Rôle de la géographie et de l'hétérogénéité de l'environnement hydrothermal profond sur la distribution et l'évolution du complexe de gastéropodes Alviniconcha spp

Jade Castel

Thèse de doctorat de Sorbonne Université

Ecole doctorale 227 : *Science de la Nature et de l'Homme : Ecologie et Evolution*

Station biologique de Roscoff – UMR 7144 Adaptation et diversité en milieu marin / DYDIV

Présentée par

Jade CASTEL

En vue de l'obtention du grade de docteur de Sorbonne Université

En génétique des populations et génomique

Rôle de la géographie et de l'hétérogénéité de l'environnement hydrothermal profond sur la distribution et l'évolution du complexe de gastéropodes *Alviniconcha* spp.

Présentée et soutenue publiquement le 23 Mai 2022,

Devant un jury composé de :

Mme Cindy Van Dover	Pr., Duke University	Rapportrice
M. Eric Pante	CR (HDR), CNRS	Rapporteur
Mme Christelle Fraisse	CR, CNRS	Examinatrice
M. François Lallier	Pr., Sorbonne Université	Examinateur
M. Stéphane Hourdez	CR (HDR), CNRS	Invité
M. Didier Jollivet	DR (HDR), CNRS	Directeur de thèse
M. Thomas Broquet	CR (HDR), CNRS	Directeur de thèse

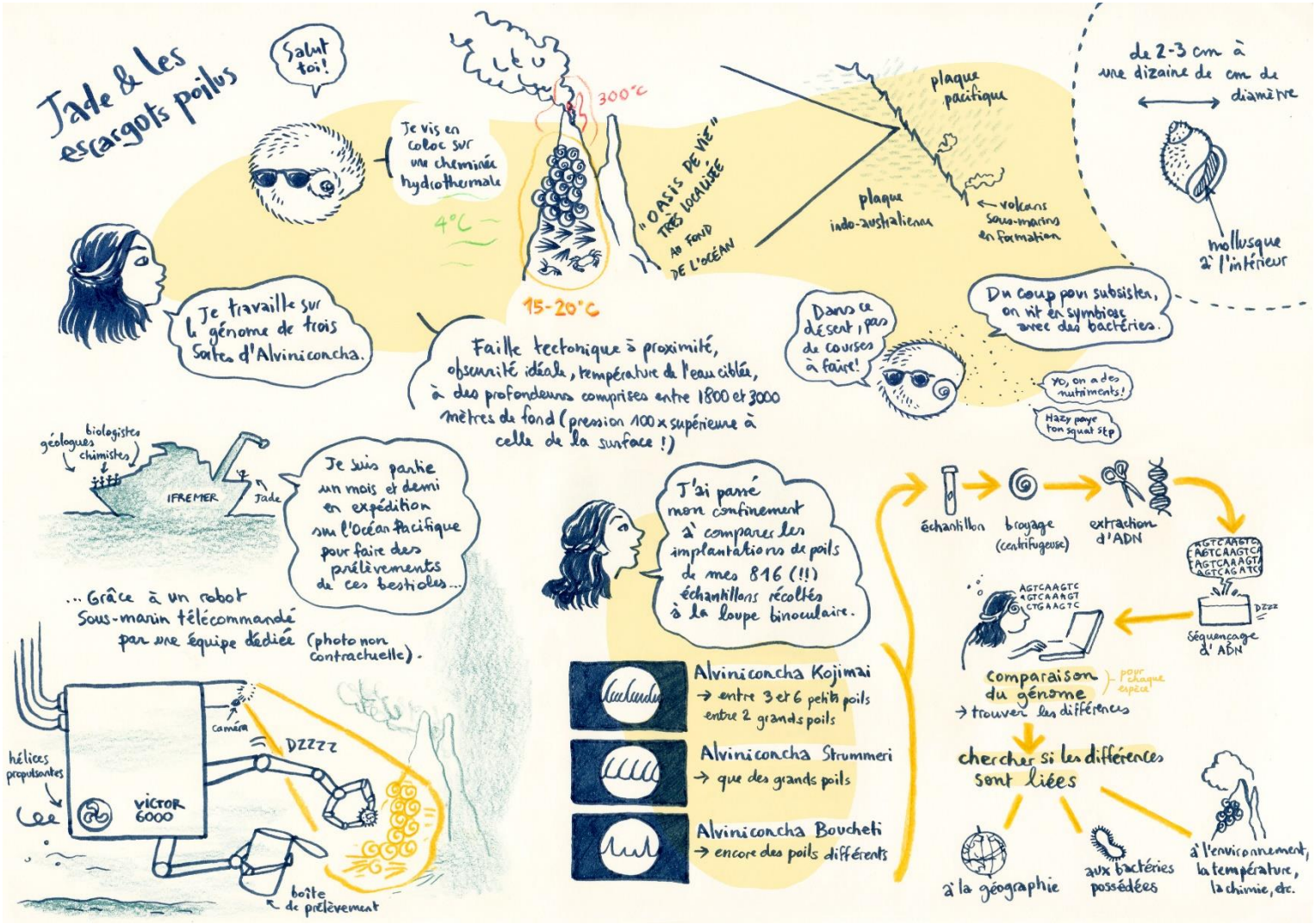


Illustration réalisée par Hélène Bléhaut lors de sa visite à la station biologique de Roscoff dans le cadre de la découverte des petites cités de caractère de Bretagne (avril-mai 2021).

Remerciements

Je voudrais tout d'abord remercier les membres de mon jury, Eric PANTE et Cindy VAN DOVER d'avoir accepté d'être rapporteurs de ma thèse, ainsi que Christelle FRAISSE, François LALLIER et Stéphane HOURDEZ d'avoir accepté de participer. Un grand merci également à Pierre-Alexandre, Florence, Stéphane, Erwan et Catherine pour leurs conseils au cours de mes comités de thèse.

Un grand merci à Stéphane et Didier pour avoir monté le projet Cerberus qui a permis de financer en partie cette thèse et qui m'a permis de participer à une super campagne d'échantillonnage. D'ailleurs, merci à tous les marins, les pilotes du ROV et les scientifiques à bord sans qui cette thèse n'aurait jamais eu lieu. Un grand nombre de personnes ont contribué à l'acquisition/dissection/extraction d'ADN des *Alviniconcha* lors de cette campagne. C'est notamment le cas de Didier, Thomas, Nicolas, Marie-Anne, Pierre-Alexandre, Florence, Jean et Valérie. Merci également à toute l'équipe DYDIV de m'avoir accueilli et soutenu autant professionnellement que personnellement pendant ces trois dernières années. Un merci tout particulier à Stéphanie, Claire, Marion et Anne-Sophie pour leur précieuse aide lors des nombreuses manips de cette thèse. Un grand merci à Thierry pour les longues discussions parfois philosophiques sur la vie, elles resteront longtemps dans mes souvenirs. Comment ne pas dire merci à l'équipe ABIMS et notamment à Erwan et Gildas pour leur formation et leur aide lors des analyses bio-informatiques, c'était pas gagné. Merci à tous ceux qui sont passés dans mon bureau, je pense à Camille, Victor, Adrien (dans le fameux bureau 116) mais plus récemment Vincent et bien évidemment Elodie pour leur bonne humeur et les discussions qui ne font pas avancer la science mais qui font du bien. Enfin bref, un grand merci à toute l'équipe.

Mais parce que la station, c'est aussi tout un tas de monde, merci à Domi pour ses gaufres et sa pâte à tartiner (mais surtout pour sa gentillesse), Fred, Ian et Céline. Merci à Stéphane E pour ses conseils bienveillants et pour avoir sauvé mon œil quand tout le monde était parti manger. Merci au Ty Pierre, au début du mercredi puis finalement de n'importe quel jour, pour leur accueil chaleureux et leur bonne humeur, merci notamment à Polo et Dorian. Et parce que, que serait le Typ Pierre sans les copains, un énorme merci à Lena, Maéva, Mael, Lisa, Ulysse, Louison, Yasmine, Emma, Camille, Dan, Marie-Morgane, Morgane, Sam, Flo … J'oublie certainement plein de monde, je suis désolé.

A Roscoff, on se fait aussi des amis, des vrais, qui sont toujours là pour remonter le moral, donc un immense merci à Damien, Bertille, Erwan, Lancelot, Charlotte, Stéphane et Mathilde. Merci aussi aux amis de la fac, Aurélien, Térence, Claire, Mathilde, Juliette, Camille, Alexis, Quentin pour les skypes du soir pour se raconter nos vies. Merci également à ma famille pour avoir toujours cru en moi et pour m'avoir laissé faire mes propres choix. Un merci tout particulier à mon papi pour son aide lors des devoirs lorsque je n'étais encore qu'une enfant, ta rigueur et ta motivation ont été un exemple à suivre et c'est certain sans toi je ne serai pas là.

Enfin, parce que sans eux, absolument rien n'est possible, merci à mes supers encadrants de thèse. Merci de m'avoir fait confiance et soutenu pendant ces trois dernières années. Merci à Didier de m'avoir fait découvrir le monde et sa passion pour la vie dans le milieu hydrothermal, merci encore également de m'avoir permis de participer à une super mission d'échantillonnage. C'était beaucoup de travail, et de massacre d'escargots, peu de sommeil, mais des souvenirs incroyables. Merci également pour ta rigueur scientifique et tes connaissances immenses, c'était un honneur de travailler avec toi. Merci à Thomas, pour sa rigueur et ses connaissances en génétique des populations particulièrement. Mais merci aussi et surtout pour ton écoute, tes conseils et ta bienveillance. Ne changez rien, vous formez un super duo et je pense que grâce à vous j'ai grandi pendant cette thèse. Merci vraiment.

Un dernier merci très personnel à Martin, pour ta patience, ton écoute et ton soutien à toute épreuve surtout lors de la fin de la thèse lorsque le stress se faisait quelque peu sentir. Qu'est-ce qu'il a bien fait mon portefeuille de se téléporter dans ta veste … Merci simplement d'être là pour moi au quotidien, tel que tu es.

Table des matières :

Liste des figures :

Introduction générale

Echantillonnage et acquisition des données génétiques

Chapter 1

Appendix 1

Chapter 2

Chapter 3

Chapter 4

Discussion générale

Liste des tableaux :

Echantillonnage et acquisition des données génétiques

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Introduction générale

1- La spéciation : théorie et mécanismes d'isolement

La spéciation est un processus évolutif conduisant à l'émergence de nouvelles espèces à partir d'une espèce ancestrale. Elle est actuellement considérée comme un processus plus ou moins continu le long duquel une population panmictique aboutit au final à deux populations reproductivement isolées. Le long de ce continuum, c'est la mise en place progressive de barrières à la reproduction qui amène les individus de la population ancestrale à limiter leurs échanges génétiques jusqu'à la formation de deux espèces distinctes lorsque l'isolement reproducteur est complet. Même si le concept biologique de l'espèce tel que défini par Mayr (1942) comme *un ensemble d'individus effectivement ou potentiellement capables de se reproduire et d'engendrer une descendance viable et féconde, tout en étant reproductivement isolés d'autres groupes similaires d'individus* reste toujours d'actualité, il n'est pas rare d'employer le terme d'espèces différentes alors même que les individus desdites espèces sont encore capables d'hybridation localement ou artificiellement remises en contact (concept de pseudo-espèces ; O'Mullan et al., 2001 ; Springer and Crespi, 2007). D'ailleurs le concept d'espèce biologique tel qu'adopté par Coyne et Orr (2004) adopte la possibilité d'hybridation introgressive limitée. On suppose alors que les barrières à la reproduction sont suffisamment fortes pour maintenir l'intégrité génétique de chaque population jusqu'à ce que l'isolement reproducteur soit complet alors même que des échanges sont encore possibles localement dans le génome (notion de barrière semi-perméable). Ainsi, l'étude de la spéciation revient à étudier les mécanismes conduisant à l'isolement reproductif qui limite le flux de gènes entre les populations.

1.1- Les différents mécanismes d'isolement reproducteur

1.1.1- Barrière pré-zygotique

Les barrières pré-zygotiques empêchent ou limitent la fécondation entre les gamètes mâles et femelles et peuvent survenir avant même l'accouplement. Ces barrières peuvent être de différentes natures (Figure 1) :

- Écologique : les populations « parentales » ne partagent pas les mêmes habitats (barrières géographique ou environnementale) ou les mêmes aires ou périodes de reproduction. Ainsi les individus n'ont pas l'occasion de se rencontrer pendant leurs périodes de reproduction pour s'accoupler ou frayer.

- Comportementale : les partenaires diffèrent par leurs caractères attractifs (sonores, visuels ou olfactifs ou encore ont une parade nuptiale différente) qui orientent le choix du partenaire dans son groupe d'origine (sélection sexuelle).
- Morphologique : les organes reproducteurs présentent des incompatibilités morphologiques ou physiologiques, ce qui empêche l'accouplement ou la survie des gamètes.
- Gamétique : les gamètes mâles hétérospécifiques présentent des problèmes de motilité ou de viabilité dans les voies génitales femelles dans le cas d'espèces à fécondation interne ou présentent des incompatibilités dans la fusion de l'acrosome avec la zone pellucide de l'ovocyte.

Chez les poissons Cichlidae par exemple, la couleur du poisson mâle est un caractère utilisé par les femelles pour choisir leur partenaire, ce qui favorise la reproduction au sein des groupes dans une même aire de distribution (Selz et al., 2014). Certaines études ont également montré que les différences dans les interactions symbiotiques entre bactéries/parasites et hôtes peut être une source d'isolement reproductif pré-zygotique entre les populations, en agissant sur l'habitat (isolement écologique) mais aussi sur le comportement des hôtes (Brucker et Bordenstein, 2012 ; Ezenwa et al., 2012 ; Shropshire et Bordenstein, 2016). C'est par exemple le cas chez le puceron du pois (*Acyrthosiphon pisum*) ou l'acquisition chez certains individus d'un symbionte de la classe des *Gamma-proteobacteria* a permis une augmentation de la fitness des individus sur les plantes de trèfle blanc (*Trifolium repens*). Cette adaptation a entrainé une modification de la niche écologique chez ces individus qui a conduit à de l'isolement écologique entre les populations de pucerons (Tsuchida et al., 2004).

1.1.2- Barrière post-zygotique

Une fois que la fécondation a eu lieu, les barrières post-zygotiques interviennent tout au long du cycle de vie de l'individu hybride en limitant sa survie ou son potentiel reproductif. Il existe des barrières intrinsèques post-zygotiques empêchant le bon développement de l'embryon ou engendrant la stérilité/mortalité des hybrides (Brannock et Hilbish, 2010 ; Maheshwari et Barbash, 2011) et des barrières post-zygotiques extrinsèques liées à l'environnement, comme le fait que les hybrides soient maladaptés à l'environnement (baisse de la fitness par rapport aux lignées parentales) ou le fait que les hybrides se reproduisent moins bien ou moins souvent en raison de leur attractivité plus faible (isolation sexuelle) (Hauser, 2002). Les barrières post-zygotiques intrinsèques sont dues à un dysfonctionnement

entre les génomes parentaux qui se retrouvent ensemble au sein de l'individu hybride, c'est par exemple le cas lorsque le système reproductif est différent entre les parents. En effet, chez certaines espèces, la survie des hybrides après l'accouplement entre des individus gonochoriques et hermaphrodites dépend du sens du croisement entre géniteurs. Chez les nématodes du genre *Caenorhabditis*, les mâles hybrides ne sont viables que si la « mère » provient de l'espèce gonochorique. A l'inverse, un mâle qui descendrait d'un croisement entre une femelle hermaphrodite et un mâle gonochorique ne survivrait pas (Coyer et al., 2002 ; Woodruff et al., 2010).

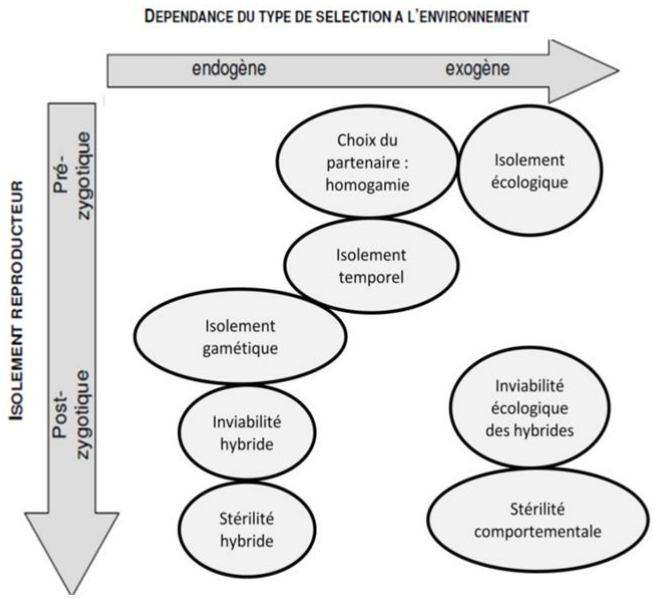


Figure 1: Mécanismes d'isolement reproducteur classés en fonction du moment du cycle du vie où ils agissent (pré ou post-zygotique) et de leur dépendance à l'environnement. Adapté de Thomas et al. (2010) dans Gay (2006).

Dobzhansky (Dobzhansky, 1937) et Muller (Muller, 1942) ont proposé un modèle permettant d'expliquer l'apparition d'un isolement reproductif au travers des interactions épistatiques (interactions existantes entre deux ou plusieurs gènes). Si on prend l'exemple des interactions entre deux locus bi-alléliques alors deux populations vont pourvoir fixer indépendamment deux nouveaux allèles à deux locus différents (a et b) (Figure 2). Ces deux allèles sont neutres ou avantageux dans leur fond génétique d'origine (c'est-à-dire respectivement associé à B et A), mais leur association peut être délétère dans les génotypes

hybrides (aA et bB). La fixation de mutations faiblement délétères par dérive génétique dans l'une des populations, suivie par la fixation de mutations compensatoires (Kimura, 1985) peut également être à l'origine d'incompatibilités Dobzhansky-Muller (DMI) (parfois également appelées BDMI pour Bateson-Dobzhansky-Muller incompatibilities (Bateson, 1909)). De plus, si les interactions entre allèles sont plus compliquées avec des relations de dominance des allèles ancestraux, les effets délétères ne seront révélés que si les allèles dérivés sont présents à l'état homozygote, ce qui peut se produire dans les générations d'hybridation plus tardives (post F1). Un exemple d'interactions épistatiques négatives entre gènes d'isolement reproductif est celui observé entre les gènes mitochondriaux et nucléaires impliqués dans des fonctions mitochondriales, à l'origine de DMI (Burton et Barreto, 2012 ; Telschow et al., 2019). Un exemple de DMI a pu être montré chez les bivalves *Macoma balthica* où des incompatibilités génétiques cyto-nucléaires existent chez les individus hybrides (Pante et al., 2012).

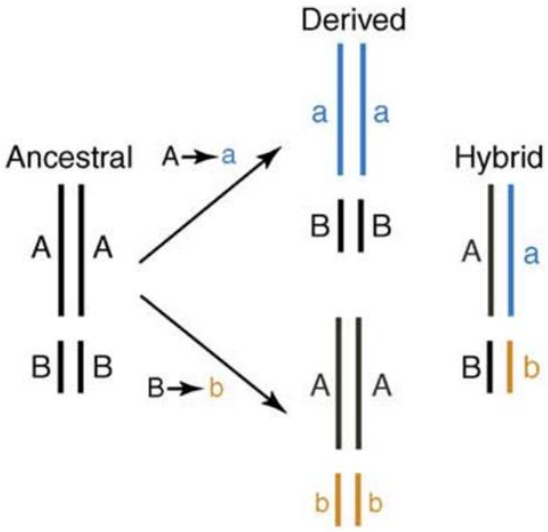


Figure 2: Modèle bi-locus d'une incompatibilité de Bateson-Dobzhansky-Muller (BDMI). Deux lignées divergent l'une de l'autre à partir d'une lignée ancestrale et chaque population fixe un nouvel allèle à un locus différent. La formation de génotypes hybrides entre ces deux lignées crée de nouvelles combinaisons alléliques qui peuvent être incompatibles.

Les mécanismes d'isolement pré et post-zygotiques ne sont pas mutuellement exclusifs et peuvent agir de manière simultanée (Bierne et al., 2013), et il est même admis que la complétion de la spéciation nécessite un couplage entre plusieurs mécanismes d'isolement,

qui, seuls, ne pourraient provoquer un isolement reproducteur complet (Butlin et Smadja, 2018).

1.2- Contexte géographique

Le contexte spatial joue un rôle primordial dans l'isolement des populations et la formation d'espèces car il va influencer l'intensité des forces évolutives mises en jeu (Coyne et Orr, 2004). Il existe principalement trois contextes géographiques de spéciation : allopatrique, sympatrique et parapatrique.

Lors d'une spéciation allopatrique, la divergence accumulée entre les deux populations est facilitée par une barrière physique (une chaîne de montagne par exemple) qui empêche la migration des individus et donc le flux de gènes entre génomes (Coyne et Orr, 2004). Les allèles polymorphes hérités de la population ancestrale vont évoluer différemment dans chacune des populations isolées sous l'effet de la dérive génétique (tri des lignées alléliques) et de la mutation, ce qui va augmenter la divergence génétique inter-populationnelle. Ce phénomène d'évolution des allèles entre deux populations isolées se nomme le tri des lignées ancestrales, et on considère qu'il se termine après environ $6N_e$ générations (Rosenberg, 2003). La divergence est d'autant plus accentuée que de nouvelles mutations apparaissent indépendamment dans chaque population. Ce processus va être à l'origine de l'isolement reproducteur car l'accumulation de la divergence n'est pas contrée par le flux génique. Si par exemple les conditions environnementales sont différentes des deux côtés de la barrière physique, alors la sélection naturelle va pouvoir accélérer l'accumulation de la divergence initiée par la dérive génétique en favorisant des balayages sélectifs de part et d'autre de cette barrière (Coyne et Orr, 2004). Ce mode de spéciation est un phénomène courant et bien documenté, c'est par exemple le cas chez les nudibranches du genre *Dendronotus* où les périodes glaciaires durant le Pliocène-Pléistocène ont entraîné la fermeture/ouverture du détroit de Béring, ce qui a conduit à la spéciation entre les populations de l'Arctique et du Pacifique (Ekimova et al., 2019). Ce processus de spéciation allopatrique a pour la première fois était décrit dans *l'Origine des espèces* (Darwin, 1859), où l'auteur prend l'exemple de l'isolement géographique des différentes espèces de Pinsons trouvés sur les îles des Galapagos pour expliquer ce procédé.

Un cas particulier de la spéciation allopatrique est la spéciation péripatrique. Dans ce contexte géographique, une faible proportion de la population se retrouve séparée par une

barrière physique. La dérive génétique sur cette petite population va être forte car le faible nombre d'individus entraîne une baisse de la diversité génétique et donc un tri plus rapide de fréquences alléliques (effet fondateur) conduisant à l'augmentation de la différenciation génétique (Mayr, 1942 ; Rocha et Bowen, 2008).

Lors d'une spéciation sympatrique, la divergence entre les populations va s'accumuler malgré l'absence de barrières aux flux de gènes. Une sélection disruptive est donc nécessaire pour initier la divergence étant donné que la dérive génétique est contrée par la migration entre les populations (Gavrilets, 2004). L'adaptation locale à deux niches écologiques distinctes peut théoriquement être à l'origine d'une sélection disruptive, sous conditions d'un environnement hétérogène entre les individus. Dans ce cas-là, des groupes d'individus peuvent se spécialiser dans des environnements distincts grâce à l'apparition de mutations et/ou de recombinaisons favorables qui vont être sélectionnées différentiellement suivant l'environnement des individus. Ainsi, chaque individu aura une valeur sélective plus élevée dans l'habitat dans lequel il s'est spécialisé et une valeur sélective réduite dans les autres habitats. Néanmoins, ces mécanismes ne peuvent se mettre en place que dans la mesure où la migration entre individus est faible ou que les coefficients de sélection sont extrêmement forts pour éviter que la recombinaison ne puisse s'opposer au maintien des allèles avantageux dans chaque habitat (Dieckmann et Doebeli, 1999 ; Via, 2009). Les migrants non adaptés vont introduire dans la population des combinaisons d'allèles défavorables et donc des génotypes mal adaptés ce qui va contrer l'adaptation locale et la divergence (Lenormand, 2002). Ce type de spéciation pourra donc se dérouler uniquement s'il existe des mécanismes limitant les échanges génétiques et la recombinaison entre les populations (comme par exemple un réarrangement chromosomique). Plusieurs modèles théoriques ont été proposés pour expliquer ce mode de spéciation, dont le modèle d'association entre gènes sous sélection et gènes de l'isolement reproducteur (Kondrashov et Kondrashov, 1999), la gestion de conflits sexuels (Gavrilets et Waxman, 2002), ou l'accumulation différentielle de mutations délètères (Kawecki, 1997). Plus récemment, des modèles basés sur l'architecture génomique de la spéciation, et notamment la mise en place d'îlots d'adaptation, ont également été proposés pour expliquer l'atténuation progressive des flux de gènes entre habitats (Feder et al., 2012, 2014 ; Feder et Nosil, 2010 ; Flaxman et al., 2014). La spéciation sympatrique est cependant considérée comme rare (Coyne et Orr, 2004) et peu d'exemples non controversés existent dans la littérature. L'exemple le plus connu est celui de la diversification des poissons de la famille des Cichlidés dans les lacs africains (Schliewen et al., 1994) ou les épinoches

(Hohenlohe et al., 2010). L'isolement des populations en sympatrie peut également être lié à un effet temporel des habitats sur la reproduction, en introduisant des comportements reproducteurs différents ou des décalages des appariements ou des pontes comme précédemment suggéré chez l'échinoderme *Acrocnida brachiata* par Muths et al. (2006) sur la base du modèle d'isolement par le temps développé par Hendry et Day (2005).

Lorsque la spéciation par l'habitat s'établit le long d'un gradient environnemental (spéciation parapatrique), il s'opère alors un couplage entre l'atténuation du flux génique par la distance géographique et la sélection disruptive qui œuvre de concert. Ce mode de spéciation apparaît alors comme un mécanisme intermédiaire d'isolement entre la spéciation allopatrique et sympatrique (Doebeli et Dieckmann, 2003). Avec l'adaptation locale, un gradient environnemental induit une différenciation spatiale au cours du temps, il y a donc une corrélation entre la localisation spatiale et le phénotype des individus. Ainsi, une compétition va se mettre en place entre les individus spatialement proches et sera moins importante entre les individus phénotypiquement distants ce qui va au cours du temps augmenter la divergence entre les populations (Doebeli et Dieckman, 2003). Ainsi, la divergence entre populations s'accumule en présence de migration mais avec un flux génétique limité. L'isolement reproducteur augmente si l'effet de la migration est plus faible que celui de la dérive génétique et de la sélection naturelle (Gavrilets et al., 2000). Comme lors de la spéciation sympatrique, une sélection disruptive est nécessaire pour initier la divergence, ce qui suppose par exemple une hétérogénéité environnementale. Ce type de spéciation va notamment se réaliser lorsqu'un gradient environnemental existe et où les populations vont finir par se spécialiser au cours du temps aux deux extrémités de ce gradient (Rolán-Alvarez, 2007).

D'une façon générale, les nombreuses études sur la spéciation montre qu'il existe un continuum entre ces différents modes de spéciation selon que les mécanismes d'isolement pré- et post-zygotiques intrinsèques et extrinsèques s'établissent dans un ordre ou dans un autre au cours du temps (Seehausen et al., 2014). L'étude de ces mécanismes d'isolement doit en effet prendre en compte que les populations ont évolué dans l'espace au cours du temps en fonction notamment des changements climatiques passés (Hewitt, 1996). En effet, des espèces actuellement en sympatrie ne l'étaient peut-être pas dans le passé, elles ont pu être remises en contact secondairement (ou même avec une histoire complexe d'isolements/ contacts répétés),

donnant lieu à des juxtapositions de populations où de l'hybridation peut avoir lieu si les mécanismes pré-zygotiques d'isolement ne sont pas complets.

Une zone d'hybridation se définit comme une zone géographique où deux populations génétiquement distinctes se rencontrent, se reproduisent et engendrent une descendance plus ou moins fertile (Harrison, 1993). Ces zones d'hybridation peuvent se former de deux manières : soit les populations ont divergé sans barrière géographique par l'effet de la sélection disruptive liée notamment à l'environnement (contact primaire, ou intergradation primaire), soit la différenciation s'est faite lors d'une période en allopatrie et les populations se sont re-mélangées suite à un contact secondaire. Il semblerait toutefois que la majorité des zones d'hybridation se soient formées suite à des contacts secondaires (Barton and Hewitt, 1985). Dans ces zones d'hybridation, la rencontre des deux populations où l'isolement reproductif est partiel peut engendrer deux scénarios dépendant de la vigueur des mécanismes d'isolement déjà mis en place. Dans le cas où les populations ne sont pas fortement isolées, on s'attend à une ré-homogénéisation des fonds génétiques de celles-ci. Dans le cas inverse où l'isolement reproducteur est plus fort, on observe un maintien des deux populations caractérisées par des individus hybrides plus ou moins fertiles selon le degré d'isolement reproducteur des lignées parentales. Il peut donc y avoir des individus dits "introgressés" si les hybrides de première génération F1 peuvent se rétro-croiser avec les lignées parentales, ou entre eux, avec plus ou moins d'efficacité. Ces individus possèdent les allèles issus de leur lignée aux locus d'isolement reproducteur (ou à ceux qui leur sont liés génétiquement ou physiquement) et des allèles des deux lignées aux locus neutres (Bierne et al., 2003). L'hybridation entre deux populations peut avoir des effets positifs comme l'introgression adaptative, qui peut se produire lorsqu'une population possède des allèles avantageux pour l'autre population. Dans ce cas, ces allèles vont facilement traverser la zone hybride et être positivement sélectionnés dans l'autre population par le biais des rétro-croisements (Racimo et al., 2017). Il est également possible que les individus hybrides, hétérozygotes, aient une meilleure valeur sélective que leurs parents, c'est ce qu'on appelle la vigueur hybride. Cependant, l'hybridation peut également avoir des effets négatifs, en ayant un effet délétère sur la valeur sélective des hybrides. On parle alors de dépression d'hybridation. Cet effet délétère peut même conduire à un isolement reproductif complet entre les espèces s'hybridant (Kawecki et Ebert, 2004).

Une manière d'étudier le contexte géographique dans lequel s'est déroulée la spéciation entre deux lignées est d'inférer l'histoire démographique de ces dernières à partir de données génétiques. Cette approche étudie le polymorphisme d'un grand nombre d'individus de deux lignées distinctes afin d'obtenir des informations à la fois sur les aspects temporels et démographiques de leur divergence. Elle permet de comparer des données réelles observées à des données simulées représentant un modèle simplifié du processus de divergence avec ou sans migration et d'évaluer le modèle le plus vraisemblable ayant conduit aux données observées. Les premiers modèles développés avaient pour objectif de distinguer des scénarios simples de spéciation allopatrique et sympatrique grâce au modèle de Strict Isolement sans flux génique (SI) et d'Isolement avec Migration (IM) (Figure 3). Ces méthodes ont permis d'explorer l'histoire évolutive des allèles entre et au sein des populations par coalescence pour les données réelles et celles simulées par les modèles et d'ajuster les différents paramètres des modèles (temps de divergence, taille efficace des populations, intensité des échanges génétiques, etc.) par maximisation de la vraisemblance. Ces modèles se complexifiant pour mieux prendre en compte la temporalité des événements de migration et les effets démographiques, ont permis de distinguer l'effet de la migration de celui du tri des lignées alléliques à partir du polymorphisme ancestral, permettant de discriminer une spéciation en cours par réduction du flux génique (AM) d'un modèle de contact secondaire (SC). Ces modèles restent cependant des simplifications de la réalité même si aujourd'hui il est possible d'estimer si les barrières aux flux de gènes sont hétérogènes ou homogènes le long des génomes autorisant la variation génomique des paramètres de migration et de taille efficace (Gagnaire, 2020 ; Gagnaire et al., 2013).

Les mécanismes d'isolement reproducteur et les contextes géographiques sont nombreux et peuvent intervenir simultanément, ce qui rend complexe l'étude de la spéciation. Dans cette étude, le rôle de la géographie, du déterminisme du sexe et de l'hétérogénéité spatiale des habitats sur certains mécanismes d'isolement conduisant à la spéciation seront étudiés. Notamment, le rôle de la sélection divergente dans l'histoire évolutive des espèces et l'adaptation locale au sein des espèces seront examinés.

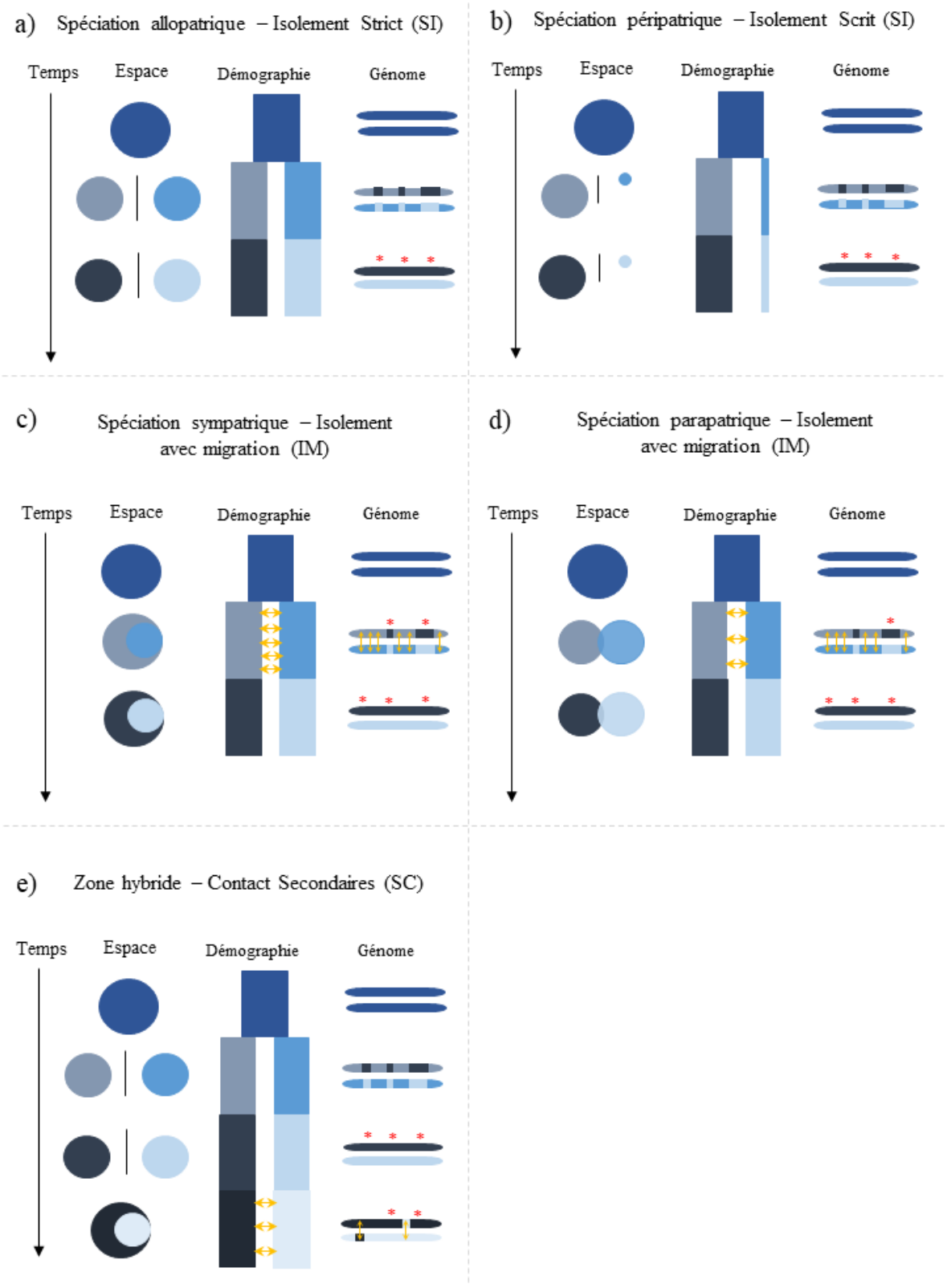


Figure 3: Effets des différents contextes géographiques sur la différenciation des populations et la formation d'espèces. Une population ancestrale se divise en deux populations filles qui accumulent progressivement des différences génétiques et des locus d'isolement reproducteur (étoiles rouges) jusqu'à former deux espèces distinctes. Ces espèces peuvent après un contact secondaire se retrouver

en sympathie. La spéciation peut se dérouler en (a) allopatrie, (b) en péripatrie, (c) en sympatrie, (d) en parapatrie, (e) avec contacts secondaires. Modifié d'après une figure de Maud Duranton (2019). Les flèches jaunes représentent l'intensité du flux de gènes.

1.3- Architecture génomique de la divergence

Une fois le contexte géographique de la divergence étudié, l'objectif est d'identifier les régions génomiques, voire même les locus, impliqués dans l'isolement reproducteur des populations. Ces locus vont réduire localement dans le génome le flux génétique et ainsi permettre à la différenciation génétique (différence de fréquence allélique) de s'accumuler entre les lignées. Pour identifier ces régions, des études utilisent le F_{ST} de Wright (1951) qui permet de mesurer le niveau de différenciation génétique entre deux sous-populations. Il est défini comme $F_{ST} = (\pi_{total} - \pi_{intra\text{-}population}) / \pi_{total}$ où π correspond à la diversité génétique mesurée pour chaque population ($\pi_{intra\text{-}population}$) ou pour l'ensemble des individus (π_{total}).

A l'échelle des génomes complets, il a pu être montré la présence de patrons de différenciation génétique hétérogènes le long des génomes (Nosil et al., 2009 ; Gagnaire et al., 2013). En effet, tant que l'isolement reproducteur et géographique sont incomplets, alors le génome est perméable aux flux de gènes (Wu, 2001). Ces patrons sont caractérisés par la présence de régions génomiques faiblement différenciées qui sont entre-coupées par des régions où la différenciation est forte, appelées îlots génomiques de différenciation (Harr, 2006). Ces îlots à forte différenciation génomique sont des marqueurs de la présence de locus d'isolement reproducteur et sont donc nommés des îlots de spéciation (Turner et al., 2005). Les îlots de différenciation peuvent se former progressivement lorsque deux populations divergent puis s'adaptent à deux environnements différents, ou lors d'un contact secondaire. En effet, après une période en allopatrie où les populations vont accumuler des locus impliqués dans l'isolement reproducteur, lors de la reprise du flux de gènes les échanges vont permettre d'éroder la différenciation précédemment accumulée exceptée à ces locus impliqués dans l'isolement reproducteur qui agissent comme des barrières à l'introgression (Barton et Bengtsson, 1986).

1.4- Mécanismes permettant l'émergence de la divergence génétique entre populations

La variation génétique est due aux mutations spontanées (et à la recombinaison). Le devenir des variants créés par ces mutations (c'est-à-dire leur fixation ou leur élimination) va

dépendre des forces évolutives qui en font évoluer la fréquence dans les populations : la dérive génétique, la migration et la sélection naturelle. La théorie de l'évolution, qui découle directement des travaux menés par Charles Darwin, a tout d'abord proposé que la sélection naturelle est la force dominante. Ainsi, les nouvelles mutations génétiques ont soit un effet positif soit négatif sur la valeur sélective des individus. La valeur sélective étant la capacité d'un individu à survivre et à se reproduire dans un environnement donné, le devenir d'une mutation est dépendant de l'environnement dans lequel elle se trouve, ce qui entraîne des adaptations ou maladaptations environnementales. Dans ces conditions, les mutations délétères sont négativement sélectionnées et donc vont avoir tendance à disparaître avec le temps, alors qu'à l'inverse, les mutations favorables vont être sélectionnées et augmenter en fréquence avec le temps dans la population (Huxley, 1942). Ainsi, au cours des générations, seules les mutations positives vont s'accumuler et conditionner la partition de la variation génétique en fonction des habitats : ceci représente le processus d'adaptation.

L'effet de la sélection naturelle sur un gène donné peut être étudié en se basant sur la divergence des régions codantes entre espèces et/ou individus isolés en comparant le nombre de changements synonymes et non-synonymes sur celles-ci (Nei et Gojobori, 1986). En effet, lorsqu'une mutation apparaît dans une séquence codante d'ADN, soit elle entraîne un changement dans la séquence en acides aminés d'une protéine, soit la séquence protéique reste identique en raison de la redondance du code génétique, ce code ayant été lui-même optimisé au cours du temps par la sélection naturelle (Spencer et Barral, 2012). Si la mutation n'entraîne pas de changement d'acide aminé on parle de mutation synonyme, dans le cas contraire on parle de mutation non-synonyme. Les changements synonymes sont considérés en théorie comme neutres alors que les changements non-synonymes peuvent être avantageux, neutres ou délétères et sont donc soumis à l'action de la sélection naturelle. Environ un tiers des mutations non-synonymes sont considérées comme neutre, presque deux tiers ont un effet délétère, et seulement 2 à 3% des mutations sont considérées comme avantageuses (Eyre-Walker et al., 2002 ; Wright et al., 2005). Ainsi, pour déterminer si un gène subit des pressions de sélection positive (diversifiante) ou négative (purifiante), on peut estimer la divergence entre deux lignées distinctes en estimant le taux de substitutions synonymes (d_S) et non-synonymes (d_N) dans leur compartiment respectif (i.e. nombre total de sites potentiellement synonymes *S* et non-synonymes *N*). Le ratio d_N/d_S permet d'observer s'il y a eu plus de remplacements non-synonymes que de remplacements synonymes. Si aucune pression de sélection n'agit sur le gène alors ce ratio sera égal à 1, si le gène est sous

sélection purifiante le d_N/d_S sera inférieur à 1 (plus de mutations synonymes que de mutations non-synonymes) et si le gène est sous sélection positive, le d_N/d_S sera supérieur à 1 (plus de mutation non-synonymes que de mutations synonymes). Il convient cependant de nuancer ce propos en précisant que les protéines étant par nature sous sélection purifiante ($d_N/d_S<1$) pour préserver leur fonction et structure 3D, la détection de sélection positive à plusieurs sites (codons) doit être suffisamment élevée pour contre-balancer la sélection négative moyenne sur les autres sites, et donc souvent difficile à évaluer correctement.

Une théorie opposée, ou complémentaire, à celle de l'évolution des espèces par la sélection naturelle est proposée par M. Kimura avec la théorie neutre de l'évolution. Cette théorie repose sur le fait que la majorité des nouvelles mutations génétiques sont neutres, c'est-à-dire qu'elles n'ont pas d'effets sur la valeur sélective d'un individu. L'évolution des espèces dépend alors essentiellement de la dérive génétique (Kimura, 1983). La dérive est un processus aléatoire qui va permettre ou non la fixation d'une nouvelle mutation dans la population. Cependant, l'effet de la dérive génétique est proportionnel à la taille efficace (N_e) de la population de l'espèce étudiée. En effet, plus une population va être petite (taille efficace faible), plus la dérive génétique sera forte, et le temps de fixation moyen d'un allèle nouvellement apparu (à la fréquence 1/2N) est d'environ $4N_e$ générations (Kimura et Ohta, 1969).

