., 1995; Teixeira et al., 2011, 2012; Won et al., 2003; Yahagi et al., 2019). Many vent invertebrates possess long-range planktonic larvae that can rapidly (re)colonize newly formed sites (Mullineaux et al., 2010). This high larval dispersal capacity leads to local colonization processes following a stepping-stone mechanism of exchanges, or to the formation of more complex metapopulation dynamics where local extinctions and migration may vary greatly according to the geotectonic context of the venting sites (Audzijonyte & Vrijenhoek, 2010; Jollivet et al., 1999; Vrijenhoek, 1997, 2010). Population connectivity can, however, be severely hampered by physical barriers to larval dispersal such as transform faults, diverging ocean currents or microplates (Johnson et al., 2006; Plouviez et al., 2009; Plouviez, Schultz, et al., 2013).

The hydrothermal ecosystems found in the Southwest Pacific are mainly associated with the formation of back-arc basins (BABs). BABs result from complex subduction processes between several plates, leading to a discontinuous and nonlinear distribution of venting sites. Hence, the question arises as to the degree of connectivity between populations inhabiting these BABs,

noticeably to address the issue of their resilience with respect to deep-sea mining projects (Gena, 2013; Niner et al., 2018). In this context, only a few studies to date have focused on understanding the general patterns of spatial genetic connectivity in ecologically vulnerable hydrothermal species. For instance, Thaler et al. (2011) showed that the gastropod *Ifremeria nautilei* is genetically differentiated between the Manus and North-Fiji/Lau basins. Similar results have been reported in other species, such as the limpet *Lepetodrilus schrolli* (Plouviez et al., 2019), the shrimp *Rimicaris variabilis* and the squat lobster *Munidopsis lauensis* (Thaler et al., 2014). Moreover, this latter species is characterized by additional intra-basin structuring. In contrast, the limpet *Shinkailepas tollmani* does not show any differentiation at any of these scales (Yahagi et al., 2020). However, due to the use of only a limited number of markers, none of these studies have reached the resolution necessary for the fine-scale assessment of connectivity in these species.

During the last decade, the development of next-generation sequencing (NGS) and associated techniques have increased the quantity and accessibility of population genomic data, particularly in non-model species. Analyzing these large datasets with thousands of markers along the entire genome using demo-genetic inference methods helps reveal the complex demographic histories of species (Excoffier et al., 2013; Feutry et al., 2020; Gutenkunst et al., 2009; Rougeux et al., 2017; Tine et al., 2014). NGS datasets also give access to unprecedented statistical power to detect non-neutral genetic variation (outlier loci) that can potentially provide finer scale spatial information on connectivity, dispersal and possibly local adaptation (Gagnaire et al., 2015; Milano et al., 2014; Wyngaarden et al., 2017). This information can potentially help distinguish situations of genetic connectivity—whereby local populations are demographically independent (*i.e.* mainly replenished by local propagules) but long-range gene flow mediated by a small number of propagules is sufficient to ensure the circulation of genetic variation among them—from situations of demographic connectivity where a substantial fraction of local population size is made up of immigrants (Lowe & Allendorf, 2010). The consequences of this difference in population connectivity in terms of resilience to local extinction or habitat destruction are quite obvious, with prompt recolonization being expected only in the second situation.

The aim of the present study was to elucidate fine-scale population structure and connectivity using high-throughput double-digest restriction-site associated DNA (ddRAD) sequencing on *Ifremeria nautilei*, whose known distribution covers the Southwest Pacific, from the Manus BAB in Papua New Guinea to the Lau BAB off the Tonga Islands. This hydrothermal gastropod

of family Provannidae harbors chemoautotrophic symbiotic bacteria in its gills to produce organic matter and forms dense aggregations around diffuse fluid venting at temperatures lower than 15°C (Borowski et al., 2002; Windoffer & Giere, 1997). The species is gonochoric with a nearly equal sex ratio and females brood their progeny in a metapodial pouch until the lecithotrophic embryos (several thousands of similar size) reach a specific and unique pre-veliger stage known as Warèn's larva (Reynolds et al., 2010; Warèn & Bouchet, 1993). The gastropod reproduces via internal fertilization leading to a patchwork of brooding and nonbrooding females throughout the year due to asynchronous spawning. Brooding is a reproductive trait that usually limits the dispersal ability of species; however, the lifespan of *I. nautili* pelagic larvae is not known, nor is the maximum distance veligers can travel prior to settlement on a new venting site. Constituting a large portion of the biomass and harboring other species such as *S. tollmanni* and *L. schrolli*, *I. nautili* is a keystone species important for these deep-sea ecosystems. Furthermore, it is classified as endangered by the IUCN (<https://www.iucnredlist.org/species/145380421/145380604>) along with some other hydrothermal species. Therefore, studying the connectivity patterns of *I. nautili* is a flagship endeavor to assess the potential impact of deep-sea mining on this keystone species and its associated fauna.

For this study, extensive sampling was carried out on 29 sites in 17 hydrothermal vent fields across five basins distributed over 5000 km in the Southwest Pacific.

Materials and Methods

Sample collection and DNA extraction

A hierarchical sampling plan was deployed on board the French oceanographic vessel RV *L'Atalante* during the ChubacArc 2019 oceanographic cruise using the remotely operated underwater vehicle (ROV) *Victor 6000*. A total of 684 individuals were collected in the Southwestern Pacific Ocean from 29 sampling sites distributed among 17 vent localities or hydrothermal fields across four BABs and one volcanic submarine area (Futuna), spanning the entire known geographical distribution range of *I. nautili* (Figure 1, SI Table S1). This sampling includes samples from the newly discovered active site La Scala in the Woodlark basin (Boulart et al., in press).

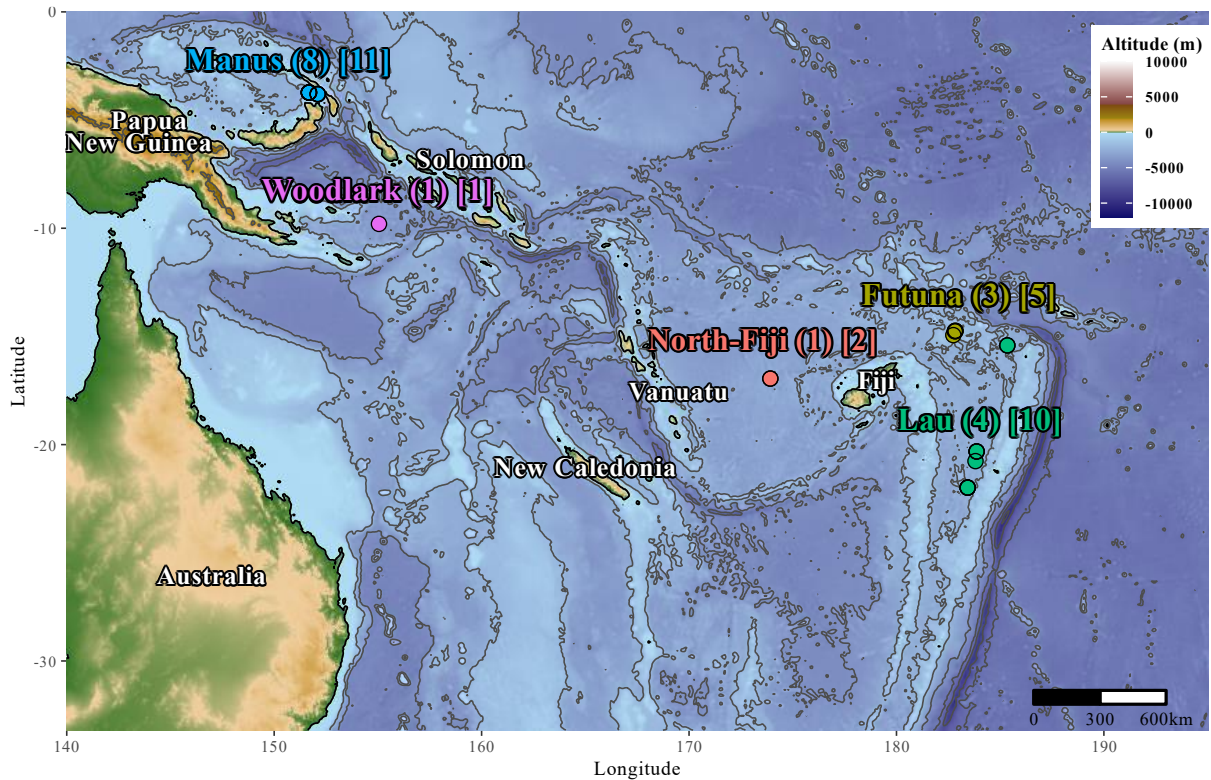


Figure 1: Colors: Back-arc basins. Sampling map of *Ifremeria nautiliei* in the Southwestern Pacific Ocean. The number of localities is given in parentheses and the total number of sampled sites in brackets.

In addition, 27 individuals were added from collections of previous oceanographic cruises, with 22 samples obtained by S. Hourdez during the Lau basin 2009 oceanographic cruise with few at the now-extinct Kilo Moana site and 5 samples obtained during the Manus basin 2009 cruise kindly provided by C. L. Van Dover (SI Table 1). Altogether, a total of 456 unique samples were used and analyzed in this study, of which 362 remained after filtering the sequence dataset (SI Table 2).

Once on board, the snails were dissected and various tissues were preserved in EtOH or frozen at -80°C . Genomic DNA was extracted from fresh foot tissue to limit DNA contamination by symbionts hosted in the gills. A fraction of the tissue samples was stored in 90% EtOH for backup. Samples from C. L. Van Dover's collection were preserved in 90% EtOH, and those from S. Hourdez were kept frozen at -80°C . Genomic DNA was extracted using the NucleoSpin Tissue kit according to the manufacturer's protocol (Macherey-Nagel); some samples were extracted using a standard phenol-chloroform method.

Preparation and sequencing of ddRAD libraries.

DNA extracts were visualized on 0.8% agarose gels and each concentration was standardized to between 10 and 50 ng. μ l⁻¹ after a fluorometric quantification with the QuantiFluor dsDNA system (Promega). Individual double-digest restriction-site associated DNA (ddRAD) *Pst*I-*Mse*I libraries were prepared following (Brelsford et al., 2016) after modifications detailed in (Daguin-Thiébaud et al., 2021). Five pooled libraries were prepared with a combination of four to eight Illumina indexes and 24 barcodes per index, multiplexing a total of 486 samples, including 27 pairs of replicates for quality control, representing 456 individuals. The sequencing effort was sized to produce on average 3×10^6 read pairs per individual. Each genomic pool was sequenced on one lane of a HiSeq4000 Illumina sequencer (paired end, 150 bp) at the Genoscope sequencing facility (Centre National de Séquençage, Evry, France).

De novo ddRAD-tag assembly, SNP calling and filtering

Fastqc (V.0.11.9) was used only for quality control of raw reads, no filters were applied on them. Individual reads were demultiplexed using the “Processradtag” pipeline in Stacks (V.2.52) (Rochette et al., 2019). Due to the lack of a reference genome for *I. nautiliei*, reads were assembled *de novo* using each Stacks module one by one (ustacks, cstacks, sstacks, tsv2bam, gstacks and populations). To identify the most appropriate assembly parameters, we followed previously published recommendations (Mastretta-Yanes et al., 2015; Paris et al., 2017) (See SI and SI Figures 1 - 6 for details). Briefly, we used the genotyping error rate between replicates, the number of variants (SNP), polymorphic loci (ddRAD-tags) and nucleotide diversity (π estimated in Stacks) as a function of the assembly parameters (m and $M = n$) determined with a subset of individuals covering all basins and localities ($n = 84$). The selected parameters were as follows: $m = 4$ (the minimum number of reads to assemble a stack), $M = 8$ (the maximum number of mismatches between putative alleles within individuals), $n = 8$ (the maximum number of mismatches between putative loci within the catalog of individuals) and $R = 0.8$ (the minimum percentage of individuals sharing a locus across all populations in the “populations” module).

After *de novo* assembly, several filters were applied using VCFtools (V.0.1.16) (Danecek et al., 2011) to reduce missing data and to account for potential paralogs (see SI Table 3). Briefly, we removed 1 of each of the 27 replicated individuals with the highest value of missing data. Then, we excluded SNPs with heterozygosity > 0.6 , SNP and only individuals with less than 10% missing data were kept. Variants with a mean coverage higher than 80X were excluded. Using

VCFtools, we excluded loci with a minor allele frequency (MAF) lower than 5% (alternative allele), followed by those that deviated significantly from Hardy-Weinberg equilibrium (p -value ≤ 0.05). We then kept only one randomly chosen SNP per ddRAD-tag to avoid short distance linkage disequilibrium between SNPs. PGDSpider (V.2.1.1.5) (Lischer & Excoffier, 2012) was used to convert the final VCF into the formats required for subsequent analyses.

Population structure and diversity

Principal component analysis (PCA) was first performed on the final dataset with the R package SNPrelate (V.1.24.0) (Zheng et al., 2012). Pairwise fixation indices (F_{ST}) were calculated in Arlequin (V.3.5.2.2) (Excoffier & Lischer, 2010). AMOVA (Excoffier et al., 1992) was performed with 10 000 permutations of genotypes between populations by considering hierarchical geographic structure of localities within basins (See SI for parameters). Co-ancestry analyses were performed through ADMIXTURE (V.1.3.0) (Alexander & Lange, 2011) with 10 independent runs for $K = 1$ to 5. The best K value was selected by using the cross-validation error as recommended by the authors. Runs of ADMIXTURE were grouped using CLUMPAK (Kopelman et al., 2015), graphical visualizations of the results were plotted using library ggplot2 (V.3.3.3) in R (V.4.0.1). TreeMix (V.1.13) (Pickrell & Pritchard, 2012) was performed with 10 independent runs with migration events ranging from 0 to 5. The optimal number was selected according to the log-likelihood of each model. F_3 admixture tests (Reich et al., 2009) were done using the THREEPOP programs implemented in TreeMix (V.1.13) package with default values.

To detect potential kinship, SNPrelate was used to compute identity-by-state between pairs of individuals. We used this approach to (1) minimize the risk of labeling error/exchange during the process of library construction and sequencing and (2) infer the level of kinship structure between non-replicated individuals because the existence of undetected underlying kinship structure can distort the population structure estimated by the gene genealogies.

Observed heterozygosity (H_o), expected heterozygosity (H_e), heterozygote deficiency (F_{IS}), nucleotide diversity (π) and raw nucleotide divergence (D_{xy}) were calculated with the population module of Stacks using all sites from all ddRAD-tags in the final dataset.

Demo-genetic history of the species

The demographic history of the species targeting past and present gene flow between metapopulation clusters was inferred using a modified version of $\partial a \partial i$ (V.2.1) (Diffusion

Approximations for Demographic Inference; Gutenkunst et al., 2009), with a dual annealing optimization function. This software simulates the joint allele frequency spectrum (JAFS) of two (or more) interacting populations according to different demo-genetic scenarios. Here, we considered 28 distinct scenarios built according to the population models used in Rougeux et al. (2017) with very few modifications, detailed below.

Basically, all these models derive from four basic models representing strict isolation (SI), isolation with migration (IM), ancient migration (AM), and secondary contact (SC). Each of them consists of an ancestral population of N_{anc} size that splits into two sister populations of effective size N_1 and N_2 during time T_s for the (SI) and T_{sm} for the (IM) model, $T_{am}+T_s$ for the (AM) model and T_s+T_{sc} for the (SC) model, where T_s is the time spent since the split of the two populations without migration, T_{sm} , the time spent since the split of the two populations with migration, T_{am} , the duration of the ancient migration period after the split of the ancestral population and before the emergence of strict isolation (T_s) and T_{sc} the duration of a secondary contact after a period T_s of strict isolation. Directional migration between populations is allowed at rates m_{12} and m_{21} from population 2 to population 1 and vice versa.

Further complexity was introduced as in Rougeux et al. (2017), by adding several processes occurring after the split, such as population expansion or contraction (G), the effect of linked selection reducing the effective population size of loci over a certain fraction of the genome ($2N$) and the effect of semipermeable genetic barriers (*i.e.* partial reproductive isolation) reducing the effective migration rate of loci over a certain fraction of the genome ($2m$). Furthermore, to dissociate the effect of the effective population size (genetic drift) and migration (gene flow), we only allowed the growth parameters (G) to vary during the migration phase of each model. Graphical representation of the four basic models and the three more complex models are displayed in SI Figure 7.

For the input dataset, we considered the two metapopulations defined by the global analyses (see Results), which corresponded to Manus/Woodlark and North-Fiji/Futuna/Lau. We used the folded joint allele frequency spectrum (folded JAFS), because no external group was available to identify the allelic ancestral states. All models were fitted independently of the dataset using dual-annealing optimization and run 10 times independently each to check convergence. Model comparisons were based on the Akaike information criterion (AIC). Using the best selected models, we then converted demographic parameters into biological units. In the absence of precise information on mutation rate and generation time for this species, we used 10^{-8} as the

mutation rate per site per generation. This widely used value falls within the range proposed by (Lynch, 2010), although admittedly the real value may be much larger or much smaller, as recalled in the Discussion. Parameters estimated using $\partial a \partial i$ are scaled from the ancestral effective population size (N_{anc}), which was estimated using the following formula:

$$N_{anc} = \frac{\theta}{(4 \times \mu \times L)}$$

where L represents the total length of the DNA sequence used in $\partial a \partial i$:

$$L = \frac{z \times y \times 275}{x}$$

where z represents the number of SNPs used, y the number of RAD-tags of 275 bp, and x the initial number of SNPs ($z = 17\,365$, $y = 17\,365$ and $x = 250\,502$, $L = 331\,032$).

Estimated times were calculated in units of $2 \times N_{anc}$ generations and the migration parameters (m_{12} and m_{21}) were divided by $2 \times N_{anc}$ to obtain the number of migrants in each population per generation. The standard deviations were estimated using the Fisher information matrix (FIM) method implemented in $\partial a \partial i$.

Outlier loci and detection of fine-scale structure

To test whether fine-scale genetic structure exists within each genetic cluster defined as a result of the global analyses described in the preceding paragraph (see Results for details), we used several genome-scanning methods to identify candidate outlier SNPs (*i.e.* loci showing higher or lower levels of differentiation than expected under assumed neutrality). Such loci may be informative about fine-scale population structure and connectivity patterns (Gagnaire et al., 2015). Four different outlier detection approaches were used. The rationale behind this multiple testing is that these methods operate with somewhat different underlying assumptions or test statistics and are known to have varying discriminatory power depending on the situations to which they are applied (Villemereuil et al., 2014). Outlier loci were selected according to statistical thresholds (p -value ≤ 0.05 and 0.01) in each software package, while checking that candidate loci outnumbered the number of loci expected to fall outside the distribution by chance only (false positives). Furthermore, to focus on the relevant scale and avoid the detection of false positives due strong geographic structuring, these programs were run independently on each Manus/Woodlark and North-Fiji/Futuna/Lau metapopulation previously defined in the

global analyses, while considering populations either by basin or by locality within these groups.

Four methods were used. (1) BayeScan (V.2.0) (Foll & Gaggiotti, 2008) detects potential outlier loci by using differences in allele frequencies under a simple island model in a Bayesian framework. Five independent runs were performed with the default settings. (2) PCAdapt (V.4.3.3) (Luu et al., 2017) uses the correlation of SNPs with the first principal components of the PCA to detect outliers by computing a Mahalanobis distance between their z-score on each PC. (3) Arlequin (V.3.5.2.2) (Excoffier & Lischer, 2010) detects outlier SNPs under a non-hierarchical finite island model integrating F_{ST} and heterozygosity through 20 000 coalescence simulations of the neutral distribution with 100 demes each. (4) The core model of Baypass (V.2.1) (Gautier, 2015) based on a hierarchical Bayesian model in which loci that are more differentiated than expected under a non-equilibrium drift model are identified through the distribution of a statistic similar to F_{ST} corrected to account for demographic history. Baypass was run five times independently with default settings under the core model.

PCA, ADMIXTURE and F_3 tests were then performed on the various outlier subsets to explore the information they convey.

In addition, outlier loci identified at the threshold ($p \leq 0.05$) were first blasted (BLASTN, E-value threshold: 10^{-5}) against the *Alviniconcha boucheti* transcriptome, which was previously assembled using rnaSPAdes (V.3.13.1) (Bankevich et al., 2012) (cf. Castel et al., in prep). Transcript hits with a size greater than 300 bp were subsequently blasted (BLASTX, E-value threshold: 10^{-5}) against the NCBI UniProtKB/Swiss-Prot database using the software Geneious Prime® 2021.2.2.

Results

De novo assembly and data filtering

De novo assembly resulted in a dataset of 38 608 ddRAD-tags with a mean coverage of 14X for 486 samples. The mean genotyping error rate was 0.48% and the maximum value was 1.06% from all pairs of replicates. These ddRAD-tags contained 649 106 SNPs. Following SNP filtering, the final dataset resulted in a VCF file containing 362 individuals with 10 570 unlinked bi-allelic variants with an individual mean coverage of 17.7X and a maximum of 10% of missing data per individual and variant.

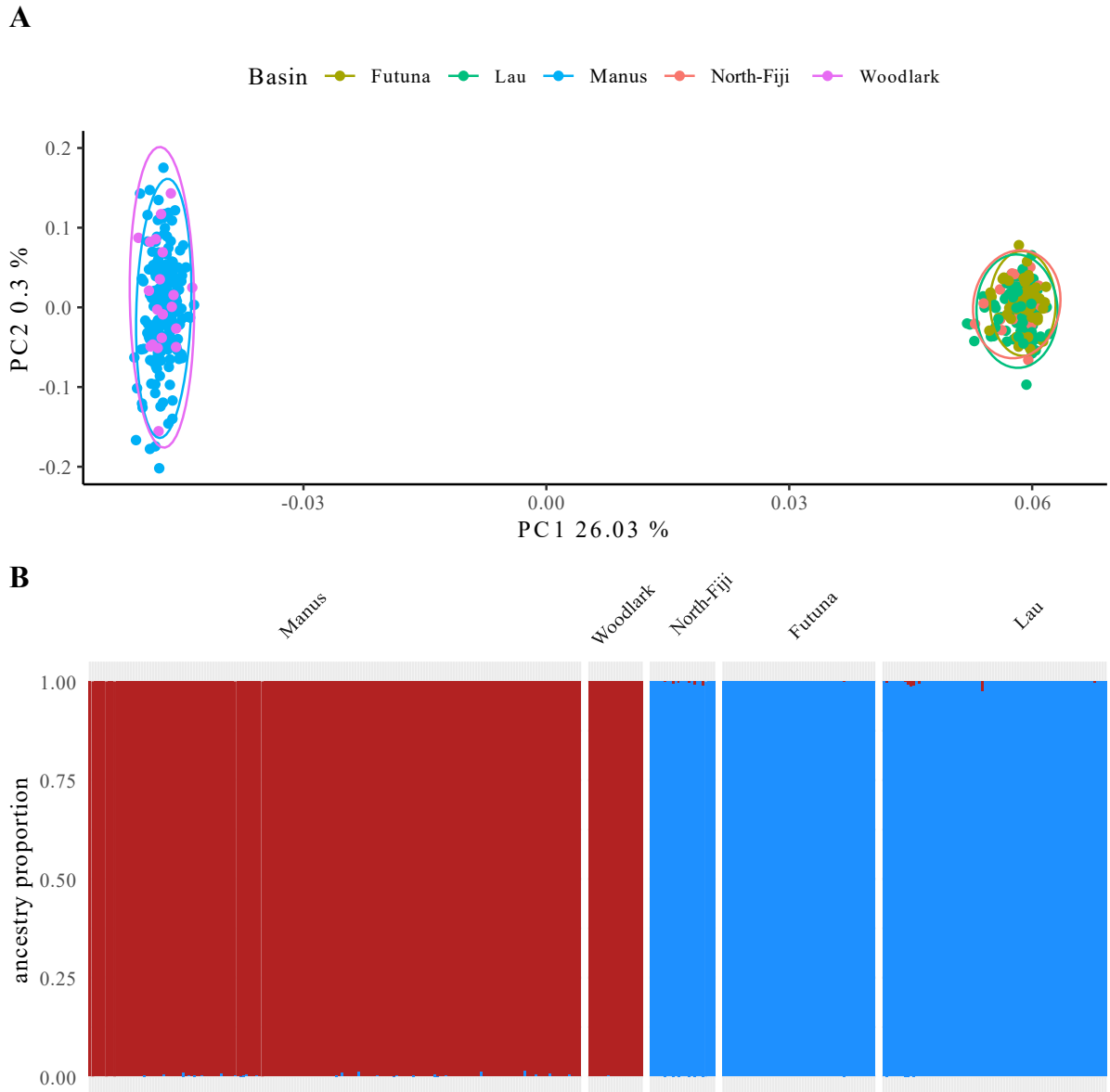


Figure 2: (A) Principal component analysis plot of 362 *Ifremeria nautilei* individuals from five hydrothermal basins scored at 10 570 SNPs, open circles represent the multivariate normal distribution of each group (basins) at 95%. (B) ADMIXTURE plot for each individual with their ancestry proportions obtained on the final dataset for the best K ($K=2$). Individuals are grouped according to their basin of origin.

Population structure analyses considering the global dataset

A PCA was performed to explore the level of population structure over the five western Pacific BABs (Figure 2, A). This analysis showed a very clear geographical separation with two distinct clusters, one corresponding to the Manus and Woodlark basins and the other to the North-Fiji, Futuna and Lau basins. The first component (PC1) explained 26.03% of the total variance; the second one carried only 0.03% of the total variance (Figure 2 and SI Figure 8). This pattern was consistent with the AMOVA results (Table 1), which also showed a strong and significant

genetic differentiation between Manus/Woodlark and North-Fiji/Futuna/Lau, but no differentiation among basins and localities within these two groups (between Manus/Woodlark and North-Fiji/Futuna/Lau, $F_{ST} = 0.387$, p -value < 0.001 , inter-basins within Manus/Woodlark and North-Fiji/Futuna/Lau, $F_{CT} = -0.050$, NS). In addition, the between-basins pairwise F_{ST} (Table 2) were only significant between Manus/Woodlark and North-Fiji/Futuna/Lau pairs. No significant pairwise F_{ST} values were observed between localities within either of the two groups Manus/Woodlark and North-Fiji/Futuna/Lau (SI Table 3).

This finding is also consistent with the ADMIXTURE (Figure 2 B & SI Figures 9–10) clustering results and strongly supports the same two distinct clusters ($K = 2$) with very few individuals showing very low percentages of mixed ancestry (from 0.1% to 3%). The identity-by-state distribution (SI Figure 11) did not show evidence of any internal structure due to kinship.

Table 1: Analysis of molecular variance (AMOVA) on the final dataset with 10 000 permutations (***: $p < 0.001$, **: $p < 0.01$ *: $p < 0.05$)

<i>Manus/Woodlark vs. North-Fiji/Futuna/Lau</i>	0.38773***	F_{ST}
<i>Basins in M/W and NF/F/L</i>	-0.05	F_{CT}
<i>Localities within Basins</i>	-0.00011	F_{SC}
<i>Individuals within Localities</i>	-0.05084	F_{IS}

M/W: Manus/Woodlark, NF/F/L: North-Fiji/Futuna/Lau

Table 2: Pairwise (between basins) F_{ST} matrix on the final dataset with 10 000 permutations after Bonferroni correction (***: $p < 0.001$, **: $p < 0.01$, * $p < 0.05$).

	<i>Lau</i>	<i>Futuna</i>	<i>North-Fiji</i>	<i>Manus</i>	<i>Woodlark</i>
<i>Lau</i>	0.00000				
<i>Futuna</i>	-0.00040	0.00000			
<i>North-Fiji</i>	0.00029*	-0.00004	0.00000		
<i>Manus</i>	0.38350***	0.38275***	0.37651***	0.00000	
<i>Woodlark</i>	0.39986***	0.39656***	0.38647***	-0.00016	0.00000

TreeMix analyses showed an optimal number of two migration events, whereas additional events did not improve the likelihood (SI Figure 12 A). Displaying the first migration event showed a very low migration weight from Manus/Woodlark towards North-Fiji (SI Figure 12 B), and adding the second migration event indicated a very slight differentiation between Woodlark and Manus (SI Figure 12 C). The F_3 statistics showed a significant admixture signal, with source populations from each genetic cluster only when North-Fiji was chosen as the focal population (SI Figure 13).

The estimated genetic diversity of the populations considering all DNA positions of the 10 570 ddRAD-tags was bimodal, with slightly higher nucleotide diversity in Manus/Woodlark ($\pi = 0.00572$) compared with North-Fiji/Futuna/Lau ($\pi = 0.00535$), regardless of the populations being considered by geographic basin or by genetic cluster (SI Figure 14). The raw nucleotide divergence (D_{xy}) between the two genetic clusters was estimated to be 0.0136.

Hence, the analyses of the complete SNP dataset of *I. nautili* indicate the co-occurrence of two quasi-panmictic metapopulations, one associated with the Manus/Woodlark basins and the other with the North-Fiji/Futuna/Lau basins, on either side of the Solomon Islands/New Hebrides arc. Thus, these two metapopulations are both sufficiently homogeneous and differentiated from each other to be analyzed using $\hat{d}a\hat{d}i$ demo-genetic inference, which aims at summarizing the global genome-wide history of divergence/contact between them over a long period of time.

Inference of demographic history and gene flow

1. Model comparisons

The folded JAFS in Figure 3 (A) shows how allele frequencies are shared between the Manus/Woodlark and North-Fiji/Futuna/Lau metapopulations. The $\partial a \partial i$ framework can fit population models on the observed dataset and compares them based on their AIC values (Figure 4).

Among the four simplest models (SI, AM, IM and SC), SC was significantly the best fitting model. Increasing complexity by adding the parameters G, 2m, 2N independently improved the AIC values regardless of the basic model used. However, capturing the effect of linked selection (2N) explained the data much better than models with population growth (G) and heterogeneous gene flow (2m) only (Figure 4).

Conversely, the combination of these parameters led to only a slight improvement in the AIC values. Nevertheless, models including the effect of linked selection (2N) were still better than the other models (Figure 4).

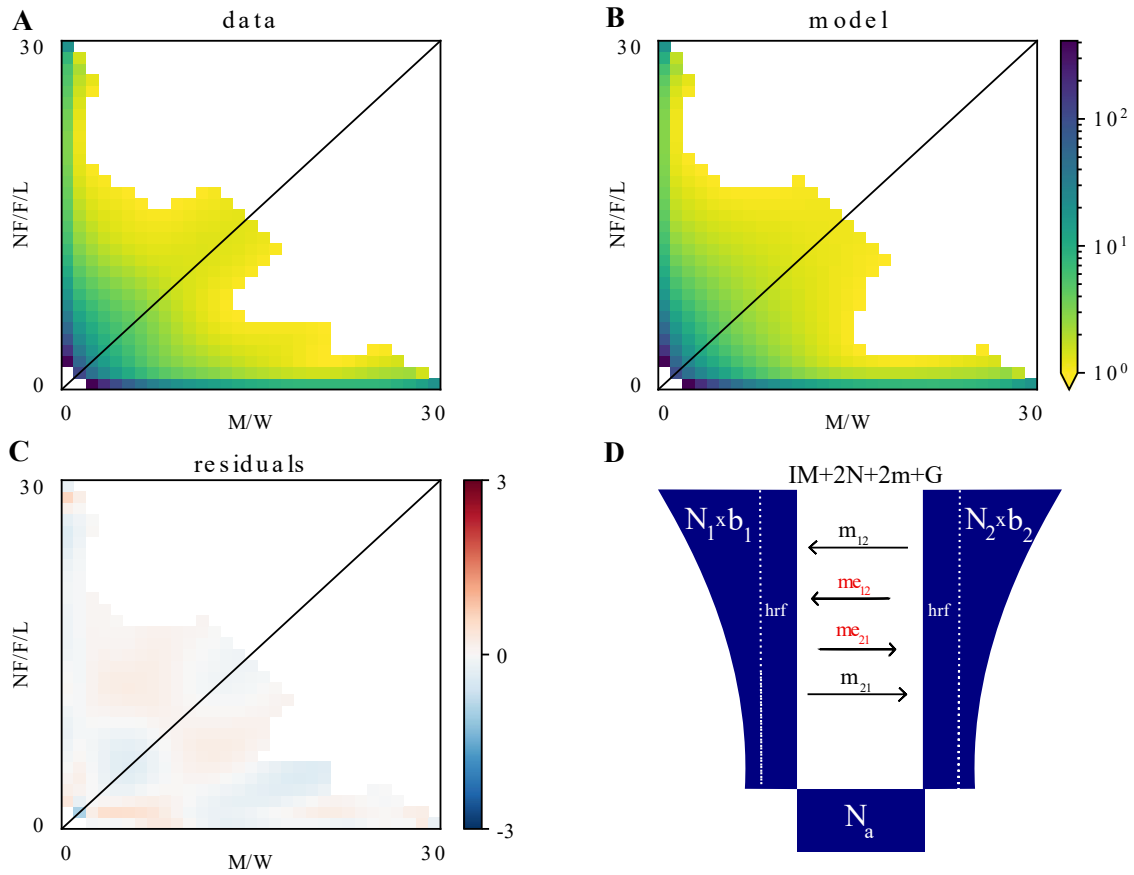


Figure 3: (A) Joint allele frequency spectrum (JAFS) between the Manus/Woodlark (M/W) and North Fiji/Futuna/Lau (NF/F/L) basin systems. (B) Simulated JAFS under the IM2N2mG model (see Figure 4), the log scale indicates the density of SNPs in each frequency class. (C) Residuals of the fit of the simulated model on the data. (D) Representation of the fitted model. (N represents population size; b , population growth factor; hrf , the Hill-Robertson factor, which simulates linked selection; m , unrestricted migration rate; me , restricted migration rate, which simulates barrier loci)

Hence, considering all models together, the best ones were those that took all parameters ($2N$, $2m$, G) into consideration, followed by models with only $2N+G$ (Figure 4). Moreover, for the $2N+2m$ and $2N+2m+G$ models, the proportion $1-P$ of the genome that evolves under restricted migration in $2m$ models amounted to 0.52–0.56 for the best AIC simulation among all runs, meaning that the proportion of the genome that evolves under a reduced effective migration rate (barrier loci) may be quite substantial. With the increasing number of population parameters, the secondary contact hypothesis was no longer the best evolutionary scenario explaining our genetic dataset: the models IM+2N+2m+G, SC+2N+2m+G, AM+2N+2m+G and AM+2N+G, IM+2N+G, SC+2N+G showed very similar AICs ($\Delta_{AIC} \leq 10$, Figure 4).

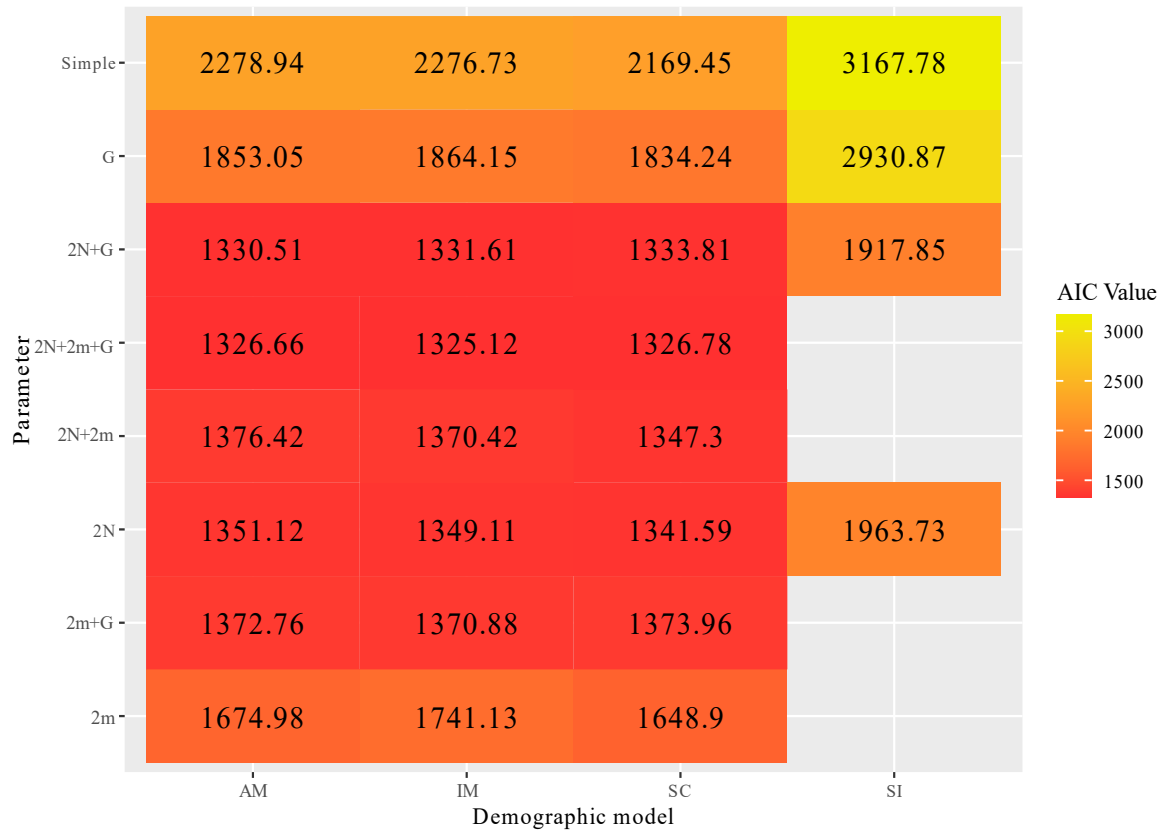


Figure 4: Heat-map of the best Akaike information criterion (AIC) value for each parameter combination (population expansion or contraction (G), effect of linked selection ($2N$) and heterogeneous migration ($2m$)) and demographic model (strict isolation (SI), isolation with migration (IM), ancient migration (AM), and secondary contact (SC)).

2. Inferences of model parameter values

According to the best models based on AIC, the two metapopulations may have diverged due to early (AM), late (SC) or constant (IM) gene flow and it is difficult to distinguish among these three possibilities. However, these models have some interesting features in common. First, although the standard deviations (SDs) are rather large, the effective population sizes of the two derived populations estimated since the split (N_1 & N_2) indicate a demographic expansion (b_1 and $b_2 > 1$), regardless of the model, including a temporal change in population size (G). Second, the local effect of selection at linked sites seems to affect a very large proportion of loci ($Q = 0.99$) with a small value of hrf (Hill-Robertson factor = ~ 0.02) (Table 3). Third, the number of contemporary migrants (estimated by $(N_1 * b_1 * m_{12})/2$ and $(N_2 * b_2 * m_{21})/2$) shows asymmetrical, but weak flow between the two metapopulations, slightly higher from North-Fiji/Futuna/Lau to Manus/Woodlark (4.2–4.6) than in the opposite direction (2.9–3.2). Fourth, nearly half of loci show a restricted migration rate ($0.52 \leq 1-P \leq 0.56$). Fifth, the total estimated

divergence time expressed as the number of generations are very similar, regardless of the model (Tsm, Tam+Ts, Ts+Tsc) and estimated to be between 66 951 and 70 295 generations.

Table 3: Parameters estimated from $\partial\text{a}\partial\text{i}$ for the three best models (IM2N2mG, SC2N2mG and AM2N2mG*) with their standard deviations (SD) estimated using a Fisher information matrix. (*isolation with migration (IM), secondary contact (SC), ancient migration (AM) and with parameters describing effective population size (2N), migration rate (2m) and population growth (between basins))

Parameter	IM+2N+2m+G	SD	SC+2N+2m+G	SD	AM+2N+2m+G	SD
N_1 (NF/F/L)	0.435	0.187	0.913	0.573	0.390	0.127
N_2 (M/W)	0.411	0.157	0.840	0.573	0.356	0.119
$b1$	30.410	13.176	16.947	8.127	34.367	12.399
$b2$	25.097	8.683	13.714	8.451	29.288	10.523
hrf	0.023	0.006	0.021	0.006	0.022	0.006
Ts			0.443	0.527	0.001	0.024
$Tsm/Tsc/Tam$	1.631	0.280	1.470	0.379	1.681	0.335
$m12$	0.444	0.147	0.422	0.145	0.461	0.137
$m21$	0.825	0.192	0.810	0.261	0.825	0.198
$me12$	0.038	0.020	0.038	0.024	0.039	0.018
$me21$	0.283	0.063	0.270	0.088	0.300	0.070
P	0.483	0.103	0.439	0.093	0.471	0.095
Q	0.990	0.136	0.990	0.185	0.990	0.136
Θ	271.772	31.55	243.283	64.91	273.222	32.988

N represents the population size of each population; b , the population growth factor; hrf , the Hill-Robertson factor; Ts , the time of strict divergence; $Tm/Tsc/Tam$, the time of divergence with migration; $m12$, represents the unrestricted migration rate from the population 2 towards population 1; $me12$, the restricted migration rate (e.g. barrier loci) from population 2 towards population 1; Q , the proportion of loci that are under the effect of linked selection (i.e. Hill-Robertson effect); P , the proportion of loci that have unconstrained migration; M/W , Manus/Woodlark; $NF/F/L$, North-Fiji/Futuna/Lau

Table 4: Estimates of the effective number of migrants (N_m) exchanged per generation between metapopulations, total time of divergence since the population split and effective population size (N_e) for three demographic models (isolation with migration (IM), ancient migration (AM), and secondary contact (SC)).

	$N_{m2 \rightarrow 1}$	$N_{m1 \rightarrow 2}$	T	N_1	N_2
	(M/W->NF/F/L)	(NF/F/L->M/W)	(in generations)	(NF/F/L)	(M/W)
IM+2N+2m+G	2.935	4.255	66 951	271 506	211 708
SC+2N+2m+G	3.265	4.665	70 295	284 279	211 653
AM+2N+2m+G	3.09	4.305	69 380	276 561	215 142

Metapopulation M/W, Manus/Woodlark; metapopulation NF/F/L, North-Fiji/Futuna/Lau.

Outlier loci and detection of fine-scale structure

Despite the absence of geographic structuring within each metapopulation depicted in the global analysis, several outlier loci were identified in each metapopulation at the thresholds of $p \leq 0.05$ and $p \leq 0.01$ (SI Table 4).

BayeScan identified much fewer candidate outliers than expected by chance only, and hence was not considered further. PCAdapt was largely out of its working range because it searches for loci that exceed the possible differentiation level captured by the very first principal component as opposed to the second-order axes. However, all axes, except axis 1 which separates the two geographic metapopulations, belong to this second category due to the lack of internal structure. Hence, all the second-order axes primarily captured noise, and they were unable to reveal additional structuring (SI Figure 15,16). Therefore, Baypass and Arlequin were the only two methods considered further. To increase the probability of considering true positives only, we only kept loci identified in both approaches (predicted by the intersection depicted in the Venn diagram in Figure 5 at the thresholds $p \leq 0.05$ and $p \leq 0.01$ in SI Figure 17). Only 458 and 223 outlier loci were shared between the methods Baypass and Arlequin in Manus/Woodlark and North-Fiji/Futuna/Lau at the threshold of $p \leq 0.05$ respectively.

PCA based on these different sets of outlier loci helped to visualize their contribution to the internal heterogeneity of each metapopulations (Figure 5 and SI Figure 18). Interestingly, although outliers were defined within each metapopulation, the inter-metapopulation differentiation was retrieved in all cases. Nevertheless, a clear signal of differentiation was highlighted within both regions. Individuals from the Manus and Woodlark basins showed clear genetic differentiation with no overlap on PC2 (Figure 6 A). Individuals from North-Fiji were slightly pulled towards Manus/Woodlark individuals based on PC1 (Figure 6 B), but they were also shifted on PC2 when outliers were considered at the threshold of $p \leq 0.05$.

The ADMIXTURE analyses based on the outlier SNPs datasets with the threshold of $p \leq 0.05$ displayed optimal K values at $K=2$. With North-Fiji/Futuna/Lau outliers, North-Fiji displayed an admixture proportion from Manus/Woodlark ranging between 5 and 15% (Figure 6 C). For Manus/Woodlark outliers, we also found an admixture proportion from North-Fiji/Futuna/Lau in Woodlark (Figure 6 D). However, very similar values of cross-validation errors were obtained with $K=3$ (SI Figure 19,20,21), North-Fiji and Woodlark each being individualized as the third cluster in their respective runs.

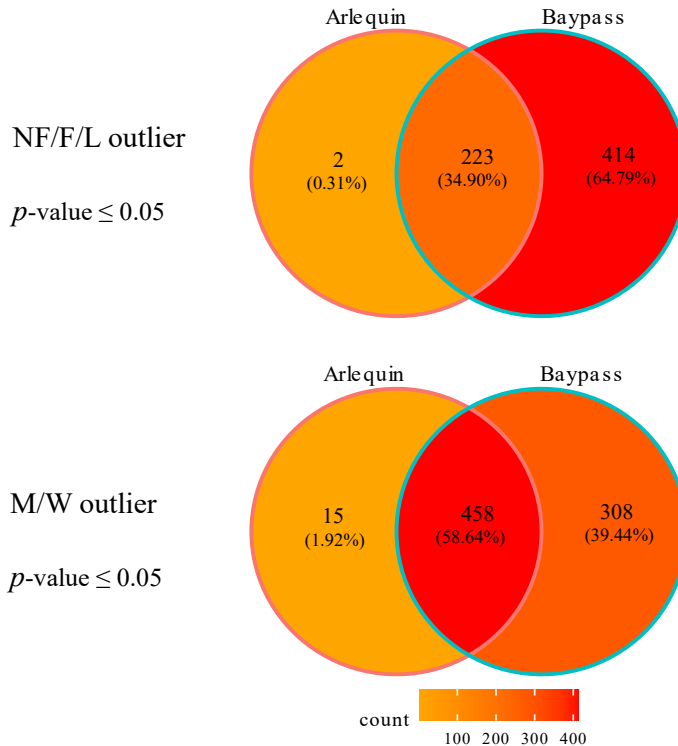


Figure 5: Venn diagram of shared outlier loci identified in Arlequin and Baypass with a p -value less than or equal to 0.05 within each metapopulation (Manus/Woodlark, M/W and North Fiji/Futuna/Lau, NF/F/L) independently.

The F_3 statistics calculated based on outliers only did not provide any additional information (SI Figure 22).

When blasted onto the *A. boucheti* transcriptome, 30% of outlier loci identified at the threshold of $p \leq 0.05$ matched with the coding sequences of transcribed regions. This number was greater than expected by chance from randomly picked ddRAD loci along the *I. nautilei* genome. Half of these 30% of outlier loci (129 transcripts) had annotations on the protein database. Among these annotations, many involved genes that encode for DNA/RNA replication and repair enzymes, transmembrane carriers and synapse/microtubule biosynthesis, but also genes involved in the exocytosis/endocytosis regulation, and more especially the GTPase regulation pathway (SI Table 7). In addition, two genes involved in spermatogenesis were also detected.

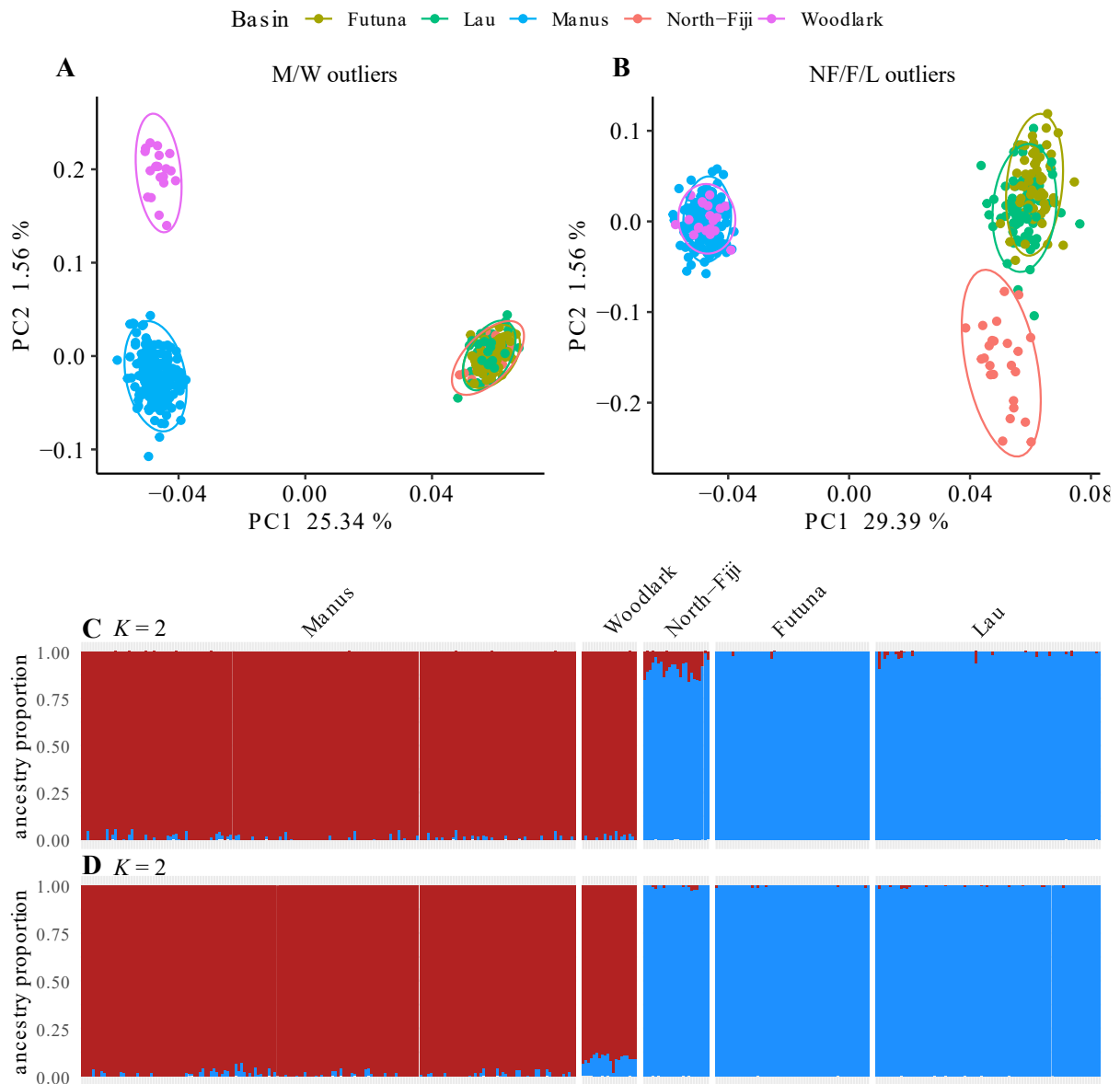


Figure 6: Principal component analysis on all individuals with the outlier loci found in both Arlequin and Baypass at a 0.05 p-value threshold. (A) Outlier loci detected in Manus/Woodlark (M/W). (B) Outlier loci detected in North-Fiji/Futuna/Lau (NF/F/L). Plot of ancestry proportion inferred with Admixture at $K = 2$ on all individual at the 0.05 p-value threshold, for (C) North-Fiji/Futuna/Lau outliers and (D) for Manus/Woodlark outliers.

Discussion

Previous work by Thaler et al. (2011) using microsatellites and mitochondrial *cox1* sequences demonstrated that the southwestern Pacific deep-sea hydrothermal vent gastropod *Ifremeria nautilei* is genetically structured into two distinct populations from Manus and North-Fiji/Lau BABs. Our study extends these previous results to a finer scale, owing to our larger, nested sampling design that includes the newly discovered La Scala vent field in the Woodlark basin (Boulart et al., in press), the Futuna volcanic arc (Konn et al., 2016) and the newly discovered northernmost Mangatolo site at the entrance of the Lau basin. Using a 10 570 SNP genome-wide dataset—unprecedented for a hydrothermal species—, we confirm that *I. nautilei* is structured into two loosely connected metapopulations corresponding to two BAB ensembles. These ensembles display an almost complete internal genetic homogeneity; however, our analyses of outlier loci nevertheless revealed fine-scale differentiation among basins within each metapopulation. We discuss below the possible implications of these results in terms of larval dispersal and demographic connectivity and ultimately their consequences on the resilience of hydrothermal communities.

Long-term gene flow and history of differentiation

One metapopulation comprises the Manus and Woodlark basins (*i.e.* the Manus/Woodlark BAB) west of the Salomon/New Hebrides arc, whereas the other extends east of it with the North-Fiji basin, Futuna volcanic arc and Lau basin (*i.e.* the North-Fiji/Futuna/Lau BAB). The genetic divergence between the two metapopulations is relatively strong (average $F_{ST} = 0.387$, $p \leq 0.001$, $D_{xy} = 0.0136$), but each of these two ensembles appears to be panmictic ($F_{CS} = -0.05$, NS, SI Table 3). The demo-genetic inferences gleaned from $\partial a \partial i$ suggest that the two metapopulations diverged with only a brief period of isolation (T_s was found to vary between 0.001 and 0.443 in the AM and SC models), although the existence of constant gene flow (IM) could not be formally excluded. The incorporation of several other demographic parameters (2N, 2m, G) produced a clear improvement in model fit. Considering each parameter independently, the effect of linked selection (2N) had a much greater influence on AIC than the other two parameters (2m and G), suggesting that a non-negligible proportion of loci may be influenced by linked selection. For the best models (2N+2m+G), this proportion approaches $Q = 0.99$ (which seems to be unrealistic), whereas only half of the markers appear to be under the influence of heterogeneous migration ($0.52 \leq 1-P \leq 0.56$). Nevertheless, these models are very close to the 2N+G models ($\Delta_{AIC} \leq 10$), which estimate a proportion of loci under linked selection

($0.56 \leq Q \leq 0.58$) and do not take into account the effect of heterogeneous migration. Disentangling these two effects is thus difficult and suggests that there are many genomic regions, possibly with lower recombination rates, where background selection and possibly selective sweeps have accelerated the rate of lineage sorting during divergence (Rougeux et al., 2017). This strong bimodality between two classes of loci affected or not by linked selection is also captured by the distribution of F_{ST} , which shows a clear trough and then a peak around 0.15–0.2 (SI Figure 23). However, this bimodality reduces the ability to distinguish between the isolation-with-migration, the secondary contact, or the ancient migration scenarios in the more complex models (IM+2N+2m+G, AM+2N+2m+G and SC+2N+2m+G).

Considering an average DNA mutation rate of 10^{-8} /site/generation, we estimated the time for the onset of divergence between the two metapopulations to be 60 000–70 000 generations (but admittedly this could as well be 10 times greater if the mutation rate is 10 times smaller). The generation time of *I. nautiliei* is still unknown. Nevertheless, most hydrothermal species display an *r*-strategy suggesting short generation times (1-2 years) as an adaptation to the unstable and ephemeral nature of their habitat (Tyler & Young, 1999). Hence, we can suppose that the two populations started to diverge between 60 and 140 thousand ago (*kya*) for a mutation rate per site and per generation of 10^{-8} and 10 times more with a mutation rate of 10^{-9} . However, these estimates correspond to discrete non-overlapping generations and the reproduction of older cohorts may increase the equivalent generation time and, as a result, the divergence estimates. These values may be tentatively compared with estimates from the *cox1* sequences in Boulart et al. (in press) (net divergence 0.615% estimated on all sites). This latter value would amount to ~ 0.550 million years ago (*mya*) considering the widely used divergence rate of 1.4%/million years (*myr*) for mitochondrial DNA (Knowlton & Weigt, 1998), but can reach 1.2 *mya*, depending on the average mitochondrial substitution rate considered for vent species (0.2-0.3%/myr (Chevaldonné et al., 2002; Breusing et al., 2020; Castel et al. in prep.) . Although these estimates are notoriously highly variable and error-prone (see for instance Breusing et al., 2020), divergence time could range between 0.5 and 1 *mya*. This estimate is in rough agreement with Martinez & Taylor, (1996) who showed that the center of the Manus BAB started to spread quite recently (~ 0.78 *mya*), suggesting that hydrothermal vents within the spreading center may be younger than this estimate. Thus, it cannot be excluded that the divergence history of *I. nautiliei* is relatively recent and not linked to the formation of BABs, but instead to regional modifications of surface and deep-sea currents during previous glacial maxima in relation to the extension of the Antarctic ice sheet which culminated around 0.126 *mya* (Barrows et al., 2011; Joy et al., 2014).

In addition to these divergence time estimates, the models allowed us to quantify the existence of an ongoing bidirectional and asymmetrical gene flow, with migration from North-Fiji/Futuna/Lau to Manus/Woodlark being higher than in the opposite direction. Despite this slight asymmetry, a genetic influence of the Manus/Woodlark metapopulation was detected in North-Fiji, which shows foreign alleles coming from the former rather than from the Lau/Futuna populations (also observed at mtDNA *cox1* gene in Thaler et al., 2011 and Boulart et al., (in press), but not the other way around (*i.e.* influence of North-Fiji/Futuna/Lau on Woodlark, but see below). This result is consistent with the geography of the region, because North-Fiji and Woodlark are the closest BABs between the two metapopulations. Connectivity through larval dispersal between these two BAB ensembles has been tested by Mitarai et al. (2016) who simulated larval dispersal through entrainment of particles by oceanic currents prevailing at depths of 1000 m in the western Pacific. That study inferred a weak stepping-stone connection through a long planktonic larval duration (PLD of 170 days), provided that active hydrothermal sites in the Solomon and New Hebrides/Vanuatu arcs act as a relay. Such fields are known to exist, mostly associated with seamounts such as Nifonea, Tinakula or Stanton along the New Hebrides/Vanuatu arc (McConachy et al., 2005; Schmidt et al., 2017). The larval dispersal model developed by Mitarai et al. (2016) suggests a scenario where dispersal is mainly oriented from east to west: a situation also depicted in this region by Yearsley & Sigwart (2011) for a non-hydrothermal species at several depths (800 and 1400 m) and with various PLD lengths (27–151 days). However, when looking at surface countercurrents between Manus/Woodlark and North-Fiji/Futuna/Lau, Ganachaud et al. (2014) found surface currents oriented mainly from west to east through Solomon Islands and Vanuatu waters.

Cases of asymmetrical bidirectional gene flow between two metapopulations have also been found in two other hydrothermal gastropod species occurring sympatrically with *I. nautili*, *L. schrolli* (Plouviez et al., 2019) and *A. boucheti* (Breusing et al., 2021). But, in contrast to *I. nautili*, the predominant gene flow is oriented eastward from Manus to Lau, ($m_{M \rightarrow L} = 0.625$, $m_{L \rightarrow M} = 0.1725$ for *L. schrolli* and $m_{M \rightarrow L} = 12$, $m_{L \rightarrow M} = 2.6$ for *A. boucheti*).

Similarly to *I. nautili*, *L. schrolli* is considered to possess lecithotrophic larvae (Berg, 1985; Craddock, Lutz, & Vrijenhoek, 1997; Tyler et al., 2008)). As for *A. boucheti*, its larval stage remains unknown, although both its morphology (Warèn & Bouchet, 1993) and the eDNA detection of *Alviniconcha* larvae close to the surface suggest planktonotrophy (Sommer et al., 2017). Provided that our estimates reflect ongoing migration, we hypothesize that *I. nautili* larvae may be influenced by deep as well as surface currents, which could explain bidirectional

gene flow, one direction being slightly stronger than the other. This asymmetry suggests vertical migration of larvae. However, further investigations including oceanographic modeling and laboratory experiments are needed to address this hypothesis. For example, larvae of the hydrothermal gastropod *Shinkailepas myojinensis* (Yahagi et al., 2017) are able to migrate through the water column, and there is evidence of hydrothermal species' larvae in near-surface waters (Arellano et al., 2014, Sommer et al., 2017). Nevertheless, although many unknowns remain, our results indicate that *I. nautili* has a complex dispersal strategy and pattern.

Fine-scale population structure and connectivity

The high homogeneity of the two clearly distinct *I. nautili* metapopulations necessarily entails that the intra-metapopulation migration (*i.e.* inter-localities within each BAB and inter-BABs within each metapopulation) is strong or extremely recent. Moreover, no kinship-related structure was detected in the SNP dataset, indicating that there is either no self-recruitment even though females brood their larvae to the trochophoran stage, or that population sizes are so large that the probability of detecting potential kin is too small (Table 4). Consequently, genetic connectivity within each metapopulation appears to be high, with evenly distributed polymorphisms among sampled sites despite the patchy distribution of hydrothermal vents and the inter-site distances, which may vary from hundreds of meters to more than a thousand kilometers within each metapopulation. This genetic connectivity therefore suggests that *I. nautili* larvae are able to disperse within the range of each metapopulation after spawning.

The question is now whether this genetic homogeneity of each metapopulation arises from demographic connectivity (*i.e.* recruitment at one site being strongly influenced by the exportation of propagules from other sites) or is due to sporadic/rare larval exchanges able to counterbalance very attenuated genetic drift due to large local population sizes. The mechanism behind the observed genetic homogeneity has strong implications in terms of conservation biology, because demographic connectivity can play a crucial role in the resilience of populations faced with local extinction potentially exacerbated by deep-sea mining. The global analysis relying on a panel of primarily neutral markers indicates no differentiation at the metapopulation scale, but—as advocated by Gagnaire et al., (2015)—a few loci markers potentially undergoing direct, or indirect selective pressures may locally harbor distinct allele frequency in the recipient population. This pattern can be explained by local selection for foreign alleles that are better adapted or less loaded by deleterious mutations than resident ones, or by resolving intrinsic asymmetrical incompatibilities between divergent genomes (Simon et

al., 2021) creating local soft sweeps through linked selection. These processes result in enhanced local introgression of certain marker loci, a common pattern observed in blue mussels (Fraïsse et al., 2016) or European sea bass (Robinet et al., 2020), for example. These markers will appear as F_{ST} outliers that may indicate recent dispersal events.

Our outlier analyses indeed suggest introgression of some loci. In Figure 6 B and SI Figure 18 B, individuals from the North-Fiji basin seem to be closer to Manus/Woodlark than Lau/Futuna on PC1, which may correspond to introgression in some of the outlier loci. An introgression pattern was confirmed by the F_3 tests performed with the North-Fiji basin as the focal populations (significant negative value of the F_3 statistic, SI Figure 13) and the ADMIXTURE analyses (Figure 6 C). These results indicate that some alleles found at high frequency in North-Fiji individuals are the consequence of long-range migration from Manus/Woodlark. Interestingly, with Manus/Woodlark outliers, although not visually detectable on the PCA (Figure 6 A and 18 A), Woodlark individuals exhibit some level of admixture from the North-Fiji/Futuna/Lau metapopulation (Figure 6 D). This low-level admixture corroborates our inference of ongoing bidirectional gene flow. However, it is not yet clear as to why its impact appears stronger in populations of the North-Fiji basin, against the predominant direction according to our $\hat{d}a\hat{d}i$ inferences. Although we are unable to determine the precise mechanisms behind these frequency changes, these alleles have not diffused from North-Fiji to Lau/Futuna, indicating a subtle—but real—limitation in connectivity between the former and the latter. The same reasoning applies for the traces of admixture detected in Woodlark that appear to have not diffused to Manus.

Another kind of differentiation depicted by outlier loci seems to be explained by intra-metapopulation divergence. The question arises as to the origin of these slight divergences on a PC axis orthogonal to the main inter-metapopulation divergence, which does not necessarily proceed from gene flow between differentiated populations as described above. Allele frequency differences for outlier loci between Manus and Woodlark are detectable on PC2 (Figure 6 A, SI Figure 18 A, ADMIXTURE $K = 3$, SI Figure 20). The same question applies to the eastern North-Fiji/Futuna/Lau metapopulation, with differences between North-Fiji and Futuna/Lau (Figure 6 B, ADMIXTURE $K = 3$, SI Figure 21). This pattern can be due to any combination of drift and/or selection. Local selection may result from major differences in depth or vent fluid composition. The fact that the fraction of outliers mapping on transcribed regions is greater than by chance and targets a few metabolic/regulatory pathways suggests their possible involvement in local adaptation to depth or different fluid chemistry, but this remains

to be investigated (Jennings et al., 2013; Liu et al., 2021). In the absence of high demographic connectivity required to ensure the interdependency of local populations, this local differentiation can remain detectable for several generations before being shuffled among all metapopulation demes.

These subtle limitations in connectivity between basins can be associated with abyssal plains, which may limit gene flow in a disconnected ridge system such as that found at the regional scale of these BABs. Physical barriers in other parts of the world, such as transform faults and microplates, have already been shown to greatly impede the effective migration of deep-sea vent species at a much more restricted spatial scale (Johnson et al., 2006; Plouviez et al., 2009; Plouviez, et al., 2013). However, regarding the populations of the Futuna volcanic arc and Lau basin, our in-depth scrutiny of outliers did not reveal any sign of genetic differentiation. Hence, the hypothesis of demographic correlation between these two regions cannot be rejected, although we cannot infer with certitude the directionality of the exchanges.

Conclusions

Overall, our analyses revealed a clear genetic differentiation of *Ifremeria nautiliei* populations between the Manus/Woodlark and the North-Fiji/Futuna/Lau BABs, with very high gene flow within each of these two metapopulations as well as higher genetic diversity in Manus/Woodlark. Despite an in-depth scrutiny of genome-wide genetic variation, no geographic substructure was detected between or within localities sampled within each individual ridge system. This genetic connectivity probably indicates high local (re)colonization capacity for this hydrothermal vent species due to the ephemeral nature of active sites in this region, at least at the scale of a given back-arc basin.

However, our outlier analyses revealed that this genetic connectivity does not necessarily equate with demographic connectivity at the larger inter-basin intra-metapopulation scale. The specific investigation of outlier loci illustrates how a few loci in a large genome-wide dataset can carry useful information about actual barriers to dispersal in high gene flow species. Deep-sea mining holds the potential to exacerbate dispersal barriers and limit population resilience, because if a large proportion of the vent habitat is destroyed locally, population rescue from other basins will be restricted.

Furthermore, our demographic simulations indicated a long period of divergence during the Quaternary period (several tens of thousands of generations) associated with restricted long-range gene flow over a large fraction of the genome. Although the effects of linked selection and reduced migration (barrier loci) are not clearly distinguishable, our results suggest that the effect of the latter is less pronounced. This interpretation agrees with the fact that the global divergence among the two metapopulations is still quite low (net nuclear nucleotide divergence, 0.81%). This divergence perhaps reflects the very beginning of an ongoing speciation process, where a handful of barrier loci may already exist and at the same time overall genetic differentiation is not hampered by weak contemporary and asymmetrical gene flow between metapopulations.

Acknowledgments

We are deeply grateful to the R/V *L'Atalante* crew as well as the ROV *Victor 6000* crew during the ChuBacArc 2019 cruise without whom nothing would have been possible. Many thanks also to Cindy L. Van Dover for sharing some Manus 2008 samples and to John Parianos and Paul Lahari from Nautilus Minerals group for sharing information on their Solwara mining prospects in the Manus basin. W. We also thank Gwenn Tanguy for advice and, and access to the Biogenouest Genomer Platform at the Station Biologique de Roscoff. Data were stored and analyzed at the Biogenouest AbiMS Bioinformatics platform, which provided data storage and computing resources, as well as at the Montpellier Bioinformatics Biodiversity platform (LabEx CeMEB). We also thank Khalid Belkhir and Christelle Fraïsse for their debugged version of $\partial a \partial i$. We are most grateful to four anonymous reviewers whose detailed and insightful comments significantly improved the quality and clarity of the manuscript.

Ship time was supported by the French Oceanographic Fleet program (CHUBACARC cruise doi [10.17600/18001111](https://doi.org/10.17600/18001111) to D. J. and S. H.). Travel expenses of ChuBacArc participants and logistic expenses were funded by CNRS-Institut Ecologie et Environnement. Lab work and A. T. L. Y. PhD fellowship were funded through the Agence Nationale de la Recherche ANR 'CERBERUS' (contract number ANR-17-CE02-0003 to S. H.). Sequencing was integrated in the eDNAbyss project (contract AP2016-228 to S. A. H.) funded by France Génomique (ANR-10-INBS-09) and Genoscope-CEA.