Ces deux théories ont été aujourd'hui validées chez bon nombre d'espèces. En effet la théorie neutre ne remet bien sûr pas en cause l'existence de la sélection naturelle, mais suppose que les mutations neutres et délétères doivent être plus fréquentes que les mutations avantageuses qui se trouvent dans une région codante. Il est néanmoins important de comprendre que plus la taille efficace d'une espèce sera grande et plus la probabilité d'apparition d'un variant avantageux le sera. Dans ce cas la fixation de ce variant sera beaucoup plus rapide qu'attendu sous les seuls effets de la dérive, avec un temps de fixation dépendant principalement du coefficient de sélection appliqué au variant dans l'habitat ou il est favorisé.

2- Le milieu hydrothermal : un environnement singulier pour les études de spéciation

La zone aphotique (inférieure à 1000m de fond) a longtemps été considérée comme un désert de vie (Anderson et Rice, 2006). En effet, il est difficile de comprendre comment des organismes peuvent vivre dans des conditions si particulières notamment en l'absence de lumière et sous l'action de fortes pressions hydrostatiques. Entre 1872 et 1876, l'expédition du Challenger a cependant permis la découverte d'espèces abyssales lors de chalutages profonds. Plus tard, avec la mise au point des bathyscaphes, des hommes ont pu explorer ces milieux profonds. C'est notamment le cas d'Auguste Piccard, qui à bord du bathyscaphe Trieste est descendu à 10 916 mètres dans la fosse des Mariannes et a observé la présence d'échinodermes sur le fond (1960). Malgré les recherches menées sur le milieu profond, les zones bathyale et abyssale situées entre 1000 et 6000 mètres de fond ont longtemps été considérées comme des zones à forte biodiversité mais faible biomasse, en raison des faibles apports en matière organique sédimentant de la surface (Gage et Tyler, 1991 ; Laubier, 1992). Ainsi, la découverte des sources hydrothermales présentées comme de véritables oasis de vie ont été une véritable avancée pour le milieu profond. Depuis, d'autres environnements eux-aussi basés sur une production primaire chimio-autotrophe comme les zones de suintements froids, d'hydrocarbures, d'agrégations de bois coulés ou de carcasses de baleines ont permis de montrer que l'océan profond n'est pas aussi vide qu'il n'y paraissait (Alfaro-Lucas et al., 2018 ; Ockelmann et Dinesen, 2011 ; Von Cosel et Olu, 1998).

2.1- Découverte et formation

Les communautés hydrothermales ont été découvertes aux Galapagos en 1976 lors d'une expédition géologique par le submersible américain Alvin recherchant des signes d'activité hydrothermale sur l'axe des dorsales océaniques (Lonsdale, 1977). A cette époque, les géologues remarquent qu'un fluide, chargé en minéraux, en méthane et en sulfure d'hydrogène, s'échappe par les orifices des monticules présents. Lors de cette expédition, les biologistes ont également découvert un écosystème riche et varié complètement nouveau (Williams et al., 1979). Par la suite, les nombreuses campagnes qui se sont succédées le long des dorsales océaniques ceinturant le globe ont permis de montrer que l'environnement hydrothermal ou les zones de suintements froids était présent sur pratiquement toutes les zones tectoniquement actives (Boulart et al., 2022 ; Desbruyeres, 1982 ; Hessler et al., 1988 ;

Jollivet et al., 1989 ; Langmuir et al., 1997 ; Rogers et al., 2012 ; Segonzac, 1992 ; Tunnicliffe et al., 1986 ; Van Dover et al., 2001).

Les sources hydrothermales sont associées à la formation des dorsales océaniques, lesquelles produisent la nouvelle croûte océanique qui compense l'éloignement des plaques océaniques au cours du temps. L'écartement de ces plaques va entraîner la formation de la nouvelle croûte océanique par épanchement du magma en surface (volcanisme effusif) et sa fracturation tectonique en failles et anfractuosités par où de l'eau de mer froide et oxygénée va pouvoir s'infiltrer en profondeur (Figure 4). En pénétrant en profondeur, l'eau va monter en pression et en température par proximité avec la chambre magmatique sous-jacente. La circulation rapide du liquide va induire un lessivage des roches chaudes (150 – 200°C) environnantes ce qui va également entraîner une modification de la chimie de l'eau de mer pour la transformer en un fluide chaud. En effet, elle va se libérer des ions calcium, sulfate et magnésium par précipitation de sulfate de calcium et de minéraux riches en magnésium. Au contact des roches, le fluide va également s'appauvrir en oxygène et se charger en éléments réduits (fer ferreux, sulfures, méthane) et en métaux (manganèse, silicium, cuivre, zinc ; Lalou, 1991). Le fluide moins dense et chaud va alors pouvoir remonter et jaillir par les interstices présents dans le plancher océanique. Les éléments minéraux contenus dans le fluide vont précipiter et créer des dépôts minéraux au contact de l'eau de mer froide. Ces édifices sont de formes variées et dépendent de l'intensité du mélange du fluide et de l'eau de mer ainsi que de la composition des roches sous-jacentes. Les fumeurs noirs, qui sont les émissions les plus chaudes (350°C), anoxiques et acides sont issus de la précipitation d'un fluide non dilué en sulfures polymétalliques et en sulfate de calcium (Hannington et al., 1995). Les fumeurs blancs sont issus d'une précipitation initiée avant l'émission en sub-surface avec une séparation de phases entre gaz et éléments métalliques. Il y a donc une dilution du fluide hydrothermal par de l'eau de mer au sein de réseaux de crevasses avec une précipitation d'une bonne partie des métaux. Comme cette précipitation est initiée en profondeur, les fumeurs blancs sont des édifices formés par des fluides appauvris en sulfures polymétalliques et en sulfate de calcium mais riche en silicates, baryum et anhydrite (Hannington et al., 1995). Des formations minérales dites de diffusion existent également. Elles sont présentes lorsque l'eau de mer pénètre moins profondément dans la croûte océanique ou lorsque de l'eau de mer se mélange en sub-surface avec les fluides hydrothermaux. On obtient donc un fluide plus pauvre en métaux et en sulfure qui va diffuser lentement par des failles ou par des structures minérales poreuses à des températures inférieures à 50°C (Hannington et al., 1995).

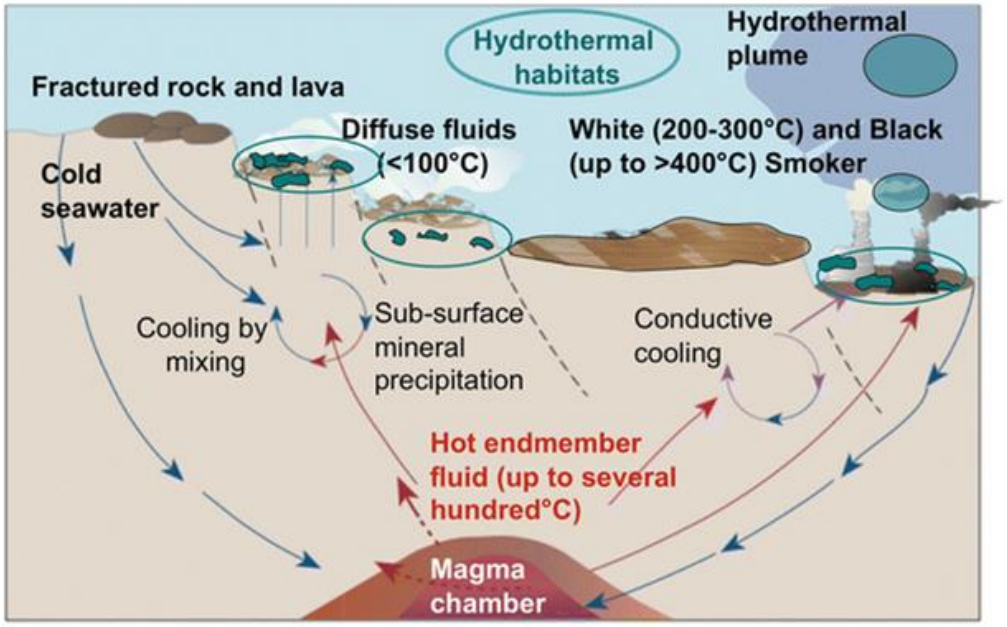


Figure 4: Schéma simplifié montrant l'évolution des fluides hydrothermaux dans un site géologiquement actif (Hein et Mizell, 2013).

2.2- Caractéristiques de l'environnement hydrothermal et effets sur les populations

L'environnement hydrothermal se distribue sur toutes les dorsales océaniques de la planète (Van Dover, 2000). Les champs hydrothermaux sont formés de différents sites composés de cheminées et de zones d'émissions diffuses. Ces sites sont regroupés autour de zones ou l'activité est plus intense selon la proximité de la chambre magmatique, le niveau de fissuration du plancher basaltique et la proximité de failles transformantes qui décalent le rift (Watremez et Kervevan, 1990). Ces zones d'activité peuvent être très éloignées les unes des autres, ce qui explique que l'habitat hydrothermal soit fragmenté. Si les effets de la fragmentation ne sont pas contrés par une capacité de dispersion larvaire massive des populations, alors elle conduit à une augmentation de la différentiation génétique entre les populations locales (effet de dérive). Lorsque l'isolement géographique des populations est complet ou limite fortement le flux génique il peut aboutir à la formation d'espèces au cours du temps. La spéciation allopatrique semble être courante dans l'environnement hydrothermal (Wilson et Hessler, 1987). C'est par exemple le cas chez l'amphipode *Ventiella sulfuris*, où l'espèce trouvée au Galapagos est actuellement différente de celle de l'EPR (France et al., 1992). La présence d'un grand nombre d'espèces cryptiques est également favorisée par des événements de fondation en allopatrie chez les espèces des sources hydrothermales qui

dispersent de proche en proche ou sur des événements rares de dispersion longue distance (Borda et al., 2013 ; Kojima et al., 2001).

Les sources hydrothermales sont non seulement des environnements fragmentés mais aussi éphémères, car le mouvement permanent des plaques entraîne une dynamique d'extinction/création de cheminées. En général, la durée de vie des sites hydrothermaux n'excède pas 100 ans sur les dorsales à taux d'accrétion rapide (e.g. EPR ; Lalou et al., 1995) et plusieurs milliers d'années sur les dorsales lentes (e.g. MAPR ; Baker et al., 2001).

Cette dynamique peut causer la disparition d'une population si celle-ci n'est pas suffisamment dispersante ou permettre la remise en contact de deux populations initialement isolées. Cette remise en contact peut entraîner un brassage génétique des populations ou à l'inverse s'il y a un isolement reproductif permettre la présence de deux espèces cryptiques en sympatrie (Jollivet et al., 1999).

Même si les sources hydrothermales existent depuis la naissance des océans et semblent donc être un environnement pérenne à l'échelle de la dorsale elle-même avec des champs d'activité pouvant perdurer sur des centaines de milliers d'années, les conditions thermo-chimiques rencontrées sont variables et très hétérogènes à l'échelle locale, voire régionale (Du Preez et Fisher, 2018). En effet, ces conditions varient de façon chaotique selon le mélange du fluide hydrothermal (acide pH=2, anoxique, riche en CO_2, H_2S, H_2 et CH_4 ; Le Bris et al., 2003 ; Von Damm, 1990) et de l'eau de fond, froide et bien oxygénée mais également selon la nature des roches sous-jacentes et les processus de séparation de phase en sub-surface. Plus un individu va être proche de l'émission, plus les conditions de vie seront extrêmes (température élevée, manque d'oxygène, forte concentration en CO_2, H_2S, H_2 et CH_4). Il existe donc un gradient qui conditionne la distribution locale des espèces dans l'espace selon leur niveau d'adaptation à la toxicité du fluide (Desbruyeres et Laubier, 1982 ; Figure 5) et dans le temps (atténuation progressive de l'émission de sa naissance à son extinction) ce qui fournit, outre des successions temporelles d'espèces au cours de la vie d'un site, un habitat propice à la sélection disruptive entre individus d'une même population grâce à la formation d'une multitude de niches écologiques.

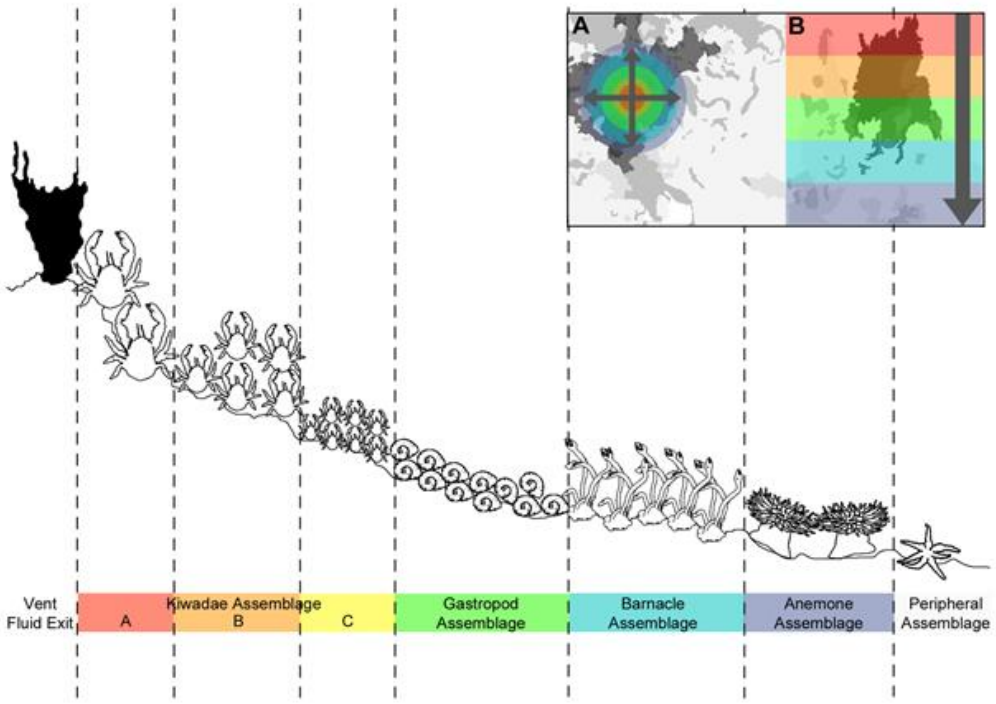


Figure 5: Exemple de distribution spatiale des communautés hydrothermales dans l'océan Antarctique (Marsh et al., 2012).

De plus, des adaptations spécifiques ont rendu la vie possible dans ces milieux extrêmes. Notamment au niveau du système respiratoire (branchies hypertrophiées, pigments respiratoires très diversifiés ; Hourdez et Lallier, 2007 ; Sell, 2000), chaîne respiratoire mitochondriale adaptée à la haute température couplée à la détoxification du sulfure d'hydrogène (Powell et Somero, 1986 ; Zierenberg et al., 2000), arsenal spécifique pour faire face aux contraintes chimiques et thermiques (stress oxydatif : Dilly et al., 2012), mode de reproduction avec la mise en place de formes dispersantes sur de longue durée (Arellano et al., 2014 ; Marsh et al., 2001) et des systèmes immunitaires évoluant avec les symbiontes bactériens (Tasiemski et al., 2014). Cette dernière adaptation a permis à certains organismes de contrôler la chimioautotrophie *via* la symbiose avec des bactéries (sulfooxydantes ou méthanotrophes) vivant dans le milieu hydrothermal (Papot et al., 2017). Au vu de l'hétérogénéité spatiale des conditions thermo-chimiques, l'étude des différents types d'adaptation (système immunitaire, système respiratoire, système de détoxification etc.) est donc importante pour comprendre le rôle de l'adaptation disruptive dans la spéciation des espèces hydrothermales.

L'environnement hydrothermal, de par sa nature fragmentée, éphémère, écologiquement hétérogène et extrême, offre donc des conditions particulières pour l'étude de la spéciation puisque ces caractéristiques géographiques et écologiques ont pu jouer un rôle clef dans la spéciation et la distribution actuelle des espèces.

En ce qui concerne la biologie de ces milieux, les espèces sont encore aujourd'hui mal décrites. Il y a donc de nombreuses inconnues quant au rôle de l'hétérogénéité du milieu hydrothermal (géographie, environnement) sur la distribution et les mécanismes de spéciation des espèces qui y sont inféodées. L'environnement hydrothermal, malgré son intérêt, est un milieu difficile d'accès. En effet, la recherche en milieu profond empêche toute manipulation directe, ce qui entraîne l'utilisation, lorsque les conditions de mer le permettent, de robots téléguidés. L'accessibilité restreinte sur les différents sites entraîne un nombre limité de prélèvements journaliers. De plus, malgré l'intensification des études sur les sources hydrothermales, la distribution des sites reste mal connue (inventaire incomplet des zones actives), et la nature éphémère de cet environnement rend leur localisation peu fiable à long terme.

2.3- Les bassins arrière-arc

Les sources hydrothermales du Pacifique Ouest sont surtout observées au niveau des dorsales ou rides qui ont conduit à l'ouverture et l'extension des bassins arrière-arc (Hall, 2002) mais elles peuvent également être trouvées dans les calderas de volcans hors axe en arrière des zones de subduction (cf Nifonea ou Kulo lasi : Fouquet et al., 2018 ; McConachy et al., 2005). Les bassins arrière-arc du Pacifique Ouest se sont formés à la suite de mécanismes de subduction de la plaque Pacifique sous la plaque Indo-Australienne. Cette subduction provoque une fonte partielle du manteau de la plaque sous-jacente en magma qui va remonter à la surface en créant un arc volcanique. Cette activité a également pour conséquence la mise en place d'une cellule de convection magmatique en arrière de l'arc volcanique. La montée du magma va provoquer la formation d'un rift qui va créer une accrétion de croûte océanique pour compenser les effets de subduction inter-plaques. Par conséquent, ce rift permet de compenser, par création d'une nouvelle croûte, la partie de la plaque qui est subductée de part et d'autre, ce qui va conduire à la formation d'un bassin (Figure 6).

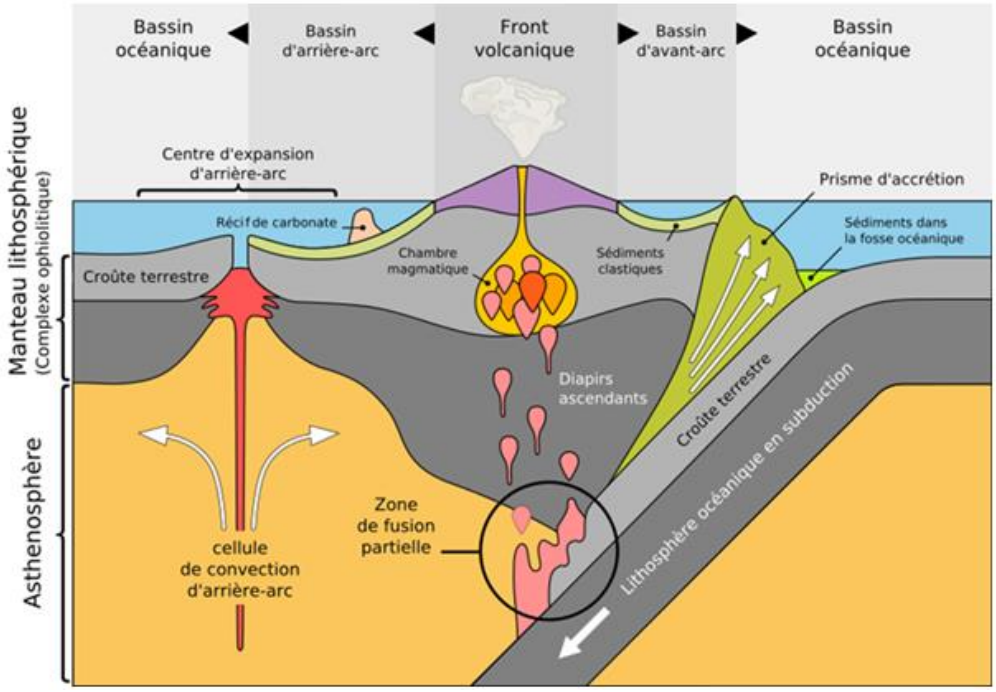


Figure 6: Formation des bassins arrière-arc (Extrait de https://fr.wikipedia.org/wiki/Bassin_arrière-arc)

Les bassins arrière-arc du Pacifique Ouest sont formés de quatre bassins. Ils se répartissent du Nord au Sud comme suit : le bassin de Manus, la ride de Woodlark, le bassin Nord-Fidjien et le bassin de Lau (Figure 7).

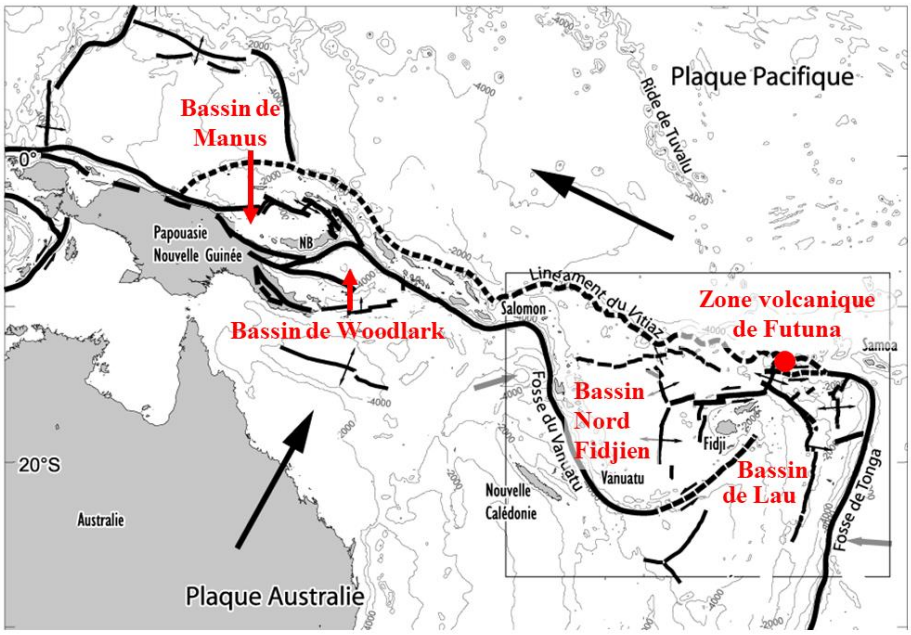


Figure 7: Bassins arrière-arc présents dans le Pacifique Sud-Ouest. La croix rouge symbolise le point triple observé dans le bassin Nord-Fidjien. Modifié de Hall et al. (2002).

Le bassin de Manus, situé en mer de Bismarck en Papouasie Nouvelle Guinée, est limité au Nord par la fosse de subduction fossile de Manus, et au Sud par la fosse active de Nouvelle Bretagne. L'ensemble des terres émergées entourant le bassin (îles de la Nouvelle Bretagne et la Nouvelle Irlande) constitue un ancien arc volcanique tertiaire. L'ouverture de ce bassin est estimée entre 3,5 et 4 Ma et résulte du démantèlement de l'arc volcanique situé à la frontière convergente des plaques pacifique et australienne (Hall et al., 2002). Le bassin de Woodlark s'étend quant-à-lui de la pointe orientale de la Nouvelle Guinée à la Nouvelle Géorgie en passant par les îles Salomon. Son ouverture est estimée à 6-7 Ma (Goodliffe et al., 1997 ; Hall et al., 2002) sous l'effet d'un rift continental, c'est-à-dire d'un amincissement d'une plaque tectonique sous l'effet de forces d'étirement associées à l'arc de subduction située le long des îles Salomon. Le bassin Nord-Fidjien est de forme triangulaire et est délimité au Nord par l'ancienne zone de subduction Vityaz, à l'Ouest par l'arc des Vanuatu, au Sud par la fracture de Matthew-Hunter et à l'Est par les îles Fidji. Ce bassin est apparu il y a moins de 10 Ma et provient de la rotation simultanée de l'arc des Vanuatu dans le sens horaire et des îles Fidjiennes dans le sens anti-horaire. Ce bassin est caractérisé par la présence d'un point triple situé au centre de celui-ci, caractérisant une région où trois fosses tectoniques se rejoignent (croix rouge sur la Figure 7). Seule la dorsale au Nord du point triple apparaît encore active en se propageant vers le Nord depuis 1-2 Ma (Auzende et al., 1994) au moment

de l'ouverture du bassin de Lau. L'expansion du plancher océanique au centre de ce bassin a certainement été le déclencheur de l'ouverture du bassin de Lau il y a 1-6 Ma (Auzende et al., 1988 ; Hall, 2002 ; Lafoy, 1989). Le bassin de Lau a une forme de trapèze et est limité au Nord par la zone de fracture Nord Fidjienne, cette forme particulière suggère une ouverture du Nord vers le Sud.

3- Le genre *Alviniconcha* comme modèle d'étude

La première espèce du genre *Alviniconcha* (Provannidae, Caenogastropoda), *Alviniconcha hessleri* (Figure 8), a été décrite en 1988 à partir d'individus récoltés au niveau des sites hydrothermaux actifs dans le bassin arrière-arc des Mariannes (Pacifique Nord-Ouest) à 3650 mètres de fond (Okutani et Ohta, 1988). Cette espèce qui mesure au maximum 5 cm présente une coquille plutôt élastique et peu calcifiée, ornementée de rangées de "poils" périostracaux régulièrement espacés. Les rangées principales composées de « longs poils » sont intercalées avec de nombreuses rangées secondaires présentant des poils plus courts. Les individus de *A. hessleri* présentent une columelle plutôt mince et tronquée à l'extrémité pour former un canal siphonal peu profond. La radula est de type taenioglosse puisqu'elle présente des rangées de 7 dents avec une dent centrale entourée de chaque côté par une dent latérale et deux dents marginales (2 : 1 : 1 : 1 : 2). Cette radula est typique des Caenogastropodes et associée à un comportement de broutage. Cette première description précise également que l'apex de la coquille n'est jamais présent chez les individus adultes récoltés (Okutani et Ohta, 1988).

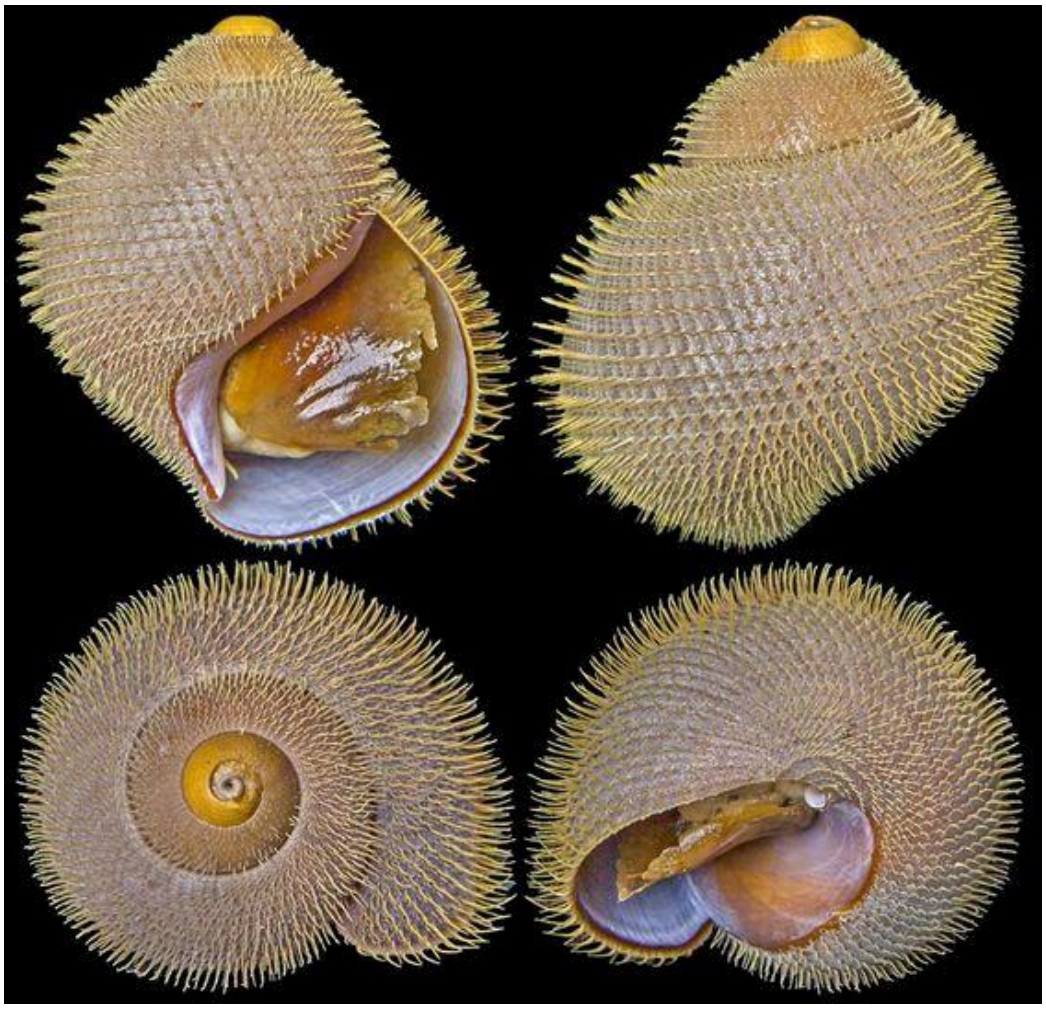


Figure 8: *Alviniconcha hessleri* (© Chong Chen's)

En 1993, des individus assimilés à *Alviniconcha hessleri* sont échantillonnés pour la première fois dans les bassins Nord-Fidjien et Lau (Warèn et Bouchet, 1993). Cette étude, qui complète celle réalisée en 1988, fait une description détaillée des stades larvaires mais également de l'anatomie de l'espèce. Elle permet notamment grâce à la forme du pied, de la radula et de l'estomac de classer des *Alviniconcha* dans la famille des Provannidae (Warèn et Bouchet, 1993). Elle permet également, au vu de la grosseur de la branchie et *a contrario* de la faible taille de l'estomac, d'affirmer que ces organismes vivent en symbiose avec des bactéries chimioautotrophes (Fiala-Médioni, 1984). Ces deux premières études semblent montrer que l'espèce *Alviniconcha hessleri* occupe une large aire de répartition sur les 3 bassins échantillonnés. Cependant, la même année, une étude génétique basée sur des allozymes révèle la présence de deux formes d'*Alviniconcha* spp. génétiquement distinctes

dans les bassins Nord-Fidjien et Lau. Ces formes semblent être génétiquement isolées entre les deux bassins arrière-arc (Denis et al., 1993).

En 2001, une première étude de séquences partielles du gène mitochondrial *Cox1* montre la présence de trois lignées génétiquement distinctes au sein de l'espèce *A. hessleri*, mais ces différentes formes semblent morphologiquement indifférenciables (Kojima et al., 2001). En l'absence de critères morphologiques fiables pour identifier les espèces, les études génétiques se sont ensuite multipliées. Les études les plus récentes, basées sur le gène mitochondrial *Cox1* et quelques gènes nucléaires, montrent la présence de six espèces distinctes (Breusing et al., 2020 ; Johnson et al., 2015 ; Figure 9) sans que le critère d'isolement reproducteur soit vérifié. Ces espèces géographiques définies sur le critère phylogénétique de monophylies réciproques de plusieurs gènes mitochondriaux et nucléaires conservés se nomment *A. kojimai*, *A. strummeri*, *A. boucheti* (présentes dans les bassins de Lau, Manus et Nord-Fidjien : espèces étudiées durant cette thèse), *A. hessleri* (bassin des Mariannes), *A. adamantis* (ride des Mariannes) et *A. marisindica* (dorsale centrale indienne) (Figure 10). En supposant que le taux d'évolution du gène *Cox1* soit resté constamment faible (environ 0,0015 substitutions par site et par million d'années sur la base d'une calibration fossile), la scission entre les deux principaux clades d'*Alviniconcha* (séparant *A. boucheti* et *A. marisindica* de *A. kojimai*, *A. hessleri* et *A. strummeri*) est ancienne et estimée à 38 Ma (Breusing et al., 2020) (Figure 11). Dans le clade comportant le plus d'espèces, la séparation entre *A. strummeri* et *A. kojimai*/*A. hessleri* est plus récente et datée à environ 25 Ma (Breusing et al., 2020).

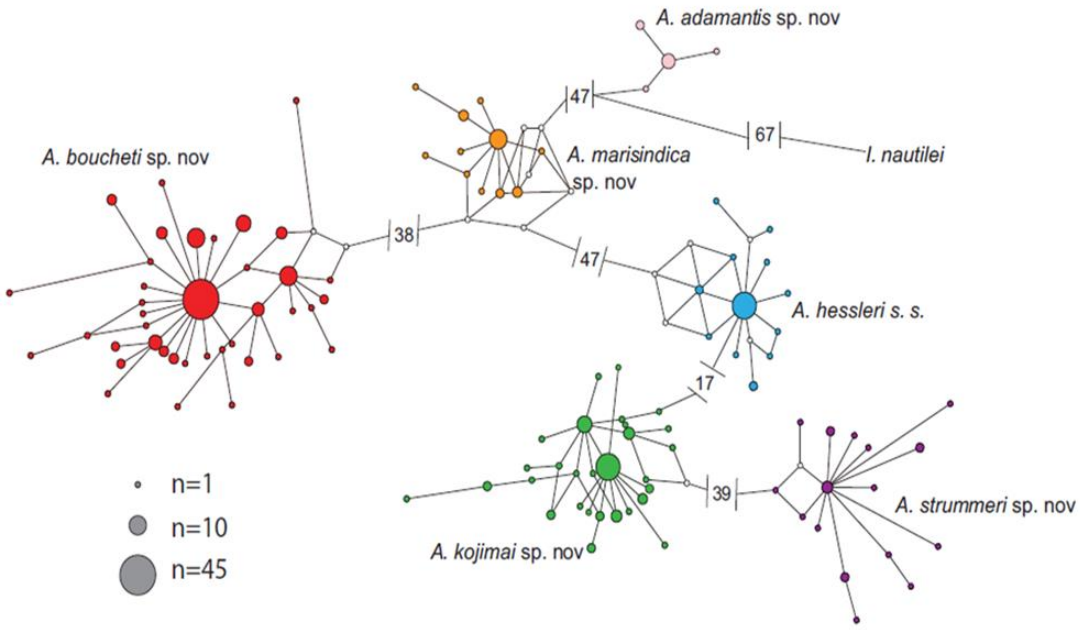


Figure 9: Réseau d'haplotypes à partir du gène mitochondrial *Cox1* du genre *Alviniconcha* enraciné avec *Ifremeria nautilei*. La taille des cercles indique les fréquences alléliques relatives. Les cercles blancs représentent les haplotypes manquants. Les segments représentent les divergences observées entre haplotypes et le nombre indiqué sur les segments représente le nombre de mutations entre deux espèces. Issus de Johnson et al. (2015).

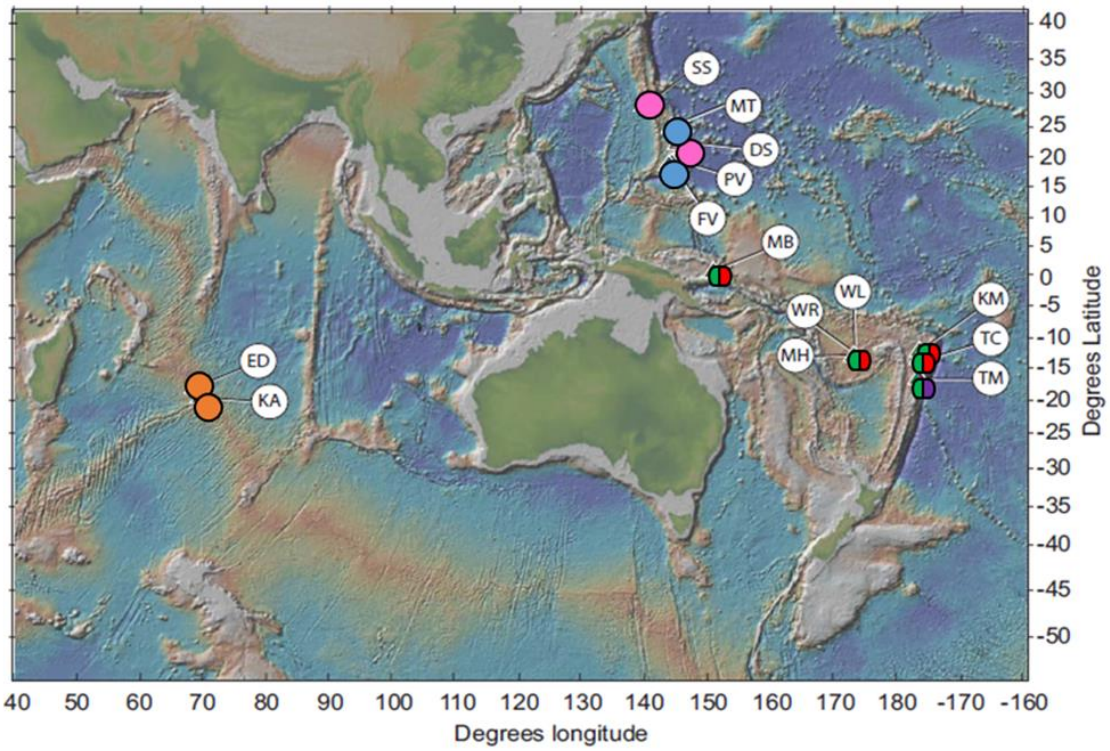


Figure 10: Répartition des différentes espèces de *Alviniconcha* dans le Pacifique Ouest et l'océan Indien. En orange : *A. marisindica* ; violet : *A. strummeri* ; vert : *A. kojimai* ; rouge : *A. boucheti* ; rose : *A. adamantis* ; bleu : *A. hessleri*. D'après Johnson et al., 2015.

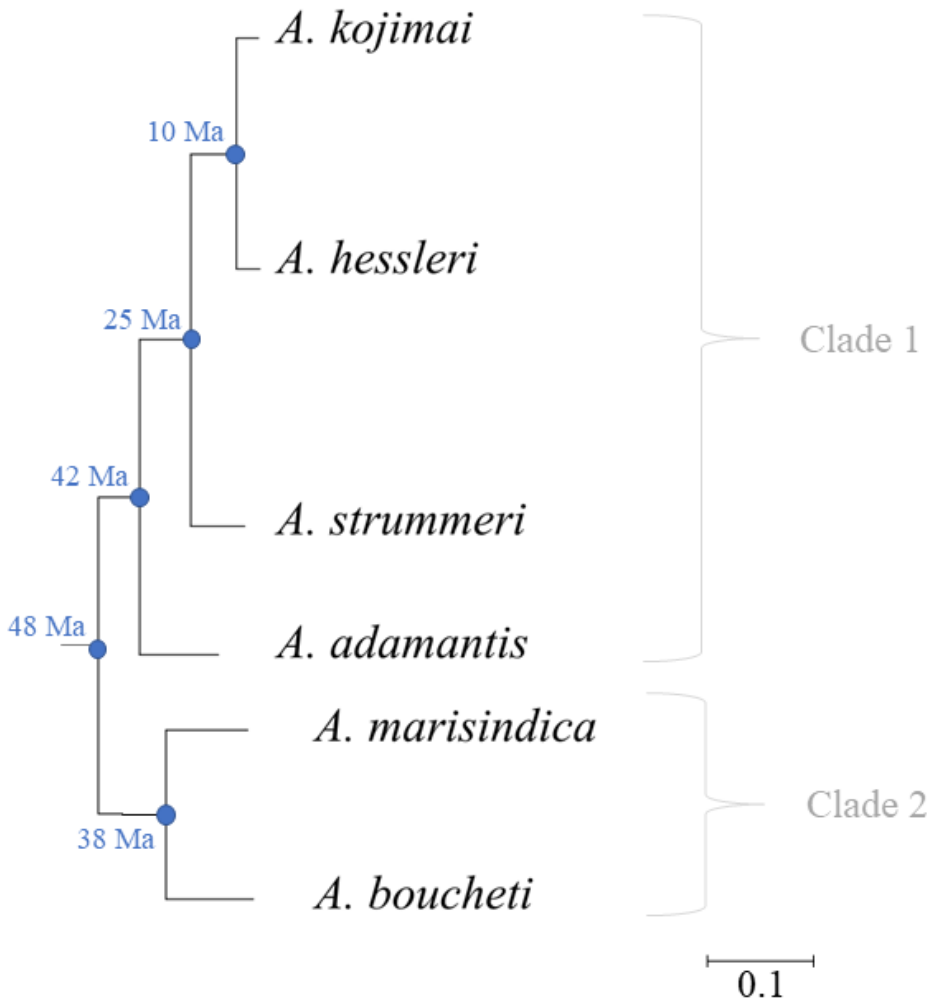


Figure 11: Arbre phylogénétique simplifié avec l'estimation de la période de séparation des espèces d'*Alviniconcha*, d'après Breusing et al. (2020).

Jusqu'à très récemment, aucune étude n'avait cherché à trouver des caractères morphologiques permettant de discriminer les espèces du genre *Alviniconcha*, à l'exception de celle de Laming et al. parue en 2020 après l'expédition CHUBACARC (expédition qui a fourni l'essentiel des échantillons utilisés pour ce travail de thèse, voir ci-dessous). Dans cette étude basée uniquement sur des individus récoltés dans la zone volcanique active de Futuna, les auteurs ont pu montrer des différences morphologiques entre trois des six espèces d'*Alviniconcha*, notamment au niveau de la forme de la radula, de la columelle, de la couleur de la coquille, de la forme du pied ainsi que de l'ornementation de la coquille. Ils ont donc montré que les espèces de la zone active de Futuna génétiquement apparentées à *A. boucheti, A. strummeri et A. kojimai* ne sont pas réellement des espèces cryptiques (Laming et al., 2020).

Depuis plusieurs années, un certain nombre d'études portent sur la relation hôte/symbiontes chez les *Alviniconcha* spp. Ces gastéropodes sont des consommateurs primaires qui présentent une endosymbiose utilisant l'énergie d'oxydation du sulfure d'hydrogène comme base de la synthèse carbonée (Urakawa et al., 2005 ; Warèn et Bouchet, 1993). Les bactéries majoritaires chez ces organismes sont des *Campylobacteria* et plusieurs phylotypes de *Gamma-proteobacteria* (Beinart et al., 2012, 2014 ; Denis et al., 1993 ; Johnson et al., 2015 ; Kojima et al., 2001 ; Suzuki et al., 2005, 2006). Cependant ces phylotypes bactériens varient en proportion suivant les sites d'échantillonnage, l'espèce et même en fonction des individus (Beinart et al., 2012). En effet, les individus de *A. boucheti* présentent majoritairement des symbiontes faisant partie des *Campylobacteria* alors que les individus de *A. kojimai* et de *A. strummeri* eux présentent des *Gamma-proteobacteria* comme phylotypes majoritaires (Beinart et al., 2012). L'étude la plus récente sur la relation hôte/symbiontes chez les *Alviniconcha* spp. est basée sur le séquençage des génomes complets de chaque type bactérien dominant chez les trois espèces d'*Alviniconcha*, *A. strummeri*, *A. kojimai* et *A. boucheti* (Beinart et al., 2019). Après annotation des génomes, malgré l'éloignement taxonomique des symbiontes séquencés (*Gamma-proteobacteria* et *Campylobacteria*), tous les génomes bactériens présentent des gènes permettant la fixation du carbone, l'oxydation du sulfure et de l'hydrogène, ainsi que des gènes participant à la respiration aérobique (en présence d'oxygène) et anaérobique (nitrate et diméthylsulfure). Ainsi des fonctions métaboliques comparables sont présentes au sein des différents symbiontes (Beinart et al., 2019). Cependant, les *Gamma-proteobacteria* et les *Campylobacteria* utilisent une voie de fixation du carbone inorganique très différente : les *Gamma-proteobacteria* utilisent le cycle de Calvin-Benson-Bassham (CBB) tandis que les *Campylobacteria* utilisent le cycle de l'acide tricarboxylique (rTCA) pour la synthèse de la matière organique. Des différences de gènes au niveau de la mobilité, de l'adhésion, de l'excrétion et des sécrétions sont également notables (Beinart et al., 2019).

Au niveau écologique, les *Alviniconcha* spp. forment un complexe d'espèces architectes qui occupent le pôle chaud (7-42°C), soufré (250 µM) et peu oxygéné (< 50 µM) de l'environnement hydrothermal, indiquant que ces animaux ont une tolérance plus forte à la température et au sulfure d'hydrogène que les autres espèces des communautés du Pacifique Ouest, exception faite des *Alvinellidae* (Podowski et al., 2009, 2010). Ceci est confirmé par Henry et al. (2008), qui ont montré en laboratoire une survie de 19 heures en présence de températures comprises entre 35°C et 40°C. Les *Alviniconcha* spp. se retrouvent

aussi bien sur un substrat composé de basalte ou d'andésite où ils forment des agrégations de petite taille (0.25-1 m^2) et partagent souvent leur habitat avec d'autres espèces de gastéropodes (majoritairement *Ifremeria nautilei*) et de bivalves (*Bathymodiolus brevior)* (Desbruyères et al., 1994 ; Podowski et al., 2009 ; Sen et al., 2014) (Figure 12). Ils peuvent également se trouver autant sur des zones de diffusion que sur les hauteurs des cheminées hydrothermales. Les trois espèces, bien que retrouvées localement en mélange (e.g. site Tui Malila, bassin de Lau : Beinart et al., 2012) semblent cependant résider dans des environnements pouvant être légèrement différents. En effet, *A. boucheti* semble vivre dans un environnement plus chaud, plus riche en hydrogène et en sulfure que *A. strummeri* et *A. kojimai* (Beinart et al., 2012). Cette différence de niche écologique entre les espèces pourrait s'expliquer par les différences dans la tolérance des symbiontes aux conditions de milieu, d'autant plus si les symbiontes sont acquis horizontalement (Beinart et al., 2012).

Une étude de Welschmeyer (2009) comparant les stratégies reproductives entre *A. hessleri* et *I. nautilei* a émis l'hypothèse que les *Alviniconcha* possèdent des petites larves planctotrophes. Cette hypothèse a été confirmée par certains travaux ayant retrouvé des larves de *A. marisindica* et probablement de *A. hessleri* en surface (0-300m) dans l'océan Indien et proche de Hawaï respectivement (Kim et al., 2022 ; Sommer et al., 2017). Ces récentes découvertes suggèrent des capacités de dispersion larvaire à longue distance, mais ceci reste encore aujourd'hui à démontrer.

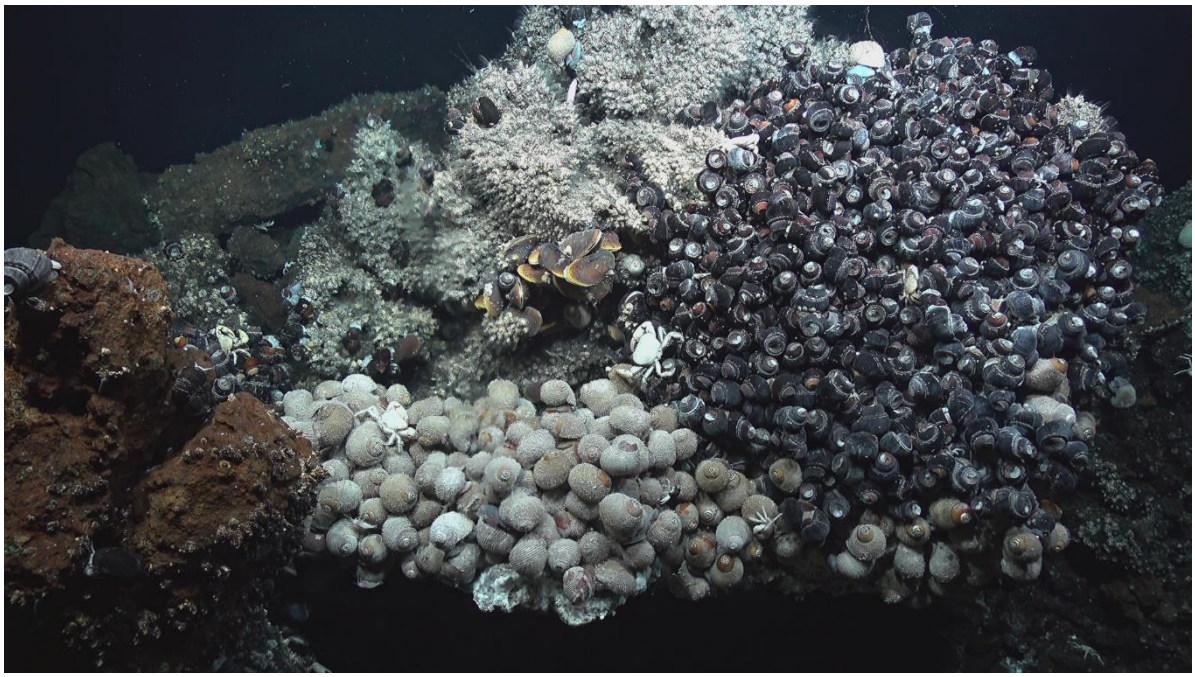


Figure 12: Arrangement des *Alviniconcha* spp., d'*Ifremeria nautilei* et de *Bathymodiolus brevior* sur le site Tui Malila dans le bassin de Lau.

Les particularités du genre *Alviniconcha* notamment en termes de (1) distribution des espèces, (2) composition symbiotique et (3) colonisation d'environnements hétérogènes (composition du sol, plus ou moins éloigné du point d'émission) rend ce complexe d'espèces intéressant pour des études de spéciation. En effet, d'après les connaissances actuelles chez ces espèces, la divergence mitochondriale semble élevée, mais nous n'avons que très peu d'information en ce qui concerne le génome nucléaire. L'hétérogénéité de l'environnement hydrothermal et la formation de niches écologique distinctes (comprenant la composition symbiotique) peuvent être une source de spéciation écologique. Cette hypothèse est prônée par Beinart et al. (2012) afin d'expliquer la spéciation des espèces d'*Alviniconcha*, mais aucune preuve n'a à l'heure actuelle été apportée. Ainsi, dans cette thèse, la divergence génomique et l'origine de la spéciation chez ce genre de gastéropode vont être étudiées.

4- Objectifs de la thèse

Dans cette introduction de thèse, il a pu être montré que le milieu hydrothermal est un environnement très particulier pour les études de spéciation, et des contributions de l'allopatrie et de la sympatrie dans l'évolution des espèces et que le genre *Alviniconcha* est un groupe intéressant pour ce genre d'études. Ainsi, cette thèse vise à questionner le rôle de l'hétérogénéité géographique et écologique de l'environnement hydrothermal profond sur la

distribution et la spéciation de trois espèces du genre *Alviniconcha* : les espèces *A. kojimai, A. strummeri* et *A. boucheti* dans les bassins arrière-arc du Pacifique Ouest. Pour tenter de répondre à cette question générale, plusieurs objectifs intermédiaires ont été définis :

1. Décrire la distribution actuelle des espèces cibles d'*Alviniconcha* dans le Pacifique Ouest.
2. Estimer la divergence de ces espèces à la fois sur les génomes nucléaire et mitochondrial.
3. Vérifier la présence de caractères morphologiques discriminants entre les espèces.
4. Identifier les scénarios démographiques historiques de séparation / contact entre les espèces pouvant expliquer la divergence observée.
5. Établir le rôle de l'adaptation locale au sein de chacune des espèces et son rôle éventuel sur le maintien de la différenciation inter-espèces.
6. Identifier le rôle de la sélection disruptive sur l'origine de la divergence entre espèces.
7. Etant donnée l'observation inattendue d'une forte différenciation génétique entre mâles et femelles chez *A. boucheti,* un objectif supplémentaire a été de définir le déterminisme du sexe chez les trois espèces étudiées.

5- Déroulé de la thèse

Certains des objectifs de la thèse cités précédemment ont été rassemblés au sein des différents chapitres constituant ce manuscrit. Ainsi, dans cette partie je présenterai le déroulé des chapitres de thèse avec un contenu résumé des méthodes utilisées et des principaux résultats escomptés.

Une partie sur l'échantillonnage général, après l'introduction générale, présente les stratégies d'échantillonnage et les méthodes utilisées dans les chapitres suivants.

Chapitre 1 : Ce chapitre de la thèse, qui fait l'objet d'un article soumis, a pour objectifs de décrire la distribution actuelle de *A. kojimai*, *A. strummeri* et *A. boucheti* sur l'ensemble de leur aire de répartition connue et d'inférer l'histoire de la divergence entre ces espèces. La distribution a été évaluée en combinant les échantillonnages effectués durant la campagne CHUBACARC (détaillée ci-dessous dans la partie échantillonnage) avec un barcode de(s)

(l')espèce(s) présente(s) à partir de l'analyse génétique du gène mitochondrial *Cox1*. Ce chapitre pose la question d'évaluer l'intégrité génétique des différentes espèces en présence à travers une estimation de la divergence génétique inter-spécifique entre différents compartiments génomiques (mitochondrial et nucléaire) et en testant la possibilité d'échanges génétiques inter-spécifiques. Pour cela nous avons estimé la divergence absolue (d_{XY}) à la fois sur le gène mitochondrial *Cox1*, sur les transcriptomes spécifiques, ainsi que sur un grand nombre de marqueurs nucléaires génotypés par ddRAD-seq. Cette notion d'intégrité a aussi été l'occasion d'évaluer la relation entre la divergence observée entre espèces et la fixation de critères morphologiques diagnostiques en combinant données génétiques, mesures morphométriques des coquilles et arrangements des soies périostracales de la coquille. Enfin, une analyse démographique à partir du logiciel DILS (Fraïsse et al., 2021) a été effectuée pour comprendre l'histoire démographique des espèces et évaluer les échanges passés et présents entre celles-ci.

Les données génétiques du gène mitochondrial *Cox1* obtenues pour ce premier chapitre ont été utilisées pour décrire la diversité de la faune sur un nouveau site du bassin de Woodlark. Ainsi, cette collaboration a permis la publication d'un article auquel j'ai contribué (voir annexe du chapitre 1).

A ce premier chapitre s'ajoute également une seconde annexe questionnant l'observation de double pics (DP) sur les séquences du gène mitochondrial *Cox1*. L'origine des DP a été investiguée en testant différentes hypothèses, notamment celles d'une contamination, d'un cancer transmissible et d'une double héritabilité des mitochondries (DUI). D'autres hypothèses restent encore à tester comme la présence d'une copie d'un fragment du gène mitochondrial dans le génome nucléaire (Numt) ou encore une duplication du gène *Cox1* dans le génome mitochondrial.

Chapitre 2 : Suite à des observations inattendues concernant le sexe des individus et la différenciation génétique entre mâles et femelles, dans ce chapitre de thèse nous aborderons le déterminisme du sexe chez les trois espèces d'*Alviniconcha*. Pour cela nous avons combiné des observations anatomiques réalisées par l'équipe de Florence Pradillon (Ifremer Brest) et nos résultats de génotypage afin de tester l'hypothèse d'un déterminisme génétique du sexe chez *A. boucheti*, *A. kojimai*, et *A. strummeri*.

Chapitre 3 : Ce chapitre vise à déterminer le rôle de l'environnement sur la différenciation génétique des populations de chacune des espèces d'*Alviniconcha*. Pour cela, une étude sur l'adaptation locale a été réalisée par une analyse GEA (genome-environment association) pour chacune des espèces. L'objectif ici est de tester si les différences génomiques trouvées entre populations des différentes espèces s'expliquent par le simple éloignement géographique ou de l'adaptation locale à des facteurs abiotiques (température, profondeur, habitat, nature chimique du fluide) et biotiques (composition bactérienne des symbioses). Pour cela, une étude préalable des différents facteurs environnementaux a été réalisée (via plusieurs collaborations détaillées dans le chapitre suivant) sur chaque échantillon récolté et des mesures isotopiques (δ^{13} C, δ^{15} N, δ^{34} S) ont été réalisées sur un sous-échantillon d'individus pris dans chaque boîte de récolte. Cette étude intra-spécifique nous a par la suite permis d'extrapoler à l'échelle inter-spécifique si l'adaptation locale a pu jouer un rôle dans le maintien de la différenciation entre les espèces. Une partie de cette étude repose sur l'analyse de la composition symbiotique des individus et a été réalisée en collaboration avec l'équipe américaine du Dr. R. Beinart (University of Rhode Island, USA). Cette collaboration a abouti à un article publié décrivant la composition symbiotique individuelle des espèces d'*Alviniconcha* sur l'ensemble de leur aire de répartition (article présenté en annexe du chapitre 3).

Chapitre 4 : Dans ce dernier chapitre, nous avons cherché à évaluer le rôle de la sélection divergente sur la formation ou la spécialisation des espèces actuelles d'*Alviniconcha* en cherchant à savoir à travers l'annotation des gènes quels mécanismes d'isolement (incompatibilités gamétiques, incompatibilités mito-nucléaires, adaptation différentielle à l'habitat hydrothermal) ont pu prévaloir au cours du temps depuis la séparation initiale des populations ancestrales. Pour cela, des transcriptomes de référence ont été séquencés et assemblés pour chaque espèce et comparés entre eux (groupes d'orthologie) afin d'examiner le nombre et la fonction des gènes sous sélection positive par l'analyse des mutations synonymes (d_S) et non-synonymes (d_N) le long des branches menant à chaque espèce. La fonction des gènes identifiée a permis d'évaluer l'impact de la divergence disruptive ou de la spécialisation sur l'évolution des espèces d'*Alviniconcha*.

Une discussion générale a été ensuite conduite pour intégrer les différents résultats obtenus et mettre en évidence les mécanismes évolutifs ayant eu un rôle clef dans la séparation des espèces et leur évolution au sein des bassins arrière-arc du Pacifique Ouest.

Echantillonnage et acquisition des données génétiques à partir d'approches par séquençage haut-débit

1- Stratégie d'échantillonnage

1.1- Campagne d'échantillonnage CHUBACARC

La campagne CHUBACARC (Connectivité et Histoire des commUnautés hydrothermales des BAssins/volCans arrière-ARC du Pacifique) s'est déroulée durant 70 jours répartis en deux legs entre le 25 mars et le 8 juin 2019. Cette campagne où j'ai eu la chance de participer était co-dirigée par S. Hourdez (Observatoire Océanologique de Banyuls, France) et D. Jollivet (Station Biologique de Roscoff, France). Cette campagne a été évaluée pour programmation par la commission nationale de la flotte hauturière (CNFH), et la partie des recherches menées sur la structure et l'évolution des espèces et de leurs communautés fait l'objet d'un financement plus ciblé dans le cadre de l'ANR Cerberus coordonné par S. Hourdez (Observatoire Océanologique de Banyuls, France). La campagne, financée partiellement (logistique) par le fonds de soutien de l'IFREMER (L'Institut Français de Recherche pour l'Exploitation de la Mer) et l'ANR Cerberus, avait pour objectif l'étude de la diversité et de la connectivité à l'échelle régionale des peuplements hydrothermaux des bassins arrière-arc du Pacifique Ouest et de la résilience des communautés face à l'exploitation minière, notamment dans les bassins de Manus et de Lau. Pour répondre à cet objectif, différents axes de recherche ont été lancés à partir des échantillons récoltés et des mesures physico-chimiques effectuées *in situ* sur 3 types de communautés (escargotières à *Ifremeria* et à *Alviniconcha* et moulières). Ces axes étaient : (1) estimer les diversités locales et régionales des différentes communautés hydrothermales profondes présentes dans ces bassins, (2) retracer l'histoire de la colonisation des différentes dorsales par les faunes présentes et (3) quantifier les degrés d'échanges génétiques entre et au sein des bassins chez plusieurs espèces cibles (en partant des micro-organismes jusqu'aux macro-organismes). Cette campagne a été caractérisée par sa multidisciplinarité en mélangeant géoscience et biologie/écologie des communautés profondes, notamment pour mieux évaluer le rôle des interactions entre chimie du fluide et composition faunistique des communautés et la recherche et caractérisation de nouveaux sites actifs, notamment au niveau de la ride de Woodlark (leg2) et de la zone Futuna (leg1). Plusieurs disciplines telles que la géochimie, la cartographie des sites, l'écologie des communautés, la génétique, la biologie des populations et l'évolution des espèces ont donc été associées sur la totalité de la campagne. C'est dans ce cadre que se situe le travail de thèse présenté ici.

La mission CHUBACARC menée sur le navire océanographique l'Atalante (Genavir, Ifremer) a permis l'exploration ainsi que l'échantillonnage de quatre bassins arrière-arc (Manus, Woodlark, Nord-Fidjien et Lau) ainsi que la zone hydrothermale active associée à l'arc volcanique de Futuna. Elle a également permis la découverte de deux nouveaux sites hydrothermaux, le premier nommé La Scala dans le bassin de Woodlark (Voir article en annexe du Chapitre 1) et le second nommé Mangatolo situé à l'extrémité Nord du bassin de Lau (Figure 13). Au sein de chaque bassin, la récolte de la faune a été réalisée à différentes échelles d'espace selon un plan d'échantillonnage hiérarchisé :

- Bassin = échantillonnage d'un à plusieurs champs d'activité
- Champ = collection de sites actifs (quelques centaines de mètres à un ou deux kilomètres de dorsale)
- Site = une à deux zones d'émission ou cheminées hydrothermales avec réplicats d'habitat pour une communauté donnée (par exemple pour les communautés à *Alviniconcha*, 2 à 3 échantillonnages par site, cf. détails ci-dessous)
- Echantillon = plusieurs individus d'espèces mélangées prélevés dans quelques dizaines de cm^2 (au sein d'un patch homogène d'individus, en général) et placés dans une même boîte de prélèvement isotherme (biobox).

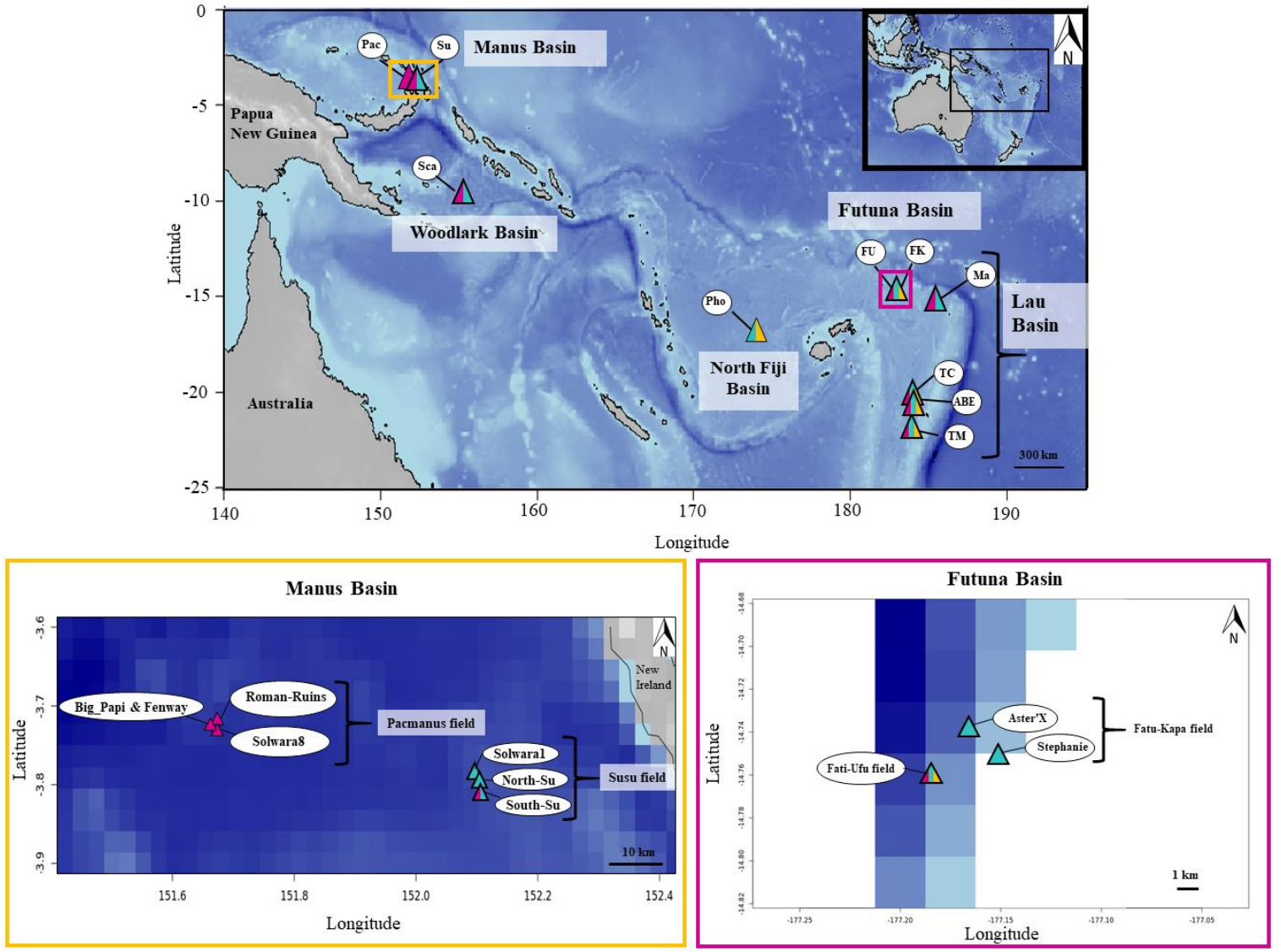

Manus Basin
Pac
Su
Papua
New Guinea
Sca
Woodlark Basin
Futuna Basin
FU
FK
Ma
Lau
Basin
Pho
North Fiji
Basin
TC
ABE
TM
Australia
300 km
N
Latitude
Longitude
0
-5
-10
-15
-20
-25
140
150
160
170
180
190
Manus Basin
Big_Papi & Fenway
Roman-Ruins
Solwara8
Pacmanus field
Solwara1
North-Su
South-Su
Susu field
New
Ireland
10 km
-3.6
-3.7
-3.8
-3.9
151.6
151.8
152.0
152.2
152.4
Futuna Basin
Aster'X
Stephanie
Fatu-Kapa field
Fati-Ufu field
1 km
-14.68
-14.70
-14.72
-14.74
-14.76
-14.78
-14.80
-14.82
-177.25
-177.20
-177.15
-177.10
-177.05

Figure 13: Aire d'échantillonnage de la mission CHUBACARC. Le bassin de Lau, Nord-Fidjien et la zone active de Futuna ont été échantillonnés durant le premier leg (avril-mai 2019), tandis que les bassins de Woodlark et de Manus ont été explorés dans un second temps (mai-juin 2019).

1.2- Utilisation du ROV Victor6000 pour les mesures et prélèvements

Durant la campagne CHUBACARC, l'ensemble des mesures et prélèvements ont été effectués grâce au ROV Victor6000 (Remote Operated Vehicle). Ce robot télé-opéré depuis la surface grâce à l'utilisation d'un câble est conçu pour l'exploration des grands fonds marins jusqu'à 6000 mètres de profondeur. Il est équipé de deux bras manipulateurs, d'un panier de prélèvement (qui permet entre autres de contenir les boîtes de prélèvements -Biobox- et divers équipements), de projecteurs et de caméras CCD. Ce « gros robot » (dimension : H=3,1m × l=1,8m × L=2,1m, Figure 14a) se déplace sur le fond à une vitesse maximale de 1,5 nœuds (i.e. environ 0,77 m/s). Les prélèvements de fluides chauds et tièdes sont effectués à partir de seringues en titane et du multi-préleveur PIF (Figure 14b), tandis que les prélèvements biologiques et géologiques sont eux réalisés grâce à une pince à godet et des boîtes de prélèvements (Figure 14c et 14d). La petite faune vagile est quant à elle obtenue avant et après défaunement à partir d'un aspirateur à faune sur barillet avec 6 bols de récupération. Des analyses du fluide chimique (T, pH, Fe, S) sont également effectuées de façon ciblée sur les trois communautés d'organismes cibles avec le même protocole en utilisant le préleveur-analyseur *in situ* Chemini. Durant la campagne CHUBACARC, 18 plongées comprises entre 12 à 72 heures ont été réalisées et ont permis la récolte de plus de 6 000 individus pour les 10 espèces cibles de la génétique des populations. Durant les longues plongées, un système d'ascenseur est mis en place et permet environ toutes les 12h d'échanger du matériel avec le fond. Ces manipulations permettent notamment d'échanger des boîtes de prélèvement pleines contre des vides et ainsi d'avoir toutes les demi-journées des remontées de faune à bord du navire.

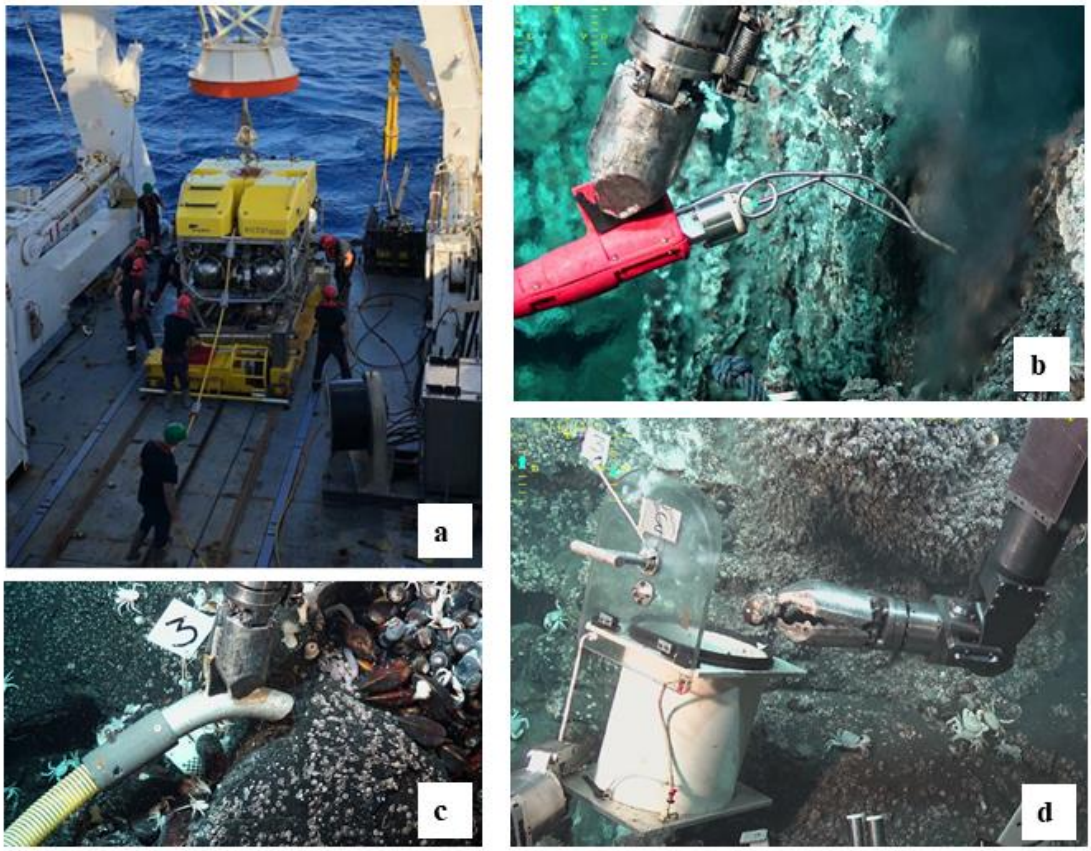


Figure 14: a) ROV Victor 6000 ; b) seringue titane ; c) aspirateur à faune ; d) pince à godet et boîte de prélèvement « biobox ».

1.3- Stratégie d'échantillonnage pour l'ensemble de la campagne

Pour chaque point de prélèvement (y compris pour les communautés à *Alviniconcha*), des mesures physico-chimiques (gaz, Fe, H_2S, Mn, composés organiques, pH, O_2) à quelques centimètres des organismes ont été réalisés de façon systématique avant et après en plus les prélèvements faunistiques. A l'échelle du champ lui-même, des prélèvements de roche (cheminées, laves et sulfures polymétalliques), de fluides chauds à la sortie des conduits de cheminées (fumeurs noirs) ont également été effectuées pour avoir une bonne connaissance de la composition chimique, minéralogique et microbiennes des roches et des fluides. A cela se sont ajoutées dans le bassin de Lau, Nord-Fidjien et dans la zone active de Futuna des récoltes de larves à l'aide de pompes autonomes SALSA placées dans la colonne d'eau à proximité des communautés.

1.4- Échantillonnage des trois espèces d'*Alviniconcha*

Une stratégie d'échantillonnage pour les *Alviniconcha* spp. a été mise en place au préalable de la mission CHUBACARC. Cette stratégie, respectée lorsque cela a été possible, a consisté à prélever ces gastéropodes et les communautés qui leur sont associées sur différents types d'habitats afin d'ultérieurement étudier si l'environnement joue un rôle dans la distribution et l'adaptation des différentes espèces. Des échantillons d'individus ont donc été récoltés avec la pince godet sur des zones de diffusion associées aux laves fracturées ainsi que sur plusieurs hauteurs de cheminées hydrothermales. Néanmoins, dans le cadre d'une approche écologique comparative et pour des raisons de faisabilité des mesures, seules les communautés à *Alviniconcha* des zones de diffusion sur les laves et au pied des cheminées ont fait l'objet d'un protocole strict de mesures physico-chimiques avant et après le prélèvement. Ainsi, durant CHUBACARC, pour chaque site hydrothermal visité, deux assemblages d'*Alviniconcha* issues de zones de diffusion et un assemblage sur une cheminée hydrothermale ont été échantillonnés dans la mesure du possible, les mesures physico-chimiques (hormis la température) n'étant effectuées que sur les zones de diffusion (Figure 15). Le récapitulatif du nombre d'individus échantillonnés dans des boîtes de prélèvements (Biobox) pour chaque agrégat de gastéropodes en fonction du type d'habitat est détaillé dans le Tableau 1.

Au total 816 *Alviniconcha* spp. ont été récoltés avec le ROV Victor6000 durant la campagne et répartis comme suit : 320 individus ont été échantillonnés dans le bassin de Lau, 125 dans la zone active de Futuna, 81 dans le bassin Nord-Fidjien (un seul site découvert au point triple, les sites White lady/Ivory Tower s'étant éteints), 72 dans le bassin de Woodlark (nouveau site La Scala découvert pendant l'expédition) et 218 dans le bassin de Manus sur la ride de Pual et les volcans Susu (Figure 13 ; Tableau 1). Au sein des différents bassins, plusieurs sites ont pu être explorés et/ou découverts pendant l'expédition (Mangatolo, Phoenix, et La Scala) et au final 16 sites ont permis la récolte d'*Alviniconcha* spp. Les différents sites échantillonnés se situent entre 1218 et 3388 mètres de profondeur (Tableau 1).

Les 816 individus d'*Alviniconcha* se répartissent au sein de 3 espèces *A. kojimai*, *A. boucheti* et *A. strummeri*. Tous les individus des 3 espèces ont été préalablement photographiés, mesurés et classifiés à bord selon plusieurs critères morphologiques empiriques (poils, couleur des branchies, columelle) puis identifiées formellement *a*

posteriori sur la base du gène mitochondrial *Cox1* (la divergence à ce gène permettant une bonne identification spécifique, comme cela avait été proposé par Johnson et al. (2015), et vérifié dans cette thèse (cf. Chapitre 1). Le nombre d'individus de chaque espèce récoltée au sein de chaque boîte de prélèvement et au sein de chaque bassin est résumé dans le Tableau 1. Dans ce tableau, il faut noter que l'espèce *A. strummeri*, absente des bassins de Manus et de Woodlark (leg 2), est moins abondante que les deux autres dans les bassins de Lau et Nord-Fidjien. De plus, cette espèce est présente exclusivement en mélange avec au moins une des deux autres et trouvée plutôt en périphérie des 'patches' de gastéropodes. Dans les bassins de Lau et Nord-Fidjien, ainsi que dans la zone active de Futuna, l'espèce *A. kojimai* semble généralement plus abondante alors que dans le bassin de Manus et de Woodlark c'est l'espèce *A. boucheti* qui est majoritairement récoltée, voire uniquement présente sur la ride de Pual (Pacmanus). Il faut également noter lors de certains prélèvements, la présence des trois espèces d'*Alviniconcha* en mélange. C'est notamment le cas sur le site de ABE (Tableau 1). Ces résultats de distribution sont présentés en détail dans le chapitre 1.

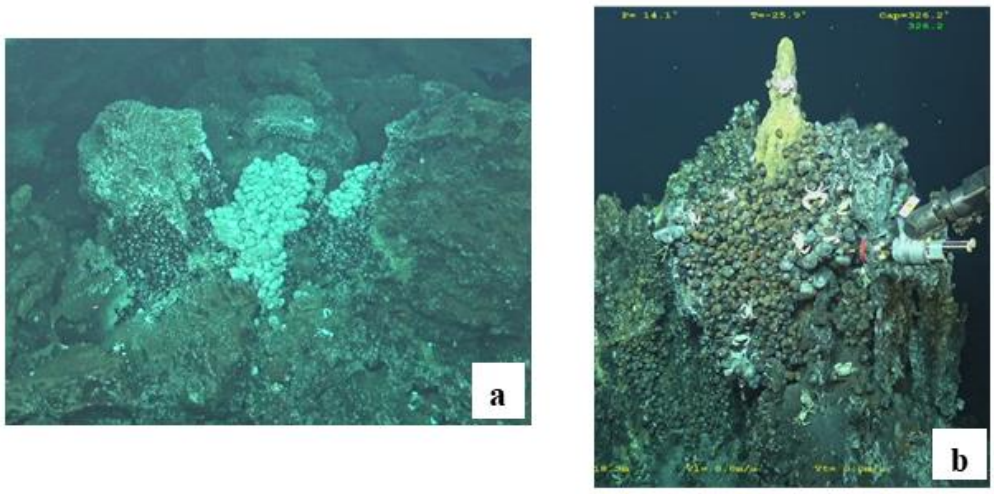


Figure 15: Communauté à *Alviniconcha* sur une zone diffuse (a) et sur une cheminée hydrothermale (b) à Fati-Ufu dans la zone volcanique de Futuna durant la campagne CHUBACARC.

Tableau 1 : Nombre d'individus de chaque espèce récoltés au sein de chaque biobox ("Population").

Population	Site	Champ	Bassin	Habitat	Nbr Individus *A. kojimai*	Nbr Individus *A. boucheti*	Nbr Individus *A. strummeri*	Non identifiés	Longitude	Latitude	Profondeur (m)	Date
721-GBT1	Tow Cam		Lau	Zone diffusion	33		1		176°08'15,4" W	20°19'04,4" S	2716	2019
721-GBT6	Tow Cam		Lau	Zone diffusion	44	5			176°08'15,8" W	20°19'05,1" S	2711	2019
721-GBT7	Tow Cam		Lau	Cheminée		20			176°08'12,7" W	20°18'59,2" S	2714	2019
722-GBT7	Tui Malila		Lau	Zone diffusion	31		5		176°34'04,2" W	21°59'15,2" S	1899	2019
722-GBT1	Tui Malila		Lau	Zone diffusion	33		19	3	176°34'05,9" W	21°59'21,2" S	1886	2019
722-GBT5	Tui Malila		Lau	Zone diffusion	32		15		176°34'05,5" W	21°59'21,4" S	1884	2019
731-GBT3	ABE		Lau	Cheminée	2	26	1	1	176°11'28,9" W	20°45'47,1" S	2149	2019
726-GBT4	Mangatolo		Lau	Zone diffusion		26			174°39'12,7" W	15°24'52,8" S	2031	2019
726-GBT3	Mangatolo		Lau	Cheminée		4			174°39'12,5" W	15°24'52,7" S	2031	2019
726-PBT6	Mangatolo		Lau	Zone diffusion	18				174°39'19,9" W	15°24'57,7" S	2039	2019
726-GBT2	Mangatolo		Lau	Zone diffusion	3				174°39'12,6" E	15°24'52,5" S	2031	2019
724-GBT4	Phoenix		North Fiji	Zone diffusion	39				173°55'7,6" E	16°57'0,0" S	1961	2019
724-PBT4	Phoenix		North Fiji	Zone diffusion	33		9		173°55'4,7" E	16°56'57,8" S	1973	2019
727-GBT2	AsterX	Fatu Kapa	Futuna	Zone diffusion	26				177°09'07,9" W	14°45'06,5" S	1562	2019
727-GBT4	Stephanie	Fatu Kapa	Futuna	Cheminée	16		1		177°09'57,6" W	14°44'14,7" S	1547	2019
728-GBT2		Fati Ufu	Futuna	Cheminée	27		3	2	177°11'07,0" W	14°45'35,8" S	1519	2019
728-PBT4		Fati Ufu	Futuna	Zone diffusion	40		1	2	177°11'04,9" W	14°45'35,3" S	1519	2019
728-GBT6		Fati Ufu	Futuna	Cheminée	3	4			177°11'05,9" W	14°45'35,2" S	1518	2019
733-GBT2	Big Papi	Pacmanus	Manus	Cheminée		24			151°40'20,1" E	3°43'43,9" S	1708	2019
733-GBT8	Fenway	Pacmanus	Manus	Zone diffusion		24			151°40'22,4" E	3°43'41,2" S	1696	2019
733-GBT9	Solwara8	Pacmanus	Manus	Cheminée		13			151°40'27,5" E	3°43'49,3" S	1737	2019
733-PBT7	Solwara8	Pacmanus	Manus	Cheminée		5			151°40'26,6" E	3°43'50,1" S	1734	2019
734-GBT9		Pacmanus	Manus	Cheminée		28			152°6'2,8" E	3°43'17,2" S	1659	2019
736-GBT3	Solwara1	Susu	Manus	Cheminée	24				152°5'47,0" E	3°47'22,1" S	1505	2019
736-GBT10	North Su	Susu	Manus	Zone diffusion	22				152°6'2,8" E	3°47'56,0" S	1218	2019
737-GBT10	South Su	Susu	Manus	Zone diffusion		24			152°6'18,6" E	3°48'35,0" S	1353	2019
737-PBT5	South Su	Susu	Manus	Cheminée	17	13			152°6'17,5" E	3°48'29,8" S	1300	2019
737-GBT7	South Su	Susu	Manus	Zone diffusion		24			152°6'17,9" E	3°48'31,8" S	1343	2019
738-GBT10	Scala		Woodlark	Cheminée	24				155°03'09,6" E	9°47'56,7" S	3388	2019
739-GBT10	Scala		Woodlark	Cheminée	24				155°03'07,0' E	9°47'56,3' S	3344	2019
739-PBT5	Scala		Woodlark	Cheminée		24			155°03'08,1" E	9°47'56,0" S	3353	2019

1.5- Protocole détaillé de traitement des individus à bord

Durant la campagne CHUBACARC, un protocole pour le traitement des *Alviniconcha* a été mis en place afin d'organiser le travail des chercheurs en écologie, en microbiologie, et en génétique. En effet, lors des plongées les plus longues, des remontées via l'ascenseur permettaient une arrivée toutes les 12h de boîtes de prélèvements contenant des gastéropodes. Une fois à bord, les boîtes de prélèvement étaient placées en chambre froide où les écologistes réalisaient un premier tri. Les *Alviniconcha* spp. « nettoyés » de la faune associée étaient stockés dans des aquariums oxygénés dans une eau à 8°C. Ce premier tri réalisé, une tentative d'identification rapide à l'œil a été effectuée notamment en fonction de la couleur des branchies des individus, de l'ornementation de la coquille et de la forme de la columelle. A chaque remontée, une dizaine d'individus par espèce a été récupérée par les microbiologistes, F. Pradillon et M.A Cambon-Bonavita (leg1) et V. Cueff (leg2) pour une dissection complète en vue d'analyses ultérieures portant sur la symbiose et l'analyse isotopique des tissus. Une fois les individus récupérés par les microbiologistes, le reste de la population échantillonnée à fait l'objet d'un échantillonnage spécifique de tissus pour la génétique des populations. Au début du premier leg (leg auquel j'ai participé), comme l'identification rapide n'était pas fiable (les critères d'identification ayant été améliorés en cours de campagne, cf. chapitre 1), tous les individus remontés sur le bateau ont été gardés et ont subi le traitement présenté ci-dessous. Pendant le second leg, l'identification à l'œil s'étant améliorée, seuls 24 individus par espèce et par échantillon (boîte de prélèvement) ont été choisis pour l'échantillonnage des tissus. Au-delà de ce nombre, les individus restants ont été conservés entiers dans des bidons remplis d'éthanol 100%.

Chaque *Alviniconcha* a été mesuré avec un pied à coulisse à 0,1 mm près (Figure 16), pris en photo et un fragment de coquille a été découpé et séché dans le but de pouvoir ultérieurement regarder l'ornementation des épines périostracales (poils) de la coquille. Par la suite, chaque individu a été disséqué sur glace et un fragment de branchie et de pied ont été conservés dans de l'éthanol 80% pour des analyses ultérieures comprenant le métabarcode de la population symbiotique du gastéropode. Un fragment de branchie a également été placé dans du RNAlater ou congelé à -80°C en vue d'une possible extraction d'ARN pour des analyses de transcriptomique. Pour la plupart des spécimens récoltés, une extraction d'ADN (Kit NucleoSpin® Tissue de Macherey Nagel ou CTAB-PVPP (Doyle et Doyle, 1987)) a été réalisée à bord sur du matériel frais pour obtenir des ADNs de qualité optimale en grande

quantité afin d'avoir suffisamment de matériel pour la construction des banques individuelles ddRAD-seq.