References

- Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, 12(1), 246. <https://doi.org/10.1186/1471-2105-12-246>
- Arellano, S. M., Van Gaest, A. L., Johnson, S. B., Vrijenhoek, R. C., & Young, C. M. (2014). Larvae from deep-sea methane seeps disperse in surface waters. *Proceedings of the Royal Society B: Biological Sciences*, 281(1786), 20133276. <https://doi.org/10.1098/rspb.2013.3276>
- Audzijonyte, A., & Vrijenhoek, R. C. (2010). When Gaps Really Are Gaps :Statistical phylogeography of hydrothermal vent invertebrates. *Evolution*, 64(8), 2369-2384. <https://doi.org/10.1111/j.1558-5646.2010.00987.x>
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes : A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19(5), 455-477. <https://doi.org/10.1089/cmb.2012.0021>
- Barrows, T. T., Hope, G. S., Prentice, M. L., Fifield, L. K., & Tims, S. G. (2011). Late Pleistocene glaciation of the Mt Giluwe volcano, Papua New Guinea. *Quaternary Science Reviews*, 30(19-20), 2676-2689. <https://doi.org/10.1016/j.quascirev.2011.05.022>
- Berg, C. J. (1985). Reproductive strategies of mollusks from abyssal hydrothermal vent communities. *Bulletin of The Biological Society of Washington*, 185-197.
- Boulart, C., Rouxel, O., Scalabrin, C., Le Meur, P., Pelleter, E., Poitrimol, C., Thiébaud, E., Matabos, M., Castel, J., Tran Lu Y, A., Michel, L.N., Cathalot, C., Cheron, S., Boissier, A., Germain, Y., Arnaud-Haond, S., Bonhomme, F., Broquet, T., Cuff-Gauchard, V., Le Layec, V., L'Haridon, S, Mary, J., Le Port, A-S., Tasiemski, A., Kuama, D.C., Hourdez, S., Jollivet, D. (in press). Active hydrothermal vents in the Woodlark Basin may act as dispersing centres for hydrothermal fauna *Communications Earth & Environment*.
- Borowski, C., Giere, O., Krieger, J., Amann, R., & Dubilier, N. (2002). New aspects of the symbiosis in the provannid snail *Ifremeria nautilei* from the North Fiji Back Arc Basin. *Cahiers de Biologie Marine*, 5. https://www.researchgate.net/publication/27262835_New_aspects_of_the_symbiosis_in_the_provannid_snail>Ifremeria_nautilei_from_the_North_Fiji_Back_Arc_Basin
- Brelsford, A., Dufresnes, C., & Perrin, N. (2016). High-density sex-specific linkage maps of a European tree frog (*Hyla arborea*) identify the sex chromosome without information on offspring sex. *Heredity*, 116(2), 177-181. <https://doi.org/10.1038/hdy.2015.83>
- Breusing, C., Biastoch, A., Drews, A., Metaxas, A., Jollivet, D., Vrijenhoek, R. C., Bayer, T., Melzner, F., Sayavedra, L., Petersen, J. M., Dubilier, N., Schilhabel, M. B., Rosenstiel, P., & Reusch, T. B. H. (2016). Biophysical and population genetic models predict the presence of “phantom” stepping stones connecting Mid-Atlantic Ridge vent ecosystems. *Current Biology*, 26(17), 2257-2267. <https://doi.org/10.1016/j.cub.2016.06.062>
- Breusing, C., Johnson, S. B., Tunnicliffe, V., Clague, D. A., Vrijenhoek, R. C., & Beinart, R. A. (2020). Allopatric and sympatric drivers of speciation in alviniconcha hydrothermal vent snails. *Molecular Biology and Evolution*, 37(12), 3469-3484. <https://doi.org/10.1093/molbev/msaa177>
- Castel, J., Hourdez, S., Pradillon, F., Daguin-Thiébaud, C., Ballenghien, M., Ruault, S., Corre, E., Tran Lu Y, A., Mary, J., Comtet, T., Gagnaire, P-A., Bonhomme, F., Bierne, N., Breusing, C., Broquet, T., & Jollivet D. (2022): A story of divergence and inter-specific exchanges in deep-sea hydrothermal vent gastropods *Alviniconcha*. *Genes*.Manuscript submitted for publication.
- Cayuela, H., Rougemont, Q., Prunier, J. G., Moore, J.-S., Clobert, J., Besnard, A., & Bernatchez, L. (2018). Demographic and genetic approaches to study dispersal in wild animal populations: A methodological review. *Molecular Ecology*, 27(20), 3976-4010. <https://doi.org/10.1111/mec.14848>

- Chevaldonne, P., Jollivet, D., Desbruyeres, D., Lutz, R., & Vrijenhoek, R. (2002). Sister-species of eastern Pacific hydrothermal vent worms (Ampharetidae, Alvinellidae, Vestimentifera) provide new mitochondrial COI clock calibration. *CBM - Cahiers de Biologie Marine*, 43(3-4), 367-370. <https://archimer.ifremer.fr/doc/00000/895/>
- Chevaldonne, P., Jollivet, D., Vangriesheim, A., & Desbruyères, D. (1997). Hydrothermal-vent alvinellid polychaete dispersal in the eastern Pacific. 1. Influence of vent site distribution, bottom currents, and biological patterns. *Limnology and Oceanography*, 42(1), 67-80. <https://doi.org/10.4319/lo.1997.42.1.0067>
- Craddock, C., Hoeh, W. R., Gustafson, R. G., Lutz, R. A., Hashimoto, J., & Vrijenhoek, R. J. (1995). Evolutionary relationships among deep-sea mytilids (Bivalvia : Mytilidae) from hydrothermal vents and cold-water methane/sulfide seeps. *Marine Biology*, 121(3), 477-485. <https://doi.org/10.1007/BF00349456>
- Craddock, C., Lutz, R. A., & Vrijenhoek, R. C. (1997). Patterns of dispersal and larval development of archaeogastropod limpets at hydrothermal vents in the eastern Pacific. *Journal of Experimental Marine Biology and Ecology*, 210(1), 37-51. [https://doi.org/10.1016/S0022-0981\(96\)02701-3](https://doi.org/10.1016/S0022-0981(96)02701-3)
- Daguin-Thiébaud, C., Ruault, S., Roby, C., Broquet, T., Viard, F., & Brelsford, A. (2021, septembre 21). *Construction of individual ddRAD libraries*. Protocols.io. <https://www.protocols.io/view/construction-of-individual-ddrad-libraries-bv4tn8wn>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156-2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP Data. *PLOS Genetics*, 9(10), e1003905. <https://doi.org/10.1371/journal.pgen.1003905>
- Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5 : A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), 564-567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
- Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, 131(2), 479-491. <https://www.genetics.org/content/131/2/479>
- Feutry, P., Devloo-Delva, F., Y, A. T. L., Mona, S., Gunasekera, R. M., Johnson, G., Pillans, R. D., Jaccoud, D., Kilian, A., Morgan, D. L., Saunders, T., Bax, N. J., & Kyne, P. M. (2020). One panel to rule them all: DArTcap genotyping for population structure, historical demography, and kinship analyses, and its application to a threatened shark. *Molecular Ecology Resources*, 20(6), 1470-1485. <https://doi.org/10.1111/1755-0998.13204>
- Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A bayesian perspective. *Genetics*, 180(2), 977-993. <https://doi.org/10.1534/genetics.108.092221>
- Fraïsse, C., Belkhir, K., Welch, J. J., & Bierne, N. (2016). Local interspecies introgression is the main cause of extreme levels of intraspecific differentiation in mussels. *Molecular Ecology*, 25(1), 269-286. <https://doi.org/10.1111/mec.13299>
- Gagnaire, P.-A., Broquet, T., Aurelle, D., Viard, F., Souissi, A., Bonhomme, F., Arnaud-Haond, S., & Bierne, N. (2015). Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evolutionary Applications*, 8(8), 769-786. <https://doi.org/10.1111/eva.12288>
- Ganachaud, A., Cravatte, S., Melet, A., Schiller, A., Holbrook, N. J., Sloyan, B. M., Widlansky, M. J., Bowen, M., Verron, J., Wiles, P., Ridgway, K., Sutton, P., Sprintall, J., Steinberg, C., Brassington, G., Cai, W., Davis, R., Gasparin, F., Gourdeau, L., ... Send, U. (2014). The southwest Pacific Ocean circulation and climate experiment (SPICE). *Journal of Geophysical Research: Oceans*, 119(11), 7660-7686. <https://doi.org/10.1002/2013JC009678>
- Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*, 201(4), 1555-1579. <https://doi.org/10.1534/genetics.115.181453>

- Gena, K. (2013). Deep sea mining of submarine hydrothermal deposits and its possible environmental impact in Manus basin, Papua New Guinea. *Procedia Earth and Planetary Science*, 6, 226-233. <https://doi.org/10.1016/j.proeps.2013.01.031>
- Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLOS Genetics*, 5(10), e1000695. <https://doi.org/10.1371/journal.pgen.1000695>
- Hannington, M., Jamieson, J., Monecke, T., Petersen, S., & Beaulieu, S. (2011). The abundance of seafloor massive sulfide deposits. *Geology*, 39(12), 1155-1158. <https://doi.org/10.1130/G32468.1>
- Hurtado, L. A., Lutz, R. A., & Vrijenhoek, R. C. (2004). Distinct patterns of genetic differentiation among annelids of eastern Pacific hydrothermal vents. *Molecular Ecology*, 13(9), 2603-2615. <https://doi.org/10.1111/j.1365-294X.2004.02287.x>
- Jennings, R. M., Etter, R. J., & Ficarra, L. (2013). Population differentiation and species formation in the deep sea: The potential role of environmental gradients and depth. *PLOS ONE*, 8(10), e77594. <https://doi.org/10.1371/journal.pone.0077594>
- Johnson, S. B., Young, C. R., Jones, W. J., Warén, A., & Vrijenhoek, R. C. (2006). Migration, isolation, and speciation of hydrothermal vent limpets (Gastropoda; Lepetodrilidae) across the Blanco transform fault. *The Biological Bulletin*, 210(2), 140-157. <https://doi.org/10.2307/4134603>
- Jollivet, D., Chevaldonne, P., & Planque, B. (1999). Hydrothermal-vent alvinellid polychaete dispersal in the eastern Pacific. 2. A metapopulation model based on habitat shifts. *Evolution*, 53(4), 1128-1142. <https://doi.org/10.1111/j.1558-5646.1999.tb04527.x>
- Jollivet, D., Desbruyères, D., Bonhomme, F., & Moraga, D. (1995). Genetic differentiation of deep-sea hydrothermal vent alvinellid populations (Annelida : Polychaeta) along the East Pacific Rise. *Heredity*, 74(4), 376-391. <https://doi.org/10.1038/hdy.1995.56>
- Jones, G. P., Planes, S., & Thorrold, S. R. (2005). Coral Reef Fish Larvae settle close to home. *Current Biology*, 15(14), 1314-1318. <https://doi.org/10.1016/j.cub.2005.06.061>
- Joy, K., Fink, D., Storey, B., & Atkins, C. (2014). A 2-million-year glacial chronology of the Hatherton Glacier, Antarctica and implications for the size of the East Antarctic Ice Sheet at the Last Glacial Maximum. *Quaternary Science Reviews*, 83, 46-57. <https://doi.org/10.1016/j.quascirev.2013.10.028>
- Knowlton, N., & Weigt, L. A. (1998). New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1412), 2257-2263. <https://doi.org/10.1098/rspb.1998.0568>
- Konn, C., Furré, E., Jean-Baptiste, P., Donval, J. P., Guyader, V., Birot, D., Alix, A. S., Gaillot, A., Perez, F., Dapoigny, A., Pelleter, E., Resing, J. A., Charlou, J. L., & Fouquet, Y. (2016). Extensive hydrothermal activity revealed by multi-tracer survey in the Wallis and Futuna region (SW Pacific). *Deep Sea Research Part I: Oceanographic Research Papers*, 116, 127-144. <https://doi.org/10.1016/j.dsr.2016.07.012>
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak : A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5), 1179-1191. <https://doi.org/10.1111/1755-0998.12387>
- Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider : An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28(2), 298-299. <https://doi.org/10.1093/bioinformatics/btr642>
- Liu, R., Liu, J., & Zhang, H. (2021). Positive selection analysis reveals the deep-sea adaptation of a hadal sea cucumber (*Paelopatides* sp.) to the Mariana Trench. *Journal of Oceanology and Limnology*, 39(1), 266-281. <https://doi.org/10.1007/s00343-020-0241-0>
- Lowe, W. H., & Allendorf, F. W. (2010). What can genetics tell us about population connectivity? *Molecular Ecology*, 19(15), 3038-3051. <https://doi.org/10.1111/j.1365-294X.2010.04688.x>
- Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt : An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, 17(1), 67-77. <https://doi.org/10.1111/1755-0998.12592>
- Lynch, M. (2010). Evolution of the mutation rate. *Trends in Genetics*, 26(8), 345-352. <https://doi.org/10.1016/j.tig.2010.05.003>

- Martinez, F., & Taylor, B. (1996). Back arc spreading, rifting, and microplate rotation, between transform faults in the Manus Basin. *Marine Geophysical Researches*, 18(2-4), 203-224. <https://doi.org/10.1007/BF00286078>
- Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, 15(1), 28-41. <https://doi.org/10.1111/1755-0998.12291>
- McConachy, T. F., Arculus, R. J., Yeats, C. J., Binns, R. A., Barriga, F. J. A. S., McInnes, B. I. A., Sestak, S., Sharpe, R., Rakau, B., & Tevi, T. (2005). New hydrothermal activity and alkalic volcanism in the back arc Coriolis Troughs, Vanuatu. *Geology*, 33(1), 61-64. <https://doi.org/10.1130/G20870.1>
- Milano, I., Babbucci, M., Cariani, A., Atanassova, M., Bekkevold, D., Carvalho, G. R., Espiñeira, M., Fiorentino, F., Garofalo, G., Geffen, A. J., Hansen, Jakob. H., Helyar, S. J., Nielsen, E. E., Ogden, R., Patarnello, T., Stagioni, M., Consortium, F., Tinti, F., & Bargelloni, L. (2014). Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). *Molecular Ecology*, 23(1), 118-135. <https://doi.org/10.1111/mec.12568>
- Mitarai, S., Watanabe, H., Nakajima, Y., Shchepetkin, A. F., & McWilliams, J. C. (2016). Quantifying dispersal from hydrothermal vent fields in the western Pacific Ocean. *Proceedings of the National Academy of Sciences*, 113(11), 2976-2981. <https://doi.org/10.1073/pnas.1518395113>
- Mullineaux, L. S., Adams, D. K., Mills, S. W., & Beaulieu, S. E. (2010). Larvae from afar colonize deep-sea hydrothermal vents after a catastrophic eruption. *Proceedings of the National Academy of Sciences*, 107(17), 7829-7834. <https://doi.org/10.1073/pnas.0913187107>
- Niner, H. J., Ardron, J. A., Escobar, E. G., Gianni, M., Jaekel, A., Jones, D. O. B., Levin, L. A., Smith, C. R., Thiele, T., Turner, P. J., Van Dover, C. L., Watling, L., & Gjerde, K. M. (2018). Deep-sea mining with no net loss of biodiversity—An Impossible Aim. *Frontiers in Marine Science*, 5. <https://www.frontiersin.org/article/10.3389/fmars.2018.00053>
- Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space : A road map for stacks. *Methods in Ecology and Evolution*, 8(10), 1360-1373. <https://doi.org/10.1111/2041-210X.12775>
- Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genetics*, 8(11), e1002967. <https://doi.org/10.1371/journal.pgen.1002967>
- Pinsky, M. L., Jr, H. R. M., & Palumbi, S. R. (2010). Using isolation by distance and effective density to estimate dispersal scales in anemonefish. *Evolution*, 64(9), 2688-2700. <https://doi.org/10.1111/j.1558-5646.2010.01003.x>
- Plouviez, S., Faure, B., Guen, D. L., Lallier, F. H., Bierne, N., & Jollivet, D. (2013). A new barrier to dispersal trapped old genetic clines that escaped the easter microplate tension zone of the pacific vent mussels. *PLOS ONE*, 8(12), e81555. <https://doi.org/10.1371/journal.pone.0081555>
- Plouviez, S., LaBella, A. L., Weisrock, D. W., Meijenfildt, F. A. B. von, Ball, B., Neigel, J. E., & Dover, C. L. V. (2019). Amplicon sequencing of 42 nuclear loci supports directional gene flow between south pacific populations of a hydrothermal vent limpet. *Ecology and Evolution*, 9(11), 6568-6580. <https://doi.org/10.1002/ece3.5235>
- Plouviez, S., Schultz, T. F., McGinnis, G., Minshall, H., Rudder, M., & Van Dover, C. L. (2013). Genetic diversity of hydrothermal-vent barnacles in Manus Basin. *Deep Sea Research Part I: Oceanographic Research Papers*, 82, 73-79. <https://doi.org/10.1016/j.dsr.2013.08.004>
- Plouviez, S., Shank, T. M., Faure, B., Daguin-Thiebaut, C., Viard, F., Lallier, F. H., & Jollivet, D. (2009). Comparative phylogeography among hydrothermal vent species along the East Pacific Rise reveals vicariant processes and population expansion in the South. *Molecular Ecology*, 18(18), 3903-3917. <https://doi.org/10.1111/j.1365-294X.2009.04325.x>
- Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, 461(7263), 489-494. <https://doi.org/10.1038/nature08365>
- Reynolds, K. C., Watanabe, H., Strong, E. E., Sasaki, T., Uematsu, K., Miyake, H., Kojima, S., Suzuki, Y., Fujikura, K., Kim, S., & Young, C. M. (2010). New molluscan larval form :

- Brooding and development in a hydrothermal vent gastropod, *Ifremeria nautilei* (Provannidae). *The Biological Bulletin*, 219(1), 7-11. <https://doi.org/10.1086/BBLv219n1p7>
- Robinet, T., Roussel, V., Cheze, K., & Gagnaire, P.-A. (2020). Spatial gradients of introgressed ancestry reveal cryptic connectivity patterns in a high gene flow marine fish. *Molecular Ecology*, 29(20), 3857-3871. <https://doi.org/10.1111/mec.15611>
- Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2 : Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, 28(21), 4737-4754. <https://doi.org/10.1111/mec.15253>
- Rougeux, C., Bernatchez, L., & Gagnaire, P.-A. (2017). Modeling the multiple facets of speciation-with-gene-flow toward inferring the divergence history of lake whitefish species pairs (*Coregonus clupeaformis*). *Genome Biology and Evolution*, 9(8), 2057-2074. <https://doi.org/10.1093/gbe/evx150>
- Schmidt, K., Garbe-Schönberg, D., Hannington, M. D., Anderson, M. O., Bühring, B., Haase, K., Haruel, C., Lupton, J., & Koschinsky, A. (2017). Boiling vapour-type fluids from the Nifonea vent field (New Hebrides Back-Arc, Vanuatu, SW Pacific): Geochemistry of an early-stage, post-eruptive hydrothermal system. *Geochimica et Cosmochimica Acta*, 207, 185-209. <https://doi.org/10.1016/j.gca.2017.03.016>
- Simon, A., Fraïsse, C., Ayari, T. E., Liautard-Haag, C., Strelkov, P., Welch, J. J., & Bierne, N. (2021). How do species barriers decay? Concordance and local introgression in mosaic hybrid zones of mussels. *Journal of Evolutionary Biology*, 34(1), 208-223. <https://doi.org/10.1111/jeb.13709>
- Sommer, S. A., Van Woudenberg, L., Lenz, P. H., Cepeda, G., & Goetze, E. (2017). Vertical gradients in species richness and community composition across the twilight zone in the North Pacific Subtropical Gyre. *Molecular Ecology*, 26(21), 6136-6156. <https://doi.org/10.1111/mec.14286>
- Teixeira, S., Cambon-Bonavita, M.-A., Serrão, E. A., Desbruyères, D., & Arnaud-Haond, S. (2011). Recent population expansion and connectivity in the hydrothermal shrimp *Rimicaris exoculata* along the Mid-Atlantic Ridge. *Journal of Biogeography*, 38(3), 564-574. <https://doi.org/10.1111/j.1365-2699.2010.02408.x>
- Teixeira, S., Serrão, E. A., & Arnaud-Haond, S. (2012). Panmixia in a Fragmented and Unstable Environment: The Hydrothermal Shrimp *Rimicaris exoculata* Disperses Extensively along the Mid-Atlantic Ridge. *PLOS ONE*, 7(6), e38521. <https://doi.org/10.1371/journal.pone.0038521>
- Thaler, A. D., Plouviez, S., Saleu, W., Alei, F., Jacobson, A., Boyle, E. A., Schultz, T. F., Carlsson, J., & Dover, C. L. V. (2014). Comparative population structure of two deep-sea hydrothermal-vent-associated decapods (*Chorocaris sp. 2* and *Munidopsis lauensis*) from Southwestern Pacific Back-Arc Basins. *PLOS ONE*, 9(7), e101345. <https://doi.org/10.1371/journal.pone.0101345>
- Thaler, A. D., Zelnio, K., Saleu, W., Schultz, T. F., Carlsson, J., Cunningham, C., Vrijenhoek, R. C., & Van Dover, C. L. (2011). The spatial scale of genetic subdivision in populations of *Ifremeria nautilei*, a hydrothermal-vent gastropod from the southwest Pacific. *BMC Evolutionary Biology*, 11(1), 372. <https://doi.org/10.1186/1471-2148-11-372>
- Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R. S. T., Hecht, J., Knaust, F., Belkhir, K., Klages, S., Dieterich, R., Stueber, K., Piferrer, F., Guinand, B., Bierne, N., Volckaert, F. A. M., Bargelloni, L., Power, D. M., Bonhomme, F., ... Reinhardt, R. (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications*, 5(1), 5770. <https://doi.org/10.1038/ncomms6770>
- Tunnicliffe, V., McArthur, A. G., & McHugh, D. (1998). A biogeographical perspective of the deep-sea hydrothermal vent fauna. Dans J. H. S. Blaxter, A. J. Southward, & P. A. Tyler (Éds.), *Advances in Marine Biology* (Vol. 34, p. 353-442). Academic Press. [https://doi.org/10.1016/S0065-2881\(08\)60213-8](https://doi.org/10.1016/S0065-2881(08)60213-8)
- Tyler, P. A., Pendlebury, S., Mills, S. W., Mullineaux, L., Eckelbarger, K. J., Baker, M., & Young, C. M. (2008). Reproduction of gastropods from vents on the East Pacific Rise and the Mid-Atlantic Ridge. *Journal of Shellfish Research*, 27(1), 107-118. [https://doi.org/10.2983/0730-8000\(2008\)27\[107:ROGFVO\]2.0.CO;2](https://doi.org/10.2983/0730-8000(2008)27[107:ROGFVO]2.0.CO;2)

- Tyler, P. A., & Young, C. M. (1999). Reproduction and dispersal at vents and cold seeps. *Journal of the Marine Biological Association of the United Kingdom*, 79(2), 193-208. <https://doi.org/10.1017/S0025315499000235>
- Villemereuil, P. de, Frichot, É., Bazin, É., François, O., & Gaggiotti, O. E. (2014). Genome scan methods against more complex models: When and how much should we trust them? *Molecular Ecology*, 23(8), 2006-2019. <https://doi.org/10.1111/mec.12705>
- Van Dover, C. L., Humphris, S. E., Fornari, D., Cavanaugh, C. M., Collier, R., Goffredi, S. K., Hashimoto, J., Lilley, M. D., Reysenbach, A. L., Shank, T. M., Damm, K. L. V., Banta, A., Gallant, R. M., Götz, D., Green, D., Hall, J., Harmer, T. L., Hurtado, L. A., Johnson, P., ... Vrijenhoek, R. C. (2001). Biogeography and ecological setting of Indian ocean hydrothermal vents. *Science*, 294(5543), 818-823. <https://doi.org/10.1126/science.1064574>
- Vrijenhoek, R. C. (1997). Gene flow and genetic diversity in naturally fragmented metapopulations of deep-sea hydrothermal vent animals. *Journal of Heredity*, 88(4), 285-293. <https://doi.org/10.1093/oxfordjournals.jhered.a023106>
- Vrijenhoek, R. C. (2010). Genetic diversity and connectivity of deep-sea hydrothermal vent metapopulations. *Molecular Ecology*, 19(20), 4391-4411. <https://doi.org/10.1111/j.1365-294X.2010.04789.x>
- Warèn, A., & Bouchet, P. (1993). New records, species, genera, and a new family of gastropods from hydrothermal vents and hydrocarbon seeps*. *Zoologica Scripta*, 22(1), 1-90. <https://doi.org/10.1111/j.1463-6409.1993.tb00342.x>
- Windoffer, R., & Giere, O. (1997). Symbiosis of the hydrothermal vent gastropod *Ifremeria nautilei* (Provannidae) with endobacteria-structural analyses and ecological considerations. *The Biological Bulletin*, 193(3), 381-392. <https://doi.org/10.2307/1542940>
- Won, Y., Hallam, S. J., O'Mullan, G. D., & Vrijenhoek, R. C. (2003). Cytonuclear disequilibrium in a hybrid zone involving deep-sea hydrothermal vent mussels of the genus *Bathymodiolus*. *Molecular Ecology*, 12(11), 3185-3190. <https://doi.org/10.1046/j.1365-294X.2003.01974.x>
- Wyngaarden, M. V., Snelgrove, P. V. R., DiBacco, C., Hamilton, L. C., Rodríguez-Ezpeleta, N., Jeffery, N. W., Stanley, R. R. E., & Bradbury, I. R. (2017). Identifying patterns of dispersal, connectivity and selection in the sea scallop, *Placopecten magellanicus*, using RADseq-derived SNPs. *Evolutionary Applications*, 10(1), 102-117. <https://doi.org/10.1111/eva.12432>
- Yahagi, T., Fukumori, H., Warén, A., & Kano, Y. (2019). Population connectivity of hydrothermal-vent limpets along the northern Mid-Atlantic Ridge (Gastropoda : Neritimorpha: Phenacolepadidae). *Journal of the Marine Biological Association of the United Kingdom*, 99(1), 179-185. <https://doi.org/10.1017/S0025315417001898>
- Yahagi, T., Thaler, A. D., Dover, C. L. V., & Kano, Y. (2020). Population connectivity of the hydrothermal-vent limpet *Shinkailepas tollmanni* in the Southwest Pacific (Gastropoda : Neritimorpha: Phenacolepadidae). *PLOS ONE*, 15(9), e0239784. <https://doi.org/10.1371/journal.pone.0239784>
- Yahagi, T., Watanabe, H. K., Kojima, S., & Kano, Y. (2017). Do larvae from deep-sea hydrothermal vents disperse in surface waters? *Ecology*, 98(6), 1524-1534. <https://doi.org/10.1002/ecy.1800>
- Yearsley, J. M., & Sigwart, J. D. (2011). Larval transport modeling of deep-sea invertebrates can aid the search for undiscovered populations. *PLOS ONE*, 6(8), e23063. <https://doi.org/10.1371/journal.pone.0023063>
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28(24), 3326-3328. <https://doi.org/10.1093/bioinformatics/bts606>

Data Accessibility and Benefit-Sharing

Individual fastq files are available at the European Nucleotide Archive (study accession number PRJEB47533). SNP data (VCF) and associated metadata are available at Dryad : <https://doi.org/10.5061/dryad.ffbg79cwq> (preview https://datadryad.org/stash/share/WfF-hiYEO6nnKdwjG_E78-dA67mmAgnJ4i4Q6lL9-JU) Scripts (R, *daði*) are available in a public Github repository (<https://github.com/Atranluy/Scripts-Ifremeria>).

We contacted the authorities of the different countries (Papua-New Guinea, Fiji, and Tonga) and territories (Wallis and Futuna) for benefit-sharing where sampling was performed. The data generated will be accessible on public databases. The results obtained will also be communicated to these authorities which may have to make decision regarding conservation of deep-sea hydrothermal vent communities in their EEZs. Observers for the different countries who took part of the on-board activities will be informed of our findings.

Author's contribution:

D. J. and S. H. designed the CHUBACARC and CERBERUS projects, F. B. supervised the genetic work. A. T. L. Y., S. R., C. D. T., J. C., P. W. and A. P. performed laboratory work. A. T. L. Y. performed bioinformatics statistical analyses with the contribution of F. B., D. J., P. A. G., N. B. and T. B. A. T. L. Y., F. B. wrote the manuscript with feedback of T. B., D. J., P. A. G., N. B., S. A. H. and C. D. T. All authors approved the manuscript.

IV. CONCLUSIONS ET PERSPECTIVES

Nous avons exploré l'histoire évolutive des populations d'*I. nautiliei* et mis en lumière que cette espèce a divergé en deux métapopulations tout en affichant du flux génique faible mais toujours existant. Nous avons aussi en partie permis de révéler des subtiles limites de connectivité entre bassins au sein de chaque métapopulation.

Néanmoins, un certain nombre de points peuvent faire l'objet de futures perspectives d'études et de discussion. En effet, bien que ces approches par représentation réduite du génome permettent d'élucider beaucoup de points majeurs concernant l'histoire évolutive (Gagnaire, 2020), une approche complémentaire par utilisation d'un génome complet serait possiblement une prochaine étape et pourrait aider à résoudre un certain nombre de points. L'accès à quelques génomes d'individus de chaque métapopulation et d'individus apparentés (première voire deuxième génération) permettrait d'avoir accès au taux de mutation moyen du génome par site et par génération, en plus d'obtenir des informations concernant l'architecture génomique (haplotypes phasés, synténie, carte de recombinaison). En complément à cela, l'utilisation d'un ou deux groupes externes (espèces sœurs) permettrait d'identifier l'état ancestral des allèles et donc d'utiliser le spectre déplié des fréquences alléliques qui permet d'améliorer les inférences démographiques.

Ces quelques points d'amélioration pourraient d'une part nous aider à distinguer les différents modèles de divergence (IM, AM, SC), mais aussi mieux estimer la temporalité de la divergence qui est dépendante d'une part du taux de mutation et d'autre part du temps de génération des espèces. Dans notre cas, l'incertitude sur ces deux paramètres ne nous permettait pas réellement de trancher sur la temporalité (70 000 ou 700 000 générations) mais avec un des paramètres correctement fixés, nous pourrions potentiellement mieux comprendre cette temporalité. Ces limitations pointent aussi vers des lacunes importantes de connaissances sur les espèces profondes et leur biologie de façon générale : temps de génération moyen, âge de première reproduction, recouvrement de générations, écologie larvaire (déplacement vertical, durée de vie larvaire, ...).

Dans notre approche de détection de locus « outliers », nous n'avons exploré qu'au sein des métapopulations sans regarder le cas de ceux qui se retrouvent entre ces deux dernières. Bien

que les méthodes utilisées et leurs modèles sous-jacents prennent en compte la structuration des populations, on peut s'attendre à un excès de faux positifs et donc de locus aberrants dus à l'effet de la structuration et des modalités de l'histoire évolutive associée. Ainsi, il serait nécessaire d'explorer ce cas-ci avec des approches de contrôle du taux de faux positif (q -value).

Enfin, nous nous sommes principalement intéressés à la structuration et à l'histoire évolutive à l'échelle de l'aire de répartition de cette espèce. Or on a observé qu'il y avait une forte différenciation entre les deux métapopulations, il est donc tout à fait possible qu'une partie du polymorphisme privé à chaque métapopulation n'ait pas été échantillonnée lors de l'assemblage final des données et au filtre qui nécessite que les locus soient présents chez au moins 80% des individus. Il serait donc intéressant d'explorer en détail le polymorphisme au sein de chaque métapopulation, voire de chaque bassin (Lau et Manus, pour lesquels nous avons un nombre de sites important) en générant un jeu de données indépendant, bien que l'on ait déjà capturé des locus liés à de la différenciation entre bassins à l'échelle de chaque métapopulation avec le jeu de données global.

**CHAPITRE II : PHYLOGEOGRAPHIE
COMPAREE DES ESPECES
HYDROTHERMALES DU PACIFIQUE
OUEST**

I. INTRODUCTION

La phylogéographie cherche à explorer et à comprendre les liens entre la géographie historique, biogéographie des espèces et les mécanismes de vicariance, permettant ainsi de comprendre l'évolution des populations dans un contexte géographique, tout en testant des hypothèses concernant différents rôles de processus historiques ou évolutifs sur les patrons observés entre ces dernières mais seulement à l'échelle d'une espèce (Avice, 2009; Avice et al., 1987; Hickerson et al., 2010). Etendre celle-ci par une approche comparative, en comparant des espèces ayant une co-occurrence de distribution dans une région géographique donnée peut être particulièrement informatif, car un certain nombre de patrons de différenciation génétique peuvent être dus à des processus évolutifs intrinsèques à l'espèce, alors que les patrons partagés entre espèces peuvent quant à eux révéler une histoire évolutive commune en lien avec des barrières physiques aux flux génique. La distinction entre ces processus est importante pour la compréhension des mécanismes de vicariance à l'échelle micro, macro évolutive et à celle des communautés d'espèces (Dawson, 2014; Pascual et al., 2017). De plus, ceci peut être utile pour prendre des mesures de conservation à la lumière des différences phylogéographiques à l'échelle de communautés d'espèces dans une région donnée.

Dans le cas des bassins arrière-arc du Pacifique Ouest, plusieurs études de génétique des populations se sont intéressées aux patrons de différenciation des espèces hydrothermales. La plupart ont mis en évidence une nette structuration des populations entre bassins, mais pas nécessairement pour toutes les espèces car certaines n'affichent aucune différenciation. Dans les rares cas où cela a été possible, ces études ont pu estimer un flux génique orienté principalement vers l'Est (Tableau 1).

Bien que ces travaux aient permis de révéler un certain nombre de différences et concordances dans les patrons de différenciation des populations de plusieurs espèces, elles affichent un certain nombre de limites : d'une part leur limitation dans l'échantillonnage (nombre d'individus, bassins échantillonnés et lieux exacts) et d'autre part par l'utilisation d'un nombre limité de marqueurs génétiques. Ce nombre limité de marqueurs, dans de nombreux cas, ne permet de refléter qu'une petite partie de l'histoire évolutive des populations, tout en pouvant être sensiblement affectés par différents processus évolutifs ou

intrinsèques aux espèces. Ces limitations peuvent donc mener à une difficulté dans la comparaison des patrons phylogéographiques des espèces hydrothermales du Pacifique Ouest.

Tableau 1 : Récapitulatif des études de structure génétique des populations dans le Pacifique Ouest pour différentes espèces hydrothermales (tiré et adapté de Plouviez et al., 2019). La couleur rouge indique qu'une structuration est observée et le noir indique l'absence de structure.

Références	Organisme	Marqueurs	Différenciation entre bassin	
(Kojima et al., 2000)	<i>Ifremeria nautilei</i>	COX 1	M VS NF	
(Kojima et al., 2001)	<i>Alviniconcha spp1</i>	COX 1		
(Suzuki et al., 2006)	<i>Alviniconcha spp2</i>	COX 1		
(Kyuno et al., 2009)	<i>Bathymodiolus</i>	ND4	M VS NF	NF vs L
	<i>Bathymodiolus</i>	COX 1	M VS NF	NF vs L
(Thaler et al., 2011)	<i>Ifremeria nautilei</i>	COX 1	M VS NF	NF vs L
	<i>Ifremeria nautilei</i>	MSAT	M VS NF	NF vs L
(Thaler et al., 2014)	<i>Chorocaris sp 2</i>	COX 1	M VS NF	
	<i>Chorocaris sp 2</i>	MSATs	M VS NF	
	<i>Munidopsis lauensis</i>	COX 1	Manus vs Lau	
	<i>Munidopsis lauensis</i>	MSAT	M VS NF	
(Lee et al., 2019)	<i>Austinogrea alayseae</i>	COX 1	M VS NF-Tonga	
		28S		
(Yahagi et al., 2020)	<i>Shinkailepas tollmanni</i>	COX 1		
(Breusing et al., 2021)	<i>Alviniconcha boucheti</i>	COX 1		
		ATPsynthase subunit	M/V vs NF	
			M/V vs Lau	
			M vs Niua South	
			NF vs Lau (1 sites)	
			Intra NF	
	<i>Alviniconcha kojimai</i>	COX 1		
		ATPsynthase subunit	M vs L	
	<i>Alviniconcha strummeri</i>	COX 1		
		ATPsynthase subunit	Mariana back -arc vs L	
	<i>Alviniconcha strummeri</i>	COX 1		
		ATPsynthase subunit	Intra Mariana back-arc	

M : Manus , NF : Nord-Fidji, L : Lau et MSAT : marqueurs microsatellites

Dans ce chapitre, nous allons donc explorer les patrons phylogéographiques des espèces hydrothermales du Pacifique Occidental par une approche comparative multi-spécifique de génomique des populations. Cette approche permettra de s'affranchir des principales limitations précédemment citées, par d'une part l'utilisation d'un important échantillonnage comparable au sein de tous les bassins arrière-arc et d'autre part l'utilisation de milliers de marqueurs génomiques indépendants permettant de proposer une reconstruction de l'histoire évolutive des populations des différentes espèces hydrothermales dans cette région.

II. MATERIEL ET METHODES

L'ensemble des méthodes et des approches ont été réalisées de manière similaire à ce qui a été fait pour *I. nautili* dans le premier chapitre. Cette même approche méthodologique permet une comparaison entre les résultats des différentes espèces, bien que les jeux de données associés aient été préparés et analysés de manière indépendante.

Collection d'échantillon et extraction d'ADN

Les échantillons pour *A. kojimai*, *S. tollmanni* et *E. ohtai* ont été collectés durant la mission océanographique Chubacarc en 2019 à bord du navire *l'Atalante*, comme pour ce qui a été réalisé pour *I. nautili* dans le premier chapitre. Cette stratégie d'échantillonnage a fourni des échantillons provenant des mêmes bassins, champs hydrothermaux ou localités, permettant d'avoir lorsque les espèces ont été observées, un échantillonnage hiérarchique et similaire entre espèces.

Au final, 294 individus ont été échantillonnés pour *A. kojimai*, 971 pour *S. tollmanni* et 696 pour *E. ohtai*, dans l'ensemble des bassins (sauf à Woodlark pour *E. ohtai* où ce dernier n'a pas été trouvé). L'ensemble des individus se répartit dans 11, 22 et 9 champs hydrothermaux pour *A. kojimai*, *S. tollmanni* et *E. ohtai*. (Figure 6). Comme pour *I. nautili*, l'échantillonnage couvre l'entièreté des aires de répartitions connues à ce jour pour nos espèces.

Une fois à bord, les individus ont été disséqués, différents tissus ont été conservés dans l'éthanol à 90° (EtOH) ou à -80°C. Les extractions d'ADN ont été réalisées à partir de tissus du manteau pour *A. kojimai*, pour limiter les contaminations avec les symbiontes dans les

branchies. Pour *S. tollmanni* et *E. ohtai*, les extractions en majorité ont été réalisées à partir des individus entiers à l'exception de certains dont une partie a été conservée dans de l'éthanol à 90%.

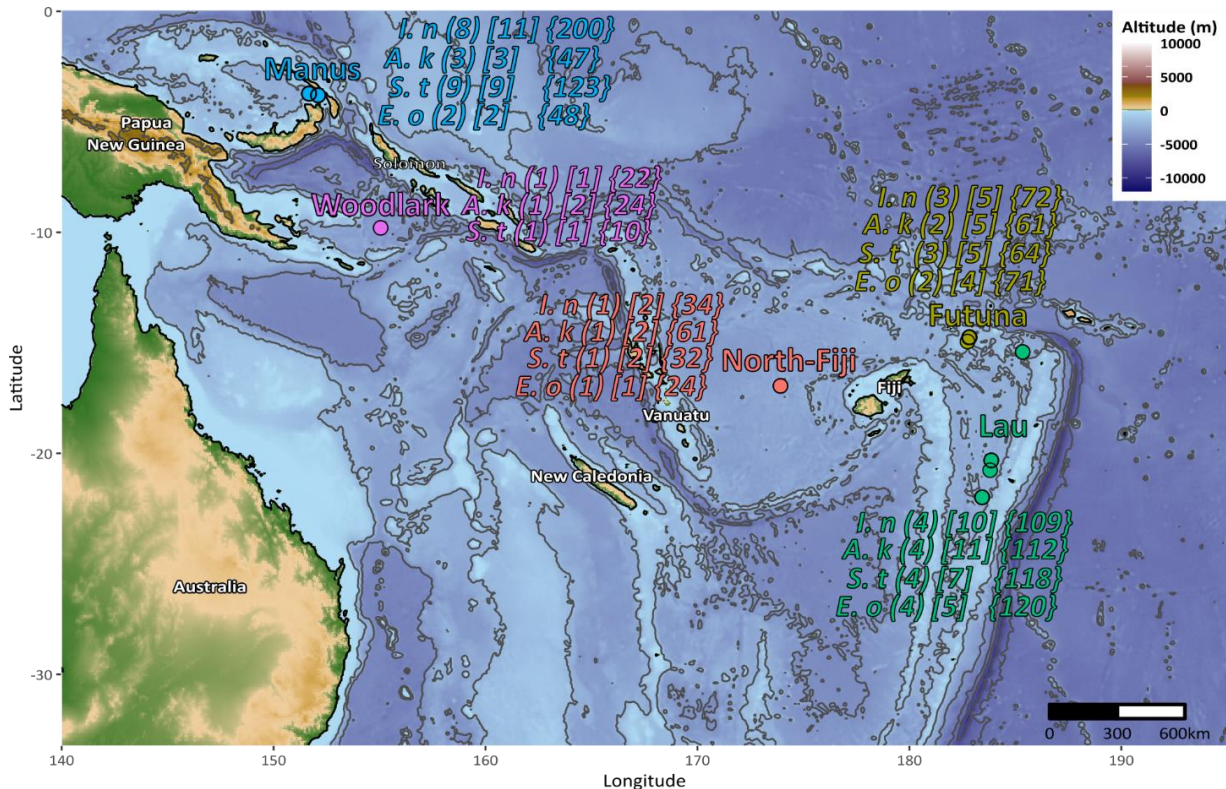


Figure 6 : Carte des échantillons utilisés par espèces et par bassin dans le Pacifique Ouest. La couleur des caractères représente les différents bassins. Les nombres entre parenthèses représentent le nombre de localités échantillonnées, ceux entre les crochets les sites et entre accolades le nombre total d'individus utilisés par bassin dans les banques ddRAD. *I. n* : *Ifremeria nautilei* ; *A. k* : *Alviniconcha kojimai* ; *S. t* : *Shinkailepas tollmanni* et *E. o* : *Eochionelasmus ohtai*.

La qualité des ADN a été vérifiée par dépôt et migration dans un gel d'agarose à 0.8% comme pour ce qui a été fait pour *I. nautilei*. La préparation des banques génomiques ddRAD a été réalisée de manière standardisée en suivant le protocole utilisé dans le premier chapitre (Daguin-Thiébaud et al., 2021). Ces banques représentent 294, 469, 282 individus incluant 8, 39 et 14 réplicats contenant un total de 22, 83 et 33 individus répliqués pour les contrôles qualité et calibration des paramètres pour *A. kojimai*, *S. tollmanni* et *E. ohtai*. Le séquençage de ces banques a produit en moyenne 4,2 et 3,6 millions de lectures pour *S. tollmanni* et *E. ohtai*. Les banques ont été séquençées chez Novogene. Les banques et lectures d'*A. kojimai*

proviennent de Castel et al. (soumis), ayant suivi le même protocole de production de banques que dans le chapitre 1.

Assemblages des lectures, appel des variants (SNPs) et filtres.

Les lectures ("reads") ont été vérifiées avec Fastqc (V.0.11.9) pour contrôler simplement leur qualité. Pour chaque espèce, les assemblages des données ont été réalisés de manière indépendante mais en suivant le même protocole et méthodologie que décrit dans le premier chapitre. Comme pour *I. nautiliei*, aucun génome de référence n'est disponible actuellement pour ces espèces, ainsi les lectures ont été assemblées de manière « *de novo* » en utilisant chaque module de stacks de manière indépendante. Pour identifier les paramètres d'assemblage les plus appropriés, nous avons la même méthodologie que dans le chapitre 1 (Voir Figures 1 à 23 Annexe 2). Ainsi, nous avons choisi d'utiliser les paramètres d'assemblage $m = 6$, $m = 5$ et $m = 4$ et $M = 7$, $M = 11$ et $M = 7$ pour *A. kojimai*, *S. tollmanni* et *E. ohtai* respectivement.

Comme pour *I. nautiliei*, les jeux de données ont été filtrés pour réduire le nombre de données manquantes et potentiels paralogues (Tableau 1 Annexe 2). Pour *S. tollmanni*, nous n'avons gardé que les individus ayant moins de 15% de données manquantes afin de minimiser le pourcentage de données manquantes, tout en gardant suffisamment d'individus, RADtags et SNPs.

Pour obtenir des résultats comparables bien qu'analysés indépendamment, nous avons appliqué exactement les mêmes analyses, programmes et paramètres que nous avons utilisés dans le chapitre 1. Cette approche a été utilisée pour chaque espèce, que ça soit pour les analyses de structuration des populations, indices de diversités, tests d'*ADMIXTURE* et inférences démo-génétiques avec *dadi*. Néanmoins, nous n'avons pas exploré les analyses de locus aberrants « outliers » comme nous l'avons fait pour *I. nautiliei* (cf. Chapitre 1).

III. RESULTATS

Les assemblages « *de novo* » des lectures obtenues pour les différentes espèces ont généré plusieurs jeux de données. Après application de filtres similaires utilisés dans le chapitre 1 sur *Ifremeria nautilei*, pour *Alviniconcha kojimai*, nous avons obtenu 12 957 SNPs indépendants pour une couverture moyenne de 26,56X. Pour *Shinkailepas tollmanni*, le jeu de données final contient 3074 SNPs, 45 984 SNPs pour *Eochionelasmus ohtai* pour une couverture moyenne respective de 24,48X et de 22,54X, en plus des 10 570 SNPs utilisés pour *Ifremeria nautilei* dans le chapitre 1.

1. Structuration des populations

Analyses inter-métapopulations

Dans cette partie, nous distinguerons ce que l'on appelle résultats des analyses inter-métapopulations, qui correspondent aux principaux signaux de différenciation, des résultats qui affichent des subtiles différences entre les populations et au sein de ces dernières.

Les analyses en composantes principales (ACP) montrent une tendance commune à toutes les espèces. La première composante principale (PC1) sépare les individus en deux sous-ensembles distincts, dont le premier rassemble les individus de Manus et de Woodlark, le second ceux du bassin Nord-Fidjien, de l'arc volcanique de Futuna et du bassin de Lau. Cette première composante explique une grande partie de la variance totale expliquée par la géographie, allant de 26,03 % à 1,94% en fonction des espèces (Figure 7). La deuxième composante (PC2) est bien plus faiblement discriminante avec 0,3-0,6 % de variance expliquée. Dans le cas de *A. kojimai*, ce deuxième axe de différenciation permet de séparer les individus de Lau/Futuna avec ceux du bassin Nord-Fidjien, ce qui semble assez paradoxal car cette espèce est celle qui apparaît la moins structurée géographiquement sur la première composante (1,94 % de variance expliquée). Chez *S. tollmanni* (Figure 7 C), les populations se divisent aussi en 2 sous-groupes géographiques ou métapopulations. Cependant les individus du bassin de Woodlark se scindent nettement entre ces 2 sous-ensembles et un individu est positionné à distance équivalente entre ces 2 groupes géographiques sur l'axe PC1, tout en

affichant une très faible quantité de données manquantes (Figure 24 Annexe 2). Les autres composantes n'apportent pas d'informations supplémentaires (Figure SI 25 pour PC3 et PC4).

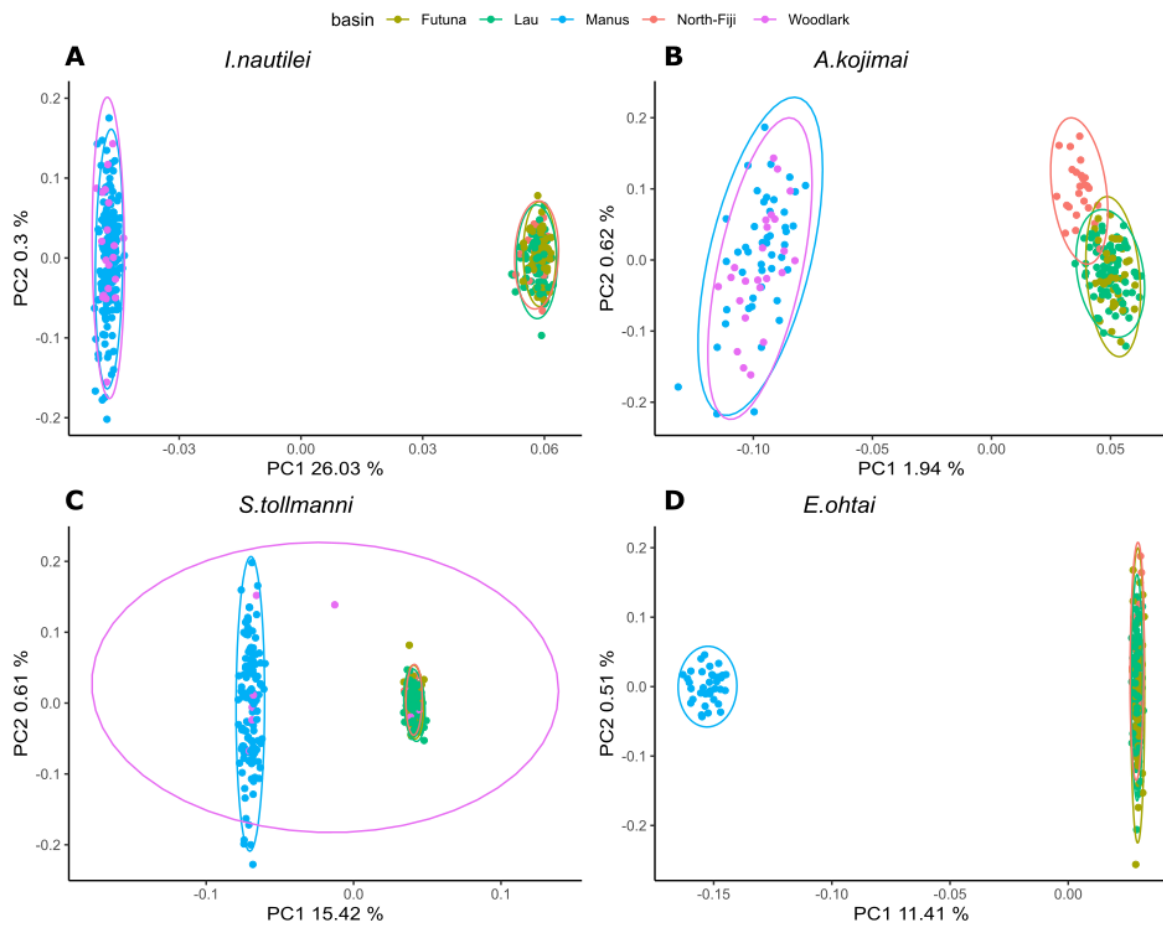


Figure 7 : Analyses en composantes principales (ACP) par espèce pour les deux premières composantes. Les points représentent les individus et les couleurs représentent les bassins arrière-arc. Le cercle représente la distribution normale multivariée à 95% de chaque groupe (bassin). (A) *I. nautili* (B) *Alviniconcha kojimai*, (C) *Shinkailepas tollmanni* et (D) *Eochionelasmus ohtai*.

Cette structure en deux ensembles géographiques est cohérente avec les valeurs de différenciation obtenues avec les AMOVA et les matrices de F_{st} par paire de bassins (Tableau 2 et Tableau SI 2 Annexe 2), où l'on observe des fortes valeurs significatives entre les deux ensembles, allant de 0,387 pour *I. nautili*, 0,271 pour *S. tollmanni*, 0,203 pour *E. ohtai* et de 0,0184 pour *A. kojimai*. Ces résultats sont corroborés par les analyses *ADMIXTURE* qui sélectionnent $K = 2$ comme la meilleure valeur du nombre d'entités génétiques distinctes

possibles au sein de nos espèces (Figure 8 et Figure SI 26 Annexe 2) avec une très nette différenciation des groupes. Il y a dans l'ensemble très peu d'ascendance partagée entre les groupes, sauf en ce qui concerne la population de *S. tollmanni* à Woodlark. Les individus Nord-Fidjiens d'*A. kojimai* possèdent aussi une signature minoritaire Manus/Woodlark.

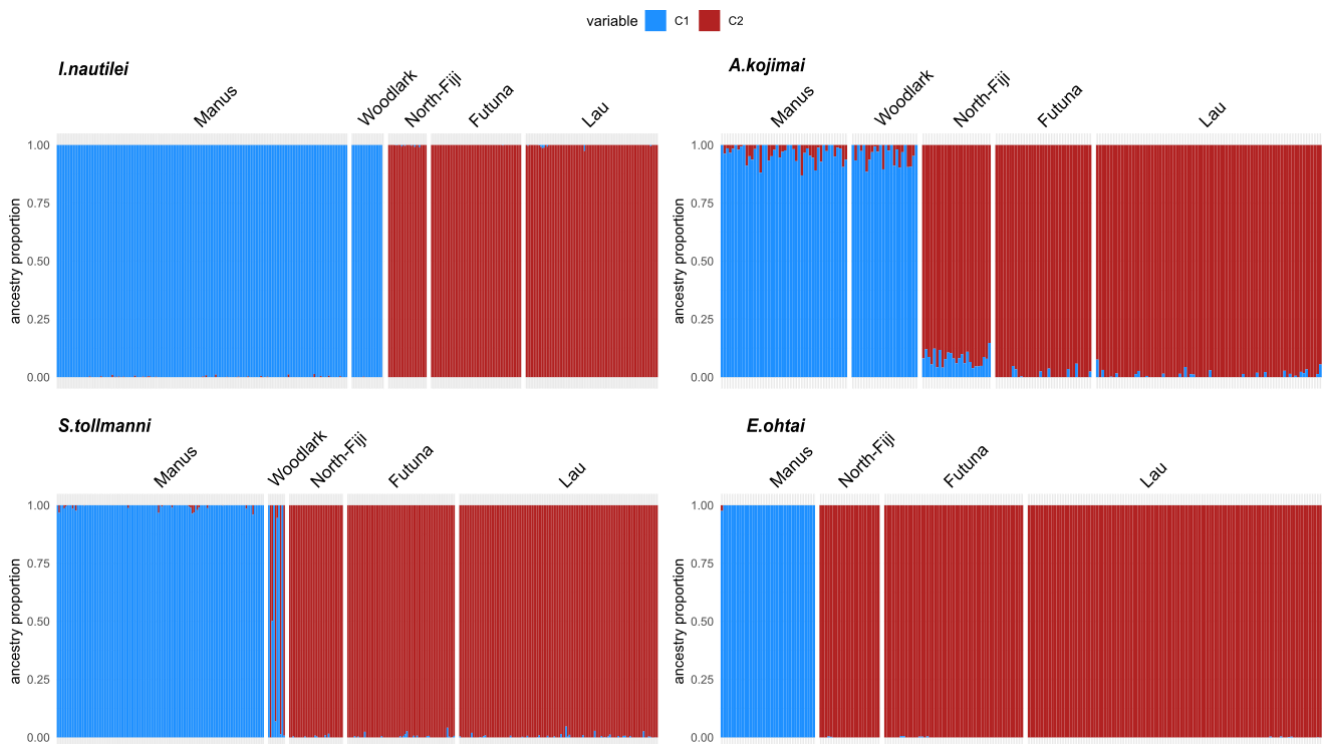


Figure 8 : analyses ADMIXTURE pour K=2 pour chaque espèce en séparant les individus par bassin géographique

Analyses intra-métapopulations

Les résultats semblent n'afficher aucune structure au sein de chaque ensemble géographique (est/ouest), à l'exception d'*A. kojimai* et de *S. tollmanni*. Dans les analyses en composantes principales, aucune structure n'est observée sur les autres composantes (PC2-PC4) à l'exception de la population de Nord-Fidji pour *A. kojimai*, qui affiche une légère différenciation sur les deux premières composantes (PC1 et PC2), se démarquant ainsi de Futuna/Lau sur la PC2 et se rapprochant légèrement de l'autre ensemble Manus/Woodlark sur la PC1, cohérent avec les valeurs de F_{st} entre bassins (Tableau SI 2,3,4 Annexe 2), où Nord-

Fidji affiche une valeur faible mais significative de F_{st} avec Futuna/Lau et une différenciation moins importante avec Manus/Woodlark.

Les analyses de la variance moléculaire (AMOVA) n'affichent aucune structure significative au sein de chaque ensemble : entre les bassins ($F_{ct} = NS$ ou ≤ 0.001 , Tableau 2) ainsi qu'à l'échelle des localités au sein de chaque bassin. Ces résultats sont aussi corroborés avec ceux obtenus avec *ADMIXTURE* où la valeur de K optimale était de 2 et aucun autre cluster n'a été identifié à l'exception de $K=3$ pour Nord-Fidji chez *A. kojimai* (Figure SI 27 Annexe 2).

Tableau 2 : Analyse moléculaire de la variance (AMOVA) sur les jeux de données par espèces avec 10 000 permutations (***: $p < 0.001$, **: $p < 0.01$ *: $p < 0.05$)

		<i>I. nautiliei</i>	<i>A. kojimai</i>	<i>S. tollmanni</i>	<i>E. ohtai</i>
Entre metapopulation	F_{ST}	0.38773***	0.01847***	0.271***	0.203***
Basins dans chaque metapopulation	F_{CT}	-0.05	0.00076***	-0.00014*	0.00005***
Localités dans les Basins	F_{SC}	-0.00011	0.00025	0.00037	-0.00017
Individus dans les Localités	F_{IS}	-0.05084	-0.02695	-0.06205	-0.05124

2. Introgression

Les analyses d'*ADMIXTURE* révèlent une très faible proportion de fond génétique partagé entre les deux ensembles génétiques chez *I. nautiliei*, *S. tollmanni* et *E. ohtai*. Cependant, l'espèce *A. kojimai* affiche une proportion de fond génétique partagé bien plus élevée entre les deux ensembles génétiques et plus particulièrement chez les individus de Nord-Fidji, où l'on observe une part plus importante d'ascendance génétique Manus/Woodlark par rapport aux individus de Futuna/Lau.

Les analyses complémentaires *Treemix* et tests F_3 révèlent un certain nombre de résultats importants. Tout d'abord, le nombre d'évènements optimaux de migration pour *I. nautiliei* et

A. kojimai est de deux ($m = 2$), alors que pour *S. tollmanni* et *E. ohtai* celui-ci est d'un ($m = 1$). L'ajout d'évènements de migration complémentaire ($m = 3$ à 5) n'améliore pas la vraisemblance des modèles (Figure SI 28 Annexe 2).

Le premier évènement de migration est consistant avec une faible valeur de migration d'une métapopulation à l'autre (Figure 9). Pour *I. nautiliei*, celui-ci affiche un évènement de migration allant de Manus/Woodlark vers Nord-Fidji. Cependant pour *S. tollmanni* et *A. kojimai*, celui-ci révèle une migration dans la direction opposée, allant de Nord-Fidji/Futuna/Lau vers Manus/Woodlark. Pour la dernière espèce, *E. ohtai*, malgré l'amélioration de la vraisemblance, l'ajout d'un premier évènement de migration ($m = 1$) n'apporte aucune information concernant la direction du flux génique. Le deuxième évènement de migration ($m = 2$) n'affiche qu'un flux génique entre Manus et Woodlark pour *I. nautiliei* et de la branche principale de Nord-Fidji/Futuna/Lau vers Lau pour *A. kojimai*.

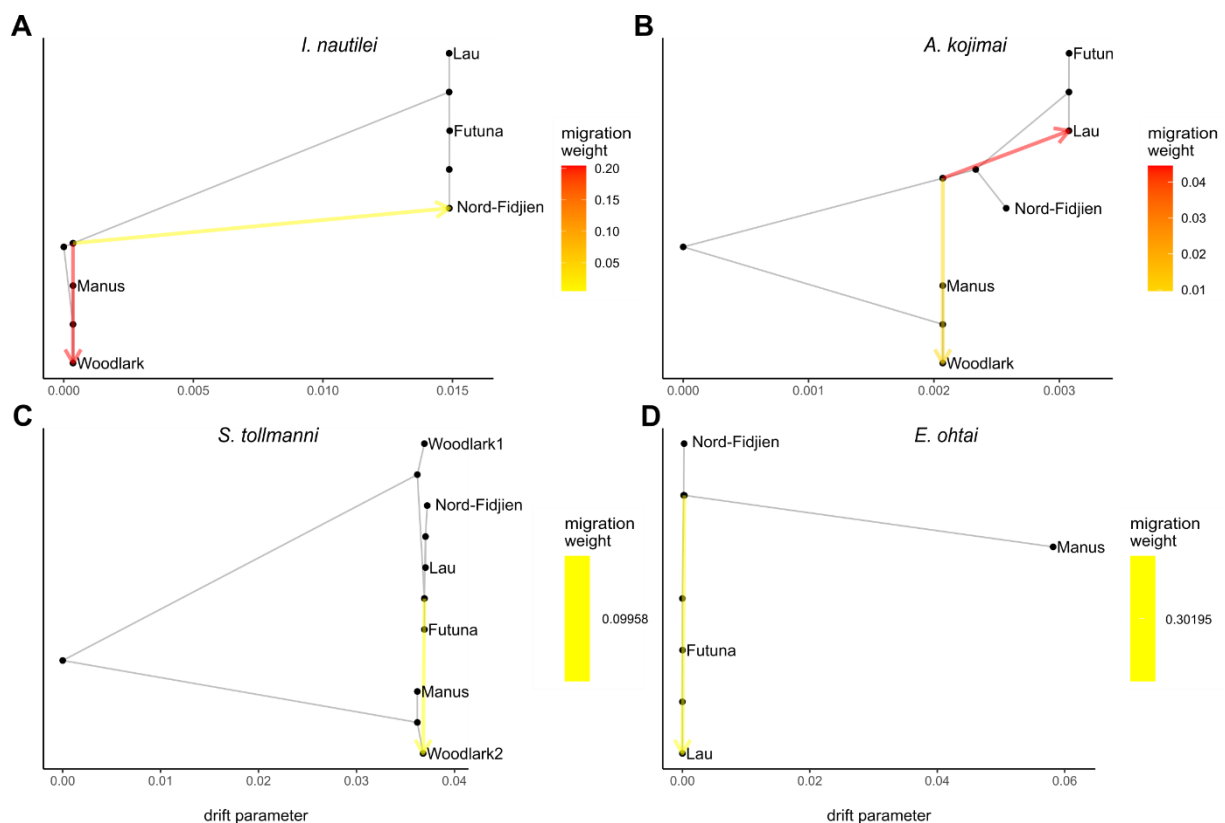


Figure 9 : Analyses Treemix avec les valeurs de m optimale pour chaque espèce.

La statistique F_3 montre seulement des valeurs négatives et significatives pour les espèces *I. nautiliei* et *S. tollmanni* (Figure SI 29 Annexes 2). Pour la première, les valeurs significatives sont seulement dans la configuration où Nord-Fidji est la population focale (cible) et les populations sources sont issues de chaque ensemble génétique (Manus/Woodlark et Lau/Futuna, respectivement). Pour *S. tollmanni*, les valeurs sont significatives uniquement lorsque les individus de Woodlark sont séparés en 2 groupes et en prenant les individus de Woodlark qui se regroupent avec les individus de Manus comme population focale. Dans ce cas, les populations sources sont les mêmes que celles trouvées pour *I. nautiliei* (Manus et Woodlark1/Futuna/Lau/Nord-Fidji).

3. Diversité génétique

Les descripteurs de la diversité génétique sont assez variables d'une espèce à l'autre et selon la métapopulation considérée. L'espèce *I. nautiliei* possède une diversité légèrement plus élevée, quelle que soit la statistique (π , Het exp, Het Obs), pour la métapopulation Manus/Woodlark (population source de la migration préférentielle) que pour celle de Nord-Fidji/Futuna/Lau. L'inverse est observé chez les autres espèces, où la métapopulation Nord-Fidji/Futuna/Lau (population source de la migration préférentielle) est génétiquement plus diversifiée que pour Manus/Woodlark. On observe aussi que l'espèce *S. tollmanni* montre la plus forte diversité génétique parmi les espèces analysées, quelle que soit la métapopulation (Figure 10). La mesure de divergence nette (D_a), qui correspond à la mesure de la divergence brute (D_{xy}) mais corrigée par la diversité nucléotidique moyenne (π) par la formule suivante :

$$\left(D_a = D_{xy} - \frac{\pi x + \pi y}{2} \right)$$

Elles affichent des valeurs assez proches et pas très élevées pour *I. nautiliei*, *A. kojimai* et *E. ohtai* (0.80 %, 0.67% et 0.73%). A l'opposé, *S. tollmanni* possède une valeur de divergence nette comparativement élevée (1.26%).

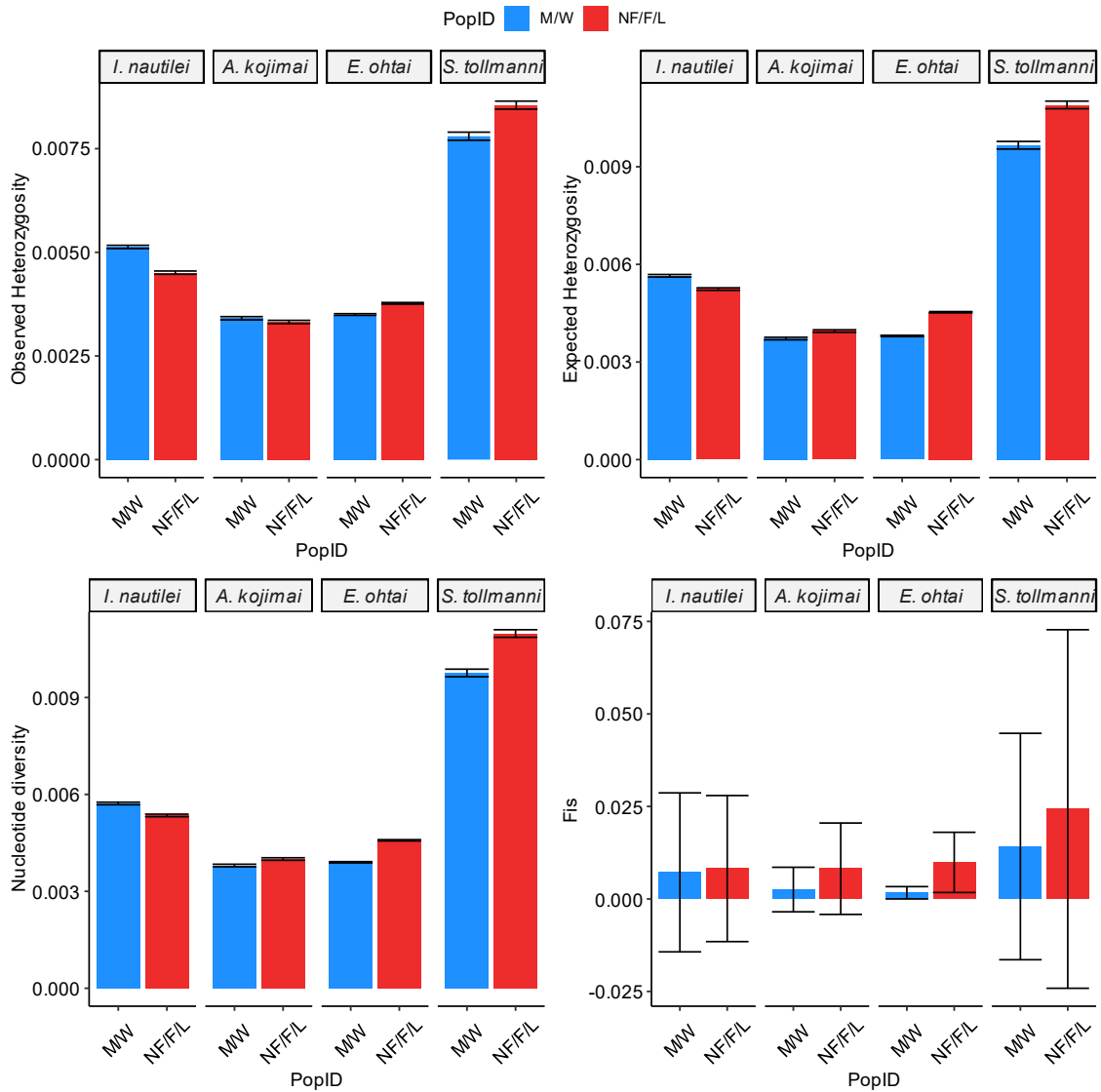


Figure 10 : Indices de diversité estimé par Stacks sur l'ensemble des RADtags de chaque jeu de données pour chaque espèce et chaque métapopulation. Les barres représentent l'intervalle de confiance à 95%. M/W : Manus/Woodlark, NF/F/L : Nord-Fidji/Futuna/Lau

4. Inférence de l'histoire de divergence

Les spectres joints des fréquences alléliques repliés montrent le partage des allèles entre métapopulations pour les différentes espèces (Figure 11). On observe que pour toutes les espèces, des allèles sont partagés (allèles en fréquence intermédiaire) entre les deux métapopulations et que pour toutes à l'exception de *A. kojimai*, elles affichent des allèles fixés et privés à chaque métapopulation. *S. tollmanni* semble néanmoins afficher moins d'allèles partagés que les autres espèces.