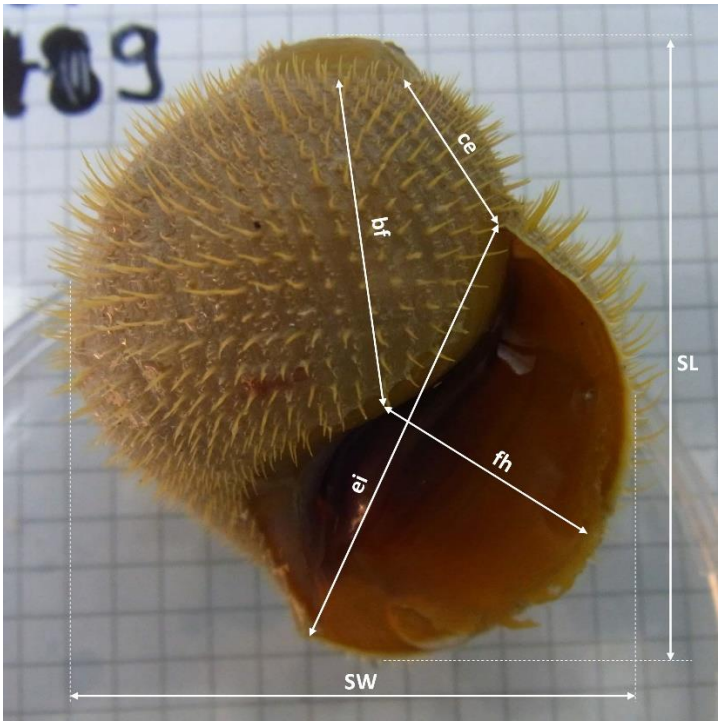


Figure 16 : Description des six mesures effectuées sur les coquilles d'*Alviniconcha* durant la campagne CHUBACARC d'après Chiu et al. (2002).

1.6- Échantillonnage complémentaire

Durant la campagne CHUBACARC, des mesures *in situ* de la température ainsi que de la concentration en sulfure et en fer ont été réalisées sur chaque communauté échantillonnée à l'aide de l'analyseur *in situ* Chemini (Vuillemin et al., 2009) par les chimistes C. Cathalot, E. Rinnert, O. Rouxel (IFREMER) et C. Boulart (SBR). Ces mêmes chimistes ont également utilisé le multi-préleveur PIF (dérivé du préleveur PEPITO ; Sarradin et al., 2008) pour échantillonner des fluides dilués sur chaque quadrat de faune échantillonné. Ces données sur l'environnement chimique des gastéropodes ont été pour partie dépouillées à bord ou obtenues par spectrométrie dans le laboratoire de géoscience de l'IFREMER pour être ensuite utilisées notamment dans le chapitre 3 de ce mémoire de thèse. Des analyses isotopiques du carbone ($\delta^{13}C$), souffre ($\delta^{34}S$) et azote ($\delta^{15}N$) ont également été effectuées par spectrométrie de masse sur des tissus congelés de pied associés à nos échantillons de gastéropodes afin de les relier à nos facteurs environnementaux à travers une collaboration développée avec L. Michel (IFREMER) durant cette thèse.

Dans le cadre de ce projet de thèse, nous avons pu utiliser d'autres échantillons d'*Alviniconcha* spp. issus de campagnes précédentes qui ont permis de compléter notre échantillonnage et effectuer les banques RNAseq avant le démarrage de la campagne CHUBACARC. C'est le cas des 10 échantillons utilisés pour l'analyse comparative des transcriptomes des trois espèces d'*Alviniconcha* dans le chapitre 4 de ce mémoire (Tableau 2). Ces tissus congelés à -80°C correspondant aux branchies de quatre individus de *A. boucheti*, de trois individus de *A. kojimai* ainsi que de trois individus de *A. strummeri* ont été récoltés par le ROV Jason dans le bassin de Lau et le bassin de Manus par S. Hourdez lors de précédentes campagnes en 2009 et 2011 (Campagne Lau2009 et BAMBUS2011 respectivement). Des échantillons supplémentaires de pied conservés dans de l'éthanol 80% correspondant à 9 individus de *A. boucheti* du site de Nifonea se trouvant au Sud-Ouest du bassin Nord-Fidjien ont également été gracieusement envoyés par V. Tunnicliffe (University of Victoria) suite à la campagne SO-229 du R/V Sonne (Zielske et Haase, 2014) et utilisés pour compléter les données génétiques déjà récoltées. Ce site est très intéressant car il constitue une pierre de gué entre les bassins de Woodlark et Nord-Fidjien échantillonnés durant la mission CHUBACARC (Figure 17).

Tableau 2: Individus utilisés pour obtenir les données RNAseq pour l'assemblage des transcriptomes de référence des 3 espèces *A. kojimai*, *A. boucheti* et *A. strummeri* du chapitre 4 de la thèse.

Individus	Espèce	Site	Bassin	Mission	Date de prélèvement	Prélèvement par
SH092-269-L22	*Alviniconcha boucheti*	ABE	Lau	Lau2009	2009	S. Hourdez
SH092-167-L10	*Alviniconcha boucheti*	Tow Cam	Lau	Lau2009	2009	S. Hourdez
SH092-172-L12	*Alviniconcha boucheti*	Tow Cam	Lau	Lau2009	2009	S. Hourdez
SH112-311-M24	*Alviniconcha boucheti*	North Su	Manus	BAMBUS	2011	S. Hourdez
SH092-600-L37	*Alviniconcha kojimai*	Tu'i Malila	Lau	Lau2009	2009	S. Hourdez
SH092-985-L62	*Alviniconcha kojimai*	Pete's experiment	Lau	Lau2009	2009	S. Hourdez
SH092-603-L30	*Alviniconcha kojimai*	Tu'i Malila	Lau	Lau2009	2009	S. Hourdez
SH092-991-L63	*Alviniconcha strummeri*	Pete's experiment	Lau	Lau2009	2009	S. Hourdez
SH092-040-L05	*Alviniconcha strummeri*	Killo Moana	Lau	Lau2009	2009	S. Hourdez
SH092-651-L68	*Alviniconcha strummeri*	Tu'i Malila	Lau	Lau2009	2009	S. Hourdez

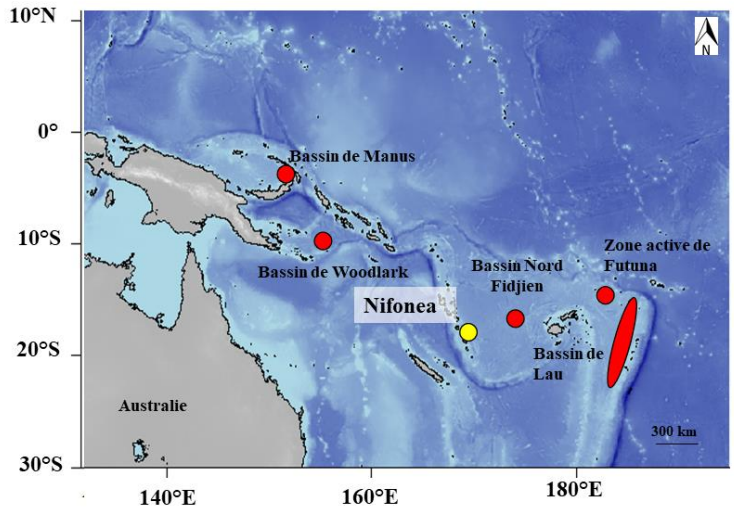


Figure 17: Principaux bassins échantillonnés pour obtenir les collections d'*Alviniconcha*. En rouge : échantillons récoltés pendant la campagne CHUBACARC 2019. En jaune : les échantillons supplémentaires du site Nifonea (campagne SO-209, 2013).

2- Obtention et traitement des données génomiques

2.1- Analyse ddRAD-seq

Pour étudier à partir des localités échantillonnées la diversité et la divergence génétiques, l'histoire démographique, l'adaptation locale ainsi que le déterminisme du sexe des *Alviniconcha*, une approche de génotypage par séquençage à haut débit a été choisi. Cette méthode correspond à un séquençage de fragments d'ADN correspondant à la digestion de l'ADN génomique par 2 enzymes de restriction spécifiques et appelé « RAD sequencing ». Cette technique, basée sur le séquençage de petits fragments d'ADN adjacents aux sites de coupure des enzymes de restriction, permet une réduction de la complexité du génome tout en conservant une très grande couverture de celui-ci, et présente l'avantage de pouvoir être mise en œuvre en l'absence de génome de référence.

Pour cela, l'ADN de 555 individus d'*Alviniconcha* (Campagne CHUBACARC + individus de Nifonea) a été extrait à partir d'un fragment de pied grâce à une extraction CTAB (Doyle et Dickson, 1987 modifié dans Jolly et al., 2003) ou avec le kit NucleoSpin® Tissue (Macherey-Nagel). La qualité de l'ADN génomique a été évaluée par migration sur gel d'agarose 0,8%. La construction des banques ddRAD réalisée par mes soins a été faite selon le protocole détaillé par Daguin-Thiebaut et al. (2021) et modifié du protocole de Peterson et

al. (2012)(Figure 18). Pour cela une double digestion d'environ 20 ng d'ADN par individu a été faite avec les enzymes *PstI* et *MseI*. Le multiplexage des individus en trois banques génomiques (fragments RAD) a nécessité l'utilisation combinée de 24 barcodes et de 12 index illumina. Ainsi, des adaptateurs illumina ont été ligués de chaque côté des fragments préalablement digérés. L'adaptateur P1, correspondant au site de coupure *Pst1* possède un barcode spécifique à chaque individu. Après la ligation, une étape de purification a été réalisée à l'aide de billes magnétiques AM micropure (Beckman™), afin d'éliminer les fragments d'ADN trop petits et les adaptateurs en excès. La suite a consisté en l'ajout d'un index illumina à chaque pool de 24 individus permettant d'augmenter la capacité de multiplexage au sein d'une banque. Pour cela une PCR normalisante a été réalisée afin d'obtenir une collection de fragments caractérisés par les 2 sites de restriction et correspondant à une combinaison index/barcode unique pour chaque individu au sein de la même banque. Après vérification de la PCR par électrophorèse sur gel d'agarose 2.5%, les individus ont été rassemblés et répartis dans trois banques différentes contenant une collection de fragments entre 300-600 bp sélectionnés après une électrophorèse Pippin Blue et visualisée à l'aide d'un BioAnalyser Agilent™. Ces banques ont ensuite été envoyées à Novogene (Cambridge, Royaume-Uni) pour un séquençage Illumina sur NovaSeq6000.

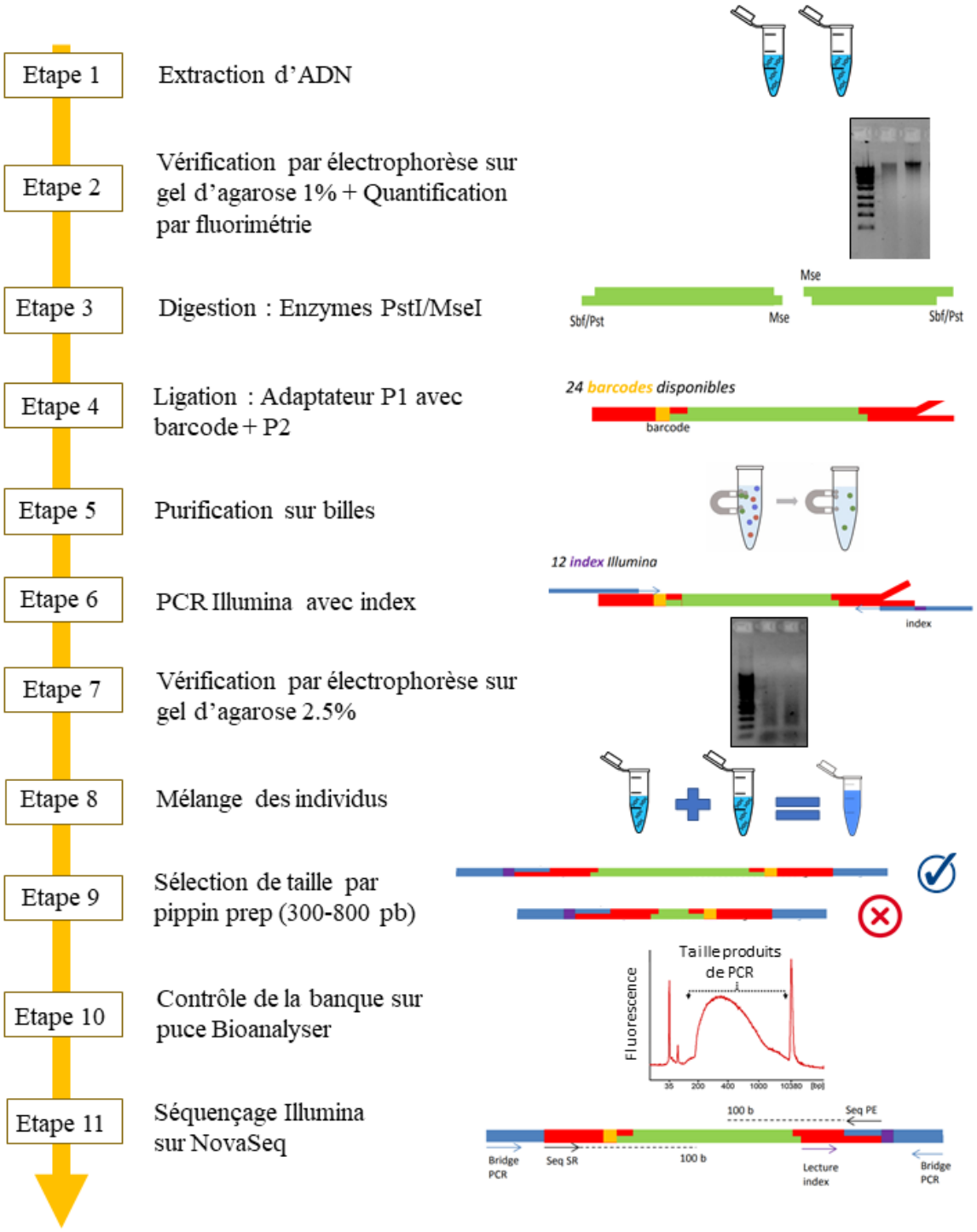


Figure 18: Protocole de construction des banques ddRAD-seq d'après Daguin-Thiebaut et al. (2021).

Après séquençage, un démultiplexage des individus a été réalisé à l'aide du programme process-radtags inclus dans le logiciel Stacks (Catchen et al., 2013). Cette procédure permet également d'éliminer les lectures de mauvaise qualité et d'éliminer les barcodes et adaptateurs dans les séquences. La qualité des lectures démultiplexées est vérifiée par multiQC v.1.7. Après l'élimination des lectures non eucaryotes par Kraken (Wood et Salzberg, 2014), l'assemblage des lectures restantes a été réalisé en *de novo* avec le module Stacks (Catchen et al., 2013). Les paramètres d'assemblage utilisés étant différents suivant l'analyse, ils seront détaillés dans chacun des chapitres de résultats. Après l'assemblage des lectures en locus, une filtration supplémentaire est réalisée dans R (R Core Team, 2020) afin de conserver uniquement les sites polymorphes (SNPs) génotypés chez 80% des individus et seulement les individus présentant au moins 80% de l'ensemble des SNPs. Les analyses de génétique de populations sont ensuite réalisées à partir des fichiers vcf produits par Stacks et filtrés dans R. Ces analyses que j'ai pu réaliser avec l'aide de mes encadrants sont résumées dans la Figure 21 et correspondent notamment au calcul de la divergence inter-espèces, l'organisation populationnelle des individus (snmf (R) ; ACP), les inférences démographiques (DILS), les indices de différenciation inter et intra espèces et entre les sexes ou encore l'analyse de l'Association Génome-Environnement (GEA). Lors des analyses GEA ou pour l'identification des locus dit 'outliers' présentant une forte différenciation entre sexes ou entre habitats, des annotations de certains locus ont été effectuées par Blastx sous Geneious dans les banques de données NCBI. Pour cela, une cartographie des gènes annotés sur les transcriptomes de référence des espèces d'*Alviniconcha* a été réalisée par Blastn avec le même logiciel (Voir section suivante). Ces différentes méthodes sont décrites au sein de chaque chapitre où elles sont employées.

2.2- Analyse RNA-seq

Pour étudier la divergence des séquences codantes de gènes entre les 3 espèces d'*Alviniconcha* et l'empreinte de la sélection sur ces derniers, un séquençage à haut débit sur des fragments d'ADNc a été réalisé. Pour cela, les ARN totaux de 10 individus (4 *A. boucheti*, 3 *A. kojimai* et 3 *A. strummeri*) ont été extrait grâce au kit d'extraction NucleoSpin® RNA Plus (Macherey-Nagel). Après vérification de la qualité des ARN extraits par migration des ARN totaux sous bioanalyzer, les banques de cDNA fragmentés (300-600 bp) ont été construites et séquencé en Illumina par la plateforme Genome Québec (Canada) sur une 1/2 ligne de NovaSeq. L'élaboration des banques a consisté à synthétiser l'ADN complémentaire

des ARNm extraits puis, à fragmenter les ADNc avant de rajouter les adaptateurs illumina possédant un barcode unique par individu afin de pouvoir rassembler les individus en un seul pool avant l'amplification par PCR et leur séquençage.

Une analyse de la qualité des séquences a été réalisée sur les données RNAseq avec fastQC et a conduit à l'élimination de 3 individus (un de chaque espèce nommé ABM24, AKL05 et ASL62 dans la Figure 19). Les lectures ont ensuite été nettoyées en supprimant les adaptateurs et les séquences de mauvaise qualité avec Trimmomatic (Bolger et al., 2014). Après l'élimination des lectures non eucaryotes par Kraken (Wood et Salzberg, 2014) au sein de chaque individu, un assemblage par espèce a été réalisé avec rnaSPAdes en laissant les paramètres par défaut (Bankevich et al., 2012). La qualité de l'assemblage a été évaluée en effectuant une assignation des lectures brutes sur les transcrits précédemment obtenues et en réalisant une analyse busco, qui vérifie si le nombre de gènes universellement connus chez les eucaryotes est retrouvé dans les transcriptomes assemblés. Au début de cette thèse, le logiciel Trinity (Grabherr et al., 2011) avait été utilisé afin d'assembler les données, or, les résultats d'assemblage étaient mauvais avec une fragmentation importante des gènes transcrits et un niveau de duplication élevé de ces gènes malgré plusieurs tests de paramétrage, c'est pourquoi seuls les assemblages rnaSPAdes ont été retenus pour mener l'analyse comparative des transcriptomes entre espèces (Figure 19).

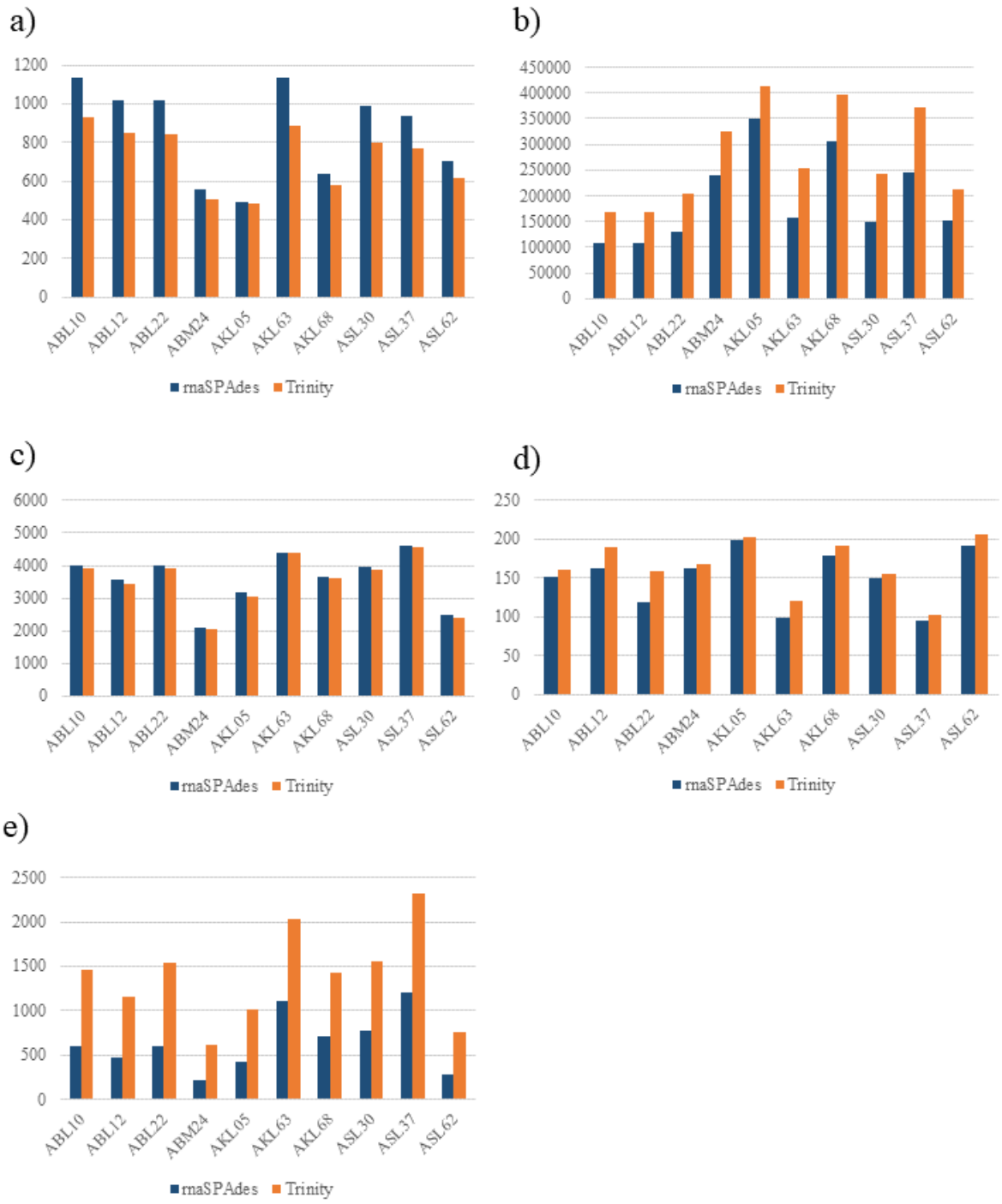


Figure 19: Comparaison des assemblages RNAseq obtenus par les logiciels rnaSPAdes et Trinity pour les 10 individus d'*Alviniconcha* spp. (a) longueur moyenne des transcrits, (b) nombre de transcrits, (c) nombre de gènes identifiés par l'analyse busco chez les mollusques à partir des 5295 gènes 'universels' de référence, (d) nombre de gènes de référence fragmentés identifié par cette même analyse busco et, (e) nombre de gènes dupliqués identifiés par l'analyse busco.

Après l'assemblage d'un transcriptome par espèce, nous avons utilisé le pipeline AdaptSearch (https://github.com/abims-sbr/adaptsearch ; Figure 20) afin d'identifier les gènes orthologues entre nos 3 espèces afin d'évaluer la divergence et le nombre de gènes sous sélection positive. Cette analyse a été complétée par une annotation par Blastx de ces transcrits pour connaître leur fonction métabolique. La première étape de ce pipeline consiste à filtrer les transcrits pour ne conserver qu'une seule séquence par gène en utilisant à la fois la longueur et le score de qualité des transcrits et rassembler certains fragments orthologues chevauchant pour un même gène dans un assemblage CAP3 supplémentaire au sein de chaque espèce (Filter_Assemblies). Les transcrits sont ensuite traduits en protéines pour être assemblés en orthogroupes avec Orthofinder (Emms et Kelly, 2015). Les séquences des orthogroupes sont ensuite alignées avec BlastAlign (Belshaw et Katzourakis, 2005) et remis dans le bon cadre de lecture à partir du script python CDSsearch (dans la suite AdaptSearch) qui vérifie le cadre de lecture en ne gardant que les alignements sans codon stop et en recherchant une annotation éventuelle par Blastx dans les banques NCBI. Enfin, les alignements de séquences codantes sont concaténés sur l'ensemble du transcriptome partagé et l'alignement utilisé pour reconstruire un arbre phylogénétique avec RaxML (Stamatakis, 2014). Cet alignement des séquences codantes est directement utilisé pour estimer la divergence absolue (d_{XY}) entre les espèces. Les divergences accumulées aux sites non synonymes (d_N) et synonymes (d_S) le long des branches terminales menant aux espèces et leur rapport est ensuite calculé avec CodeML (Yang et dos Reis, 2011). De plus, une analyse du d_N/d_S et des divergences entre espèces est également menée gène à gène pour identifier les gènes (et leur fonction) sous sélection positive. Un BlastX contre les banques du NCBI puis une annotation GO réalisée par D. Jollivet a permis d'identifier la fonction biologique des gènes présentant une divergence trop élevée ou un $d_N/d_S > 1$ pour chaque espèce lorsque le modèle sous sélection diversifiante M1 (n paramètres) est préféré au modèle d'évolution neutre M0 (p paramètres) après les avoir comparés par un Likelihood Ratio Test (LRT) pour un nombre de degrés de liberté équivalent à n-p-1 paramètres estimés entre modèles).

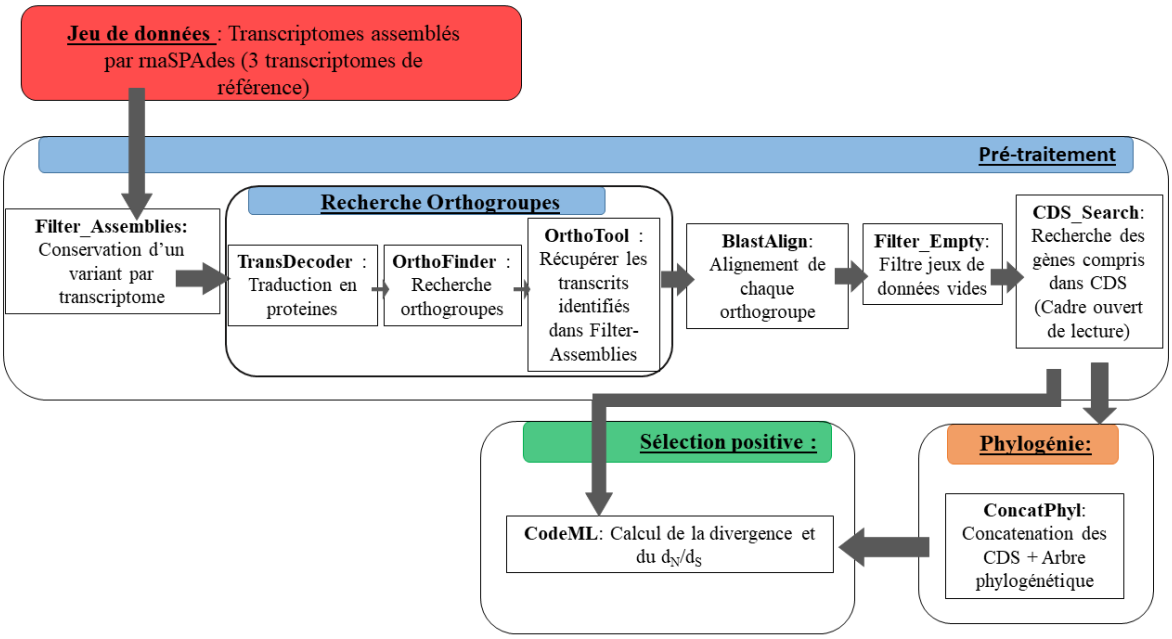


Figure 20: Structure algorithmique du pipeline AdaptSearch utilisé dans le cadre de cette thèse (d'après Mataigne et al. in prep).

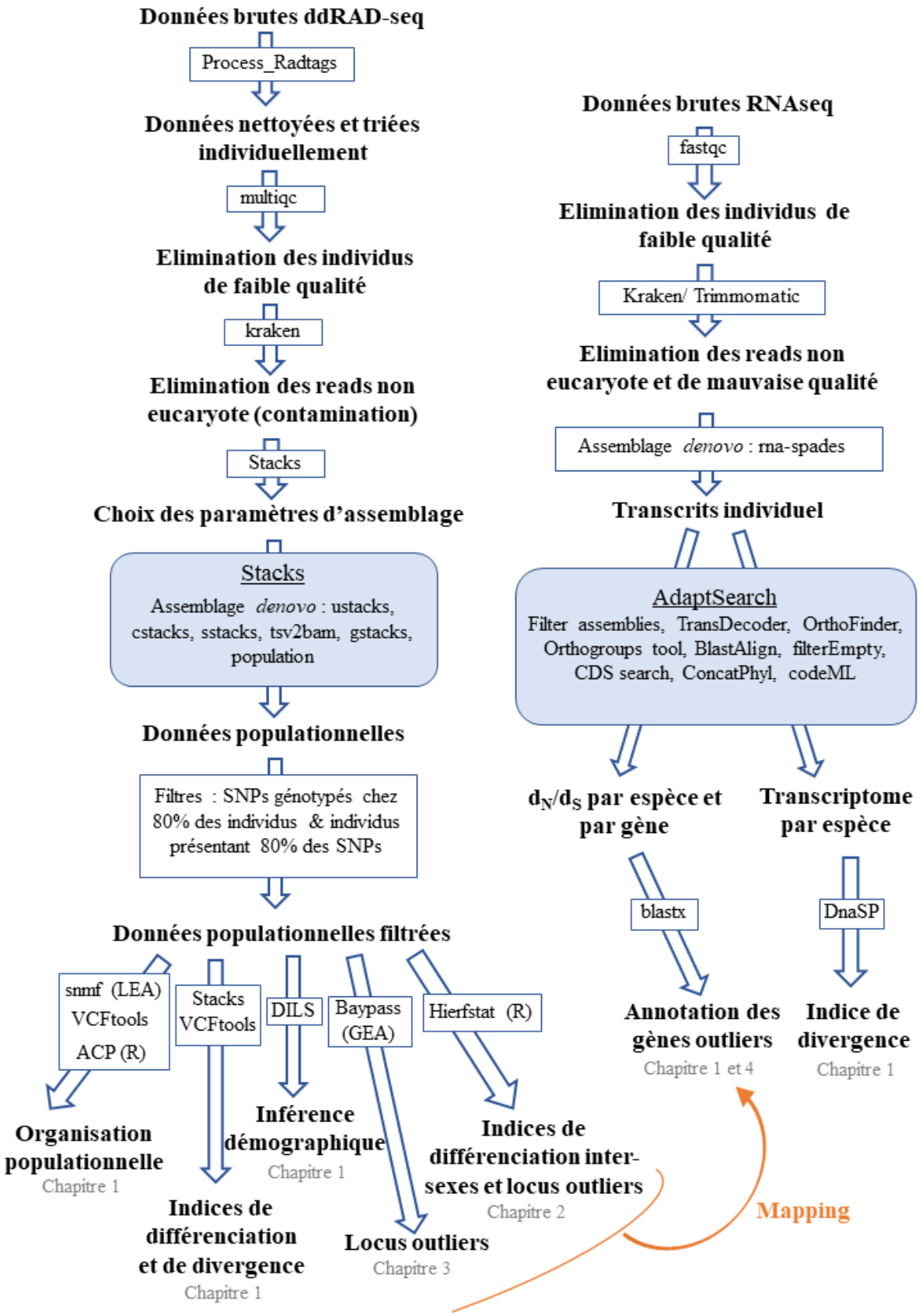


Figure 21: Schéma récapitulatif des analyses génomiques effectuées dans chacun des chapitres de résultats de cette thèse.

Chapitre 1:

Inter-specific genetic exchange despite strong divergence in deep-sea hydrothermal vent gastropods of the genus *Alviniconcha*

Chapter realized in collaboration with Hourdez S, Pradillon F, Daguin-Thiébaut C, Ballenghien M, Ruault S, Corre E, Tran Lu Y A, Mary J, Gagnaire P-A, Bonhomme F, Bierne N, Breusing C, Broquet T and Jollivet D.

1- Introduction

Pre- and post-zygotic isolating mechanisms that prevent or limit the exchange of genetic material between populations are expected to become stronger as lineages diverge (Wu, 2001; Coyne and Orr, 2004). However, the observed level of divergence between populations is the result of two distinct processes. First, it comes in part from the sorting of allelic lineages that were segregating before populations separated. Assuming mutation-drift equilibrium and an infinite-site mutation model, an average of this first component is theoretically equal to 4Ne.µ mutational events for a given gene, where Ne is the effective size of the ancestral population but only 2 Ne.µ (Q/2) when solely looking at only two alleles randomly drawn from this ancestral population (Edwards and Beerli, 2000). Under these conditions, 1.66 Ne and 6 Ne generations of allelic lineage sorting are required to initiate and achieve reciprocal monophyly between species for 99% of the whole genome, providing that the ancestral population split into two equal population sizes with the same demographic histories (Rosenberg, 2003). Second, additional divergence (referred to as the net divergence) accumulates at the rate of 2μ per generation since the population split, regardless of effective population size. It is during this divergence process that genetic incompatibilities can accumulate, and emerging adaptive alleles may be fixed, participating in the build-up of reproductive isolation between populations. Following these assumptions, population/species divergence time is likely to be different from zero and thus fixed when greater to 4.*Ne*/n where n is the number of loci examined (Edwards and Beerli, 2000).

One must also consider that nuclear and organelle genomes can evolve at different rates. In animals, the mitochondrial genome evolves generally much faster than the nuclear genome (Havird and Sloan, 2016). In the period following the population split, when lineage sorting is incomplete, this difference can be due to the fact that mtDNA is inherited uniparentally in most animals and thus has a fourfold smaller effective size providing that sex ratio is 1:1 with an equal reproductive success among males and females (Rosenberg, 2003). This means that mtDNA will complete the lineage sorting process four times faster than nuDNA (Hudson and Turelli, 2003). However, beyond this period, the difference between

mitochondrial and nuclear evolutionary rates mostly relies on selective pressures (e.g. purifying selection and selective sweeps) acting differently on the two genomic compartments and on nuclear compensation (Havird and Sloan, 2016). This difference in the speed at which genetic differences accumulate between the two genomic compartments is likely to promote mito-nuclear incompatibilities (Decker and Ammerman, 2020; Toews and Brelsford, 2012), which often create post-zygotic barriers during the speciation process (Lima et al., 2019; Pante et al., 2019).

As a consequence, when divergent lineages come into secondary contact, speciation processes may be attenuated or reinforced, depending on the time spent since the ancestral population split and how they are still able to hybridise and maintain viable and fertile hybrids (Orr and Turelli, 2001). In this context, inter-specific genetic exchanges due to local hybridization are, among other factors, dependent on the number of genetic incompatibilities accumulated and become rare when the time elapsed since the first isolation largely exceeds the time needed for allelic sorting (i.e. T >>2N generations). As a consequence, most species in the speciation phase and still capable of hybridization have relatively low divergence that is typically no more than a few percent in the mitochondrial genome (Abbott et al., 2013; Stelkens and Seehausen, 2009) and, even less in the nuclear genome (Roux et al., 2016). This observation is, however, challenged in marine invertebrate species for which a number of cryptic but highly divergent taxa (*Cox1* divergence >15%) still have porous nuclear genomes while hybridising locally (Jolly et al., 2005; Muths et al., 2009; Roux et al., 2013). These inter-specific exchanges between rather divergent genomes can be explained by the way most marine invertebrates reproduce (i.e., broadcast spawning and the propensity of a great variety of marine taxa - and especially those living under extreme conditions - for morphological stasis affecting among others the reproductive apparatus) (Lindholm, 2014; Parsons, 1994). Species crypticism may thus account for possible inter-specific pairing (but also see how marine invertebrates pheromones and egg-sperm recognition molecules affect cross-species mating (Palumbi, 1994; Vacquier et al., 1990).

The tectonic history and fragmentation along oceanic ridges of deep-sea hydrothermal ecosystems provides an interesting situation to study the evolution of divergence and speciation. Since the discovery of the first faunal communities associated with hydrothermal vents on the Galapagos rift in the late 1970s (Lonsdale, 1977), many endemic species have been described. But because sampling in the deep sea is difficult, our understanding of the evolutionary and ecological mechanisms that have led to the present-day distribution of vent-

associated fauna is still relatively limited. The continuous movement of the Earth's tectonic plates since the formation of the oceans has acted as an evolutionary force to partition the vent fauna over geological time scales (Matabos and Jollivet, 2019; Moalic et al., 2012; Tunnicliffe, 1991). Tectonically-driven allopatric speciation thus appears to be common in hydrothermal systems (Vrijenhoek, 1997). Hydrothermal vents are not only fragmented in space but also in time, as tectonic movements and volcanism lead to the recurrent birth and extinction of vent sites. This dynamics is likely to lead to extinction/recolonization events in the vent populations which in turn promotes either population isolation or secondary contact (Jollivet et al., 1999). In addition, thermal and chemical conditions (high temperature, low oxygen, high concentration of CO_2, H_2S, H_2 and CH_4 (Le Bris et al., 2003) may induce strong purifying selection on the molecular arsenal to adapt to this hypoxic and highly toxic environment (Fontanillas et al., 2017) and could thus be a way of favouring morphological and functional stasis. Despite the homogenising and convergent effect of living under sulfidic conditions, mineral effluents could however vary locally depending on the nature of the oceanic crust and associated fluid chemistry (Ondréas et al., 2012; Reeves et al., 2011) and should affect the distribution, interactions and speciation mechanisms of the species present. A large number of hydrothermal species develop symbiosis with microorganisms. In this case, environmental factors act on microbiomes, where interactions between host and microbiota can be affected and participate in local adaptation mechanisms.

As part of deep-sea hydrothermal vent communities, large symbiotic gastropods of the genus *Alviniconcha* (Gastropoda: Abyssochrysoidea), inhabit warm (7-42°C), sulphur-rich (250 µM) and poorly oxygenated (< 50 µM) diffuse venting environments where they represent a group of engineer species (Podowski et al., 2009, 2011). To date, six species of *Alviniconcha* have been described using as a proxy their high level of genetic divergence of the mitochondrial gene *Cox1* (Johnson et al., 2015). Assuming that the evolutionary rate of the *Cox1* gene has been constantly low (around 0.0015 substitutions per site and million years), the speciation events that led to the current species are thought to have occurred between 38 and 10 million years ago (Breusing et al., 2020). This suggests a long history of allopatric speciation that may be linked to the plate history of the region and variations of environmental conditions that could have modified the nature of symbiotic interactions.

Here, we focus on three species that correspond to lineages found across a series of back-arc basins in the Western Pacific: *A. kojimai* (Johnson et al., 2015) and *A. boucheti* (Johnson et al., 2015) partially overlap in the Manus, North Fiji, Futuna, and Lau back-arc

Basins, and *A. strummeri* (Johnson et al., 2015) which is currently restricted to the Futuna volcanic arc and the most Southern part of the Lau Basin and Futuna arc (Breusing et al., 2020; Johnson et al., 2015; Laming et al., 2020). Like many hydrothermal species, *Alviniconcha* hairy snails have long been considered cryptic (i.e. species that cannot be separated based on their morphology but that are genetically distinct) (Denis et al., 1993; Johnson et al., 2015; Kojima et al., 2001). However, the recent study of Laming et al. (2020) described morphological and anatomical differences between the three above species in the active hydrothermal zone near to the volcanic arc of Wallis and Futuna.

The species of *Alviniconcha* have so far been distinguished solely on the basis of two mitochondrial markers (*Cox1* and 16S) and a few slowly-evolving nuDNA genes (18S, 28S), as no diagnostic morphological features existed until the very recent study by Laming et al. (2020). Thus, there is still a need to verify the delineation and evolutionary history of these species by examining both their morphological differences and inter-specific divergence over the species distribution ranges and possible local genetic exchanges between their genomes. This study based on three species of *Alviniconcha* aims at 1) describing the current distribution of *A. kojimai*, *A. boucheti*, and *A. strummeri* in the Western Pacific back-arc basins, 2) estimating their genetic divergence on different genomic datasets, mtDNA, nuDNA and transcribed sequences (i.e. transcriptomes) to test whether these divergences are congruent and proportional to the time elapsed since estimated species separation, 3) understanding the demographic history of these species and whether they have diverged in allopatry, and 4) investigating whether these species still display fixed morphological differences over their overall species range (i.e. five Western Pacific back-arc basins) using morphological traits previously highlighted as diagnostic in the volcanic arc of Wallis and Futuna (Laming et al., 2020).

2- Materials and Methods

2.1- Sampling

A total of 816 individuals of *Alviniconcha* spp. were sampled during the CHUBACARC expedition (https://campagnes.flotteoceanographique.fr/campagnes/18001111/) conducted in May-June 2019 at 18 different vent fields from 5 back-arc Basins of the Western Pacific Ocean on board the N/O L'Atalante (chief scientists: S. Hourdez and D. Jollivet). Species were identified *a posteriori* following a barcoding approach using the mitochondrial *Cox1* gene

(see results, Table S1 and Figure 22) and the diagnostic reference sequences of Johnson et al. (2015). Sampling was conducted with the tele-manipulated arm of the remotely operated vehicle (ROV) Victor6000. The snails were scooped either on diffuse venting sites or on the wall of active chimneys and transferred into insulated biological boxes ("bioboxes"). Upon recovery on board, samples were temporarily stored in tanks containing refrigerated sea water (4°C) before being examined and dissected. Several soft tissues (gill, foot, mantle) were preserved in both 80% ethanol and RNALater, and fresh foot tissue was also used for immediate DNA extraction on board (see below). Geographic information about the gastropod collection used during this study is summarised in Table S1.

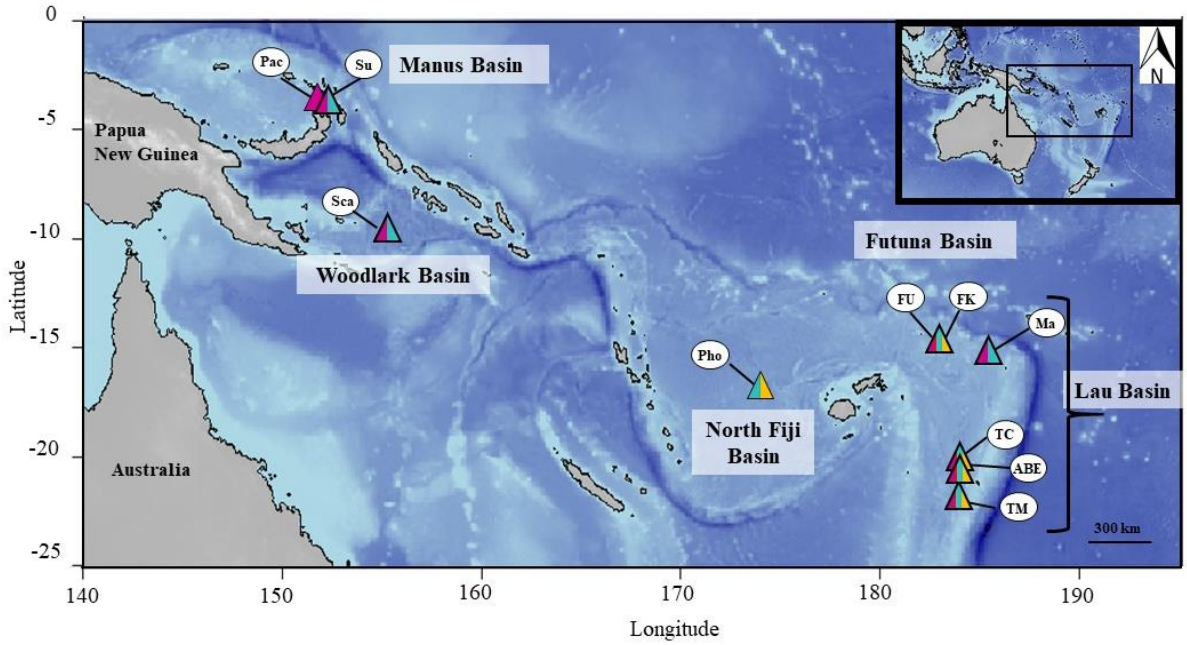


Figure 22: Sampled localities of *Alviniconcha* species during the CHUBACARC expedition. Manus Basin: Susu (Su); Pacmanus (Pac). Woodlark Basin: Scala (Sca). North Fiji Basin: Phoenix (Pho). Futuna arc: Fatu-Kapa (FK); Fati-Ufu (FU). Lau Basin: Mangatolo (Ma); Tow Cam (TC); ABE (ABE); Tu'i Malila (TM). Colours indicate species occurrence, purple: *A. boucheti*, turquoise: *A. kojimai*, and yellow: *A. strummeri*. The map background was obtained using the R package marmap (Pante and Simon-Bouhet, 2013).

Ten frozen (-80°C) individuals from the TN 235 Lau expedition (2009) on board the R/V Thomas G. Thompson (chief scientist: C.R. Fisher) were used for total RNA extraction in order to produce RNAseq datasets and subsequent transcriptome assemblies.

2.2- Morphological analysis

Prior to the dissection of the animals, the shell was photographed and a series of six distinct shell traits (length, width, spires and aperture lengths, as detailed in Chiu et al. (2002) and in Figure S1) were measured with a calliper. Because specimens within and between collections had different sizes, we used the total shell length to standardise the value of the remaining five traits. After checking there was no growth allometry for these five traits, they were scaled and centred so that they all had a mean of 0 and a standard deviation of 1 to prepare the following statistical analyses. These five transformed variables were then analysed in a principal component analysis (PCA) and a linear discriminant analysis (LDA). All analyses were performed in R using functions from packages ade4 (Thioulouse et al., 2018) and MASS (Ripley et al., 2013).

A shell fragment (approximately 4 by 3 cm) was cut off near the shell aperture in a subset of 732 individuals to evaluate the 'hairy' periostracal ornamentation. It consists of bristles arranged in rows. Bristles were classified into three different types: small, medium and large, which have been used to discriminate species (Laming et al., 2020). These three types of bristles were arranged in seven different ways (noted a-g) for a given row (Figure 23a). Because each individual may bear several arrangements depending on the examined rows, individuals were characterised according to the different bristle arrangements they had (e.g. ab, abc, etc.). The presence of either a double- or single-twisted columella was also observed and recorded for each individual (Laming et al., 2020).

2.3- Mitochondrial Cox1 sequence analysis

DNA was extracted on board from a small piece of fresh foot tissue of the 816 individuals using either the NucleoSpin® Tissue kit (Macherey-Nagel) or the CTAB protocol (Doyle and Dickson (1987) modified in Jolly et al. (2003) for the amplification of the mitochondrial *Cox1* gene and the preparation of ddRAD libraries. Genomic DNA quality was checked on return to the laboratory following a 1% agarose gel electrophoresis and DNA extracts were quantified with a fluorimetric method with the Quantifluor Promega kit in a Spark plate reader (Tecan). A 709-bp fragment of the mitochondrial cytochrome *c* oxidase subunit I (*Cox1*) gene was amplified from diluted DNA extracts with the LCO1490 (5'-GGTCAACAAATCATAAAGATATTGG-3') and HCO2198 (5'-TAAACTTCAGGGTGACCAAAAAATCA-3') primer pair (Folmer et al., 1994). PCR amplifications were conducted in a final volume of 25 µl using 2.5 µl of diluted template

DNA, 0.1 µM each primer, 50 µM of each dNTP, 2 mM $MgCl_2$, 0.5 U Flexi GoTaq® polymerase (Promega), 0.1mg/ml of Bovine Serum Albumin in 1X Green GoTaq® Reaction Buffer (Promega). The amplification protocol was as follows: 94 °C for 3 min, followed by 35 cycles with 94 °C for 30 s, 52 °C for 45 s, 72 °C for 1 min, and a final elongation step at 72 °C for 5 min, in a T100 thermocycler (Biorad). Because of the co-amplification of an additional non-specific small fragment for some specimens with the forward primer, PCR products were sequenced only from the HCO2198 primer, with the Sanger method Eurofins Genomics, Ebersberg, Germany). The *Cox1* gene sequences were manually checked for polymorphic sites using CodonCode aligner (v. 5.1.5, Codon Code Corporation), edited in Bioedit (Hall, 1999) and aligned by ClustalX (Thompson et al., 1997). One reference sequence representing each of the described *Alviniconcha* species stemming from Johnson et al. (2015) and Suzuki et al. (2006) (Genbank accession numbers: AB235216; KF467675; KF467921; KF467741; KF467873; KF467896) was added to the dataset. Sequences of five individuals of the species *Ifremeria nautilei* also collected during the CHUBACARC cruise were used as outgroup. Haplotype and nucleotide diversities (*Hd* and π, respectively), the substitution rates at non-synonymous (d_N) and synonymous sites (d_S) as well as the absolute population divergence (d_{XY}) were estimated with DnaSP 6.0 (Nei and Li, 1979; Rozas et al., 2017) between the three pairs of *Alviniconcha* species. McDonald-Kreitman tests were used to evaluate the degree of adaptive evolution of the *Cox1* gene within and between species. Inter-specific pairwise F_{ST} were calculated from haplotype frequencies with ARLEQUIN V.3.5 (Excoffier and Lischer, 2010). A haplotype network was created with the median-joining method (Bandelt et al., 1999) using the software PopArt (Leigh and Bryant, 2015) to depict phylogenetic relationships based on the mitochondrial sequences.

2.4- Nuclear genome analysis

2.4.1- ddRAD library preparation

A subset of foot fragment from 532 individuals were used for the production of individual double digested restriction site associated DNA (ddRAD) libraries (Peterson et al., 2012; Brelsford et al., 2016) following the protocol fully detailed in Daguin-Thiébaut et al. (2021). Briefly, genomic DNA of each individual (~60 ng) was digested with the restriction enzymes *Pst I* and *Mse I*, ligated to Illumina Truseq adapters containing a 6-bp barcode, and purified with AMPure XP beads (Beckman Coulter) and prior to PCR amplification with Illumina indexed primers. Individual PCR products were checked on agarose gels and then

pooled in three distinct groups of multiplexed individuals (24 barcodes and 8 illumina indices) before performing a final fragment size-selection (300-800 bp) using a Pippin Prep system (Sage Science). Distributions of DNA fragment sizes were checked in a high-sensitivity dsDNA chip using a BioAnalyzer 2100^{TM} instrument (Agilent). The three pools were sent to Novogene Europe (Cambridge UK) for 150-bp paired-end sequencing on an Illumina Novaseq6000 sequencer.

2.4.2- Bioinformatic filtering of Illumina reads

Raw reads were first demultiplexed using the process_radtags module of the Stacks software version 2.52 (Catchen et al., 2013; Catchen et al., 2011) that also removed adapters and low quality reads. Average sequence quality per read and GC-content were checked using multiQC version 1.7. Paired-end 144-base reads were assembled using the *de novo* pipeline in Stacks v2.52 (Catchen et al., 2013; Catchen et al., 2011). Assembly parameters (*m* 6; *M* 14; *n* 14) were chosen after empirical testing over a range of values (*m*: 2–6; *M*: 2–12; *n*: 2–18) on a subset of 23 individuals that included the three *Alviniconcha* species. Among these 23 individuals, 13 were triplicated for a total of 49 samples, which made it possible to evaluate the genotyping error rate according to the parameters tested. The applied parameter settings resulted in the greatest number of loci retained while maintaining a minimum genotyping error rate (see details in supplementary material; Figure S2). Trimmed reads were aligned in unique stacks (RAD-seq equivalent of alleles) if six or more identical reads were found within an individual (*m* = 6). Alleles were then compiled within each individual into sample-specific loci if they differed by less than 14 nucleotides (~10% divergence; *M* = 14) to consider the divergence of putative introgressed alleles between species. To assess the inter-specific divergences and shared polymorphisms, sample-specific loci were then assembled between individual samples of the three species into homologous loci if they differed by less than the same number of nucleotide changes (~10%, *n* = 14). A high value of *n* was chosen, despite the risk of assembling paralogues, because previous studies showed a strong divergence between these species (Breusing et al., 2020; Johnson et al., 2015). To identify single nucleotide polymorphism sites (SNP), clustering was performed with the denovo_map pipeline of Stacks-2.52, using a popmap of 570 individuals and the parameters defined previously (Catchen et al., 2013; Catchen et al., 2011). First, a sub-popmap consisting of 150 individuals distributed among the three species (one species constituting one population) and the different basins sampled was used for the construction of the inter-individual catalogue

(ctacks). Second, all individuals were mapped to this catalogue (sstacks). The raw SNPs data were filtered against the following thresholds: minimum individuals sharing a locus in a population $r \geq 0.8$ and minor allele count MAC ≥ 4. Using this first dataset (hereafter referred to as the "raw" dataset), we examined the number of RAD loci and SNPs that were genotyped within each species or were shared between species, to get a picture of the impact of divergence between species in the RAD-seq dataset.

The SNPs identified by Stacks were then further filtered for missing data using R scripts to only keep the SNPs genotyped in 90% of the individuals and to keep only the individuals which were genotyped at more than 85% of the SNPs. In the end, a high-quality dataset of 60 577 SNPs for 532 individuals (267 *A. kojimai*, 228 *A. boucheti* and 37 *A. strummeri*) was obtained.

2.4.3- Population structure, divergence and admixture

The population structure associated with the SNPs shared between the three *Alviniconcha* species was visualised in R using a principal component analysis (PCA; adegenet package; Jombart, 2008). The population divergence (d_{XY}) between species, the nucleotide diversity (π) and the net divergence ($d_A = d_{XY} - (\pi_X+\pi_Y)/2$) was estimated following Nei and Li (1979) using the --fstat option in Stacks-2.52. F_{ST} between species was calculated with the R package Hierfstat (Goudet, 2020). Individual admixture coefficients were estimated using the function snmf from the R package LEA (Frichot and François, 2015). Assuming K ancestral populations, the R function snmf provides least-square estimates of ancestry proportions for all analysed individuals (Frichot et al., 2014). This analysis was carried out in 20 runs with the 532 individuals obtained previously during the ddRAD-seq analysis and K=3 (derived from PCA analyses and the best entropy value, see results).

2.4.4- Demographic inference

The ABC software DILS (Demographic Inferences with Linked Selection by using ABC; Fraïsse et al., 2021) was used to determine which historical scenario of isolation *versus* gene flow might best explain the genetic structure observed today between the three *Alviniconcha* species in the Western Pacific ocean. In an approximate Bayesian computational framework, DILS compares summary statistics from simulated and observed datasets to identify plausible demographic scenarios and to jointly estimate population parameters such as effective population sizes, time of isolation, and past and contemporary gene flow between previously-separated populations. It also takes into consideration changes in population size,

local selective effects along the genome (variation in the effective size N_e among loci), and semi-permeable barriers to gene flow (variation in the effective migration rate *me* among loci). Because DILS is designed to simulate divergence scenarios for a pair of populations, we reanalysed our RAD-seq dataset for the three pairs of species by running Stacks' *population* function using the common catalogue of loci described above, with $r = 0.8$ and MAC = 4. Then we only kept SNPs genotyped in at least 90% of the individuals. For the demographic inference, 5 655 RAD loci and 265 individuals were used for the *A. boucheti/A. strummeri* pair, 8 808 RAD loci and 495 individuals for the *A. kojimai/A. boucheti* pair, and 5 775 RAD loci and 304 individuals for the *A. kojimai/A. strummeri* pair. Because DILS uses a random subsample of 1000 loci to estimate observed summary statistics, we ran 10 independent analyses for each pair of species, under the following main models: Strict Isolation (SI), Ancient Migration (AM), Isolation with Migration (IM) and Secondary Contact (SC). During these runs, the method accounted for two events of instantaneous population size change (one when the ancestral population splits into two daughter populations, and one at any later time). Finally, for each run the program also tested whether N_e and *me* are distributed heterogeneously along the genome (therefore accounting for the effect of linked selection and local barrier effects on gene flow along the genome). The goodness-of-fit of each run was estimated using the Euclidean distance between the observed and simulated summary statistics calculated with an accompanying python script. Detailed parameter settings are provided in the supplementary material (see example of detailed parameter settings section).

2.5- Transcriptome analysis

Total RNA was extracted using the NucleoSpin® RNA Plus kit (Macherey-Nagel) according to the manufacturer's protocol from the frozen gill tissue of two *A. kojimai*, two *A. strummeri* and three *A. boucheti* individuals. RNA integrity was confirmed using a 2100 Bioanalyzer (Agilent Technologies). RNA extracts were sent to Genome Québec (Montréal, Canada) for RNAseq library construction and sequencing on half a lane of Illumina Novaseq6000 in PE150. The raw reads were cleaned by removing adaptor sequences, empty reads and low-quality sequences (including reads with unknown sequences 'N') with Trimmomatic (Bolger et al., 2014). A taxonomic assignment was made with Kraken v.2 (Wood and Salzberg, 2014) from the reads to only keep the eukaryotic reads for the assembly. This taxonomic-based clean-up was conducted because gills of our target species contain

large amounts of endosymbiotic bacteria whose RNA was not totally discarded from the polyA hybridization technique associated with the library preparation. Following the Kraken filtering, around 5 to 10% of the initial reads of each individual were assigned to prokaryotes. The 90 to 95% remaining reads were then assembled in rnaSPAdes 3.13.1 using default parameters (Bankevich et al., 2012). The quality of the assembly was checked by looking at the remapping rate and by performing a busco analysis. After sequence assembly, the resulting contigs were used in the Galaxy pipeline AdaptSearch (repository https://**github**.com/abims-sbr/**adaptsearch**) to find orthologous sequences between species, estimate species divergence from the coding sequences and find genes under positive selection. The first step of this pipeline is to filter transcripts to only keep one sequence per gene using both the length and quality score of the transcripts as a proxy and to perform an additional CAP3 assembly (Huang and Madan, 1999). Transcripts of the different species were then put together in orthogroups with Orthofinder (Emms and Kelly, 2015). The sequences of orthogroups were aligned using BlastAlign (Belshaw and Katzourakis, 2005) and the reading frame was identified from the species alignment with the in-house python script CDSsearch (in the AdaptSearch suite) using Blastx against NCBI libraries. Finally, gene alignments were concatenated over the whole shared transcriptome and the alignment was used to reconstruct a phylogenetic tree with RAxML (Stamatakis, 2014). This alignment of coding sequences was then used to estimate the standard diversity index (π) within species and the absolute divergence (d_{XY}) between species using DnaSP 6.0 (Nei and Li, 1979; Rozas et al., 2017). The divergences accumulated at non-synonymous (d_N) and synonymous sites (d_S) along terminal branches and their ratio was calculated with CodeML (Yang and dos Reis, 2011). The values of d_A were estimated from d_{XY} and the values of the nucleotide diversity (π) found within each species from 2-3 individuals.

3- Results

3.1- Species distribution

The *Cox1* barcode assignment allowed us to get a better understanding of the distribution range for each species (Figure 22). *Alviniconcha kojimai* was the most abundant species sampled during the 2019 CHUBACARC expedition. It was present throughout the study area (with the exception of the Pacmanus field (Manus)) and dominates over the two other *Alviniconcha* species in the Futuna arc, and the Lau, and North Fiji Basins where it was present in 15 out of 18 populations samples (Figure 22; Table S1). *A. strummeri*, found in

only 7 population samples from the Futuna arc, and the Lau, and North Fiji Basins, was the most geographically restricted species. It was always mixed with at a low abundance *A. kojimai* and typically found at the periphery of snail patches. *A. boucheti* was sampled at all locations except the North Fiji Basin (Figure 22; Table S1). The species dominated *Alviniconcha* assemblages in the Manus and Woodlark Basins where it represented 71% in abundance of the samples (present in 8 populations samples) and was also the only species found in the PacManus field. Although mixed with *A. kojimai* at some locations, *A. boucheti* was found preferentially (in 60% of the samples; Table S1) on the wall of active hydrothermal chimneys as previously suggested in Beinart et al. (2012).

3.2- Morphology

We measured six shell traits from 700 *Alviniconcha* individuals. The distribution of the values of shell traits is shown in supplementary Figure S3. All measurements were significantly different between species (ANOVA, p-values < 0.001), with *A. boucheti* having on average a larger shell (at all measured traits) than *A. kojimai*, and *A. strummeri* being the smallest species. Only a small fraction of the variance in shell traits seemed to be related to species identity (see PCA axis 2 in Fig. S4). This was due to the low power in discriminating species using solely shell dimensions measured (see LDA in Fig. S5). Since the reclassification of species using the whole dataset was correct for only 75% of the individuals (and below 2 % for *A. strummeri* in particular), we did not evaluate further the capacity of the LDA using training and test subsets of the data.

By contrast, the shell ornamentation of bristles was clearly distinct between species except for a slight overlap between *A. kojimai* and *A. boucheti* (Figure 23). *A. strummeri* was characterised by only one category of bristles (type a) which corresponds to only long and uniformly sized spikes. *A. boucheti* was more polymorphic with several categories of rows (ab, b, bc, bcd, bd and abc) corresponding to rows showing repetitions of 0 to 3 smaller bristles between two long ones. Finally, *A. kojimai* was the most polymorphic species in terms of row categories but with a shifted number of smaller bristles from 2 to 6 between two subsequent long ones. Despite this heterogeneous arrangement of bristles in *A. boucheti* and *A. kojimai*, periostracal ornamentation only overlapped at the bristle arrangement bcd and d between these two species (Figure 23b), which characterised 30 *A. kojimai* and 43 *A. boucheti* individuals over their whole geographic range.

In addition, our morphological observation of the three species also indicated that *A. boucheti* was the only species displaying a single twisted columella as opposed to *A. kojimai* and *A. strummeri* which exhibited a double twisted columella, as previously reported by Laming et al. (2020) for individuals from the volcanic arc of Wallis and Futuna.

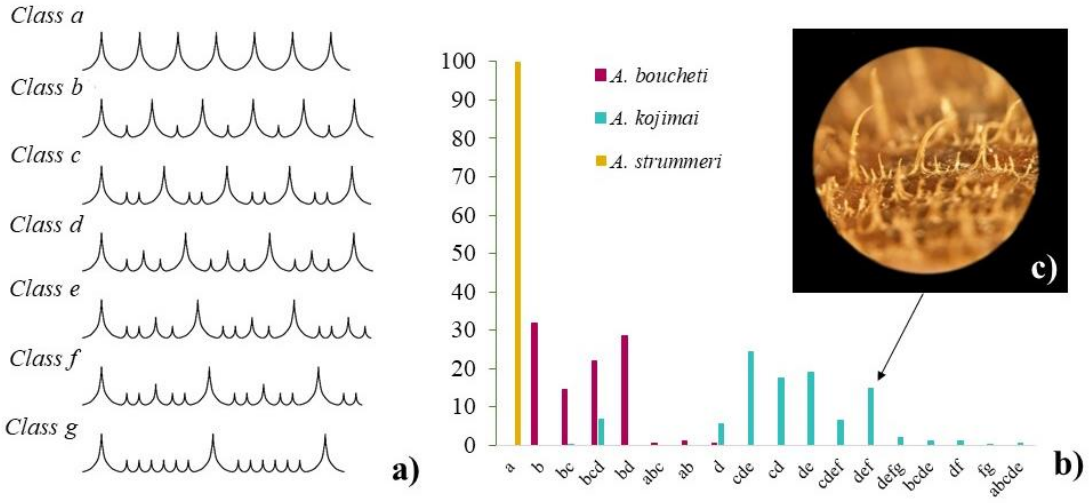


Figure 23: a) Arrangement of small, medium, and large bristles on *Alviniconcha* shells (detailed in Figure S6); b) Percentage of bristles arrangement categories observed in each *Alviniconcha* species. Different categories of bristle arrangements may be encountered on the shell of a given specimen depending on the row considered, which has been noted ab, bcd etc to account for such heterogeneity; c) Example of bristles arrangement for an individual of *Alviniconcha kojimai*.

3.3- Mitochondrial Cox1 gene analysis

An alignment of 599 bp was obtained for 722 individuals of *Alviniconcha* spp. for the mitochondrial *Cox1* gene. Haplotype diversity (*Hd*) ranged from 0.8 to 0.96 across species while nucleotide diversities (π) ranged from 0.003 to 0.008 (Table 3). The mitochondrial *Cox1* haplotype network (Figure 24) indicated that the three studied *Alviniconcha* species display complete mitochondrial lineage sorting over their whole geographic range, as previously reported (Breusing et al., 2020; Johnson et al., 2015). The net divergence d_A was smallest between *A. kojimai* and *A. strummeri* (d_A =8.6%) while the two other species pairs displayed d_A values 1.4 times greater (d_A = 11.8% and 12.3%; Table 4). The nearly fixed F_{ST} indices between the species pairs are shown in Table 4. The d_N/d_S ratios estimated between pairs of species was very low with values ranging between 0.003 and 0.014. The McDonald-Kreitman test performed between the three pairs of species was not significant, suggesting that this gene evolved neutrally during the separate evolution of the three species.

Table 3: Intraspecific genetic diversity of the three species for the three genomic datasets: *Cox1* mitochondrial gene, ddRAD-seq nuclear SNPs, and RNA-seq nuclear SNPs. *N*: sample size; *K*: number of polymorphic sites; *H*: number of haplotypes; *Hd*: haplotype diversity and π: nucleotide diversity.

	N	***K***	***H***	***Hd***	$\boldsymbol{\pi}$
Mitochondrial (Cox1 - 599 pb - 722 ind)					
A. kojimai	454	89	102	0.8	0.003
A. boucheti	243	67	62	0.91	0.004
A. strummeri	25	31	19	0.96	0.008
Transcriptomes (RNAseq - 1 186 131 pb - 7					
A. kojimai	2	10672	2	1	0.004
A. boucheti	3	10851	3	1	0.003
A. strummeri	2	12010	2	1	0.005
Genome (ddRAD-seq - 532 ind)					
A. kojimai	267	21397			0.0013
A. boucheti	228	40879			0.0014
A. strummeri	37	25801			0.0014

Table 4: Interspecific genetic divergence and differentiation for the three genetic datasets: *Cox1* mitochondrial gene (599 bp), ddRAD-seq SNPs (640 002 SNPs) and nuclear encoding RNAseq (1 186 131 bp). d_A: net divergence; d_{XY}: absolute divergence; F_{ST}: genetic differentiation; d_N: rate of non-synonymous substitutions; d_S: rate of synonymous substitutions.

	$\boldsymbol{d_A}$	$\boldsymbol{d_{XY}}$	$\boldsymbol{F_{ST}}$	$\boldsymbol{d_N/d_S}$	$\boldsymbol{d_N}$	$\boldsymbol{d_S}$
Mitochondrial (*Cox1* - 599 pb – 722 ind)						
A. kojimai/ A. boucheti	0.123	0.126	0.974	0.006	0.005	0.784
A. kojimai/ A. strummeri	0.086	0.091	0.961	0.015	0.007	0.468

A. boucheti/ A. strummeri	0.118	0.124	0.967	0.003	0.002	0.793
Transcriptomes (RNAseq - 1 186 131 pb - 7 ind)						
A. kojimai/ A. boucheti	0.028	0.031		0.133	0.013	0.097
A. kojimai/ A. strummeri	0.016	0.020		0.124	0.008	0.062
A. boucheti/ A. strummeri	0.027	0.031		0.134	0.013	0.096
Genome (ddRAD-seq - 640 002 SNPs - 532 ind)						
A. kojimai/ A. boucheti	0.031	0.032	0.922			
A. kojimai/ A. strummeri	0.018	0.018	0.842			
A. boucheti/ A. strummeri	0.031	0.032	0.917			

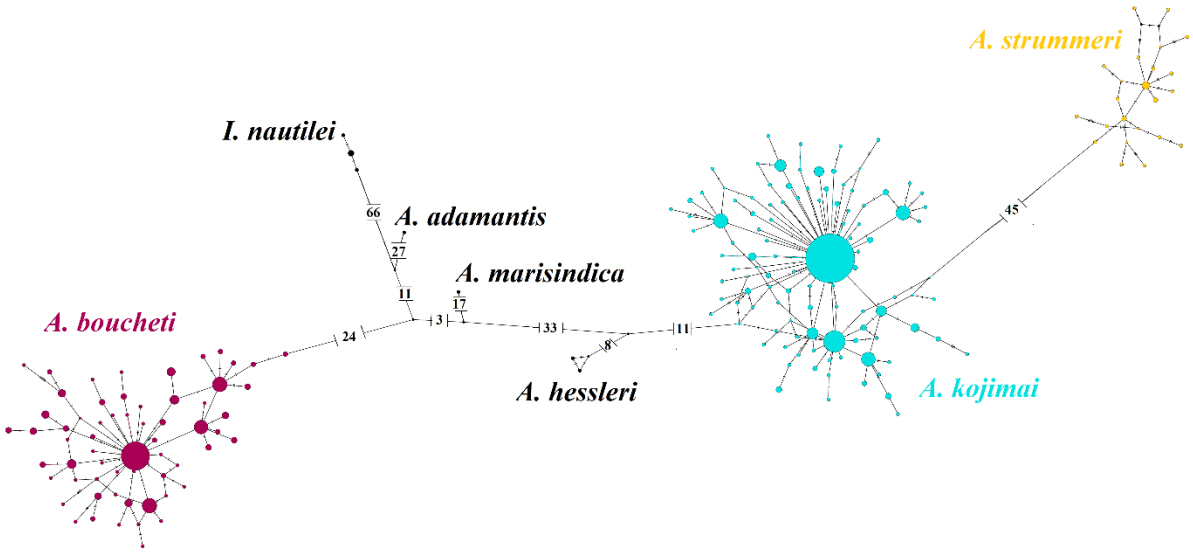


Figure 24: Haplotype network of 722 individuals of *Alviniconcha* spp for *Cox1* partial mitochondrial sequence. Circles represent individual haplotypes, while circle size is proportional to haplotype frequency. The numbers on the branches indicate the number of mutations between haplotypes. Purple: *A. boucheti*; Turquoise: *A. kojimai* and Yellow: *A. strummeri*.

3.4- ddRAD-seq analysis along the nuclear genome

The “raw” dataset (where a SNP is retained if genotyped in at least 80% of individuals within any species) contained 640 002 SNPs from 94 215 RAD loci. The distribution of these loci among species is shown in Figure 25. Almost 76% of the RAD loci sequenced in *A. strummeri* were also sequenced in at least one of the two other species, while this fraction was 57 % in *A. kojimai* and 33% in *A. boucheti*.

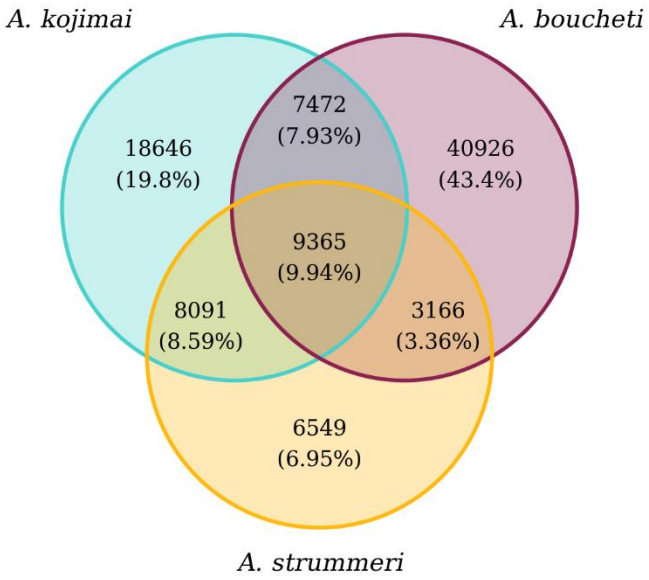


Figure 25: Number of RAD loci sequenced in each species (under the condition that a locus is considered in any species if it is sequenced in at least 80% of the individuals of that species). There are 94 215 RAD loci overall (640 002 SNPs). The 9 365 loci shared among all species were then further filtered to produce the final dataset used in downstream analyses.

Further filtering on missing data of the 9 365 loci in common across the three species led to a final dataset containing 60 577 biallelic SNPs on 4 377 RAD loci genotyped in 532 individuals. Among these SNPs, 63% were polymorphic within a single species while being fixed (differentially or not) in the two others (Figure 26a). Only 5 930 SNPs (9.8%) were polymorphic in all three species. However, these numbers strongly depend on sample size. With only 37 *A. strummeri* individuals genotyped, the smallest measurable allelic frequency at any SNP in this species is 0.0135, while it is much smaller for the two other species which have larger sample sizes. Looking at the distribution of polymorphisms after reclassifying the

SNPs as polymorphic only if they had a minimum allelic frequency of at least 0.0135, we see that the total polymorphism in *A. boucheti* (as shown in Fig. 26b) was due to relatively rare alleles, and there is in fact much more polymorphism shared between *A. strummeri* and *A. kojimai* than between *A. boucheti* and either species (Fig. 26b).

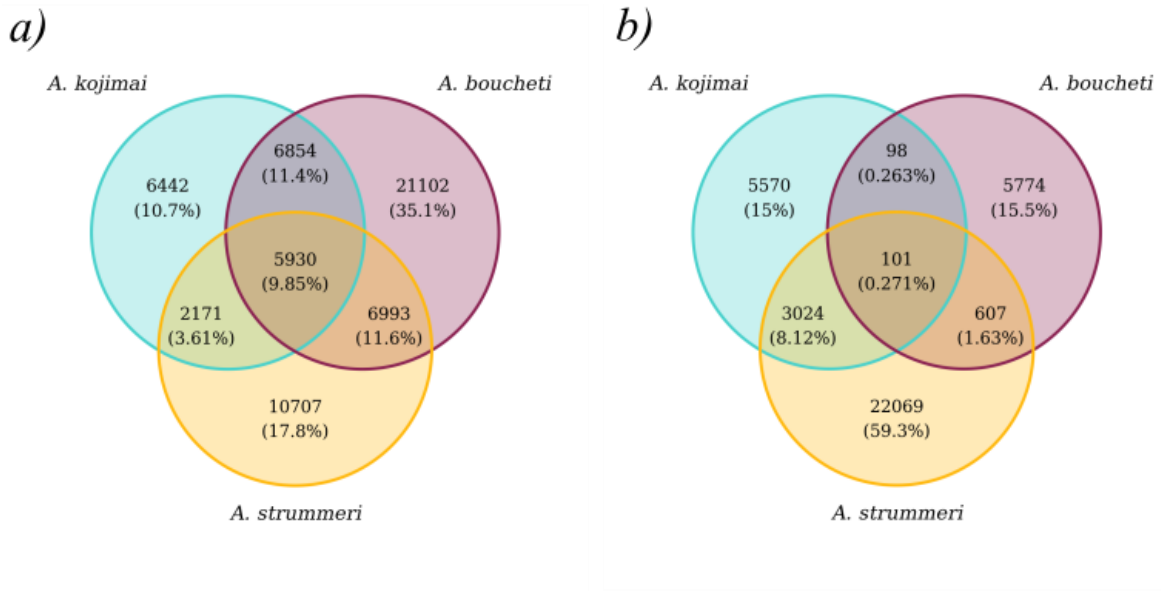


Figure 26: Distribution of polymorphism in the 60 577 SNPs shared by *A. kojimai* (*n*=267), *A. boucheti* (*n*=228), and *A. strummeri* (*n*=37). In panel a), the numbers exclusive to each circle show the number of SNPs that are polymorphic within one species and fixed in the two others (for instance, there are 21 102 SNPs that are polymorphic within *A. boucheti* only). Panel b) shows the same data but considering a SNP as polymorphic only if the allelic frequency of either allele was above 0.0135 (minimum allele frequency observable in *A. strummeri*, the species with the smallest sample size).

The nucleotide diversity (π) was the same between species even though the geographic distribution of *A. strummeri* was more restricted (0.0013 in *A. kojimai* and 0.0014 in *A. strummeri* and *A. boucheti*) (Table 3). The PCA analysis using all 60 577 SNPs clearly separated the three species of *Alviniconcha* along the first two principal components (explaining 51.1% and 17.3% of the total variance, respectively) (Figure 27). The net divergence (d_A) was 1.8% between *A. kojimai* and *A. strummeri* and about 1.6 times greater (3.1%) between the two other pairs of species. The total genetic differentiation (as estimated by F_{ST}) between populations of *A. kojimai* and *A. strummeri* was high (0.84), but lower than the nearly fixed F_{ST} values obtained for the two other species pairs (Table 4). These values were close to one as expected for species having nearly completed their allelic lineage sorting (Table 4; Figure S6). The admixture analysis clearly separated the three *Alviniconcha* species,

in agreement with the PCA analysis and results from the mitochondrial *Cox1* gene (Figures 24 and 27). The admixture bar plot (Figure 28), however, still exhibited traces of shared polymorphism or introgressed alleles between the three species. Indeed, some individuals displayed up to 20% of hetero-specific ancestry (especially in *A. strummeri*). In *A. strummeri*, individuals with the highest hetero-specific ancestry were collected at both Futuna and Lau but not at the Phoenix site on the North Fiji triple junction. In *A. kojimai* and *A. boucheti* individuals with the highest hetero-specific ancestry were also retrieved along the Futuna volcanic arc, and the Lau Basin where the two species are found in syntopy - *A. kojimai* from Futuna being 2.3 times more hetero-specific than the others - but not on the Pual ridge of the Manus Basin (Pacmanus, Fenway, Big Papi) where only *A. boucheti* is found.

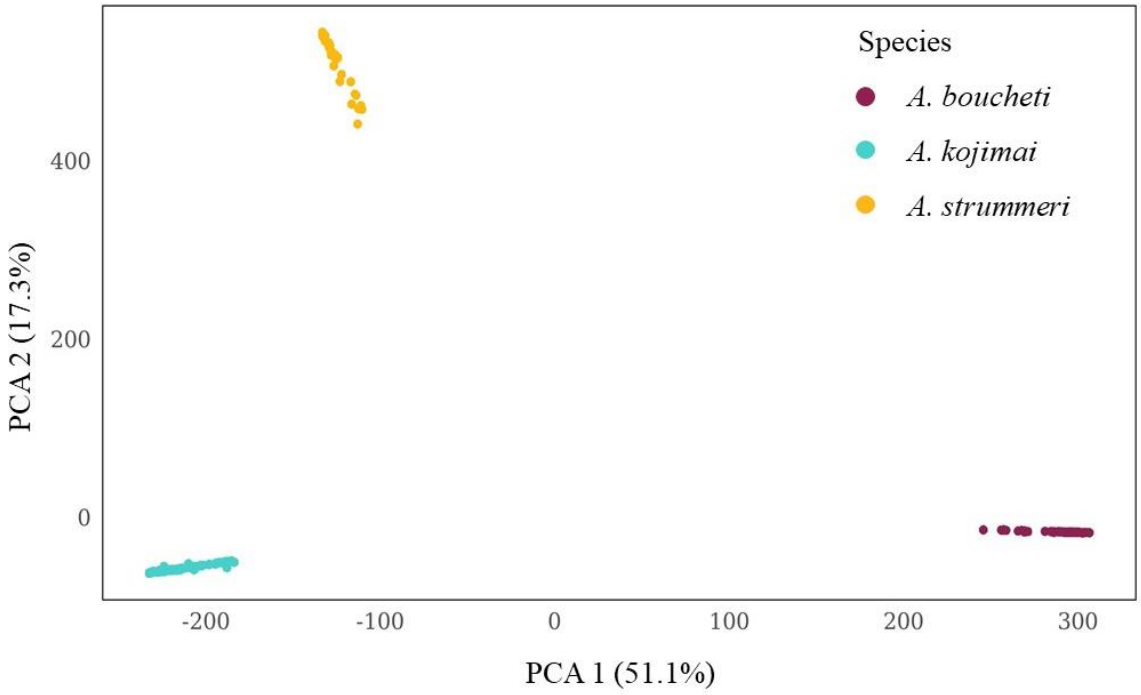


Figure 27: Principal components analysis of multilocus genotypes using 60 577 SNPs across 532 individuals of *Alviniconcha* spp. Purple: *A. boucheti*; Turquoise: *A. kojimai* and Yellow: *A. strummeri.* The attraction of individuals to the center of the PCA is due to missing data, despite the filters applied in R (see method section).

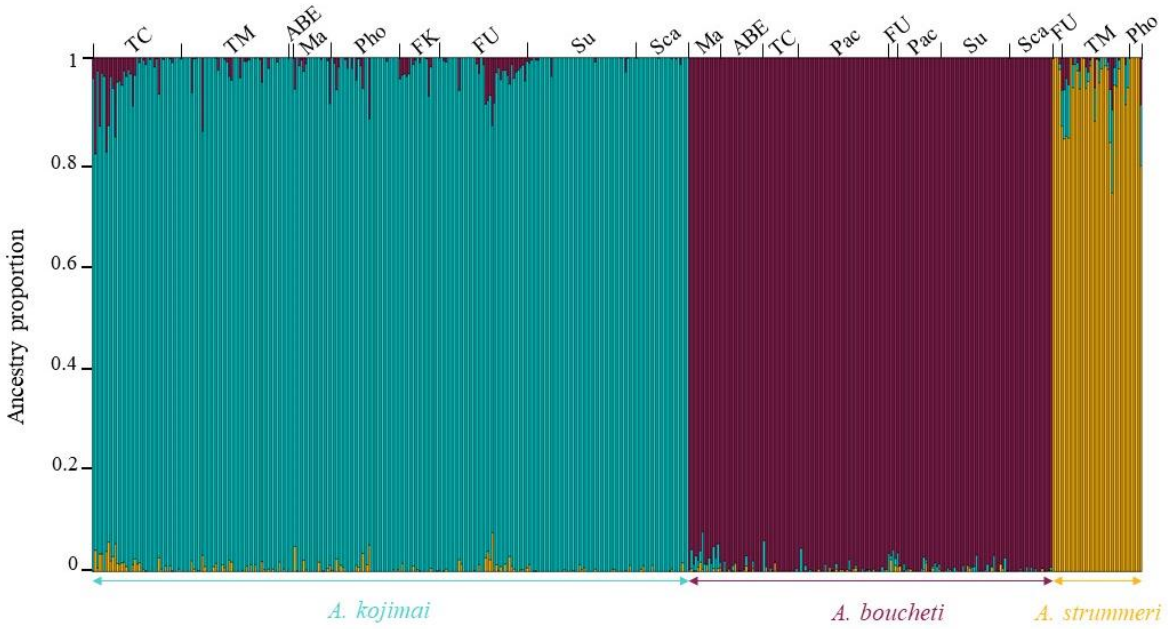


Figure 28: Ancestry coefficients bar plots representing *K*=3 for the three species of *Alviniconcha* sampled during the CHUBACARC expedition. This graphic was obtained with the snmf function in the package LEA (R software) on 60 577 SNPs for 532 individuals with 20 runs. Each vertical bar corresponds to one individual. The colours (Purple: *A. boucheti*, Turquoise: *A. kojimai*, Yellow: *A. strummeri*), were assigned according to the ancestry inferred for each cluster. Manus Basin: Susu (Su); Pacmanus (Pac). Woodlark Basin: Scala (Sca). North Fiji Basin: Phoenix (Pho). Futuna volcanic arc: Fatu-Kapa (FK); Fati-Ufu (FU). Lau Basin: Mangatolo (Ma); Tow Cam (TC); ABE (ABE); Tui Malila (TM).

Using DILS to test whether admixtures may be the result of incomplete lineage sorting or introgression, we found that the Secondary Contact model (SC) fitted the observed data better than any other model in all ten runs for *A. boucheti*/*A. kojimai* and *A. kojimai*/*A. strummeri* (see the probabilities in Table S2 and joint Site Frequency Spectrum in Figure 29). For the *A. boucheti*/*A. strummeri* species pair, SC was selected in 7 runs, while the three remaining analyses better refer to ancient migration (AM). Since DILS is thought to be more efficient at distinguishing between these two broad scenario than identifying further details (Fraïsse et al., 2021), we did not interpret further the results for this last species pair (although we note that goodness-of-fit was highest for a SC model with heterogeneous N_e and *me*, Table S2). All our species pair simulations were also done using a bottlenecked model of populations where the two sister populations were subjected to a size reduction after the ancestral population split. Introducing size reduction indeed clearly improved the goodness

of fit of all the most sophisticated models implementing linked selection (2N) and semi-permeable barriers to gene flow (2m).

For the two other species pairs (*A. boucheti*/*A. kojimai*; *A. kojimai*/*A. strummeri*), increasing the complexity of the SC model by adding a heterogeneous effective population size (2N) among loci improved the fit between simulations and observed data (in all but one run, Tables S2 and S3), suggesting that a non-negligible fraction of loci is affected by linked selection or other interferences. Variation in migration rates among loci across the genome (2m) was supported 7 times out of 10 for each species pair (Table S2) but not in the analysis yielding the best fit for both species pairs (Table S2). This can be explained by the fact that the portion of the genome still permeable to interspecific gene flow was probably very small between each pair of species.

The estimated demographic parameters of the best models (using the random forest prediction method) are summarised in Table S3, and a graphical representation of the models with the best goodness-of-fit is presented in Figure 30. Note that most parameter estimates depend on mutation rate, which was set to $\mu = 10^{-8}$ as suggested by Wares and Cunningham (2001) in *Littorina obtusata*, but the uncertainty of this parameter is large (Lynch, 2010). The effective size of the ancestral population was always higher than the effective sizes of the populations of the two daughter species after the split (Nf_1 and Nf_2 at T_{split}; i.e strong population bottlenecks). The size of daughter populations is then predicted to have increased at T_{dem} (time of expansion of the populations in this study) to reach current population sizes N_1 and N_2. Quantitative estimates of the timing of demographic change and population sizes vary strongly between runs and thus are quite imprecise. In most simulations the sizes of the current populations remain, however, much lower than that of the ancestral population (Table S3). This finding may suggest that species isolation was also linked with a reduction of species range to geographic isolates, or that the ancestral population size between pairs of species is overestimated due to the lack of intermediate species such as *A. hessleri*.

Separation between the ancestral population and the two daughter populations was old (for analysis yielding the best fit, T_{split}: around 141 340 and 133 200 generations for *A. boucheti/A. kojimai* and *A. kojimai/A. strummeri*, respectively) and preceded the period of population expansion (T_{dem1}: 35 140/ T_{dem2}: 10 200 and T_{dem1}: 25 740/ T_{dem2}: 55 500 generations for *A. boucheti/A. kojimai* and *A. kojimai/A. strummeri*, respectively). Then the period of recovery of gene flow occured between the populations (T_{sc}: about 6 900 and 12

000 generations for *A. boucheti/A. kojimai* and *A. kojimai/A. strummeri*, respectively). For each pair of species, the T_{sc}/T_{split} ratio was of the same order of magnitude indicating that the recovery of gene flow was very recent as compared to the formation of the species.