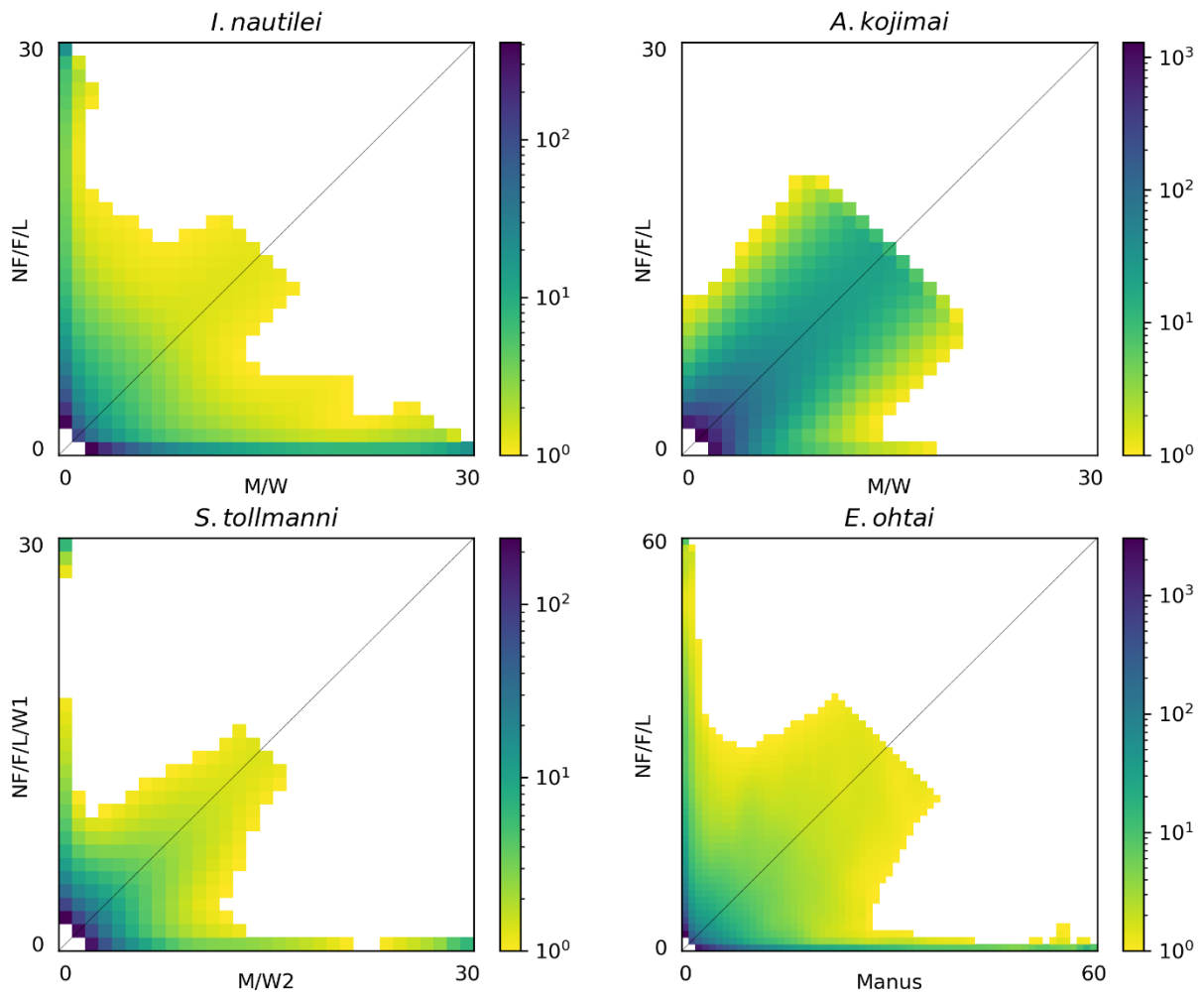


Figure 10 : Spectre des fréquences alléliques par paire de métapopulations pour les quatre espèces.

Modèles démo-génétiques

Pour les quatre modèles de base de *dadi* (SI/IM/AM/SC), quelles que soient les espèces, les modèles avec migration présentent une meilleure valeur d'AIC que le modèle d'isolement strict (SI). Le modèle de contact secondaire est préféré pour *I. nautili* et *E. ohtai* alors que pour *A. kojimai* le modèle d'isolement avec migration a une meilleure valeur d'AIC. Pour *S. tollmanni*, *dadi* n'a pas réussi à ajuster ces modèles de base au spectre observé (Figure 11).

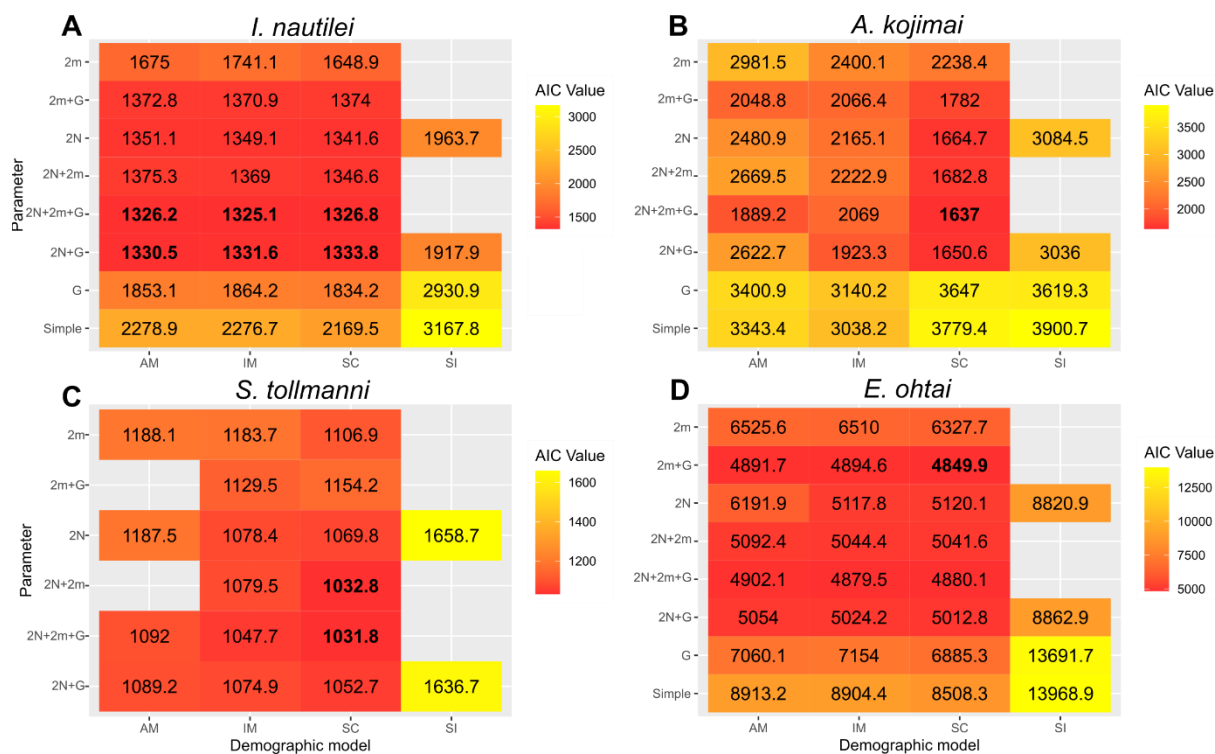


Figure 11 : Carte de chaleur pour les meilleures valeurs du critère d'information d'Akaike (AIC) par espèces, pour chaque combinaison des paramètres (expansion ou contraction de la population (G), effet de la sélection liée (2N) et migration hétérogène (2m)) et modèle démographique (isolement strict (SI), isolement avec migration (IM), migration ancienne (AM) et contact secondaire (SC)). Les meilleurs modèles sont indiqués par les valeurs en gras.

Pour toutes les espèces, l'ajout de complexité par la prise en compte d'effets tels que la sélection en liaison (2N), l'hétérogénéité de la migration (2m) et expansion ou contraction des populations (G), améliore l'ajustement aux données observées, et cela quels que soient les modèles simulés. La contribution indépendante de chaque paramètre (2N, 2m et G) ne diffère pas en fonction des espèces, l'effet de la sélection en liaison contribue plus à l'ajustement des modèles simulés que l'hétérogénéité de la migration (2m) et la variation de la taille des populations (G). Ce dernier paramètre contribue bien moins à l'amélioration des valeurs d'AIC des modèles simulés que les deux autres paramètres (2m et 2N).

La prise en compte des différents paramètres de manière simultanée améliore légèrement les valeurs d'AIC de chaque modèle. A l'exception de *E. ohtai*, les modèles avec l'effet de la sélection en liaison (2N) restent meilleurs. Les modèles incluant tous les paramètres (2N+2m+G) sont les meilleurs pour *S. tollmanni*, *A. kojimai* et *I. nautiliei* (Figure 11 & Figure SI 30,31 Annexe 2). Pour cette dernière espèce tous les modèles (IM/AM/SC) affichent des

valeurs d'AIC similaires, alors que pour *S. tollmanni*, le modèle de contact secondaire (SC) produit une meilleure valeur d'AIC ($\Delta_{AIC} > 10$ avec les autres). Pour *E. ohtai*, le modèle expliquant le mieux les données observées est aussi un modèle de contact secondaire (SC) mais ne prenant en compte que de l'hétérogénéité dans la migration et avec une variation de taille des populations (2m+G) (Figure SI 32 Annexe 2). De plus, il est à noter que les modèles 2N+G pour *I. nautiliei*, sont très proches en valeurs d'AIC (<10) des meilleurs modèles (2N+2m+G) pour cette même espèce (voir la discussion dans le chapitre 1 correspondant à l'Article Tran Lu Y et al 2022).

Temporalité de la divergence des métapopulations

En accord avec les résultats des meilleurs modèles de divergence en fonction de leur valeur d'AIC, les métapopulations de toutes les espèces étudiées ont divergé selon un modèle de contact secondaire, avec une phase d'isolement strict suivi par une phase de divergence avec migration, à l'exception de *I. nautiliei* pour laquelle les différents modèles de divergence avec migration (IM/AM/SC) n'arrivent pas à être distingués sur la base des valeurs d'AIC. Malgré cela, on observe des résultats similaires mais différents entre espèces concernant les temps de divergence et de contact secondaire (Tableau 3). Tout d'abord, on observe que les temps de divergence totale varient entre les espèces, allant de 3 695 à 118 000 générations en considérant un taux de mutation identique pour toutes les espèces (10^{-8}). Seul *A. kojimai* affiche une temporalité de divergence très récente, estimée à 3 695 générations alors que les autres espèces ont des temps estimés très similaires entre eux. Le temps estimé varie entre 67 000 -70 000 générations pour les modèles (2N+2m+G) et 112 000 – 118 000 générations pour ceux (2N+G) chez *I. nautiliei*. Des valeurs similaires ont été estimées chez *S. tollmanni* et *E. ohtai*, allant de 105 00 à 116 987 générations (Tableau 3).

Lorsque l'on regarde la temporalité du contact secondaire pour les meilleurs modèles des différentes espèces, des valeurs assez différentes sont observées entre espèces, allant de 1015 à 111 530 générations en fonction des espèces. Les espèces *A. kojimai* et *S. tollmanni* montrent des valeurs récentes de contact secondaire avec 1015 et 9 759 générations, alors que les espèces *E. ohtai* et *I. nautiliei* affichent des valeurs plus élevées de 70 574 et 54 027 (2N+2m+G) à 111 530 (2N+G) générations, respectivement.

Tableau 3 : Estimation des différents temps de divergence en générations (estimé avec un taux de mutation de 10^{-8} /site/génération) et de leurs étagements pour les différents modèles. T_s : temps de divergence stricte, $T_{sm}/T_{am}/T_{sc}$: temps de divergence avec migration pour les différents modèles et T total : le temps total de divergence

Espèce	Modèle	T_s	$T_{sm}/T_{am}/T_{sc}$	T total
<i>I. nautili</i>	IM+2N+2m+G		66 951,08	66 951,08
<i>I. nautili</i>	SC+2N+2m+G	16 278,50	54 016,70	70 295,20
<i>I. nautili</i>	AM+2N+2m+G	41,27	69 371,70	69 412,96
<i>I. nautili</i>	SC+2N+G	10 344,46	103 444,60	113 789,06
<i>I. nautili</i>	AM+2N+G	1 175,43	111 527,16	112 702,58
<i>I. nautili</i>	IM+2N+G		118 023,89	118 023,89
<i>A. kojimai</i>	SC+2N+2m+G	2 680,13	1 015,16	3 695,28
<i>S. tollmanni</i>	SC+2N+2m+G	94 668,77	10 508,14	105 176,91
<i>S. tollmanni</i>	SC+2N+2m	96 767,24	8 305,00	105 072,24
<i>E. ohtai</i>	SC+2m+G	46 308,20	70 591,77	116 899,97

Flux génique

En considérant les meilleurs modèles pour toutes les espèces, ces derniers comprennent tous du flux génique entre les deux métapopulations et le modèle d'isolement strict (SI) est dans tous les cas le moins bon. Le flux génique pour toutes les espèces est bidirectionnel et avec une légère asymétrie, avec une valeur légèrement plus importante Est-Ouest, de Nord-Fidji/Futuna/Lau vers Manus/Woodlark (Tableau 4 & Tableau SI 5 Annexe 2).

Pour *I. nautili*, le flux génique non restreint (m_{12}) allant de Manus/Woodlark (2) vers Nord-Fidji/Futuna/Lau (1) est estimé entre 0,421 et 0,444 alors que dans le sens opposé (m_{21}), celui-ci est estimé être entre 0,809 et 0,825 pour les modèles (2N+m+G). Pour cette même espèce, les modèles (2N+G) affichent une valeur entre 0,176 et 0,217 pour m_{12} et entre 0,489 et 0,495 pour m_{21} . Des valeurs et asymétries similaires sont observées chez *S. tollmanni* et *E. ohtai*, où le flux génique (m_{12}) est respectivement estimé à 4,04 à 4,747 et 0,364 chez ces espèces et de 12,882 à 14,381 et de 1,386 dans le sens inverse (m_{21}). La dernière espèce, *A. kojimai* arbore aussi cette asymétrie mais avec pour valeur de m_{12} estimé à 5,230 et m_{21} à 5,871.

L'ensemble des espèces suivent des modèles avec du flux génique restreint (2m), néanmoins pour *I. nautili*, ces modèles (2N+2m+G) sont très proches de ceux sans flux génique restreint (2N+G) (voir discussion chapitre 1). Pour les autres espèces, la proportion de ce flux génique restreint (1-P) reste très faible, allant de 0,014 à 0,039 pour *A. kojimai* et *S. tollmanni*. Néanmoins pour *E. ohtai*, qui est la seule espèce qui favorise le modèle avec seulement de la migration réduite, celui a une valeur de 1-P de 0,492.

Tableau 4 : Estimations des taux de migration entre populations (estimé avec dadi) pour chaque espèce et pour le meilleur modèle. P représente la proportion de locus migrant avec les valeurs m_{12} et m_{21} . 1-P représente ceux avec les valeurs me_{12} et me_{21} . Q la proportion de locus influencé par la sélection en liaison.

Espèce	Modèle	m_{12}	m_{21}	me_{12}	me_{21}	P	Q
		NF/F/L<-M/W	M/W <-NF/F/FL	NF/F/L<-M/W	M/W <-NF/F/FL		
<i>I. nautili</i>	IM+2N+2m+G	0,444	0,825	0,038	0,283	0,483	0,99
<i>I. nautili</i>	SC+2N+2m+G	0,422	0,81	0,038	0,27	0,439	0,99
<i>I. nautili</i>	AM+2N+2m+G	0,461	0,825	0,039	0,3	0,471	0,99
<i>I. nautili</i>	SC+2N+G	0,18	0,49				0,566
<i>I. nautili</i>	AM+2N+G	0,217	0,492				0,584
<i>I. nautili</i>	IM+2N+G	0,178	0,495				0,57
<i>A. kojimai</i>	SC+2N+2m+G	5,230	5,871	2,394	6,845	0,990	0,301
<i>S. tollmanni</i>	SC+2N+2m+G	4,04	12,882	0,338	2,697	0,975	0,418
<i>S. tollmanni</i>	SC+2N+2m	4,334	13,866	0,321	3,314	0,986	0,434
<i>E. ohtai</i>	SC+2m+G	0,364	1,386	0,013	0,124	0,508	

IV. DISCUSSION

Cette discussion est organisée en deux sections, la première est centrée autour des résultats de notre approche génomique de phylogéographie comparative des quatre espèces, puis la seconde partie est une comparaison de nos résultats avec la littérature et études basées soit uniquement sur le marqueur *Cox 1*, soit sur une combinaison de ce dernier et avec quelques marqueurs nucléaires.

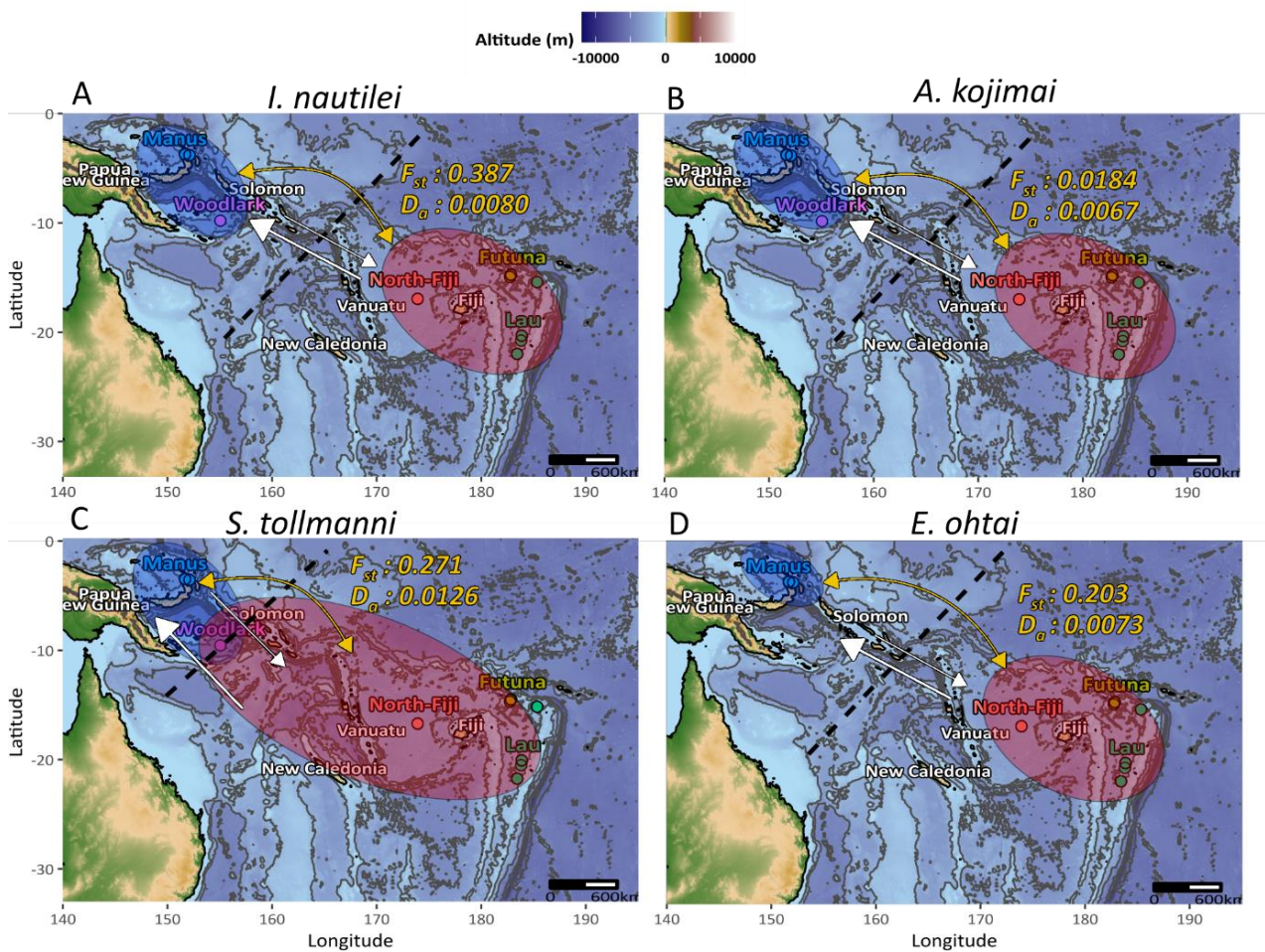


Figure 12 : Récapitulatif schématique des résultats de structure des populations et d'inférence démographique par espèces. Les valeurs en jaune représentent les mesures de différenciation (F_{st}) et de divergence (D_a). Les flèches blanches représentent le sens du flux génique. Les cercles de couleurs représentent les métapopulations et la ligne en pointillé une barrière perméable au flux génique.

Tableau 5: Récapitulatif simplifié des résultats de structuration et d'inférence démographique

Espèces	Métapopulations	F_{st} / D_a	Modèle démographique	Paramètres du modèle	Asymétrie du flux génétique	Temps de divergence (en génération)
<i>I. nautiliei</i>	M/W et NF/F/L	0,387/0,008	SC/IM/AM	2N+2m+G/ 2N+G	Ouest	66 -70 K/ 112- 118K
<i>A. kojimai</i>	M/W et NF/F/L	0,0184/0,0067	SC	2N+2m+G	Ouest	3,7 K
<i>S. tollmanni</i>	M/W et NF/F/L	0,271/0,0126	SC	2N+2m+G	Ouest	105 K
<i>E. ohtai</i>	M et NF/F/L	0,203/0,0073	SC	2mG	Ouest	116 K

M: Manus, W: Woodlark, NF: Nord-Fidji, F: Futuna, L: Lau

1. Comparaison entre espèces par approche génomique (ddRAD)

Les résultats de structuration des populations et les inférences démographiques ont révélé des patrons similaires de connectivité, où les populations sont quasi-homogènes génétiquement à l'échelle de l'ensemble des bassins avec une forte différenciation de part et d'autre d'une zone barrière au niveau des arcs volcaniques et de subduction des Salomon et Vanuatu. Ces deux entités génétiques ou métapopulations sont néanmoins capables d'échanges génétiques en affichant un faible flux génétique bi-directionnel et avec une légère asymétrie orientée d'est en ouest (Figure 12).

Bien qu'une quasi-homogénéité au sein de ces entités génétiques soit observée, des subtilités et dissimilitudes de structure entre espèces existent entre bassins au sein des métapopulations. Ces variations et disparités peuvent être la conséquence des histoires évolutives particulières de chaque espèce, influencées probablement par différents facteurs allant des traits d'histoire de vie des espèces, au degré de fragmentation des habitats, en passant par des différences au niveau de l'arrivée et/ou apparition des espèces dans l'aire géographique.

Ainsi dans notre approche, nous avons essayé de mieux évaluer la contribution relative de ces différents facteurs sur les patrons de distribution des espèces. Bien que ces espèces vivent dans le même environnement et affichent des aires de distribution similaires, ces patrons sont-ils le résultat (1) d'une histoire géologique commune (fronts de migration à travers la mise en place des bassins ou mise en place de barrières physiques à la dispersion) ou sont-ils

(2) plus spécifiquement sous l'influence des traits d'histoire de vie des espèces, selon leur capacité de dispersion et influencé par la courantologie actuelle et passée ?

Bien que plusieurs études pionnières aient été réalisées sur certains de ces organismes, celles-ci n'apportent qu'une partie de l'information, en n'examinant la différenciation génétique des populations que sur un petit nombre de locus (*Cox 1* et quelques microsatellites) et en n'étudiant plus spécifiquement que les bassins de Manus et de Lau. Si le Bassin de Nord-Fidji a parfois été échantillonné, les populations du Bassin de Woodlark et de l'Arc Volcanique de Futuna n'ont alors jamais été étudiées pour la connectivité. D'autre part, ces travaux pionniers n'ont pas été effectués dans le cadre d'une approche comparative plurispécifique et n'ont pas exploré l'histoire évolutive possible des populations et leurs connectivités, ayant conduit à la colonisation des bassins du Pacifique Ouest. Notre approche à l'échelle génomique (utilisation de milliers de locus génomiques indépendants) combinée à un échantillonnage hiérarchique et similaire entre espèces, permet d'aller plus loin et de mieux comprendre comment se structurent les populations, d'inférer leur niveau de connectivité passée et présente ainsi que leur histoire démographique.

Existe-t-il des déterminants géomorphologiques concordants pour les différentes espèces ?

(1) DIFFERENTIATION EST/OUEST

Nos résultats pointent de manière générale vers l'existence d'une zone de moindre perméabilité génétique entre les deux métapopulations (Manus/Woodlark et Nord-Fidji/Futuna/Lau), se trouvant probablement au niveau des îles Vanuatu/Salomon pour la plupart des espèces (zone non échantillonnée). Cependant, *S. tollmanni* n'affiche pas cette zone de moindre perméabilité mais le site du Bassin de Woodlark apparaît comme une zone de contact actuelle entre les deux métapopulations.

Bien que l'on observe une structure en métapopulations et non par bassin, des subtiles différenciations génétiques entre populations de certains bassins proches existent pour certaines espèces, comme entre le bassin Nord-Fidjien et Lau/Futuna (*A. kojimai* & *I. nautilei*) mais sont aussi retrouvées chez *A. strummeri* (cf Thèse de J.Castel), entre Manus et Woodlark (*I. nautilei*) et Woodlark avec les autres bassins (*S. tollmanni*).

Ces résultats permettent de réfuter l'hypothèse d'une structuration génétique de ces espèces par bassin ou chaque bassin serait complètement isolé des autres. Une telle structure aurait pu résulter de la fragmentation d'une ancienne zone d'accrétion beaucoup plus continue comme la dorsale sud-fidjienne. Cette dorsale étendue a mené à la formation du bassin des Salomon il y a 25 millions d'années lors des réorganisations tectoniques qui ont conduit à la formation des bassins actuels (Hall, 2002). Dans ce cas, l'isolement des faunes hydrothermales par bassin aurait dû se faire de façon graduelle avec de l'isolement par la distance, en contraste aux résultats obtenus ici. Néanmoins, cette zone contenant une barrière semi-perméable semble se situer au niveau des arcs volcaniques des Iles Salomon et de l'archipel des Vanuatu (Nouvelles-Hébrides) qui sont associés aux fosses océaniques de subduction. Ces arcs volcaniques constituent probablement un réservoir pour la propagation actuelle de cette faune à travers l'existence de sources hydrothermales actives associées aux calderas des nombreux volcans sous-marins présents. Cette hypothèse est d'ailleurs étayée par la découverte de sources hydrothermales sur l'arc de Futuna (présentement échantillonnées : cf (Konn et al., 2016)) mais également sur le volcan Nifonea (arc des Vanuatu :McConachy et al., 2005; Schmidt et al., 2017) ou sur l'arc des Salomon (Auzende et al., 1990). Dans d'autres régions du monde, des réorganisations tectoniques telles que les failles transformantes et/ou les microplaques, ont montré leur implication en tant que barrière physique à la dispersion (Hurtado et al., 2004; Matabos & Jollivet, 2019; Plouviez et al., 2009) ou, au contraire, à la mise en place de contacts secondaires entre des faunes préalablement isolées (Johnson et al., 2013; Plouviez et al., 2013).

Même si l'existence d'une zone de moindre perméabilité où une barrière au flux génique semble exister, toutes nos espèces n'ont pas une réponse similaire face à cette barrière en termes de différenciation génétique inter-métapopulations. Une graduation de la différenciation est en effet observée, avec trois des espèces (*I. nautilei*, *E. ohtai* et *S. tollmanni*) qui affichent des valeurs élevées de F_{st} par rapport à *A. kojimai*. Cette variation dans la différenciation des espèces pourrait être expliquée d'une part par une migration inégale face à cette barrière selon l'espèce considérée et d'autre part à des fortes différences dans les tailles efficaces des populations. De plus, les événements de l'histoire évolutive des populations ont pu se produire à différents moments, ce que nous discuterons dans la deuxième partie de la discussion.

Bien qu'il existe des similitudes générales dans la différenciation est/ouest entre espèces, des particularités notables à chacune sont observées. Tout d'abord, le cas de *S. tollmanni*, qui, bien que subdivisée en deux métapopulations, présente une zone de tension au niveau du bassin de Woodlark où les deux lignées génétiques fortement différenciées se retrouvent en syntopie. Cette population correspond à un mélange des deux lignées capables de s'hybrider (présence d'un individu correspondant à un hybride de première génération après reproduction (F1)). Cette découverte suggère que la barrière génétique observée chez toutes les espèces, ne semble pas nécessairement couplée à une barrière géographique, particulièrement chez ce gastéropode avec un début d'isolement reproductif entre les deux lignées avec des valeurs de divergence nette assez importantes (~ 1.2%). Néanmoins, cette situation pourrait aussi résulter d'un front d'expansion plus marqué et plus rapide chez cette espèce (Gilbert et al., 2022). Ce chevauchement de lignées divergentes avec un front d'hybridation est relativement bien connu chez les espèces hydrothermales associées aux dorsales médio-océaniques, notamment chez les *Bathymodiolus* (Johnson et al., 2013; O'Mullan et al., 2001) ou les gastéropodes patelliformes du genre *Lepetodrilus* (Matabos et al., 2008). A ce titre, la dorsale de Woodlark constitue également une zone de contact chez les deux lignées mitochondriales de *Lepetodrilus schrolli* (Poitrimol et al. Soumis) et que l'on retrouve également en mélange localement à l'entrée du bassin de Manus sur les volcans Susu avec le marqueur mitochondrial Cox 1 (Plouviez et al., 2019).

Chez *I. nautilei*, comme décrit dans le premier chapitre, chaque métapopulation apparaît subdivisée selon les différents bassins lorsque l'analyse génétique est menée seulement sur les locus « outliers ». Ces différenciations inter-bassins au sein d'une des deux métapopulations sont dues d'une part à des probables diminutions locales de la migration entre ces bassins et d'autre part à de l'adaptation locale ou à l'influence (introgression) de l'autre métapopulation. Les locus différenciés entre Manus et Woodlark s'alignent pour la moitié sur des gènes codant des enzymes de réparation et réplication de l'ADN, suggérant ainsi que cette différence de fréquences alléliques pourrait être potentiellement due à de l'adaptation locale et notamment au gradient de profondeur entre ces deux sites. A l'inverse, ceux qui différencient Nord-Fidji de Lau/Futuna sont en partie dus à de l'introgression d'allèles venant de Manus/Woodlark, et à une limitation des échanges entre les bassins de Nord Fidji et de Futuna/Lau. D'une façon similaire, *A. kojimai* et *A. strummeri* (thèse Jade Castel) montrent le

même patron de différenciation entre Lau et Nord-Fidji. Chez *A. kojimai*, la population Nord-Fidjienne affiche cependant une proportion du fond génétique Manus/Woodlark plus importante que celle de Futuna/Lau, suggérant aussi une introgression locale d'allèles issus de Manus/Woodlark, ce qui n'est pas le cas pour *A. strummeri* qui est restreinte aux seuls bassins Nord-Fidjien et la zone Lau/Futuna.

Ces mesures de différenciation indicatives des structures actuelles des populations peuvent être influencées d'une part par (I) les traits d'histoire de vie (mode de reproduction, dispersion) et d'autre part par (II) des facteurs environnementaux pouvant agir comme des barrières aux flux de gènes. Ainsi, des espèces présentant des structures génétiques similaires peuvent arborer des histoires évolutives nuancées.

Dans le milieu hydrothermal profond, des cas de dispersion longue distance ont déjà été décrits dans la littérature scientifique suggérant que des populations distantes de plusieurs centaines de kilomètres peuvent toujours être connectées entre elles par la phase larvaire (Young et al., 2012). Pour essayer de distinguer plus précisément quels sont les facteurs qui ont influencé l'histoire évolutive des espèces, nous avons réalisé une approche indépendante et multi-locus génomique de reconstruction démographique avec *dadi* à partir des spectres joint des fréquences alléliques repliés (JAFS). Cette approche démo-génétique appliquée de manière indépendante à chaque espèce, a permis de mettre en lumière que toutes les espèces ont ainsi divergé sous un modèle de contact secondaire (SC) à l'exception d'*I. nautiliei* où les trois modèles (IM/AM/SC) ne sont pas différenciable (cf Chapitre 1). Néanmoins elles présentent tout du flux génique avec une orientation similaire, mais dont la temporalité varie en fonction des espèces.

Dans un premier temps, nous allons discuter de l'intensité et l'orientation des flux géniques entre ces métapopulations et leur implication possible en termes de dispersion larvaire.

(II) DIRECTIONNALITE DES FLUX GENIQUES

Bien que les métapopulations Est (Manus/Woodlark, ci-après M/W) et Ouest (Nord-Fidji/Futuna/Lau, ci-après NF/F/L) soient plus ou moins différenciées pour toutes les espèces, et affichent une très faible proportion de fond génétique commun avec les analyses d'*ADMIXTURE* (exceptions faites d'*A. kojimai* et de l'analyse réduite aux marqueurs outliers pour *I. nautiliei*), les différentes analyses Treemix et F_3 , ainsi que les inférences démographiques avec *dadi*, décrivent l'existence d'un flux génique réduit entre ces deux métapopulations.

Les analyses Treemix ont mis en évidence du flux génique entre les métapopulations chez trois des quatre espèces (Figure 9). Sur ces trois espèces, seul *I. nautiliei* présente un flux génique orienté de M/W vers NF/F/L, alors que pour *A. kojimai* et *S. tollmanni*, ce flux est orienté dans la direction opposée. Complémentaires à ces analyses, les statistiques F_3 montrent que la population Nord-Fidjienne d'*I. nautiliei* est introgressée par des allèles en provenance de M/W et qu'inversement, les individus de *S. tollmanni* ayant le fond génétique Manus recevraient à Woodlark des allèles de la métapopulation NF/F/L (Figure SI 28 Annexe 2). Cette introgression s'accompagnerait de la présence et/ou de l'arrivée d'individus avec un fond génétique NF/F/L et de formation d'hybrides de première génération (F1).

Chez *A. kojimai*, les analyses Treemix affichent un flux génique de Nord-Fidji vers M/W. Cependant les individus de la population Nord-Fidjienne présentent une proportion non négligeable d'allèles issus de la métapopulation M/W dans les analyses *ADMIXTURE* d'une façon relativement semblable à l'introgression observée chez *I. nautiliei*, bien que les statistiques F_3 avec Nord-Fidji en population focale soient significativement positives, ne permettant ainsi pas de prouver qu'il n'y a pas d'introgression. Ces analyses, bien que fournissant des indications intéressantes sur l'orientation des flux géniques ainsi qu'un potentiel flux génique bidirectionnel, peuvent aussi paraître contradictoires et ne prennent pas en compte la temporalité de la divergence des populations. En effet, *A. kojimai* est l'espèce qui affiche la plus faible différenciation entre les populations, ce qui pourrait résulter d'une histoire évolutive récente ou/et a un fort flux génique.

Les reconstructions démographiques avec *dadi* permettent de prendre en compte la temporalité de la migration. Ces analyses indiquent que l'histoire évolutive des populations

de toutes les espèces s'ajuste mieux à un modèle avec une phase de contact secondaire après une période d'isolement strict dont la durée est variable en fonction des espèces. Quels que soit l'espèce ou le modèle choisi, une généralité semble se dessiner où le flux génique est bidirectionnel mais asymétrique et légèrement plus important lorsqu'il est dirigé de NF/F/L vers M/W (Tableau 4) ; Cela malgré l'existence d'une barrière géographique se trouvant dans la zone de l'arc volcanique des Vanuatu et Salomon, séparant ainsi Woodlark et Nord-Fidji, à l'exception de *S. tollmanni* pour qui cette barrière semble être au niveau du bassin de Woodlark. Cette barrière semblerait donc perméable au flux génique, permettant néanmoins une connectivité génétique entre métapopulations. De plus, *A. kojimai* présente une différenciation très faible dont l'origine pourrait être due à des forts taux de migration ou à une histoire de divergence très récente. Cette orientation du flux génique pourrait être en partie la résultante d'une stratégie similaire de dispersion ou/et influencée par les mêmes facteurs environnementaux.

La possibilité physique de connexions longue distance entre bassins a été testée par Mitarai et al. (2016) en simulant avec un modèle de courantologie les trajectoires larvaires à la profondeur de 1000 m. Ce modèle hydrodynamique montre que pour des durées de phase larvaire pélagique (PLD) longues (>190 jours), il est possible d'établir des connexions entre NF/F/L et M/W via des sources relai à proximité des Salomon et Vanuatu, mais avec une unidirectionnalité de l'Est vers l'Ouest. Ceci ne concorde pas totalement avec nos résultats d'inférence démographique qui décrivent l'existence d'un flux génique bi-directionnel et légèrement asymétrique. Le flux génique de M/W vers NF/F/L est en effet loin d'être négligeable chez toutes les espèces et l'introgession différentielle résultant d'un flux génique de M/W vers le bassin Nord-Fidjien est clairement observée chez *I. nautili* (chapitre 1) et possiblement chez *A. kojimai*. Ces résultats suggèrent donc que malgré une asymétrie Est-Ouest possiblement due à l'hydrodynamisme régional, le flux génique s'opèrerait dans les deux sens.

Ce flux génique, bien que faible est donc influencé par les caractéristiques des espèces en matière de mode de développement larvaire, des courants océaniques dans la région et donc de la position de la larve dans la colonne d'eau. De nombreuses espèces hydrothermales possèdent un développement lécithotrophe des larves avec des œufs de grande taille (càd. riches en vitellus ; (Tyler & Young, 1999, 2003)), qui serait plutôt l'apanage d'espèces dont les

larves disperseraient près du fond où les eaux sont oligotrophes et froides. Ce type de développement avec un développement fortement ralenti, voire stoppé pendant la phase migratoire constitue alors un avantage pour la dispersion à longue-distance, potentiellement orientée parallèlement à l'axe des dorsales (Chevaldonne et al., 1997; Young et al., 1994). Cependant, ce mode de dispersion peut être fortement contre-sélectionné dans un système de dorsales discontinues où les sites actifs peuvent être séparés par d'importantes étendues de plaines abyssales. Les courants ne sont alors pas forcément canalisés par la topologie du fond. De plus en plus d'études ont ainsi montré que certaines espèces peuvent aussi avoir des larves à développement planctonotrophe. Celles-ci seraient alors capables de remonter près de la surface pour être dispersées par les courants dont les vitesses sont généralement plus élevées (Arellano et al., 2014; Sommer et al., 2017; Yahagi et al., 2017). Dans nos cas, sur la base des connaissances existantes, *I. nautili* et *E. ohtai* possèderaient des larves lécithotrophes. De manière opposé, *A. kojimai* et *S. tollmanni* pourraient posséder des larves planctonotrophes. Les ovocytes sont en effet de petite taille (~ 70 µm de diamètre) chez *Alviniconcha* et pour une des espèces (*Alviniconcha hessleri*) la larve a déjà été identifiées à moins de 200 m de profondeur par barcode du méroplancton au large d'Hawaii (Sommer et al., 2017). De façon analogue, bien que la larve de *S. tollmanni* soit dans un premier temps encapsulé sur les coquilles d'*Ifremeria*, la larve véligère émise ensuite dans la colonne d'eau pourrait être similaire à celle de *Shinkailepas myojinensis*, qui elle remonterait en surface (Yahagi et al., 2017).

Cette bidirectionnalité asymétrique du flux génique orienté de la même manière entre *I. nautili*, *S. tollmanni* et *E. ohtai* par l'analyse *dadi*, suggère néanmoins que les différences de développement larvaire observées entre espèces n'impactent pas l'orientation principale de la dispersion, mais conduisent à une intensité variable de celle-ci. *A. kojimai* montre une quasi-symétrie dans le flux génique, bien que celui allant de NF/F/L vers M/W soit très légèrement supérieur que le flux génique opposé. Cette bidirectionnalité suggère donc que ces larves seraient potentiellement capables de migrer verticalement dans la colonne d'eau, tout en étant impactées non seulement par les courants profonds, expliquant ainsi le sens majoritaire du flux génique, mais aussi entraînées par les courants de surface pour expliquer le flux génique de M/W vers NF/F/L. Cette capacité de migration verticale a déjà été montrée chez d'autres espèces hydrothermales possédant des larves planctonotrophes comme les

gastéropodes *S. myojinensis*, *A. hessleri* ou encore les *bathymodiolo* (Arellano et al., 2014; Yahagi et al., 2017).

Nos résultats peuvent suggérer ainsi que les quatre espèces n'ont pas une stratégie de dispersion simplement unimodale (c'est-à-dire soit une dispersion à longue distance en surface ou une dispersion par des courants profonds permettant dans une moindre mesure l'autorecrutement ou la colonisation de proche en proche). Les stratégies révélées par les flux géniques apparaissent plus complexes et relativement semblables entre espèces. Cette complexité dans la dispersion pourrait s'expliquer d'une part par le fait que ces espèces sont sessiles et inféodées de façon stricte aux sources hydrothermales et d'autre part par l'instabilité et la fragmentation de ces dernières. Cela pourrait générer ainsi une forte pression de sélection pour permettre la colonisation d'une nouvelle source : dispersion sur le fond et autorecrutement lorsque les conditions sont bonnes localement, et capacité de dispersion à longue distance pour éviter les extinctions locales. Cela permet de suggérer que ces espèces peuvent présenter une plurimodalité de stratégies de dispersion dépendante des conditions environnementales dans lesquelles elles se trouvent. En effet, bien que très éloignés dans le règne du vivant, un certain nombre de plantes terrestres affichent une bimodalité dans leurs capacités de dispersion, avec du polymorphisme au sein des propagules où certaines sont capable de se disperser à longue distance, alors que d'autres ne se dispersent que localement (Brändel, 2007; Imbert, 2002; van Mølken et al., 2005; Venable & Levin, 1985). Ces différences dans les modalités de dispersion sont généralement sous l'influence de différents facteurs tels que la compétition interspécifique, la stabilité des patchs de la métapopulation et l'avantage à la colonisation ainsi que d'autres compromis sur des traits d'histoire de vie (Jakobsson & Eriksson, 2003; Lönnberg & Eriksson, 2013). En milieu marin, certains invertébrés sont déjà bien connus pour exhiber cette double stratégie de développement larvaire, appelée « poecilogonie », qui est due notamment à du polymorphisme dans le développement larvaire influencé par la saisonnalité. Quelques espèces possèdent cette particularité, dont l'annélide *Streblospio benedicti*, pour laquelle l'origine de ce polymorphisme larvaire proviendrait de différences génétiques pour quelques locus importants du développement et non à une histoire de divergence des populations (Zakas & Rockman, 2015). Antérieurement à cette dernière étude, Zakas & Hall (2012) ont suggéré que ce polymorphisme de développement larvaire et donc de stratégies de dispersion serait sélectionné pour le succès de survie dans un

environnement hétérogène. Dans le milieu hydrothermal, les espèces montrent un certain nombre de caractéristiques similaires leur permettant de recoloniser un habitat extrêmement mouvant où les espèces pourraient présenter des stratégies de dissémination de proche en proche lorsque l'habitat est disponible et des larves dites 'dormantes' capables de disperser à longue distance par des courants profonds ou de surface. A ce titre, Craig Young et collaborateurs (comm.pers. à D. Jollivet) suggèrent, à partir de données sur les isotopes de l'oxygène dans les coquilles larvaires (^{18}O) que ce genre de bimodalité dans la dispersion existe chez des moules hydrothermales pour lesquelles une faible proportion de larves reste près du fond alors que la majorité d'entre elles remontent à la surface pour disperser.

La position et la durée de séjour de la larve dans la colonne d'eau sont centrales pour permettre une dispersion longue distance et pourraient expliquer pourquoi *A. kojimai* est la seule des quatre espèces à afficher une très faible différenciation génétique entre les métapopulations M/W et NF/F/L. En effet, la dispersion larvaire longue distance est plus efficace dans les eaux de surface où la vitesse de déplacement des masses d'eau est plus grande. A l'inverse, les stratégies d'incubation des larves chez *I. nautiliei*, *E. ohtai* et d'encapsulation des larves chez *S. tollmani* pourraient laisser penser que chez ces espèces, une partie des larves pourraient rester plus facilement près du fond. Alors qu'une faible partie pourrait quant à elle participer à de la dispersion à longue distance, ce qui pourrait expliquer le flux génique entre métapopulations.

La patelle *S. tollmani* représente un cas très intéressant car, contrairement aux autres, sa barrière génétique ne concorde pas avec la zone où une barrière géographique se trouverait pour les autres espèces. Les deux lignées génétiques se retrouvent en syntopie au niveau du bassin de Woodlark et s'hybrident. Woodlark représenterait donc une zone de contact et de tension entre les deux lignées génétiques, l'une d'entre elle ayant donc réussi à franchir la barrière des îles Salomon / Vanuatu. Les différentes analyses ont montré que les individus de Woodlark génétiquement similaires aux individus de Manus (Woodlark2) sont introgressés par des allèles NF/F/L. Cette introgression pourrait venir des individus Manus et NF/F/L en mélange à Woodlark avec la présence d'individus hybrides de première génération (F1) seulement, suggérant un possible début d'isolement reproductif entre les deux lignées, ce qui est aussi étayé par une assez forte divergence nette de 1.2% entre les deux lignées. Les paramètres démographiques nous indiquent que Manus/Woodlark2 seraient possiblement

en déclin ($b_2 < 1$) alors que Nord-Fidji/Futuna/Lau/Woodlark1 serait en expansion ($b_1 > 1$), ceci étant recoupé par les informations d'introgression et de flux génique au niveau de la zone de contact. Ces résultats semblent suggérer que ces deux lignées ne sont pas à l'équilibre stable au niveau de Woodlark ce qui suggérerait le remplacement possible de la lignée Manus/Woodlark2 par la lignée Nord-Fidji/Futuna/Lau/Woodlark1 avec une contre-sélection pour les hybrides de première génération (F1). Cette expansion irait alors plutôt dans le sens Est vers Ouest accréditant le modèle actuel de dispersion larvaire de (Mitarai et al., 2016). Cependant, la découverte d'un individu hybride (F1) dans la zone de contact laisse aussi penser qu'une partie des larves produites recrutent localement à moins que cette zone de contact ne s'étende au-delà de la dorsale de Woodlark, ou provient d'une autre zone de contact proche.

Les différentes approches d'estimations du flux génique ont permis de mettre en évidence de manière indirecte des comportements potentiels de dispersion des espèces (migration verticale des larves, longue PLD ou stratégie mixte de dispersion) ainsi que les facteurs extrinsèques qui influencent potentiellement la connectivité à longue distance des populations (courants profonds et de surface). Il convient cependant de s'intéresser maintenant aux facteurs ayant entraîné l'isolement de ces (méta)populations à l'est et l'ouest de la région en examinant leur niveau de divergence. De nombreuses études de génomique de la spéciation ont mis en évidence l'importance de la temporalité des flux géniques dans la mise en place de la divergence des populations en cours d'isolement, tout en mettant en lumière le rôle de la sélection en liaison et l'apparition d'incompatibilités génétiques dans cet isolement (Roux et al., 2016). Les temps de divergence estimés concordent-ils avec la mise en place des bassins ou ont-ils été plutôt influencés par les cycles climatiques ?

Existe-t-il une temporalité de la divergence commune à toutes les espèces ?

(III) TEMPORALITE DE LA DIVERGENCE DES METAPOPOPULATIONS

Dans un premier temps, nous nous sommes intéressés au temps d'isolement des deux métapopulations et par la suite aux cas où cet isolement s'est effectué avec de la migration ou a été récemment interrompu par des remises en contact.

A l'exception de *A. kojimai*, le temps de divergence depuis la séparation des populations T total (Tsm :IM, Tam+Ts :AM, Ts+Tsc :SC) est du même ordre de grandeur, quelles que soient les espèces considérées (Tableau 3). Cette durée est de l'ordre de la centaine de milliers de générations pour un taux de mutation de 10^{-8} par site et par génération. Sur la base d'un temps de génération court de l'ordre de 1-2 ans caractérisant généralement les espèces à stratégie r des milieux hydrothermaux (Appeltans et al., 2012), ces temps d'isolement remonteraient au minimum à 100 000 ans, ce qui concorderait plus ou moins avec les derniers cycles glaciaires (Tableau 5 ; Barrows et al., 2011). Ces derniers ayant impacté le niveau des océans, cela aurait pu entraîner une modification des courants océaniques suivant l'expansion de la banquise de l'Antarctique il y a plus de 126 000 ans dans cette région du monde (Joy et al., 2014). En plus de ces variations du niveau de la mer, lors d'un interglaciaire la fonte du glacier Mont Giluwe en Papouasie Nouvelle-Guinée a pu s'accompagner d'arrivées importantes d'eau douce et de produits terrigènes (érosion) en milieu marin qui pourraient avoir perturbé plus localement la circulation marine et la dispersion des larves.

Tableau 5 : Chronologie de l'expansion du glacier sur le volcan Mont Giluwe (Papouasie – Nouvelle Guinée) à la fin de Pléistocène (d'après Barrows et al., 2011).

Nom de la glaciation	Age (ka)
Gogon	293 - 306
Mengane	136 - 158
Komia	62
Tongo	>20,3 - 11,5

Cependant, l'estimation du nombre de générations est basée sur un taux de mutation par site et par génération de 10^{-8} , valeur choisie en l'absence d'un taux connu chez nos espèces. Les valeurs de taux de mutation peuvent néanmoins varier d'un facteur 10 selon les espèces (10^{-10} à 10^{-8} ; Lynch 2010). L'estimation du taux de mutation mitochondrial (*Cox 1*) chez les espèces hydrothermales est de 0,2-0,3% pour les invertébrés (Breusing et al., 2020; Chevaldonné et al., 2002; Johnson et al., 2006; Matabos & Jollivet, 2019), au lieu de près de 10 fois plus (1,2-2%) chez les espèces trans-panaméennes (Knowlton & Weigt, 1998). Si l'on considère un taux de 10^{-9} , les temps estimés avec *dadi* sont alors autour du million de générations, soit des temps de séparation estimés à un ou deux millions d'années, plus en phase avec l'ouverture et le début des accrétions des bassins de Lau et Nord-Fidjien (Tableau 6). La formation des bassins arrière-arc actuels dans le Pacifique Ouest a été progressive entre 1 à 7 millions d'années avec des histoires tectoniques complexes et étroitement imbriquées (Hall 2002, Shellart et al 2006).

Tableau 6 : Récapitulatif des temps de formation et d'accrétion de certains grand ensembles géologiques du pacifique Ouest (tiré et adapté de Schellart et al 2006). Ma : millions d'année

Ensemble géologique	Temps de formation	Temps d'accrétion	Références
Bassin de Lau	7-6 Ma	5-6 Ma (dépend de la zone)	(Parson & Hawkins, 1994; Ruellan & Lagabrielle, 2005)
Bassin de Manus	~ 1 -2 Ma	0,7-1 Ma	(Martinez & Taylor, 1996)
Bassin de Woodlark	5-7 Ma	3,5 Ma (dépend de la zone)	(Benes et al., 1994; Weissel et al., 1982)
Bassin Nord-Fidjien	12 Ma début rotation et 6 Ma	3 Ma (zone centrale mais d'anciennes zones d'accrétions au nord)	(Auzende et al., 1990; de Alteriis et al., 1993; Ruellan & Lagabrielle, 2005)
Nouvelle-Hébrides/Vanuatu	Fin Miocène 5,3 Ma avec la rotation, Nord-Fidjien		(Greene et al., 1994; Musgrave & Firth, 1999)
Arc Volcanique Futuna	2,5- 0,78 Ma	2,5- 0,78 Ma	(Szitkar et al., 2020)

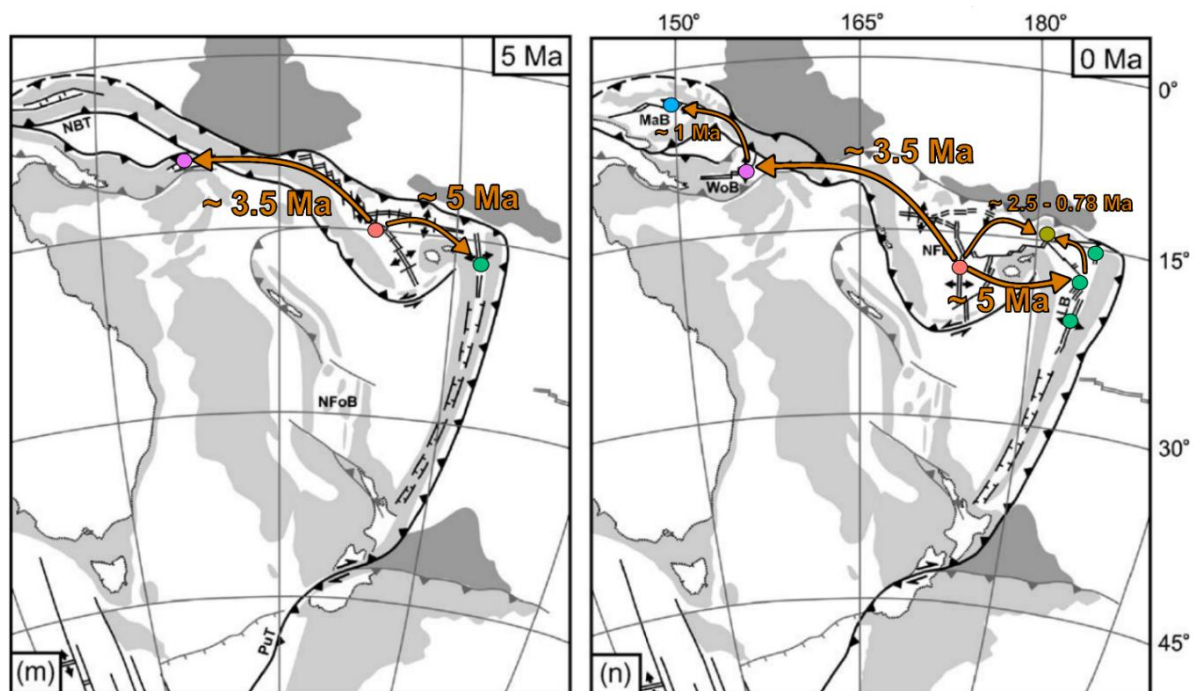


Figure 12 : Reconstruction hypothétique de l'histoire récente de la colonisation des sources hydrothermales pour nos quatre espèces, au cours des 5 derniers millions d'années. (Adapté de Schellart et al 2006 et de Boulart et al 2022). Les flèches orange représentent les possibles voies de colonisation et datation en lien avec le début de l'accrétion des bassins arrière-arc.

Cette concordance possible entre le début de l'accrétion des bassins de Nord-Fidji et Lau et le temps de séparation des deux métapopulations pourrait laisser suggérer que cette divergence aurait pu commencer à se mettre en place lors de la colonisation des derniers bassins formés à travers des zones relais et notamment l'expansion de la proto-dorsale nord-fidjienne (Figure 12) (Boulart et al. 2022). Breusing et al. (2020) ont effectué des datations moléculaires des événements de spéciation ayant conduit à la formation des différentes espèces du genre *Alviniconcha* à partir de données multilocus calibrées par des données fossiles. Ils ont ainsi proposé que le dernier événement de spéciation au sein de ce genre ayant séparé *A. kojimai* de l'espèce *A. hessleri*, actuellement située dans le bassin des Mariannes, daterait de moins de 10 millions d'années et serait probablement concomitant de la mise en place du système actuel de dorsales du Pacifique Ouest. D'un autre côté, la séparation des populations Est et Ouest d'*A. kojimai* apparaît beaucoup plus récente que d'avec les autres espèces et la divergence pourrait avoir commencé durant les dernières périodes glaciaires (entre 10 000 et

100 000 ans selon le taux de mutation et de génération choisis). Chez cette espèce, on pourrait alors penser que la colonisation des bassins arrière-arc du Pacifique Ouest serait très récente, rapide et pourrait ne pas avoir la même origine que les autres.

Bien que les métapopulations Est et Ouest (*A. kojimai* excepté) se soient sans doute séparées suite à un événement vicariant (en supposant un taux de mutation équivalent pour toutes les espèces), les temporalités estimées des phases avec migration ou depuis un contact secondaire apparaissent comme très variables d'une espèce à l'autre. Dans tous les cas de contact secondaire, *I. nautiliei* et *A. kojimai* montrent une faible période d'isolement strict (T_s entre 41 et 16 000 générations pour *I. nautiliei* et 2 680 pour *A. kojimai*). Cependant sous l'hypothèse du contact secondaire *I. nautiliei* aurait eu une beaucoup plus longue période d'isolement avec migration (T_{sc} entre 54 000 et 120 000 générations comparé à *A. kojimai* ($T_{sc} \sim 1000$ générations)). *I. nautiliei* est cependant la seule espèce pour laquelle un modèle AM est équiprobable et dans ce cas, la séparation des populations aurait été très progressive avec un isolement strict très récent, qui le rend très similaire au modèle IM.

En comparaison, *E. ohtai* et *S. tollmanni* ont divergé de façon stricte pendant beaucoup plus longtemps. Chez *E. ohtai*, T_s est estimé à environ 46 000 générations, tout en arborant un contact secondaire d'une durée aussi importante d'environ 70 000 générations. Cette dernière pourrait être aussi potentiellement associée à la colonisation de la dorsale la plus récente du bassin de Manus au début de son accréation, il y a 780 000 ans (Martinez & Taylor, 1996).

L'espèce *S. tollmanni* montre elle aussi une durée d'isolement strict importante ($T_s \sim 95 000$ générations) avec un contact secondaire très récent (entre 8 300 et 10 500 générations). La temporalité du contact secondaire pour *S. tollmanni* concorderait avec les derniers cycles glaciaires (8 300 – 10 500 ou 83 000 - 105 000).

Ces résultats mettent en lumière des histoires de divergence très différentes en termes de contacts secondaires malgré une histoire de vicariance semblable si l'on exclut *A. kojimai*. Néanmoins, il est nécessaire d'augmenter le nombre des espèces cibles pour obtenir une image plus claire de ces temporalités à l'échelle des communautés d'espèces. A ce titre, cette étude devrait être renforcée par l'analyse en cours d'espèces supplémentaires comme les bivalves *Bathymodiolus manusensis* et *septemderium*, le gastéropode *Lepetodrilus schrolli* et

au moins deux espèces de polychètes de la famille des Polynoidae, dont l'une est associée aux *Bathymodiolus*. Les travaux déjà réalisés sur *L. schrolli* présentent un certain nombre d'analogies avec *S. tollmanni* du point de vue des échanges entre métapopulations Est et Ouest (Plouviez et al., 2019) et du positionnement d'une zone de contact à Woodlark (Poitrimol et al. Soumis) ce qui renforce notre modèle de vicariance.

Une autre approche qui reflète la temporalité de cette divergence est la mesure de divergence nette (D_a). Dans notre cas, trois des quatre espèces (*I. nautilei*, *A. kojimai*, *E. ohtai*) ont une divergence nette inter-métapopulations (D_a) relativement faible entre 0,6-0,8% bien qu'affichant des fortes différences en termes de différenciation (F_{st} 0,387 - 0,012). A l'inverse, *S. tollmanni* a une divergence nette élevée à 1.2% en cohérence avec la valeur élevée de différenciation ($F_{st} = 0,271$). L'analyse conjointe des données de divergence et de différenciation génétique aux locus nucléaires mais également mitochondriaux pose cependant question. *A. kojimai* qui présente une divergence relativement importante (0,67%) entre métapopulations n'est que très peu différenciée à cette même échelle d'observation ($F_{st} = 0,012$). A contrario, *I. nautilei* et *E. ohtai* pour des divergences nettes (D_a) semblables ont de fortes valeurs de différenciation ($F_{st} = 0,27$) que l'on retrouve également au niveau mitochondrial (Thaler et al., 2011) pour *Ifremeria* mais pas pour *Eochionelasmus* ($D_{xy} Cox 1 = 0 - 0,9\%$: Herrera et al., 2015). Cette disparité entre gènes nucléaires et mitochondriaux est également retrouvée pour *S. tollmanni* pour laquelle une forte différenciation est détectée sur les locus nucléaires alors qu'elle est inexistante à la même échelle sur les données mitochondriales (Yahagi et al., 2020 , Poitrimol et al. soumis). Néanmoins, les valeurs de F_{st} et D_a sont cohérentes au niveau nucléaire pour *S. tollmanni*, pour laquelle un possible début d'isolement reproducteur entre les deux clades génétiques en sympatrie est observé. La non-concordance avec les valeurs de différenciation-divergence génomique, pourrait provenir de plusieurs points. L'effet de la sélection en liaison pourrait expliquer en partie la différence entre F_{st} et D_a , cet effet est omniprésent à différent degrés dans nos données à l'exception d'*E. ohtai*. Cet effet peut influencer les estimations de F_{st} , à cause d'une forte hétérogénéité des taux de recombinaison au sein du génome, (Charlesworth, 1998; Charlesworth et al., 1997) .

C'est ce qui semble être observé entre les mesures de différenciation (F_{st}) et la quantification de l'effet de la sélection en liaison (Q) dans nos modèles démo-génétiques et pour nos

espèces, le degré d'influence de la sélection en liaison à la même gradation que celle des valeurs de F_{st} , c'est-à-dire, *I. nautiliei* avec Q entre, 0,52 et 0,56, *S. tollmanni* avec Q entre 0,416 et 0,436 et *A. kojimai* avec Q = 0,301. Une autre possible explication serait que l'estimation de la diversité nucléotidique pourrait être sous-estimée, en raison de la perte allélique lors de la sélection des fragments d'ADN et l'assemblage *de novo* (Arnold et al., 2013; Gautier et al., 2013).

2. Comparaison des résultats avec la littérature existante

(IV) STRUCTURATION ET FLUX GENIQUE DES GENOMES NUCLEAIRES ET MITOCHONDRIAUX

Les études antérieures ont principalement utilisé des marqueurs mitochondriaux ou/et quelques marqueurs génomiques (gènes, amplicons, microsatellites) pour mettre en évidence de manière indépendante des patrons de différenciation génétique pour les mêmes espèces. Bien que ces marqueurs puissent être influencés par différents processus, ils restent néanmoins très informatifs dans les approches de phylogéographie comparative, permettant de donner un premier aperçu des ensembles géographiques et d'identifier les barrières à la migration. Cependant, de nombreuses études de génétique comparative (mtDNA, cpDNA) ont montré qu'il existait des discordances entre le génome des organelles (mitochondries et chloroplastes) et le génome nucléaire. Ces discordances peuvent être la conséquence de différents mécanismes comme l'hybridation (capture), la sélection diversifiante, les extinctions non aléatoires de lignées, des captures d'allèles exogènes par balayage sélectif entre espèces (Toews & Brelsford, 2012). Dans la suite de cette partie, nous allons discuter espèce par espèce des résultats issus des marqueurs nucléaires et mitochondriaux, ainsi que des données génétiques d'autres espèces hydrothermales qui n'ont pas encore été analysées avec des marqueurs RAD.

Comme précédemment discuté dans le chapitre 1 sur *I. nautiliei*, Thaler et al. (2011) ont montré que cette espèce était bien structurée entre le bassin de Manus et les bassins Lau/Nord-Fidji, avec quelques individus de Lau/Nord-Fidji porteurs d'haplotypes *Cox 1* typiques de Manus. Cette observation a été confirmée par Boulart et al (2022) qui ont de plus rajouté la population de Woodlark. Ils ont montré la présence d'haplotypes Manus/Woodlark dans des individus échantillonnés à Nord-Fidji/Futuna/Lau. Ces travaux sont en accord avec

nos résultats sur le génome. On peut cependant souligner des signaux d'introgression différentielle au niveau du bassin nord-fidjien, ce qui confirme un flux génique existant entre métapopulations.

Cette concordance entre génomes mitochondrial et nucléaire n'est pas toujours évidente pour les autres espèces.

Chez *A. kojimai*, Breusing et al. (2020) ont montré qu'il existait des subdivisions subtiles au sein de Manus et de Nord-Fidji/Lau sur quelques gènes nucléaires et mitochondriaux sans pour autant observer une nette différenciation entre ces deux métapopulations. En réutilisant ces mêmes données, Breusing et al. (2021) ont pu néanmoins montrer qu'*A. kojimai* était structuré entre Manus et Lau (F_{st} nucléaire = 0.034 – 0.1095), dépeignant une très faible connectivité entre populations, en contraste avec nos données RAD. Cependant, les données mitochondriales des travaux précédents (Kojima et al., 2001) et plus récemment Boulart et al (2022 : suppl. mat.), ont montré sur la base du Cox 1 que cette espèce n'affichait pas de structure génétique claire à l'échelle du Pacifique Ouest.

L'absence de différenciation spatiale au niveau du marqueur mitochondrial *Cox 1* est donc peu compatible avec nos données RAD qui ont pu détecter une structure en deux sous-ensembles ainsi qu'une légère différenciation entre les bassins de Lau et Nord-Fidji. Cette disparité peut sans doute s'expliquer par un pouvoir de résolution plus fin de nos marqueurs nucléaires dans la discrimination d'une histoire de divergence des populations très récente et/ou d'un flux génique très important entre les populations. Cependant notre étude et celle de Breusing et al. (2021) montrent des orientations de flux opposé. Cela pourrait s'expliquer par les marqueurs nucléaires conservés utilisés par cette étude n'ont sans doute pas pu capter l'information du flux de génique contemporain entre bassins. En effet, l'estimation du flux génique a été réalisée avec très peu de marqueurs génétiques, la plupart étant des gènes codant pour des protéines fonctionnelles, donc potentiellement sous l'effet d'une sélection purifiante plus ou moins marquée.

De la même façon, les travaux antérieurs ont montré que *S. tollmanni* ne présentait aucune structuration génétique à l'échelle régionale sur le marqueur mitochondrial *Cox 1* (Yahagi et al., 2020; Poitrimol et al, soumis) alors que les données RAD montrent de façon non-ambigüe que l'espèce est sub-divisée en deux ensembles génétiques en voie d'isolement reproductif

avec une zone d'hybridation à Woodlark. Ces non-concordances entre nos données génomiques et mitochondriales sont difficiles à expliquer par la seule absence d'une résolution suffisante sur le génome mitochondrial eut égard à la forte différenciation génétique et les valeurs de divergence observées sur le nucléaire. Une première explication pourrait être la possible capture mitochondriale d'une lignée par l'autre bien que ce genre de capture s'accompagne en général d'une perte de diversité génétique liée à l'effet de balayage associé (Currat et al., 2008).

Néanmoins, un balayage trans-lignées mitochondriales impliquerait un flux génique entre les deux lignées entre tous les bassins. Or, chez *S. tollmani*, le flux génique est principalement observé à l'échelle génomique au niveau de la zone de contact où la lignée Nord-Fidji/Futuna/Lau/Woodlark1 est en expansion alors que l'autre semble en déclin. Cette situation assez complexe nous laisserait imaginer que ces deux lignées partiellement isolées génétiquement, se soient possiblement retrouvées en compétition lors de leur remise en contact dans le bassin de Woodlark avec une des lignées mitochondriales qui se serait répandue bien plus rapidement que le reste du génome.

Le gastéropode patelliforme *Lepetodrilus schrolli* qui vit de manière analogue aux *S. tollmani* sur les grands gastéropodes et moules hydrothermales présente lui aussi une structuration des populations en deux lignées (Manus vs. Lau) tout en affichant un flux génique bidirectionnel asymétrique vers l'Est (de Manus vers Lau) (Plouviez et al., 2019). Cette orientation du flux est elle aussi plutôt opposée à celle que l'on observe avec nos données génomiques pour *S. tollmani* et *A. kojimai*. Néanmoins, Poitrimol et al (soumis) montrent que les individus des deux lignées génétiques différenciées sont également en mélange dans le Bassin de Woodlark mais également à l'entrée du bassin de Manus sur les volcans Susu (Plouviez et al., 2019). Ce dernier résultat permet de mieux appréhender le cas de *S. tollmani*, en proposant soit l'existence d'une barrière commune à ces deux espèces et sans doute une histoire évolutive similaire sans qu'on sache si les lignées de cette espèce sont capables de s'hybrider. Dans les deux cas, une approche génomique plus approfondie permettrait de mieux interpréter ces résultats et de confirmer si ces espèces ont une histoire évolutive commune.

Des espèces relativement proches et vivant aussi au sein des sources hydrothermales peuvent présenter des différences importantes de patrons de structuration génétique sur la base du marqueur mitochondrial *Cox 1*. Des études sur d'autres espèces récoltées sur les mêmes sites suggèrent qu'un certain nombre de gastéropodes semblent soit privilégier la non-dispersion et d'autres la dispersion longue distance, soit ont eu une histoire évolutive complètement différente, notamment pour les genres *Shinkailepas* et *Symmetromphalus/Lamellomphalus* (Poitrimol et al. soumis). De la même façon, chez les annélides Polynoidae, certaines espèces ont une distribution très large alors que d'autres sont limitées à certaines zones géographiques (Chabert, stage de M2).

Chez le Polynoidae *Branchinotogluma trifurcus* (Chabert, stage de M2), la crevette hydrothermale *Rimicaris variabilis* et le crabe *Austinogrea alayseae* (Chabert, stage de M2) (Lee et al., 2019) on retrouve une structuration très similaire à ce que l'on a observé chez *I. nautilei* et *A. kojimai*, avec deux lignées génétiques distinctes (Manus/Woodlark et Nord-Fidji/Futuna/Lau) et quelques individus Nord-Fidjiens ayant la signature génétique Manus/Woodlark. Néanmoins, d'autres espèces, comme le Polynoidae *Branchinotogluma segonzaci* n'affichent aucune structuration. Cette situation similaire à *S. tollmani* ou *A. kojimai* laisse supposer qu'ils ont un fort potentiel dispersif, ou sont de récents colonisateurs de la région. Nos données génétiques montrent néanmoins que cette apparente homogénéité génétique pourrait être due à une capture mitochondriale ou un comportement particulier de la mitochondrie dans les fronts de migration, comme nous l'avons observé pour *S. tollmanni* pour laquelle les données nucléaires montrent clairement deux lignées distinctes.

(V) COMPARAISON AVEC D'AUTRES ETUDES PHYLOGEOGRAPHIQUES DU MONDE HYDROTHERMAL

La situation que nous avons observée avec des lignées distinctes n'est pas spécifique au système discontinu formé par les bassins arrière-arc. En effet, une étude sur le système linéaire de la dorsale du Pacifique Oriental (« *East Pacific Rise* », EPR) avec des discontinuités formées par les failles transformantes et les microplaques montre des similitudes avec notre situation (Plouviez et al., 2009). Cette étude a comparé à partir du gène mitochondrial *Cox 1*, plusieurs espèces arborant des traits d'histoire de vie contrastés et appartenant à différentes

familles, allant de gastéropodes (*Lepetodrilus elevatus*, *L. ovalis*, *Eulepetopsis vitrea*) à des vers polychètes (*Alvinella pompejana*, *Hesiolyra bergi*, *Branchiopolynoe symmytilida*) et aux bivalves (*Bathymodiolus thermophilus*). Dans cette approche comparative, les auteurs ont mis en évidence des patrons de structuration génétique similaires entre espèces, bien que des différences sur le positionnement des barrières puisse apparaître selon les espèces et les gènes étudiés (Plouviez et al., 2010, 2013). Hormis ces exceptions, la plupart des espèces affichent deux lignées génétiques distinctes et assez bien différenciées le long de la dorsale de part et d'autre d'une grande zone de transition se positionnant entre 17°S et l'Équateur. Malgré ces similitudes, ces auteurs ont montré et confirmé que toutes les espèces arborent de subtiles différences dans leur structure géographique intra-lignée génétique, allant d'une faible différenciation au sein d'une lignée à l'absence de différenciation le long de la dorsale, qui ont été déjà observées chez certaines espèces (Hurtado et al., 2004; Shank & Halanych, 2007).

De manière complémentaire avec une méthode de calcul bayésien par approximation n (ABC), les auteurs ont estimé les temps de divergence moléculaire sur la base du *Cox 1*, un premier événement de divergence génétique entre ces deux lignées est ainsi estimé à 1.3 million d'années (Ma) pour toutes les espèces, cependant pour une des espèces (*L. elevatus*), un deuxième événement de spéciation plus ancienne est daté à 11.6 Ma, formant des espèces "cryptiques". En recoupant ces résultats, les auteurs ont pu reconstruire une possible histoire phylogéographique de plusieurs espèces hydrothermales de la Dorsale Est Pacifique, suggérant ainsi que ces divergences auraient pu être en lien avec des événements tectoniques. La divergence la plus ancienne estimée à 11.6 Ma pour seulement *L. elevatus*, pourrait être en lien avec l'importante réorganisation géologique de la partie nord de la dorsale Est Pacifique, par la formation d'un système de dorsale parallèle et de microplaque il y a 12.5-11 Ma. Ceci a été confirmé plus tard par Matabos & Jollivet (2019) alors que l'évènement commun de divergence considéré à 1.3 Ma pourrait être la conséquence de la formation d'une série de failles transformantes entre 14°S et 5°S et d'un effet gyrotoire des courants profonds à l'équateur.

Ces précédentes études ont permis de mettre en avant d'une part l'importance d'entités géologiques ainsi que les conséquences associées de leurs formations et des discontinuités des courants sur la divergence des populations hydrothermales le long d'une dorsale

océanique. Cela montre que non seulement des facteurs extrinsèques (barrière géologique et courants) ont un rôle dans l'isolement des populations et les mécanismes de spéciation mais que la dynamique elle-même de déplacement de l'activité hydrothermale à l'axe des dorsales pourrait aussi moduler les échanges génétiques au sein des métapopulations en favorisant tantôt des isolats tantôt des remises en contact de ces derniers (Jollivet et al., 1999).

Et la spéciation dans tout ça ?

Le concept d'espèce fournit une définition simple pour les lignées évolutives suffisamment éloignées. Cependant pour le cas de lignées génétiquement proches n'ayant pas suffisamment divergé, cette notion devient difficilement applicable. En effet, si ces deux lignées se retrouvent en contact, elles peuvent toujours s'échanger des gènes par hybridation. La zone de contact géographique entre ces deux lignées forme ce qu'on appelle une zone d'hybridation où se trouvent de part et d'autre les populations parentales distinctes. Il a été montré que ces zones hybrides sont assez fréquentes dans la nature (Harrison, 1990 ; Hewitt, 2001, 2011). D'un point de vue génétique, le processus de divergence conduisant à la spéciation se traduit par une accumulation progressive de locus barrière contribuant à l'isolement reproductif, menant ainsi à des situations intermédiaires qui constituent une zone grise dans laquelle la spéciation n'est pas achevée, ceci expliquant que la nomenclature des espèces est souvent controversée (Roux et al., 2016).

Qu'en est-il des espèces hydrothermales du Pacifique Ouest ?

Pour les quatre espèces étudiées, nous avons identifié des lignées génétiques distinctes, qui affichent actuellement une assez faible divergence nette (D_a), se trouvant probablement au tout début de cette zone grise, à l'exception de la patelle *S. tollmanni* pour qui le degré de divergence est plus fort et serait donc probablement plus avancée dans ce processus de spéciation.

Nos espèces affichent le même modèle de divergence (contact secondaire, à l'exception *I. nautiliei*) dont la temporalité de la remise en contact varie entre espèces. Ces inférences ont aussi mis en évidence l'important effet de la sélection en liaison mais aussi la présence de certains locus barrières.

E. ohtai représente aussi un cas particulier car cette espèce n'affiche pas une forte divergence entre ses populations mais la moitié des locus apparaissent dans les reconstructions démogénétique comme ayant un taux de migration réduit.

De son côté, la forte divergence de *S. tollmanni* pourrait s'expliquer par une plus longue durée d'isolement strict qui aurait permis une accumulation plus importante de locus barrières entre les deux lignées ce qui explique leur sympatrie partielle (Woodlark). Leur divergence n'est cependant pas suffisante pour empêcher la formation d'hybrides (F1) dans la zone de contact.

Une autre patelle *L. schrolli* affiche aussi cette zone de contact pour les lignées mitochondriales (Poitrimol et al., soumis). Le bassin de Woodlark représenterait ainsi une zone charnière de contact pour ces espèces. Pour les autres espèces, la zone de contact hypothétique se trouverait quelque part au niveau des îles Vanuatu et Salomon, où l'existence des sources et communautés hydrothermales est avérée. Ainsi, il serait intéressant dans le futur d'obtenir des échantillons de ces zones charnières.

DISCUSSION GENERALE

Comparaison avec les modèles phylogéographiques Atlantique/Méditerranée

Il existe un nombre important d'études et d'approches de phylogéographie comparée accompagnées d'une grande variété de réponses en fonction des facteurs intrinsèques et extrinsèques qui contrôlent le devenir d'une espèce et de ses populations. Cette diversité de facteurs génère ainsi une grande diversité de patrons, avec des similitudes et discordances génétiques entre espèces. Ces études ont ainsi permis de documenter un certain nombre de ruptures phylogéographiques et de définir des provinces biogéographiques, particulièrement dans le monde marin où ces dernières ne sont pas évidentes à définir (Bowen et al 2016). Un certain nombre de ruptures biogéographiques existent dans tous les océans, allant de modification de front polaire à la fermeture ou l'ouverture d'isthmes ou de détroits il y a plusieurs millions d'années, jusqu'aux modifications climatiques du quaternaire (Avice, 2000; Hewitt, 2000; Johannesson et al., 2020; Olsen et al., 2004). Au sein de ces nombreux exemples, celui de la fracture phylogéographique Atlantique/Méditerranée affiche un certain nombre de similitudes avec notre étude.

Cette zone est bien documentée pour histoire géologique avec la ré-ouverture de la connexion avec l'Atlantique au niveau du détroit de Gibraltar il y a 5.33 millions d'années, mettant fin à la crise saline messinienne. Cette ré-ouverture a permis la colonisation et la dispersion d'une faune marine dans la mer Méditerranée, menant à la formation de nouvelles populations. Ainsi, de nombreuses études génétiques se sont intéressées à comprendre la structure génétique des espèces atlantico-méditerranéennes de part et d'autre de cette ancienne barrière qu'est Gibraltar. Ces études ont montré que de nombreuses espèces sont structurées en deux lignées génétiques, une méditerranéenne et une atlantique affichant une forte variabilité des divergences atlantico-méditerranéennes, allant d'espèces n'affichant aucune structure à des espèces cryptiques ayant initié leur isolement reproducteur (Patarnello et al., 2007). Cette zone de transition d'une lignée génétique à l'autre, ne semble pas concorder avec le détroit de Gibraltar mais plutôt au front Almeria-Oran, à l'Est de Gibraltar en mer d'Alboran,

bien que pour certaines espèces cette zone de transition se place à l'Ouest de Gibraltar dans l'Océan Atlantique. Après une revisite des données moléculaires, les auteurs ont estimé que la temporalité de cette divergence tombe dans une fenêtre temporelle en concordance avec les derniers cycles glaciaires du Quaternaire, bien plus récente que la fin de la crise messinienne, montrant ainsi que les cycles glaciaires (par leur effet sur le niveau des océans) ont aussi impacté l'histoire évolutive des espèces atlantico – méditerranéennes.

Des études récentes en génomique des populations s'intéressant plus précisément à des poissons téléostéens marins ont montré que malgré des fortes divergences génomiques (0,01 - 1,06%), un certain nombre d'espèces, même relativement divergentes, présentent une introgression génétique d'une lignée à l'autre (souvent de l'Atlantique vers la Méditerranée) au niveau d'une zone de contact (Patarnello et al., 2007 ; Thèse P.Barry). Plus particulièrement, pour certaines espèces, l'étude de l'histoire de la différenciation Atlantique/Méditerranée a mis en évidence des patrons hétérogènes le long du génome mettant ainsi en lumière des îlots génomiques de forte divergence qui ont probablement émergé durant la période allopatrique par l'effet de la sélection en liaison et contenant probablement des locus barrières (Duranton et al., 2018; Tine et al., 2014). Bien que ces études se focalisent principalement sur l'étude des mécanismes de spéciation le long du génome, ces analyses ont permis d'estimer la temporalité des arrêts et des reprises d'échanges génétiques entre différentes populations géographiques.

Cette très brève description et comparaison de la fracture phylogéographique des espèces Atlantique/Méditerranée avec celle de nos espèces hydrothermales montre bien que malgré l'hypothèse d'un effet géotectonique sur l'évolution des espèces, celle-ci ne reflète pas nécessairement l'histoire évolutive récente des espèces et des populations hydrothermales, tout mettant en lumière l'importance d'autres facteurs tels que les cycles climatiques et nuancé par les traits d'histoire de vie dans la compréhension des patrons phylogéographiques d'une région donnée.

Biologie de la conservation, connectivité génétique, démographique et résilience

Il est désormais bien connu que les changements écologiques dus à l'activité humaine ont accru les menaces pesant sur les populations et les espèces naturelles. Ces nouvelles pressions, allant de la destruction de l'habitat à l'introduction d'espèces non indigènes, en passant par la surexploitation des ressources ont généré une forte augmentation des pressions sur les espèces et populations naturelles, résultant ainsi en une augmentation du taux d'extinction de ces dernières. Ces changements ont déjà mené à une perte de la biodiversité et considéré comme la « 6^{ème} extinction de masse » dans l'histoire de la Terre (Ceballos et al., 2015). Cette perte de biodiversité est un problème central en biologie de la conservation qui cherche à mieux protéger les populations naturelles et espèces en utilisant différents outils et champs disciplinaires scientifiques variés, allant de l'écologie à l'utilisation de méthodes génétiques par exemple. Ces derniers ont déjà prouvé leur utilité pour la compréhension et identification d'unités de gestion, mais aussi pour mieux évaluer la taille, la connectivité des populations, l'hybridation d'espèces, le nombre d'espèces cryptiques et les cas d'adaptation (Allendorf et al., 2010; Funk et al., 2012; Hohenlohe et al., 2021; Palsbøll et al., 2007). Ces informations populationnelles constituent un des questionnements fondamentaux de la génétique des populations, mais aussi tout autant en biologie de la conservation qui consiste à d'abord identifier les populations, comprendre leur structure et connexion entre elles. La caractérisation de la variabilité et diversité génétique intra et inter-populations est aussi un autre point clé d'étude pour mieux gérer les espèces et leurs populations. La délimitation des populations naturelles (structure et connectivité) n'est cependant pas évidente. En effet, la notion de connectivité semble évidente mais sa quantification ne l'est pas. Il est ainsi nécessaire de distinguer la notion de connectivité génétique et démographique, bien qu'elles soient intrinsèquement liées (Crooks & Sanjayan, 2006; Lowe & Allendorf, 2010). Cette première représente le nombre de migrant effectif ou absolu entre les populations (c'est-à-dire l'apport d'information génétique extérieur à la

population réceptrice), alors que la connectivité démographique est la résultante des contributions relatives des différentes dynamiques entre populations, comme le taux de dispersions, le recrutement local ainsi que la croissance des populations, impliquant ainsi une non-indépendance entre deux populations lorsqu'il existe une connectivité démographique entre ces dernières. Ces deux concepts, comme dit précédemment sont intrinsèquement liés, surtout dans le cas d'une connectivité démographique implique une connectivité génétique, bien que le cas inverse ne le soit pas nécessairement.

La génomique des populations, bien que son objectif principal soit d'étudier les processus évolutifs façonnant l'architecture génomique, peut permettre d'estimer la connectivité génétique par l'utilisation un nombre très important de marqueurs neutres et potentiellement sous sélection (Gagnaire, 2020; Gagnaire et al., 2015). L'utilisation de ces deux « types » de marqueurs a déjà permis de mettre en lumière des limitations de connectivité et des délimitations cryptiques de populations chez plusieurs espèces (Milano et al., 2014; Wyngaarden et al., 2017). Néanmoins ces approches sont généralement réalisées à l'échelle d'une espèce, alors que dans certains cas, ce sont les écosystèmes ou un ensemble d'espèces ou populations qui sont à gérer. C'est particulièrement le cas des écosystèmes hydrothermaux profonds du Pacifique Ouest qui arborent une faune riche et endémique avec des espèces n'ayant pas forcément toutes les mêmes capacités dispersives. Du fait de leur dépôt polymétalliques ces écosystèmes sont aussi sujets à prospection minière menaçant possiblement la biodiversité associée (Gena, 2013; Levin et al., 2016; Van Dover et al., 2017; Washburn et al., 2019).

Ainsi, il important de délimiter les populations et de comprendre leurs connectivités à différences échelles comme nous l'avons réalisé dans le chapitre 1 sur *Ifremeria nautilei*, où nous avons recherché des locus dit « outliers » qui permettent d'identifier des limitations subtiles à la connectivité des populations à l'échelle des bassins.

Mais il est aussi nécessaire d'étendre ces approches de manière comparative à l'échelle des communautés d'espèces vivant dans ces écosystèmes, pour tenter de mieux appréhender les risques que ces dernières encourent. A ce titre, la structuration en métapopulations révélée chez toutes les espèces que nous avons étudiées traduit sans doute l'histoire vicariante des communautés hydrothermales dans la région du Pacifique Ouest. Cependant la mise en

évidence de remises en contact n'ayant pas la même temporalité et les subtiles différences de structure entre espèces peuvent traduire aussi une fragilité des dites communautés selon les traits d'histoire de vie des différentes espèces. Ceci serait particulièrement vrai dans notre cas où la connectivité génétique faible entre métapopulations et la quasi-homogénéité entre bassins au sein de ces dernières ne reflètent pas nécessairement une connectivité démographique suffisante pour assurer une possible recolonisation par d'autres populations après perte des populations locales.

CONCLUSION GENERALE ET PERSPECTIVES

Bilan des résultats

Au cours de cette thèse et des travaux qui y sont associés, j'ai cherché tout d'abord à explorer les patrons de différenciation phylogéographiques et de connectivité de plusieurs espèces inféodées aux sources hydrothermales des bassins d'arrière-arc du Pacifique Occidental. Ces sources constituent un habitat discontinu car entrecoupé par des arcs volcaniques, des fosses de subduction et des plaines abyssales. Les écosystèmes associés à ces sources sont composés d'une faune abondante d'espèces dépendant entièrement de la chimiosynthèse bactérienne. Dans ce contexte très fragmenté se pose la question du degré de connectivité des peuplements des différentes espèces hydrothermales ainsi que leur histoire phylogéographique.

Au commencement de ma thèse, quelques études pionnières de génétique des populations avaient mis en évidence des patrons de structure assez variés entre les espèces hydrothermales vivant dans le Pacifique Ouest (Tableau 1) (Kojima et al., 2000; Lee et al., 2019; Plouviez et al., 2019; Thaler et al., 2011, 2014; Yahagi et al., 2020). La plupart de ces études n'ont utilisé que le marqueur mitochondrial *Cox 1*. Celui-ci étant souvent assez résolutif pour détecter la structuration des populations. Ce marqueur particulier peut cependant aussi être assujéti à différents processus évolutifs pouvant brouiller le signal. Peu étudiées jusqu'à présent, nous explorons ces questions avec les méthodologies de la génomique des populations qui permettent de s'affranchir des principales limites liées à l'utilisation d'un faible nombre de marqueurs et apportent une meilleure résolution des processus démographiques.

Mon premier objectif a donc été d'explorer le patron de distribution phylogéographique et de connectivité chez le gastéropode emblématique et clé de voute de ces écosystèmes, *Ifremeria nautilei* (Provannidea) échantillonné sur l'ensemble de son aire de répartition. A partir de 10 570 marqueurs génomiques (ddRAD) indépendants nous avons permis d'identifier une structure en deux métapopulations d'ensemble de bassin d'arrière-arc séparées par les Iles Salomon et Vanuatu et quasi-homogènes génétiquement de part et d'autre. Ensuite, par une approche de reconstructions démographiques avec *dadi*, nous avons montré que ces métapopulations auraient commencé à diverger il y a environ 70 000 générations en présence d'un flux génique faible, bidirectionnel et légèrement asymétrique. Cependant une analyse

détaillée des locus ne se conformant pas à la distribution neutre attendue (« outliers ») montre de subtiles différenciations à l'échelle des bassins au sein des métapopulations. Ces résultats ont permis de montrer que la quasi-homogénéité génétique entre bassin affichant un très fort flux génique n'est pas nécessairement en adéquation avec une forte connectivité démographique. Ceci peut avoir des conséquences concernant la recolonisation des sites en cas de destruction massive.

L'objectif principal de la thèse était de comparer les patrons phylogéographiques et de connectivité entre quatre espèces (*I. nautiliei*, *A. kojimai*, *S. tollmanni* et *E. ohtai*) avec les mêmes méthodologies que celles utilisées pour le chapitre 1. Cette approche comparée a permis d'identifier un certain nombre de similitude et de différences entre espèces.

Comme pour *I. nautiliei* (chapitre 1), les autres espèces affichent une structure similaire en deux métapopulations dont le degré de différenciation varie en fonction des espèces. Les reconstructions indépendantes des histoires démo-génétiques ont montré l'existence d'un flux génique bi-directionnel et légèrement asymétrique entre les métapopulations. Cette bi-directionnalité commune à toutes les espèces pourrait suggérer que les stratégies de dispersion larvaire ne sont pas unimodales et donc plus complexes que celles proposées jusqu'à présent. La plupart des espèces auraient divergé sous un modèle de contact secondaire dont la divergence initiale des métapopulations aurait débuté il y a entre 70 000 – 100 000 générations à l'exception du gastéropode *A. kojimai* dont la divergence serait très récente. Les temporalités de remise en contact secondaire et du niveau du flux génique varient cependant d'une espèce à l'autre, des différences probablement attribuables à l'interaction entre l'histoire géologique de la région et leurs traits d'histoires de vie qui eux-mêmes conditionnent leur distribution biogéographique.

Pour accéder à une meilleure compréhension phylogéographique de cette région, nous nous proposons de rajouter d'autres espèces ayant aussi les mêmes distributions biogéographiques comme le Polynoidea *Branchinotogluma segonzaci*, les moules hydrothermales *Bathmodiolus septenderium* et *manusensis* ainsi que la patelle *Lepetodrilus schrolli*, espèces pour lesquelles un échantillonnage a également été réalisé lors de la campagne Chubacarc

Outils pour aborder les questions de résilience et perspectives liées à l'obtention de génomes complets

Plusieurs perspectives d'analyses génomiques complémentaires ont déjà proposées et discutées dans les perspectives du chapitre 1. Elles peuvent bien entendu être aussi appliquées aux autres espèces étudiées, comme par exemple l'approche de détection et d'utilisation de marqueurs « outliers » dans chaque métapopulation pour chaque espèce. Cela pourrait possiblement aider à détecter des subtiles limites de connectivités entre les bassins ou autres ensembles de populations. Ainsi une approche comparée de ces limites pourrait mettre en lumière des facteurs (courant, ensemble géologique) réduisant possiblement la dispersion et donc la connectivité démographique pour l'ensemble des espèces.

Nous avons proposé une histoire possible de colonisation depuis le plus ancien bassin arrière-arc du Pacifique Occidental (Nord-Fijidi). Cependant il est nécessaire d'explorer cette hypothèse concernant l'origine et l'orientation de celle-ci pour les différentes espèces. Pour cela, il est possible d'utiliser différentes approches. Ces colonisations ou expansions spatiales par des effets de fondation en série, laissent des empreintes caractéristiques au sein des génomes et des indices de diversité génétique, comme prédit dans la littérature théorique (Austerlitz et al., 1997; Excoffier et al., 2009; Slatkin & Excoffier, 2012). Ainsi ils peuvent laisser une trace caractéristique, comme une baisse de diversité génétique avec l'accroissement de la distance avec l'origine de l'expansion spatiale pouvant ainsi être utilisée pour inférer l'origine de cette dernière (Ramachandran et al., 2005). De manière analogue, l'indice de directionnalité (ψ) proposé par Peter & Slatkin (2013) est une mesure basée sur la différence en proportion des allèles dérivés partagés par paire de populations et donc cette proportion en fréquences alléliques dérivés sera d'autant plus forte lorsque l'on s'éloigne du centre d'origine (Peter & Slatkin, 2013, 2015).

L'utilisation combinée de ces approches permettrait possiblement d'explorer les origines possibles de la colonisation de ces bassins pour chaque espèce. Comme *A. kojimai* affiche une

temporalité de divergence très récente et complètement différente des autres espèces, il est possible que cette espèce ait une origine phylogéographique différente.

Cependant ces approches nécessitent l'accès à l'information des états ancestraux des allèles, ainsi qu'une estimation assez précise des indices de diversité génétique. Ces approches de représentation réduite du génome comme déjà mentionné permettent d'accéder à un grand nombre de marqueurs et d'individus à la fois avec un coût réduit, (Figure 12)

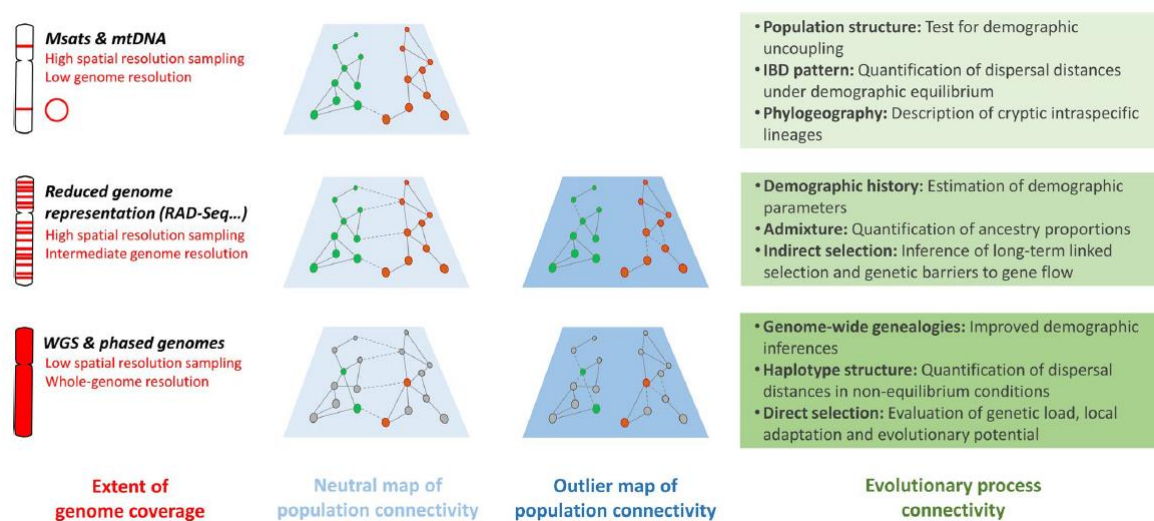


Figure 13 : Approches moléculaires pour comprendre la connectivité des populations allant des marqueurs microsatellites et mitochondriale aux génomes phasés. Les points colorés représentent les individus de chaque population. Tiré de (Gagnaire, 2020), Figure 1.

Cependant avec la très rapide l'évolution des techniques moléculaires et informatiques, couplée à une réduction drastique du coût de séquences, les génomes complets sont devenus particulièrement accessibles pour les espèces non-modèles. Ces génomes entiers, lorsqu'ils sont correctement assemblés, permettent d'accéder à de nouvelles données et statistiques concernant l'architecture du génome, telles que la phase haplotypique, l'estimation de blocs d'introgression entre populations ou espèces différentes, où les cartes de recombinaison qui peuvent être utilisées à l'aide différentes approches (*Relate*, *TSinfer/TSdate*) pour retracer l'histoire évolutive génomique ancienne et d'estimer la connectivité très récente entre populations (Duranton et al., 2019; Gagnaire, 2020; Kelleher et al., 2019; Leitwein et al., 2020; Speidel et al., 2019; Supple & Shapiro, 2018). Nous espérons que ce sera la prochaine étape pour ces espèces lorsqu'elles seront dotées de génomes complets.

RÉFÉRENCES

- Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, *11*(10), 697-709.
<https://doi.org/10.1038/nrg2844>
- Appeltans, W., Ah Yong, S. T., Anderson, G., Angel, M. V., Artois, T., Bailly, N., Bamber, R., Barber, A., Bartsch, I., Berta, A., Błażewicz-Paszkowycz, M., Bock, P., Boxshall, G., Boyko, C. B., Brandão, S. N., Bray, R. A., Bruce, N. L., Cairns, S. D., Chan, T.-Y., ... Costello, M. J. (2012). The Magnitude of Global Marine Species Diversity. *Current Biology*, *22*(23), 2189-2202. <https://doi.org/10.1016/j.cub.2012.09.036>
- Arellano, S. M., Van Gaest, A. L., Johnson, S. B., Vrijenhoek, R. C., & Young, C. M. (2014). Larvae from deep-sea methane seeps disperse in surface waters. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1786), 20133276.
<https://doi.org/10.1098/rspb.2013.3276>
- Arnold, B., Corbett-Detig, R. B., Hartl, D., & Bomblies, K. (2013). RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, *22*(11), 3179-3190. <https://doi.org/10.1111/mec.12276>
- Arteaga, A., Pyron, R. A., Peñafiel, N., Romero-Barreto, P., Culebras, J., Bustamante, L., Yáñez-Muñoz, M. H., & Guayasamin, J. M. (2016). Comparative Phylogeography Reveals Cryptic Diversity and Repeated Patterns of Cladogenesis for Amphibians and Reptiles in Northwestern Ecuador. *PLOS ONE*, *11*(4), e0151746.
<https://doi.org/10.1371/journal.pone.0151746>
- Audzijonyte, A., & Vrijenhoek, R. C. (2010). When Gaps Really Are Gaps : Statistical Phylogeography of Hydrothermal Vent Invertebrates. *Evolution*, *64*(8), 2369-2384.
<https://doi.org/10.1111/j.1558-5646.2010.00987.x>
- Austerlitz, F., Jung-Muller, B., Godelle, B., & Gouyon, P.-H. (1997). Evolution of Coalescence Times, Genetic Diversity and Structure during Colonization. *Theoretical Population Biology*, *51*(2), 148-164. <https://doi.org/10.1006/tpbi.1997.1302>
- Auzende, J.-M., Honza, E., Boespflug, X., Deo, S., Eissen, J.-P., Hashimoto, J., Huchon, P., Ishibashi, J., Iwabuchi, Y., Jarvis, P., Joshima, M., Kisimoto, K., Kuwahara, Y., Lafoy, Y., Matsumoto, T., Maze, J.-P., Mitsuzawa, K., Monma, H., Naganuma, T., ... Urabe, T.

- (1990). Active spreading and hydrothermalism in North Fiji Basin (SW Pacific). Results of Japanese French Cruise Kaiyo 87. *Marine Geophysical Researches*, 12(4), 269-283. <https://doi.org/10.1007/BF02428198>
- Avise, J. C. (2000). *Phylogeography : The History and Formation of Species*. Harvard University Press.
- Avise, J. C. (2009). Phylogeography : Retrospect and prospect. *Journal of Biogeography*, 36(1), 3-15. <https://doi.org/10.1111/j.1365-2699.2008.02032.x>
- Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A., & Saunders, N. C. (1987). Intraspecific Phylogeography : The Mitochondrial DNA Bridge Between Population Genetics and Systematics. *Annual Review of Ecology and Systematics*, 18(1), 489-522. <https://doi.org/10.1146/annurev.es.18.110187.002421>
- Bacon, C. D., Silvestro, D., Jaramillo, C., Smith, B. T., Chakrabarty, P., & Antonelli, A. (2015). Biological evidence supports an early and complex emergence of the Isthmus of Panama. *Proceedings of the National Academy of Sciences*, 112(19), 6110-6115. <https://doi.org/10.1073/pnas.1423853112>
- Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O. U., Swartz, B., Quental, T. B., Marshall, C., McGuire, J. L., Lindsey, E. L., Maguire, K. C., Mersey, B., & Ferrer, E. A. (2011). Has the Earth's sixth mass extinction already arrived? *Nature*, 471(7336), 51-57. <https://doi.org/10.1038/nature09678>
- Barry, P., Broquet, T., & Gagnaire, P.-A. (2022). Age-specific survivorship and fecundity shape genetic diversity in marine fishes. *Evolution Letters*, 6(1), 46-62. <https://doi.org/10.1002/evl3.265>
- Barton, N., & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridizing populations. *Heredity*, 57(3), 357-376. <https://doi.org/10.1038/hdy.1986.135>
- Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian Computation in Population Genetics. *Genetics*, 162(4), 2025-2035. <https://doi.org/10.1093/genetics/162.4.2025>
- Beinart, R. A., Nyholm, S. V., Dubilier, N., & Girguis, P. R. (2014). Intracellular Oceanospirillales inhabit the gills of the hydrothermal vent snail *Alviniconcha* with chemosynthetic, γ -Proteobacterial symbionts. *Environmental Microbiology Reports*, 6(6), 656-664. <https://doi.org/10.1111/1758-2229.12183>

- Benes, V., Scott, S. D., & Binns, R. A. (1994). Tectonics of rift propagation into a continental margin : Western Woodlark Basin, Papua New Guinea. *Journal of Geophysical Research: Solid Earth*, *99*(B3), 4439-4455. <https://doi.org/10.1029/93JB02878>
- Bonatelli, I. A. S., Gehara, M., Carstens, B. C., Colli, G. R., & Moraes, E. M. (2022). Comparative and predictive phylogeography in the South American diagonal of open formations: Unravelling the biological and environmental influences on multitaxon demography. *Molecular Ecology*, *31*(1), 331-342. <https://doi.org/10.1111/mec.16210>
- Boulart, C., Rouxel, O., Scalabrin, C., Le Meur, P., Pelleter, E., Poitrimol, C., Thiébaud, E., Matabos, M., Castel, J., Tran Lu Y, A., Michel, L.N., Cathalot, C., Cheron, S., Boissier, A., Germain, Y., Arnaud-Haond, S., Bonhomme, F., Broquet, T., Cueff-Gauchard, V., Le Layec, V., L'Haridon, S, Mary, J., Le Port, A-S., Tasiemski, A., Kuama, D.C., Hourdez, S., Jollivet, D. (*in press*). Active hydrothermal vents in the Woodlark Basin may act as dispersing centres for hydrothermal fauna *Communications Earth & Environment*.
- Brändel, M. (2007). Ecology of Achene Dimorphism in *Leontodon saxatilis*. *Annals of Botany*, *100*(6), 1189-1197. <https://doi.org/10.1093/aob/mcm214>
- Breusing, C., Biastoch, A., Drews, A., Metaxas, A., Jollivet, D., Vrijenhoek, R. C., Bayer, T., Melzner, F., Sayavedra, L., Petersen, J. M., Dubilier, N., Schilabel, M. B., Rosenstiel, P., & Reusch, T. B. H. (2016). Biophysical and Population Genetic Models Predict the Presence of “Phantom” Stepping Stones Connecting Mid-Atlantic Ridge Vent Ecosystems. *Current Biology*, *26*(17), 2257-2267. <https://doi.org/10.1016/j.cub.2016.06.062>
- Breusing, C., Johnson, S. B., Mitarai, S., Beinart, R. A., & Tunnicliffe, V. (2021). Differential patterns of connectivity in Western Pacific hydrothermal vent metapopulations: A comparison of biophysical and genetic models. *Evolutionary Applications*. <https://doi.org/10.1111/eva.13326>
- Breusing, C., Johnson, S. B., Tunnicliffe, V., Clague, D. A., Vrijenhoek, R. C., & Beinart, R. A. (2020). Allopatric and Sympatric Drivers of Speciation in Alviniconcha Hydrothermal Vent Snails. *Molecular Biology and Evolution*, *37*(12), 3469-3484. <https://doi.org/10.1093/molbev/msaa177>
- Browning, S. R., & Browning, B. L. (2011). Haplotype phasing : Existing methods and new developments. *Nature Reviews Genetics*, *12*(10), 703-714. <https://doi.org/10.1038/nrg3054>

- Browning, S. R., Browning, B. L., Zhou, Y., Tucci, S., & Akey, J. M. (2018). Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. *Cell*, *173*(1), 53-61.e9. <https://doi.org/10.1016/j.cell.2018.02.031>
- Brüniche-Olsen, A., Kellner, K. F., & DeWoody, J. A. (2019). Island area, body size and demographic history shape genomic diversity in Darwin's finches and related tanagers. *Molecular Ecology*, *28*(22), 4914-4925. <https://doi.org/10.1111/mec.15266>
- Burri, R. (2017). Interpreting differentiation landscapes in the light of long-term linked selection. *Evolution Letters*, *1*(3), 118-131. <https://doi.org/10.1002/evl3.14>
- Cann, R. L., Stoneking, M., & Wilson, A. C. (1987). Mitochondrial DNA and human evolution. *Nature*, *325*(6099), 31-36. <https://doi.org/10.1038/325031a0>
- Carranza, O. (2004). Late Tertiary terrestrial mammals from Central Mexico and their relationship to South American immigrants. *Revista Brasileira de Paleontologia*, *7*(2), 249-261. <https://doi.org/10.4072/rbp.2004.2.19>
- Castel, J., Hourdez, S., Pradillon, F., Daguin-Thiébaud, C., Ballenghien, M., Ruault, S., Corre, E., Tran Lu Y, A., Mary, J., Comtet, T., Gagnaire, P-A., Bonhomme, F., Bierne, N., Breusing, C., Broquet, T., & Jollivet D. (2022): A story of divergence and inter-specific exchanges in deep-sea hydrothermal vent gastropods *Alviniconcha*. *Genes*. Manuscript submitted for publication.
- Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human-induced species losses : Entering the sixth mass extinction. *Science Advances*, *1*(5), e1400253. <https://doi.org/10.1126/sciadv.1400253>
- Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, *15*(5), 538-543. <https://doi.org/10.1093/oxfordjournals.molbev.a025953>
- Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetics Research*, *70*(2), 155-174. <https://doi.org/10.1017/S0016672397002954>
- Chevaldonne, P., Jollivet, D., Desbruyeres, D., Lutz, R., & Vrijenhoek, R. (2002). Sister-species of eastern Pacific hydrothermal vent worms (Ampharetidae, Alvinellidae,

- Vestimentifera) provide new mitochondrial COI clock calibration. *CBM - Cahiers de Biologie Marine*, 43(3-4), 367-370. <https://archimer.ifremer.fr/doc/00000/895/>
- Chevaldonne, P., Jollivet, D., Vangriesheim, A., & Desbruyères, D. (1997). Hydrothermal-vent alvinellid polychaete dispersal in the eastern Pacific. 1. Influence of vent site distribution, bottom currents, and biological patterns. *Limnology and Oceanography*, 42(1), 67-80. <https://doi.org/10.4319/lo.1997.42.1.0067>
- Cody, S., Richardson, J. E., Rull, V., Ellis, C., & Pennington, R. T. (2010). The Great American Biotic Interchange revisited. *Ecography*, 33(2), 326-332. <https://doi.org/10.1111/j.1600-0587.2010.06327.x>
- Colinvaux, P. A., De Oliveira, P. E., Moreno, J. E., Miller, M. C., & Bush, M. B. (1996). A Long Pollen Record from Lowland Amazonia : Forest and Cooling in Glacial Times. *Science*, 274(5284), 85-88. <https://doi.org/10.1126/science.274.5284.85>
- Craddock, C., Hoeh, W. R., Gustafson, R. G., Lutz, R. A., Hashimoto, J., & Vrijenhoek, R. J. (1995). Evolutionary relationships among deep-sea mytilids (Bivalvia : Mytilidae) from hydrothermal vents and cold-water methane/sulfide seeps. *Marine Biology*, 121(3), 477-485. <https://doi.org/10.1007/BF00349456>
- Crawford, A. J., Bermingham, E., & Carolina, P. S. (2007). The role of tropical dry forest as a long-term barrier to dispersal : A comparative phylogeographical analysis of dry forest tolerant and intolerant frogs. *Molecular Ecology*, 16(22), 4789-4807. <https://doi.org/10.1111/j.1365-294X.2007.03524.x>
- Crooks, K. R., & Sanjayan, M. (2006). *Connectivity Conservation*. Cambridge University Press.
- Currat, M., Ruedi, M., Petit, R. J., & Excoffier, L. (2008). The Hidden Side of Invasions : Massive Introgression by Local Genes. *Evolution*, 62(8), 1908-1920. <https://doi.org/10.1111/j.1558-5646.2008.00413.x>
- Daguin-Thiébaud, C., Ruault, S., Roby, C., Broquet, T., Viard, F., & Brelsford, A. (2021, septembre 21). *Construction of individual ddRAD libraries*. Protocols.io. <https://www.protocols.io/view/construction-of-individual-ddrad-libraries-bv4tn8wn>
- Dawson, M. N. (2014). Natural experiments and meta-analyses in comparative phylogeography. *Journal of Biogeography*, 41(1), 52-65. <https://doi.org/10.1111/jbi.12190>
- De Queiroz, K. (2007). Species Concepts and Species Delimitation. *Systematic Biology*, 56(6), 879-886. <https://doi.org/10.1080/10635150701701083>

- de Alteriis, G., Ruellan, E., Auzende, J.-M., Ondréas, H., Bendel, V., Gràcia-Mont, E., Lagabrielle, Y., Huchon, P., & Tanahashi, M. (1993). Propagating rifts in the North Fiji Basin (southwest Pacific). *Geology*, *21*(7), 583-586. [https://doi.org/10.1130/0091-7613\(1993\)021<0583:PRITNF>2.3.CO;2](https://doi.org/10.1130/0091-7613(1993)021<0583:PRITNF>2.3.CO;2)
- Desbruyères, D., Hashimoto, J., & Fabri, M.-C. (2006). Composition and Biogeography of Hydrothermal Vent Communities in Western Pacific Back-Arc Basins. Dans *Back-Arc Spreading Systems : Geological, Biological, Chemical, and Physical Interactions* (p. 215-234). American Geophysical Union (AGU). <https://doi.org/10.1029/166GM11>
- Desbruyeres, D., Segonzac, M., Bright, M., & Decapoda, A. (2006). Handbook of deepsea hydrothermal vent fauna. *Denisia*.
- d’Horta, F. M., Cuervo, A. M., Ribas, C. C., Brumfield, R. T., & Miyaki, C. Y. (2013). Phylogeny and comparative phylogeography of *Sclerurus* (Aves : Furnariidae) reveal constant and cryptic diversification in an old radiation of rain forest understorey specialists. *Journal of Biogeography*, *40*(1), 37-49. <https://doi.org/10.1111/j.1365-2699.2012.02760.x>
- Durantón, M., Allal, F., Fraïsse, C., Bierne, N., Bonhomme, F., & Gagnaire, P.-A. (2018). The origin and remolding of genomic islands of differentiation in the European sea bass. *Nature Communications*, *9*(1), 2518. <https://doi.org/10.1038/s41467-018-04963-6>
- Durantón, M., Bonhomme, F., & Gagnaire, P.-A. (2019). The spatial scale of dispersal revealed by admixture tracts. *Evolutionary Applications*, *12*(9), 1743-1756. <https://doi.org/10.1111/eva.12829>
- Edwards, S. V., Robin, V. V., Ferrand, N., & Moritz, C. (2022). The Evolution of Comparative Phylogeography : Putting the Geography (and More) into Comparative Population Genomics. *Genome Biology and Evolution*, *14*(1), evab176. <https://doi.org/10.1093/gbe/evab176>
- Edwards, S. V., Shultz, A. J., & Campbell-Staton, S. C. (2015). Next-generation sequencing and the expanding domain of phylogeography. *Folia Zoologica*, *64*(3), 187-206. <https://doi.org/10.25225/fozo.v64.i3.a2.2015>
- Esselstyn, J. A., Oliveros, C. H., Moyle, R. G., Peterson, A. T., McGuire, J. A., & Brown, R. M. (2010). Integrating phylogenetic and taxonomic evidence illuminates complex biogeographic patterns along Huxley’s modification of Wallace’s Line. *Journal of Biogeography*, *37*(11), 2054-2066. <https://doi.org/10.1111/j.1365-2699.2010.02378.x>

- Ewing, G. B., & Jensen, J. D. (2016). The consequences of not accounting for background selection in demographic inference. *Molecular Ecology*, *25*(1), 135-141.
<https://doi.org/10.1111/mec.13390>
- Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 481-501.
<https://doi.org/10.1146/annurev.ecolsys.39.110707.173414>
- Fouquet, A., Courtois, E. A., Baudain, D., Lima, J. D., Souza, S. M., Noonan, B. P., & Rodrigues, M. T. (2015). The trans-riverine genetic structure of 28 Amazonian frog species is dependent on life history. *Journal of Tropical Ecology*, *31*(4), 361-373.
<https://doi.org/10.1017/S0266467415000206>
- Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, *27*(9), 489-496.
<https://doi.org/10.1016/j.tree.2012.05.012>
- Gagnaire, P.-A. (2020). Comparative genomics approach to evolutionary process connectivity. *Evolutionary Applications*, *13*(6), 1320-1334.
<https://doi.org/10.1111/eva.12978>
- Gagnaire, P.-A., Broquet, T., Aurelle, D., Viard, F., Souissi, A., Bonhomme, F., Arnaud-Haond, S., & Bierne, N. (2015). Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evolutionary Applications*, *8*(8), 769-786. <https://doi.org/10.1111/eva.12288>
- Gasse, F. (2000). Hydrological changes in the African tropics since the Last Glacial Maximum. *Quaternary Science Reviews*, *19*(1), 189-211. [https://doi.org/10.1016/S0277-3791\(99\)00061-X](https://doi.org/10.1016/S0277-3791(99)00061-X)
- Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdelhué, C., Pudlo, P., Cornuet, J.-M., & Estoup, A. (2013). The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology*, *22*(11), 3165-3178.
<https://doi.org/10.1111/mec.12089>
- Gena, K. (2013). Deep Sea Mining of Submarine Hydrothermal Deposits and its Possible Environmental Impact in Manus Basin, Papua New Guinea. *Procedia Earth and Planetary Science*, *6*, 226-233. <https://doi.org/10.1016/j.proeps.2013.01.031>

- Gilbert, K. J., Moinet, A., & Peischl, S. (2022). Gene surfing of underdominant alleles promotes formation of hybrid zones. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1846), 20210006. <https://doi.org/10.1098/rstb.2021.0006>
- Gladenkov, A. Y., Oleinik, A. E., Marincovich, L., & Barinov, K. B. (2002). A refined age for the earliest opening of Bering Strait. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 183(3), 321-328. [https://doi.org/10.1016/S0031-0182\(02\)00249-3](https://doi.org/10.1016/S0031-0182(02)00249-3)
- Gokcumen, O. (2020). Archaic hominin introgression into modern human genomes. *American Journal of Physical Anthropology*, 171(S70), 60-73. <https://doi.org/10.1002/ajpa.23951>
- Grant, P. R. (1999). *Ecology and Evolution of Darwin's Finches*. Princeton University Press.
- Grant, P. R., & Grant, B. R. (2009). The secondary contact phase of allopatric speciation in Darwin's finches. *Proceedings of the National Academy of Sciences*, 106(48), 20141-20148. <https://doi.org/10.1073/pnas.0911761106>
- Greene, H. G., Collot, J.-Y., Fisher, M. A., & Crawford, A. J. (1994). Neogene tectonic evolution of the New Hebrides island arc : A review incorporating ODP drilling results. *Proceedings of the Ocean Drilling Program, Scientific Results*, 134, 19-46.
- Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLOS Genetics*, 5(10), e1000695. <https://doi.org/10.1371/journal.pgen.1000695>
- Hall, R. (2002). Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific : Computer-based reconstructions, model and animations. *Journal of Asian Earth Sciences*, 20(4), 353-431. [https://doi.org/10.1016/S1367-9120\(01\)00069-4](https://doi.org/10.1016/S1367-9120(01)00069-4)
- Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, 25(11), 2454-2466. <https://doi.org/10.1111/mec.13582>
- Haug, G. H., & Tiedemann, R. (1998). Effect of the formation of the Isthmus of Panama on Atlantic Ocean thermohaline circulation. *Nature*, 393(6686), 673-676. <https://doi.org/10.1038/31447>
- Herrera, S., Watanabe, H., & Shank, T. M. (2015). Evolutionary and biogeographical patterns of barnacles from deep-sea hydrothermal vents. *Molecular Ecology*, 24(3), 673-689. <https://doi.org/10.1111/mec.13054>

- Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, 58(3), 247-276.
<https://doi.org/10.1111/j.1095-8312.1996.tb01434.x>
- Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, 68(1-2), 87-112. <https://doi.org/10.1111/j.1095-8312.1999.tb01160.x>
- Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789), 907-913. <https://doi.org/10.1038/35016000>
- Hewitt, G. M. (2011). Quaternary phylogeography : The roots of hybrid zones. *Genetica*, 139(5), 617-638. <https://doi.org/10.1007/s10709-011-9547-3>
- Hickerson, M. J., Carstens, B. C., Cavender-Bares, J., Crandall, K. A., Graham, C. H., Johnson, J. B., Rissler, L., Victoriano, P. F., & Yoder, A. D. (2010). Phylogeography's past, present, and future : 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution*, 54(1), 291-301. <https://doi.org/10.1016/j.ympev.2009.09.016>
- Hill, W. G., & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetics Research*, 8(3), 269-294. <https://doi.org/10.1017/S0016672300010156>
- Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology*, 30(1), 62-82.
<https://doi.org/10.1111/mec.15720>
- Hurtado, L. A., Lutz, R. A., & Vrijenhoek, R. C. (2004). Distinct patterns of genetic differentiation among annelids of eastern Pacific hydrothermal vents. *Molecular Ecology*, 13(9), 2603-2615. <https://doi.org/10.1111/j.1365-294X.2004.02287.x>
- Imbert, E. (2002). Ecological consequences and ontogeny of seed heteromorphism. *Perspectives in Plant Ecology, Evolution and Systematics*, 5(1), 13-36.
<https://doi.org/10.1078/1433-8319-00021>
- Jakobsson, A., & Eriksson, O. (2003). Trade-offs between dispersal and competitive ability: A comparative study of wind-dispersed Asteraceae forbs. *Evolutionary Ecology*, 17(3), 233-246. <https://doi.org/10.1023/A:1025526903281>
- Johannesson, K. (2003). Evolution in Littorina: Ecology matters. *Journal of Sea Research*, 49(2), 107-117. [https://doi.org/10.1016/S1385-1101\(02\)00218-6](https://doi.org/10.1016/S1385-1101(02)00218-6)

- Johannesson, K., Le Moan, A., Perini, S., & André, C. (2020). A Darwinian Laboratory of Multiple Contact Zones. *Trends in Ecology & Evolution*, *35*(11), 1021-1036. <https://doi.org/10.1016/j.tree.2020.07.015>
- Johannesson, K., Panova, M., Kemppainen, P., André, C., Rolán-Alvarez, E., & Butlin, R. K. (2010). Repeated evolution of reproductive isolation in a marine snail : Unveiling mechanisms of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1547), 1735-1747. <https://doi.org/10.1098/rstb.2009.0256>
- Johnson, S. B., Warén, A., Lee, R. W., Kano, Y., Kaim, A., Davis, A., Strong, E. E., & Vrijenhoek, R. C. (2010). Rubyspira, New Genus and Two New Species of Bone-Eating Deep-Sea Snails With Ancient Habits. *The Biological Bulletin*, *219*(2), 166-177. <https://doi.org/10.1086/BBLv219n2p166>
- Johnson, S. B., Won, Y.-J., Harvey, J. B., & Vrijenhoek, R. C. (2013). A hybrid zone between Bathymodiolus mussel lineages from eastern Pacific hydrothermal vents. *BMC Evolutionary Biology*, *13*(1), 21. <https://doi.org/10.1186/1471-2148-13-21>
- Johnson, S. B., Young, C. R., Jones, W. J., Warén, A., & Vrijenhoek, R. C. (2006). Migration, Isolation, and Speciation of Hydrothermal Vent Limpets (Gastropoda; Lepetodrilidae) Across the Blanco Transform Fault. *The Biological Bulletin*, *210*(2), 140-157. <https://doi.org/10.2307/4134603>
- Jollivet, D., Chevaldonne, P., & Planque, B. (1999). Hydrothermal-Vent Alvinellid Polychaete Dispersal in the Eastern Pacific. 2. A Metapopulation Model Based on Habitat Shifts. *Evolution*, *53*(4), 1128-1142. <https://doi.org/10.1111/j.1558-5646.1999.tb04527.x>
- Jollivet, D., Desbruyères, D., Bonhomme, F., & Moraga, D. (1995). Genetic differentiation of deep-sea hydrothermal vent alvinellid populations (Annelida : Polychaeta) along the East Pacific Rise. *Heredity*, *74*(4), 376-391. <https://doi.org/10.1038/hdy.1995.56>
- Jolly, M. T., Viard, F., Gentil, F., Thiébaud, E., & Jollivet, D. (2006). Comparative phylogeography of two coastal polychaete tubeworms in the Northeast Atlantic supports shared history and vicariant events. *Molecular Ecology*, *15*(7), 1841-1855. <https://doi.org/10.1111/j.1365-294X.2006.02910.x>
- Jones, A. W., & Kennedy, R. S. (2008). Evolution in a tropical archipelago : Comparative phylogeography of Philippine fauna and flora reveals complex patterns of colonization and diversification. *Biological Journal of the Linnean Society*, *95*(3), 620-639. <https://doi.org/10.1111/j.1095-8312.2008.01073.x>

- Joy, K., Fink, D., Storey, B., & Atkins, C. (2014). A 2 million year glacial chronology of the Hatherton Glacier, Antarctica and implications for the size of the East Antarctic Ice Sheet at the Last Glacial Maximum. *Quaternary Science Reviews*, *83*, 46-57.
<https://doi.org/10.1016/j.quascirev.2013.10.028>
- Keigwin, L. D., Jr. (1978). Pliocene closing of the Isthmus of Panama, based on biostratigraphic evidence from nearby Pacific Ocean and Caribbean Sea cores. *Geology*, *6*(10), 630-634. [https://doi.org/10.1130/0091-7613\(1978\)6<630:PCOTIO>2.0.CO;2](https://doi.org/10.1130/0091-7613(1978)6<630:PCOTIO>2.0.CO;2)
- Keith, M., Haase, K. M., Häckel, F., Schwarz-Schampera, U., Klemd, R., Hannington, M., Strauss, H., McConachy, T., & Anderson, M. (2021). Trace element fractionation and precipitation in submarine back-arc hydrothermal systems, Nifonea caldera, New Hebrides subduction zone. *Ore Geology Reviews*, *135*, 104211.
<https://doi.org/10.1016/j.oregeorev.2021.104211>
- Kelleher, J., Wong, Y., Wohns, A. W., Fadil, C., Albers, P. K., & McVean, G. (2019). Inferring whole-genome histories in large population datasets. *Nature Genetics*, *51*(9), 1330-1338. <https://doi.org/10.1038/s41588-019-0483-y>
- Kelly, R. P., & Palumbi, S. R. (2010). Genetic Structure Among 50 Species of the Northeastern Pacific Rocky Intertidal Community. *PLOS ONE*, *5*(1), e8594.
<https://doi.org/10.1371/journal.pone.0008594>
- Knowlton, N., & Weigt, L. A. (1998). New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1412), 2257-2263. <https://doi.org/10.1098/rspb.1998.0568>
- Kojima, S., Segawa, R., Fujiwara, Y., Fujikura, K., Ohta, S., & Hashimoto, J. (2001). Phylogeny of Hydrothermal-Vent-Endemic Gastropods *Alviniconcha* spp. From the Western Pacific Revealed by Mitochondrial DNA Sequences. *The Biological Bulletin*, *200*(3), 298-304. <https://doi.org/10.2307/1543511>
- Kojima, S., Segawa, R., Fujiwara, Y., Hashimoto, J., & Ohta, S. (2000). Genetic Differentiation of Populations of a Hydrothermal Vent-Endemic Gastropod, *Ifremeria nautilei*, between the North Fiji Basin and the Manus Basin revealed by Nucleotide Sequences of Mitochondrial DNA. *Zoological Science*, *17*(8), 1167-1174.
<https://doi.org/10.2108/zsj.17.1167>

- Konn, C., Fourré, E., Jean-Baptiste, P., Donval, J. P., Guyader, V., Birot, D., Alix, A. S., Gaillot, A., Perez, F., Dapoigny, A., Pelleter, E., Resing, J. A., Charlou, J. L., & Fouquet, Y. (2016). Extensive hydrothermal activity revealed by multi-tracer survey in the Wallis and Futuna region (SW Pacific). *Deep Sea Research Part I: Oceanographic Research Papers*, *116*, 127-144. <https://doi.org/10.1016/j.dsr.2016.07.012>
- Kyuno, A., Shintaku, M., Fujita, Y., Matsumoto, H., Utsumi, M., Watanabe, H., Fujiwara, Y., & Miyazaki, J.-I. (2009). Dispersal and Differentiation of Deep-Sea Mussels of the Genus *Bathymodiolus* (Mytilidae, Bathymodiolinae). *Journal of Marine Biology*, *2009*, e625672. <https://doi.org/10.1155/2009/625672>
- Laakkonen, H. M., Hardman, M., Strelkov, P., & Väinölä, R. (2021). Cycles of trans-Arctic dispersal and vicariance, and diversification of the amphi-boreal marine fauna. *Journal of Evolutionary Biology*, *34*(1), 73-96. <https://doi.org/10.1111/jeb.13674>
- Laakkonen, H. M., Lajus, D. L., Strelkov, P., & Väinölä, R. (2013). Phylogeography of amphi-boreal fish : Tracing the history of the Pacific herring *Clupea pallasii* in North-East European seas. *BMC Evolutionary Biology*, *13*(1), 67. <https://doi.org/10.1186/1471-2148-13-67>
- Lee, W.-K., Kim, S.-J., Hou, B. K., Dover, C. L. V., & Ju, S.-J. (2019). Population genetic differentiation of the hydrothermal vent crab *Austinograea alayseae* (Crustacea : Bythograeidae) in the Southwest Pacific Ocean. *PLOS ONE*, *14*(4), e0215829. <https://doi.org/10.1371/journal.pone.0215829>
- Leitwein, M., Duranton, M., Rougemont, Q., Gagnaire, P.-A., & Bernatchez, L. (2020). Using Haplotype Information for Conservation Genomics. *Trends in Ecology & Evolution*, *35*(3), 245-258. <https://doi.org/10.1016/j.tree.2019.10.012>
- Levin, L. A., Mengerink, K., Gjerde, K. M., Rowden, A. A., Van Dover, C. L., Clark, M. R., Ramirez-Llodra, E., Currie, B., Smith, C. R., Sato, K. N., Gallo, N., Sweetman, A. K., Lily, H., Armstrong, C. W., & Brider, J. (2016). Defining “serious harm” to the marine environment in the context of deep-seabed mining. *Marine Policy*, *74*, 245-259. <https://doi.org/10.1016/j.marpol.2016.09.032>
- Little, C. T. S., & Vrijenhoek, R. C. (2003). Are hydrothermal vent animals living fossils? *Trends in Ecology & Evolution*, *18*(11), 582-588. <https://doi.org/10.1016/j.tree.2003.08.009>

- Lönnerberg, K., & Eriksson, O. (2013). Rules of the seed size game : Contests between large-seeded and small-seeded species. *Oikos*, *122*(7), 1080-1084.
<https://doi.org/10.1111/j.1600-0706.2012.00249.x>
- Lowe, W. H., & Allendorf, F. W. (2010). What can genetics tell us about population connectivity? *Molecular Ecology*, *19*(15), 3038-3051. <https://doi.org/10.1111/j.1365-294X.2010.04688.x>
- Luttikhuisen, P. C., van den Heuvel, F. H. M., Rebours, C., Witte, H. J., van Bleijswijk, J. D. L., & Timmermans, K. (2018). Strong population structure but no equilibrium yet : Genetic connectivity and phylogeography in the kelp *Saccharina latissima* (Laminariales, Phaeophyta). *Ecology and Evolution*, *8*(8), 4265-4277.
<https://doi.org/10.1002/ece3.3968>
- Maggs, C. A., Castilho, R., Foltz, D., Henzler, C., Jolly, M. T., Kelly, J., Olsen, J., Perez, K. E., Stam, W., Väinölä, R., Viard, F., & Wares, J. (2008). Evaluating Signatures of Glacial Refugia for North Atlantic Benthic Marine Taxa. *Ecology*, *89*(sp11), S108-S122.
<https://doi.org/10.1890/08-0257.1>
- Marincovich, L., & Gladenkov, A. Y. (1999). Evidence for an early opening of the Bering Strait. *Nature*, *397*(6715), 149-151. <https://doi.org/10.1038/16446>
- Martinez, F., & Taylor, B. (1996). Backarc spreading, rifting, and microplate rotation, between transform faults in the Manus Basin. *Marine Geophysical Researches*, *18*(2-4), 203-224. <https://doi.org/10.1007/BF00286078>
- Matabos, M., & Jollivet, D. (2019). Revisiting the *Lepetodrilus elevatus* species complex (Vetigastropoda : Lepetodrilidae), using samples from the Galápagos and Guaymas hydrothermal vent systems. *Journal of Molluscan Studies*, *85*(1), 154-165.
<https://doi.org/10.1093/mollus/eyy061>
- Matabos, M., Thiébaud, E., Le Guen, D., Sadosky, F., Jollivet, D., & Bonhomme, F. (2008). Geographic clines and stepping-stone patterns detected along the East Pacific Rise in the vetigastropod *Lepetodrilus elevatus* reflect species crypticism. *Marine Biology*, *153*(4), 545-563. <https://doi.org/10.1007/s00227-007-0829-3>
- McConachy, T. F., Arculus, R. J., Yeats, C. J., Binns, R. A., Barriga, F. J. A. S., McInnes, B. I. A., Sestak, S., Sharpe, R., Rakau, B., & Tevi, T. (2005). New hydrothermal activity and alkalic volcanism in the back arc Coriolis Troughs, Vanuatu. *Geology*, *33*(1), 61-64.
<https://doi.org/10.1130/G20870.1>

- Milano, I., Babbucci, M., Cariani, A., Atanassova, M., Bekkevold, D., Carvalho, G. R., Espiñeira, M., Fiorentino, F., Garofalo, G., Geffen, A. J., Hansen, Jakob. H., Helyar, S. J., Nielsen, E. E., Ogden, R., Patarnello, T., Stagioni, M., Consortium, F., Tinti, F., & Bargelloni, L. (2014). Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). *Molecular Ecology*, *23*(1), 118-135.
<https://doi.org/10.1111/mec.12568>
- Mitarai, S., Watanabe, H., Nakajima, Y., Shchepetkin, A. F., & McWilliams, J. C. (2016). Quantifying dispersal from hydrothermal vent fields in the western Pacific Ocean. *Proceedings of the National Academy of Sciences*, *113*(11), 2976-2981.
<https://doi.org/10.1073/pnas.1518395113>
- Mitchell, K. J., Pratt, R. C., Watson, L. N., Gibb, G. C., Llamas, B., Kasper, M., Edson, J., Hopwood, B., Male, D., Armstrong, K. N., Meyer, M., Hofreiter, M., Austin, J., Donnellan, S. C., Lee, M. S. Y., Phillips, M. J., & Cooper, A. (2014). Molecular Phylogeny, Biogeography, and Habitat Preference Evolution of Marsupials. *Molecular Biology and Evolution*, *31*(9), 2322-2330. <https://doi.org/10.1093/molbev/msu176>
- Moalic, Y., Desbruyères, D., Duarte, C. M., Rozenfeld, A. F., Bachraty, C., & Arnaud-Haond, S. (2012). Biogeography Revisited with Network Theory : Retracing the History of Hydrothermal Vent Communities. *Systematic Biology*, *61*(1), 127-127.
<https://doi.org/10.1093/sysbio/syr088>
- Moan, A. L., Gagnaire, P.-A., & Bonhomme, F. (2016). Parallel genetic divergence among coastal–marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. *Molecular Ecology*, *25*(13), 3187-3202.
<https://doi.org/10.1111/mec.13627>
- Mondal, M., Bertranpetit, J., & Lao, O. (2019). Approximate Bayesian computation with deep learning supports a third archaic introgression in Asia and Oceania. *Nature Communications*, *10*(1), 246. <https://doi.org/10.1038/s41467-018-08089-7>
- Mullineaux, L. S. (2013). *Deep-Sea Hydrothermal Vent Communities*. 18.
- Mullineaux, L. S., Adams, D. K., Mills, S. W., & Beaulieu, S. E. (2010). Larvae from afar colonize deep-sea hydrothermal vents after a catastrophic eruption. *Proceedings of the National Academy of Sciences*, *107*(17), 7829-7834.
<https://doi.org/10.1073/pnas.0913187107>

- Musgrave, R. J., & Firth, J. V. (1999). Magnitude and timing of New Hebrides Arc rotation : Paleomagnetic evidence from Nendo, Solomon Islands. *Journal of Geophysical Research: Solid Earth*, *104*(B2), 2841-2853. <https://doi.org/10.1029/1998JB900080>
- Newman, W. A. (1985). The abyssal hydrothermal vent invertebrate fauna. A glimpse of antiquity? *Bull. Biol. Soc. Wash.*, *6*, 231-242. <https://ci.nii.ac.jp/naid/10013499951/>
- Ni, X., Yuan, K., Liu, C., Feng, Q., Tian, L., Ma, Z., & Xu, S. (2019). MultiWaver 2.0 : Modeling discrete and continuous gene flow to reconstruct complex population admixtures. *European Journal of Human Genetics*, *27*(1), 133-139. <https://doi.org/10.1038/s41431-018-0259-3>
- Nielsen, R., Akey, J. M., Jakobsson, M., Pritchard, J. K., Tishkoff, S., & Willerslev, E. (2017). Tracing the peopling of the world through genomics. *Nature*, *541*(7637), 302-310. <https://doi.org/10.1038/nature21347>
- Nielsen, R., & Wakeley, J. (2001). Distinguishing Migration From Isolation : A Markov Chain Monte Carlo Approach. *Genetics*, *158*(2), 885-896. <https://doi.org/10.1093/genetics/158.2.885>
- O’Dea, A., Lessios, H. A., Coates, A. G., Eytan, R. I., Restrepo-Moreno, S. A., Cione, A. L., Collins, L. S., de Queiroz, A., Farris, D. W., Norris, R. D., Stallard, R. F., Woodburne, M. O., Aguilera, O., Aubry, M.-P., Berggren, W. A., Budd, A. F., Cozzuol, M. A., Coppard, S. E., Duque-Caro, H., ... Jackson, J. B. C. (2016). Formation of the Isthmus of Panama. *Science Advances*, *2*(8), e1600883. <https://doi.org/10.1126/sciadv.1600883>
- Olsen, J. L., Stam, W. T., Coyer, J. A., Reusch, T. B. H., Billingham, M., Boström, C., Calvert, E., Christie, H., Granger, S., Lumière, R. L., Milchakova, N., Oudot-Le Secq, M.-P., Procaccini, G., Sanjabi, B., Serrão, E., Veldsink, J., Widdicombe, S., & Wyllie-Echeverria, S. (2004). North Atlantic phylogeography and large-scale population differentiation of the seagrass *Zostera marina* L. *Molecular Ecology*, *13*(7), 1923-1941. <https://doi.org/10.1111/j.1365-294X.2004.02205.x>
- O’Mullan, G. D., Maas, P. a. Y., Lutz, R. A., & Vrijenhoek, R. C. (2001). A hybrid zone between hydrothermal vent mussels (Bivalvia : Mytilidae) from the Mid-Atlantic Ridge. *Molecular Ecology*, *10*(12), 2819-2831. <https://doi.org/10.1046/j.0962-1083.2001.01401.x>