Despite an overall identical demographic pattern between the two pairs of species, there were slight differences (Figure 30). The ancestral effective size (N_a), as well as the post-split effective sizes (Nf_1 and Nf_2) in the *A. kojimai/A. strummeri* pair were much lower than in *A. boucheti/A. kojimai* (about half as much). Furthermore, although the T_{split} period was almost identical, the expansion of daughter population sizes was older in *A. boucheti/A. kojimai* to nearly reach contemporary effective sizes (N_1/ N_2). Since *A. kojimai* was present in both species pairs analysed, its current effective size could be estimated at about 130 000 individuals while the other two species had lower current effective sizes (about 52 500 individuals for *A. boucheti* and 57 115 individuals for *A. strummeri*). One of the main differences between the species pairs was the rate of migrants (*4.Ne.m*) exchanged between the species. Indeed, in the *A. boucheti/A. kojimai* split, this rate seemed to be low and bidirectional while for the *A. kojimai/A. strummeri* split, the migration was strongly directed from *A. kojimai* to *A. strummeri* (Table S3).

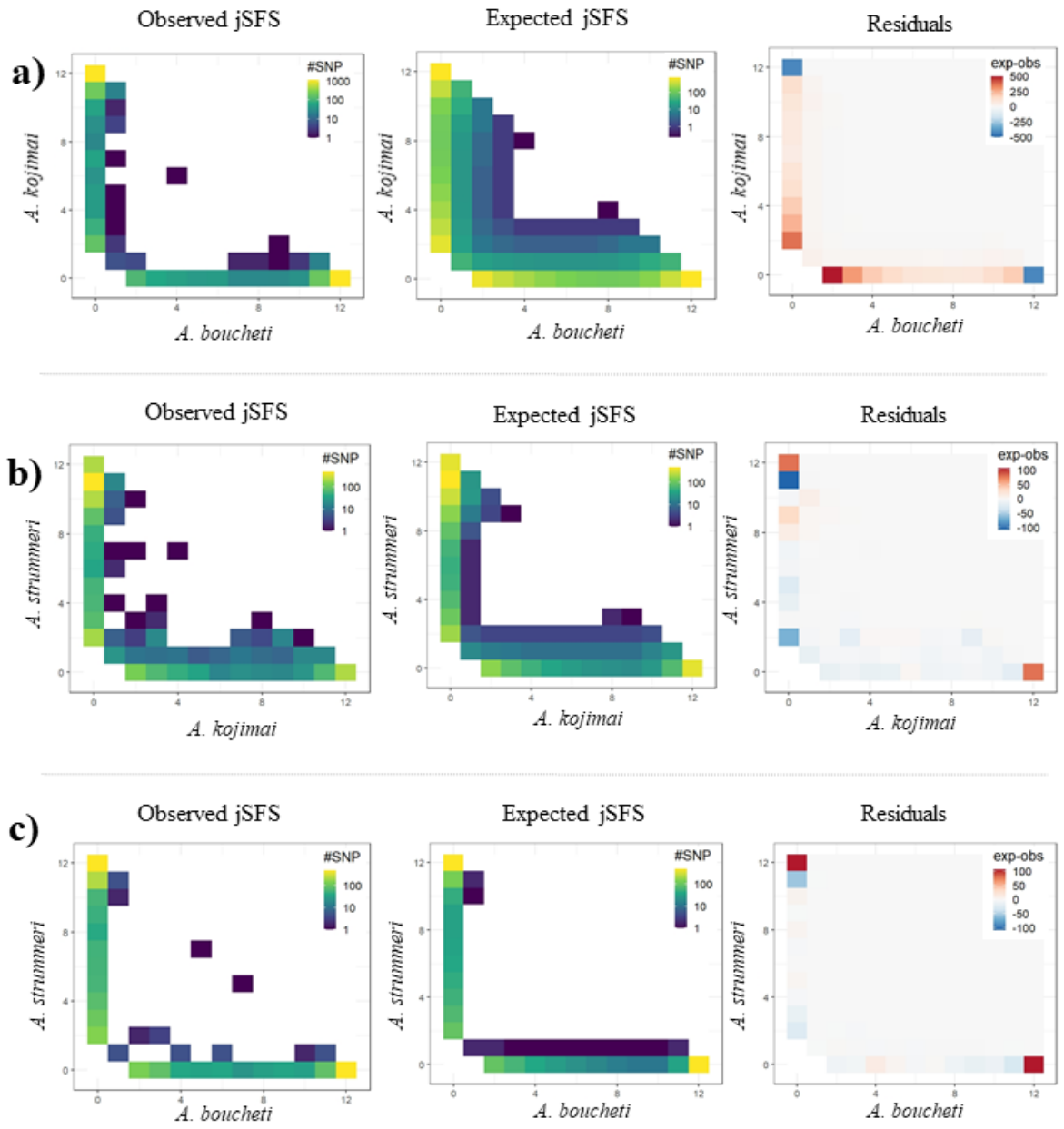


Figure 29: Comparison of the joint spectra of allele frequencies observed from the data (first column) and expected from the demographic model that provided the best fit in DILS (secondary contact with heterogeneous effective population size, second column) for the three pairs of species. The residuals (expected - observed) are displayed in the third column. a) *A. boucheti*/*A. kojimai*; b) *A. kojimai*/*A. strummeri* and c) *A. boucheti*/*A. strummeri*. For the first pair of species (a), the position in the spectrum corresponds to the allele frequencies in *A. kojimai* (on the y-axis) and *A. boucheti* (on the x-axis), and the colour scale represents the number of SNPs with these allele frequencies. The white squares show an absence of SNPs having these allelic frequencies.

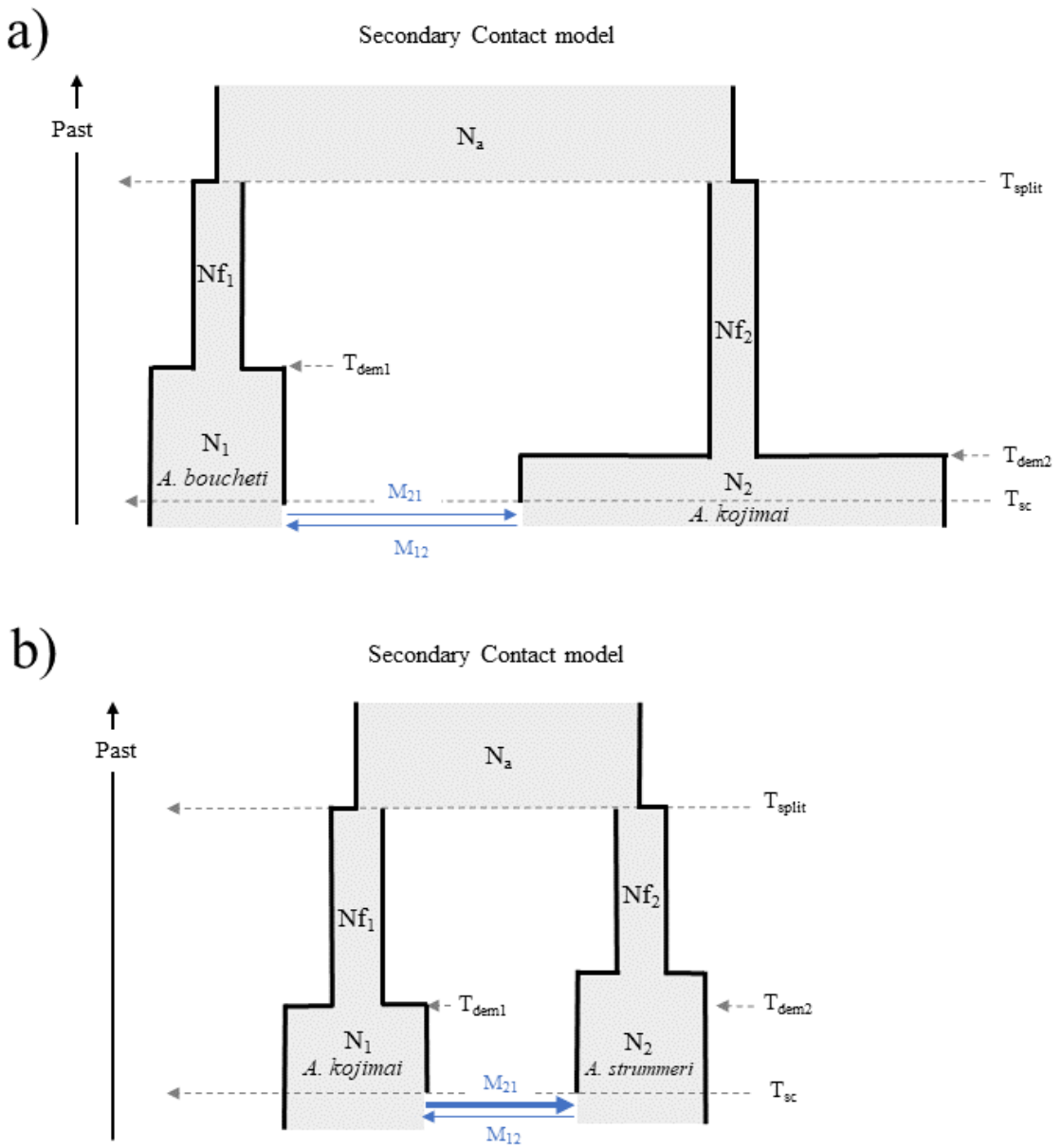


Figure 30: Demographic models that best explained the data observed in this study. N_1 and N_2: effective size of population 1 and 2; N_a: effective size of the ancestral population; Nf_1 and Nf_2 : effective size of population 1 and 2 after the split; T_{split}: time of split at which the ancestral population subdivides in two populations; T_{dem} : time of the expansion of the population; T_{sc}: time of secondary contact at which the two populations start exchanging genes; M_{12} and M_{21}: introgression rates from population 2 to 1 and from population 1 to 2, respectively (blue arrows). a) analysis for *A. boucheti*/*A. kojimai* and b) analysis for *A. kojimai*/*A. strummeri*.

3.5- Divergences in coding sequences

The concatenated coding sequence alignment of the three species' sets of transcripts corresponded to an overall alignment of 1 172 052 homologous nucleotide sites consisting of

1 705 partial CDSs in the right coding frame, with 809 317 conserved and 362 735 variated sites. The topology of the phylogenetic tree obtained is the same than that observed for the mitochondrial *Cox1* gene. The observed net divergence (d_A) between *Alviniconcha* species is of the same order of magnitude as those obtained with the ddRAD analyses: 1.6% between *A. kojimai* and *A. strummeri*, and around 2.8% between the two other species pairs (Table 4). The ratio of non-synonymous to synonymous substitutions (d_N/d_S), estimated using the free-ratio model M1 of CodeML implemented in AdaptSearch, was equal to 0.124 between *A. kojimai* and *A. strummeri* and 0.133 between the two other species pairs (Table 4). Although non-synonymous and synonymous divergences were almost identical for the species pairs *A. boucheti/A. kojimai* (d_N=0.0129, d_S=0.0972) and *A. boucheti/A. strummeri* (d_N=0.0128, d_S=0.0960) and about 1.6 times greater than the pair *A. strummeri/A. kojimai*, the branch leading to *A. boucheti* accumulated non-synonymous changes a little bit more than the two other *Alviniconcha* branches (d_N/d_S=0.166 vs d_N/d_S=0.150).

4- Discussion

4.1- Hydrothermal vents in the Western Pacific are home to three sympatric *Alviniconcha* species

Hydrothermal vent gastropods of the genus *Alviniconcha* have been found in the Indian and Western Pacific oceans, where a total of six species have been described (Breusing et al., 2020; Johnson et al., 2015). Although species were delineated using up to three mitochondrial and seven nuclear genes (Breusing et al., 2020; Johnson et al., 2015; Kojima et al., 2001), the most discriminating genetic marker was the mitochondrial *Cox1* gene (Breusing et al., 2020; Johnson et al., 2015). Focussing on the three species *A. boucheti*, *A. kojimai*, and *A. strummeri*, which inhabit the Western Pacific, we found that the genetic divergence measured across the genome was consistent with the *Cox1* mtDNA results used previously to delineate species. Reference transcriptome assemblies counting nearly 1.2 million homologous sites across the three species and a RAD-seq dataset composed of over 4 377 RAD loci sequenced in 37 to 267 individuals per species showed that the genetic divergence (measured as d_{XY}) was ca. 2% between *A. kojimai* and *A. strummeri* and ca. 3% between *A. boucheti* and the two other species (Table 4). These values are consistent with the levels of genetic divergence generally observed between reproductively isolated species (Roux et al., 2013). The SNP dataset (over 60 000 SNPs) derived from homologous RAD sequences also indicated that the genetic differentiation between species was strong (over 68% of the genetic

variance explained by the first two principal components, Fig. 27, and pairwise F_{ST} values between 0.84 and 0.92 from shared polymorphic loci, Table 4). Together with the mtDNA network and divergence estimates we obtained for the same set of individuals, all our results support the previous conclusions of, Beinart et al. (2012), Breusing et al. (2020) and Johnson et al. (2015) that *A. boucheti*, *A. kojimai*, and *A. strummeri* are genetically clearly distinct and at the end of their allelic lineage sorting (T>4*Ne* generations).

While early observations showed that individuals of different *Alviniconcha* species had very similar external morphologies, here we found that genetic divergence between species is accompanied by phenotypic differences. Cryptic species are common at hydrothermal vents (Matabos et al., 2011; Vrijenhoek, 2009) and *Alviniconcha* species have long been defined as such (Denis et al., 1993; Johnson et al., 2015; Warèn and Bouchet, 1993). Yet, our morphological analyses across the distribution range of the three species confirm and extend the recent findings of Laming et al. (2020) who described diagnostic morphological traits for these species in the hydrothermal fields of the Futuna arc. Laming et al. (2020) showed that they differed in the shape of the columella, radula, and snout, but also in the arrangement of shell bristles. Building on their work on bristle arrangement, we confirm that individuals of *A. strummeri* in the Futuna arc and North Fiji and Lau Basins all had a regular arrangement of bristles of identical length. In addition, we observe that the bristle arrangements for the other two species are more variable, but with only two common types (type "bcd" and “d” in Figure 23). We also found that the specific "double-twisted" columella (or columellar fold) described by Laming et al. (2020) for *A. kojimai* and *A. strummeri* in the Futuna arc is a criterion that extends throughout these species’ range. It is therefore possible to identify *Alvinoncha* species from the Western Pacific by combining observations of bristle arrangement and columella shape. These observations can be made on board without dissection. In contrast, external shell measurements appear to be poorly discriminant (Figure S2 and S3).

Our study brought new information regarding the distribution of *Alviniconcha* in the Western Pacific. As recently reported, *A. kojimai* was present in the Futuna arc, Lau and North Fiji Basins (Johnson et al., 2015; Laming et al., 2020) but also at the entrance to the Manus Basin on Susu Volcanoes and Woodlark Basin (Figure 22). It was absent from the Pacmanus area (Pual Ridge) of the Manus Basin, where *A. boucheti* was the only species found. *A. boucheti* was found in the Lau and Manus Basins and near the Futuna volcanic arc (Fati Ufu and Mangatolo fields). In contrast to the previous study by Johnson et al. (2015), we did

not find this species at the White Lady/Ivory Tower sites in the North Fiji Basin (both of which being now extinct), and we did not found it in the Northern part of the newly discovered Phoenix vent site (North Fiji Triple Junction). However, noting its sporadic presence in this basin from previous reports is important to understand where the different species might have met contemporaneously (see Divergence History section below). It also highlights the speed at which species distribution can change (at least locally), and thus the difficulty in interpreting the overlap of observed ranges at a given time.

A. kojimai and *A. boucheti* thus have largely overlapping distributions, *A. kojimai* being the most abundant in the study area (where it represented 72% of the individuals we sampled) except in the Manus Basin (where *A. boucheti* represented 71% of the sampled individuals). Interestingly, this large-scale overlap in the species distribution range reflects local coexistence at finer spatial scales. These two species were often found in syntopy (that is, a mixture of individuals within the same site, and even in a given patch). For instance, we found two samples with the two species in the Lau Basin, one in the Manus Basin (Susu), and one at a Futuna arc site. Both species were also found in syntopy at the base of hydrothermal vent chimneys at 3 388 m depth at the newly discovered site "La Scala" on the Woodlark Basin (Boulart et al., submitted). Based on these observations, these two species seem to follow a longitudinal gradient where *A. boucheti* is currently expanding from the West to the East, and *A. kojimai* in the opposite direction. Although mixed with *A. kojimai* at some locations, *A. boucheti* was preferentially found (in 60% of the samples) on the walls of active hydrothermal chimneys as previously suggested in Beinart et al. (2012).

The distribution of *A. strummeri* appeared more restricted. It occurred at a relatively low abundance at several diffuse vent sites of the Southern area of the Lau Basin (Tow Cam, Tu'i Malila and ABE), but also, although in lower abundance, on venting sites in the Futuna volcanic arc and North Fiji Basin (Figure 22 and Table S1). When present, *A. strummeri* was always found in syntopy with at least one of the two other species (mostly *A. kojimai*). However, *in situ* observations suggest that it may be more peripheral to vent emissions and thus perhaps less well sampled than the other two species (data not shown).

This study confirms that the three species of *Alviniconcha* gastropods coexist at small spatial scales in the Western Pacific (all three species were even found to co-occur simultaneously at the ABE vent site in the Lau Basin). In our collection boxes, 42% of samples contained at least 2 species (i.e. a mixing of individuals within a few tens of cm^2). At

the scale of a vent site, this proportion reached 60% (i.e. 9 out of 15 sampled sites). This species coexistence in sympatry and even in syntopy raises interesting questions about the ecological preferences of these species and the role of geography and habitat in their evolution. To shed light on the biogeographic history of these gastropods, we conducted further analysis of their genomic divergence and polymorphism.

4.2- A long history of divergence in allopatry

All genomic datasets (mtDNA *Cox1*, transcriptomes, and nuclear RAD-seq) indicated that the divergence (measured either as d_{XY} or d_A, Table 4) between *A. boucheti* and the ancestor of the two other species was ca. 1.4 to 1.6 times greater than the divergence between *A. kojimai* and *A. strummeri*. This is further confirmed by the fixation rates of non-synonymous (d_N) and synonymous mutations (d_S) in the three pairs of species, both of which indicating that *A. boucheti* was ca. 1.6 times more divergent than the two other species in the two coding compartments (mtDNA *Cox1* and transcriptomes). The older separation of *A. boucheti* from the two other species is also supported by the fact that it differs from its two congenerics by a different shell columella shape and a distinct endosymbiosis with *Campylobacteria* (Beinart et al., 2012, 2014; Breusing et al., 2020). The almost equal and high divergence between *A. boucheti* and the two other species in all genomic compartments suggests that mutations have been accumulated at the same rate between the two other species or that the demography of *A. kojimai* and *A. strummeri* was very similar. These results therefore suggest that the genetic divergence between species mostly reflects the time elapsed since speciation events. This finding is important because it is a critical assumption when using fossil-calibrated phylogenies for dating lineage splitting, as was done recently by Breusing et al. (2020).

With mutations accumulating globally at a constant rate along lineages, ratios of divergence between species (e.g. $d_{A,sp1}$ / $d_{A,sp2}$) appear relatively constant when measured from distinct genomic datasets, as explained above. By contrast, point divergence values (e.g. $d_{A,mtDNA}$ and $d_{A,nuDNA}$) are expected to differ due to differences in how these genomic regions are affected by evolutionary forces. Here we found that mtDNA divergence was ca. 3.8 to 4.8 times greater than nuDNA divergence (see d_A and d_{XY} values in Table 4). It is unlikely that this discrepancy is due to the fourfold difference in effective population size N between mitochondrial and nuclear loci, because intraspecific diversity (reported as π) was low compared to the divergence. This means that d_A and d_{XY} are nearly equivalent, and the time

since species split (> 25 Ma according to Breusing et al. (2020)) has long past the time needed for allelic lineage sorting between the three species (~8Ne). Rather, our results are consistent with the neutral ratio of mitochondrial to nuclear mutation rate observed in non-vertebrates (typically around 5; Havird and Sloan, 2016), although this ratio varies widely in mollusks (from 1.4 in Fissurelloidea to 91.9 in *Cristataria* (Duda, 2021), see also Allio et al. (2017)). In absence of strongly biased sex ratios, the contrast in evolutionary rate between mitochondrial and nuclear genomes could also be explained by differences in the strength of background selection. Here, based on d_N/d_S values averaged over the three species, the strength of background selection was 16 times greater in the mitochondrial genome when compared with the nuclear one. This fits well with the previous hypothesis by Havird and Sloan (2016) that a higher mitochondrial mutation rate could generate more positive selection for compensatory changes in nuclear genes interacting with the mitochondrial genome and/or that selective sweeps are more frequent in the mitochondrial genome while promoting divergence in the latter compartment.

Unexpectedly, the divergence measured from coding vs non-coding nuclear regions did not differ much (transcriptomes vs RAD sequences, Table 4). Purifying selection acting on the 3D structure and function of the encoded proteome instead predicts a lower divergence in coding regions (Fontanillas et al., 2017). Our results therefore suggest that loci have not been chosen randomly along the genome but rather in well-conserved areas probably associated with coding regions. This is possible since we had to retain only the RAD sequences that were shared between the three species (which represented 10% of all RAD sequences, while 70% were amplified in a single species). Diversity and divergence estimate from RAD data may therefore be biased downwards (a caveat that cannot be avoided when looking at strongly divergent species by selecting a subset of their genomes).

Regarding the effect of selection on transcriptomes, we found that the non-synonymous fraction of the species divergence represented about 13% of the total observed divergence. This value is similar to that found in some marine species (0.10-0.20, Gayral et al., 2013) and slightly lower than those obtained for mammals (0.20-0.25, Arbiza et al., 2006; Ohta and Ina, 1995). The non-synonymous divergence for hydrothermal *Alviniconcha* species is however much higher than those found in hydrothermal alvinellids (0.01-0.04, Fontanillas et al., 2017). These thermophilic worms indeed seem to experience stronger purifying selection due to functional constraints associated with the thermal denaturation of their proteins (Fontanillas et al., 2017). This result demonstrates that very low d_N/d_S ratios (like in alvinellids) are not

representative of all vent taxa and that species experiencing less thermal constraints such as *Alviniconcha* (Podowski et al., 2009, 2011), exhibit d_N/d_S ratios similar to other marine species. In future studies, it will be interesting to look at the d_N/d_S ratio between species for each gene to check for traces of disruptive (habitat) selection.

The analysis of genetic divergence between our three *Alviniconcha* species yielded one final intriguing observation. The mtDNA/nuDNA divergence ratio between *A. kojimai* and *A. strummeri* ($d_{A,mtDNA,strummeri} / d_{A,nuDNA,kojimai} \approx 4.6$) seemed slightly greater than between the two other species pairs (ca. 3.9). Although we cannot rule out the possibility that the strength of selection could have been greater on the mitochondrial genome of *A. kojimai* and *A. strummeri* after their separation, another hypothesis is that the nuclear genome of these two species experienced allelic rejuvenation through introgressive hybridization. Although the genetic differentiation found between *A. kojimai* and *A. strummeri* on nuDNA was strong (F_{ST} = 0.84), it was not nearly fixed as found for the two other pairs (F_{ST} = 0.92 in both cases, Table 4). This is in line with the higher level of shared polymorphism between these species: over 8% of the SNPs genotyped in all three species were polymorphic in both *A. kojimai* and *A. strummeri*, while this figure dropped below 2% in the two other pairs. These observations suggest that *A. kojimai* and *A. strummeri* did not reach reciprocal monophily despite their high level of divergence on nuclear and mitochondrial genomes, which would be compatible with low levels of inter-specific gene flow. We used ancestry analyses and demographic inferences to explore this hypothesis among other historical scenarios of divergence for the three *Alviniconcha* species.

4.3- Historical scenarios of divergence

Results from ancestry analyses are compatible with the hypothesis of gene flow between the three *Alviniconcha* species, with traces of hetero-specific ancestry reaching up to 20% in some individuals (Figure 28). The alternative hypothesis of incomplete lineage sorting is unlikely given the high levels of divergence found in the mitochondrial compartment and the fact that most individuals displayed different levels of mixed ancestry depending on geography and the level of species mixing at relevant locations: traces of admixture in all species were mostly located in the Lau Basin and Futuna arc, where the level of species mixture is also highest.

Demographic inferences performed with DILS supported the same conclusion: 27 of 30 runs (including models with the best fit) found that heterospecific gene flow occurred

relatively recently after a long period of isolation between species. With this method, it is however difficult to date speciation events and secondary contacts in terms of number of generations, as it depends directly on an unknown mutation rate that could range from 5×10^{-10} to 3×10^{-8} (Scally and Durbin, 2012; Tine et al., 2014). With this range of mutation rates, the divergence time between *A. kojimai* and *A. strummeri* (estimated from the model with the best support) would be anywhere between 45 000 and 2.7 million generations. By contrast, using a clock calibration based on fossil records, Breusing et al. (2020), estimated this split at about 25 Ma. The same comparison for the *A. boucheti* /*A. kojimai* split leads to 47 113 - 2.8 million generations (DILS) *vs* 48 Ma (Breusing et al., 2020). These values must be considered with caution as they are difficult to reconcile. Dating the separation of *A. kojimai* and *A. strummeri* to 2.7 millions generations on one hand, and 25 Myr on the other hand, leads to an approximate generation time of ca. 9 years, but all other estimates assuming a higher mutation rate lead to higher, unrealistic, generation times for vent species with high growth rate and early sexual maturation (Tyler and Young, 1999). As a consequence, reconciling these analyses would probably lead to either slow down the nuclear mutation rate of vent species by analogy to what was previously proposed for the mitochondrial *Cox1* gene (Chevaldonne et al., 2002; Johnson et al., 2006, 2015) or to revise these geotectonic- and/or fossil-driven molecular calibration dates.

One hypothetical scenario is that the collision of the Ontong-Java plateau with the Melanesian arc about 18 Ma disrupted gene flow along the formerly well-connected South-Fiji and Solomon ridges, thereby promoting allopatric diversification of the vent fauna in this region (Schellart et al., 2006). This collision led to the simultaneous rotation of the Vanuatu arc and the Fiji Islands triggering the opening of 1) the North Fiji proto Basin about 10 Ma (Hall, 2002), 2) the Woodlark Basin (one of the oldest basins in the Western Pacific), about 6 Ma as a result of continental rifting (i.e. the thinning of a tectonic plate due to stretching forces creating a volcanic zone; Hall, 2002; Taylor et al., 1995), and 3) the Manus Basin 3.5 to 4 Ma (Auzende et al., 2000). Finally, the expansion of the seafloor in the centre of the North Fiji proto Basin led to the opening of several ridge systems from which the Lau Basin was the most recently formed while expanding to the South, about 1 or 2 Ma (Hall, 2002). Given that all our estimations converge towards secondary contact representing ca 5% of the time eluded since populations split (an estimate that is independent from mutation rates), it is thus possible that secondary contact between the three species happened during the formation of the Lau Basin.

Under this hypothesis, the newly opened Lau Basin could have been subsequently colonised by a series of different allopatric species coming from the older Manus, Woodlark and North Fiji ridge systems. We can therefore hypothesise that the opening of the Lau Basin near the active zone of Futuna allowed the mixing and hybridization of the three *Alviniconcha* species, starting 1.2 Ma ago but with different timings (secondary contacts being more recent between *A. boucheti* and the two other species than between *A. strummeri* and *A. kojimai*). This hypothesis also accounts for our observation that most introgressed individuals were found in the Lau band Futuna volcanic arc (Figure 28) where the three species sometimes co-occur. This finding however does not fit well with larval dispersal modelling (Breusing et al., 2019; Mitarai et al., 2016) which rather predict that inter-basin exchanges are more likely occurring stepwise via the North Fiji Basin, suggesting that rare interspecific exchanges are either much older than thought or not specifically performed where the introgressed migrants are recruiting.

We were not expecting inter-specific exchange between the three species given the high level of mitochondrial and nuclear divergences. Usually, nuclear divergences greater than 2-3% should limit introgression between previously separated species due to the accumulation of genetic incompatibilities (Roux et al., 2016). Based on genetic and morphological differences, previous studies on *Alviniconcha* species suggested that they were reproductively isolated. However, these studies were based on a few mitochondrial and highly conserved nuclear genes (Breusing et al., 2020; Johnson et al., 2015). This is not the first time that traces of introgression have been observed between genetically distant species (Bouchemousse et al., 2016; Muths et al., 2009; Roux et al., 2013). In fact, hybridization still occasionally occurs between *Ciona robusta* and *Ciona intestinalis* despite a high transcriptome synonymous sequence divergence (14%, Roux et al., 2013) and 12-14% of mitochondrial *Cox3-Nd1* divergence (Bouchemousse et al., 2016). In gastropods and echinoderms, cases also exist (Hamilton and Johnson, 2015; Pinceel et al., 2005), as in the pulmonate gastropod genus *Rhagada* where two species diverging by 30.2% on the *Cox1* gene, are still able to produce sterile F1 (Hamilton and Johnson, 2015) or, in the ophiuroid genus *Acrocnida* where cryptic species diverging by 19% on the *Cox1* gene are still able to locally hybridise and introgress in very sheltered habitats (Muths et al., 2006).

The maintenance of distinct species against gene flow is possible when reproductive isolation barriers efficiently prevent genome remixing (Coyne and Orr, 2004). In *Alviniconcha*, pre-zygotic barriers are almost certainly a strong obstacle to gene exchange.

While we have shown here that the species are sometimes in contact, they are often found in different habitats (although at small spatial scales) because of the metabolic requirements of their symbiotic bacteria (Beinart et al., 2012, 2014). Here we found *A. kojimai* and *A. strummeri* preferentially on diffuse venting zones while *A. boucheti* was more frequently found on the wall of hydrothermal chimneys. In addition, post-zygotic barriers due to a long history of divergence in allopatry such as maladaptation (Hauser, 2002) and sterility or inviability of hybrids (Brannock and Hilbish, 2010; Maheshwari and Barbash, 2011) are likely to play a role in maintaining divergence between species. These barriers remain to be investigated.

5- Conclusions

This study showed the co-occurrence of three divergent *Alviniconcha* species (*A. kojimai*, *A. boucheti* and *A. strummeri*) across the five back-arc Basins of the Western Pacific Ocean. The number of accumulated non-synonymous and synonymous substitutions between *A. boucheti*/*A. kojimai* and *A. boucheti*/*A. strummeri* was nearly identical on all genomic datasets, suggesting that divergence is proportional to the time since species separation. However, admixture analyses and demographic inferences clearly supported a scenario in which the three species evolved without gene flow for a long period of time (different geographic origins), followed by a relatively recent secondary contact with resumption of gene flow despite the strong accumulated divergence between species. These secondary contacts could coincide with the opening of the Lau Basin from 1.2 Ma onwards. With more than 60% of substitutions fixed between species, genetic (i.e. post-zygotic) barriers to gene flow are likely to be an important factor in reproductive isolation. Finally, this study confirms that the species of the genus *Alviniconcha* can be distinguished using morphological characters.

Supplementary material of Chapter 1:

Table S1: Specimens used for each genetic analysis. Details are also available from NCBI Biosamples SAMN22059315-SAMN22060075 and SAMN22155691-SAMN22155716.

Sample	Site	Field	Basin	Habitat	Nbr Ind. *A. kojimai*	Nbr Ind. *A. boucheti*	Nbr Ind. *A. strummeri*	Not identified	Longitude	Latitude	depth (m)	Date	Analysis
721-GBT1	Tow Cam		Lau	diffuse zone	33		1		176°08'15,4" W	20°19'04,4" S	2716	2019	COI/ddRAD/ morphology
721-GBT6	Tow Cam		Lau	diffuse zone	44	5			176°08'15,8" W	20°19'05,1" S	2711	2019	COI/ddRAD/ morphology
721-GBT7	Tow Cam		Lau	chimney		20			176°08'12,7" W	20°18'59,2" S	2714	2019	COI/ddRAD/ morphology
	Tow Cam		Lau	diffuse zone		2					2715	2009	COI/RNAseq
722-GBT7	Tui Malila		Lau	diffuse zone	31		5		176°34'04,2" W	21°59'15,2" S	1899	2019	COI/ddRAD/ morphology
722-GBT1	Tui Malila		Lau	diffuse zone	33		19	3	176°34'05,9" W	21°59'21,2" S	1886	2019	COI/ddRAD/ morphology
722-GBT5	Tui Malila		Lau	diffuse zone	32		15		176°34'05,5" W	21°59'21,4" S	1884	2019	COI/ddRAD/ morphology
	Tui Malila		Lau	diffuse zone	2		2				1885	2009	COI/RNAseq
731-GBT3	ABE		Lau	chimney	2	26	1	1	176°11'28,9" W	20°45'47,1" S	2149	2019	COI/ddRAD/ morphology

	ABE		Lau	diffuse zone		1					2145	2009	COI/RNAseq
726-GBT4	Mangatolo		Lau	diffuse zone		26			174°39'12,7" W	15°24'52,8" S	2031	2019	COI/ddRAD/ morphology
726-GBT3	Mangatolo		Lau	chimney		4			174°39'12,5" W	15°24'52,7" S	2031	2019	COI/ddRAD/ morphology
726-PBT6	Mangatolo		Lau	diffuse zone	18				174°39'19,9" W	15°24'57,7" S	2039	2019	COI/ddRAD/ morphology
726-GBT2	Mangatolo		Lau	diffuse zone	3				174°39'12,6" W	15°24'52,5" S	2031	2019	COI/ddRAD/ morphology
724-GBT4	Phoenix		North Fiji	diffuse zone	39				173°55'7,6" E	16°57'0,0" S	1961	2019	COI/ddRAD/ morphology
724-PBT4	Phoenix		North Fiji	diffuse zone	33		9		173°55'4,7" E	16°56'57,8" S	1973	2019	COI/ddRAD/ morphology
727-GBT2	AsterX	Fatu Kapa	Futuna	diffuse zone	26				177°09'07,9" W	14°45'06,5" S	1562	2019	COI/ddRAD/ morphology
727-GBT4	Stephanie	Fatu Kapa	Futuna	chimney	16		1		177°09'57,6" W	14°44'14,7" S	1547	2019	COI/ddRAD/ morphology
728-GBT2		Fati Ufu	Futuna	chimney	27		3	2	177°11'07,0" W	14°45'35,8" S	1519	2019	COI/ddRAD/ morphology
728-PBT4		Fati Ufu	Futuna	diffuse zone	40		1		177°11'04,9" W	14°45'35,3" S	1519	2019	COI/ddRAD/ morphology
728-GBT6		Fati Ufu	Futuna	chimney	3	4			177°11'05,9" W	14°45'35,2" S	1518	2019	COI/ddRAD/ morphology
733-GBT2	Big Papi	Pacmanus	Manus	chimney		24			151°40'20,1" E	3°43'43,9" S	1708	2019	COI/ddRAD/ morphology

733-GBT8	Fenway	Pacmanus	Manus	diffuse zone		24	151°40'22,4" E	3°43'41,2" S	1696	2019	COI/ddRAD/ morphology
733-GBT9	Solwara8	Pacmanus	Manus	chimney		13	151°40'27,5" E	3°43'49,3" S	1737	2019	COI/ddRAD/ morphology
733-PBT7	Solwara8	Pacmanus	Manus	chimney		5	151°40'26,6" E	3°43'50,1" S	1734	2019	COI/ddRAD/ morphology
734-GBT9		Pacmanus	Manus	chimney		28	152°6'2,8" E	3°43'17,2" S	1659	2019	COI/ddRAD/ morphology
736-GBT3	Solwara1	Susu	Manus	chimney	24		152°5'47,0" E	3°47'22,1" S	1505	2019	COI/ddRAD/ morphology
736-GBT10	North Su	Susu	Manus	diffuse zone	22		152°6'2,8" E	3°47'56,0" S	1218	2019	COI/ddRAD/ morphology
737-GBT10	South Su	Susu	Manus	diffuse zone		24	152°6'18,6" E	3°48'35,0" S	1353	2019	COI/ddRAD/ morphology
737-PBT5	South Su	Susu	Manus	chimney	17	13	152°6'17,5" E	3°48'29,8" S	1300	2019	COI/ddRAD/ morphology
737-GBT7	South Su	Susu	Manus	diffuse zone		24	152°6'17,9" E	3°48'31,8" S	1343	2019	COI/ddRAD/ morphology
738-GBT10	Scala		Woodlark	chimney	24		155°03'09,6" E	9°47'56,7" S	3388	2019	COI/ddRAD/ morphology
739-GBT10	Scala		Woodlark	chimney	24		155°03'07,0' E	9°47'56,3' S	3344	2019	COI/ddRAD/ morphology
739-PBT5	Scala		Woodlark	chimney		24	155°03'08,1" E	9°47'56,0" S	3353	2019	COI/ddRAD/ morphology

Table S2: Likelihood of demographic models in the hierarchical comparison of alternate models in DILS. The probability (P.) of ongoing migration is the likelihood of current migration vs. current isolation. P. SC gives the support for secondary contact against isolation with migration (within the context of ongoing migration). The probability (P.) of ongoing isolation is the likelihood of current isolation vs. current migration. P. AM gives support for ancient migration against strict isolation (within the context of ongoing isolation). The last four columns give the likelihood of heterogeneous/ homogeneous effective population size and migration across loci along the genome.

Species 1	Species 2	P. Ongoing migration	P. Ongoing isolation	P. SC	P. AM	P. N-heterogeneous	P. N-homogeneous	P. M-heterogeneous	P. M-homogeneous
A. boucheti	*A. kojimai*	0.93		0.95		0.65		0.42	
A. boucheti	*A. kojimai*	0.87		0.96			0.57	0.60	
A. boucheti	*A. kojimai*	0.95		0.95		0.56			0.51
A. boucheti	*A. kojimai*	0.94		0.96		0.59		0.46	
A. boucheti	*A. kojimai*	0.94		0.91		0.58		0.45	
A. boucheti	*A. kojimai*	0.96		0.95		0.65			0.54
A. boucheti	*A. kojimai*	0.87		0.95		0.63			0.60
A. boucheti	*A. kojimai*	0.91		0.89		0.67		0.51	
A. boucheti	*A. kojimai*	0.94		0.93		0.79		0.47	
A. boucheti	*A. kojimai*	0.90		0.90		0.65		0.48	
A. kojimai	*A. strummeri*	0.97		0.87		0.70		0.62	
A. kojimai	*A. strummeri*	0.97		0.80		0.67		0.59	
A. kojimai	*A. strummeri*	0.98		0.90		0.84		0.65	

A. kojimai	*A. strummeri*	0.98		0.89		0.81			0.55		
A. kojimai	*A. strummeri*	0.97		0.88		0.79			0.51		
A. kojimai	*A. strummeri*	0.98		0.90		0.72				0.60	
A. kojimai	*A. strummeri*	0.97		0.88		0.71			0.64		
A. kojimai	*A. strummeri*	0.99		0.86		0.74			0.47		
A. kojimai	*A. strummeri*	0.98		0.78		0.77				0.52	
A. kojimai	*A. strummeri*	0.97		0.88		0.71				0.59	
A. boucheti	*A. strummeri*	0.90		0.92		0.66				0.47	
A. boucheti	*A. strummeri*	0.88		0.93			0.74		0.61		
A. boucheti	*A. strummeri*		0.89		0.60	0.70					
A. boucheti	*A. strummeri*	0.90		0.88		0.63				0.49	
A. boucheti	*A. strummeri*	0.87		0.87		0.65				0.39	
A. boucheti	*A. strummeri*	0.89		0.81		0.68			0.68		
A. boucheti	*A. strummeri*	0.87		0.90		0.59			0.43		
A. boucheti	*A. strummeri*	0.90		0.82		0.80			0.56		
A. boucheti	*A. strummeri*		0.90		0.72	0.78					
A. boucheti	*A. strummeri*		0.92		0.65	0.81					

Table S3: Demographic parameters estimates under the Secondary Contact model. N_1 and N_2: effective size of population 1 and 2; N_a: effective size of the ancestral population; Nf_1 and Nf_2 : effective size of population 1 and 2 after the split calculated by Na x founders1 or Na x founders2 ; shape_N_1 and shape_N_2: shape parameter α (resp.β) of the Beta (α,β) distribution for Ne; T_{split}: time of split at which the ancestral population subdivides in two populations (in generations); T_{dem} : time of the reduction of the effective size population; T_{sc}: time of secondary contact at which the two populations start exchanging genes (in generations); M_{12} and M_{21}: introgression rates from population 2 to 1 and from population 1 to 2, respectively (in number of migrants per generation); For each analysis performed, the index between brackets represents the species population.