- Palsbøll, P. J., Bérubé, M., & Allendorf, F. W. (2007). Identification of management units using population genetic data. *Trends in Ecology & Evolution*, *22*(1), 11-16.
<https://doi.org/10.1016/j.tree.2006.09.003>
- Parson, L., & Hawkins, J. (1994). 51. Two-Stage Ridge Propagation And The Geological History Of The Lau Back arc Basin1. *Proceedings of the Ocean Drilling Program, Scientific Results*, *135*, 819-828.
- Pascual, M., Rives, B., Schunter, C., & Macpherson, E. (2017). Impact of life history traits on gene flow : A multispecies systematic review across oceanographic barriers in the Mediterranean Sea. *PLOS ONE*, *12*(5), e0176419.
<https://doi.org/10.1371/journal.pone.0176419>
- Patarnello, T., Volckaert, F. a. M. J., & Castilho, R. (2007). Pillars of Hercules : Is the Atlantic–Mediterranean transition a phylogeographical break? *Molecular Ecology*, *16*(21), 4426-4444. <https://doi.org/10.1111/j.1365-294X.2007.03477.x>
- Pellegrino, K. C. M., Rodrigues, M. T., Waite, A. N., Morando, M., Yassuda, Y. Y., & Sites, J. W., JR. (2005). Phylogeography and species limits in the *Gymnodactylus darwinii* complex (Gekkonidae, Squamata) : Genetic structure coincides with river systems in the Brazilian Atlantic Forest. *Biological Journal of the Linnean Society*, *85*(1), 13-26.
<https://doi.org/10.1111/j.1095-8312.2005.00472.x>
- Peter, B. M., & Slatkin, M. (2013). Detecting Range Expansions from Genetic Data. *Evolution*, *67*(11), 3274-3289. <https://doi.org/10.1111/evo.12202>
- Peter, B. M., & Slatkin, M. (2015). The effective founder effect in a spatially expanding population. *Evolution*, *69*(3), 721-734. <https://doi.org/10.1111/evo.12609>
- Pinto-Sánchez, N. R., Ibáñez, R., Madriñán, S., Sanjur, O. I., Bermingham, E., & Crawford, A. J. (2012). The Great American Biotic Interchange in frogs : Multiple and early colonization of Central America by the South American genus *Pristimantis* (Anura: Craugastoridae). *Molecular Phylogenetics and Evolution*, *62*(3), 954-972.
<https://doi.org/10.1016/j.ympev.2011.11.022>
- Plouviez, S., Faure, B., Guen, D. L., Lallier, F. H., Bierne, N., & Jollivet, D. (2013). A New Barrier to Dispersal Trapped Old Genetic Clines That Escaped the Easter Microplate Tension Zone of the Pacific Vent Mussels. *PLOS ONE*, *8*(12), e81555.
<https://doi.org/10.1371/journal.pone.0081555>

- Plouviez, S., LaBella, A. L., Weisrock, D. W., Meijerfeldt, F. A. B. von, Ball, B., Neigel, J. E., & Dover, C. L. V. (2019). Amplicon sequencing of 42 nuclear loci supports directional gene flow between South Pacific populations of a hydrothermal vent limpet. *Ecology and Evolution*, *9*(11), 6568-6580. <https://doi.org/10.1002/ece3.5235>
- Plouviez, S., Le Guen, D., Lecompte, O., Lallier, F. H., & Jollivet, D. (2010). Determining gene flow and the influence of selection across the equatorial barrier of the East Pacific Rise in the tube-dwelling polychaete *Alvinella pompejana*. *BMC Evolutionary Biology*, *10*(1), 220. <https://doi.org/10.1186/1471-2148-10-220>
- Plouviez, S., Shank, T. M., Faure, B., Daguin-Thiebaut, C., Viard, F., Lallier, F. H., & Jollivet, D. (2009). Comparative phylogeography among hydrothermal vent species along the East Pacific Rise reveals vicariant processes and population expansion in the South. *Molecular Ecology*, *18*(18), 3903-3917. <https://doi.org/10.1111/j.1365-294X.2009.04325.x>
- Poitrimol, C., Thiébaud, E., Daguin-Thiébaud, C., Le Port, A-S., Ballenghien, M., Tran Lu Y, A., Jollivet, D., Hourdez, S., Matabos, M., Contrasted phylogeographic patterns of hydrothermal vent gastropods along Southwest Pacific: Woodlark basin, a possible contact zone and/or stepping-stone. Manuscript submitted for publication.
- Pradillon, F., Shillito, B., Young, C. M., & Gaill, F. (2001). Developmental arrest in vent worm embryos. *Nature*, *413*(6857), 698-699. <https://doi.org/10.1038/35099674>
- Ramachandran, S., Deshpande, O., Roseman, C. C., Rosenberg, N. A., Feldman, M. W., & Cavalli-Sforza, L. L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences*, *102*(44), 15942-15947. <https://doi.org/10.1073/pnas.0507611102>
- Reynolds, K. C., Watanabe, H., Strong, E. E., Sasaki, T., Uematsu, K., Miyake, H., Kojima, S., Suzuki, Y., Fujikura, K., Kim, S., & Young, C. M. (2010). New Molluscan Larval Form : Brooding and Development in a Hydrothermal Vent Gastropod, *Ifremeria nautilei* (Provannidae). *The Biological Bulletin*, *219*(1), 7-11. <https://doi.org/10.1086/BBLv219n1p7>
- Rougemont, Q., Gagnaire, P.-A., Perrier, C., Genthon, C., Besnard, A.-L., Launey, S., & Evanno, G. (2017). Inferring the demographic history underlying parallel genomic divergence

- among pairs of parasitic and nonparasitic lamprey ecotypes. *Molecular Ecology*, 26(1), 142-162. <https://doi.org/10.1111/mec.13664>
- Rougeux, C., Bernatchez, L., & Gagnaire, P.-A. (2017). Modeling the Multiple Facets of Speciation-with-Gene-Flow toward Inferring the Divergence History of Lake Whitefish Species Pairs (*Coregonus clupeaformis*). *Genome Biology and Evolution*, 9(8), 2057-2074. <https://doi.org/10.1093/gbe/evx150>
- Roze, D., & Barton, N. H. (2006). The Hill–Robertson Effect and the Evolution of Recombination. *Genetics*, 173(3), 1793-1811. <https://doi.org/10.1534/genetics.106.058586>
- Ruellan, E., & Lagabrielle, Y. (2005). Subductions et ouvertures océaniques dans le Sud-Ouest Pacifique. *Géomorphologie : relief, processus, environnement*, 11(2), 121-142. <https://doi.org/10.4000/geomorphologie.307>
- Schmidt, K., Garbe-Schönberg, D., Hannington, M. D., Anderson, M. O., Bühring, B., Haase, K., Haruel, C., Lupton, J., & Koschinsky, A. (2017). Boiling vapour-type fluids from the Nifonea vent field (New Hebrides Back-Arc, Vanuatu, SW Pacific) : Geochemistry of an early-stage, post-eruptive hydrothermal system. *Geochimica et Cosmochimica Acta*, 207, 185-209. <https://doi.org/10.1016/j.gca.2017.03.016>
- Schmitt, T. (2007). Molecular biogeography of Europe : Pleistocene cycles and postglacial trends. *Frontiers in Zoology*, 4(1), 11. <https://doi.org/10.1186/1742-9994-4-11>
- Schrider, D. R., Shanku, A. G., & Kern, A. D. (2016). Effects of Linked Selective Sweeps on Demographic Inference and Model Selection. *Genetics*, 204(3), 1207-1223. <https://doi.org/10.1534/genetics.116.190223>
- Seston, S. L., Beinart, R. A., Sarode, N., Shockey, A. C., Ranjan, P., Ganesh, S., Girguis, P. R., & Stewart, F. J. (2016). Metatranscriptional Response of Chemoautotrophic *Ifremeria nautilei* Endosymbionts to Differing Sulfur Regimes. *Frontiers in Microbiology*, 7. <https://www.frontiersin.org/article/10.3389/fmicb.2016.01074>
- Shank, T. M., & Halanych, K. M. (2007). Toward a mechanistic understanding of larval dispersal : Insights from genomic fingerprinting of the deep-sea hydrothermal vent tubeworm *Riftia pachyptila*. *Marine Ecology*, 28(1), 25-35. <https://doi.org/10.1111/j.1439-0485.2007.00146.x>

- Slatkin, M., & Excoffier, L. (2012). Serial Founder Effects During Range Expansion : A Spatial Analog of Genetic Drift. *Genetics*, *191*(1), 171-181.
<https://doi.org/10.1534/genetics.112.139022>
- Sommer, S. A., Van Woudenberg, L., Lenz, P. H., Cepeda, G., & Goetze, E. (2017). Vertical gradients in species richness and community composition across the twilight zone in the North Pacific Subtropical Gyre. *Molecular Ecology*, *26*(21), 6136-6156.
<https://doi.org/10.1111/mec.14286>
- Sousa, V. C., Carneiro, M., Ferrand, N., & Hey, J. (2013). Identifying Loci Under Selection Against Gene Flow in Isolation-with-Migration Models. *Genetics*, *194*(1), 211-233.
<https://doi.org/10.1534/genetics.113.149211>
- Speidel, L., Forest, M., Shi, S., & Myers, S. R. (2019). A method for genome-wide genealogy estimation for thousands of samples. *Nature Genetics*, *51*(9), 1321-1329.
<https://doi.org/10.1038/s41588-019-0484-x>
- Sulloway, F. J. (1982). Darwin and his finches : The evolution of a legend. *Journal of the History of Biology*, *15*(1), 1-53. <https://doi.org/10.1007/BF00132004>
- Supple, M. A., & Shapiro, B. (2018). Conservation of biodiversity in the genomics era. *Genome Biology*, *19*(1), 131. <https://doi.org/10.1186/s13059-018-1520-3>
- Suzuki, Y., Kojima, S., Sasaki, T., Suzuki, M., Utsumi, T., Watanabe, H., Urakawa, H., Tsuchida, S., Nunoura, T., Hirayama, H., Takai, K., Nealson, K. H., & Horikoshi, K. (2006). Host-Symbiont Relationships in Hydrothermal Vent Gastropods of the Genus *Alviniconcha* from the Southwest Pacific. *Applied and Environmental Microbiology*.
<https://doi.org/10.1128/AEM.72.2.1388-1393.2006>
- Szitkar, F., Dymont, J., & Fouquet, Y. (2020). Widespread volcanism Southeast of Futuna Island (SW Pacific Ocean) : Near-seafloor magnetic dating and regional consequences. *Journal of Volcanology and Geothermal Research*, *406*, 107064.
<https://doi.org/10.1016/j.jvolgeores.2020.107064>
- Tänzler, R., Toussaint, E. F. A., Suhardjono, Y. R., Balke, M., & Riedel, A. (2014). Multiple transgressions of Wallace's Line explain diversity of flightless *Trigonopterus weevils* on Bali. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1782), 20132528.
<https://doi.org/10.1098/rspb.2013.2528>

- Teixeira, J. C., & Cooper, A. (2019). Using hominin introgression to trace modern human dispersals. *Proceedings of the National Academy of Sciences*, *116*(31), 15327-15332. <https://doi.org/10.1073/pnas.1904824116>
- Teixeira, S., Cambon-Bonavita, M.-A., Serrão, E. A., Desbruyères, D., & Arnaud-Haond, S. (2011). Recent population expansion and connectivity in the hydrothermal shrimp *Rimicaris exoculata* along the Mid-Atlantic Ridge. *Journal of Biogeography*, *38*(3), 564-574. <https://doi.org/10.1111/j.1365-2699.2010.02408.x>
- Teixeira, S., Serrão, E. A., & Arnaud-Haond, S. (2012). Panmixia in a Fragmented and Unstable Environment : The Hydrothermal Shrimp *Rimicaris exoculata* Disperses Extensively along the Mid-Atlantic Ridge. *PLOS ONE*, *7*(6), e38521. <https://doi.org/10.1371/journal.pone.0038521>
- Templeton. (1998). Nested clade analyses of phylogeographic data : Testing hypotheses about gene flow and population history. *Molecular Ecology*, *7*(4), 381-397. <https://doi.org/10.1046/j.1365-294x.1998.00308.x>
- Thaler, A. D., Plouviez, S., Saleu, W., Alei, F., Jacobson, A., Boyle, E. A., Schultz, T. F., Carlsson, J., & Dover, C. L. V. (2014). Comparative Population Structure of Two Deep-Sea Hydrothermal-Vent-Associated Decapods (*Chorocaris sp. 2* and *Munidopsis lauensis*) from Southwestern Pacific Back-Arc Basins. *PLOS ONE*, *9*(7), e101345. <https://doi.org/10.1371/journal.pone.0101345>
- Thaler, A. D., Zelnio, K., Saleu, W., Schultz, T. F., Carlsson, J., Cunningham, C., Vrijenhoek, R. C., & Van Dover, C. L. (2011). The spatial scale of genetic subdivision in populations of *Ifremeria nautilei*, a hydrothermal-vent gastropod from the southwest Pacific. *BMC Evolutionary Biology*, *11*(1), 372. <https://doi.org/10.1186/1471-2148-11-372>
- Thomé, M. T. C., Zamudio, K. R., Haddad, C. F. B., & Alexandrino, J. (2014). Barriers, rather than refugia, underlie the origin of diversity in toads endemic to the Brazilian Atlantic Forest. *Molecular Ecology*, *23*(24), 6152-6164. <https://doi.org/10.1111/mec.12986>
- Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R. S. T., Hecht, J., Knaust, F., Belkhir, K., Klages, S., Dieterich, R., Stueber, K., Piferrer, F., Guinand, B., Bierne, N., Volckaert, F. A. M., Bargelloni, L., Power, D. M., Bonhomme, F., ... Reinhardt, R. (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications*, *5*(1), 5770. <https://doi.org/10.1038/ncomms6770>

- Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, *21*(16), 3907-3930.
<https://doi.org/10.1111/j.1365-294X.2012.05664.x>
- Tunnicliffe, V. (1992). The Nature and Origin of the Modern Hydrothermal Vent Fauna. *PALAIOS*, *7*(4), 338-350. <https://doi.org/10.2307/3514820>
- Tunnicliffe, V., & Bone, Q. (1988). Biogeography and evolution of hydrothermal-vent fauna in the eastern Pacific Ocean. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *233*(1272), 347-366. <https://doi.org/10.1098/rspb.1988.0025>
- Tunnicliffe, V., McArthur, A. G., & McHugh, D. (1998). A Biogeographical Perspective of the Deep-Sea Hydrothermal Vent Fauna. Dans J. H. S. Blaxter, A. J. Southward, & P. A. Tyler (Éds.), *Advances in Marine Biology* (Vol. 34, p. 353-442). Academic Press.
[https://doi.org/10.1016/S0065-2881\(08\)60213-8](https://doi.org/10.1016/S0065-2881(08)60213-8)
- Tyler, P. A., & Young, C. M. (1999). Reproduction and dispersal at vents and cold seeps. *Journal of the Marine Biological Association of the United Kingdom*, *79*(2), 193-208.
<https://doi.org/10.1017/S0025315499000235>
- Tyler, P. A., & Young, C. M. (2003). Dispersal at hydrothermal vents : A summary of recent progress. *Hydrobiologia*, *503*(1), 9-19.
<https://doi.org/10.1023/B:HYDR.0000008492.53394.6b>
- Tzedakis, P. C., Emerson, B. C., & Hewitt, G. M. (2013). Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends in Ecology & Evolution*, *28*(12), 696-704.
<https://doi.org/10.1016/j.tree.2013.09.001>
- Vacher, J.-P., Chave, J., Ficetola, F. G., Sommeria-Klein, G., Tao, S., Thébaud, C., Blanc, M., Camacho, A., Cassimiro, J., Colston, T. J., Dewynter, M., Ernst, R., Gaucher, P., Gomes, J. O., Jairam, R., Kok, P. J. R., Lima, J. D., Martinez, Q., Marty, C., ... Fouquet, A. (2020). Large-scale DNA-based survey of frogs in Amazonia suggests a vast underestimation of species richness and endemism. *Journal of Biogeography*, *47*(8), 1781-1791.
<https://doi.org/10.1111/jbi.13847>
- Van Dover, C. L. (1995). Ecology of Mid-Atlantic Ridge hydrothermal vents. *Geological Society, London, Special Publications*, *87*(1), 257-294.
<https://doi.org/10.1144/GSL.SP.1995.087.01.21>
- Van Dover, C. L., Ardron, J. A., Escobar, E., Gianni, M., Gjerde, K. M., Jaeckel, A., Jones, D. O. B., Levin, L. A., Niner, H. J., Pendleton, L., Smith, C. R., Thiele, T., Turner, P. J., Watling,

- L., & Weaver, P. P. E. (2017). Biodiversity loss from deep-sea mining. *Nature Geoscience*, *10*(7), 464-465. <https://doi.org/10.1038/ngeo2983>
- Van Dover, C. L., German, C. R., Speer, K. G., Parson, L. M., & Vrijenhoek, R. C. (2002). Evolution and Biogeography of Deep-Sea Vent and Seep Invertebrates. *Science*, *295*(5558), 1253-1257. <https://doi.org/10.1126/science.1067361>
- Van Dover, C. L., Humphris, S. E., Fornari, D., Cavanaugh, C. M., Collier, R., Goffredi, S. K., Hashimoto, J., Lilley, M. D., Reysenbach, A. L., Shank, T. M., Von Damm, K. L., Banta, A., Gallant, R. M., Götz, D., Green, D., Hall, J., Harmer, T. L., Hurtado, L. A., Johnson, P., ... Vrijenhoek, R. C. (2001). Biogeography and Ecological Setting of Indian Ocean Hydrothermal Vents. *Science*, *294*(5543), 818-823. <https://doi.org/10.1126/science.1064574>
- van Mólken, T., Jorritsma-Wienk, L. D., van Hoek, P. H. W., & de Kroon, H. (2005). Only seed size matters for germination in different populations of the dimorphic *Tragopogon pratensis* subsp. *Pratensis* (Asteraceae). *American Journal of Botany*, *92*(3), 432-437. <https://doi.org/10.3732/ajb.92.3.432>
- Venable, D. L., & Levin, D. A. (1985). Ecology of Achene Dimorphism in *Heterotheca Latifolia* : I. Achene Structure, Germination and Dispersal. *Journal of Ecology*, *73*(1), 133-145. <https://doi.org/10.2307/2259774>
- Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K., & Wilson, A. C. (1991). African Populations and the Evolution of Human Mitochondrial DNA. *Science*, *253*(5027), 1503-1507. <https://doi.org/10.1126/science.1840702>
- Vrijenhoek, R. C. (1997). Gene Flow and Genetic Diversity in Naturally Fragmented Metapopulations of Deep-Sea Hydrothermal Vent Animals. *Journal of Heredity*, *88*(4), 285-293. <https://doi.org/10.1093/oxfordjournals.jhered.a023106>
- Vrijenhoek, R. C. (2010). Genetic diversity and connectivity of deep-sea hydrothermal vent metapopulations. *Molecular Ecology*, *19*(20), 4391-4411. <https://doi.org/10.1111/j.1365-294X.2010.04789.x>
- Wallace, A. R. (1860). Notes of a Voyage to New Guinea. *The Journal of the Royal Geographical Society of London*, *30*, 172-177. <https://doi.org/10.2307/1798299>
- Warèn, A., & Bouchet, P. (1993). New records, species, genera, and a new family of gastropods from hydrothermal vents and hydrocarbon seeps*. *Zoologica Scripta*, *22*(1), 1-90. <https://doi.org/10.1111/j.1463-6409.1993.tb00342.x>

- Washburn, T. W., Turner, P. J., Durden, J. M., Jones, D. O. B., Weaver, P., & Van Dover, C. L. (2019). Ecological risk assessment for deep-sea mining. *Ocean & Coastal Management, 176*, 24-39. <https://doi.org/10.1016/j.ocecoaman.2019.04.014>
- Weissel, J. K., Taylor, B., & Karner, G. D. (1982). The opening of the Woodlark Basin, subduction of the Woodlark spreading system, and the evolution of Northern Melanesia since mid-pliocene time. *Tectonophysics, 87*(1), 253-277. [https://doi.org/10.1016/0040-1951\(82\)90229-3](https://doi.org/10.1016/0040-1951(82)90229-3)
- Wilson, A. C., Cann, R. L., Carr, S. M., George, M., Gyllenstein, U. B., Helm-Bychowski, K. M., Higuchi, R. G., Palumbi, S. R., Prager, E. M., Sage, R. D., & Stoneking, M. (1985). Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society, 26*(4), 375-400. <https://doi.org/10.1111/j.1095-8312.1985.tb02048.x>
- Wilson, O. J., Mayle, F. E., Walters, R. J., Lingner, D. V., & Vibrans, A. C. (2021). Floristic change in Brazil's southern Atlantic Forest biodiversity hotspot : From the Last Glacial Maximum to the late 21st Century. *Quaternary Science Reviews, 264*, 107005. <https://doi.org/10.1016/j.quascirev.2021.107005>
- Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics, 18*(2), 87-100. <https://doi.org/10.1038/nrg.2016.133>
- Wolff, T. (2005). Composition and endemism of the deep-sea hydrothermal vent fauna. *CBM-Cahiers de Biologie Marine., 46*(2), 97-104.
- Won, Y., Hallam, S. J., O'Mullan, G. D., & Vrijenhoek, R. C. (2003). Cytonuclear disequilibrium in a hybrid zone involving deep-sea hydrothermal vent mussels of the genus *Bathymodiolus*. *Molecular Ecology, 12*(11), 3185-3190. <https://doi.org/10.1046/j.1365-294X.2003.01974.x>
- Wynngaarden, M. V., Snelgrove, P. V. R., DiBacco, C., Hamilton, L. C., Rodríguez-Ezpeleta, N., Jeffery, N. W., Stanley, R. R. E., & Bradbury, I. R. (2017). Identifying patterns of dispersal, connectivity and selection in the sea scallop, *Placopecten magellanicus*, using RADseq-derived SNPs. *Evolutionary Applications, 10*(1), 102-117. <https://doi.org/10.1111/eva.12432>
- Yahagi, T., Thaler, A. D., Dover, C. L. V., & Kano, Y. (2020). Population connectivity of the hydrothermal-vent limpet *Shinkailepas tollmanni* in the Southwest Pacific

- (Gastropoda : Neritimorpha: Phenacolepadidae). *PLOS ONE*, 15(9), e0239784.
<https://doi.org/10.1371/journal.pone.0239784>
- Yahagi, T., Watanabe, H. K., Kojima, S., & Kano, Y. (2017). Do larvae from deep-sea hydrothermal vents disperse in surface waters? *Ecology*, 98(6), 1524-1534.
<https://doi.org/10.1002/ecy.1800>
- Yamaguchi, T., & Newman, W. A. (1997). The hydrothermal vent barnacle *Eochionelasmus* (Cirripedia, Balanomorpha) from the North Fiji, Lau and Manus Basins, South-West Pacific. 27.
- Young, C. M., Eckelbarger, K. J., & Eckelbarger, K. (1994). *Reproduction, Larval Biology, and Recruitment of the Deep-sea Benthos*. Columbia University Press.
- Young, C. M., He, R., Emler, R. B., Li, Y., Qian, H., Arellano, S. M., Van Gaest, A., Bennett, K. C., Wolf, M., Smart, T. I., & Rice, M. E. (2012). Dispersal of Deep-Sea Larvae from the Intra-American Seas : Simulations of Trajectories using Ocean Models. *Integrative and Comparative Biology*, 52(4), 483-496. <https://doi.org/10.1093/icb/ics090>
- Zakas, C., & Hall, D. W. (2012). Asymmetric Dispersal Can Maintain Larval Polymorphism : A Model Motivated by *Streblospio benedicti*. *Integrative and Comparative Biology*, 52(1), 197-212. <https://doi.org/10.1093/icb/ics055>
- Zakas, C., & Rockman, M. V. (2015). Gene-based polymorphisms reveal limited genomic divergence in a species with a heritable life-history dimorphism. *Evolution & Development*, 17(4), 240-247. <https://doi.org/10.1111/ede.12128>
- Zeisset, I., & Beebee, T. J. C. (2008). Amphibian phylogeography : A model for understanding historical aspects of species distributions. *Heredity*, 101(2), 109-119.
<https://doi.org/10.1038/hdy.2008.30>

ANNEXES

ANNEXE 1

Information supplémentaire pour l'article:

Subtle limits to connectivity revealed by outlier loci within two divergent metapopulations
of the deep-sea hydrothermal gastropod *Ifremeria nautili*

Information supplémentaire pour l'article du chapitre 1 accepté à *Molecular Ecology*

Adrien Tran Lu Y, Stéphanie Ruault, Claire Daguin-Thiébaud, Jade Castel, Nicolas Bierne,
Thomas Broquet, Patrick Wincker, Aude Perdereau, Sophie Arnaud-Haond, Pierre-
Alexandre Gagnaire, Didier Jollivet, Stéphane Hourdez, François Bonhomme

Table of Contents :

<i>De novo parameter calibration</i>	Page 150
<i>Tables</i>	Page 153
<i>Figures</i>	Page 162
<i>Settings</i>	Page 185

De novo parameter calibration

Assembling sequences *de novo* requires to explore assembly parameter to have the best representation of the genetic diversity in a given species with a good balance of parameter to ensure to not split homolog loci into several loci or assembling paralogue or chimeric loci while minimizing the bias due to genotyping errors.

To calibrate and identify the most appropriate set of parameters, we used subset of individual covering all localities and basins with 3-4 individuals at least per locality with and without replicates. This results in a dataset of 84 individuals while considering 3 cases of populations. A first one with all individual and replicate considered as one population, a second with population considered by basin but without replicate and a third one with population as localities also without replicate.

Thus, reads were assembled by *de novo* with each module of the software Stacks one by one (ustacks, cstacks, sstacks, tsv2bam, gstacks and populations) using values ranged from 3 to 8 for the minimum coverage of reads (m) to assemble a stack (a putative allele) and values ranged from 3 to 11 for both the maximum number of mismatches (M) between reads within an single individual to assemble a putative loci and the maximum divergence (n) between putative loci among individuals used for the construction of the catalog (with $M = n$). Then, all the assemblies were made with the variants present in at least 80% of the individuals ($R = 0.8$), following suggestions and recommendation of Mastretta-Yanes et al. (2015) and Paris et al. (2017).

These latter authors suggested to monitor assembly parameters m , M and n with several assembly statistic such as (i) the number of assembled loci; (ii) the number of polymorphic loci and (iii) the number of SNPs for each combination of parameters. In addition, Mastretta-Yanes et al (2015) also recommended to use pairs of replicated individuals to assesses the genotyping error rate.

In addition, we also use the nucleotide diversity (π) to monitor assembly parameters, since, it is simply a measure of the degree of polymorphism. It is defined in a population as the average number of nucleotide differences per site between two DNA sequences (Nei & Li, 1979) and if estimated from bi-allelic loci is very similar to the expected heterozygosity. This statistic is

therefore sensitive to the number of assembled sites, polymorphic or not, as well as to the number of polymorphic loci (RADtags). The nucleotide diversity (π) was calculated using the populations module of Stacks and the error rate of SNP genotyping between replicates with a house R script for each set of parameters on the assembled VCF.

In their recommendations, Paris et al. (2017) suggest taking the values of the parameters as being optimal when the different statistics used to monitor assembly parameter reach a plateau.

The number of polymorphic loci (i.e. RADtags), the number of variants (SNPs) and the total number of sites assembled show the same pattern for each value of m , M and $M=n$ (SI figure 1,2,3) and this whatever the case considered for the populations (one population, by basins or localities with or without replicas). Increasing m reduces the number of polymorphic loci, SNPs and assembled sites, but at $m = 4$, this decrease seems to be less important (SI figure 1,2,3).

For the parameter M and $n = M$, a slight increase is observed in the number of polymorphic loci and assembled sites until M reaches the value 4, then decreases (SI figure 1,2,3). In addition, the same pattern is observed for the number of SNPs but with a decrease starting from $M = 6$ (SI figure 1,2,3). Although monitoring these different parameters, any plateau is observed for the different cases. However, when examining nucleotide diversity (π), regardless of the case considered, as a single population (SI figure 1) or by basin (SI figure 2) and locality (SI figure 3) without replicates, a plateau is observed at $M = 8$. Although a plateau is observed for M with the nucleotide diversity, none is observed with the parameter m and this whatever the statistics used.

Following the recommendations of Mastretta Yanes et al., (2015), we also looked at the genotyping error rate for all pairs of replicates for all parameter sets (SI Figure 4,5). Thus, to observe the general trend for each parameter, mean genotyping error rate were also calculated across all individuals for all set of parameters (SI figure 6). The parameter m tends to decrease the genotyping error rate while M slightly increases this one but remain very low (SI Figure 4,5,6). For $m = 4$, the mean genotyping error rate was minimal at 0.96% (SI Figure 6).

Based on these approaches and recommendation, we found that $m = 4$ represents a good trade-off between polymorphism and coverage while the mean genotyping error rate remain very low. M and n parameter depends on the degree of polymorphism of the species while considering the size of the DNA fragment. M will be higher if the degree of polymorphism is important, here, we have tested a wide range of value for M (3-11) and we observe at $M = 8$, the nucleotide diversity (π) seems to reach to a plateau, regardless of the value of m and while having a low genotyping error rate

All 486 individual were assembled using the Stacks modules with these parameter values and following the suggestion of Mastretta-Yanes et al (2015) and Paris et al. (2017) with $R = 0.8$, i.e., allowing a maximum of 20% missing data per SNP called in the entire dataset.

Tables

SI Table 1 : Details for Ifremeria nautiliei samples (including replicates; total = 486) and unique individuals (total = 456) used in this study from the Chubacarc oceanographic campaign and S.Hourdez and C.L Van Dover's collections.

Basin	Locality	Depth (m)	ID_ sites	Number of sampled (indiv.)	Longitude	Latitude	Year	Collection
North-Fiji	Phoenix	1974	56	26 (22)	E 173°55.111'	S 16°56.936'	2019	Chubacarc
North-Fiji	Phoenix	1961	62	13 (12)	E 173°55.133'	S 16°57.005'	2019	Chubacarc
Futuna	Fati_Ufu	1516	10	19 (19)	W 177°10.968'	S 14°45.325'	2019	Chubacarc
Futuna	Fati_Ufu	1519	16	12 (12)	W 177°11.113'	S 14°45.601'	2019	Chubacarc
Futuna	Fatu_Kapa	1562	26	12 (11)	W 177°09.133'	S 14°45.110'	2019	Chubacarc
Futuna	Fatu_Kapa	1548	29	11 (11)	W 177°09.960'	S 14°44.243'	2019	Chubacarc
Futuna	Kulo_Lasi	1371	38	28 (19)	W 177°15.551'	S 14°56.468'	2019	Chubacarc
Lau	Abe	2153	2	22 (20)	W 176°11.479'	S 20°45.784'	2019	Chubacarc
Lau	Mangatolo	2031	46	24 (24)	W 174°39.210'	S 15°24.875'	2019	Chubacarc
Lau	Tow_Cam	2698	66	8 (8)	W 176°08.203'	S 20°19.047'	2019	Chubacarc
Lau	Tow_Cam	2711	73	9 (8)	W 176°08.250'	S 20°19.074'	2019	Chubacarc
Lau	Tui_Malila	1892	80	14 (12)	W 176°34.077'	S 21°59.280'	2019	Chubacarc
Lau	Tui_Malila	1874	81	4 (4)	W 176°34.088'	S 21°59.351'	2019	Chubacarc
Lau	Tui_Malila	1877	84	5 (4)	W 176°34.094'	S 21°59.354'	2019	Chubacarc
Lau	Tui_Malila	1886	85	7 (7)	W 176°34.097'	S 21°59.352'	2019	Chubacarc
Manus	Fenway	1703	98	22 (21)	E 151°40.342'	S 03°43.707'	2019	Chubacarc
Manus	Fenway	1699	103	23 (23)	E 151°40.367'	S 03°43.665'	2019	Chubacarc
Manus	North_Su	1216	145	24 (24)	E 152°06.046'	S 03°47.935'	2019	Chubacarc
Manus	Romans_Ruins	1659	119	10 (10)	E 151°40.469'	S 03°43.287'	2019	Chubacarc
Manus	Solwara_6	1729	123	25 (23)	E 151°40.854'	S 03°43.654'	2019	Chubacarc
Manus	Solwara_7	1769	107	23 (23)	E 151°40.374'	S 03°43.042'	2019	Chubacarc
Manus	Solwara_8	1739	111	23 (22)	E 151°40.441'	S 03°43.825'	2019	Chubacarc
Manus	Solwara_8	1734	112	2 (2)	E 151°40.443'	S 03°43.834'	2019	Chubacarc
Manus	South_Su	1353	135	23 (23)	E 152°06.310'	S 03°48.582'	2019	Chubacarc
Manus	South_Su	1341	152	25 (24)	E 152°06.291'	S 03°48.499'	2019	Chubacarc
Manus	Suzette	1506	140	21	E 152°05.783'	S 03°47.368'	2019	Chubacarc
Woodlark	Scala	3388	172	24 (22)	E 155°03.161'	S 09°47.945'	2019	Chubacarc
Lau	Kilo_Moana	2615		5 (5)	W 176°8.025'	S 20°3.175'	2009	S.Hourdez
Lau	Abe	2144		6 (6)	W 176°11.483'	S 20°45.791'	2009	S.Hourdez
Lau	Tow_cam	2703		5 (5)	W 176°8.19'	S 20°18.996'	2009	S.Hourdez
Lau	Tui_Malila	1885		6 (6)	W 176°34.079'	S 21°59.353'	2009	S.Hourdez
Manus	Solwara_8	1710		4 (4)	E 151° 40.458'	S 03° 43.824'	2008	C.L Van Dover
Manus	South-Su	1300		1 (1)	E 152° 06.144'	S 03° 48.564'	2008	C.L Van Dover

SI Table 2: Number of individuals per basin and locality, kept after Filtering data prior populations and demographic analyses.

Basin	Number of sample (by basin)	Locality	Number of sample (by locality)
Futuna	56	Fati_Ufu	25
		Fatu_Kapa	18
		Kulo_Lasi	13
Lau	82	Abe	15
		Mangatolo	22
		Tow_Cam	17
		Tui_Malila	28
Manus	180	Fenway	41
		North_Su	24
		Romans_Ruins	9
		Solwara_6	19
		Solwara_7	20
		Solwara_8	26
		South_Su	21
Suzette	20		
North-Fiji	24	Phoenix	24
Woodlark	20	Scala	20
TOTAL			362

SI Table 3: Filtering data prior populations and demographic analyses.

Filtering criterons	N° SNPs	N° radtag	N° individuals
Assembled data	649 106	38,608	486
Replicate and ind with more 20% miss data	649 106	38,608	441
SNPs with heterozygosity >0.6	648 916	38,550	441
max missing data per SNPs 10%	265 383	17381	441
max missing data per Individual 10%	265 383	17381	375
Max Read Depth (DP) > 80	265 365	17375	375
Minor allele frequency > 0.05	36 239	11140	375
Max missing data par Individual 10%	36 239	11140	362
Out of Hardy-Weinberg equilibrium p -value < 0.05	30 794	10570	362
1 SNPs per RADtag	10 570	10570	362

SI Table 4 : Pairwise Fst on the final dataset considering population as locality with 10 000 permutations (***: $p < 0.001$, **: $p < 0.01$, * $p < 0.05$) after Bonferroni correction.

	Fenway	Kulo_Lasi	Fatu_Kapa	Fati_Ufu	Phoenix	Tui_Mailila	Tow_Cam	Mangatolo	Abe	
	0.37480***	0.00091	0.00151	0.00251	0.00146	0.00090	0.00121	0.00228	0.00000	Abe
	0.38261***	-0.00197	-0.00114	-0.00129	-0.00013	-0.00023	-0.00195	0.00000		Mangatolo
	0.37391***	-0.00286	-0.00210	-0.00234	-0.00114	-0.00151	0.00000			Tow_Cam
	0.38257***	-0.00072	-0.00029	-0.00035	0.00067	0.00000				Tui_Mailila
	0.37980***	-0.00088	0.00005	-0.00072	0.00000					Phoenix
	0.38188***	-0.00216	-0.00082	0.00000						Fati_Ufu
	0.38095***	-0.00194	0.00000							Fatu_Kapa
	0.37566***	0.00000								Kulo_Lasi
	0.00000									Fenway
										North_Su
										Romans_Rui
										Solwara_6
										Solwara_7
										Solwara_8
										South_Su
										Suzette
										Scala

Scala	Suzette	South_Su	Solwara_8	Solwara_7	Solwara_6	Romans_Rui	North_Su
0.38245***	0.38948***	0.38701***	0.38276***	0.38807***	0.38831***	0.39661***	0.38481***
0.38945***	0.39449***	0.39225***	0.38881***	0.39331***	0.39402***	0.40040***	0.39064***
0.38116***	0.38722***	0.38517***	0.38088***	0.38611***	0.38687***	0.39388***	0.38329***
0.39011***	0.39462***	0.39245***	0.38903***	0.39379***	0.39445***	0.40128***	0.39034***
0.38647***	0.39177***	0.38946***	0.38594***	0.39068***	0.39114***	0.39751***	0.38750***
0.38849***	0.39329***	0.39114***	0.38774***	0.39208***	0.39289***	0.39910***	0.38914***
0.38706***	0.39312***	0.39095***	0.38726***	0.39191***	0.39226***	0.39864***	0.38888***
0.38193***	0.38906***	0.38658***	0.38272***	0.38745***	0.38823***	0.39429***	0.38488***
-0.00096	-0.00192	-0.00177	-0.00156	-0.00158	-0.00114	-0.00264	-0.00168
-0.00110	-0.00162	-0.00136	-0.00180	-0.00142	-0.00099	-0.00117	0.00000
-0.00170	-0.00080	-0.00146	-0.00227	-0.00135	-0.00266	0.00000	
-0.00117	-0.00130	-0.00102	-0.00093	-0.00046	0.00000		
-0.00054	-0.00079	-0.00093	-0.00067	0.00000			
-0.00101	-0.00132	-0.00138	0.00000				
-0.00067	-0.00102	0.00000					
-0.00166	0.00000						
0.00000							

SI Table 5: Outlier loci detected by the 4 methods at two different thresholds. (*p*-value Threshold 0.05, 0.01 and expected number of outlier (plus expected upper bound for Arlequin)

Population	N° SNPs	Bayescan	PCAdapt	Arlequin	Baypass	Threshold	Expected at the threshold	Expected at the threshold upper bound
NF/F/L	8254	0	503	225	637	0.05	413	207
M/W	9897	0	565	473	766	0.05	495	248
NF/F/L	8254	0	163	64	177	0.01	83	42
M/W	9897	0	162	94	286	0.01	99	50

SI Table 6: Gene annotation associated with outlier loci obtained at the 5% threshold after being blasted on the *Alviniconcha boucheti* transcriptome.

Accessio	Bit-	E	Organism	Description	GO Biological process
n N°	Score	Value			
XP_041	76.24	1.06e-	Gigantopelta		
357606	05	17	aegis	uncharacterized protein LOC121374568	Not determined
XP_034	55.55	3.36e-	Pangasianodon	NACHT, LRR and PYD domains-containing	innate immune
160643	23	08	hypophthalmus	protein 3-like, partial	response
XP_013	86.92	2.23e-	Biomphalaria	PREDICTED: uncharacterized protein	
078170	95	22	glabrata	LOC106064217	Not determined
XP_013	86.92	2.23e-	Biomphalaria	PREDICTED: uncharacterized protein	
078170	95	22	glabrata	LOC106064217	Not determined
XP_013	86.92	2.23e-	Biomphalaria	PREDICTED: uncharacterized protein	
078170	95	22	glabrata	LOC106064217	Not determined
WP_180	119.6	5.46e-	Bathymodiolus		DNA/RNA
337615	86	35	brooksi	reverse transcriptase family protein, partial	replication/repair
XP_041	73.13	2.42e-	Lytechinus		
480387	73	13	variegatus	uncharacterized protein LOC121427880	Not determined
XP_038	68.65	6.11e-			
052343	49	14	Patiria miniata	uncharacterized protein LOC119725050	Not determined
XP_038	68.65	6.11e-			
052343	49	14	Patiria miniata	uncharacterized protein LOC119725050	Not determined
XP_038	68.65	6.11e-			
052343	49	14	Patiria miniata	uncharacterized protein LOC119725050	Not determined

XP_033	101.4	6.91e-			
641004	11	28	Asterias rubens	uncharacterized protein LOC117301244	Not determined
XP_041	150.3	1.77e-	Gigantopelta		
369670	73	45	aegis	uncharacterized protein K02A2.6-like	Not determined
XP_005	117.9	9.18e-	Aplysia		
088965	62	32	californica	uncharacterized protein LOC101855985	Not determined
XP_025	110.7	8.03e-	Pomacea	uncharacterized protein LOC112555363	
079513	21	27	canaliculata	isoform X3	Not determined
XP_025	172.0	2.97e-	Pomacea		
077015	96	53	canaliculata	enhancer of mRNA-decapping protein 3-like	deadenylation of mRNA
XP_025	86.92	1.33e-	Pomacea	uncharacterized protein LOC112554579	
078203	95	17	canaliculata	isoform X1	Not determined
XP_025		6.26e-	Pomacea		
083764	271.4	89	canaliculata	uncharacterized protein LOC112557883	Not determined
XP_014	77.96	5.86e-	Poecilia	PREDICTED: RNA-directed DNA polymerase	DNA/RNA
915915	46	18	latipinna	from mobile element jockey-like	replication/repair
XP_025	175.1	2.94e-	Pomacea		
107354	99	52	canaliculata	uncharacterized protein LOC112572055	Not determined
XP_035	62.10	4.15e-	Branchiostoma		
684824	36	07	floridae	hemacentin-1-like	cell division
XP_025	268.2	2.70e-	Pomacea	FGFR1 oncogene partner 2 homolog isoform	positive regulation of
086588	96	95	canaliculata	X1	intracellular signalling
XP_042	61.41	7.19e-	Plectropomus	uncharacterized protein LOC121950173,	monocarboxylic acid
352033	4	12	leopardus	partial	transport
XP_041	52.79	3.91e-	Gigantopelta		
362065	39	07	aegis	uncharacterized protein LOC121378034	Not determined
XP_027	95.20	1.46e-	Panaeus		
206803	48	22	vannamei	uncharacterized protein LOC113800258	Not determined
XP_020	91.06	1.85e-	Exaiptasia		
901902	71	18	diaphana	uncharacterized protein LOC110240429	Not determined
XP_006	175.1	1.15e-	Saccoglossus	PREDICTED: uncharacterized protein	
813392	99	49	kowalevskii	LOC102810040	Not determined
XP_012	73.82	8.03e-	Aplysia		monocarboxylic acid
945111	69	12	californica	monocarboxylate transporter 7, partial	transport
XP_025	502.7		Pomacea	RNA polymerase-associated protein RTF1	DNA/RNA
104366	63	0	canaliculata	homolog	replication/repair
XP_025	194.8	2.18e-	Pomacea		endocytosis,
082792	53	52	canaliculata	cubilin-like	cholesterol transport
XP_041	382.0	6.62e-	Gigantopelta	zinc finger C3H1 domain-containing protein-	
350627	82	119	aegis	like isoform X2	RNA processing
XP_041	143.4	9.10e-	Gigantopelta	28S ribosomal protein S14, mitochondrial-	
360706	77	46	aegis	like	protein translation

XP_025	267.2	8.01e-	Pomacea	uncharacterized protein LOC112577155	
115929	62	87	canaliculata	isoform X1	Not determined
XP_025	686.5		Pomacea	calcium-responsive transcription factor-like	
098569	44	0	canaliculata	isoform X2	transcription regulation
XP_025	563.7		Pomacea		lipoprotein signal
081068	94	0	canaliculata	G-protein coupled receptor GRL101-like	transduction
XP_025	495.8	3.24e-	Pomacea	uncharacterized protein LOC112571901	
107077	67	179	canaliculata	isoform X1	Not determined
XP_025	584.4		Pomacea	uncharacterized protein LOC112555421	
079593	82	0	canaliculata	isoform X2	Not determined
XP_025	717.9		Pomacea	spermatogenesis-associated protein 5-like	
087068	21	0	canaliculata	isoform X2	spermatogenesis
XP_025	1200.		Pomacea		proteasome catabolic
115788	3	0	canaliculata	E3 ubiquitin-protein ligase MARCH6-like	process
XP_025	673.4		Pomacea	nose resistant to fluoxetine protein 6-like	
112929	41	0	canaliculata	isoform X2	lipid transport
XP_025	225.5	3.64e-	Pomacea	tripartite motif-containing protein 2-like	proteasome catabolic
081703	41	69	canaliculata	isoform X1	process
XP_025	460.6	3.17e-	Pomacea		DNA/RNA
090127	97	148	canaliculata	protein MLP1-like	replication/repair
XP_025	472.7	1.80e-	Pomacea		
087125	65	174	canaliculata	male-specific lethal 3 homolog isoform X1	chromatin organization
XP_025	970.3		Pomacea	transient receptor potential cation channel	
107409	18	0	canaliculata	subfamily V member 5-like	calcium ion transport
XP_025	814.1		Pomacea		
082925	21	0	canaliculata	PHD finger protein 14-like	transcription regulation
XP_025	1423.		Pomacea	exocyst complex component 4-like isoform	
095023	39	0	canaliculata	X5	excytosis
XP_025	919.9		Pomacea	myogenesis-regulating glycosidase-like	carbohydrate
084399	76	0	canaliculata	isoform X1	metabolic process
XP_025	607.2		Pomacea		
088996	39	0	canaliculata	sorting nexin-17-like isoform X2	endocytosis regulation
XP_025	762.4		Pomacea	putative uncharacterized zinc finger protein	hematopoietic cell
076534	01	0	canaliculata	814	synthesis
XP_025	505.1		Pomacea		
098108	77	0	canaliculata	transcription factor Sp3-like	transcription regulation
XP_031	140.7	1.72e-	Oncorhynchus	uncharacterized protein C2orf16-like	
644985	19	33	kisutch	isoform X11	not determined
XP_041	826.5		Gigantopelta	sodium- and chloride-dependent taurine	amino-acid transport,
356928	34	0	aegis	transporter-like	symbiosis
XP_025	1461.		Pomacea		DNA/RNA
084483	32	0	canaliculata	DNA excision repair protein ERCC-6-like	replication/repair

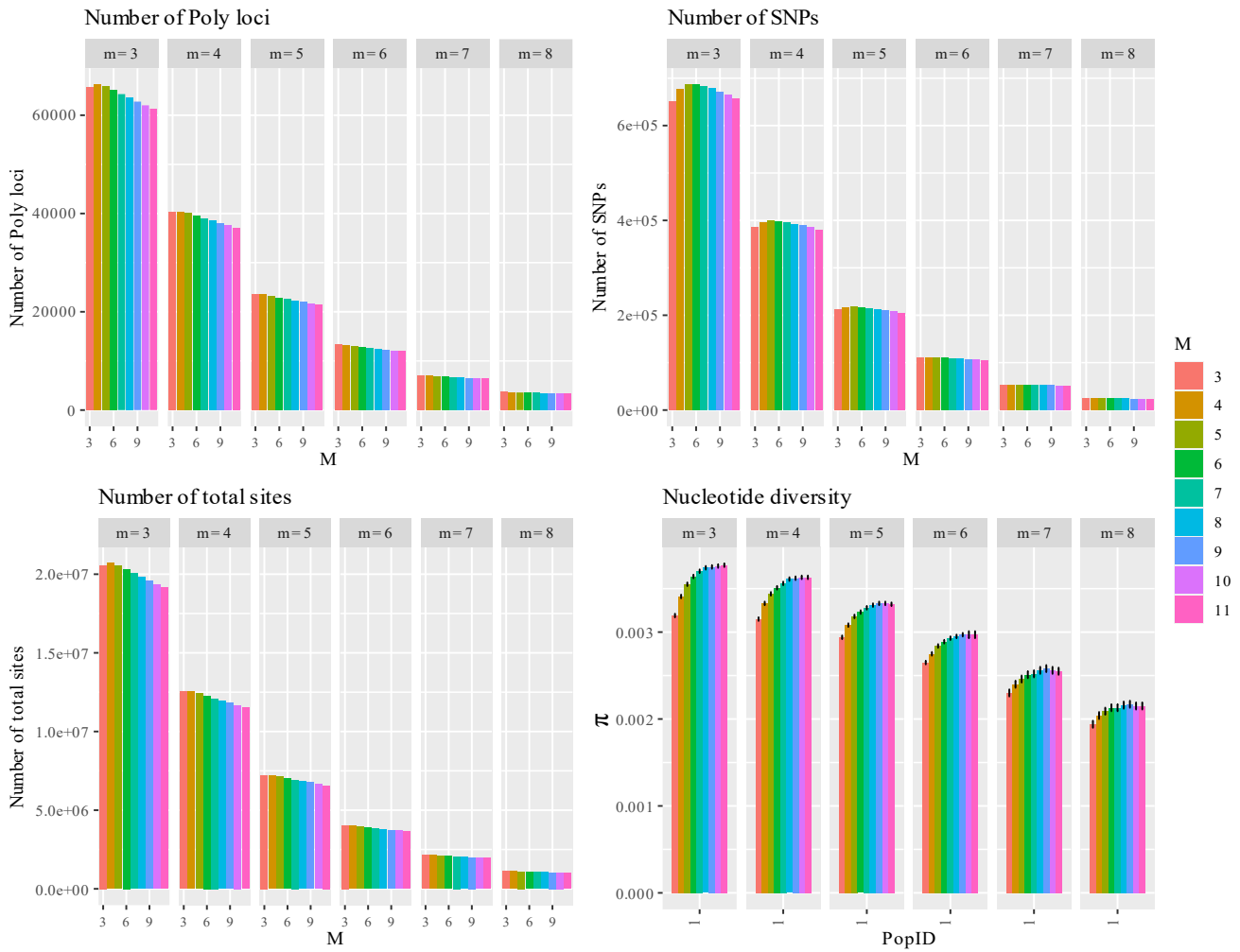
XP_041	1358.		Gigantopelta	transient receptor potential cation channel	ion transmembrane
352044	22	0	aegis	subfamily A member 1 homolog	transport
XP_025	211.0	9.79e-	Pomacea	tRNA (guanine-N(7)-)-methyltransferase	DNA damage, tRNA
085537	59	67	canaliculata	non-catalytic subunit WDR4-like isoform X1	methylation
XP_025	2038.		Pomacea	trafficking protein particle complex subunit	autophagy of
115248	18	0	canaliculata	8-like isoform X3	peroxisome
XP_025	877.2		Pomacea		proteasome catabolic
081732	21	0	canaliculata	E3 ubiquitin-protein ligase MIB2-like	process
XP_025	813.0		Pomacea		
080790	87	0	canaliculata	exosome component 10-like	RNA catabolic pathway
XP_025	1544.		Pomacea		intracellular transport
089931	76	0	canaliculata	DENN domain-containing protein 3-like	regulation
XP_025	267.2	4.63e-	Pomacea		
079669	62	85	canaliculata	uncharacterized protein LOC112555461	not determined
XP_025	1013.		Pomacea		axon morphogenesis
076913	76	0	canaliculata	semaphorin-1A-like	regulation
					amino acid
XP_041	478.6	6.89e-	Gigantopelta	sodium-coupled neutral amino acid	transmembrane
358172	27	176	aegis	transporter 9-like	transport
XP_012	678.9		Aplysia		
939858	58	0	californica	protein SFI1 homolog	mitotic cell regulation
XP_025	1764.		Pomacea		
115131	4	0	canaliculata	uncharacterized protein LOC112576696	not determined
XP_025	1195.		Pomacea		
078860	13	0	canaliculata	unconventional myosin-XVI-like	mitotic cell regulation
XP_041	52.79	3.91e-	Gigantopelta		
362065	39	07	aegis	uncharacterized protein LOC121378034	not determined
XP_013	86.92	2.23e-	Biomphalaria	PREDICTED: uncharacterized protein	
078170	95	22	glabrata	LOC106064217	not determined
XP_013	86.92	2.23e-	Biomphalaria	PREDICTED: uncharacterized protein	
078170	95	22	glabrata	LOC106064217	not determined
XP_037	55.55	1.46e-	Penaeus		
774499	23	07	monodon	uncharacterized protein LOC119571119	not determined
XP_039	62.10	5.10e-	Nilaparvata		
300680	36	10	lugens	uncharacterized protein LOC120355972	not determined
XP_038	68.65	6.11e-			
052343	49	14	Patiria miniata	uncharacterized protein LOC119725050	not determined
XP_038	68.65	6.11e-			
052343	49	14	Patiria miniata	uncharacterized protein LOC119725050	not determined
XP_033	101.4	6.91e-			
641004	11	28	Asterias rubens	uncharacterized protein LOC117301244	not determined

XP_041	96.23	2.14e-	Gigantopelta		calcium channel
361597	92	21	aegis	polycystic kidney disease protein 1-like 2	transport
XP_040	100.7	4.61e-	Gossypium		
966432	22	23	hirsutum	uncharacterized protein LOC121226908	not determined
XP_041	62.79	4.71e-	Gigantopelta		
355387	32	11	aegis	uncharacterized protein LOC121373011	not determined
XP_041	138.9	1.71e-	Lytechinus		
453152	95	39	variegatus	uncharacterized protein LOC121406196	not determined
XP_025	320.7	3.19e-	Pomacea	uncharacterized protein F13E9.13,	
115745	07	117	canaliculata	mitochondrial-like	not determined
XP_025	415.5	6.37e-	Pomacea		heat shock protein,,
082988	28	150	canaliculata	dnaJ homolog subfamily C member 16-like	DNA repair
XP_041	440.0	1.06e-	Gigantopelta	rap1 GTPase-activating protein 1-like	
378159	09	159	aegis	isoform X3	GTPase regulation
XP_027	95.20	1.46e-	Penaeus		
206803	48	22	vannamei	uncharacterized protein LOC113800258	not determined
XP_027	95.20	1.46e-	Penaeus		
206803	48	22	vannamei	uncharacterized protein LOC113800258	not determined
XP_025	100.7	2.44e-	Pomacea		
094800	22	21	canaliculata	uncharacterized protein LOC112564302	not determined
XP_025	100.7	2.45e-	Pomacea		
094800	22	21	canaliculata	uncharacterized protein LOC112564302	not determined
XP_035	227.2	1.42e-	Sander		
862389	65	74	luciperca	zinc finger protein OZF-like	transcription regulation
WP_147	243.1	2.83e-	Alcanivorax		
393938	26	83	profundi	hypothetical protein, partial	not determined
XP_025	656.8		Pomacea	TBC1 domain family member 15-like isoform	
082337	91	0	canaliculata	X3	GTPase regulation
XP_025	207.2	5.45e-	Pomacea		proteoglycan
110655	66	67	canaliculata	carbohydrate sulfotransferase 13-like	biosynthesis
XP_025	292.0	2.39e-	Pomacea		amine metabolic
092467	88	104	canaliculata	sulfotransferase 1A1-like	process
XP_025	598.2		Pomacea		
105426	74	0	canaliculata	cyclin-dependent kinase-like 1 isoform X3	protein kinase activity
XP_012	77.96	7.28e-	Aplysia		
946942	46	16	californica	uncharacterized protein LOC106014226	not determined
WP_223	84.86	8.72e-	Pseudomonas		
227296	06	21	savastanoi	hypothetical protein, partial	not determined
WP_223	84.86	8.72e-	Pseudomonas		
227296	06	21	savastanoi	hypothetical protein, partial	not determined
XP_025	292.7	1.00e-	Pomacea		
114418	78	97	canaliculata	E3 ubiquitin-protein ligase TRIM33-like	transcription regulation

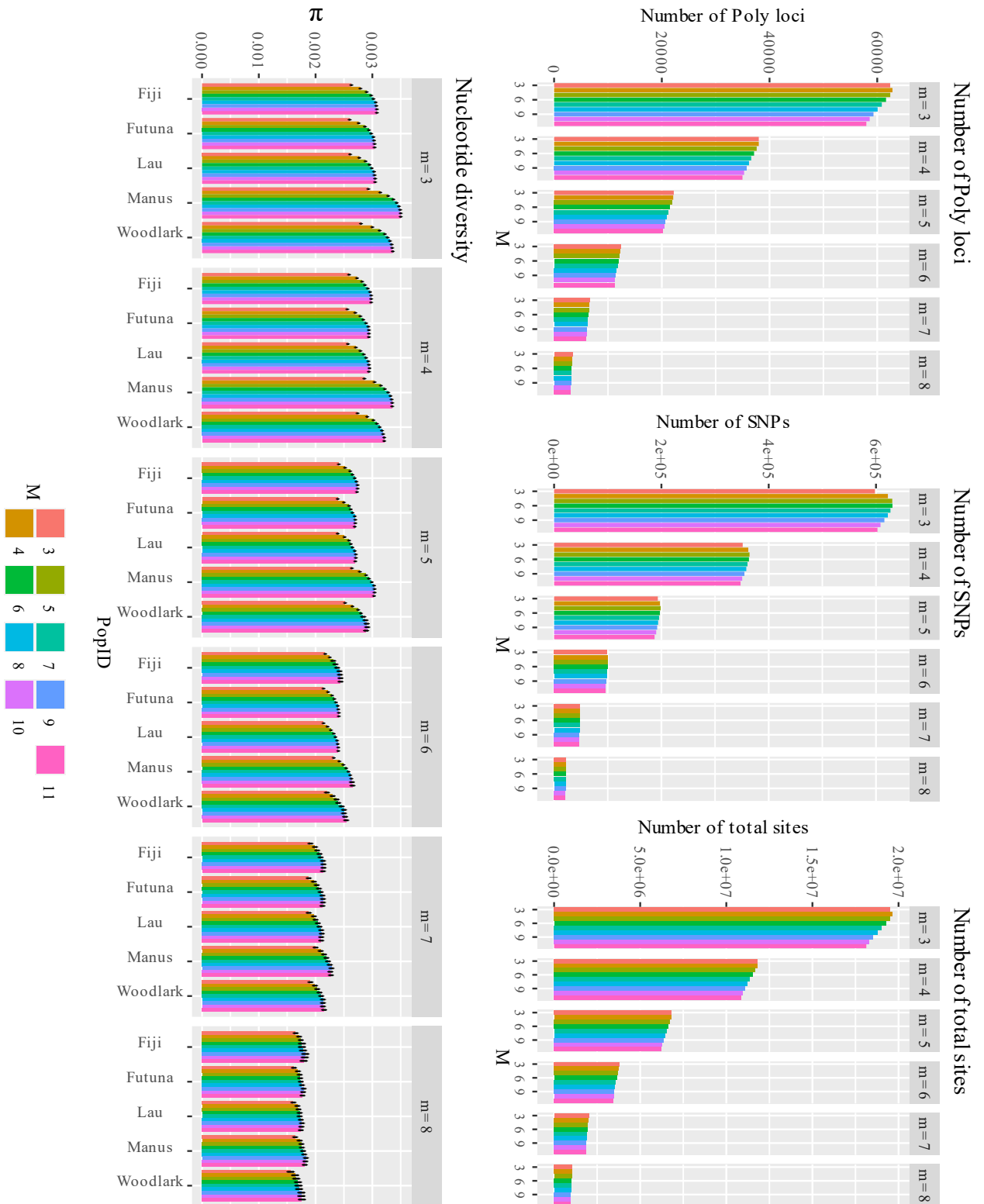
XP_025	224.8	4.86e-	Pomacea			
083620	51	64	canaliculata	zinc finger protein 280C-like isoform X2		transcription regulation
XP_025	485.1		Pomacea	protein pangolin, isoforms A/H/I/S-like		
090987	78	0	canaliculata	isoform X2		Wnt signalling pathway
XP_041	108.3	1.90e-	Alosa	uncharacterized protein LOC121677948		
913036	07	24	sapidissima	[Alosa sapidissima]		not determined
XP_025	264.1	1.53e-	Pomacea			
090237	59	92	canaliculata	ficolin-2-like		immune lectin pathway
XP_025	517.2		Pomacea	neuroepithelial cell-transforming gene 1		
106495	45	0	canaliculata	protein-like isoform X2		GTPase regulation
XP_025	562.0		Pomacea			proteasome catabolic process
104923	7	0	canaliculata	NEDD8 ultimate buster 1-like		
XP_025	913.4		Pomacea			
109240	25	0	canaliculata	radial spoke head 10 homolog B-like		not determined
XP_035	990.6		Aplysia			cell adhesion, axon organization
826375	61	0	californica	teneurin-m, partial		
XP_025	743.0		Pomacea	cytoplasmic dynein 1 intermediate chain 2-		antigen processing, microtubule movement
080467	92	0	canaliculata	like		
XP_025	874.8		Pomacea	putative pre-mRNA-splicing factor ATP-		DNA/RNA
094611	07	0	canaliculata	dependent RNA helicase PRP1 isoform X1		replication/repair
XP_025	926.1		Pomacea			transmembrane transport
087863	83	0	canaliculata	C2 domain-containing protein 5-like		
XP_025	1085.		Pomacea			
076334	48	0	canaliculata	neural-cadherin-like		axon morphogenesis
XP_025	1455.		Pomacea			
096786	46	0	canaliculata	F-actin-monooxygenase MICAL3-like		cell division, exocytosis
XP_025	1455.		Pomacea			
096786	46	0	canaliculata	F-actin-monooxygenase MICAL3-like		cell division, exocytosis
XP_025	1048.		Pomacea			protein localization to plasma membrane
105201	59	0	canaliculata	tetratricopeptide repeat protein 7B-like		protein localization to plasma membrane
XP_025	1048.		Pomacea			
105201	59	0	canaliculata	tetratricopeptide repeat protein 7B-like		
XP_025	693.7		Pomacea			
115787	85	0	canaliculata	zinc finger protein 136-like		transcription regulation
XP_025	1489.		Pomacea			
078070	59	0	canaliculata	chitin synthase chs-2-like		chitin biosynthesis
XP_025	547.5		Pomacea	myosin light chain kinase, smooth muscle-		
098546	88	0	canaliculata	like		muscle contraction
XP_025	1392.		Pomacea			
115071	01	0	canaliculata	integrator complex subunit 8-like		SnRNA transcription
XP_025	514.4	2.91e-	Pomacea	SUN domain-containing ossification factor-		
114019	87	169	canaliculata	like isoform X3		collagen biosynthesis

XP_025	144.1	1.36e-	Pomacea	peroxisome proliferator-activated receptor	adaptative
094438	67	31	canaliculata	gamma coactivator 1-alpha-like	thermogenesis, cellular respiration
XP_025	2085.		Pomacea		
115786	76	0	canaliculata	splicing factor 3B subunit 3-like	mRNA splicing
XP_025	818.2		Pomacea		cell division, exocytosis,
081036	59	0	canaliculata	kinesin-like protein unc-104 isoform X6	reproduction
XP_025	1358.		Pomacea		cell adhesion, cell
092997	57	0	canaliculata	cytospin-A-like isoform X9	division
XP_041	1029.		Gigantopelta		cell adhesion, axon
378258	97	0	aegis	calsyntenin-1-like	organization
XP_025	630.3		Pomacea		spermatogenesis,
091243	41	0	canaliculata	septin-2-like isoform X1	exocytosis regulation
XP_025	734.8		Pomacea		cell adhesion, axon
086850	16	0	canaliculata	protocadherin-1-like	organization
XP_025	962.3		Pomacea	pericentriolar material 1 protein-like	centrosome cycle,
084689	87	0	canaliculata	isoform X2	microtubule anchoring
XP_025	1129.		Pomacea		
097485	62	0	canaliculata	gelsolin-like protein 1	actin filament synthesis
XP_025	2451.		Pomacea		
089019	94	0	canaliculata	neurofibromin-like isoform X4	GTPase regulation
XP_033	629.3		Pecten		
748466	06	0	maximus	protein adenylyltransferase SeIO-like	protein adenylylation
WP_046	217.6	2.98e-	Sulfurovum		
550757	1	63	lithotrophicum	ATP-binding protein	symbiont protein
XP_025	2275.		Pomacea	WD repeat-containing protein 90-like	
111583	75	0	canaliculata	isoform X2	cilium biogenesis
XP_025	415.5	2.47e-	Pomacea		protein localization to
112244	28	133	canaliculata	protein crumbs-like isoform X3	plasma membrane
XP_005	1113.		Aplysia		
107202	07	0	californica	uncharacterized protein LOC101851669	not determined

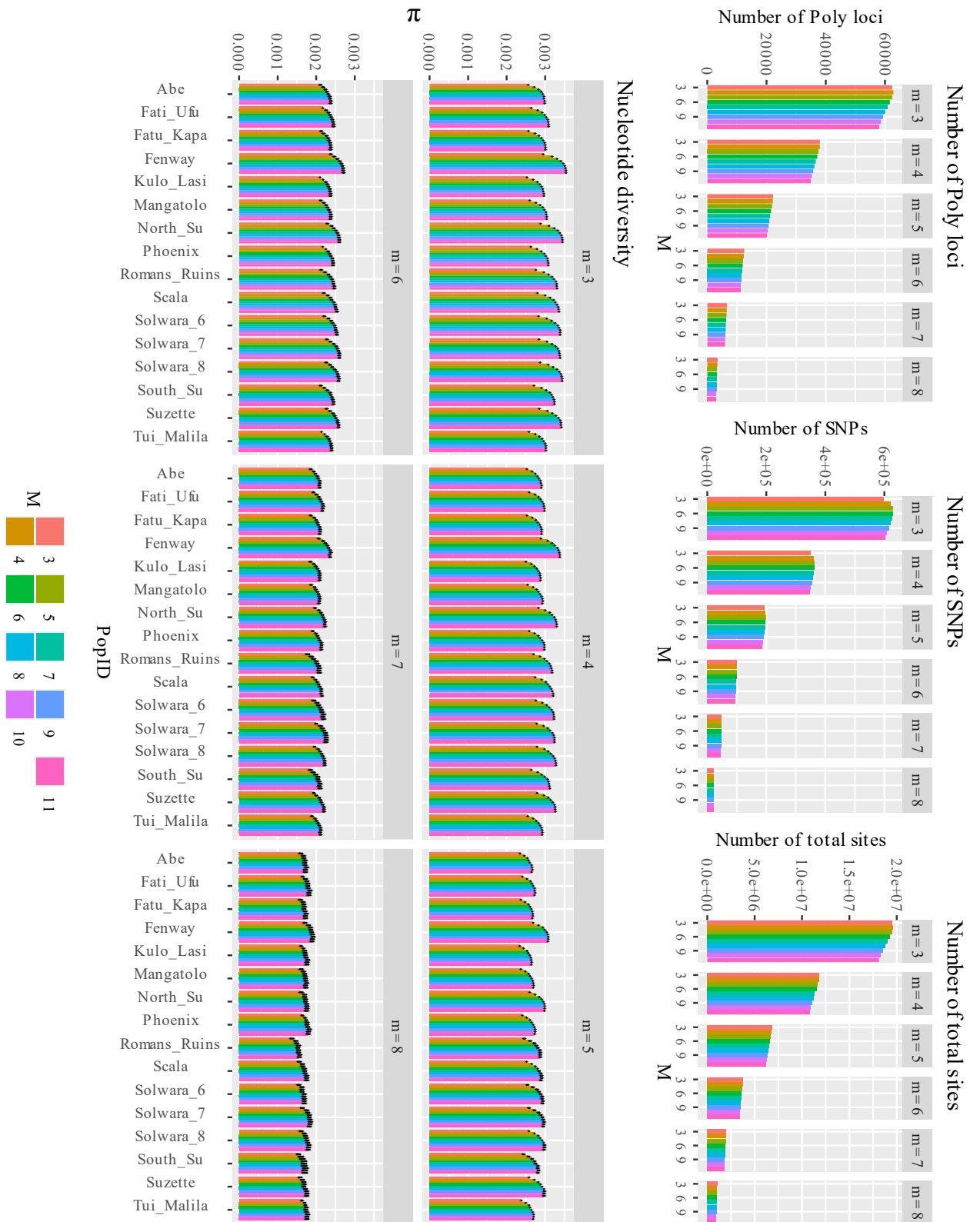
Figures

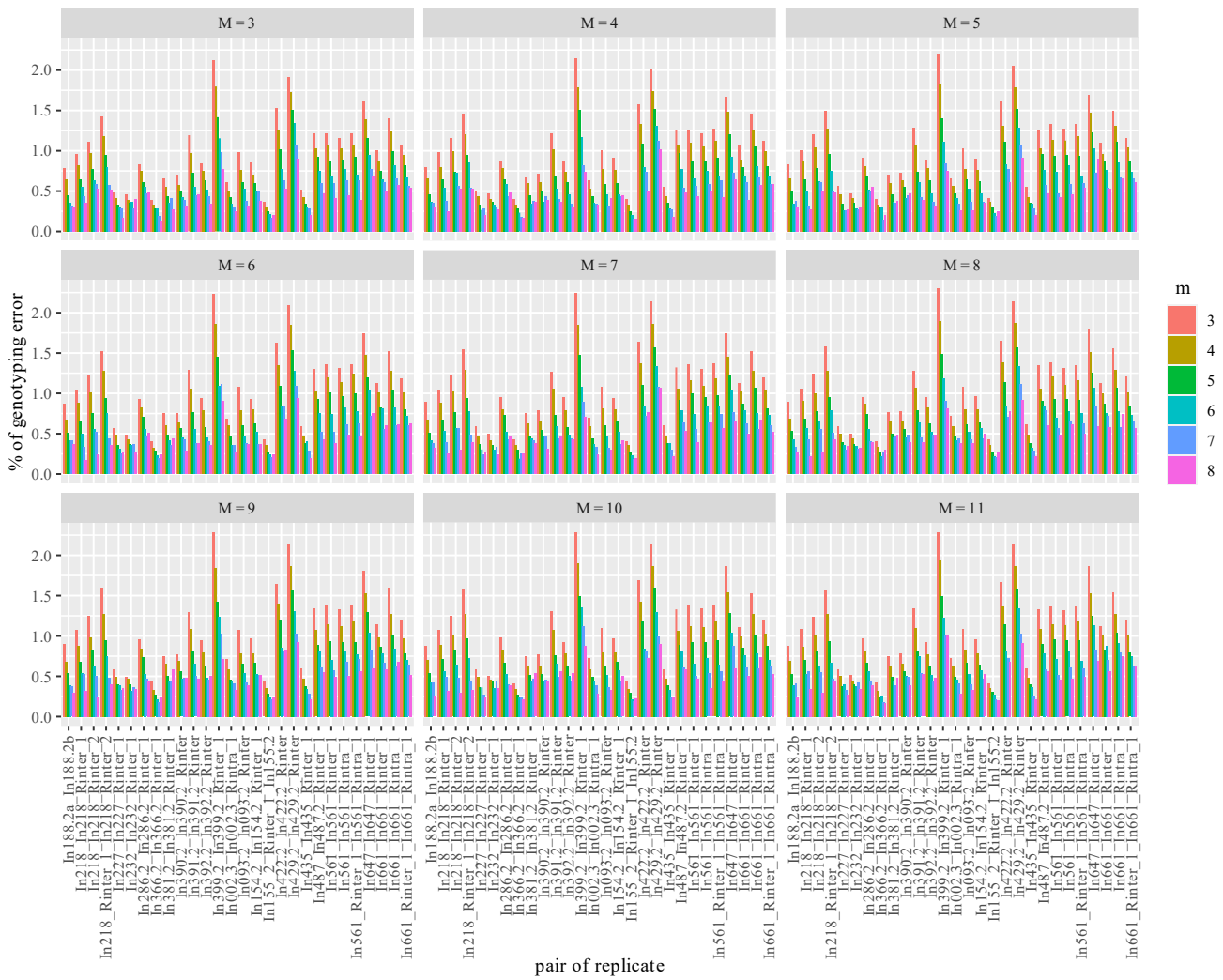


SI Figure 1: Calibration statistics for m (each tiny box represents a value of m) and M ($n=M$) considering all individuals as one population including pair of replicates. the small black bar for the nucleotide diversity represents the standard deviation at 95%)

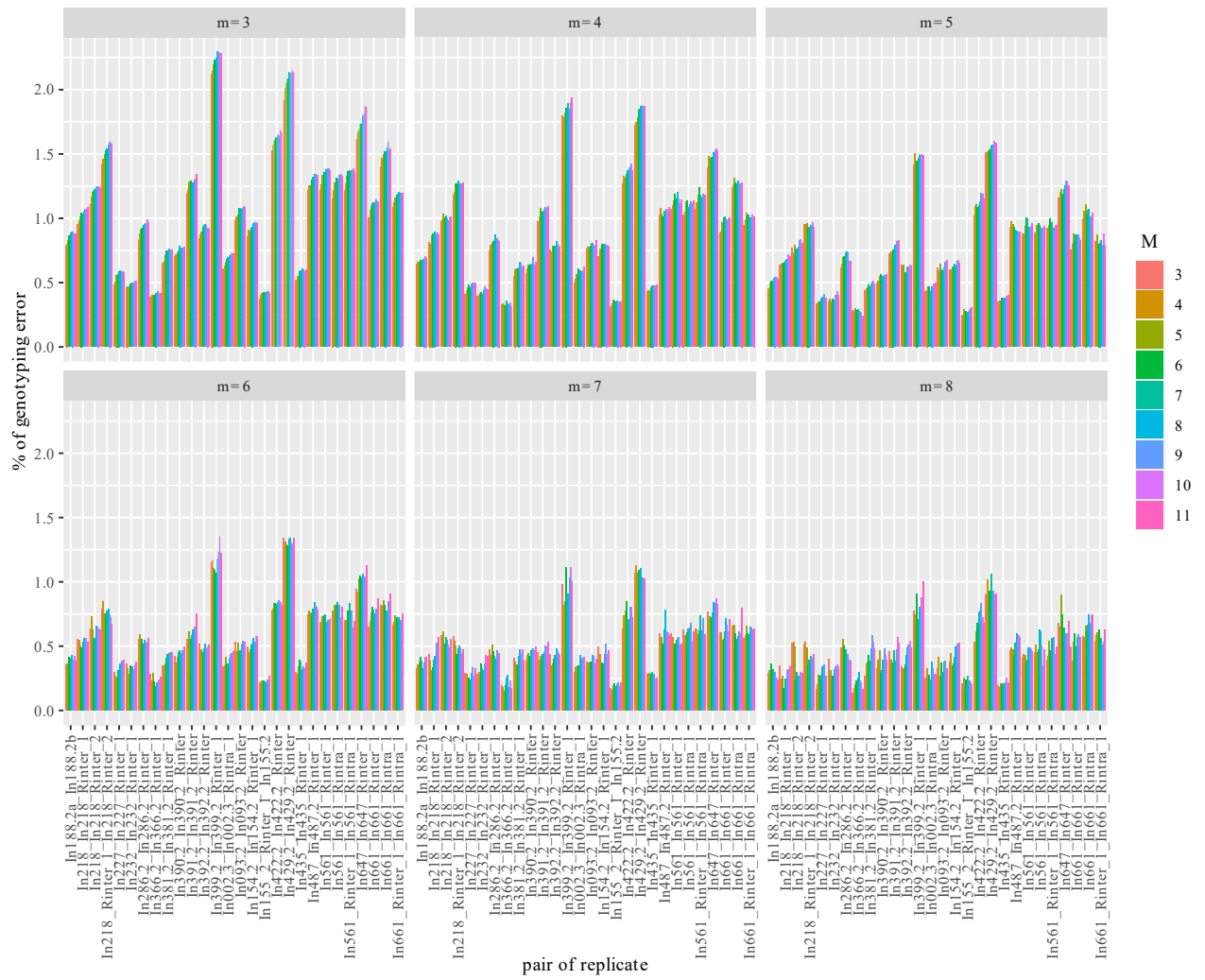


SI figure 2: calibration statistics for m (each box represents a value of m) and M ($n=M$) considering population as basin without including replicate. the small black bar for the nucleotide diversity represents the standard deviation at 95%

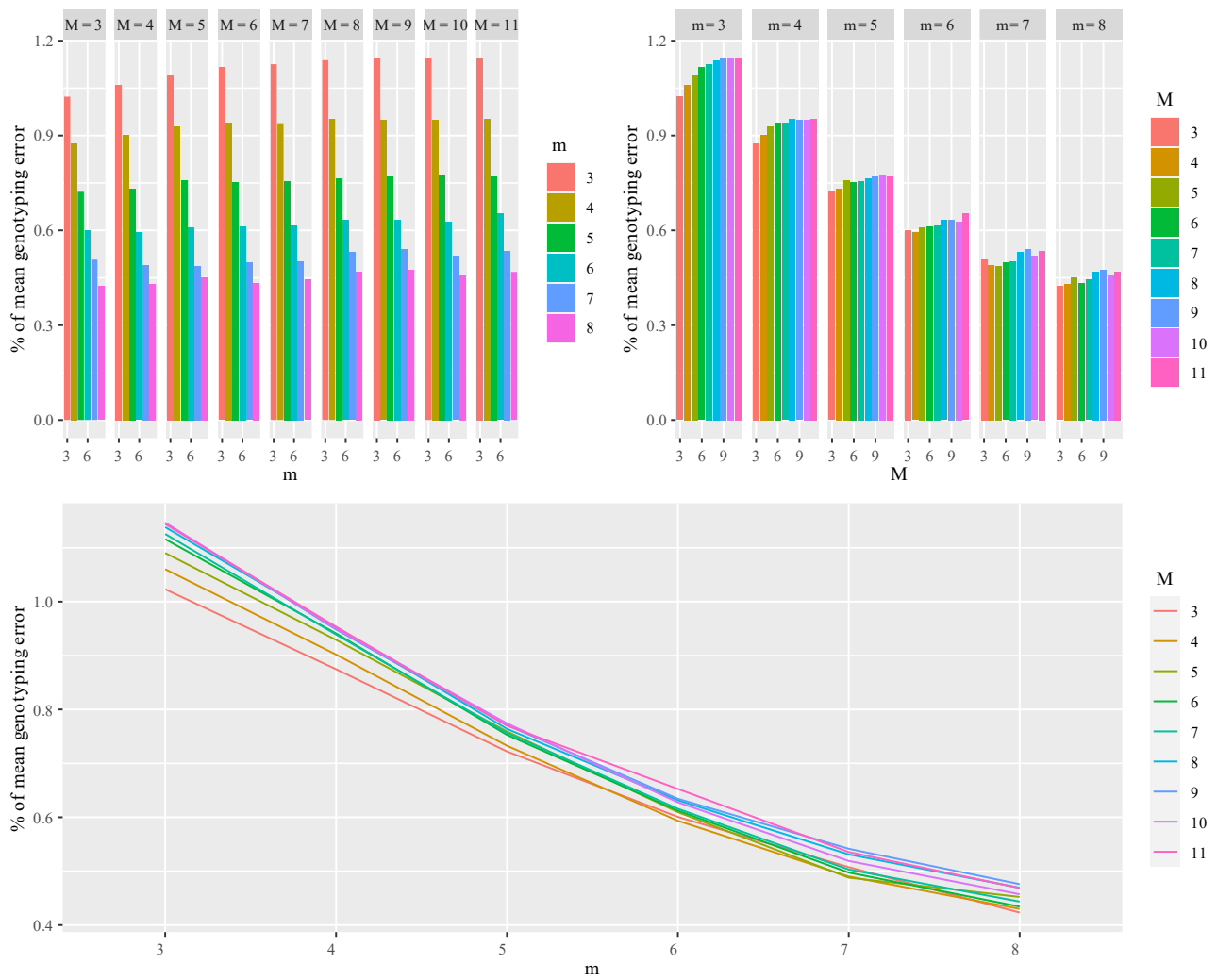




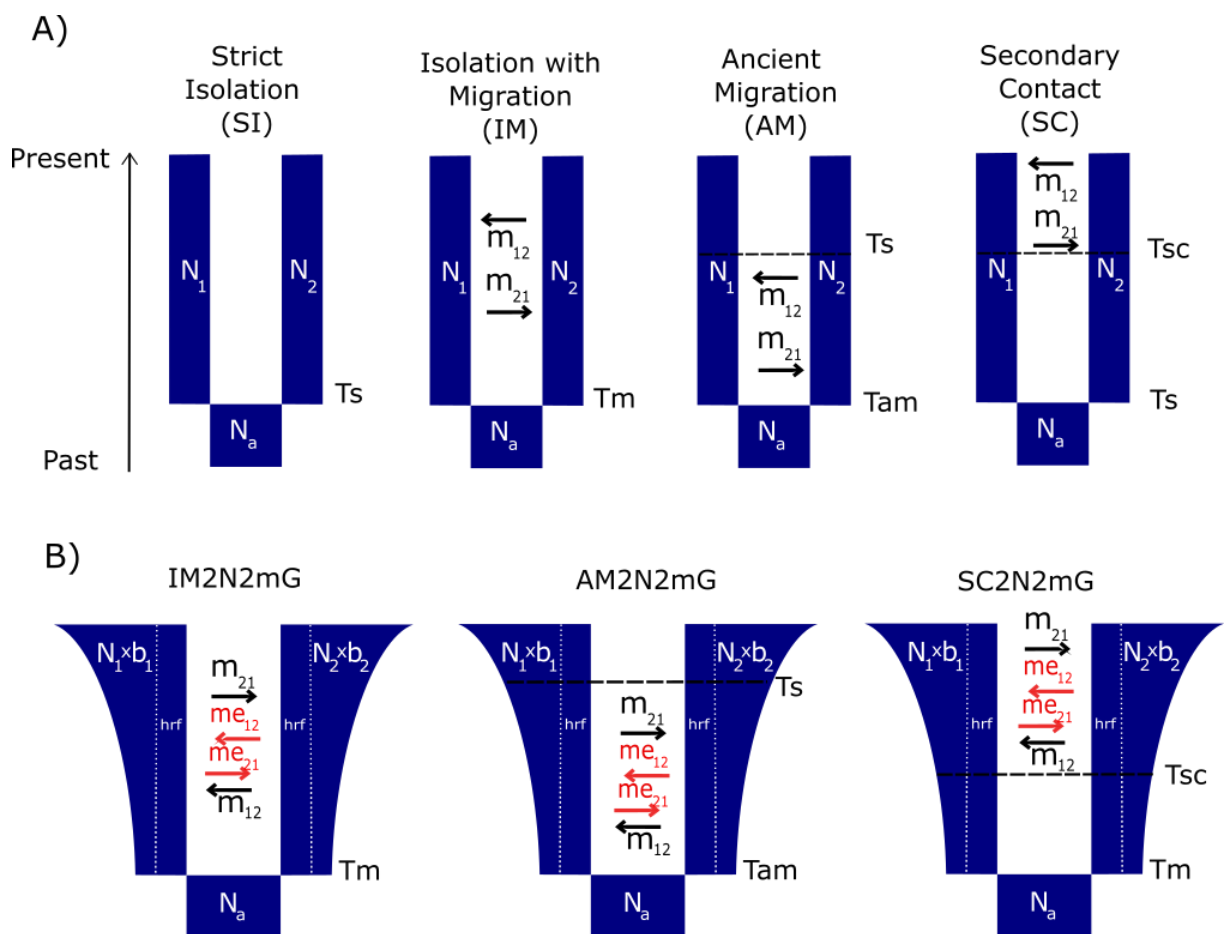
SI figure 4: Percent of genotyping error between pair of replicates for each parameter m and M ($n=M$), each box plot represents a fixed value of M



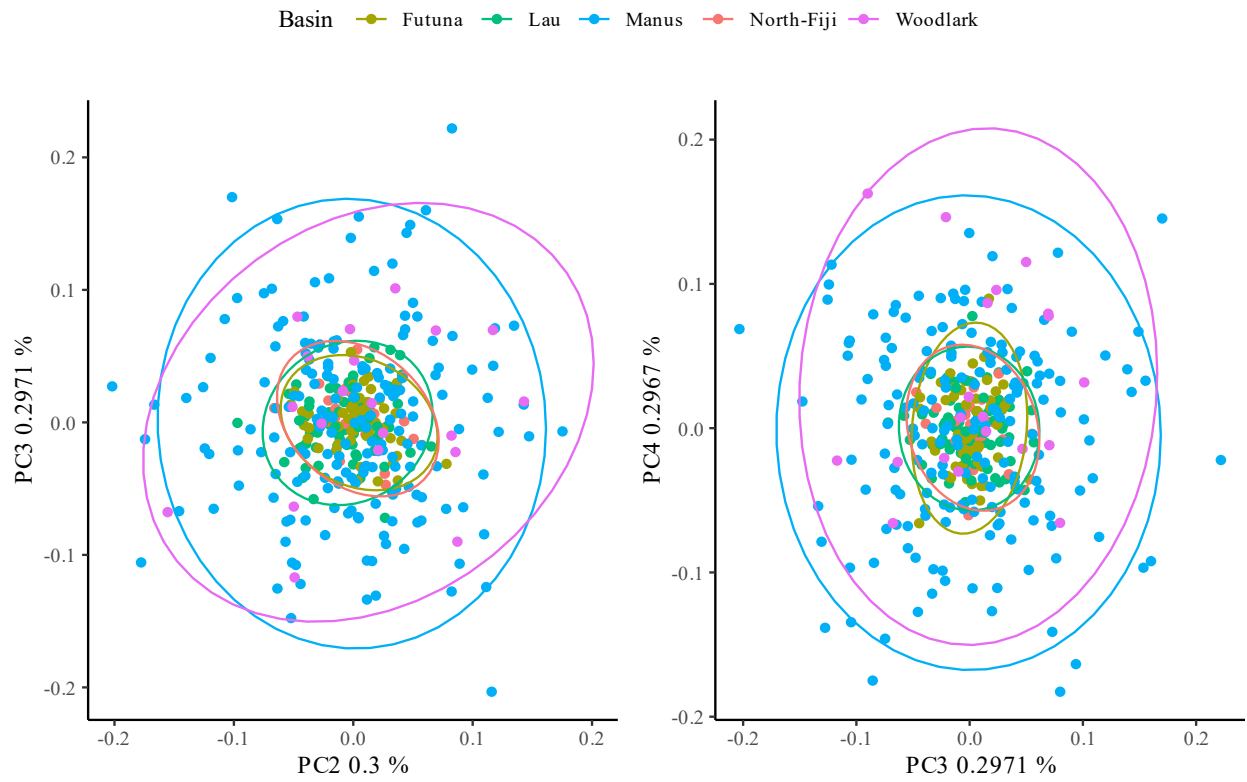
SI figure 5: Percent of genotyping error between pair of replicates for each parameter m (each box plot represents a fixed value of m) and M ($n=M$).



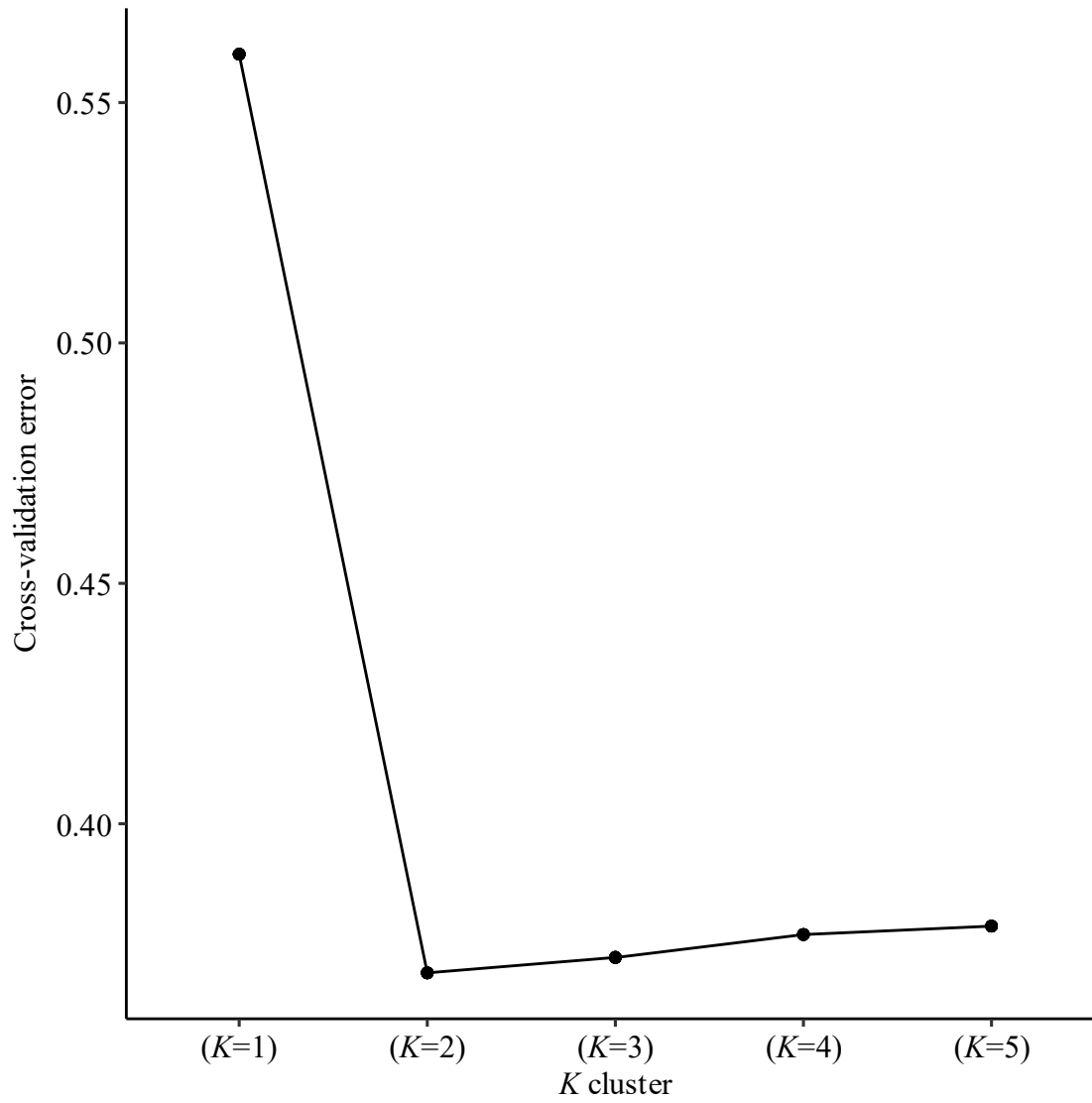
SI Figure 6: Percent of mean genotyping error over all pair of replicates for each parameter m and M ($n=M$)



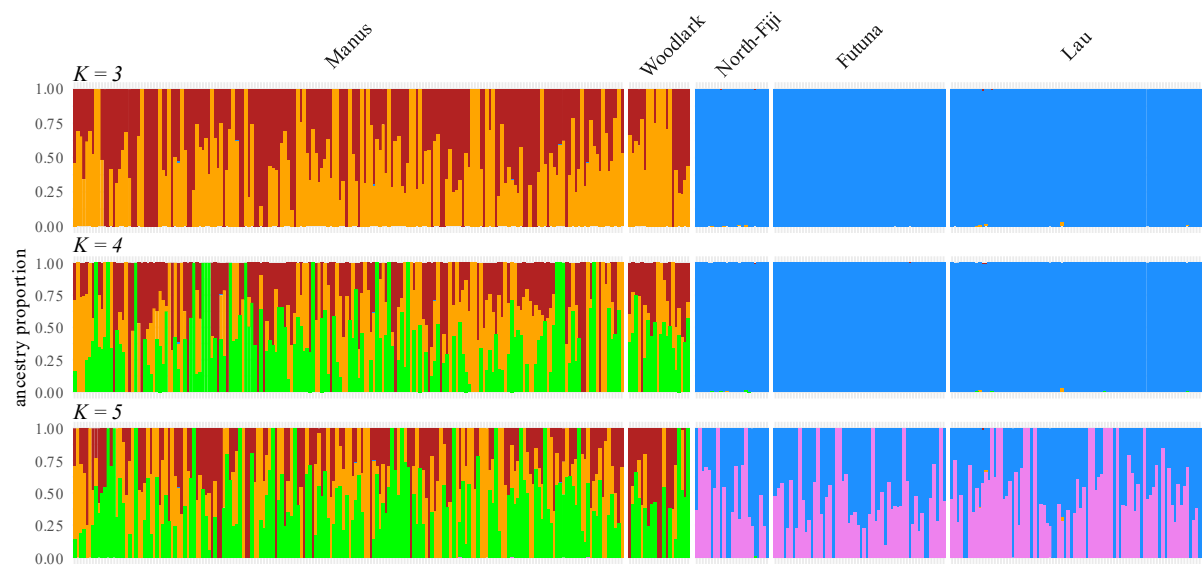
SI Figure 7: Graphical representation of *dadl* model. A) Four basic models. B) Three more complex models. N represents population size; b , population growth factor; hrf , the Hill-Robertson factor, which simulates linked selection; m , unrestricted migration rate; me , restricted migration rate, which simulates barrier loci, N_a ancestral population size, T_s time of split without migration; T_m , time of the split with migration; T_{sc} , time of the secondary contact and T_{am} , time of ancient migration



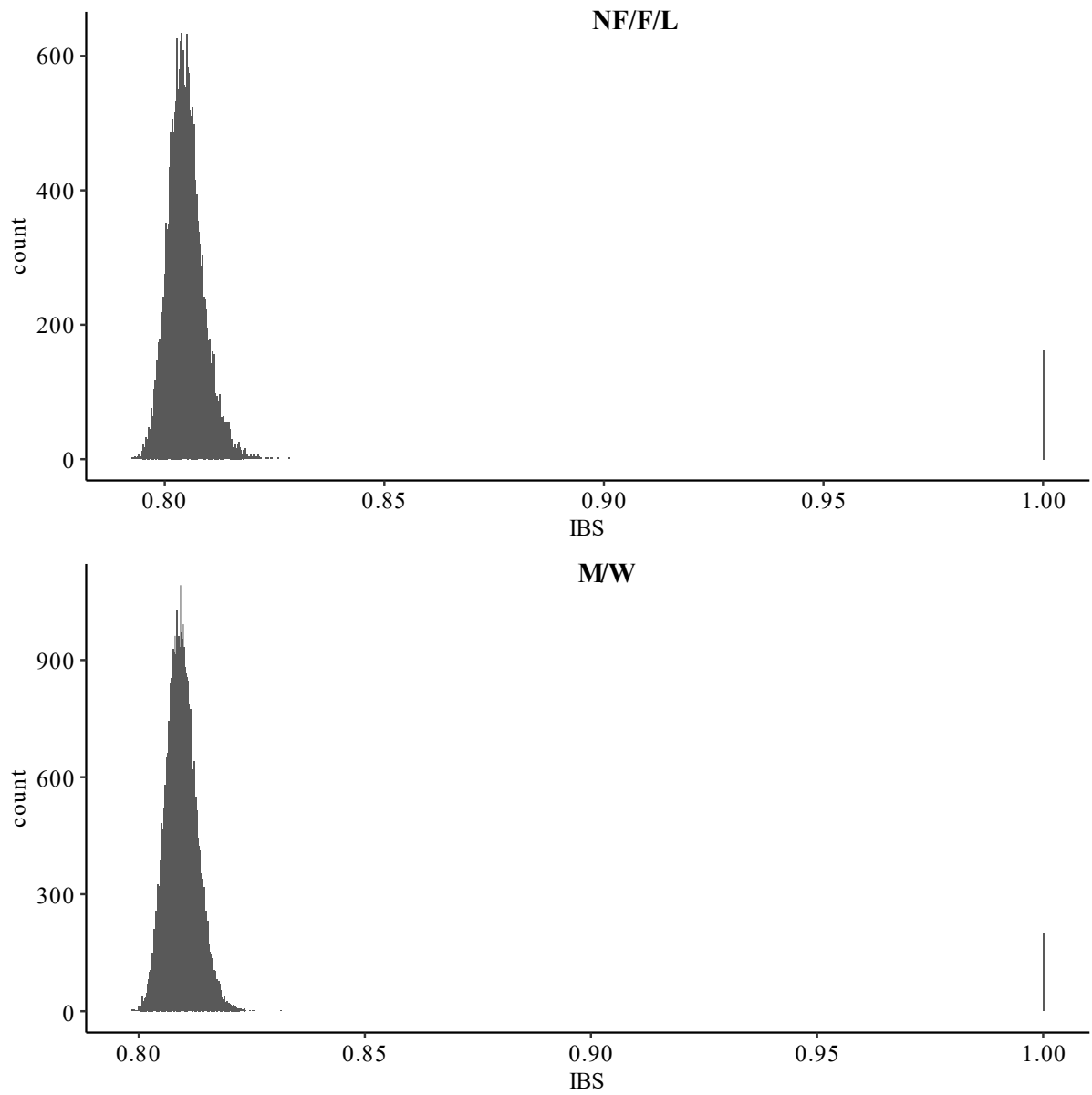
SI Figure 8: PCA analyses on the global dataset (10 570 SNPs), circle represents the multivariate normal distribution of each group of points (basins) at 95%. A) PCA plot with the PC2 and PC3. B) PCA plot with the PC3 and PC4



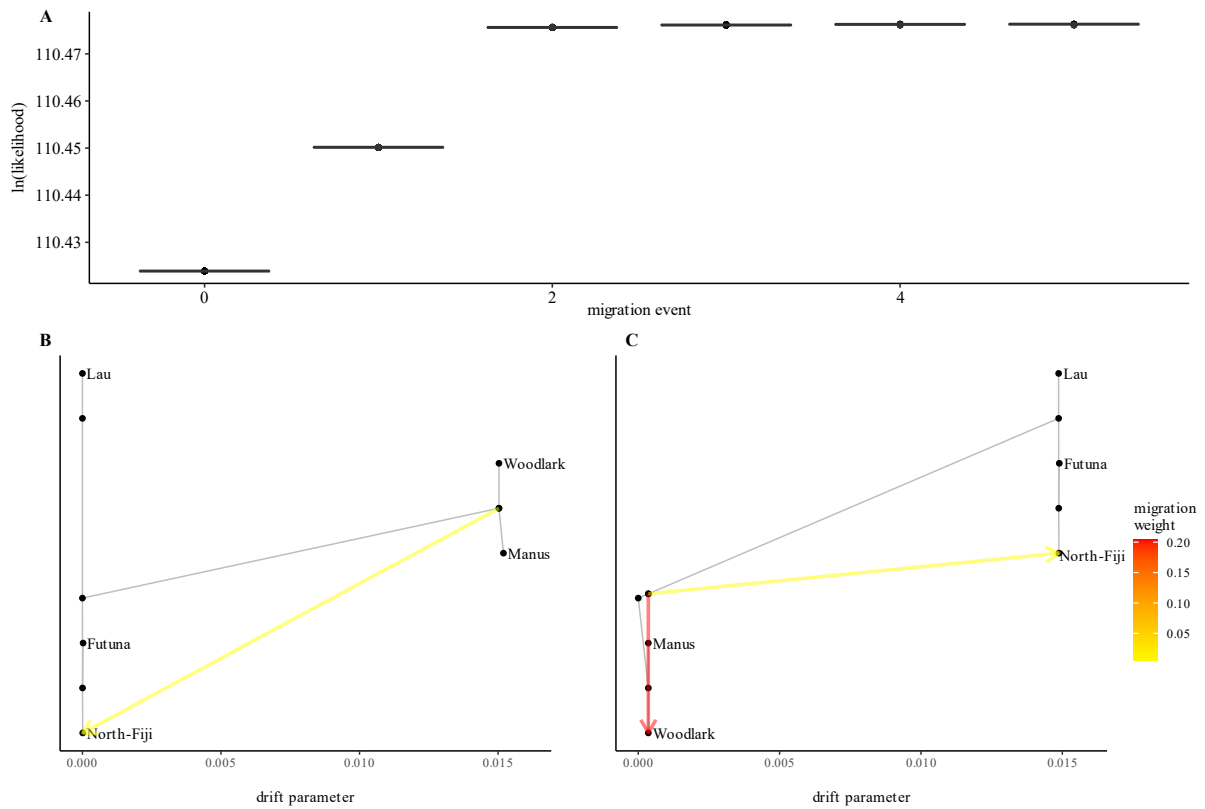
SI Figure 9: Cross-validation error for the 10 independent run for each K from K = 1 to 5 of the ADMIXTURE analysis



SI Figure 10: Admixture plot, each bar represents an individual ancestry proportions obtained with the final dataset for $K = 3$ to 5

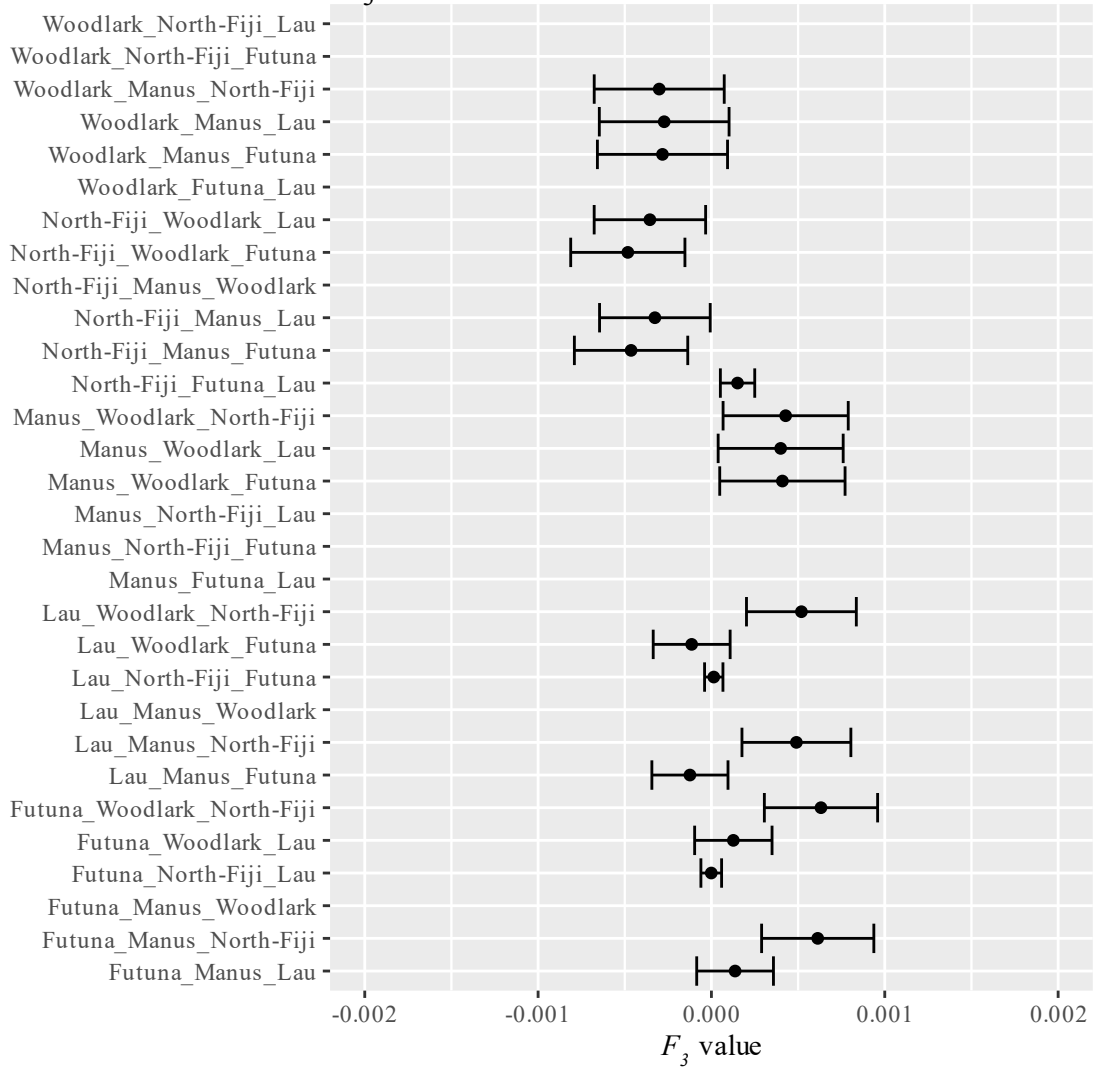


SI Figure 11: histogram of the Identity by state value between individuals. The maximum IBS value of 1 corresponds to the comparison of an individual with itself.

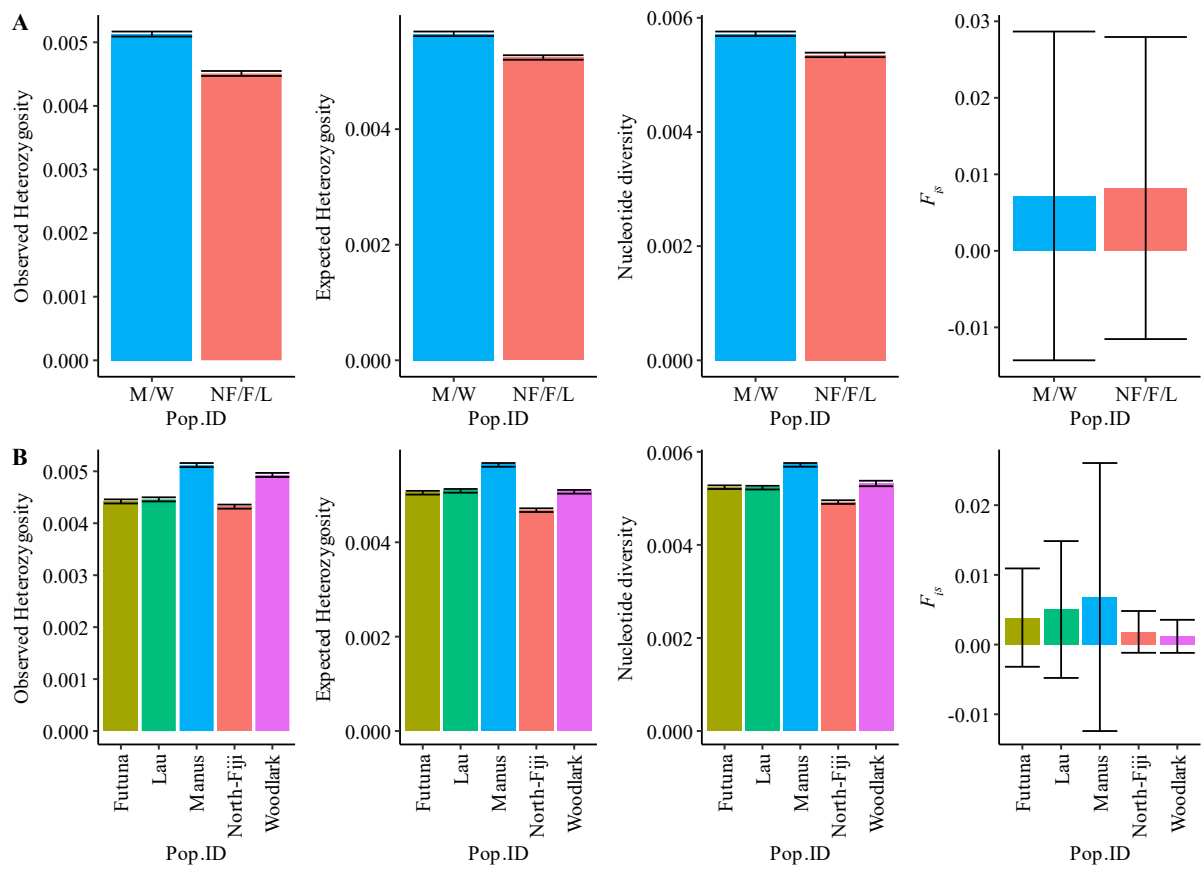


SI Figure 12: Treemix analyses on the global dataset (10 570 SNPs). A) log likelihood over the 10 runs of treemix for each number of migration events. B) Treemix tree with the first migration event/edge. C) Treemix tree with the two first migration events/edges

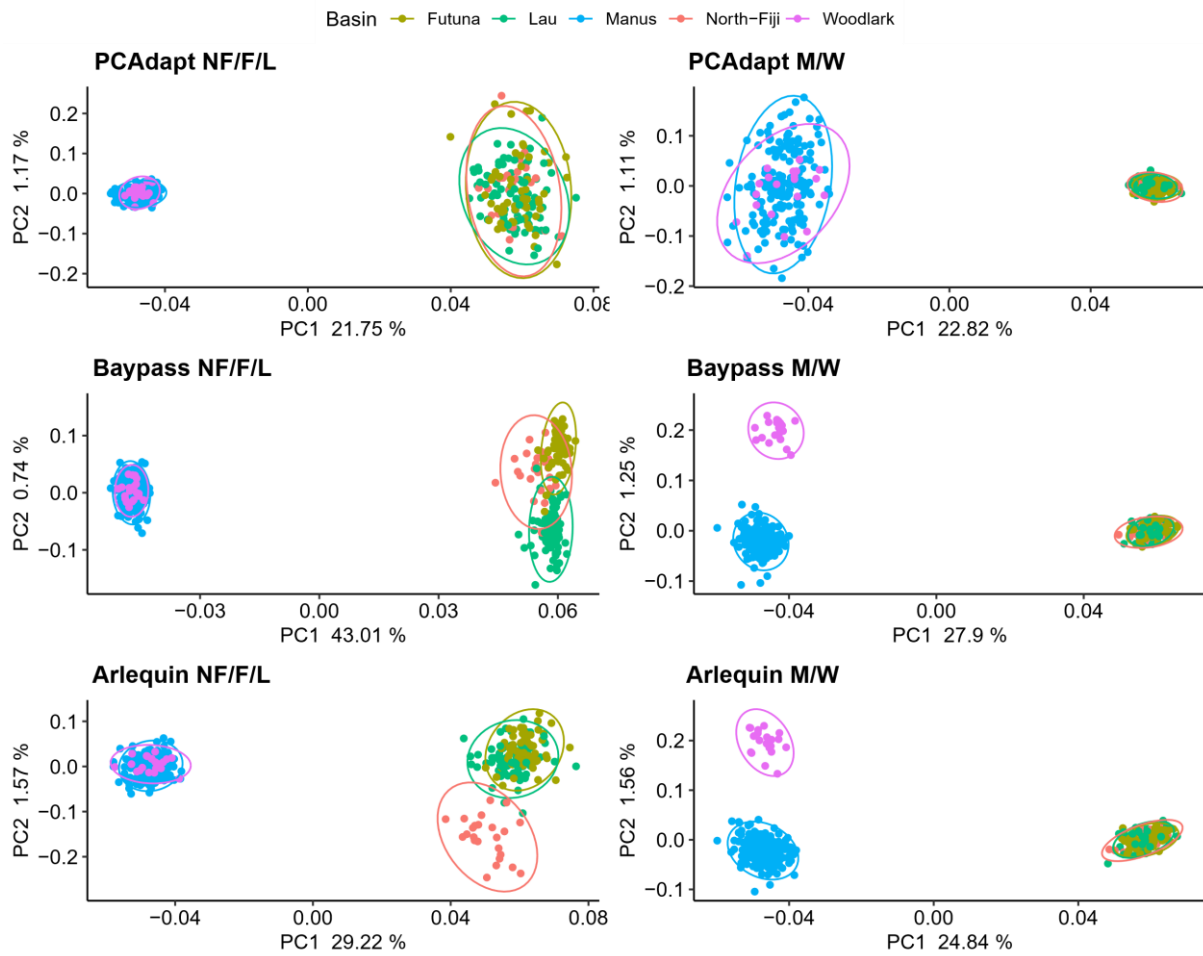
F_3 Statistics with 10 570 SNPs



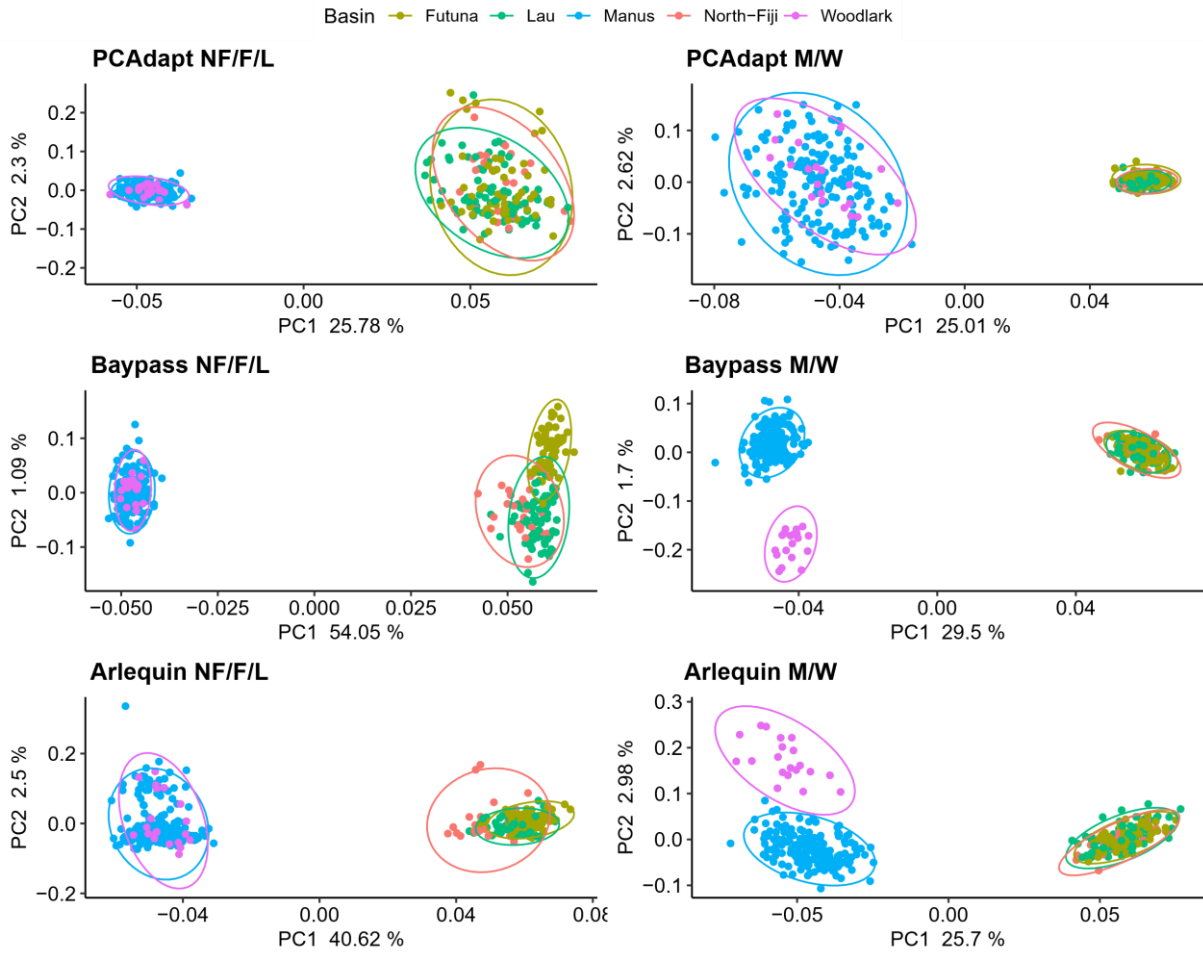
SI Figure 13: f_3 statistic with the 95% standard deviation interval for the 10 570 SNPs near to the value 0. Value that are not display are highly positive. Combination of population is (target population; population source 1, population source 2).



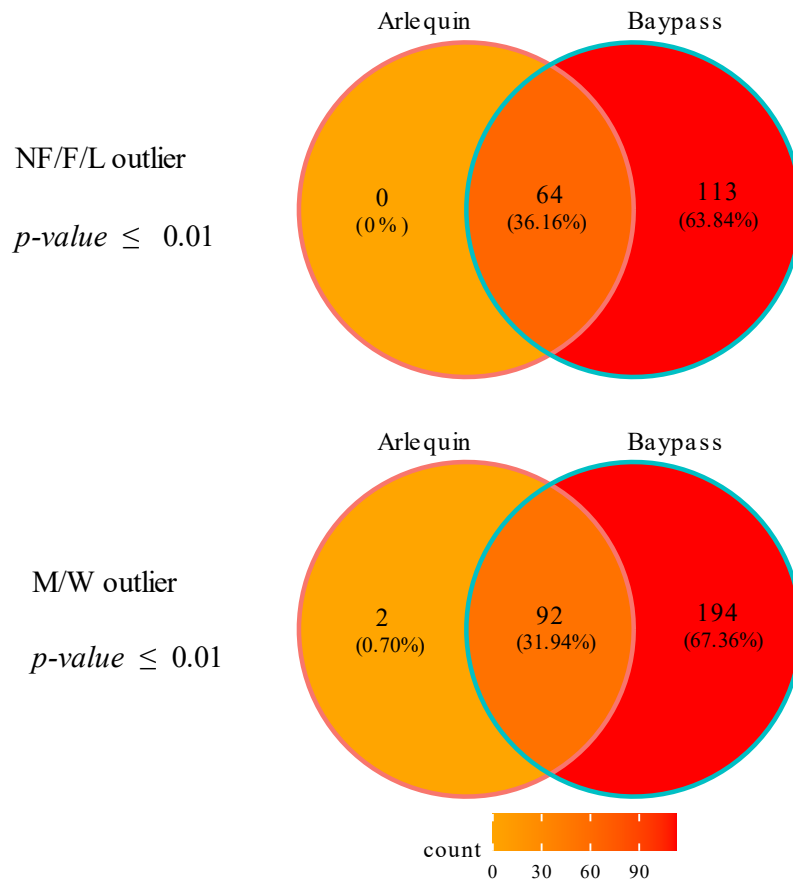
SI Figure 14 : Estimated Observed Heterozygosity (H_o), Expected Heterozygosity (H_e), Nucleotide diversity (π) and the deficit in heterozygosity (F_{is}) with standard deviation at 95% over all positions (variant and non-variants) on the 10 570 radtags loci of the final dataset. (NF/F/L) North-Fiji/Futuna/lau; (M/W) Manus/Woodlark



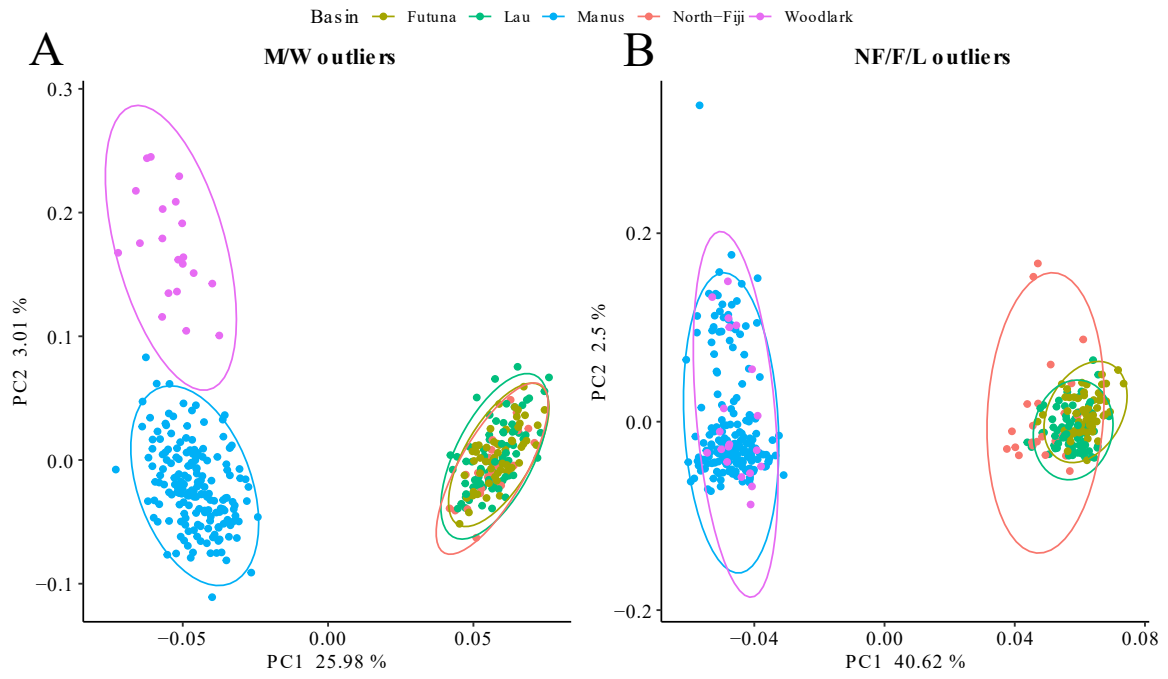
SI Figure 15: PCA for outliers detected by method (PCAdapt, Baypass and Arlequin) at a threshold <0.05 . (NF/F/L) North-Fiji/Futuna/lau; (M/W) Manus/Woodlark. circle represents the multivariate normal distribution of each group of points (basins) at 95%.



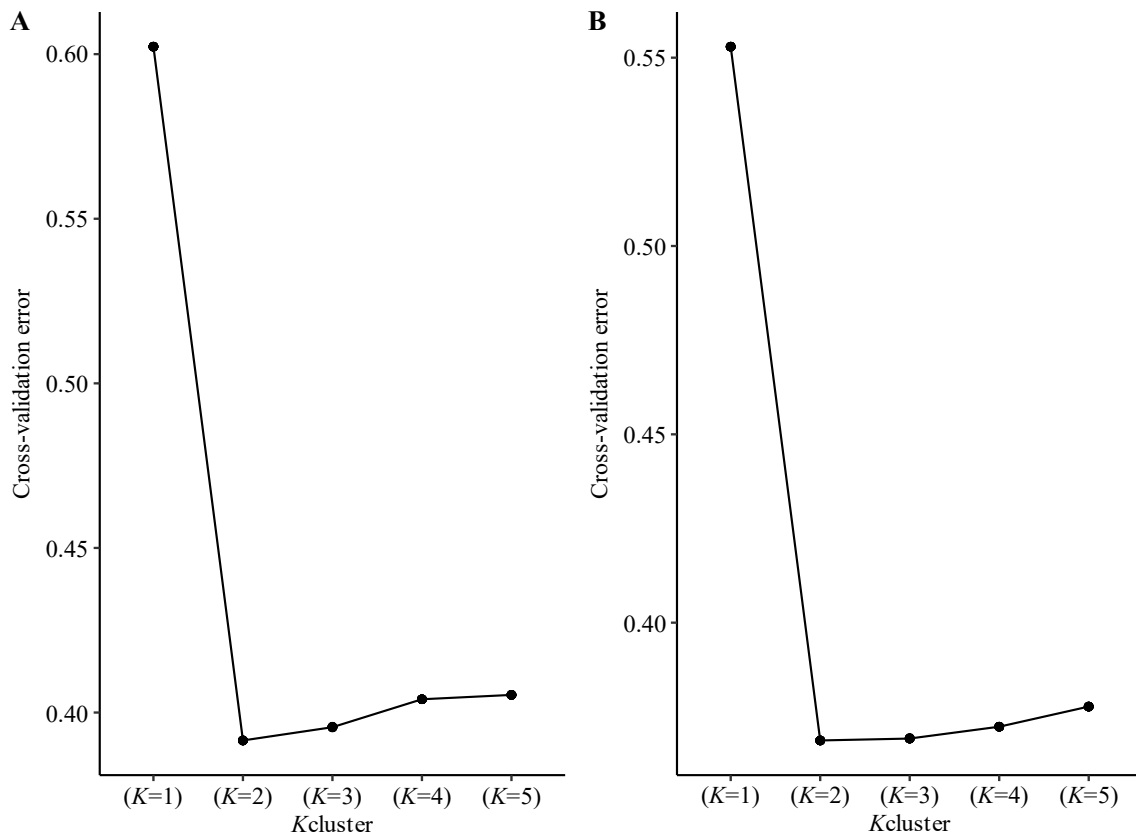
SI Figure 16: PCA with outliers detected by each method independently at the threshold of p -value < 0.01 . (NF/F/L) North-Fiji/Futuna/lau; (M/W) Manus/Woodlark. circle represents the multivariate normal distribution of each group of points (basins) at 95%.



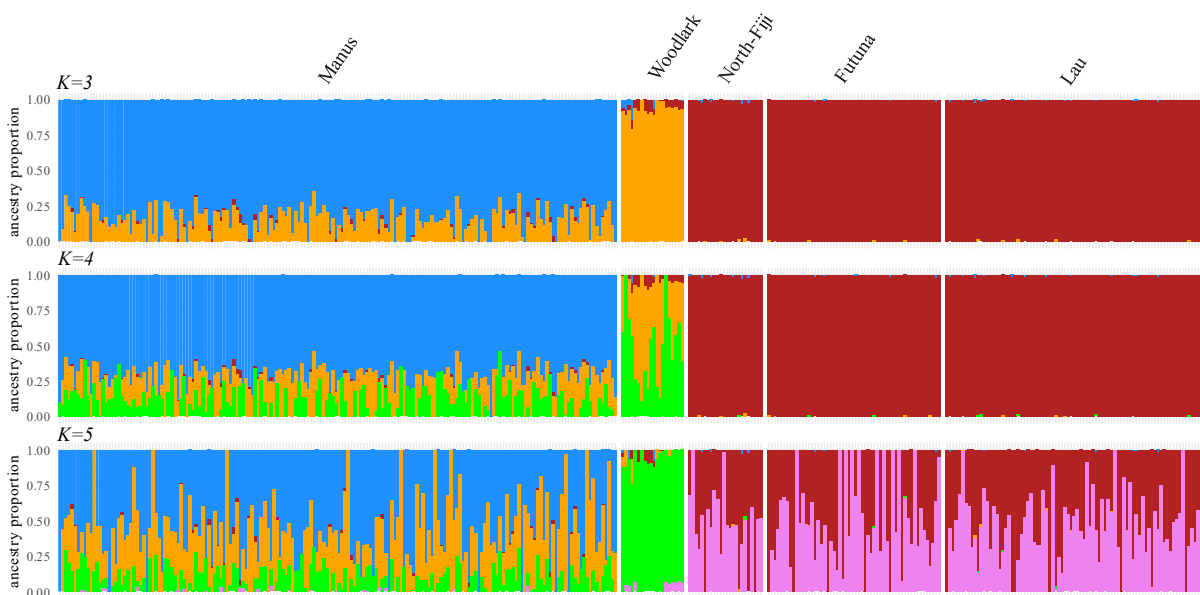
SI Figure 17: Venn diagram of both outlier loci found in Arlequin and Baypass at 0.01 p -value threshold in North-Fiji/Futuna/lau (NF/F/L) and Manus/Woodlark (M/W)



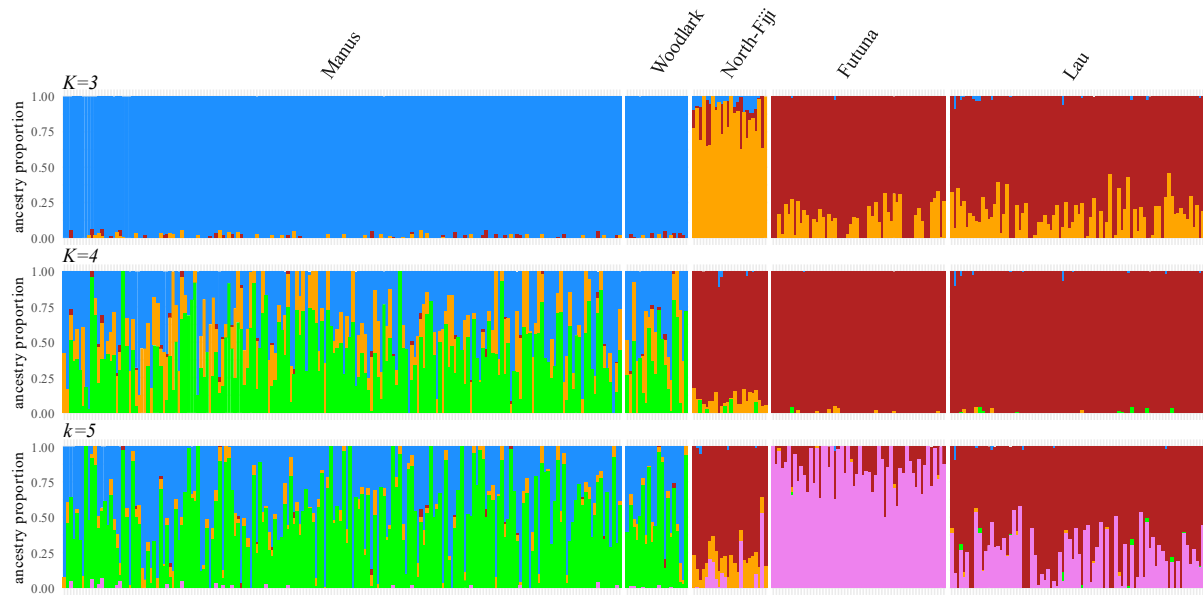
SI Figure 18: PCA with outlier found both in Arlequin and Baypass at 0.01 p -value threshold. (A) Outlier loci detected in Manus/Woodlark. (B) Outlier loci detected in North-Fiji/Futuna/Lau



SI Figure 19: Cross-validation error plot from 10 independent run of Admixture for each and each outliers dataset at 0.05% A) Manus/Woodlark outliers dataset at 0.05%. B) North-Fiji/Futuna/Lau outliers dataset at 0.05%

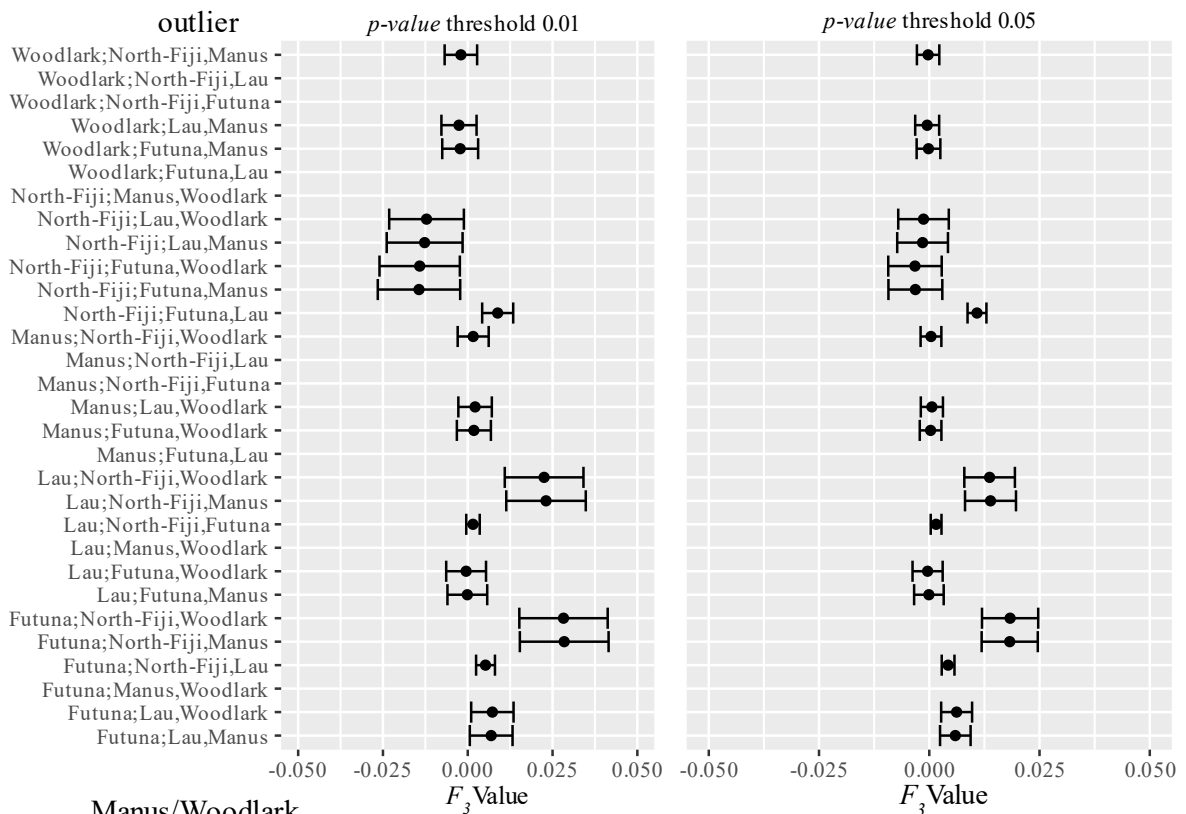


SI Figure 20: Admixture plot for the outliers M/W dataset at 0.05% for $K=3$ to $K=5$.

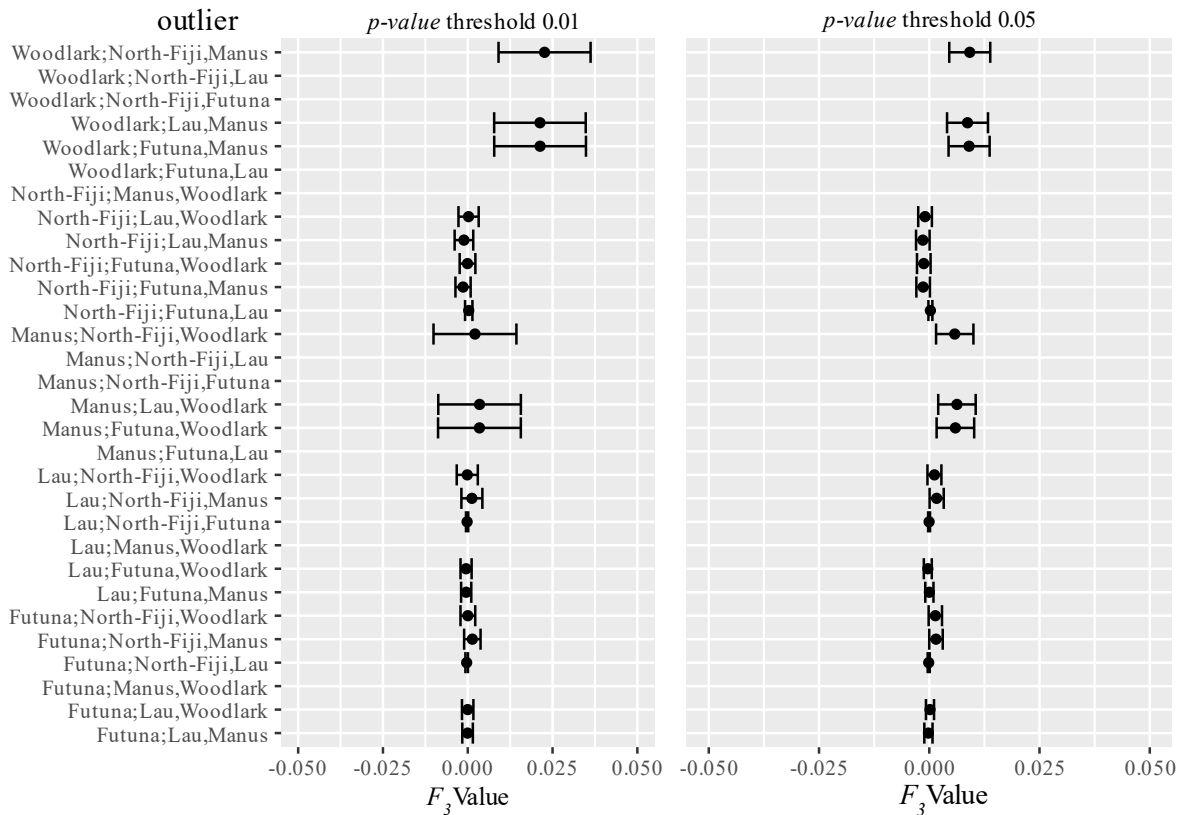


SI Figure 21: Admixture plot for the outliers NF/F/L dataset at 0.05% for K=3 to K=5.