Species 1	Species 2	Euclidean distance	N_a	Nf_1	Nf_2	N_1	N_2	Shape _N_1	Shape_ N_2	T_{split}	T_{dem1}	T_{dem2}	T_{sc}	T_{sc}/T_{split}	T_{am}	M_{12}	shape_ M12_a	shape_ M12_b	M_{21}	shape_ M21_a	shape_ M21_b
A. boucheti	*A. kojimai*	0.95	266 855	16 529	2 057	62 340	49 340	1.41	0.72	139 580	5 440	87 920	3 480	0.03		1.32	1.27	3.21	1.88	0.90	0.73
A. boucheti	*A. kojimai*	1.52	221 825	10 577	22 611	33 110	347 690	1.46	1.86	185 580	80 740	5 320	2 960	0.02		2.24	1.35	2.55	1.55	3.45	3.40
A. boucheti	*A. kojimai*	1.08	291 415	13 702	1 769	396 915	53 375			96 060	6 660	75 280	7 100	0.07		0.49	1.03	2.01	1.02	0.51	1.39
A. boucheti	*A. kojimai*	0.78	263 905	11 250	10 786	52 500	280 335	2.06	1.70	141 340	35 140	10 200	6 900	0.05		0.72			0.71		
A. boucheti	*A. kojimai*	0.99	262 445	7 797	20 072	38 640	298 155	2.37	2.21	154 160	41 980	7 820	10 040	0.07		0.72			0.42		
A. boucheti	*A. kojimai*	0.98	269 830	11 203	15 159	42 260	81 105	3.02	2.00	145 640	40 340	24 020	2 700	0.02		1.61	1.81	3.90	1.93	0.91	2.58
A. boucheti	*A. kojimai*	1.21	237 620	11 301	1 138	37 105	39 755	2.79	3.98	156 620	50 300	144 420	2 340	0.02		2.00	4.66	2.82	1.91	1.49	2.07
A. boucheti	*A. kojimai*	1.32	266 985	29 350	19 698	20 315	110 000	2.38	1.46	151 780	120 060	14 340	5 380	0.04		1.41	1.65	3.47	0.75	1.24	0.94
A. boucheti	*A. kojimai*	1.32	238 975	12 111	19 347	33 020	90 130	1.93	2.49	172 500	31 940	23 560	2 680	0.02		1.82	3.23	0.94	2.52	1.39	3.08
A. boucheti	*A. kojimai*	1.11	283 840	68 343	10 130	18 190	51 420	1.86	0.77	122 100	122 700	49 680	1 440	0.01		2.93			3.72		
A. kojimai	*A. strummeri*	0.78	130 795	11 040	4 816	103 615	61 535	1.55	2.74	126 720	22 620	69 980	10 320	0.08		0.36	1.62	1.89	4.18	4.04	2.14
A. kojimai	*A. strummeri*	0.94	145 945	14 254	10 947	112 745	59 170	3.40	2.37	102 560	18 020	42 540	7 300	0.07		0.34	2.76	1.25	4.94	4.56	1.60
A. kojimai	*A. strummeri*	0.92	115 855	14 612	4 124	52 910	46 635	1.49	3.21	143 220	35 460	103 940	7 940	0.06		0.48	4.39	1.15	5.15	3.28	2.79

A. kojimai	*A. strummeri*	0.76	114 260	7 936	21 847	64 615	169 210	1.28	2.35	143 180	60 080	8 860	8 540	0.06		0.68	0.82	2.94	4.72	4.10	0.20
A. kojimai	*A. strummeri*	1.03	118 035	5 361	10 031	83 595	108 870	1.61	4.05	106 720	33 340	36 180	5 700	0.05		1.22	2.92	0.96	5.57	5.05	1.00
A. kojimai	*A. strummeri*	0.78	127 500	11 676	12 719	53 885	67 510	2.25	4.69	125 280	49 580	42 680	8 480	0.07		0.48	0.76	1.50	4.74	4.27	3.60
A. kojimai	*A. strummeri*	0.75	109 300	12 043	12 553	64 980	74 600	1.40	1.95	136 620	37 640	35 720	8 240	0.06		0.33			4.62		
A. kojimai	*A. strummeri*	0.81	124 230	6 130	17 370	54 150	104 685	1.83	3.48	124 080	67 360	15 460	10 660	0.09		0.36	2.25	3.35	3.86	4.23	1.07
A. kojimai	*A. strummeri*	1.35	156 270	10 328	6 952	492 275	84 105	1.56	1.50	74 640	10 260	29 780	6 880	0.09		0.33			5.58		
A. kojimai	*A. strummeri*	0.72	131 210	12 654	7 038	82 960	57 115	1.55	4.70	133 200	25 740	55 500	12 000	0.09		0.37			4.08		
A. boucheti	*A. strummeri*	1.31	195 030	13 428	30 349	36 935	32 095	3.45	0.26	158 020	38 220	62 500			82 760	1.91	2.55	5.12	3.14	3.17	2.91
A. boucheti	*A. strummeri*	1.16	157 765	10 731	12 132	54 930	25 145	1.73	3.17	113 160	24 360	39 520	2 060	0.02		1.24			3.75		
A. boucheti	*A. strummeri*	1.53	182 395	16 649	21 539	546 525	1 010 655	4.48	2.27	140 740	7 660	20			63 980	2.11			3.33		
A. boucheti	*A. strummeri*	1.21	151 795	10 821	15 923	46 850	26 230	2.37	1.27	112 820	23 820	41 560	2 960	0.03		0.56	2.72	1.04	2.63	2.33	0.98
A. boucheti	*A. strummeri*	1.29	181 940	10 338	69 598	44 435	16 660			70 920	15 720	73 960	2 260	0.03		0.93	1.59	2.52	2.11	1.75	1.85
A. boucheti	*A. strummeri*	1.23	231 740	17 420	18 963	27 395	28 245	4.22	2.16	232 520	62 080	88 860			104 420	4.14			3.11		
A. boucheti	*A. strummeri*	1.01	167 975	1 925	12 828	36 065	22 980	1.65	0.67	91 580	74 900	47 720	2 320	0.03		0.68			2.77		
A. boucheti	*A. strummeri*	1.01	163 150	2 842	30 770	39 650	15 625	2.14	3.88	107 360	85 360	104 780	2 760	0.03		0.50	4.66	3.89	2.44	1.16	1.53
A. boucheti	*A. strummeri*	1.41	145 360	14 251	22 573	34 770	26 695	3.15	2.77	139 300	49 780	3 220	6 120	0.04		0.39			1.27		
A. boucheti	*A. strummeri*	1.26	167 780	17 548	21 056	83 455	21 185	2.84	1.93	108 140	360	5 760	3 200	0.03		0.72	3.67	1.92	2.19	4.01	4.51

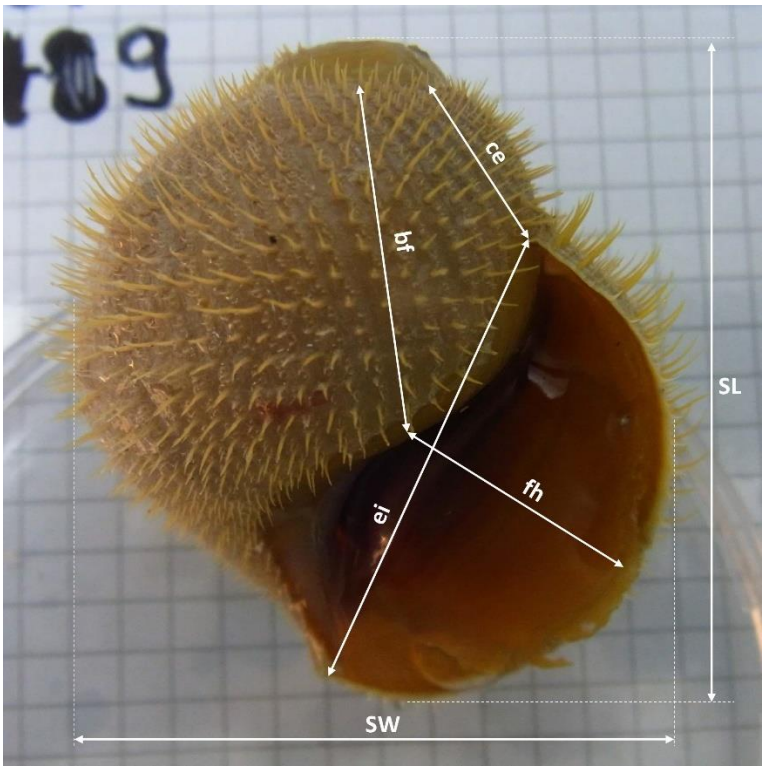


Figure S1 : Shell traits measured with a calliper on *Alviniconcha* individuals during the CHUBACARC expedition from Chiu et al. (2002).

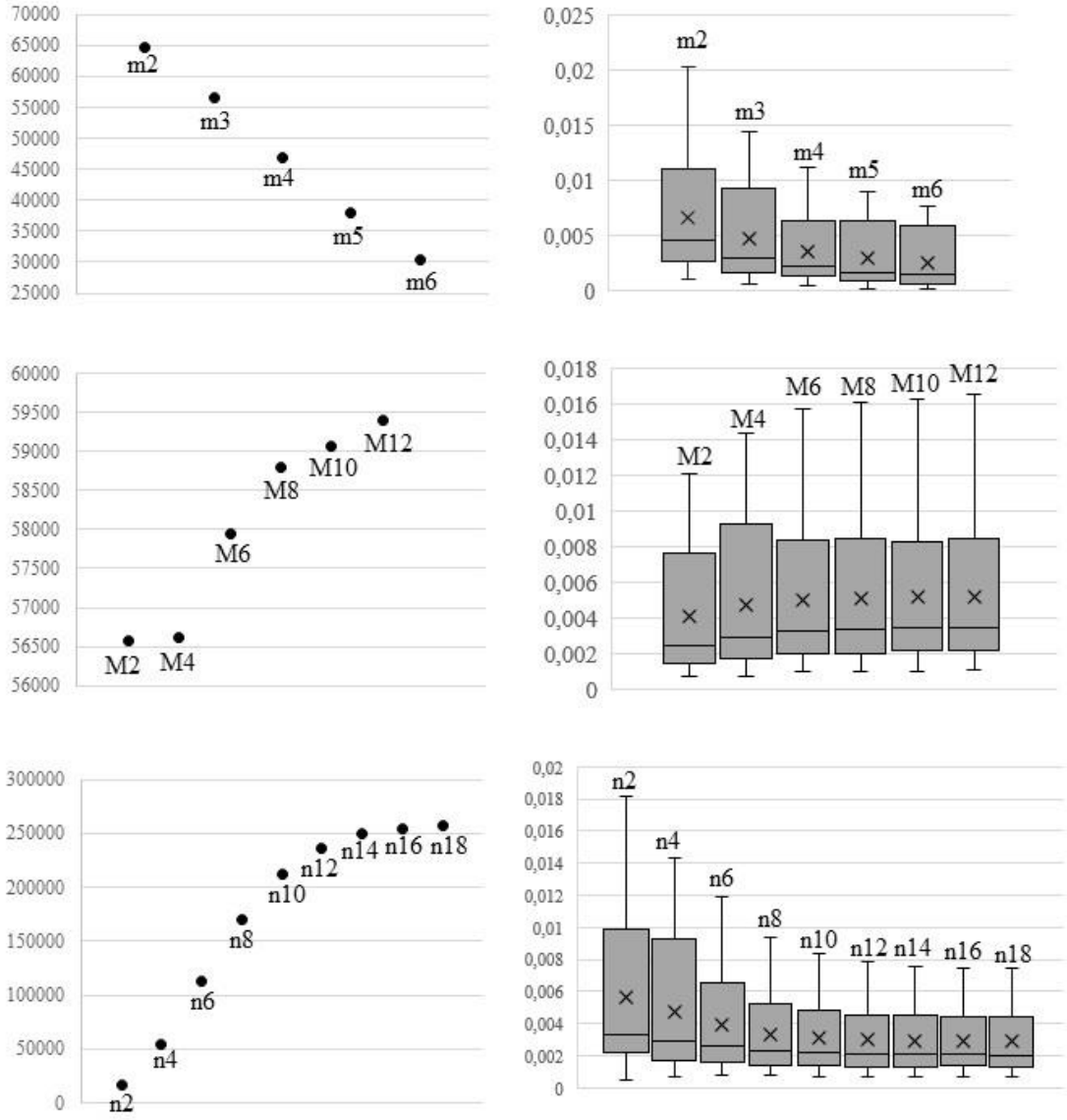


Figure S2: Number of conserved SNPs and error rate as a function of *n*, *m* and *M* in Stacks.

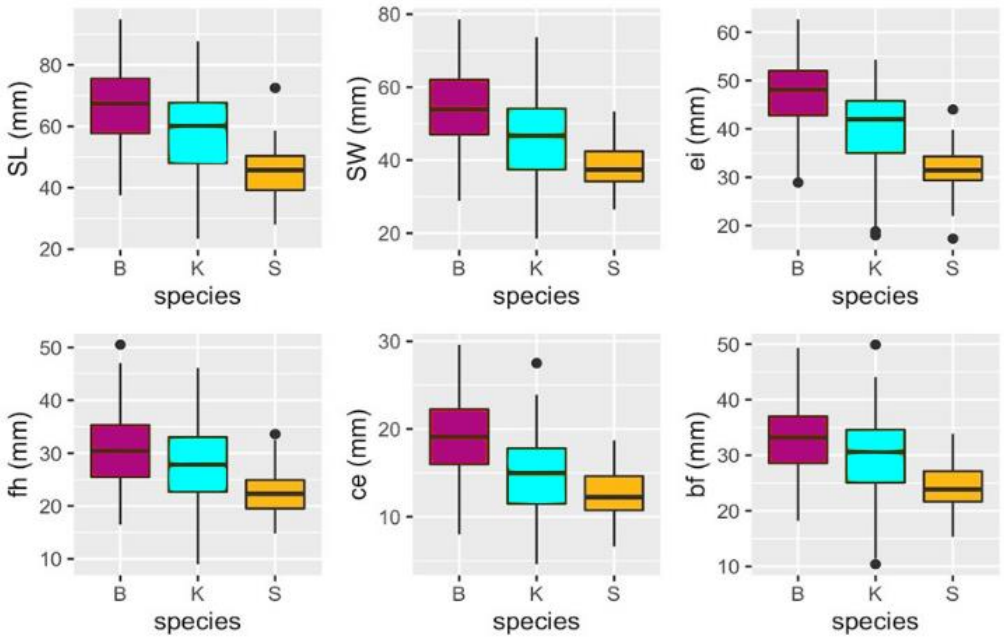


Figure S3: Distribution of six morphometric variables measured in *Alviniconcha boucheti* (purple, n=247 ind.), *A. kojimai* (turquoise, n=409), and *A. strummeri* (yellow, n=44).

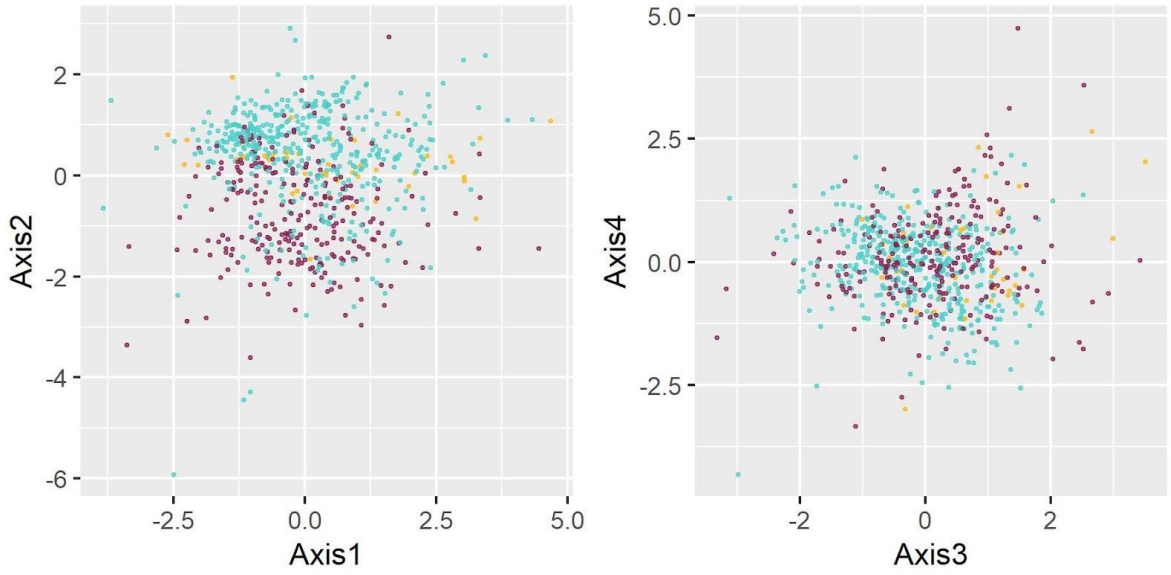


Figure S4: Principal component analysis of five transformed morphometric variables (see text) for *Alviniconcha boucheti* (purple, n=247 ind.), *A. kojimai* (turquoise, n=409), and *A. strummeri* (yellow, n=44). The first four components shown here explain 31.2%, 25.4%, 18.1%, and 16.5% of the variance, respectively. The second principal component seems to be linked mildly with species identity.

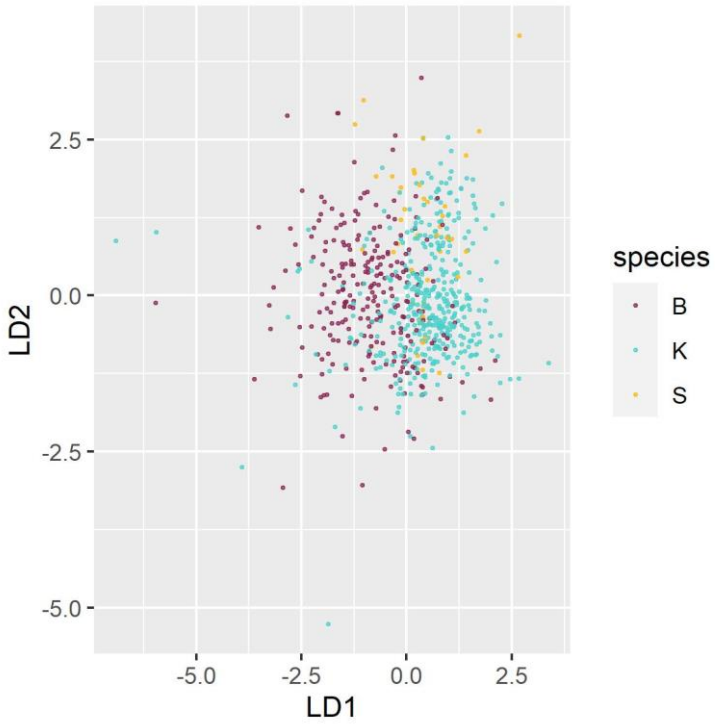


Figure S5: Linear discriminant analysis of species identity based on five transformed morphometric variables (see main text) for *Alviniconcha boucheti* (purple, n=247 ind.), *A. kojimai* (turquoise, n=409), and *A. strummeri* (yellow, n=44). The combination of linear discriminants only allows a partial reclassification of species (75%).

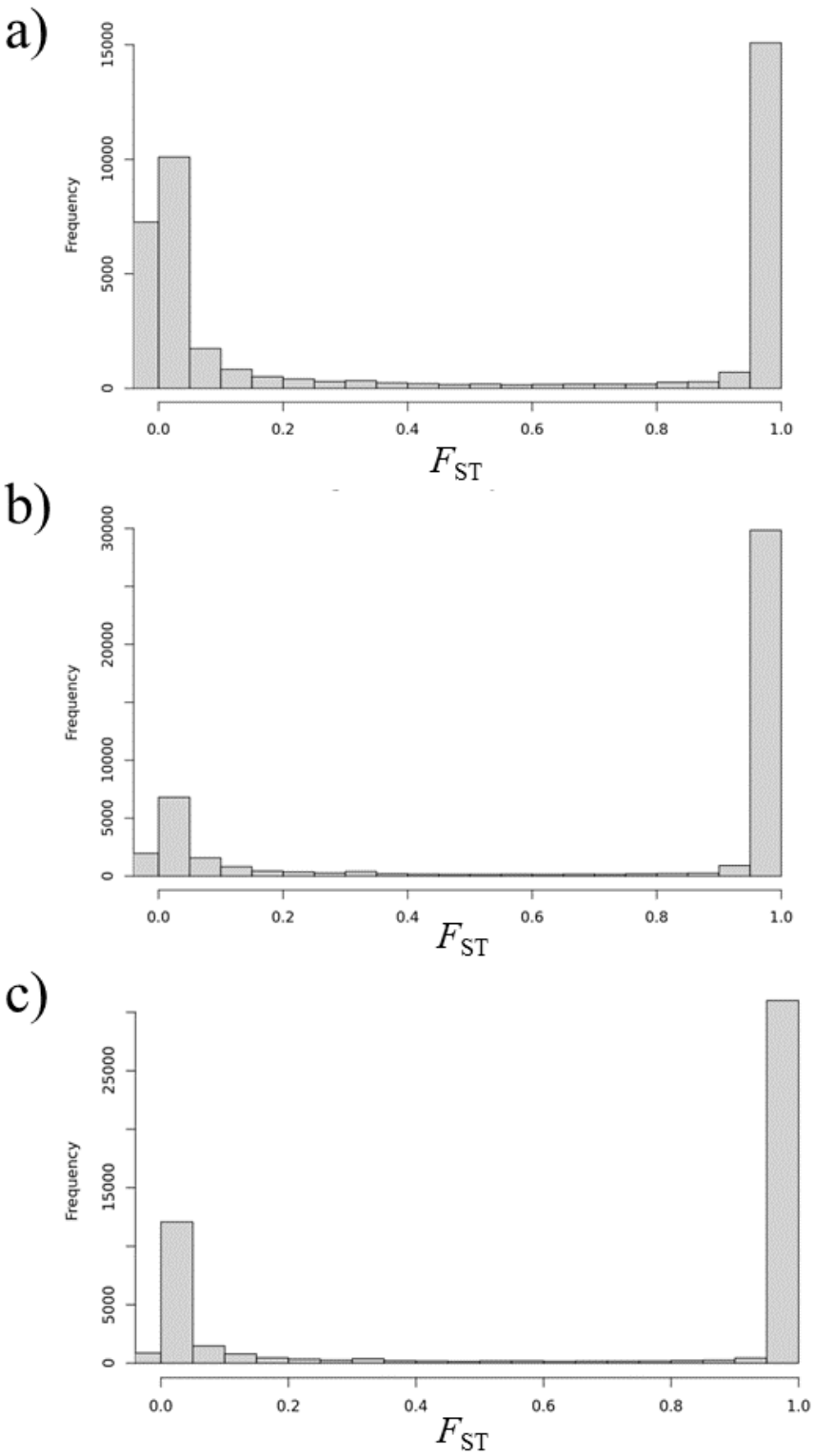


Figure S6: Distribution of F_{ST} values by species pair, a) *A. kojimai/A. strummeri* on 39 533 SNP, b) *A. boucheti/ A. strummeri* on 45 514 SNP and c) *A. kojimai/ A. boucheti* on 50 093 SNP.

Example of detailed parameter settings for *A. boucheti/A. kojimai* in DILS analysis :

infile: /beegfs/data/dils/8YVpfpJuxy/DILS_fichier.fasta

region: noncoding

nspecies: 2

nameA: *A. boucheti*

nameB: *A. kojimai*

nameOutgroup: NA

useSFS: 1

lightMode: FALSE

config_yaml: /beegfs/data/dils/8YVpfpJuxy/8YVpfpJuxy.yaml

timeStamp: 8YVpfpJuxy

population_growth: variable

modeBarrier: beta

max_N_tolerated: 0.2

Lmin: 144

nMin: 12

mu: 1e-08

rho_over_theta: 0.1

N_min: 100

N_max: 1000000

Tsplit_min: 100

Tsplit_max: 1000000

M_min: 0.4

M_ma

Appendix 1: Brief account of our exploration of the origin of unexpected "double-peaks" observed on the mitochondrial gene *Cox1* sequences

This annex reports the work that I have carried out to try understanding the observation of double peaks (DP) (Figure 31) on the sequences of the mitochondrial gene *Cox1*. This 709-bp gene fragment was amplified with the Folmer primer pair, LCO1490 (5'-GGTCAACAAATCATAAAGATATTGG-3') and HCO2198 (5'-TAAACTTCAGGGTGACCAAAAAATCA-3'). Out of 722 sequenced individuals, 48 individuals displayed double peaks, including 41 *A. kojimai* and 7 *A. strummeri* (these individuals were discarded from all the analyses presented in this PhD thesis). The absence of DPs in *A. boucheti* should be noted. These DPs can represent up to 71 positions on the *Cox1* sequence and mainly correspond to diagnostic sites between *Alviniconcha* species. We have manually separated the minor sequence in 11 individuals, i.e. at all positions with a double peak, the minor allele was kept to get an additional sequence. As can be seen in Figure 32, these reconstructed sequences are found in the centre of the *Cox1* haplotype network. This means that they do not correspond exactly to diagnostic alleles between *Alviniconcha* species.

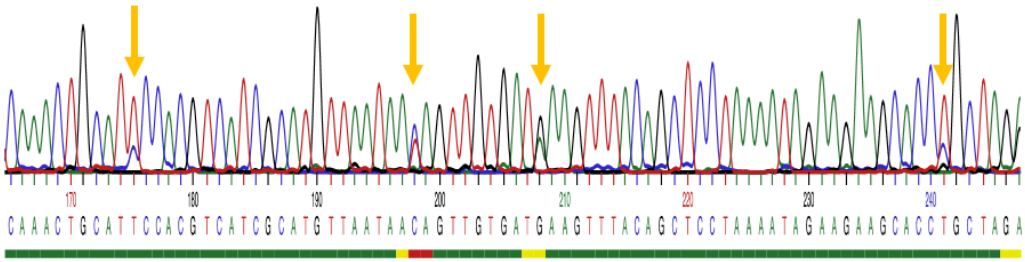


Figure 31: Chromatogram of a *Cox1* sequence showing the presence of double peaks (yellow arrows) in an individual of *A. kojimai*.

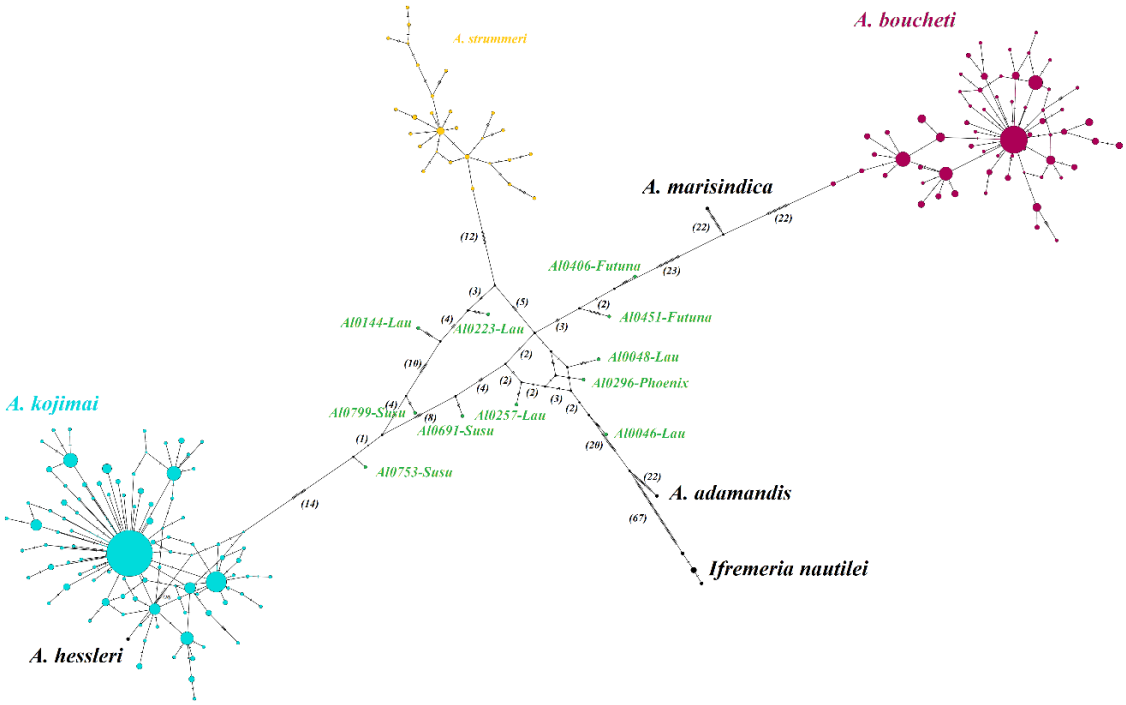


Figure 32: Haplotype network for the *Cox1* mitochondrial gene for 722 individuals from the three species of *Alviniconcha* sampled in the Southwest Pacific. The numbers in brackets represent the number of mutations on the branch. In green are represented the minor sequences of the 11 individuals showing double peaks.

The observation of double peaks led us to make several hypotheses following recommendations developed by Kmiec et al. (2006):

1) Do we have contamination of some samples?

2) Are we in the presence of a transmissible cancer?

3) Is there a double heritability of mitochondria in these species (DUI)?

4) Are these sequences represent mitochondrial heteroplasmy linked to species hybridization?

5) Are we in the presence of a nuclear copy of a mitochondrial gene (Numt)?

1- Contamination

Double peaks in our Sanger sequences could result from the amplification and sequencing of a mixture of mtDNA genomes originating from one or more contaminating

individuals. Contamination could possibly have occurred during dissection, DNA extraction, amplification, and sequencing.

This hypothesis was tested by performing a new extraction, amplification and sequencing of DNA in individuals with DPs, but the maintenance of DPs in these individuals showed that this hypothesis was not longer supported. For instance, DPs were found to be repeatable in several individuals using distinct extracts from different sources of DNA (foot, gill, gonad). Although we know for a fact that contamination occurs (e.g. one of our negative controls came back from sequencing with a readable *Alviniconcha* sequence), its low frequency, and the repeatability of DPs across replicates, prove that contamination alone cannot explain our observations. Another line of evidence against this hypothesis comes from the fact that DPs were never detected in *A. boucheti*.

2- Transmissible cancer

A single-cell lineage is called cancer when it acquires somatic mutations that promote it into a programme of continuous proliferation. Usually, natural selection favours the most prolific subclones, often directing the cancer to a more aggressive phenotype resulting in the rapid death of its host. Occasionally, very unusual cancers that have overcome the limitations of existing within a single host by acquiring the ability to spread between individuals exist. For example, Tasmanian devil facial tumor disease (DFTD) and canine transmissible venereal tumor (CTVT) are clonally transmitted cancers that spread by the physical transfer of cancer cells between hosts (Murchison, 2008). Thus, both of these cancers have continued to exist by serial transfer between hosts. Concretely in cells, this translates into the presence of both host and cancer DNA and can therefore result in the presence of two mitochondrial genomes as seen in the blue mussel (Riquet et al., 2017).

To test this hypothesis, we used our nuclear data (ddRAD-seq) because the presence of minor foreign tissue as in the case of cancer generates the presence of two nuclear genomes in the host as well. Since it represents a small amount of tissue in the organism, its copy number is lower than the host genome (Riquet et al., 2017). To identify alleles of a transmissible cancer, we therefore search for heterozygous genotypes (one allele from the host and one from the cancer) that have an unbalanced allelic sequencing depth. In summary, we tried to find false heterozygotes that could be related to the presence of foreign tissue. On the ddRAD-seq catalog common to all three species (same data set as in chapter 1), we

selected heterozygous genotypes and looked at ratio between the depth of the first allele compared to the total depth of the genotype (this ratio is expected to be centred on 0.5 in absence of foreign cells). We thus observed that the sequencing depth of each first allele in our dataset is comparable to a theoretical distribution (binomial draw). Subsequently, we repeated this calculation focusing only on the individuals which had DPs on the *Cox1* mitochondrial gene (in blue in Figure 33) but we did not observe any deviation from the expected depth of heterozygous alleles. Hence a transmissible cancer does not seem to be the cause of the observed DPs.

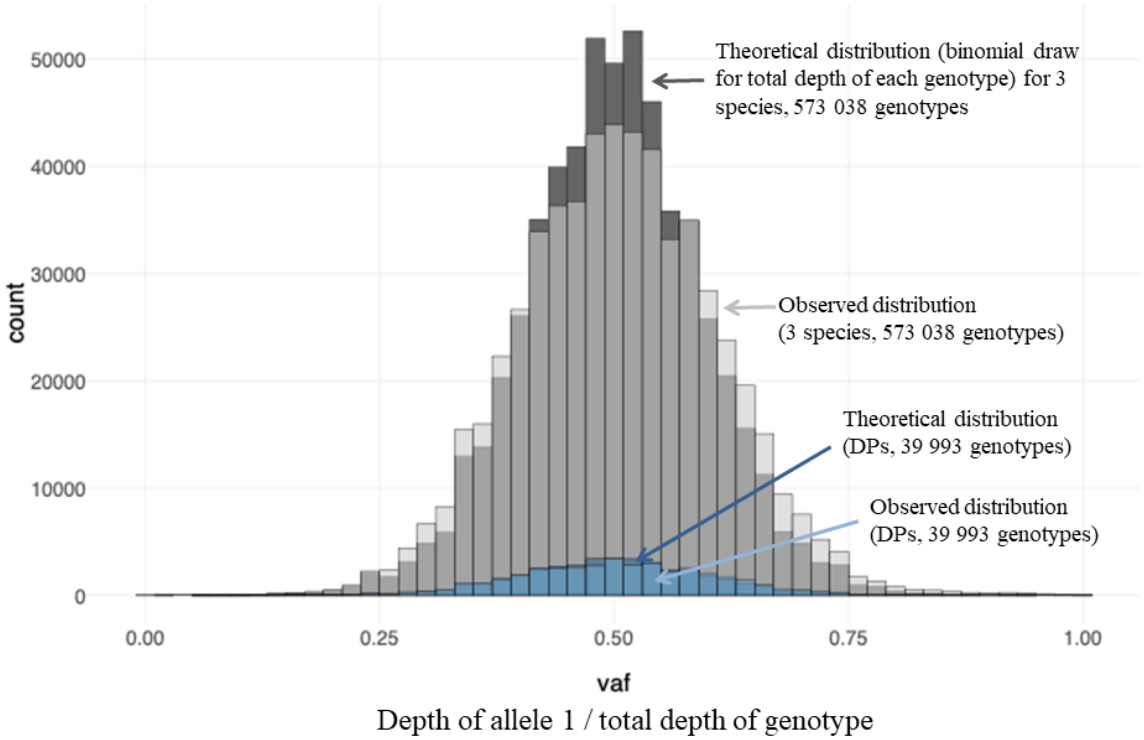


Figure 33: Comparison of the depth of the first allele over the total genotype depth between our inter-specific RAD-seq dataset and a theoretical distribution of the allele coverage. In gray, 573 038 genotypes analyzed for 450 individuals belonging to the three species of *Alviniconcha* and in blue, 39 993 genotypes analysed for the 48 individuals with double peaks.

3- Doubly Uniparental Inheritance

Another hypothesis is a DUI (Doubly Uniparental Inheritance), where females have a single *Cox1* mitochondrial sequence of maternal origin in their bodies and in males, two mitochondrial sequences are present, one of paternal origin that is found in the male gonad

and one of maternal origin that is found in all other tissues, resulting in a mitochondrial heteroplasmy if the extraction is done over the whole individual (Kmiec et al., 2006). At first, this hypothesis seemed compelling because our first detection of DP appeared potentially strongly male-biased.

To test this hypothesis, 37 individuals (9 *A. boucheti*, 24 *A. kojimai*, and 5 *A. strummeri*) were re-analysed by repeating extractions, amplification, and sequencing of the mitochondrial *Cox1* gene using both gonadal tissue, foot, and gill in male and female individuals. This experiment gave surprising results.

In *A. boucheti*, no DP were observed, and the CO1 sequence was identical across all types of tissue within each individual (in 8 individuals out of 9), thereby showing that there is nothing particular regarding *Cox1* in this species. In one individual we obtained two slightly different haplotypes (3 mutations) in the foot vs gonad/gill. We cannot exclude a contamination or other error for this particular case.

In *A. kojimai*, the haplotypes were always identical across tissues from within each individual (thus excluding the hypothesis of DUI). However, DPs were again observed in 7 individuals, with a potential sex/tissue association: all these individuals were male with DPs found in 5 gonad and 2 foot samples. The sample size is however too small to interpret further these results.

In *A. strummeri* the situation is even more complex. The five individuals used in this experiment had consistently a unique haplotype across their different tissues (again, against DUI predictions). However, 3 samples had DPs (one female foot and 2 male gonads) and 3 had hardly readable sequences (superposition of different PCR products, most likely).

4- Heteroplasmy linked to species hybridization

The hypothesis of mitochondrial heteroplasmy due to species hybridization was also rule out as all minor haplotype sequences were not falling in a specific species clade but are rather intermediate between species, leaving to the only possibility that these intermediate sequences represent Numts that could have been transferred to the nuclear genome early on during the speciation process that led to the emergence of the three species.

Conclusions and perspectives

Double-peaks are not technical artefacts: their detection is not straightforward (i.e. it depends on stochastic effects) but some of the observations were repeatable (e.g. across independent DNA extracts). Moreover, the patterns were clearly different between species, which means that the cause is biological. That is, we did amplify several loci in *A. kojimai* and *A. strummeri*.

There was no sign of variant allelic depth that could have originated from transmissible cancers, no evidence for heteroplasmy due to hybridization and there were no tissue specific patterns consistent with classical examples of double-uniparental inheritance.

- In *A. strummeri*, we think that there are two processes. In some cases a divergent sequence is co-amplified (rendering some regions of the sequences impossible to read). In addition, DPs indicate the co-amplification of a sequence closely related to *Cox1*, at least in a fraction of individuals.
- In *A. kojimai*, DP indicate co-amplification of a sequence closely related to *Cox1*, and this seems quite strongly linked with sex (24 males out of 30 in natural populations, 5 out of 7 in the additional experiment), but no link with tissue (and thus no male/female lineages within individuals).

Two main hypotheses remain: some level of mitochondrial heteroplasmy with a minority mtDNA that is not always amplified (perhaps inherited from paternal leakage), or a nuclear copy of *Cox1* (numt; Calvignac et al., 2011). The numt hypothesis would require that the divergence between the numt and the mitochondrial copy should be quite high with the Numt sequences having their own mutations. Of course, this depends on the time at which the mt sequence was transferred to the nuclear genome. Both hypotheses are difficult to reconcile with the sex-bias observed (e.g. Y-linked Numt or some kind of sex-linked heteroplasmy).

Observing DP is very difficult, time consuming, and inconsistent. To identify further the cause of these observations, one should probably try one of two options. One could either aim at isolating and sequencing specific haplotypes within individuals (using a cloning approach) or in contrary use a massive sequencing approach on mitochondrial amplicons and assemble the different locus that seem to be co-amplified in our experiments.

Appendix 2: Active hydrothermal vents in the Woodlark Basin may act as dispersing centres for hydrothermal fauna - collaborative article

communications earth & environment

ARTICLE

https://doi.org/10.1038/s43247-022-00387-9 OPEN

Active hydrothermal vents in the Woodlark Basin may act as dispersing centres for hydrothermal fauna

Cédric Boulart[1✉], Olivier Rouxel[2], Carla Scalabrin[2], Pierre Le Meur[3], Ewan Pelleter[2], Camille Poitrimol[1,4], Eric Thiébaut[1], Marjolaine Matabos[4], Jade Castel[1], Adrien Tran Lu Y[5,6], Loic N. Michel[4], Cécile Cathalot[2], Sandrine Chéron[2], Audrey Boissier[2], Yoan Germain[2], Vivien Guyader[2], Sophie Arnaud-Haond[7], François Bonhomme[5], Thomas Broquet[1], Valérie Cueff-Gauchard[8], Victor Le Layec[1,6], Stéphane L'Haridon[8], Jean Mary[1], Anne-Sophie Le Port[1], Aurélie Tasiemski[9], Darren C. Kuama[10], Stéphane Hourdez[6] & Didier Jollivet[1]

Here we report the discovery of a high-temperature hydrothermal vent field on the Woodlark Ridge, using ship-borne multibeam echosounding and Remotely Operated Vehicle (ROV) exploration. La Scala Vent Field comprises two main active areas and several inactive zones dominated by variably altered basaltic rocks, indicating that an active and stable hydrothermal circulation has been maintained over a long period of time. The Pandora Site, at a depth of 3380 m, is mainly composed of diffuse vents. The Corto site, at a depth of 3360 m, is characterized by vigorous black smokers (temperature above 360 °C). The striking features of this new vent field are the profusion of stalked barnacles *Vulcanolepas* sp. nov., the absence of mussels and the scarcity of the gastropod symbiotic fauna. We suggest that La Scala Vent Field may act as a dispersing centre for hydrothermal fauna towards the nearby North Fiji, Lau and Manus basins.

[1] UMR 7144 AD2M CNRS-Sorbonne Université, Station Biologique de Roscoff, Place Georges Tessier, 29680 Roscoff, France. [2] IFREMER REM-GM, Technopôle Brest Plouzané, 29280 Plouzané, France. [3] GENAVIR, Technopôle Brest Plouzané, 29280 Plouzané, France. [4] IFREMER REM-EEP, Technopôle Brest Plouzané, 29280 Plouzané, France. [5] ISEM CNRS UMR 5554, Université de Montpellier 2, 34095 Montpellier Cedex 5, France. [6] UMR 8222 LECOB CNRS-Sorbonne Université, Observatoire Océanologique de Banyuls, Avenue du Fontaulé, 66650 Banyuls-sur-mer, France. [7] IFREMER UMR 248 MARBEC, Avenue Jean Monnet CS 30171, 34203 Sète, France. [8] Univ. Brest, Ifremer, CNRS, Laboratoire de Microbiologie des Environnements Extrêmes UMR6197, F-29280 Plouzané, France. [9] Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019-UMR9017-CIIL-Centre d'Infection et d'Immunité de Lille, Lille, France. [10] PNG Science and Technology Secretariat, University of Papua New Guinea, Port Moresby, Papua New Guinea. ✉email: cedric.boulart@sb-roscoff.fr

Hydrothermal venting on the deep seafloor is the manifestation of heat and matter transfer from the lithosphere to the oceans, which modify their geochemical composition[1]. Upon mixing with the surrounding, cold, deep-sea water, a precipitation occurs and forms sulfide deposits on the seafloor. These hot and acidic fluids sustain diverse chemical-based ecosystems where large specialized bacterial and animal communities can thrive under extreme conditions of temperature, pressure, and pH[2]. Since the observation of the first active high-temperature vents more than 40 years ago[3], the exploration of the deep ocean has revealed the existence of hydrothermal circulation in a wide range of geological settings from fast-spreading ridges[4] to ultra-slow ones[5] as well as in intraplate hotspots[6], subduction zones[7], and back-arc basins[8].

Economic and societal interest in deep-sea hydrothermal vents has increased in the recent years because of the formation of massive polymetallic sulfide deposits[9] that are now targeted for deep-sea mining[10], which may impact the deep ocean environment. The Woodlark Basin, located South of the Solomon Islands arc region in the Western Pacific Ocean, is a rather young (~5–7 Mya) oceanic basin that is subducting beneath the New Britain-San Cristobal Trench (Fig. 1). This is in fact one of the only places on Earth where an active spreading centre expands into both continental crust to the West, and is bound by a subduction zone to the East[11,12], therefore offering a wide range of geotectonic constraints for the development of seamounts and the possibility of forming high and low temperature hydrothermal venting along the axis. This may then provide a great diversity of niches for hydrothermal fauna and microbial colonization, and the potential settlement of associated fauna, provided that venting sites remain active over a long period of time.