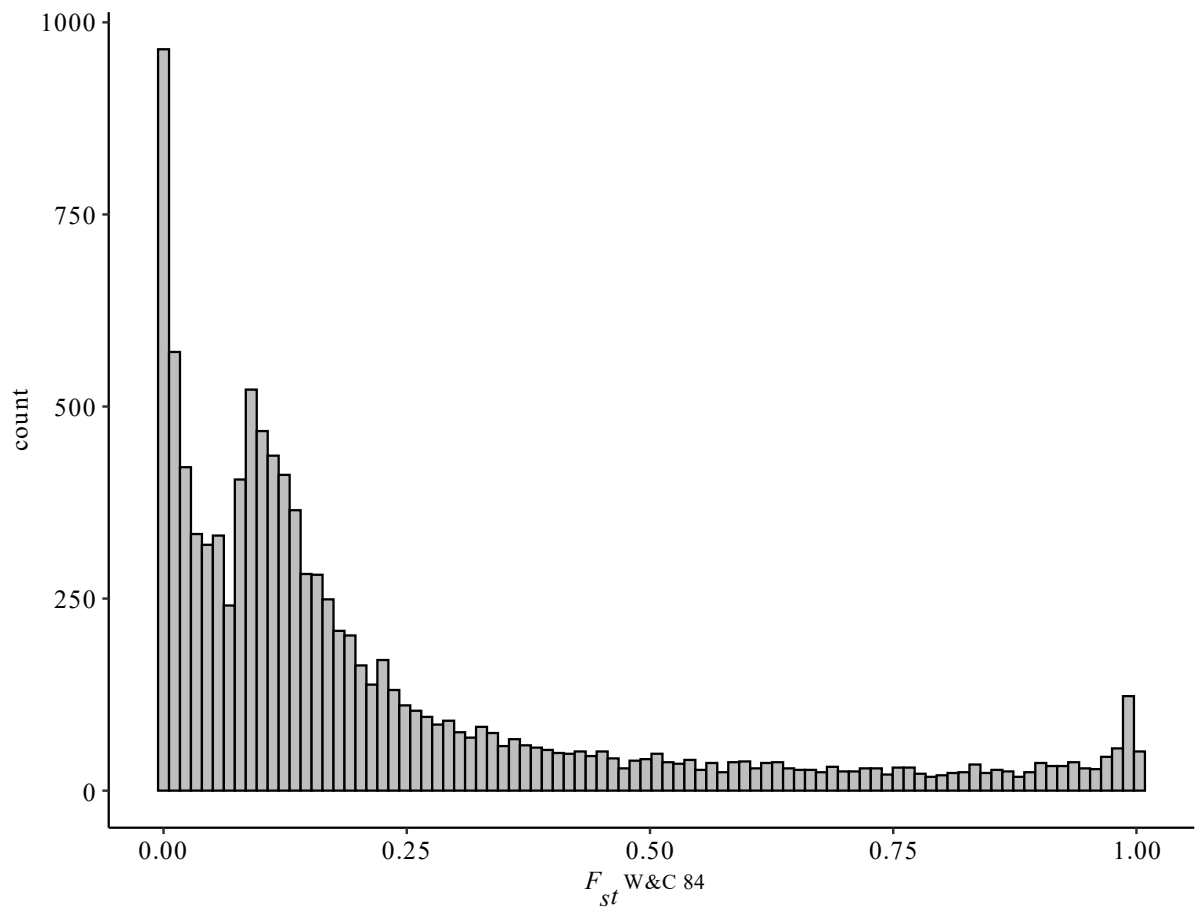
North-Fiji/Futuna/Lau



Manus/Woodlark



SI Figure 22: F_3 statistic with the 95% standard deviation interval for outlier markers near 0. Value that are not display are highly positive. Combination of population is (target population; population source 1, population source 2).



SI Figure 23: Histogram of loci F_{st} (W&C 1984) between Manus/Woodlark and North-Fiji/Futuna/Lau for all the 10 570 SNPs.

Settings

1) Arlequin settings:

General settings in Polymorphism Control section:

Allowed missing level per site = 1 (means we used all sites)

Other settings were by default

a) AMOVA settings:

10 000 permutations

1 Standard AMOVA with basins group by metapopulation

1 Standard AMOVA including individuals' levels with localities group by basins

Other settings were by default

b) Pairwise Fst:

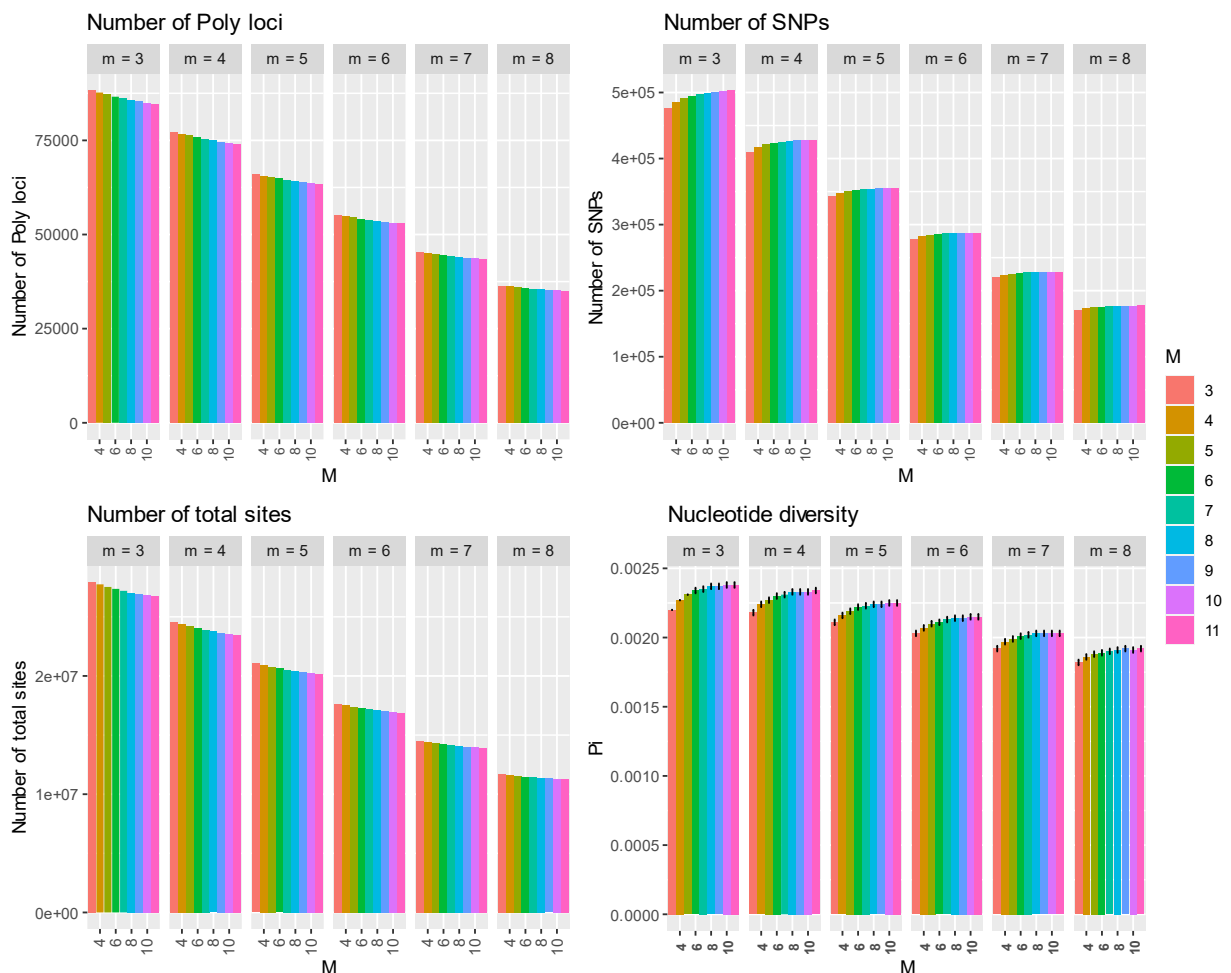
10 000 Permutations

Other settings were by default.

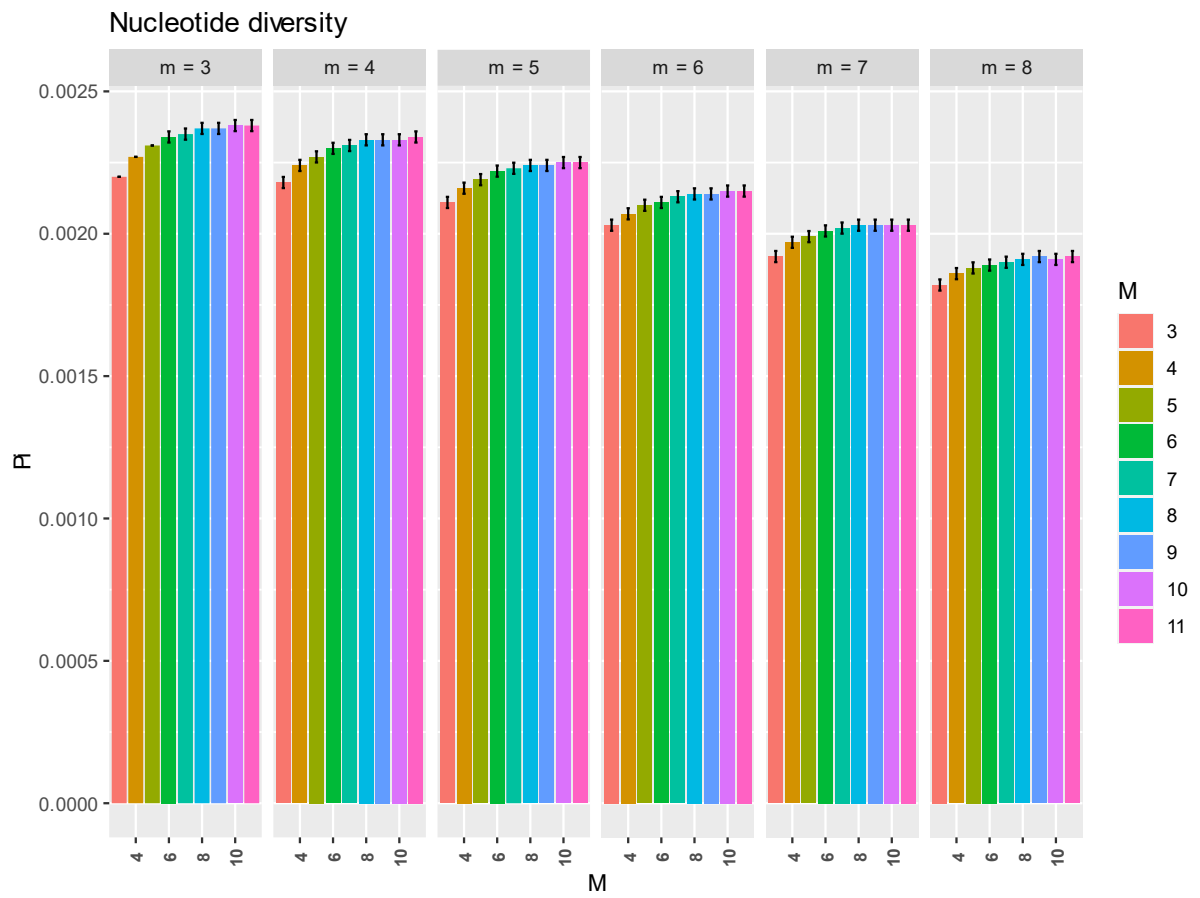
ANNEXE 2

Comme mentionné dans le chapitre 2, nous avons utilisé la même approche que pour le chapitre 1 et *I. nautiliei* pour explorer et calibrer les paramètres d'assemblage *de novo* pour chaque espèce.

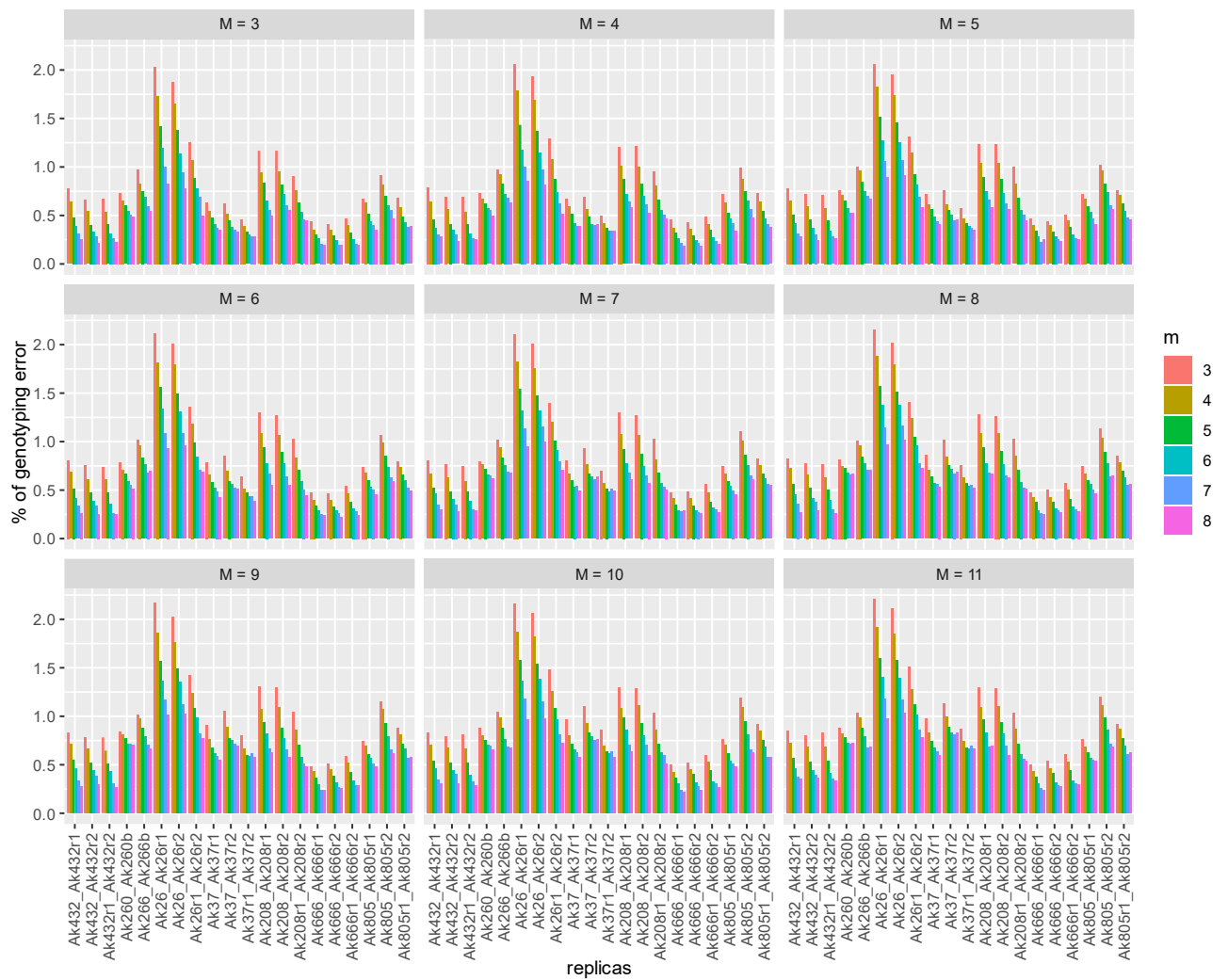
1. Assemblage de novo pour *A. kojimai*



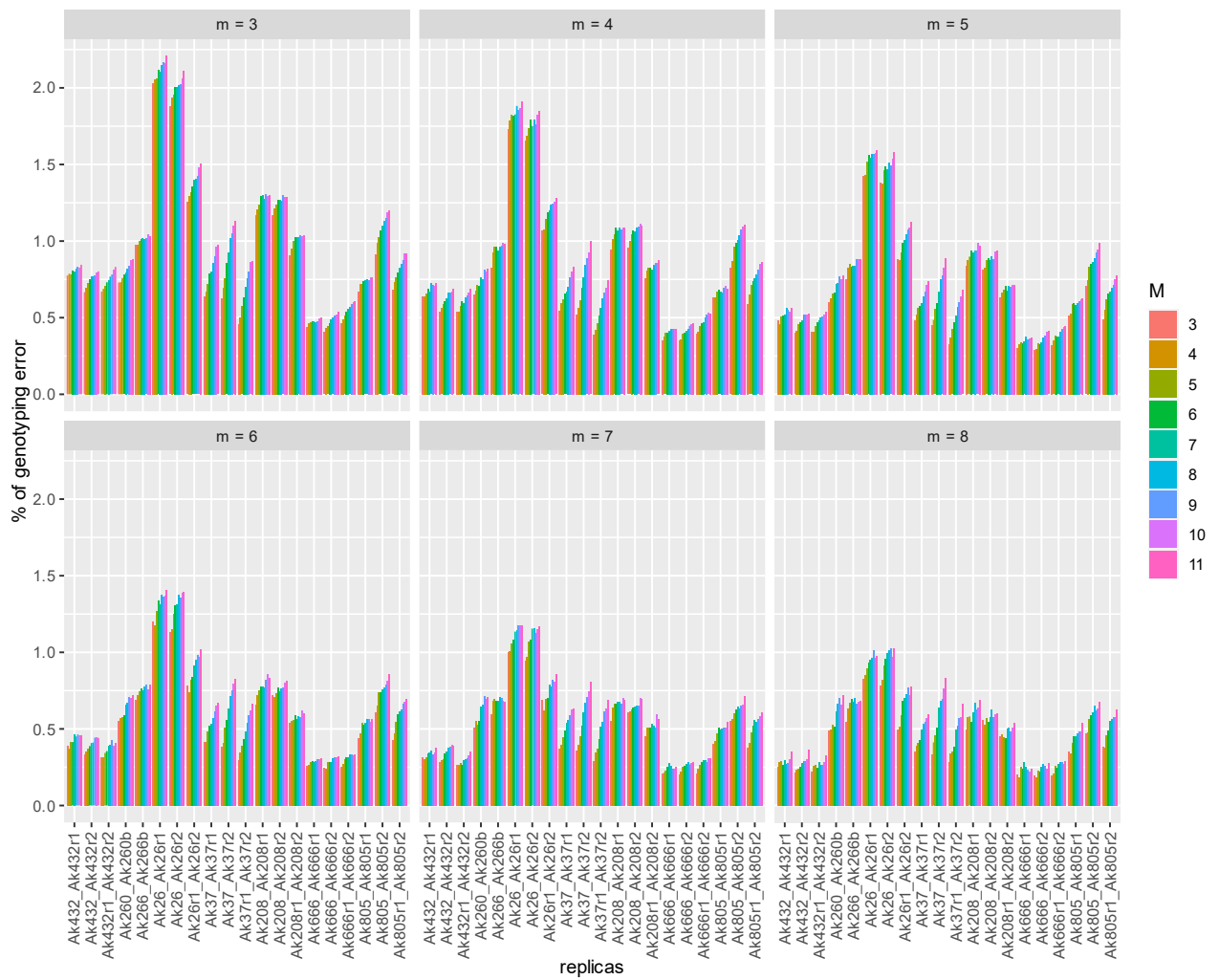
SI Figure 1: Statistiques de calibration (Nombre de locus polymorphes, Variants (SNPs), sites totaux assemblé et diversité nucléotidique (π)) pour chaque valeur de M (chaque box représente une valeur fixée de m) et $n=M$, en considérant tous les individus comme une seule population en incluant les réplicats.



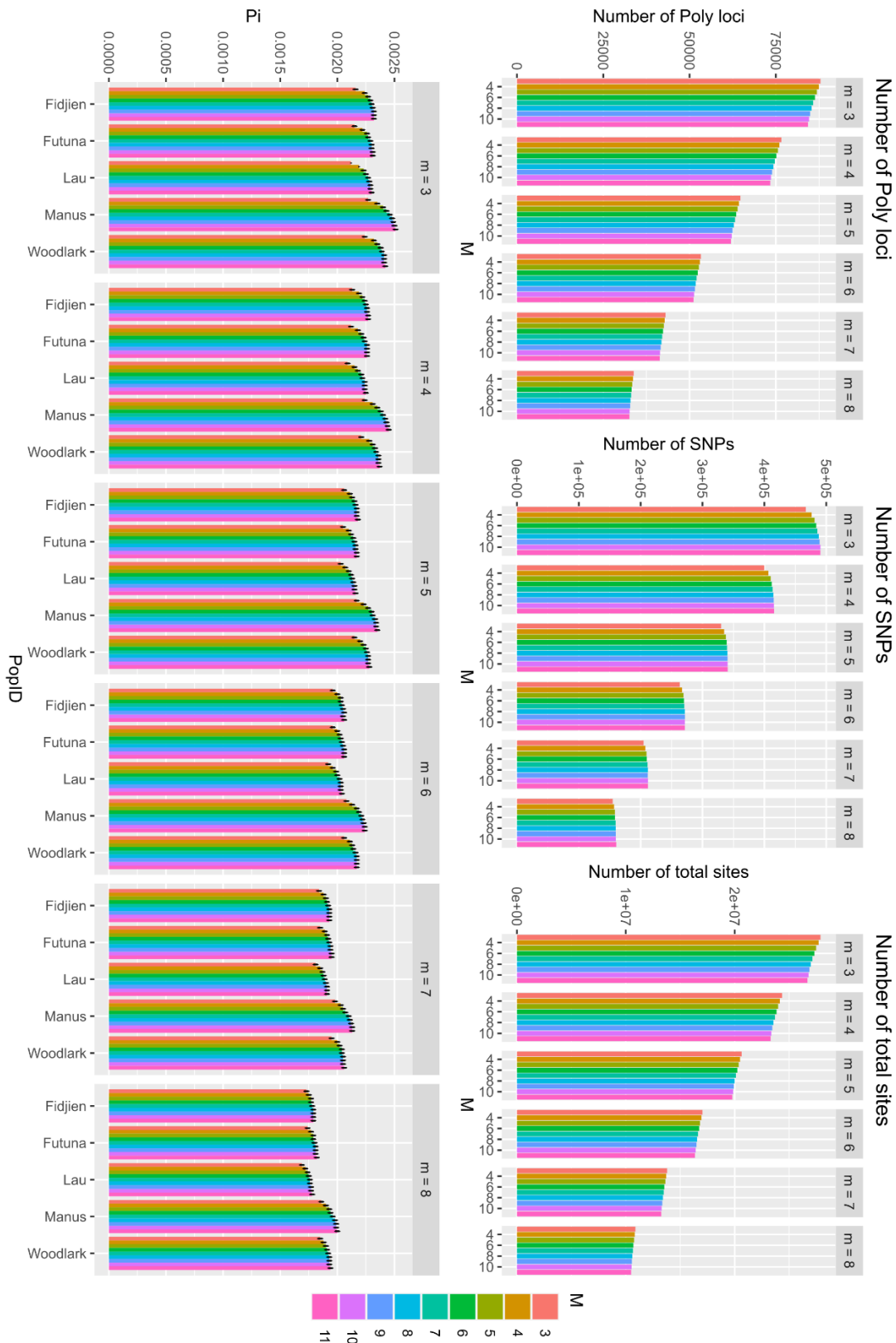
SI Figure 2: Estimation de la diversité nucléotidique (π) pour chaque combinaison des paramètres m et M , en considérant tous les individus et réplicats comme une seule population. (Chaque box représente une valeur fixée de m). Les barres noires représentent les intervalles de confiance à 95%



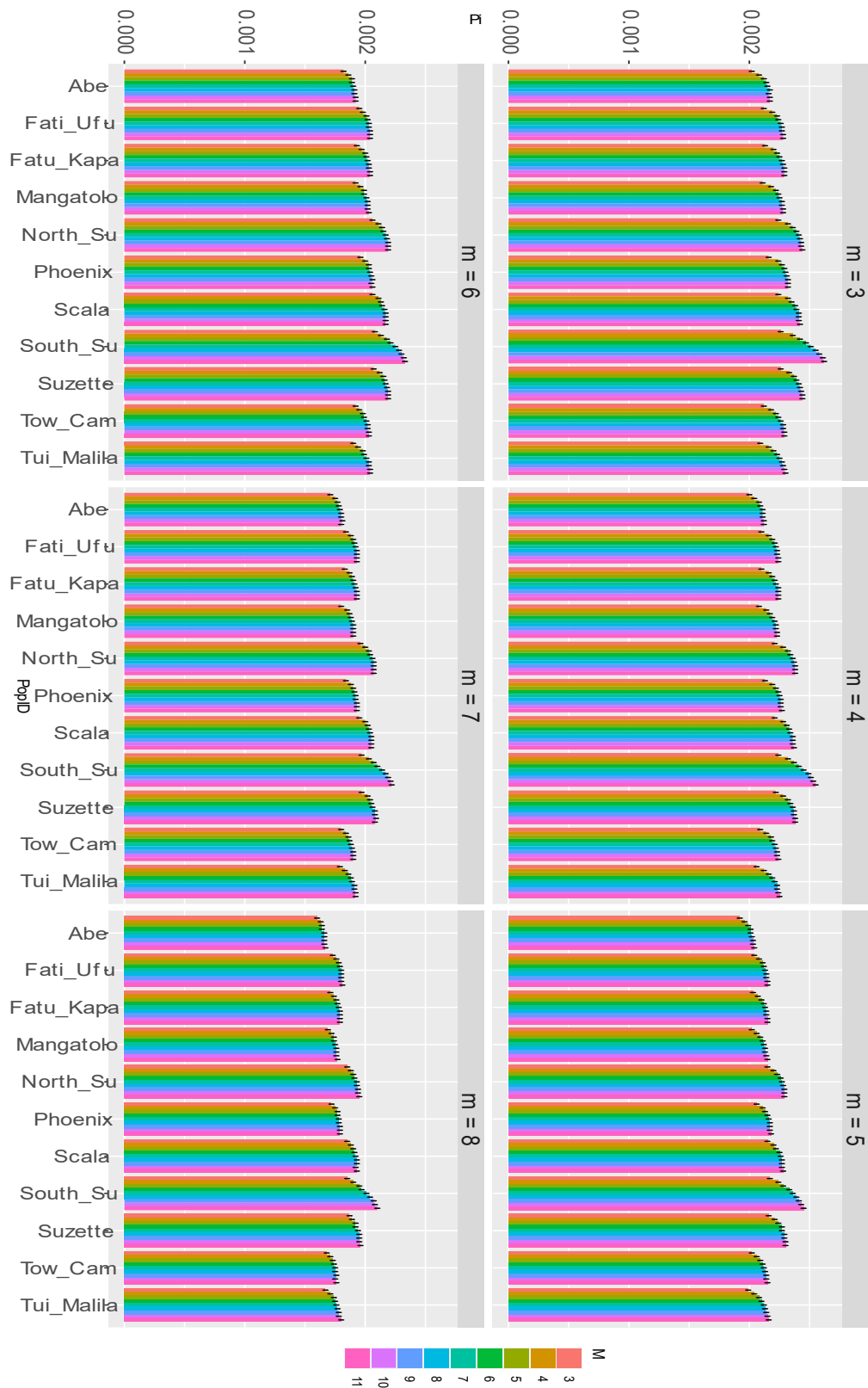
SI Figure 3: Pourcentage d'erreur de génotypage entre paire de réplicats pour chaque combinaison des paramètres m et M ($n=M$), (chaque box représente une valeur fixée de M)



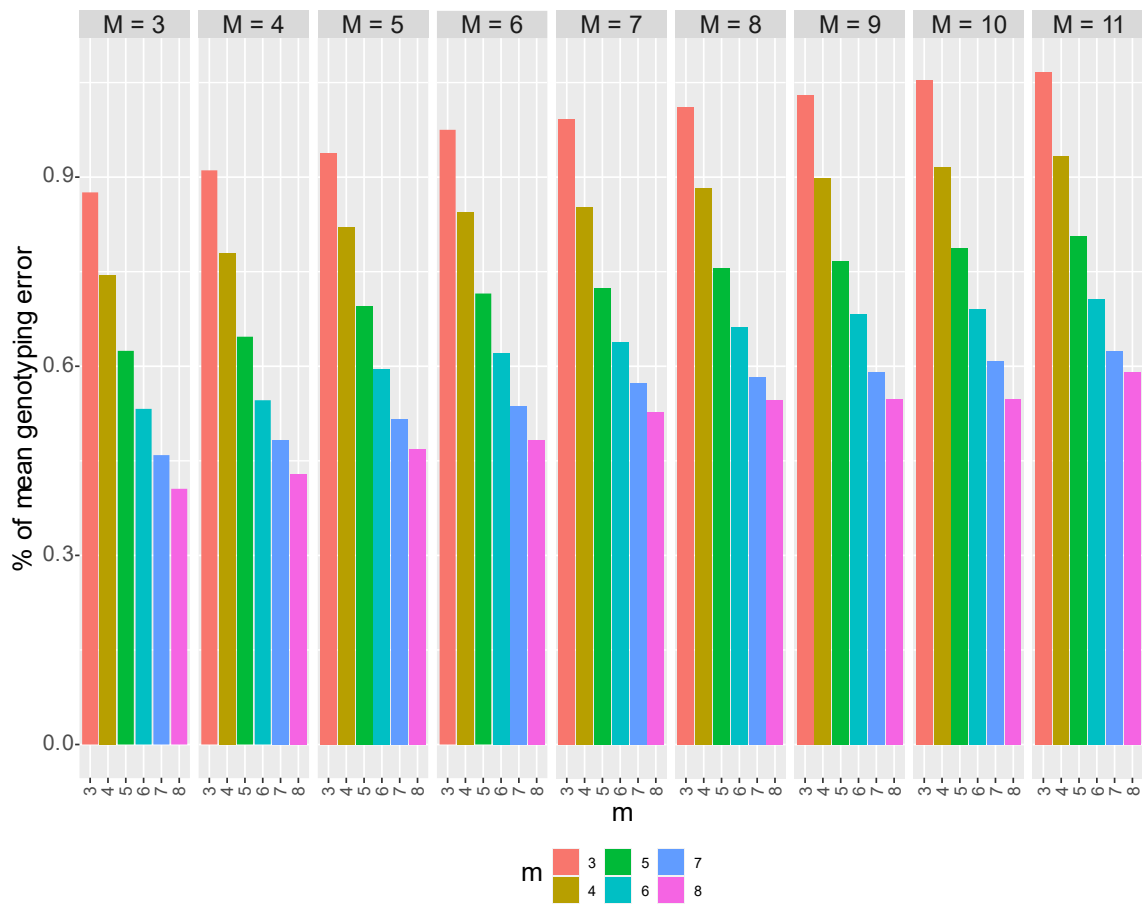
SI Figure 4: Pourcentage d'erreur de génotypage entre paire de réplicats pour chaque combinaison des paramètres m et M ($n=M$), (chaque box représente une valeur fixée de m)



SI Figure 5: Statistiques de calibration (Nombre de locus polymorphes, Variants (SNPs), sites totaux assemblés et diversité nucléotidique (π)) pour chaque valeur de M (Chaque boîte représente une valeur fixée de m) et $n=M$, en considérant les individus par bassin échantonné et en excluant les répliqués. Les barres noires représentent les intervalles de confiance à 95%



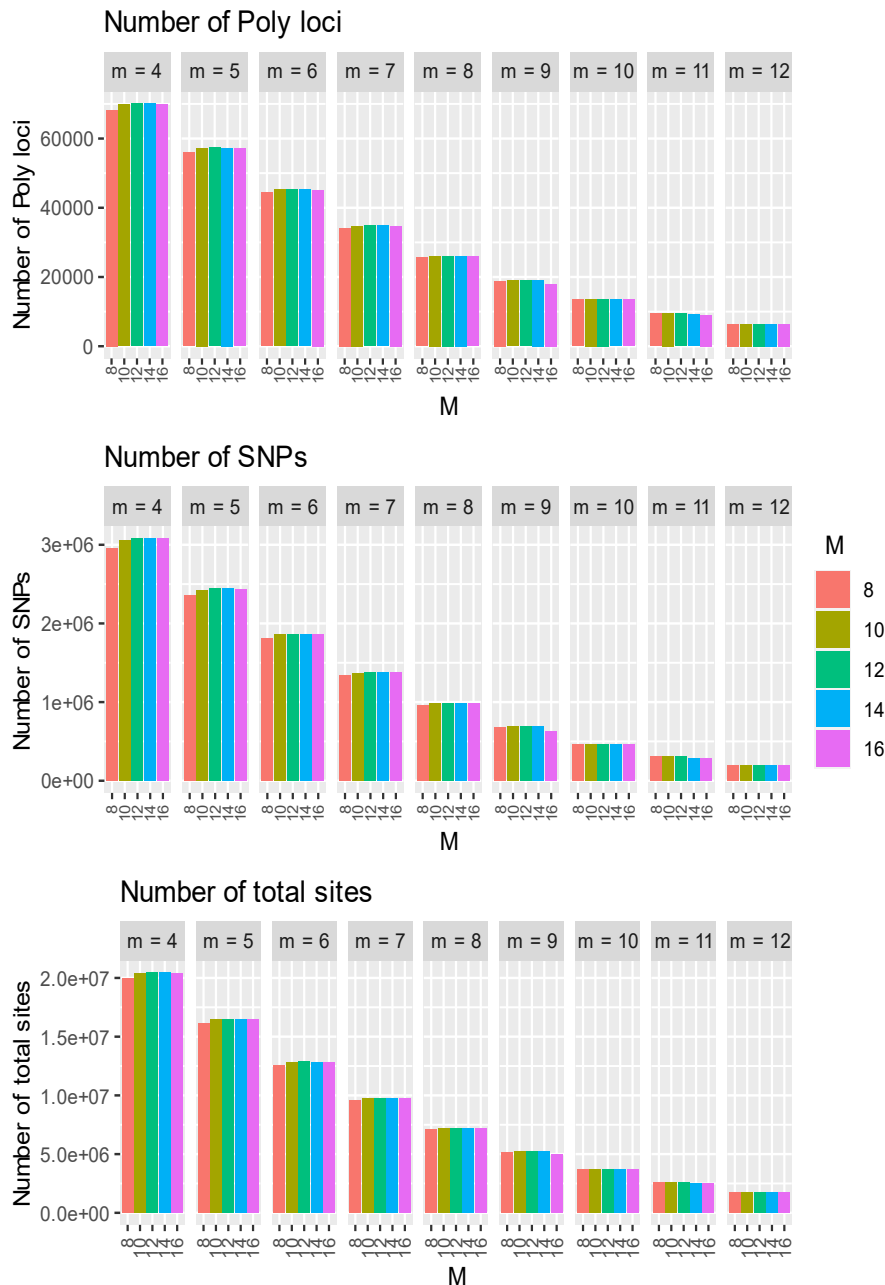
SI Figure 6: Estimation de la diversité nucléotidique (π) pour chaque combinaison des paramètres m et M en considérant les individus par localité échantillonnée et en excluant les réplicats. (Chaque boîte représente une valeur fixée de m). Les barres noires représentent les intervalles de confiance à 95%



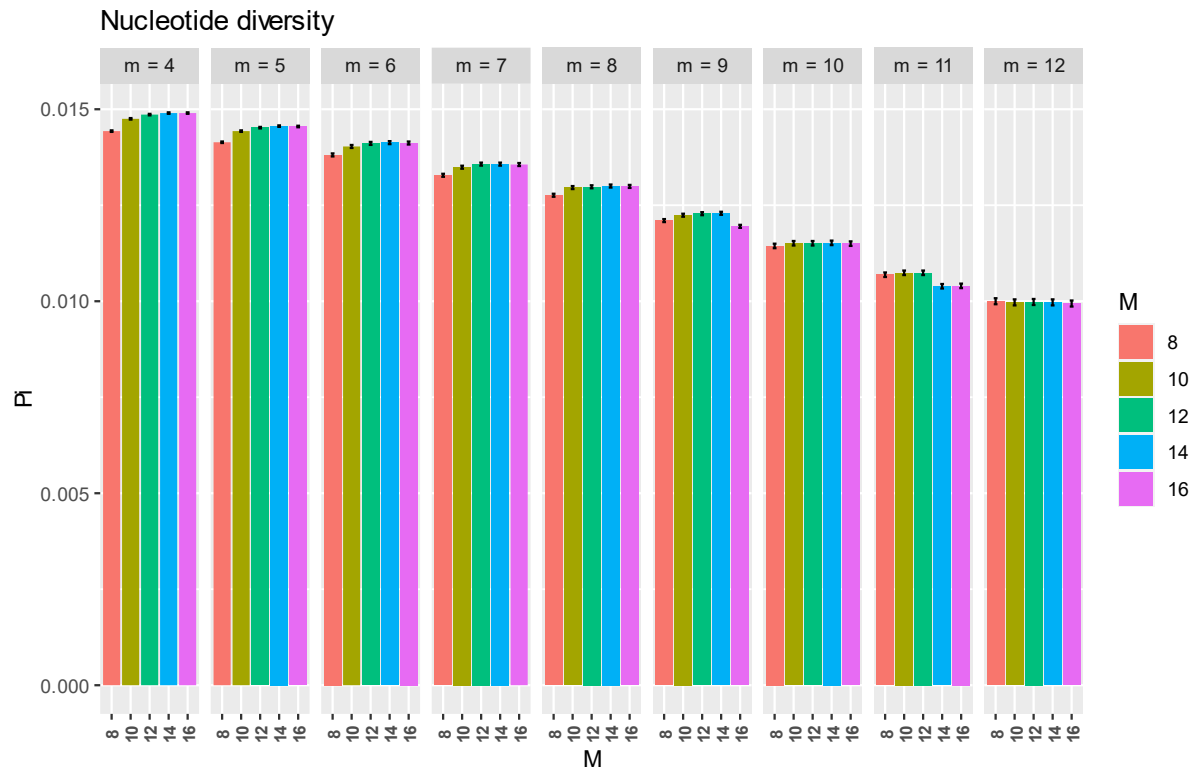
SI Figure 7: Pourcentage d'erreur moyenne de génotypage estimé sur l'ensemble de toutes les paires de réplicats et pour chaque combinaison des paramètres m et M ($n=M$) (Chaque boîte représente une valeur fixée de M)

2. Assemblage de novo pour *S. tollmanni*

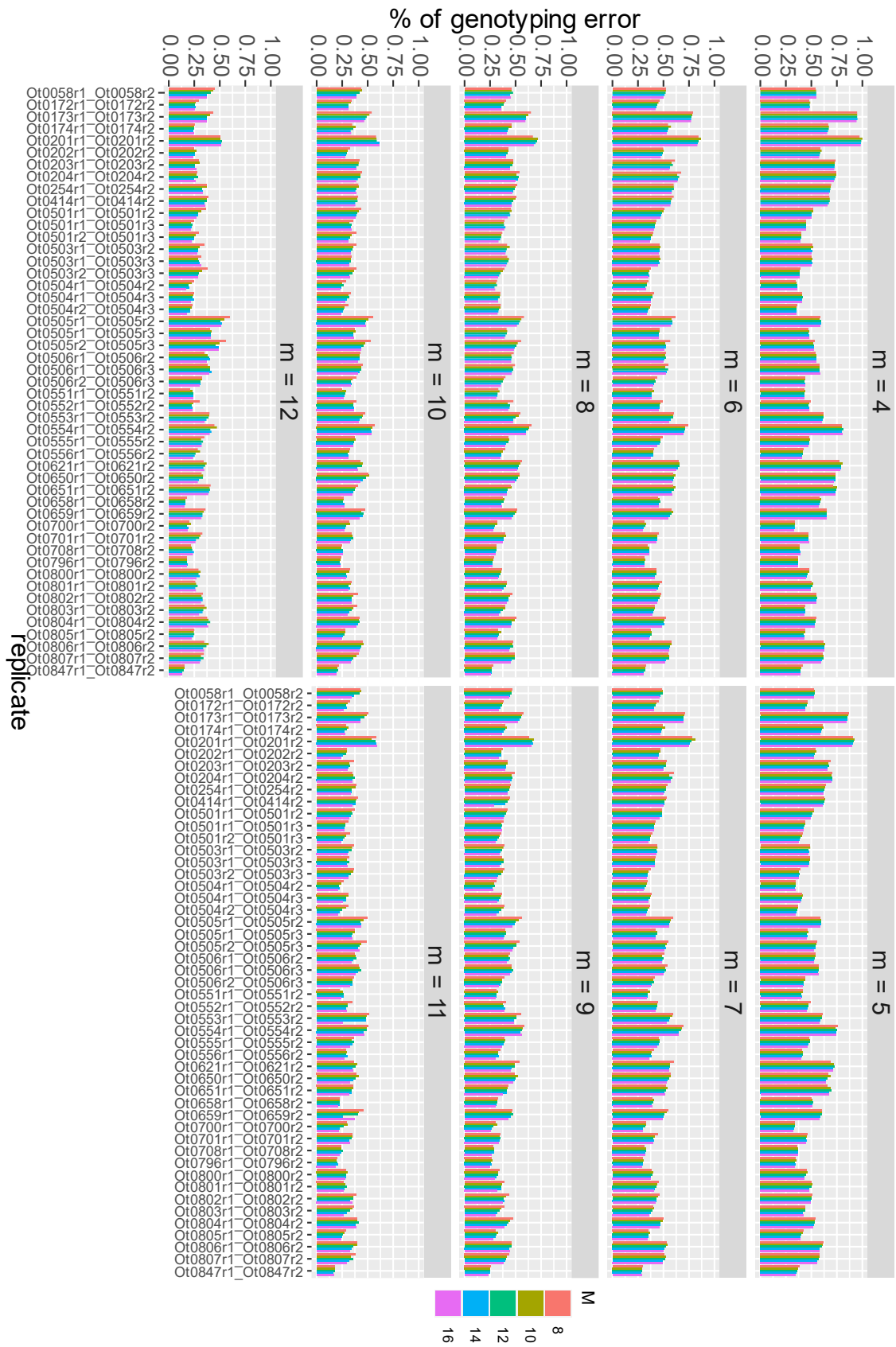
Pour *S. tollmanni*, nous avons exploré une plage de valeur concernant M et m, car cette espèce affiche un fort polymorphisme et des ajouts important de locus « *putatif* » au catalogue des locus lors de l'assemblage. Néanmoins, nous avons exploré les paramètres de la même manière que pour les autres espèces.



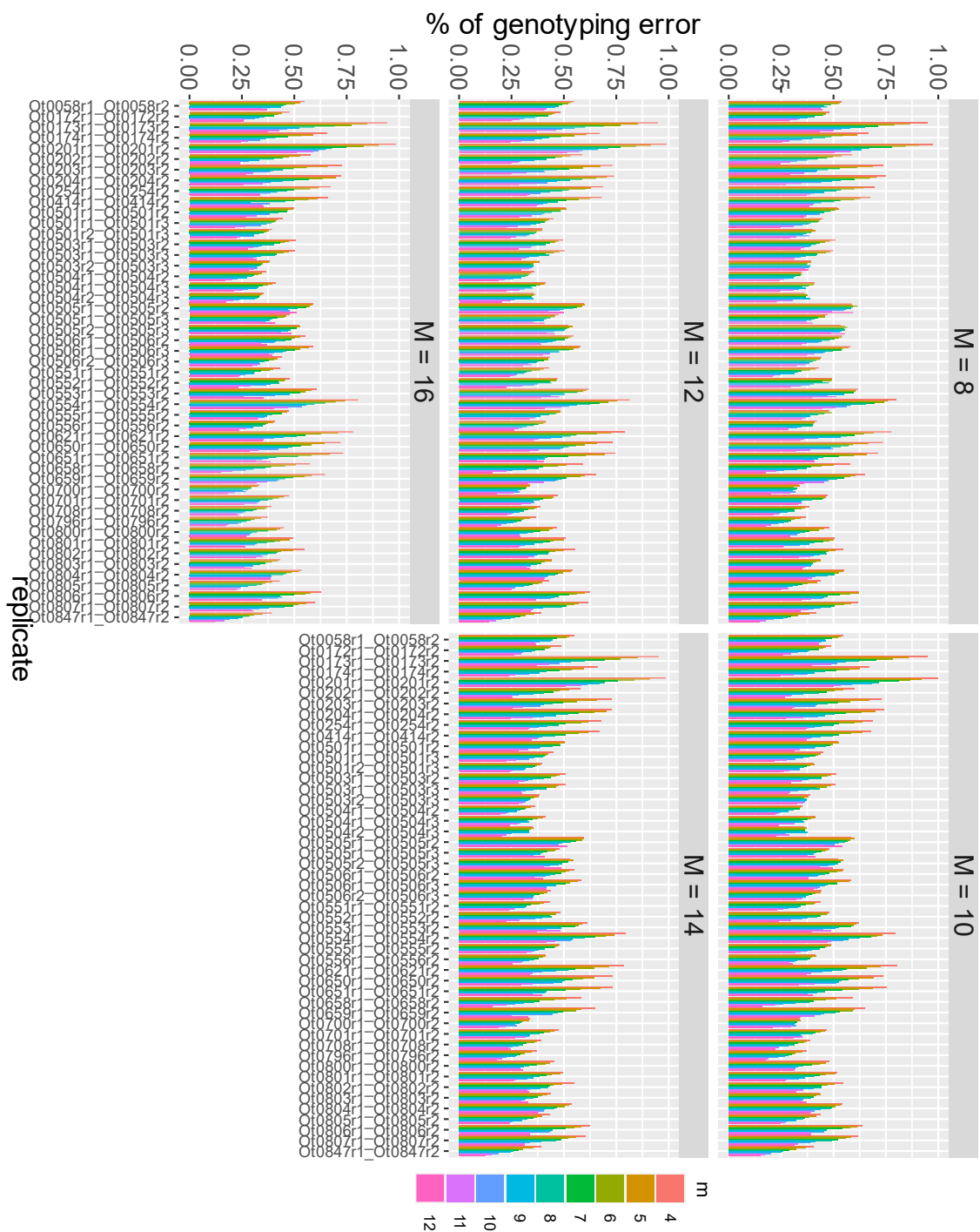
SI Figure 8: Statistiques de calibration (Nombre de locus polymorphes, Variants (SNPs) et sites totaux assemblés) pour chaque valeur de M (Chaque boîte représente une valeur fixée de m) et n=M, en considérant tous les individus comme une seule population et incluant les répliquats.



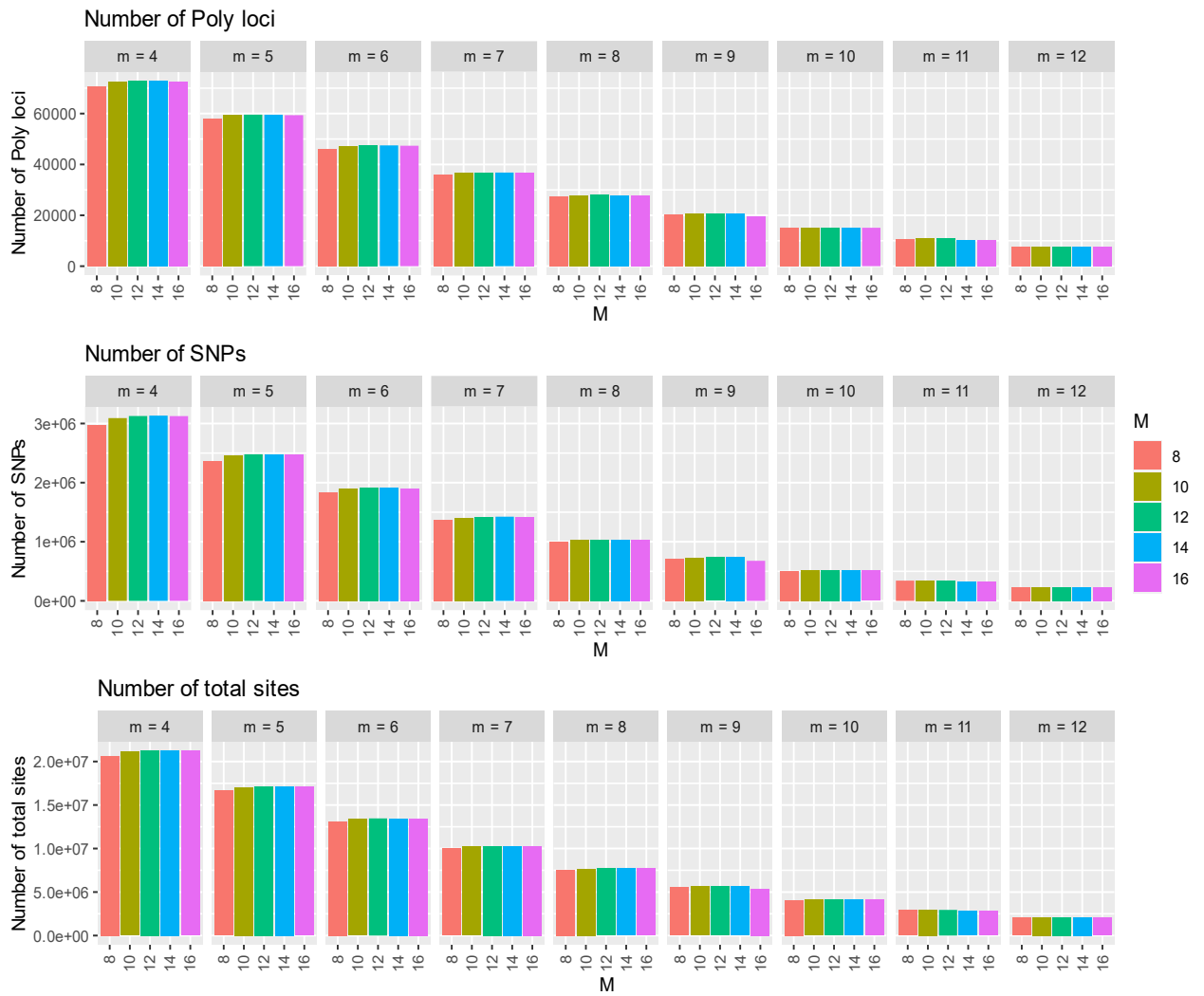
SI Figure 9: Estimation de la diversité nucléotidique (π) pour chaque combinaison des paramètres m et M en considérant tous les individus comme une seule population et en incluant les réplicats. (Chaque boîte représente une valeur fixée de m) Les barres noires représentent les intervalles de confiance à 95% .



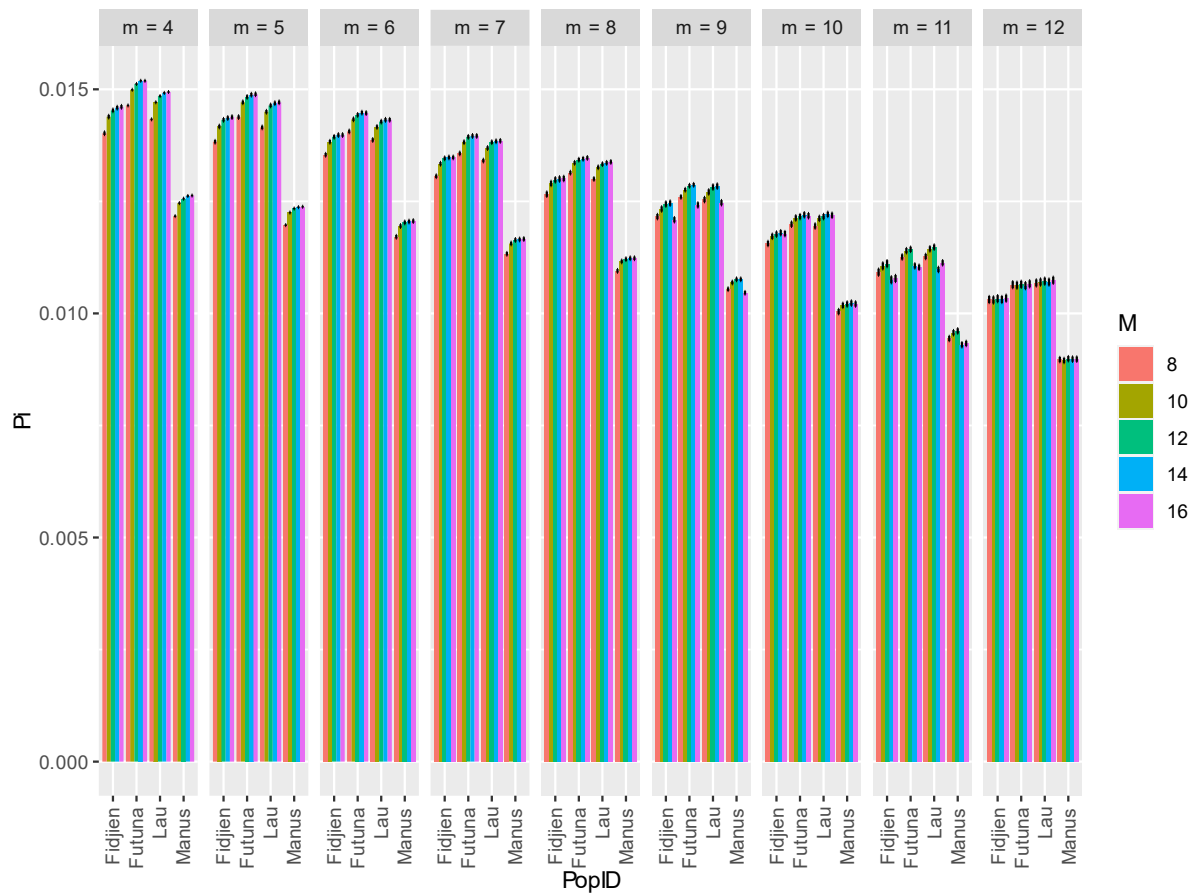
SI Figure 10: Pourcentage d'erreur moyenne de génotypage estimé sur l'ensemble des paires de réplicats et pour chaque combinaison des paramètres m et M ($n=M$) (Chaque boîte représente une valeur fixée de m)



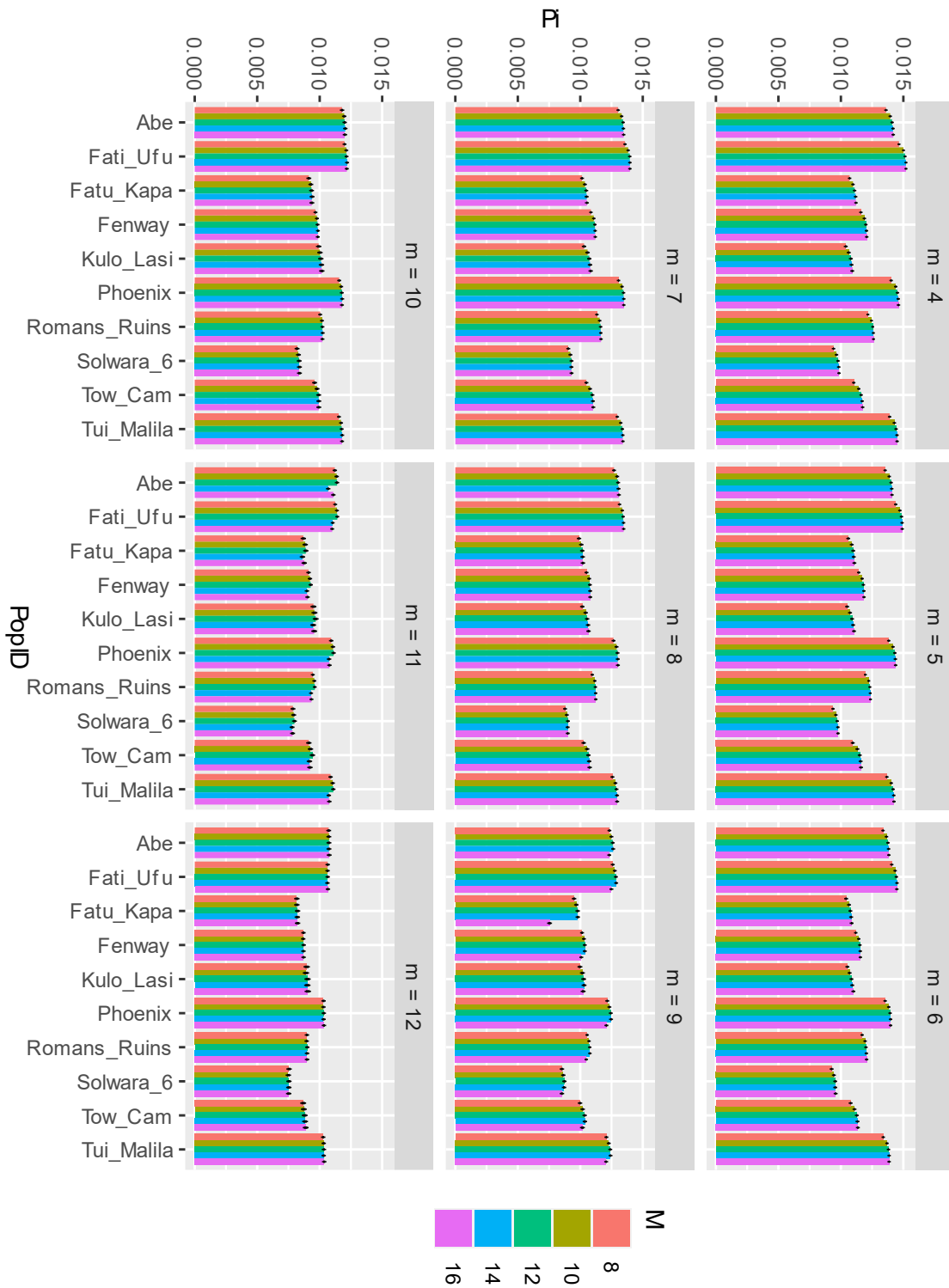
SI Figure 11: Pourcentage d'erreur moyenne de génotypage estimé sur l'ensemble des paires de réplicats et pour chaque combinaison des paramètres m et M ($n=M$) (Chaque boîte représente une valeur fixé de M)



SI Figure 12: Statistiques de calibration (Nombre de locus polymorphes, Variants (SNPs) et sites totaux assemblé) pour chaque valeur de M (Chaque boîte représente une valeur fixée de m) et $n=M$, en considérant les individus groupé par bassin échantillonné et en excluant les réplcats.

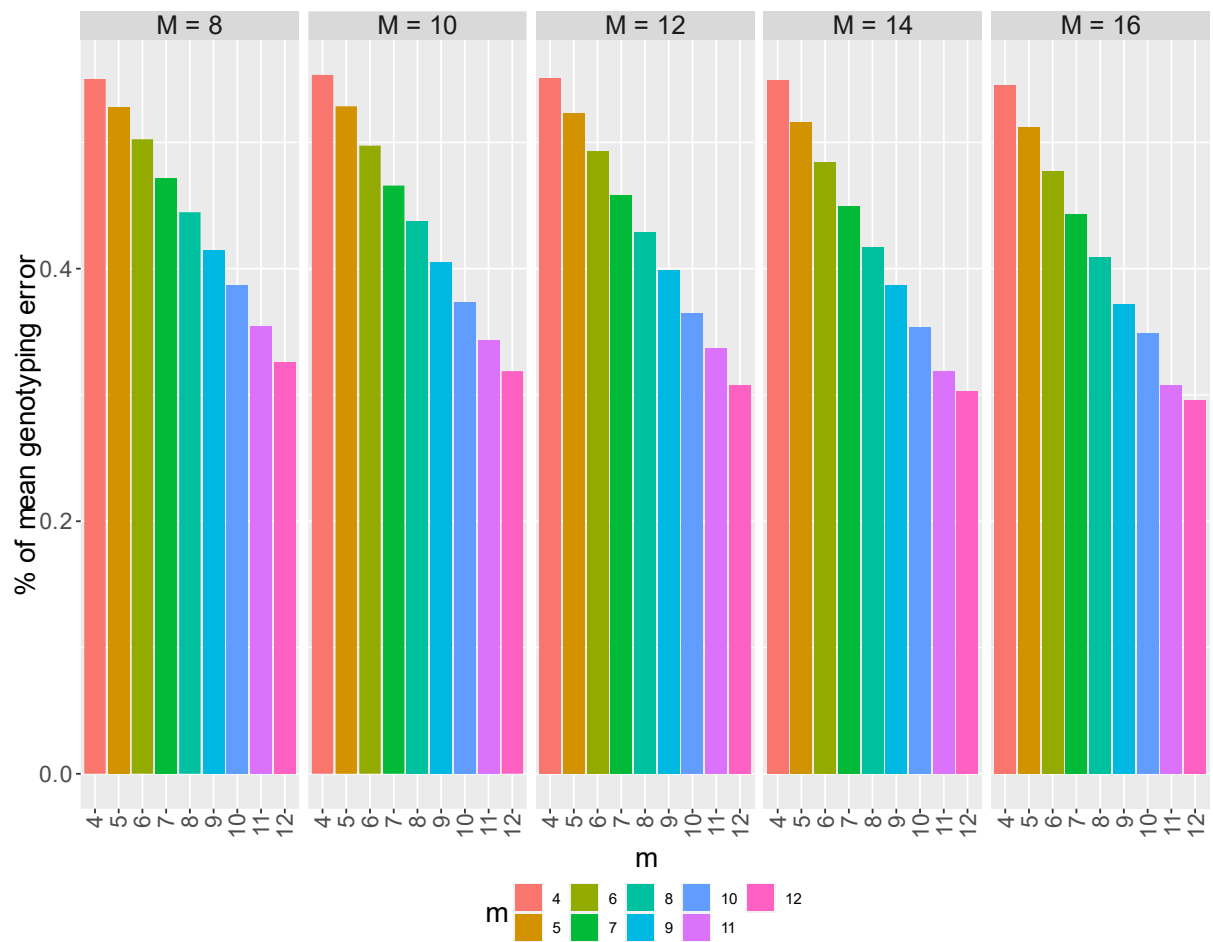


SI Figure 13: Estimation de la diversité nucléotidique (π) pour chaque combinaison des paramètres m et M en considérant les individus par bassin échantillonnée et en excluant les réplicats. (Chaque boîte représente une valeur fixée de m). Les barres noires représentent les intervalles de confiance à 95%



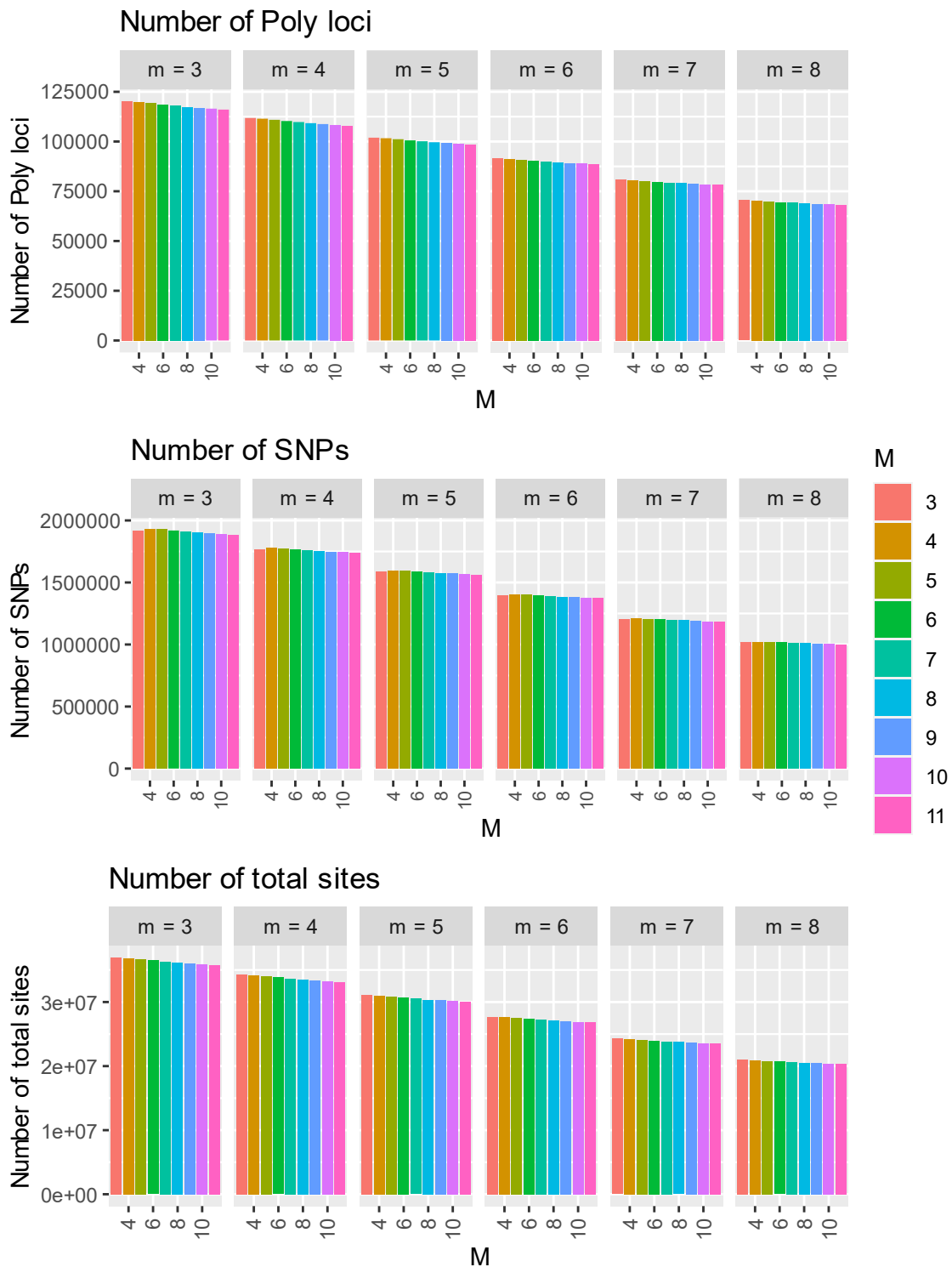
SI

Figure 14: Estimation de la diversité nucléotidique (π) pour chaque combinaison des paramètres m et M en considérant les individus par localité échantillonnée et en excluant les répliquats. (Chaque boîte représente une valeur fixée de m). Les barres noires représentent les intervalles de confiance à 95%

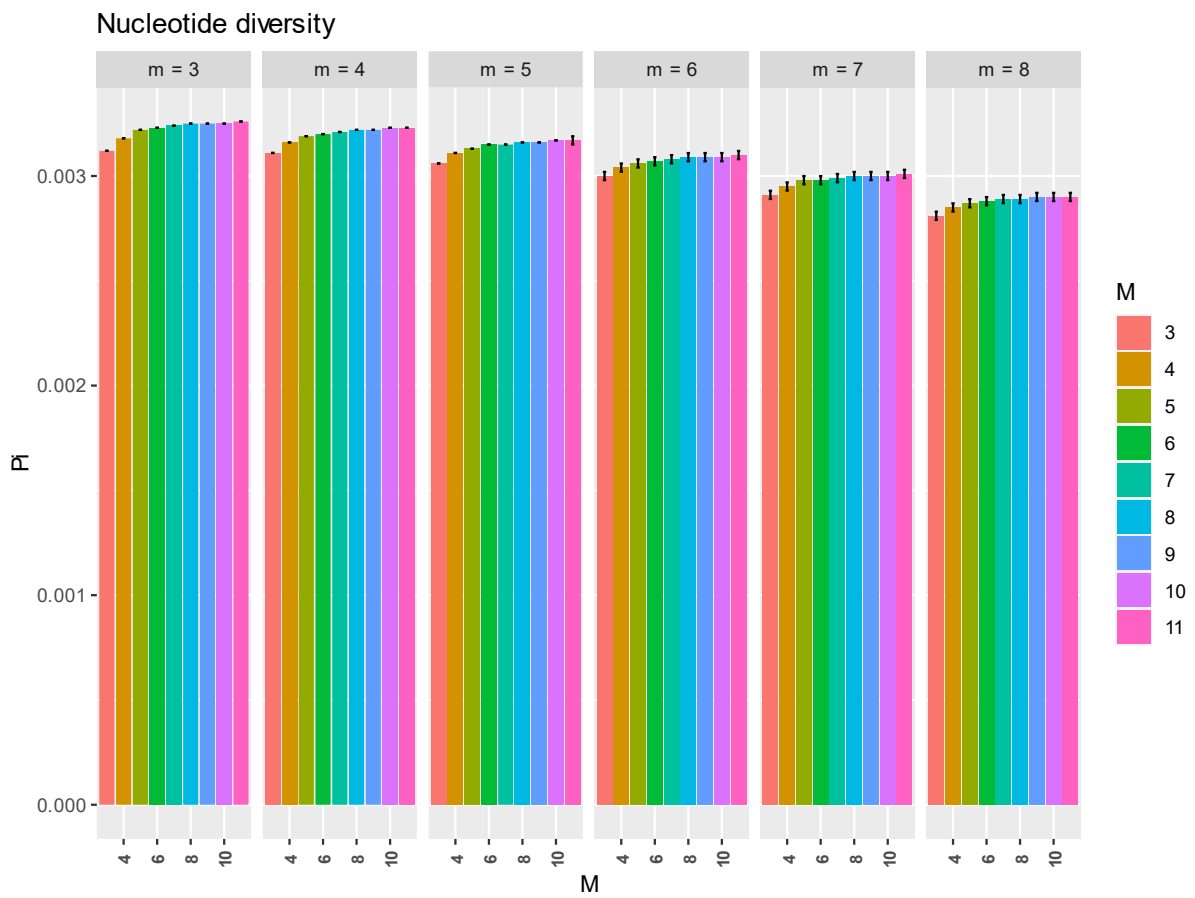


SI Figure 15: Pourcentage d'erreur moyenne de génotypage estimé sur l'ensemble des paires de réplicats et pour chaque combinaison des paramètres m et M ($n=M$) (Chaque boîte représente une valeur fixée de M)

3. Assemblage de novo pour E. ohtai



SI Figure 16 : Statistiques de calibration (Nombre de locus polymorphes, Variants (SNPs) et sites totaux assemblés) pour chaque valeur de M (Chaque boîte représente une valeur fixée de m) et n=M, en considérant tous les individus comme une seule population et incluant les répliquats.



SI Figure 17: Estimation de la diversité nucléotidique (π) pour chaque combinaison des paramètres m et M en considérant tous les individus comme une seule population incluant les réplicats (Chaque boîte représente une valeur fixée de m). Les barres noires représentent les intervalles de confiance à 95%

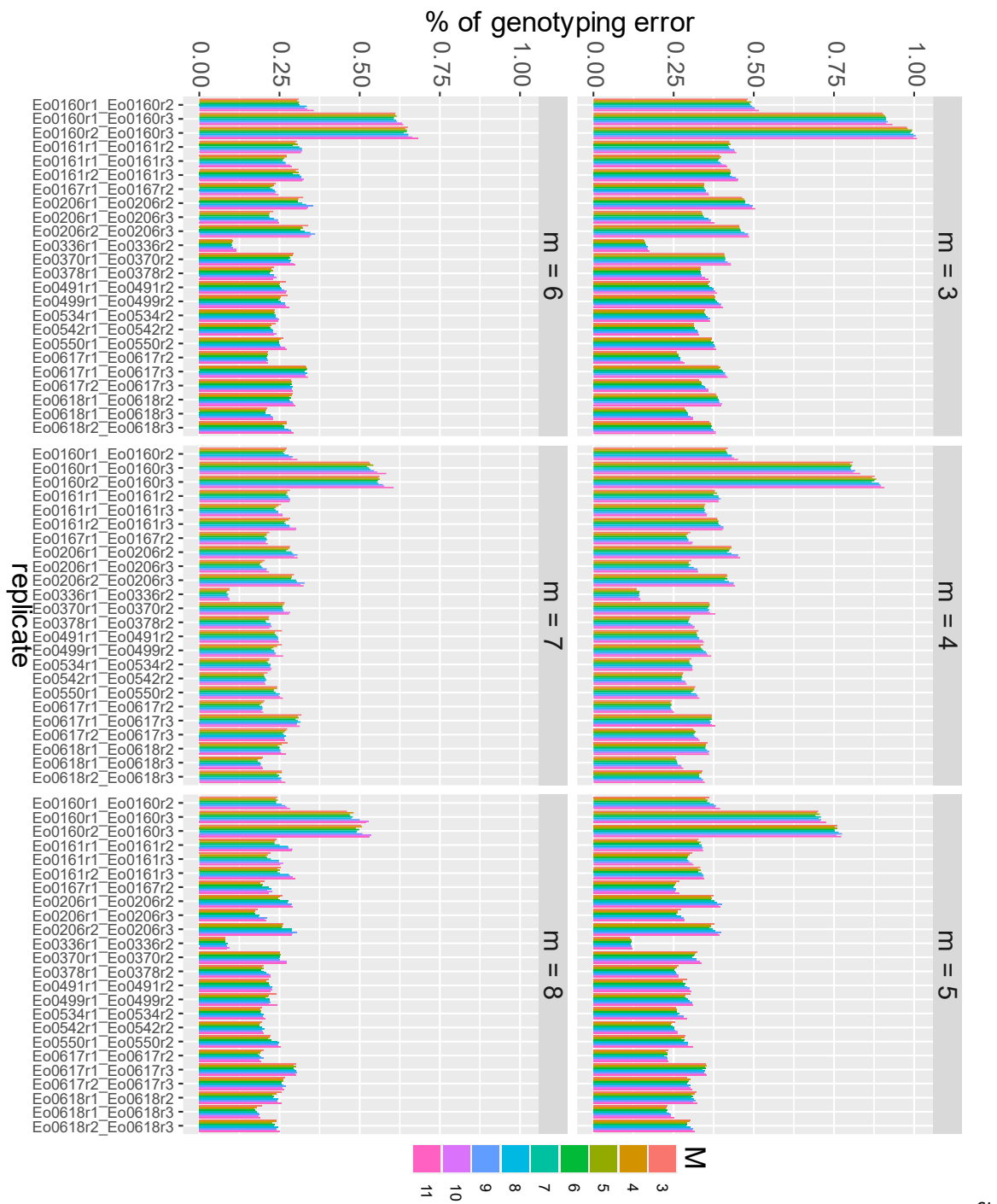


Figure 18 : Pourcentage d'erreur de génotypage entre paire de réplicats pour chaque combinaison des paramètres m et M ($n=M$), (chaque boîte représente une valeur fixée de m)

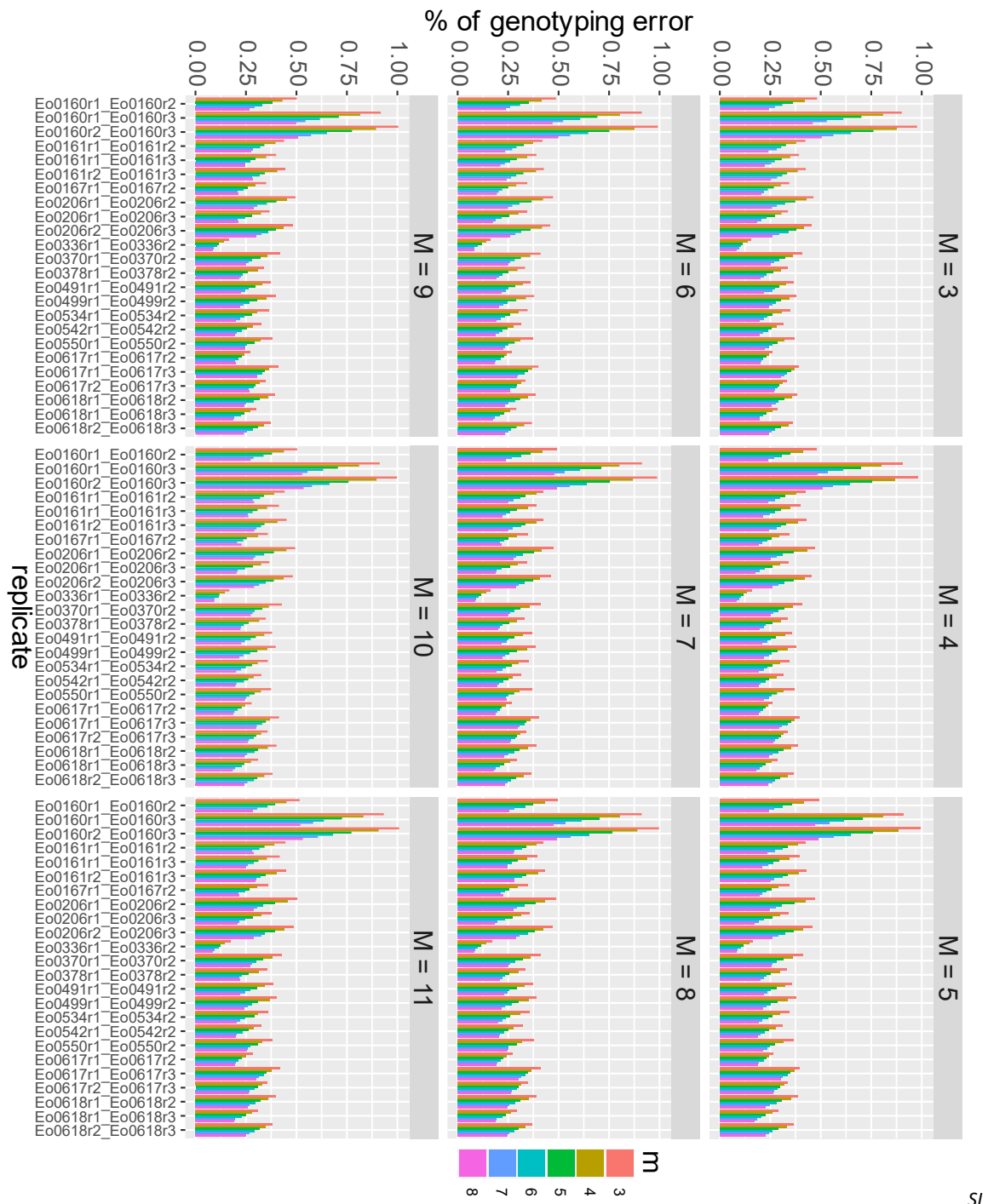
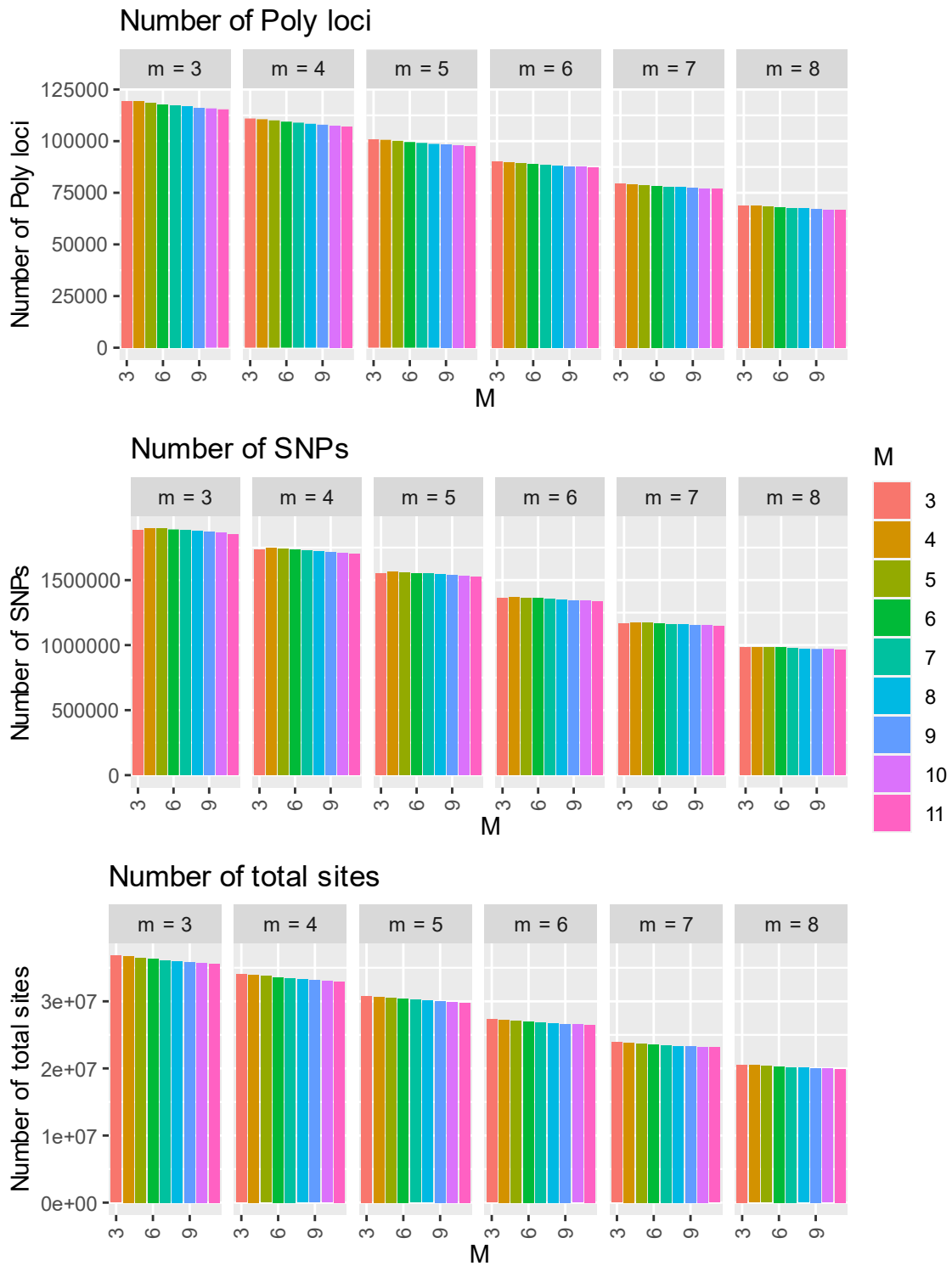
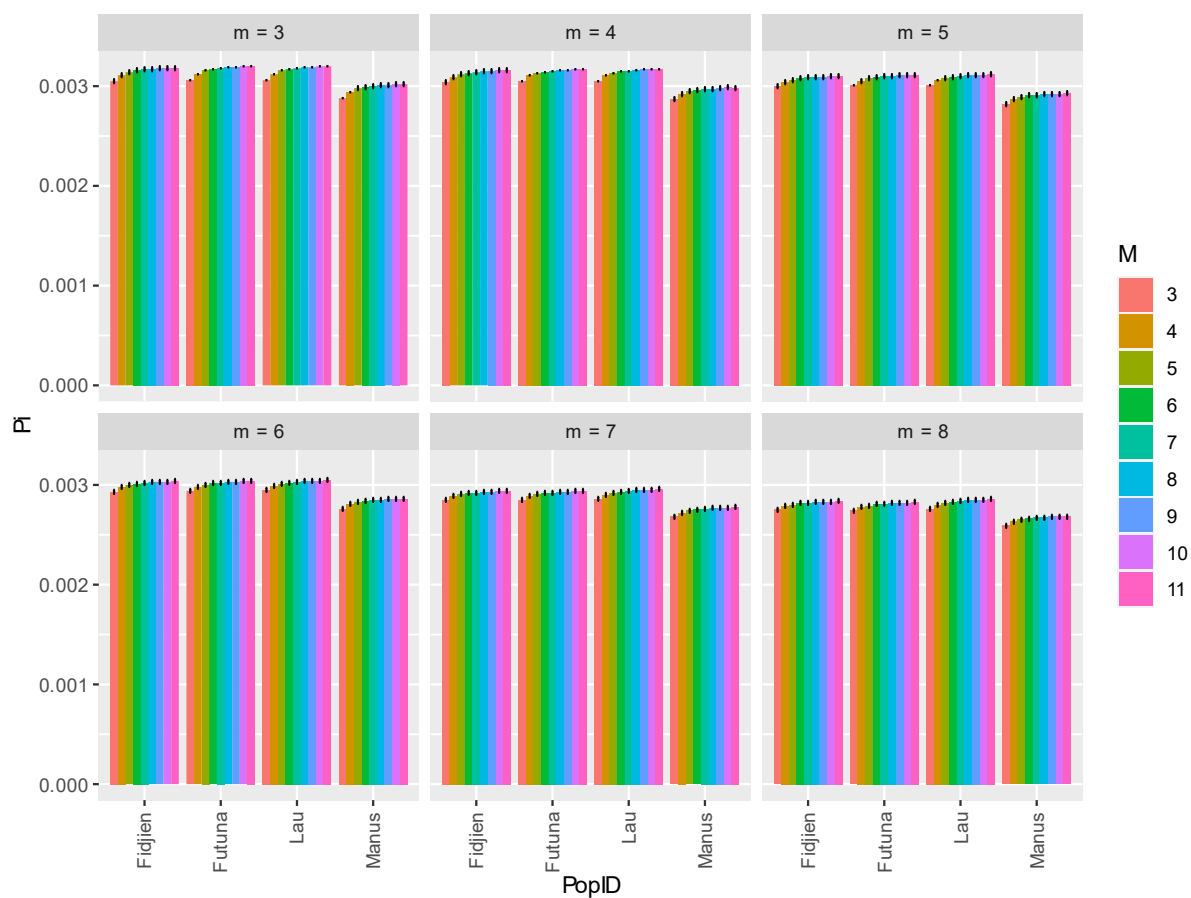


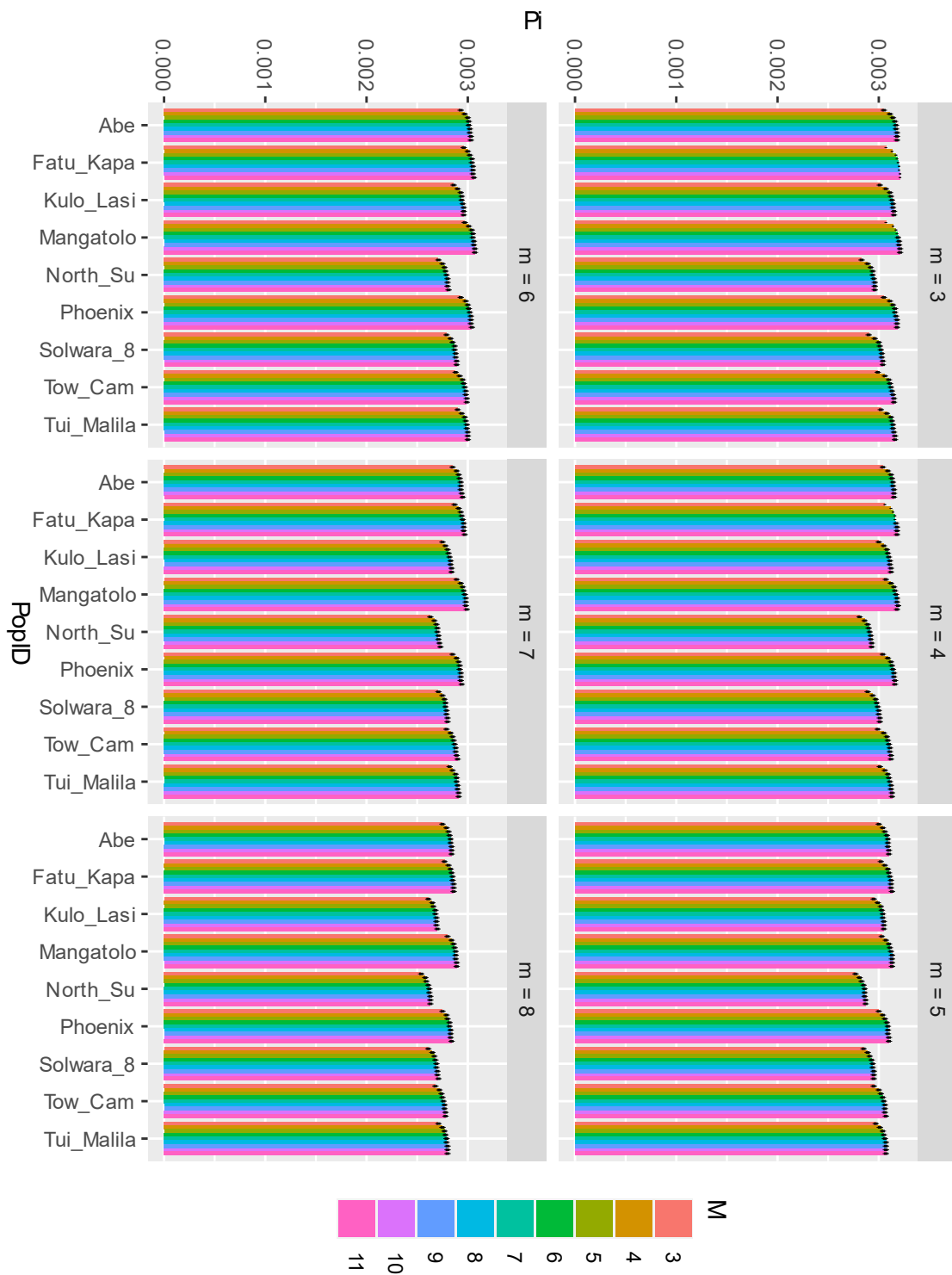
Figure 19 : Pourcentage d'erreur de génotypage entre paire de réplicats pour chaque combinaison des paramètres m et M ($n=M$), (chaque boîte représente une valeur fixée de M)



SI Figure 20: Statistiques de calibration (Nombre de locus polymorphes, Variants (SNPs) et sites totaux assemblés) pour chaque valeur de M (chaque boîte représente une valeur fixée de m) et $n=M$, en considérant les individus par bassin sans les réplicats.

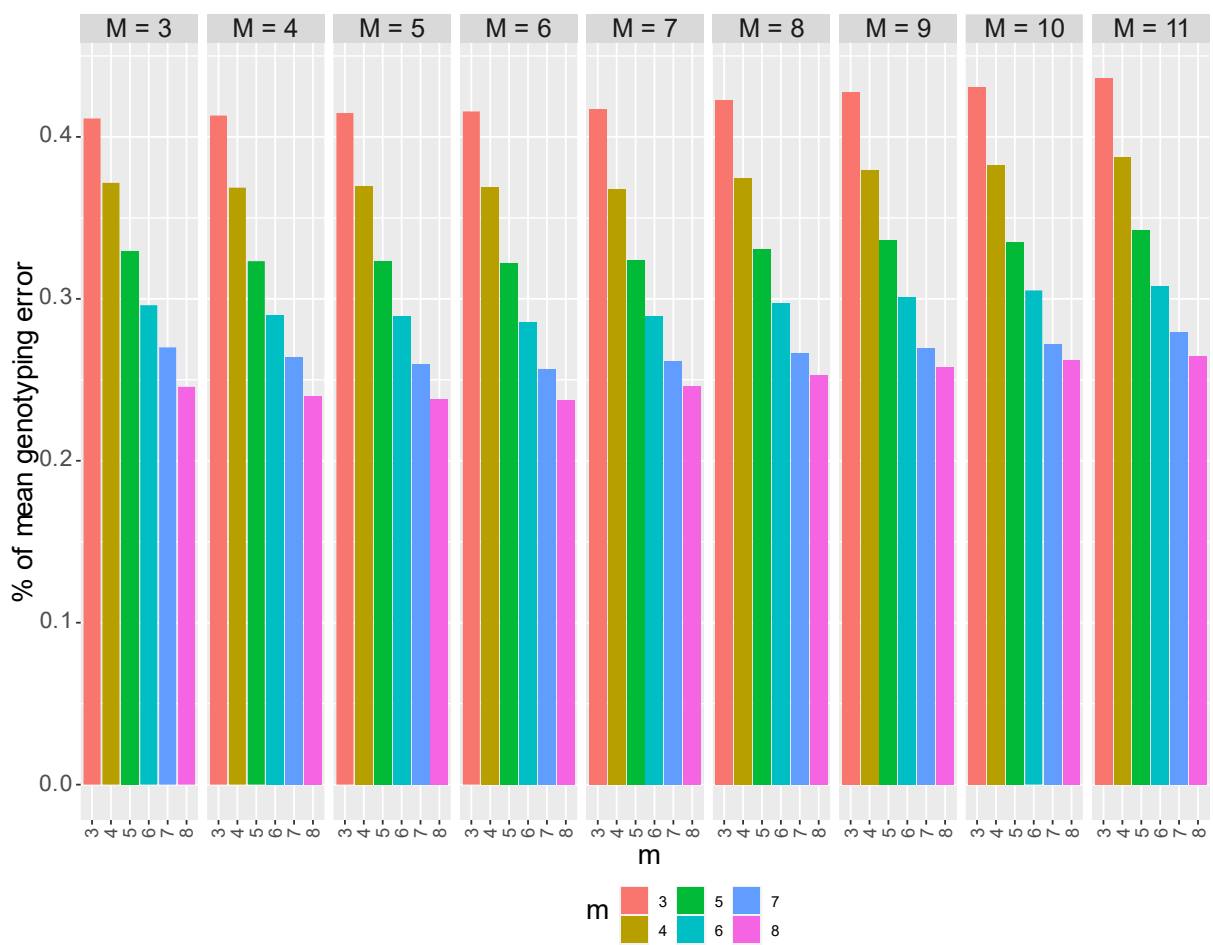


SI Figure 21 : Estimation de la diversité nucléotidique (π) pour chaque combinaison des paramètres m et M en considérant les individus par bassin et en excluant les répliquats. (Chaque boîte représente une valeur fixée de m). Les barres noires représentent les intervalles de confiance à 95%.



SI

Figure 22: Estimation de la diversité nucléotidique (π) pour chaque combinaison des paramètres m et M en considérant les individus par localité et en excluant les réplicats. (Chaque boîte représente une valeur fixée de m). Les barres noires représentent les intervalles de confiance à 95%



SI Figure 23 : Pourcentage d'erreur moyenne de génotypage estimé sur l'ensemble des paires de réplicats et pour chaque combinaison des paramètres m et M ($n=M$) (chaque boîte représente une valeur fixée de M)

4. Analyses

Tableaux

SI Tableau 1: critères de filtre appliqués aux données pour les trois espèces (*A. kojimai*, *S. tollmanni* et *E. ohtai*) avant les analyses

Filtering criterons	<i>A. kojimai</i>			<i>S. tollmanni</i>			<i>E. ohtai</i>		
	N° SNPs	N° radtag	N° individuals	N° SNPs	N° radtag	N° individuals	N° SNPs	N° radtag	N° individuals
Assembled data	286 975	39 931	294	1 142 237	21 424	384	2 144 137	103 097	282
Replicate and ind with more 20% miss data	286 975	39 931	259	1 142 237	21 424	347	2 144 137	103 097	256
SNPs with heterozygosity >0.6							2 143 451	103 077	256
max missing data per SNPs 10%	164 171	24 375	259	174 914	4 920	347	1 407 227	74 794	256
max missing data per Individual 10% (15% <i>S. tollmanni</i>)	164 144	24 363	259	174 914	4 920	347	1 407 227	74 794	240
Max Read Depth (DP) > 80	164 077	24 349	212	174 853	4 907	356	1 407 110	74 773	240
Minor allele frequency > 0.05	32 031	13 728	212	15 413	3 395	347	118 173	48 486	240
Max missing data par Individual 10% (15% <i>S. tollmanni</i>)	32 031	13 728	212	15 413	3 395	347	118 173	48 486	224
Out of Hardy-Weinberg equilibrium p -value < 0.05	28 330	12 957	212	11 059	3 074	347	104 892	45 984	224
1 SNPs per RADtag	12 957	12 957	212	3 074	3 074	347	45 984	45 984	224

SI Tableau 2 : valeurs de Fst en considérant les populations par bassin pour A.kojimai et le jeu de donnée final avec 10 0000 permutations après avoir appliqué une correction de Bonferroni. (***: $p < 0.001$, **: $p < 0.01$, * $p < 0.05$)

	Manus	Woodlark	Nord-Fidji	Lau	Futuna
Manus	0.00000				
Woodlark	-0.00079	0.00000			
Nord-Fidji	0.01651***	0.01668***	0.00000		
Lau	0.01888***	0.01911***	0.00317***	0.00000	
Futuna	0.01868***	0.01881***	0.00324***	-0.00044	0.00000

SI Tableau 3 : valeurs de Fst en considérant les populations par bassin pour S.tollmanni et le jeu de donnée final avec 10 0000 permutations après avoir appliqué une correction de Bonferroni. (***: $p < 0.001$, **: $p < 0.01$, * $p < 0.05$)

	Manus	Woodlark2	Woodlark1	Nord-Fidji	Lau	Futuna
Manus	0.00000					
Woodlark2	0.00695***	0.00000				
Woodlark1	0.26753***	0.21692***	0.00000			
Nord-Fidji	0.27114***	0.23451***	0.00146	0.00000		
Lau	0.27280***	0.23649***	-0.00248	-0.00103	0.00000	
Futuna	0.27256***	0.23492***	-0.00212	-0.00049	-0.00057	0.00000

SI Tableau 4 : valeurs de Fst en considérant les populations par bassin pour E. ohtai et le jeu de donnée final avec 10 0000 permutations après avoir appliqué une correction de Bonferroni. (***: $p < 0.001$, **: $p < 0.01$, * $p < 0.05$)

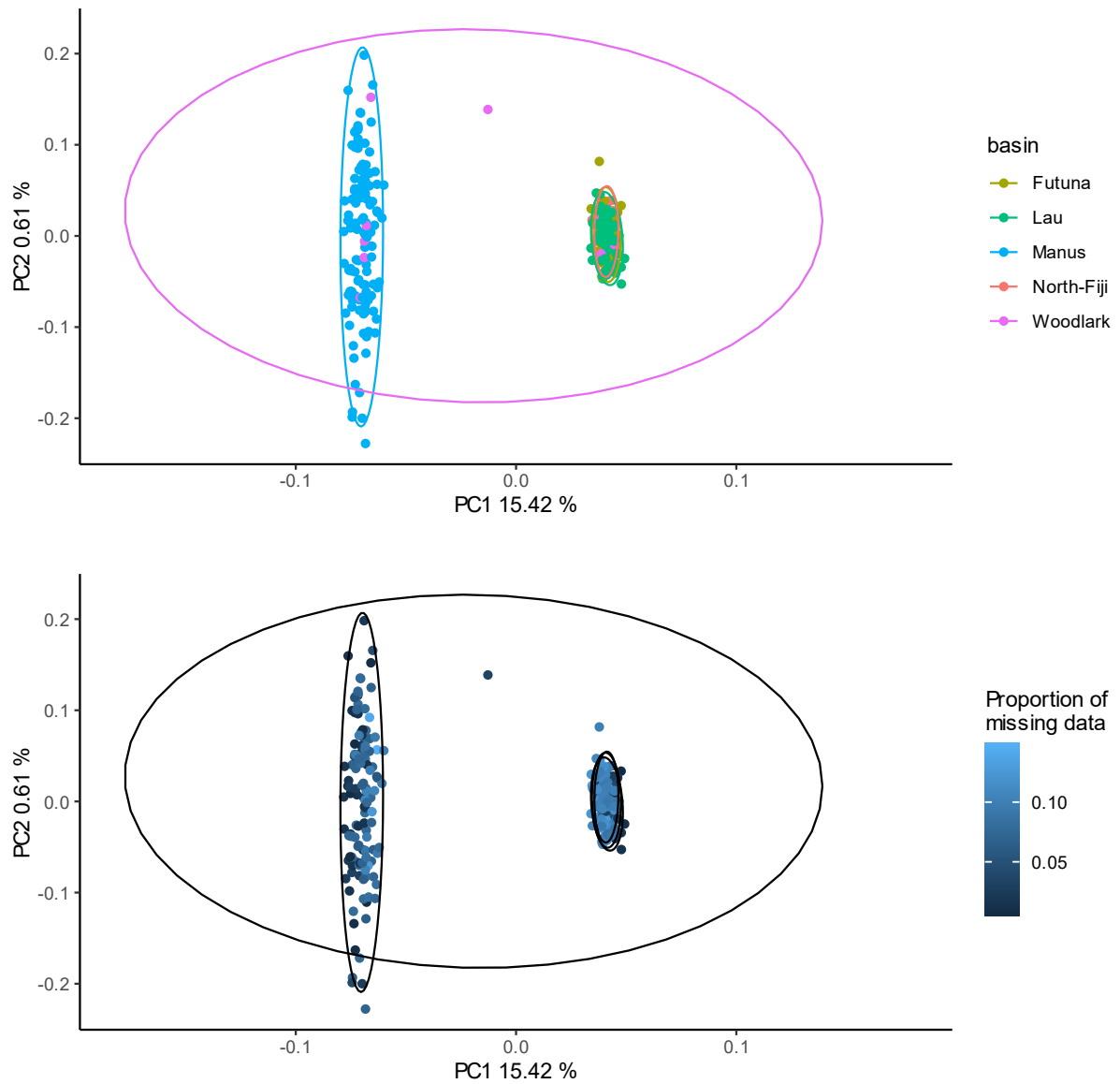
	Manus	Lau	Nord-Fidji	Futuna
Manus	0.00000			
Lau	0.20507***	0.00000		
Nord-Fidji	0.21365***	-0.00008	0.00000	
Futuna	0.20765***	-0.00022	0.00008***	0.00000

SI Tableau 5 : Paramètres dadi pour les meilleurs modèles et par espèces avec leur écart-type (SD) estimé avec la Matrice d'Information de Fisher dans dadi.

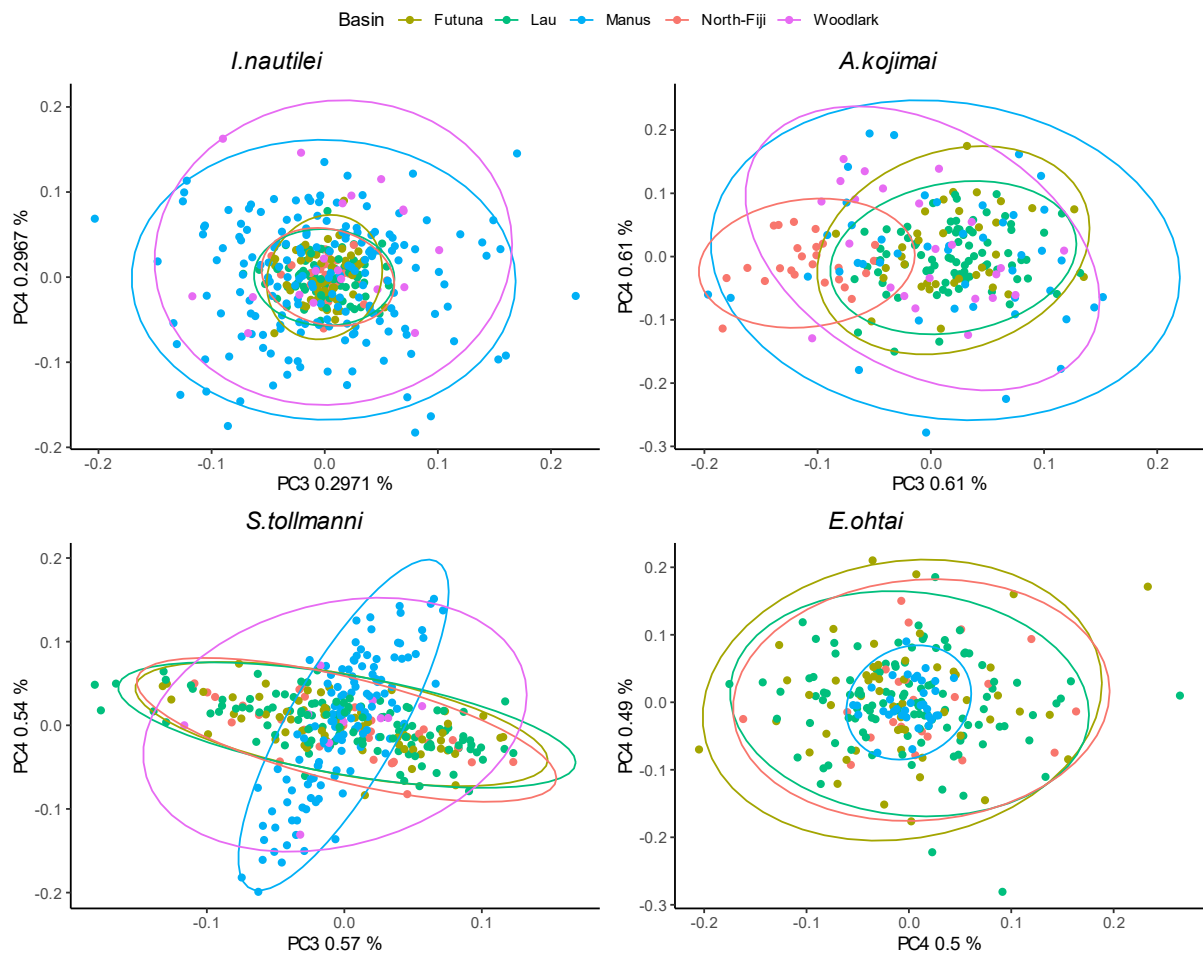
	SD	Nu1 (NFL)	Nu2 (M/W)	b1	b2	hrf	Ts	Tsm/Tsc/tam
<i>E. ohtai</i>	SC2mG	1,343	1,203	31,559	4,080		0,656	1,000
	SD	3,519	0,322			0,010	0,135	0,023
<i>S. tollmanni</i>	SC2N2m	16,057	1,527			0,042	1,037	0,089
	SD	0,516	nan	1,123	0,034	0,009	0,129	0,036
<i>S. tollmanni</i>	SC2N2mG	5,571	4,669	2,023	0,431	0,040	0,991	0,110
	SD	nan	0,715	9,117	4,797	0,005	0,039	0,074
<i>A. kojimai</i>	SC2N2mG	1,432	2,430	17,493	5,418	0,023	0,534	0,202
	SD	0,347	0,354	5,350	10,595	0,011		0,377
<i>I. nautilei</i>	IM2NG	1,081	0,808	12,614	19,360	0,063		1,767
	SD	0,228	0,143	2,536	4,455	0,007	0,013	0,214
<i>I. nautilei</i>	AM2NG	1,162	0,915	13,008	15,985	0,054	0,017	1,613
	SD	0,181	0,233	0,788	3,565	0,013	0,198	0,257
<i>I. nautilei</i>	SC2NG	1,493	1,137	8,973	13,575	0,064	0,156	1,560
	SD	0,127	0,119	12,399	10,523	0,006	0,024	0,335
	Modèle/ Paramètre	Nu1 (NFL)	Nu2 (M/W)	b1	b2	hrf	Ts	Tsm/Tsc/tam

	0,053	0,120	0,002	0,010		0,013		64,298
	0,364	1,386	0,013	0,124		0,508		1 268,530
	0,113	0,974	1,493	4,009		0,148	0,091	7,896
	0,321	3,314	4,334	13,866		0,014	0,434	107,227
	0,104	0,613	1,589	1,781		0,060	0,067	7,945
	0,338	2,697	4,040	12,882		0,025	0,418	109,771
	1,616	1,468	1,187	2,694		0,635	0,050	39,869
	5,230	5,871	2,394	6,845		0,999	0,301	977,632
	0,033	0,059					0,029	44,339
	0,178	0,495					0,570	442,216
	0,040	0,047					0,029	34,696
	0,217	0,492					0,584	457,770
	0,036	0,075					0,031	50,233
	0,180	0,490					0,566	439,020
	0,137	0,198	0,018	0,070		0,095	0,136	32,988
m12	m21	me12	me21	P	Q	Theta		

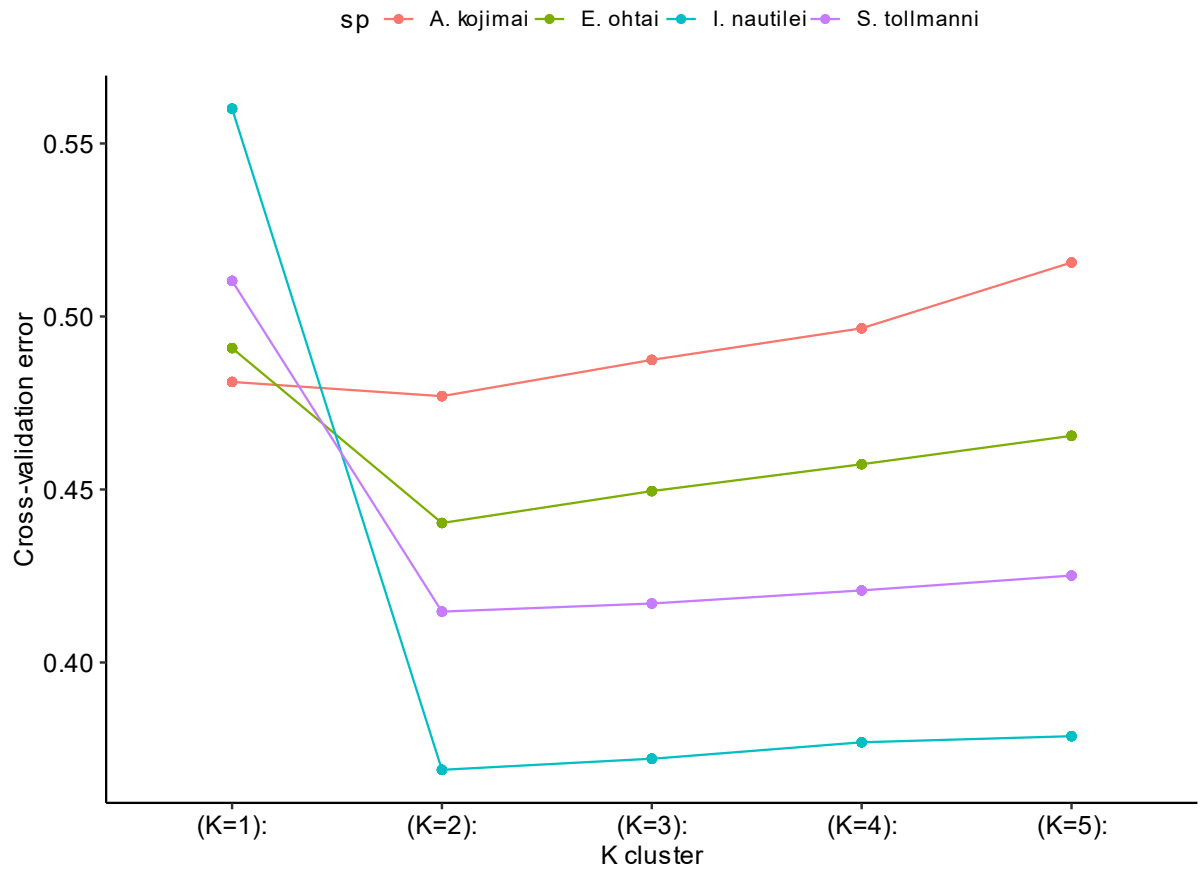
Figures



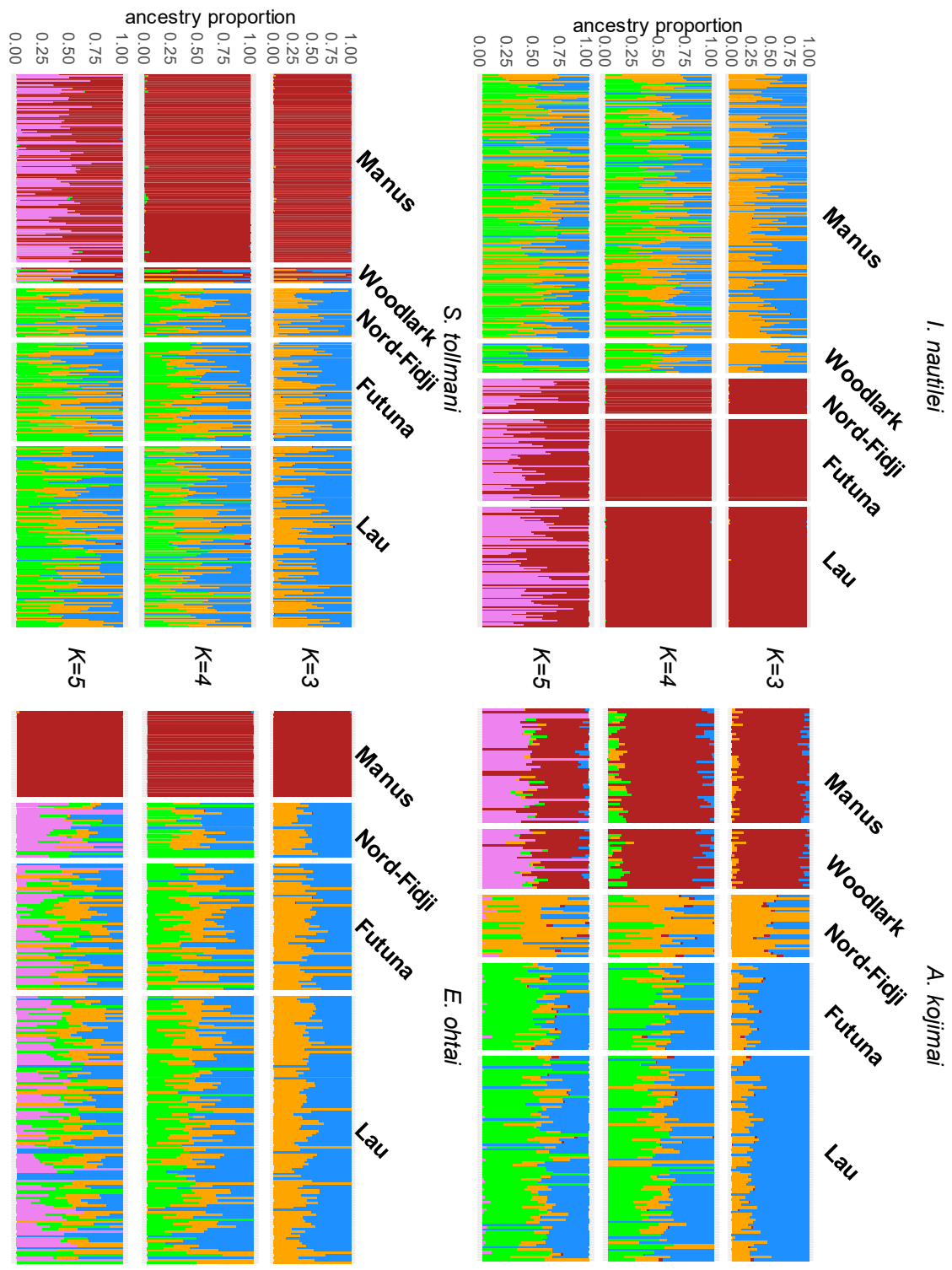
SI Figure 24 : Analyses en composantes principales (ACP) avec les deux premières composantes pour *S. tollmanni*. La première ACP représente les individus par bassin et la deuxième affiche la proportion de données manquantes par individu. Le cercle représente la distribution normale multivariée à 95% de chaque groupe (bassin).



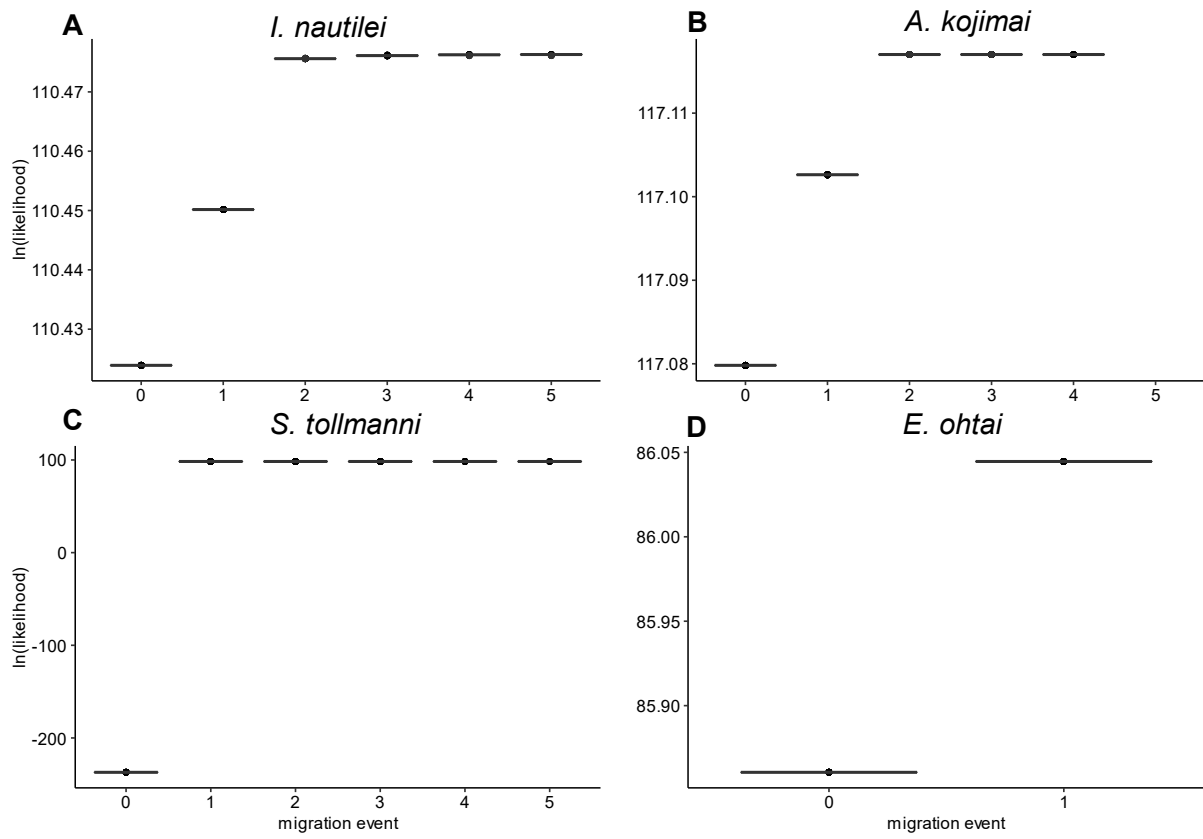
SI Figure 25 : Analyses en composantes principales (ACP) par espèce pour la troisième et quatrième composante. Les points représentent les individus et les couleurs représentent les bassins arrière-arc. Le cercle représente la distribution normale multivariée à 95% de chaque groupe (bassin).



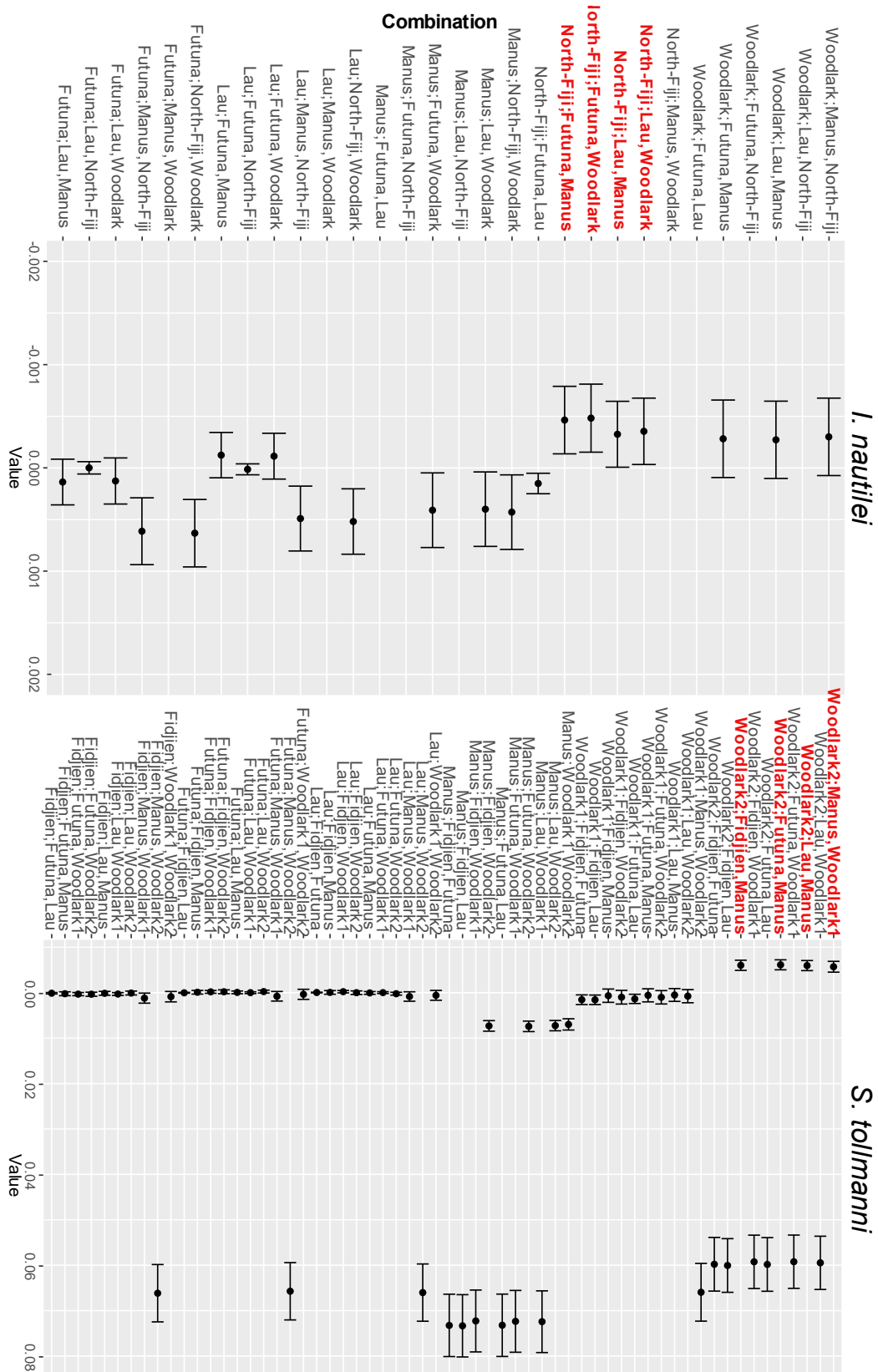
SI Figure 26 : taux d'erreur de validation croisée pour les 10 runs d'ADMIXTURE par valeur de K et par espèce. Chaque couleur représente une espèce.



SI Figure 27 : analyses ADMIXTURE pour K= 3 à 5 pour chaque espèce en séparant les individus par bassin géographique. Chaque barre représente un individu et chaque couleur sa proportion relative à chaque fond génétique inféré.

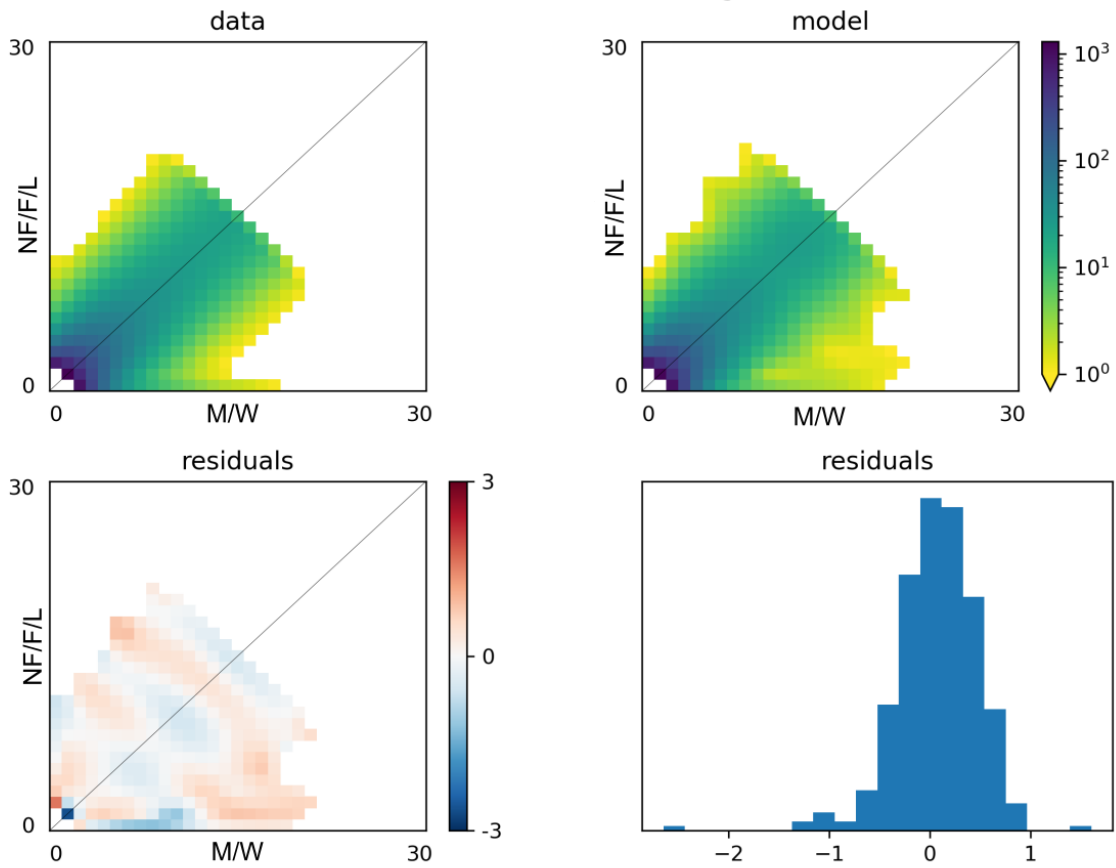


SI Figure 28 : Logarithme de la vraisemblance ($\ln(\text{likelihood})$) par espèce pour les 10 runs de Treemix pour chaque valeur d'évènement de migration (m). La barre noire représente la boîte à moustache des 10 run de Treemix.



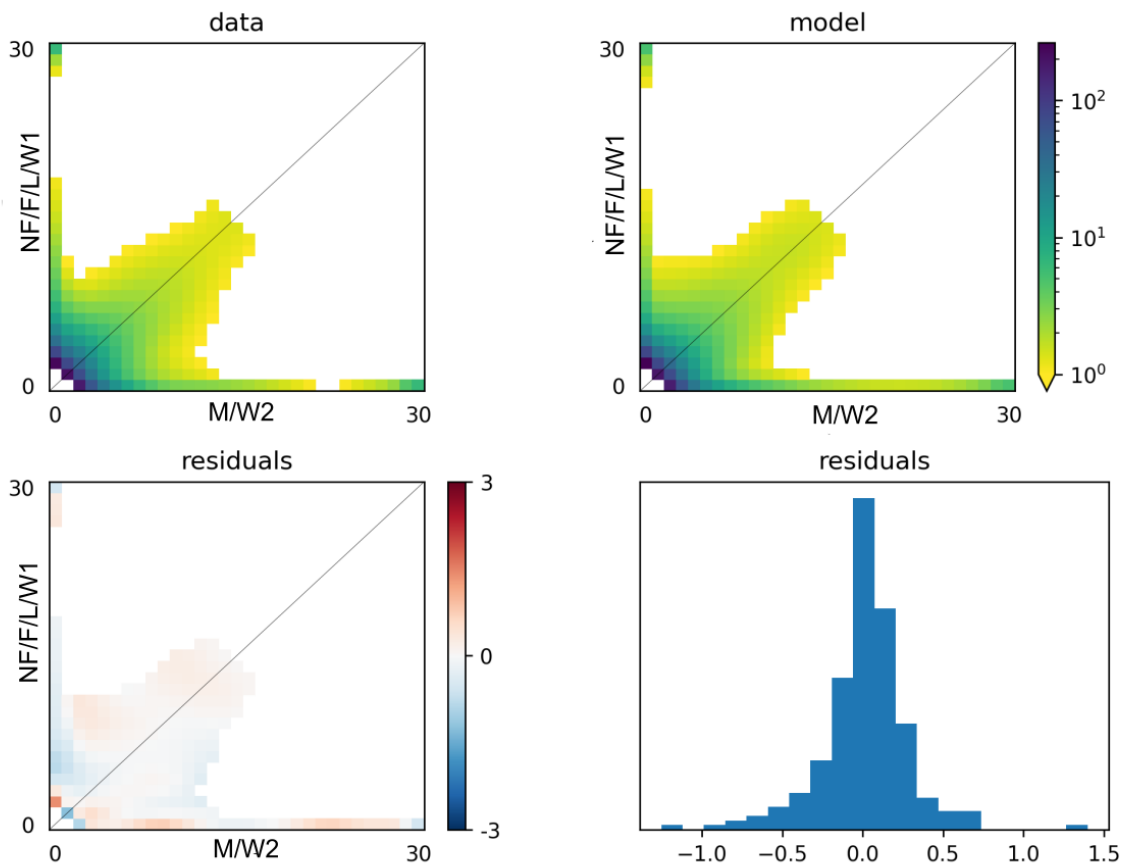
SI Figure 28 : Valeurs de la statistique F_3 (point) et son intervalle de confiance à 95% (moustache). Les valeurs non-affichées sont fortement positives. Les noms en rouge indiquent des valeurs significativement négatives. (Population focale ; population source 1, population source 2).

SC2N2mG *A. kojimai*



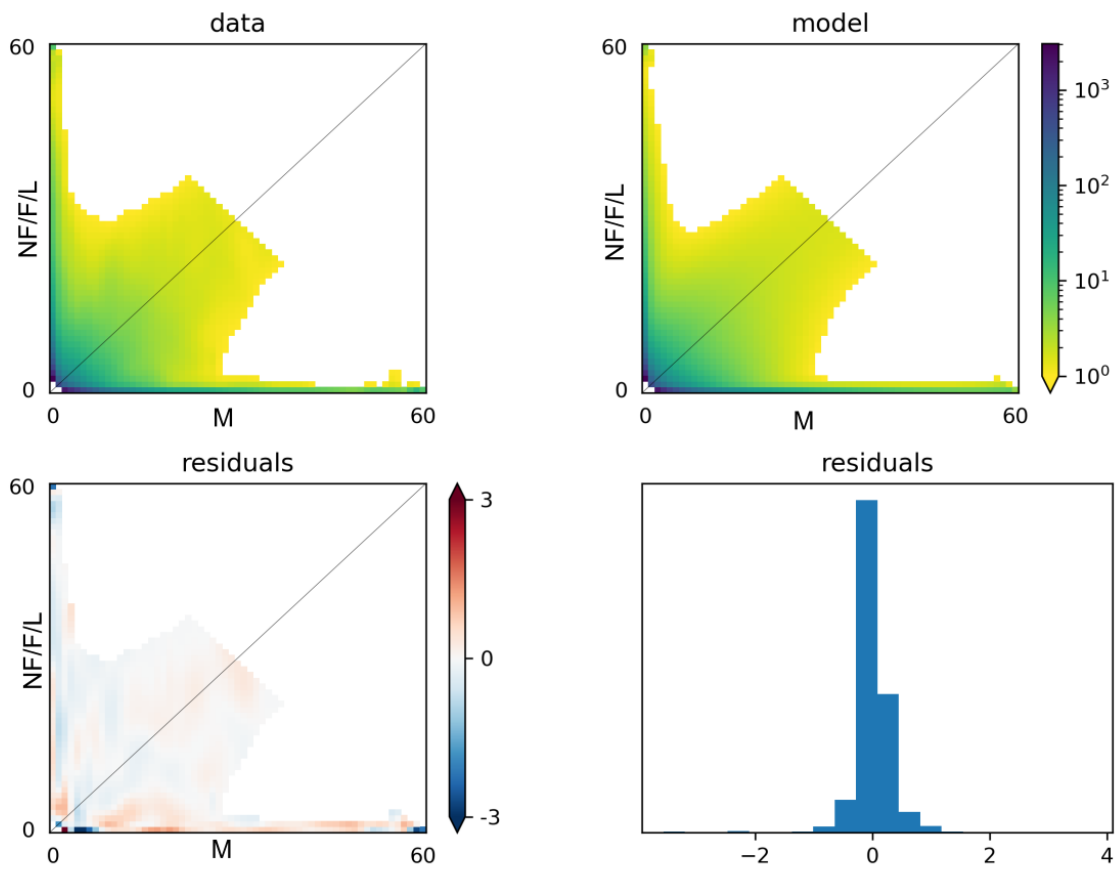
SI Figure 29 : Meilleur modèle et run pour *A. kojimai* sous le modèle SC+2N+2m+G. le spectre observé en haut à gauche. Le spectre simulé en haut à droite. Le spectre des résidus et leurs distributions en bas.

SC2N2mG *S. tollmanni*



SI Figure 30 : Meilleur modèle et run pour *S. tollmanni* sous le modèle SC+2N+2m+G. le spectre observé en haut à gauche. Le spectre simulé en haut à droite. Le spectre des résidus et leurs distributions en bas gauche et droite.

SC2mG *E. ohtai*



SI Figure 31 : Meilleur modèle et run pour *E. ohtai* sous le modèle SC+2m+G. le spectre observé en haut à gauche. Le spectre simulé en haut à droite. Le spectre des résidus et leurs distributions en bas gauche et droite.

REMERCIEMENTS

Je vais commencer en remerciant les deux personnes sans qui la réalisation de cette thèse n'aurait pas été possible : François et Stéphane. Tout d'abord merci à vous deux de m'avoir choisi pour réaliser cette incroyable aventure. Sans vous, je n'aurais jamais pu découvrir ces écosystèmes si particuliers que sont les sources hydrothermales et les espèces qui y demeurent tout en m'épanouissant dans un environnement scientifique, incroyablement riche, diversifié et motivant.

Stéphane, merci d'avoir été mon directeur de thèse, même si au début nos échanges n'étaient pas très fréquents, tu as été présent pour m'apporter ton soutien, tes points de vue et ton enthousiasme bienveillant sur cette thèse. Sans ton aide au début de la mission et ton apport sur la rédaction, la biologie et les écosystèmes des sources hydrothermales, la thèse aurait été toute autre.

François, merci d'avoir pris particulièrement le temps de me former un dernier étudiant un peu farfelu tel que moi. Merci pour ton importante implication sur cette thèse allant des manip à bord de l'Atalante à toutes les longues discussions que l'on a eu aussi bien sur la science et la génétique que sur d'autres sujets divers et variés. Je n'aurais jamais autant appris en discutant avec toi que tout ce que j'ai pu lire au cours de ces trois ans et demi. Cela n'a pas toujours été facile, mais tu as toujours pris le temps de m'expliquer les choses et d'avoir pris patience pour corriger mes erreurs. Tu as aussi su apporter une assurance concernant la réussite de ce projet de thèse et me rassurer lors des moments de doute. Bien qu'étant officiellement le dernier de tes doctorants (mohicans), j'espère que tu trouveras toujours un peu de temps pour discuter de science, car ton expérience et avis sont toujours enrichissants.

Merci Didier, même si officiellement, tu ne faisais pas partie de l'encadrement de thèse, tu as été l'acteur de l'ombre. Tu as su apporter tout autant que Stéphane sur les connaissances en biologie et génétique des sources hydrothermales sur ce projet de thèse, mais tu as su aussi me pousser à chercher dans de nouvelles directions !

À vous trois, vous avez su attiser ma curiosité concernant l'évolution et le milieu hydrothermal tout en repoussant aussi mes limites et élargissant les réflexions. Sans vous, je n'aurais pu réaliser ce super projet de thèse, malgré la situation dégradée que l'on a eu au cours de ces dernières années. J'espère aussi que nous continuerons à discuter et à échanger.

Je souhaiterais également remercier les membres de mon jury Cynthia Riginos, Sarah Samadi, Pierre Chevaldonné et Mathieu Gautier pour d'avoir accepté de faire partie de mon jury et de vous être intéressé à mes travaux. Je remercie également les membres de mes comités de thèse Renaud Vitalis, Pierre-Olivier Antoine, Jean-Christophe Avarre, Carole Smadja, Pierre-Alexandre Gagnaire et Nicolas Bierne.

Ayant commencé ma thèse projetée à Roscoff, je voudrais tout d'abord remercier toutes les personnes que j'ai croisées et qui ont contribué à la bonne ambiance de la station biologique et au projet Cerberus. Je voudrais particulièrement remercier la team Camtor Jr et Sr, maintenant Docteur(e)s, sans vous et votre accueil chaleureux dans le bureau 116 ainsi que vos petites farces de tous les jours, la vie roscovite n'aurait pas été aussi drôle et chouette. Avec dudu, merci d'avoir été mes rayons de soleil breton de tous les jours et de me donner « un petit goût de reviens-y ;) ». Merci à toutes l'équipe DYDIV, que ça soit pour leurs sourires de tous les jours et leurs soutiens lors des manip. Je remercie ainsi Thomas, Jade, Claire, Eric, Thierry, Ann, Momo, Stéphanie, Anne-Sophie, Marion, François et tous les autres membres de l'équipe. Merci aussi tous les autres doctorant(e)s, permanent ou non, avec qui j'ai passé du temps. Ce fut une opportunité unique d'avoir pu aller à Roscoff, de vous rencontrer et de réaliser cette thèse en collaboration avec les membres de DYDIV et du projet Cerberus.

Je souhaite également remercier toute l'équipe BEM, aux membres actuels ou passé, Pierre-Alexandre, Nicolas, Fred, Pierre, Maurine, Marie, Laura, Fanny, Alexis et Maud, Nelly, Stephen, Emilie, Christine, Johan et Bruno pour contribuer à la bonne ambiance, bienveillance et à toujours être là pour filer un coup de main et nous motiver à venir au labo !! Merci aux nombreuses discussions et débats autour d'un café ou à midi. C'est particulièrement agréable de pouvoir apprendre et faire de la recherche dans une ambiance aussi familiale, agréable et sans prise de tête. Un grand merci aux doctorants et docteur(e)s de l'équipe, franchement sans vous ça n'aurait pas été pareil surtout durant cette période de pandémie. Vos sourires et vos présences tous les jours m'a permis de garder la grinta !!

Je remercie aussi le personnel de l'ISEM et les copains de l'ISEM/CEFE, pour toutes les sorties durant ces dernières années. Vous avez su apporter une ambiance géniale les apéros sur le campus ou ceux du bois de montmaur !! :) Je souhaite plus particulièrement remercier Anjélica, Martin et François, mais aussi tous les autres, pour tous les moments passés qu'ils soient autour de randonnée, sortie astronomie dans les Cévennes ou autres !!

Merci à ma merveilleuse famille de m'avoir toujours soutenu et encouragée au cours de cette aventure. Malgré mon manque de nouvelle, vous avez su m'accompagner et me motiver à chaque instant, particulièrement pendant cette période difficile de pandémie et de fin de thèse.