The Woodlark Basin is characterized by an E–W active spreading axis that Goodliffe et al.[13] divided into five segments, numbered 1 through 5 from West to East (Fig. 1). The bathymetry shows major differences between the Eastern and Western parts, separated by the Moresby Transform Fault (TF), with a significantly shallower seafloor to the West, and a well-developed axial graben to the East[14]. Spreading rates vary from 38 mm yr^{-1} in the West, to 67 mm yr^{-1} on the Eastern part[15]. Hydrothermal activities are known to occur on only three segments of the ridge, which include confirmed volcanic activity on Segment 1 (Franklin Seamount[11]), and inferred hydrothermal venting thought to be due to punctual eruptive phases on Segment 3 and Segment 5[12]. Although recent exploration found consistent turbidity and redox

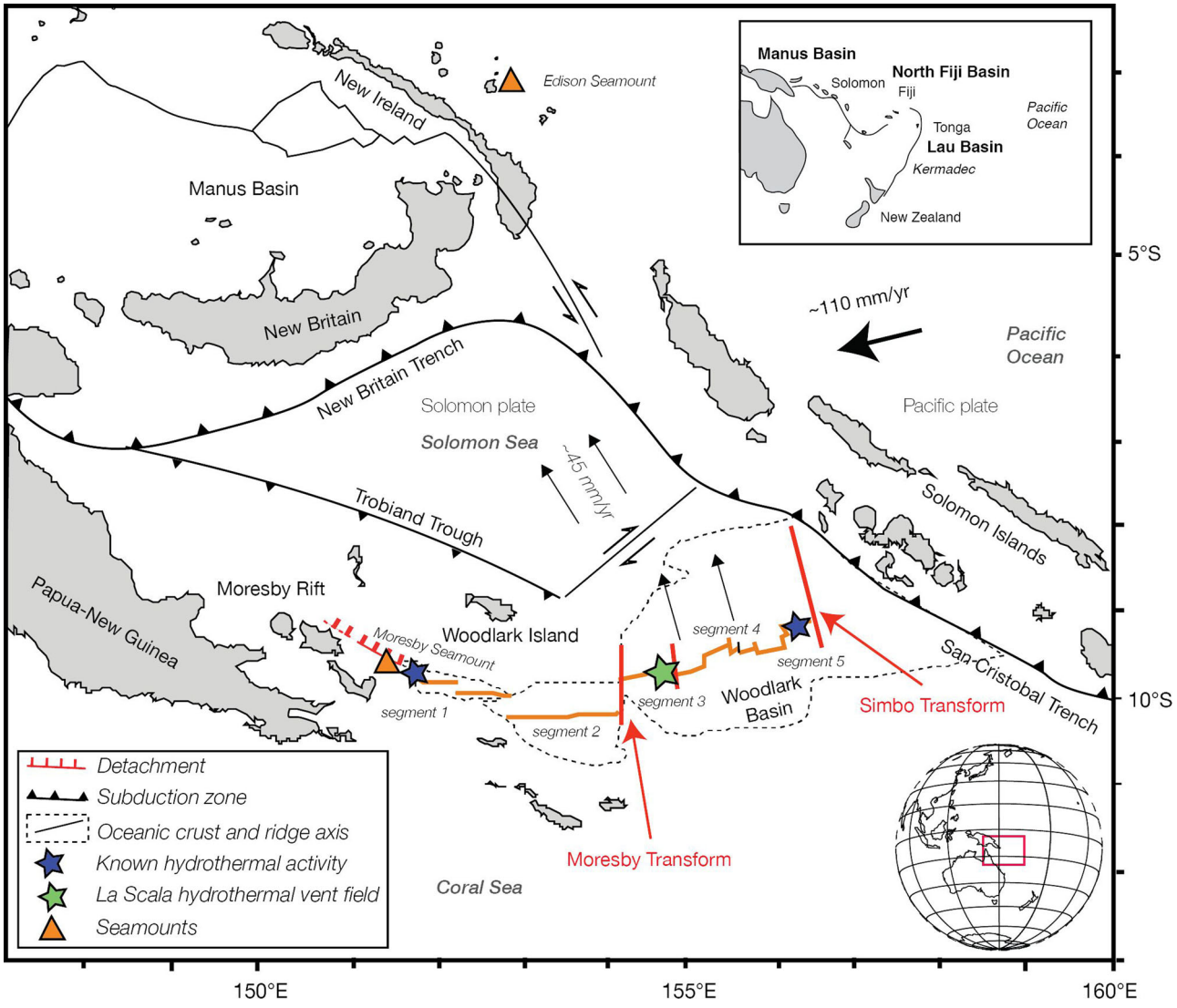


Fig. 1 Regional map of the Woodlark Basin. The main tectonic settings are highlighted as well as the location of the newly discovered 'La Scala' hydrothermal vent field (green star), modified from Laurila et al. (2012)[12]. Black arrows indicate the directions of plate motion relative to a fixed Australian plate.

potential anomalies at 2900 m below sea level (mbsl) over the eastern edge of Segment 3, Laurila et al.[12] concluded that "the complexity of the tectonics, i.e., frequent ridge jumps and re-orientation of the spreading axis, prevents high-temperature venting in one stable location and, hence the formation of sea-floor massive sulfide deposits", suggesting that stable deep-sea vent communities may not be able to establish there.

Although not reported, the presence of perennial hydrothermal vent fields in the Woodlark Basin could constitute a stepping stone for hydrothermal fauna, at the intersection between the Manus Basin, the Edison Seamount, and the North Fiji and Lau basins, because of its possible link to the Northern expansion of the North Fiji basin which started about 10–11 Mya[16]. This spreading centre, now subducted, was presumed to bridge the older -and fossil- Solomon and South Fiji ridges that shaped the region before the collision of the Melanesian arc and the Ontong Java plateau about 18 Mya[17]. This collision coincides with the oldest dates of speciation events in the complex of gastropod species *Alviniconcha* which initiated 20 Mya[18]. The Woodlark Basin is older than the currently active spreading centres of the adjacent Lau (1–2 Ma) and North Fiji (3–4 Ma) basins, and therefore, may act as a biodiversity dispersion centre for the modern hydrothermal vent fauna at a crossroad between the Manus, North Fiji, and Lau basins.

During the CHUBACARC 2019 cruise (https://doi.org/10.17600/18001111), we carried out an extensive water-column survey of the eastern edge of Segment 3, using a strategy based on ship-borne acoustic survey followed by CTD-casts, to detect both thermal and chemical anomalies near the location where Laurila et al.[12] previously reported activity. Based on these observations, we then conducted a seafloor survey using the ROV Victor 6000 that led to the discovery of several high-temperature hydrothermal black smokers. Here, we report on the acoustic and chemical characterization of the hydrothermal plumes, the composition of endmember fluids, the geological setting of the vent fields, and the composition of the fauna associated with the newly discovered vent field we named 'La Scala'.

Results and discussion

Hydrothermal plume exploration over the Eastern Woodlark Ridge. As part of the high-resolution mapping—using the ship-borne multibeam echosounder (EM122 12 kHz)—carried out over Segment 3 (Fig. 2a), the water column imaging survey revealed the presence of echoes rooted to the seafloor in the vicinity of the 'TVG-150' marker[15]. These echoes, rising 200–300 m above the seafloor (Fig. 2b, Supplementary Fig. 1), remained visible at the same location at each passage of the ship over the area and showed the same features as the plumes previously observed in the Guaymas Basin[19]. During the CHUBACARC cruise, similar signals were also observed at shallower sites known to host active vents in the Manus Basin. These echoes were therefore attributed to the presence of hydrothermal plumes. It is commonly believed that deep hydrothermal plumes cannot be detected by water column acoustic imaging because of the absence of strong scatterers such as gas bubbles or droplets[20]. However, Ondréas et al.[19] showed that hydrothermal plumes down to 2000 mbsl could be identified using ship-borne multibeam echosounding and differentiated from gas and liquid emissions. In the case of La Scala plumes, the backscattering mechanism may result from a combination of turbulence-induced temperature fluctuations within the first tens of metres of the vents and the presence of fine particles over the entire field at 100–500 m above the seafloor, which was later confirmed by the CTD surveys over the area (very strong turbidity signal and temperature anomalies >0.2 °C in the buoyant plume) and the subsequent ROV dives. By modelling the acoustic characteristics of the hydrothermal plumes, Xu et al.[21] showed that particles had a predominant role in the acoustic backscatter with the height above the vents. It is worth noting that the success of this technique of detection strongly depends on the acoustic impedance difference between the hydrothermal fluid and the surrounding water, on the acoustic frequency, and on the sea state.

The subsequent CTD/tow-yo survey (Fig. 2c) first in the N–S direction and, then followed by a NE–SW transect crossing above the fluid echo showed turbidity (or nephelometry) anomalies at 2800 and 3300 mbsl i.e., at an altitude of 300–500 m above the seafloor (Fig. 3a, b), associated to manganese (Mn) anomalies (up to ~108 nM, Supplementary Data 1) and redox potential (Eh) signals. The strongest anomalies appeared to be consistent with the location of the echoes, confirming the presence of a buoyant plume.

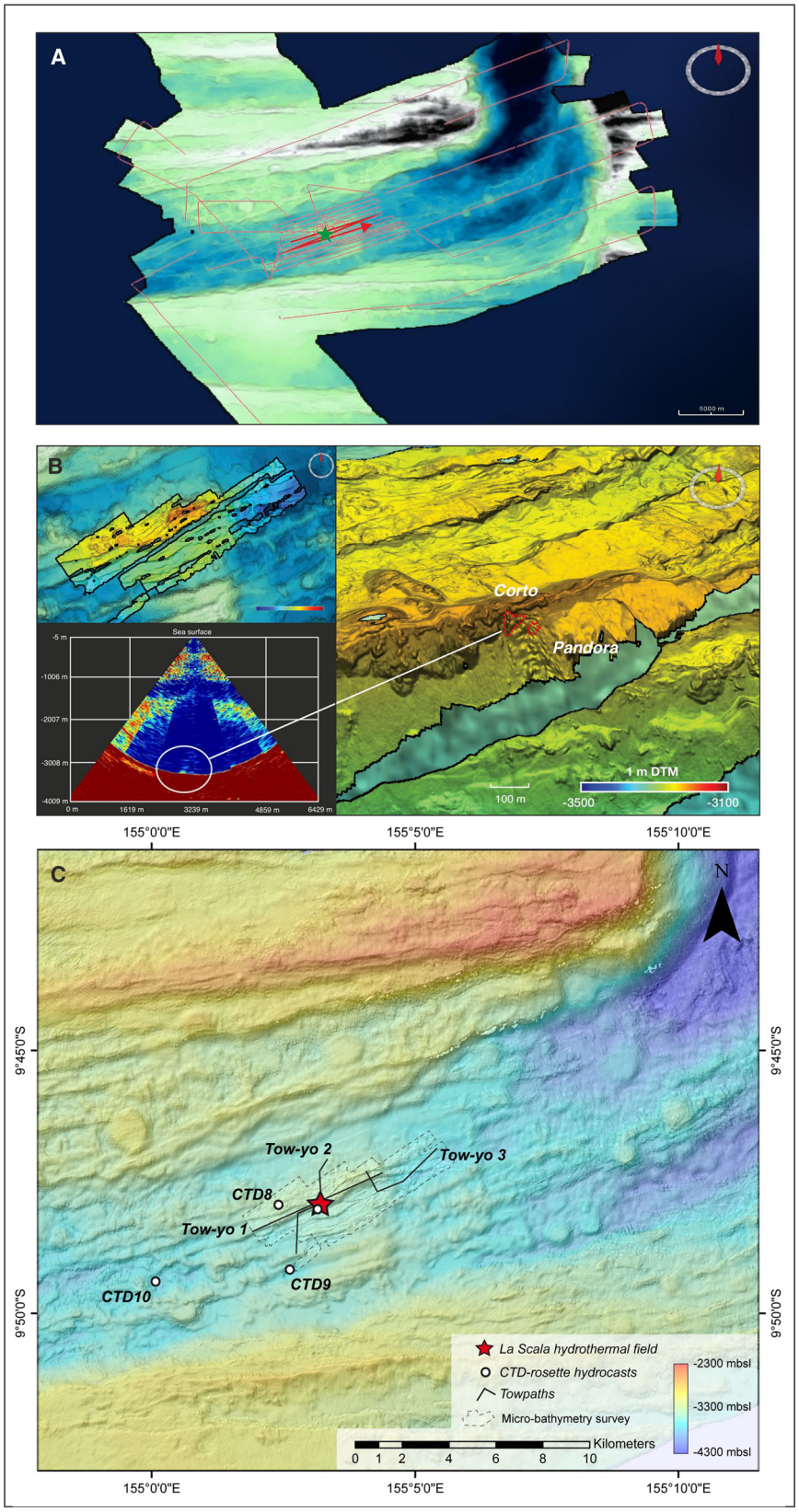
The second phase of the water column survey consisted of two vertical CTD casts (CTD-06 and CTD-07) above the strongest acoustic and chemical anomalies spotted during the MBES and the tow-yo survey. The data showed unambiguous anomalies of Eh, pH, potential temperature, density, and turbidity (Fig. 3c, d), which confirmed the presence of both buoyant and non-buoyant plumes. During CTD-06, we encountered a first turbidity anomaly (0.08 NTU) associated to an Eh signal at 2850 mbsl corresponding to a dispersing buoyant plume 500 m above seafloor. Below, a second, weaker turbidity anomaly (0.03 NTU) was found but not associated to any other chemical signal. Finally, very strong turbidity signals (up to 0.5 NTU), as well as temperature anomalies (~0.1–0.3 °C), were observed close to the seafloor (180 and 100 m above the seafloor) indicative of the presence of at least two hydrothermal plumes (Supplementary Fig. 2). The CTD-07 cast, conducted a few hours later and 60 m away from the CTD-06 location, showed a typical non-buoyant plume structure for turbidity and temperature (anomalies > 0.2 °C), as well Mn varying from 34 to 265 nM. The CTD-07 vertical profile clearly indicated the presence of several black smoker-type sources spread out on the flank of the ridge, from 3300 to 3500 mbsl. These vents seem to generate a regional neutrally-buoyant plume dispersing northeastward at 2800–2850 mbsl.

Based on these exploratory results, an ROV dive close to the bottom led to the discovery of many active hydrothermal chimneys (i.e., black smoker types) and complex spires located between 3300 and 3400 mbsl, as well as a large diffuse venting area colonized by extensive communities of stalked barnacles (Fig. 4). The new vent field, discovered at coordinates 9°47S/155°03E was named 'La Scala', in recognition of the acoustic eyes of its discoverer.

Three additional CTD casts (CTD-08, 09, and 10, Fig. 2c) were carried out to estimate the dispersion of the plumes. While the CTD-08 profile showed some clear evidence of hydrothermal plumes dispersing at 2800–2900 mbsl, no anomalies were observed on the CTD-09 and CTD-10 profiles, located at 2.4 and 5.7 km, south and southwest of CTD-07, respectively (Supplementary Data 1). This may indicate a strong control of the dispersion of the plumes by the topography.

Field observations of the La Scala Vent Field

La Scala Vent Field (LSVF) is located at a distance of 1.2 km North of the neovolcanic axis on the northern part of an old split axial volcanic ridge. The LSVF lies on a NE-trending talus slope and is composed of two main active areas and several zones with inactive chimneys (Fig. 2b). The first active site, named 'Pandora', is located Southeast of the field at 3380 mbsl and extends some 30 m by 10 m. It is characterized by predominantly diffusive vents, a few black smokers (Fig. 4a, b) and



extinct chimneys N20 aligned and colonized by an extensive coverage of stalked barnacles and patches of gastropods. The second active site of the field, named 'Corto', lies along a steeper slope at 3360 mbsl, only 15–20 m Northwest of the first site. It comprises a 50 m by 15 m sulfide zone where several 7–10 m tall, vigorous, high-temperature black smokers (Temperature ranging from 364 to 366 °C, Table 1, Fig. 4c, d) were observed at the Southern edge.

Active hydrothermal chimneys are located on steep talus composed of hydrothermally encrusted basaltic rubble, chimneys fragments, and white blocks of hydrothermally altered volcanic rocks. Near the summit, just above the Corto site, a 30 m tall cliff

Fig. 2 La Scala Vent Field bathymetry. A Chubacarc cruise bathymetry image (20 m grid DTM (Digital Terrain Model), ship-borne MBES) corresponding to the water column acoustic surveys of the Woodlark Ridge (segment 3b). The green star locates the 'La Scala' hydrothermal field. **B** Northward 3D view of the high-resolution bathymetry image (1 m grid DTM, ROV VICTOR MBES) overlaid on the greyed 20 m grid DTM corresponding to the northern part of the old split axial volcanic ridge where LSVF is located. La Scala Vent Field including Pandora and Corto sites is delimited by the red dashed line, on the southern flank of the ridge together with the post-processed echogram of the second profile showing the echo attributed to LSVF. **C** Bathymetry of the segment 3 of the Woodlark Ridge from shipboard multibeam survey showing the location of LSVF as well as the towpaths and the CTD-hydrocasts. Note that CTD06 and CTD07 are located right above LSVF. TVG-150 spot from Laurila et al. (2012)[12] is not showed for clarity as it is 140 m away from LSVF.

is composed of pillow lavas topped by massive lava flows (Fig. 4e). On this same cliff, fifty metres Southwest of the second hydrothermal site, pervasive alteration zones and possibly oxidized massive sulfides exposed after a tectonic event (Fig. 4f) have been observed, indicating past hydrothermal activity in the area. Even though high-temperature chimneys could have formed during previous hydrothermal episodes, most ancient sulfide structures were probably destroyed during a collapse. Thus, at least two hydrothermal events occurred at LSVF. Based on the spreading rate (~ 50 mm yr^{-1}) and the distance from the neo-volcanic axis (~1.2 km), the maximum age of the first event of hydrothermal activity at LSVF can be estimated to be 24,000 years ago. This calculation, along with the presence of numerous inactive chimneys and altered lava, clearly indicates that an active and stable hydrothermal circulation has been maintained over a certain period of time and may not be the result of a recent and episodic volcanic eruption as previously suggested[12].

Petrological and geochemical characteristics

Basement rocks recovered at LSVF are composed of rounded to sub-angular clasts of non-vesicular aphyric to plagioclase-pyroxene-phyric basalt (cm-size pebbles), cemented by a silica-rich matrix (quartz and cristobalite) and often coated by mm-thick iron oxyhydroxide crust. Pyrite and minor sphalerite are also identified by X-ray diffraction (XRD) confirming the importance of hydrothermal fluid circulation in basement rock alteration at LSVF. East of the vent field along the summit of the ridge crest, recovered basement rocks consisted of essentially aphyric and non-vesicular pillow lava with glassy chilled margin. Incipient to no alteration was identified. The volcanic crest is covered by thin sediment dusting composed of nanofossil-bearing clay.

Geochemical analyses of volcanic rocks show silica contents ranging from 48.4 to 48.7 wt% and low alkali contents ($Na_2O + K_2O$) from 2.7 to 2.9 wt% (Supplementary Table 1). Thus, all samples from the LSVF area are tholeiitic basalt similar to mid-ocean ridge basalt (MORB) composition (Supplementary Fig. 3). Glassy pillow margins contain 9.1–9.5 wt% MgO suggestive of rather primitive melt. To our knowledge, these are the first report of basaltic rock composition along Segment 3 of the Woodlark Basin in agreement with the previous studies along Eastern Segment 4[22,23] and consistent with MORB composition with major element variations resulting from fractional crystallization.

Sulfide deposits and hydrothermal fluids

Three different types of mineralization (Fig. 4 and Supplementary Data 2) were observed and sampled during the subsequent ROV dives: (1) active high-temperature sulfide/sulfate chimneys without significant presence of sessile animals; (2) inactive black smoker chimneys; (3) massive sulfide recovered at the base of either active or inactive complex spire densely colonized by vent organisms. Mineralogical and petrological descriptions show that anhydrite is the dominant mineral in the top section of active black smoker chimney, followed by chalcopyrite, sphalerite, and pyrite towards the inner and lower section of the chimney. Multiple vigorously venting internal conduits are lined with fine grained euhedral chalcopyrite while chimney walls are composed of mixed mineral assemblages of chalcopyrite, sphalerite, and pyrite. Inactive chimneys and massive sulfides have essentially similar characteristics to active chimney except that anhydrite is absent due to retrograde dissolution and chimney wall are consolidated by more massive pyrite/marcasite assemblages, and thinly encrusted by Fe oxyhydroxide crust.

Sulfide deposits show a wide range of copper (Cu) and zinc (Zn) concentrations reflecting the abundances of chalcopyrite versus sphalerite, while calcium (Ca), strontium (Sr), and silica (Si) reflect the abundance of anhydrite and silica respectively (Suppl. Data 3). As previously recognized[24], trace elements such as cadmium (Cd) and lead (Pb) are associated with lower-temperature sphalerite-rich mineral assemblages. Average compositions are relatively similar between sulfide deposits, except for the Carioca black smoker characterized by lower Cu and higher cobalt (Co) concentration, up to 0.6 wt% in pyrite-rich samples. High Co contents in basalt-hosted sulfide chimneys have been previously inferred as reflecting a higher magmatic contribution to the hydrothermal fluid[24]. By comparison with average compositions of sulfide mineralization in other tectonic environments in the modern seafloor, LSVF lacks any of the geochemical features of back-arc settings (e.g., gold and Pb enrichment) and is seemingly identical to other basalt-hosted hydrothermal fields along mid-oceanic ridges.

Sulfur isotope analysis was performed on the mineral fractions of pyrite and chalcopyrite of four samples, yielding a range of $\delta^{34}S$ values between 1.96‰ and 4.20‰, with an average of 3.03 ‰ (Supplementary Data 3). Slightly higher $\delta^{34}S$ values in the Carioca black smoker indicate a seawater contribution to the hydrothermal fluid in the upflow zone. Such $\delta^{34}S$ values are typical of sulfide mineralization from mid-oceanic ridge settings (Supplementary Fig. 4) and suggest no contribution of magmatic SO_2 ($\delta^{34}S < 0$‰) as encountered in back-arc basin hydrothermal systems (e.g., Lau Basin and Manus Basin). Altogether, these results are consistent with the basaltic rock composition and indicate that the magmatic activity on the ridge segment is not directly related to subduction but rather to accretion.

Hydrothermal vent fluids were recovered from three black smoker chimneys from LSFV, all showing remarkably homogeneous venting temperature of ~365 °C (Table 1 and Supplementary Table 2). All recovered vent fluids have non-zero magnesium (Mg) concentrations due to significant seawater entrainment during sampling. Endmember composition of vent fluids are conventionally assumed to be devoid of Mg, because of the quantitative removal of Mg from seawater during hydrothermal interactions with basalt[25,26]. Calculated end-member hydrothermal fluid compositions are reported in Table 1 for all measured elements.

Homogeneous end-member compositions suggest that LSVF is fed by a deep-seated hydrothermal fluid source undergoing minor subsurface fluid mixing. The lowest pH value (3.31) was recorded in Carioca vent fluids consistent with the lowest Mg concentration of 1.6 mM, which suggests minor seawater contribution during sampling. Hence, the end-member pH value is considered to be close to 3.3 for all vents. All end-member vent fluids have

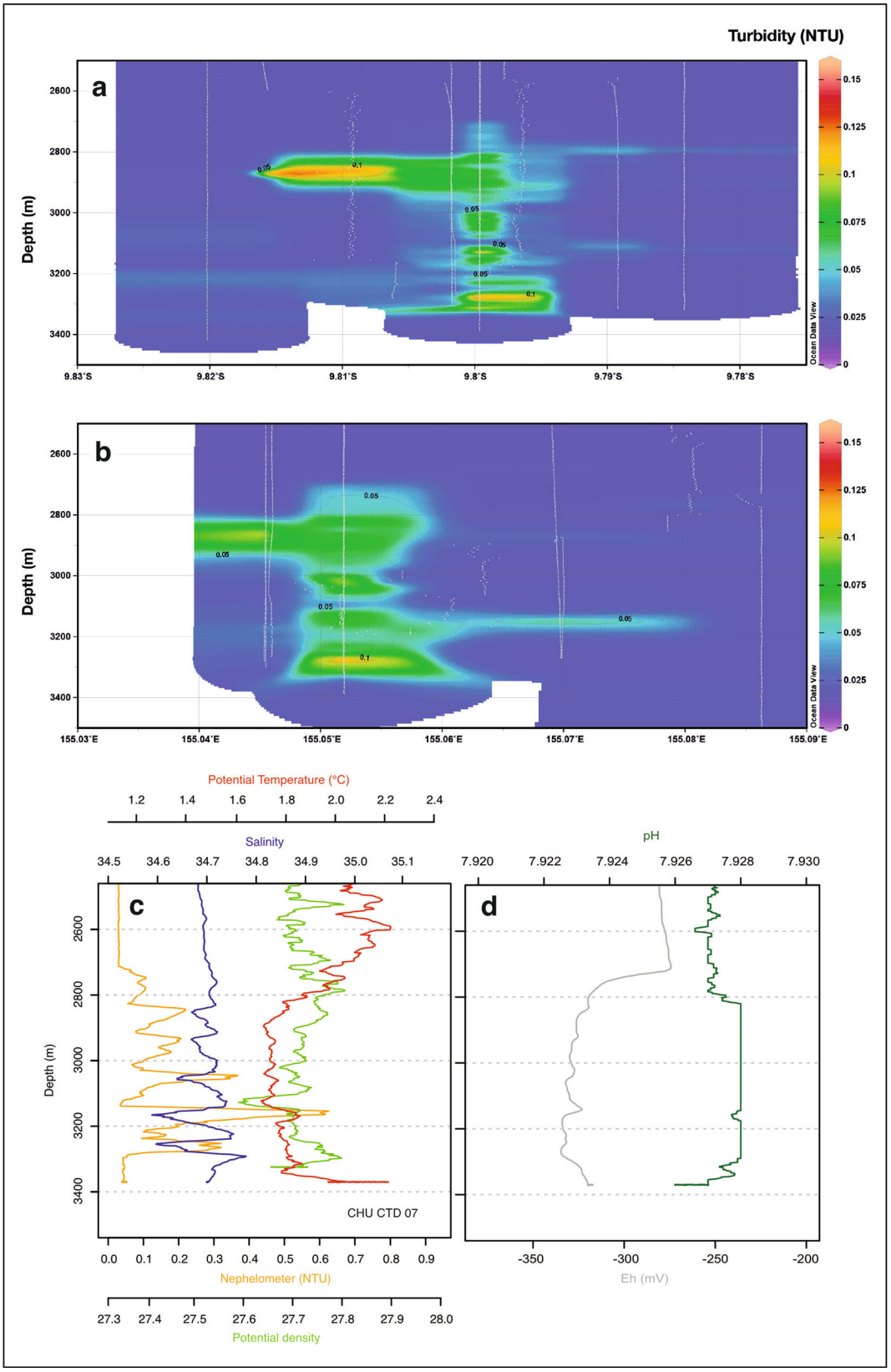


Fig. 3 Physico-chemical characteristics of hydrothermal plumes above La Scala Vent Field. N-S (**a**) and NE-SW (**b**) sections of turbidity anomalies above Segment 3. The section compiles data from tow-yo surveys (CTD-01 to CTD-03) and vertical casts (CTD-06 and CTD-07) (grey lines). **c**, **d** Nephelometry (yellow), density (light green), salinity (blue), temperature (red), Eh (grey), and pH (dark green) vertical profiles in the non-buoyant plume of 'La Scala' vent site (CTD-07).

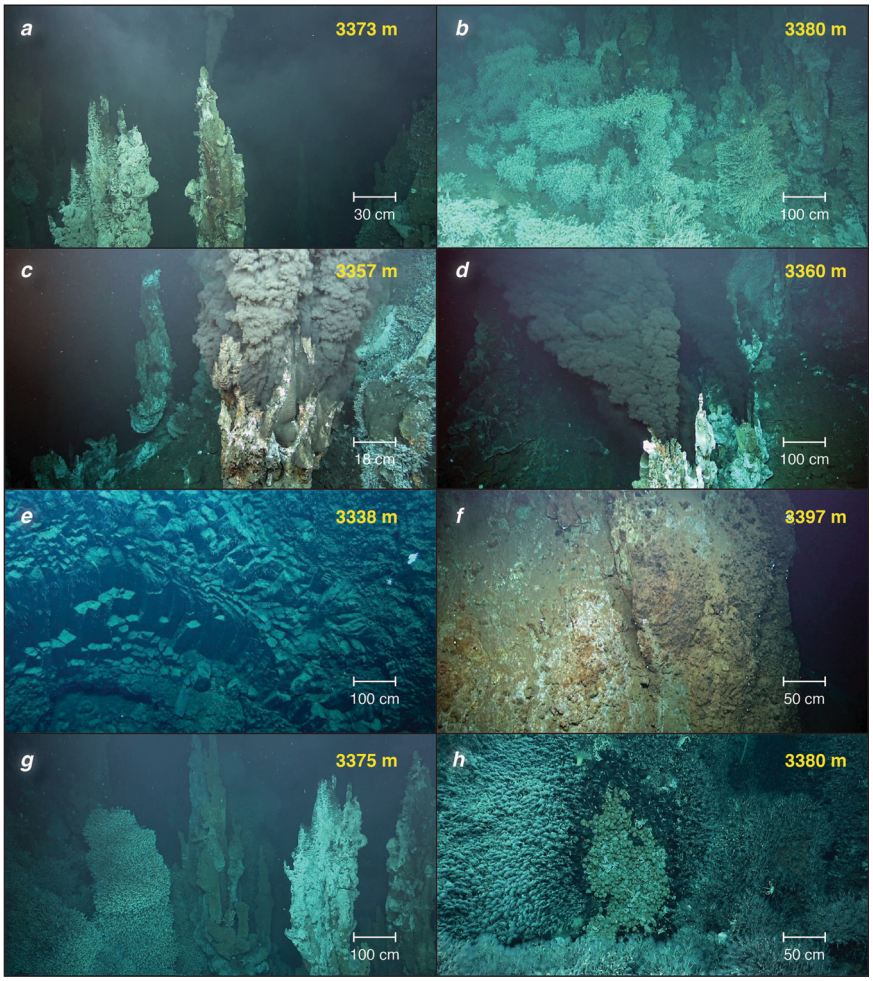


Fig. 4 La Scala Vent Field. a, **b** Chimney cluster ('Pandora Site'), SE of the vent field at 3380 m depth characterized by a few black smokers and predominant weakly diffusive vents. This site is surrounded by cirripeds to the West, North and East. **c**, **d** Second site ('Corto Site') located on the flank of a cliff controlled by a N55-60 fault. The South area at 3360 m depth is the most active of the hydrothermal field. **e** Massive prismatic lava. **f** Highly altered basalts due to past high temperature hydrothermal circulation. **g** Large colonies of *Vulcanolepas* sp. colonizing inactive chimneys. **h** A patch of *Alviniconcha* endosymbiotic gastropods surrounded by *Ifremeria nautilei* and *Vulcanolepas* sp.

chlorinities higher than seawater suggesting they have undergone sub-critical phase separation. Alkali concentrations (K≈25.4 mM; Rb ≈ 15.6 μM) are similar to most basaltic-hosted vent sites consistent with LSVF geological setting. Silica (Si), iron (Fe) and Mn concentrations are ≈20 mM, ≈1.7 mM, and ≈1.0 mM, respectively. Fe/Mn ratio of 1.7 is however lower than values expected from the vent temperature[27], suggesting possible subsurface sulfide precipitation. Endmember Cu and H_2S concentration of 21 μM and 4.1 mM respectively, are also lower than typical high temperature (i.e., >350 °C) vent fluids, lending further support to this hypothesis.

Hydrothermal vent fauna

LSVF is characterized by several very active black smoker chimneys whose surface is occupied by a sparse population of *Rimicaris* shrimp and very large beds of a new species of stalked barnacles, *Vulcanolepas* sp. nov., that colonizes mildly active diffuse venting areas and old inactive chimneys or sulfide mounds (Fig. 4g). The dominant engineer species and their associated fauna were comparable to other Southwest Pacific vent communities but differed in relative abundance[28–30]. Stalked barnacles covered vast areas (c.a. >300 m^2) on the seafloor while large symbiotic gastropods were rare and restricted to patches smaller than in other Western Pacific back-arc basins. Patches of small-sized gastropods *Ifremeria nautilei* and *Alviniconcha* spp. were observed in the most active diffuse areas with temperatures varying from 2.4 to 7.3 °C (average of 4.6 ± 1.3 °C) and 2.5–20.3 °C (average 8.5 ± 4.4 °C), respectively (Fig. 4h). Cirripeds *Eochinelasmus ohtai* and *Imbricaverruca* sp. were found next to *Ifremeria nautilei* and *Alviniconcha* spp. communities (Fig. 5). Aggregations of alvinocaridid shrimp *Rimicaris variabilis*, *Branchinotogluma segonzaci* polynoids, and *Shinkailepas tufari* phenacolepadid gastropods were also observed on active chimneys (Fig. 5f). Extensive bacterial mats were also noticed at the periphery of the vents, as well as bamboo corals (Isididae), squat lobsters (Munidopsidae), brisingid starfish, crinoids, and sea anemones (Supplementary Fig. 5). Morphological identification of the benthic fauna collected and observed from videos was compiled as a list of at least 45 taxa, including 23 families and 23 genera (Supplementary Data 4). The macrofauna associated with *Vulcanolepas* sp. nov. was mainly composed of the barnacle *Imbricaverruca* sp., members of the polychaete families Ampharetidae (*Amphisamytha* cf. *vanuatuensis*), Maldanidae (*Nicomache* spp.), and Spionidae, and the provannid gastropod *Provanna* sp. (Supplementary Data 5). Many copepods and nematodes were also part of this community. Holothurians *Chiridota* sp., anemones (Actiniaria), squat lobsters, *Austinograea*

Table 1 Temperature, pH, and chemical composition of La Scala fluid endmembers.

	Temp (°C)	pH (21 °C)	H_2S mM	Mg mM	Cl mM	SO_4 mM	Na mM	S mM	K mM	Ca mM	Sr µM	Li µM	B µM	Rb µM	Cd µM	Ba µM	Si µM	Mn µM	Fe µM	Cu µM	Zn µM
Bottom SW		7.5	0.00	55.75	546	28.2	474	29.8	9.1	10.6	99	27	435	1.5	<0.1	0.08	122	<0.1	<1	<0.1	<0.1
Black smoker 6	364	<4.31	>2.2	0	633.2	2.0	561	2.7	25.3	31.7	115	1163	763	15.6	0.2	9.4	20074	1019	1602	39.9	108.1
Black smoker 5	366	3.3	4.0	0	634.9	1.4	543	2.1	25.4	31.0	115	1208	749	15.6	0.1	17.1	19795	1019	1848	10.8	50.6
Black smoker 8	365	3.3	4.1	0	647.9	1.5	548	5.4	25.6	31.5	115	1198	757	15.6	0.1	14.6	20122	1028	1707	12.6	39.8

crabs and *Phymorhynchus* sp. gastropods were also found among and next to *Vulcanolepas* sp. nov. thickets. A sample collected in an *Ifremeria nautilei* patch showed that the associated macrofauna was dominated by an undescribed peltospirid limpet and polynoid scaleworms, and formed the most diverse assemblage in term of species richness. *Amphysamytha* cf. *vanuatensis* and alvinellid worms (genus *Paralvinella*) were also found in *Ifremeria nautilei* and *Alviniconcha* spp. communities. Copepods, bythograeid crabs (*Austinograea* sp., which is likely to represent a new species), *Rimicaris variabilis* shrimp and polychaetes were the dominant taxa found in *Alviniconcha* spp. patches.

Most of the fauna had $\delta^{34}S < 10‰$ (Fig. 6), suggesting they primarily depend (either directly or indirectly) on chemosynthetic vent production and sulfide oxidation for their nutrition[31]. This was notably the case of organisms considered as peripheral fauna, or not strictly found at vents, such as *Vulcanolepas* sp. nov., anemones, or *Phymorhynchus* sp.. Brisingidae starfish, holothurians *Chiridota* sp., and carnivorous sponges had $\delta^{34}S$ between 10 and 15‰, suggesting a mixed diet, depending on both chemosynthesis- and photosynthesis-derived organic matter[31]. Only the Isididae bamboo corals had a $\delta^{34}S$ greater than 15‰, suggesting they primarily feed on exported photosynthetic production[31]. Overall, the dependence of organisms on vent endogenous production seemed significant, and spanned all sampled taxonomic and functional groups. Similar observations have been made in active[32] and inactive[33] hydrothermal vents from the Manus basin. In contrast, the dependence of hydrothermal fauna on exported photosynthetic matter is higher at shallower vents of the Southern Tonga arc[34]. The increased importance of endogenous production for deeper sites matches patterns observed in other chemosynthesis-based habitats[35,36].

While $\delta^{34}S$ values ruled out major photosynthetic contributions, the wide ranges of $\delta^{13}C$ and $\delta^{15}N$ measured in Woodlark fauna suggest that animal communities show considerable trophic diversity, and depend on several production mechanisms (Fig. 6). Some species had very negative $\delta^{13}C$ (e.g., *Alviniconcha kojimai*, *Ifremeria nautilei* or *Provanna* sp.), suggesting they mostly depend on sulfide oxidizers using the Calvin–Benson–Bassham (CBB) cycle for their nutrition[37]. Conversely, some species, such as *Alviniconcha boucheti* or *Branchinotogluma segonzaci* had very positive $\delta^{13}C$, suggesting they acquire most of their organic matter from sulfide oxidizers using the reverse tricarboxylic acid (rTCA) cycle[37]. Interestingly, many taxa had intermediate $\delta^{13}C$ values comprised between those two end-members, suggesting a co-reliance on both bacterial metabolisms in variable proportions. Similar findings have been reported for the Manus Basin[32,38]. However, in Manus, communities seem to be mostly supported by the CBB cycle, particularly at the Solwara 1 site[32]. In Woodlark, our results suggest that inter- (and sometimes intra-) taxon differences in feeding preferences could lead to a more evenly balanced continuum of $\delta^{13}C$ values, and therefore the importance of the two production mechanisms for the food web. Finally, $\delta^{34}S$ of Woodlark fauna was similar to some hydrothermal vents from the Manus Basin (Solwara 1, PACMANUS), but markedly higher than other Manus sites (e.g., South Su)[32]. This suggests that local changes in sulfur geochemistry could influence this parameter and the way secondary consumers are locally distributed.

Genetic characterization of the main ecological engineer species

As other vent communities of the western Pacific back-arc basins, engineer species observed at LSVF are mainly composed of stalked barnacles, and provannid gastropods that belong to the genera *Alviniconcha* and *Ifremeria*. *Bathymodiolus* mussel beds

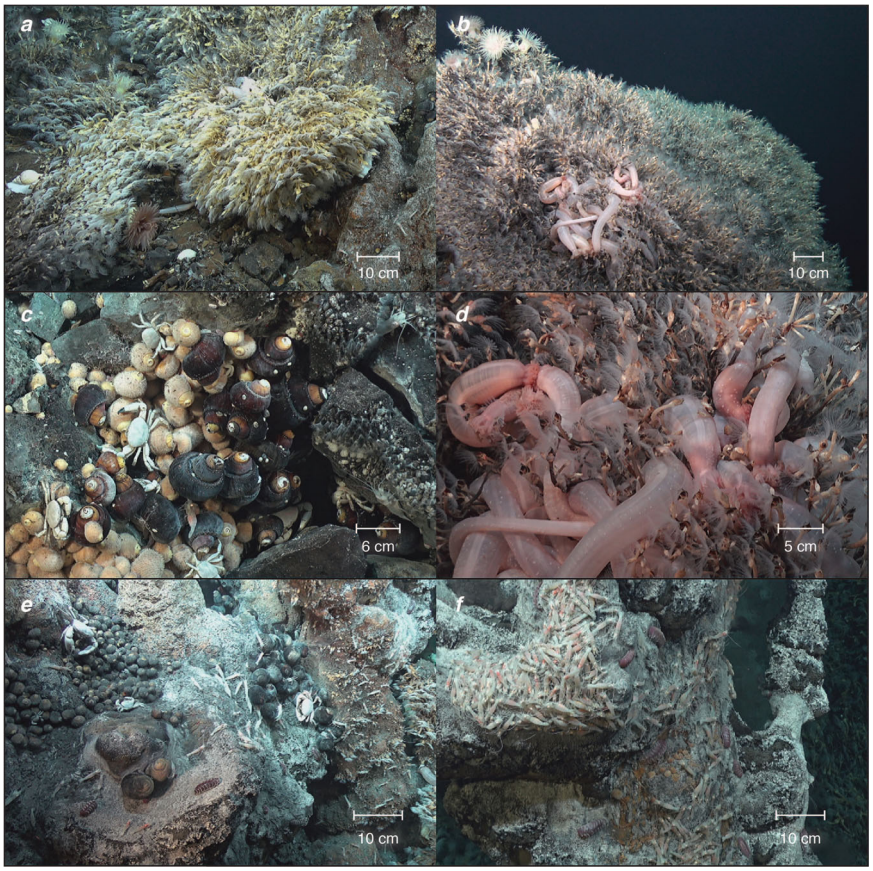


Fig. 5 Hydrothermal fauna from La Scala Vent Field. a *Vulcanolepas* sp. n. thickets, anemones and gastropod *Phymorhynchus* sp. **b** Anemones and holothurian *Chiridota* sp. in *Vulcanolepas* sp. n. thickets. **c** Aggregation of gastropods *Ifremeria nautilei* and *Alviniconcha kojimai* with *Austinograea* crabs and cirripeds *Eochionelasmus ohtai* and *Imbricaverruca* sp. (right side of the picture). **d** *Chiridota* sp. in *Vulcanolepas* sp. n. thicket. **e** Aggregation of *Alviniconcha kojimai*, polynoids, alvinocarids shrimps, and *Austinograea* crabs. **f** Aggregation of alvinocarid shrimp *Rimicaris variabilis*, polynoids *Branchinotogluma segonzaci* and gastropod *Shinkailepas tufari* on active chimney.

were surprisingly not found at Woodlark. A first genetic examination of stalked barnacles using a 658-bp fragment of the mitochondrial cytochrome c oxidase (*Cox1*) gene clearly indicated that, while falling into the *Vulcanolepas* clade (support 0.66), specimens from Woodlark form a distinct phylogenetic clade (support 0.99, within clade pairwise Jukes–Cantor distances 0.22 ± 0.21%) and therefore represent a new putative species (Fig. 7). This new species is surprisingly most closely related to the species found at vents on the Kermadec volcanic arc (*V. osheai*, 6.86 ± 0.23% average Jukes–Cantor distances), and a bit more distantly to the species found at vents in the Lau Basin (*V. buckeridgei*, 7.35 ± 0.33% average Jukes–Cantor distances) or the species *Leucolepas longa* from Lihir island[39]. Interestingly, for this genus, each sampled area seems to host a distinct species. *Alviniconcha* and *Ifremeria* gastropods were not numerous (distributed as small patches at the base of chimneys) and of rather small size, suggesting that they may represent genetically differentiated populations when compared to other described species elsewhere. All individuals collected at Woodlark were thus barcoded using a 659 bp-fragment of mitochondrial *Cox1* gene to build haplotype networks and test their relationships with previously described species. For *Alviniconcha*, the use of previously published sequences yielded 5 clusters that correspond to the currently described species. *Alviniconcha* specimens collected at La Scala fall into two equally represented but distinct species, *A. boucheti* and *A. kojimai*, but are not geographically distinct from populations collected in other basins, at least using this genetic marker (Supplementary Fig. 6). *Ifremeria* specimens fall into two main clusters separated by five fixed substitutions (divergence = 0.75%, Fig. 8): one that mainly comprises specimens from the Manus and Woodlark basins and one that comprises specimens from other basins (Lau, North Fiji, and Futuna) (Fig. 8 and Supplementary Fig. 6). Remarkably, a few specimens from Futuna and Lau however, fall into the Manus/Woodlark cluster. This suggests a possible recent migration from the Woodlark ridge to the Northern part of the Lau basin (including Futuna) following a first geographic isolation event of the Manus and Lau/North Fiji/Futuna populations in allopatry. These preliminary results clearly indicate that vent species have contrasted population histories. Some of the species are clearly endemic of the Woodlark ridge (a possible separation due to the greater depth) whereas others may use it as a stepping stone during the colonization of the present-day back-arc basins of the Western Pacific.

The Woodlark Ridge: A potential biological cornerstone?

The tectonic history of the western Pacific back-arc basins is complex[17]. The opening of some basins such as Manus, Lau, and North Fiji is recent (less than 4–3 millions years)[17,40,41], and likely simultaneous, even if the formation of the North Fiji protobasin initiated before about 10–12 Mya[16]. This points towards the existence of older relay-ridges for the vent fauna before their opening of the present-days basins as previously mentioned to explain the spatial distribution patterns of the symbiotic gastropod *Alviniconcha* spp[18]. In this context, the present-day distribution of the hydrothermal vent fauna likely results from the partition of an older hydrothermal fauna originating from ridges

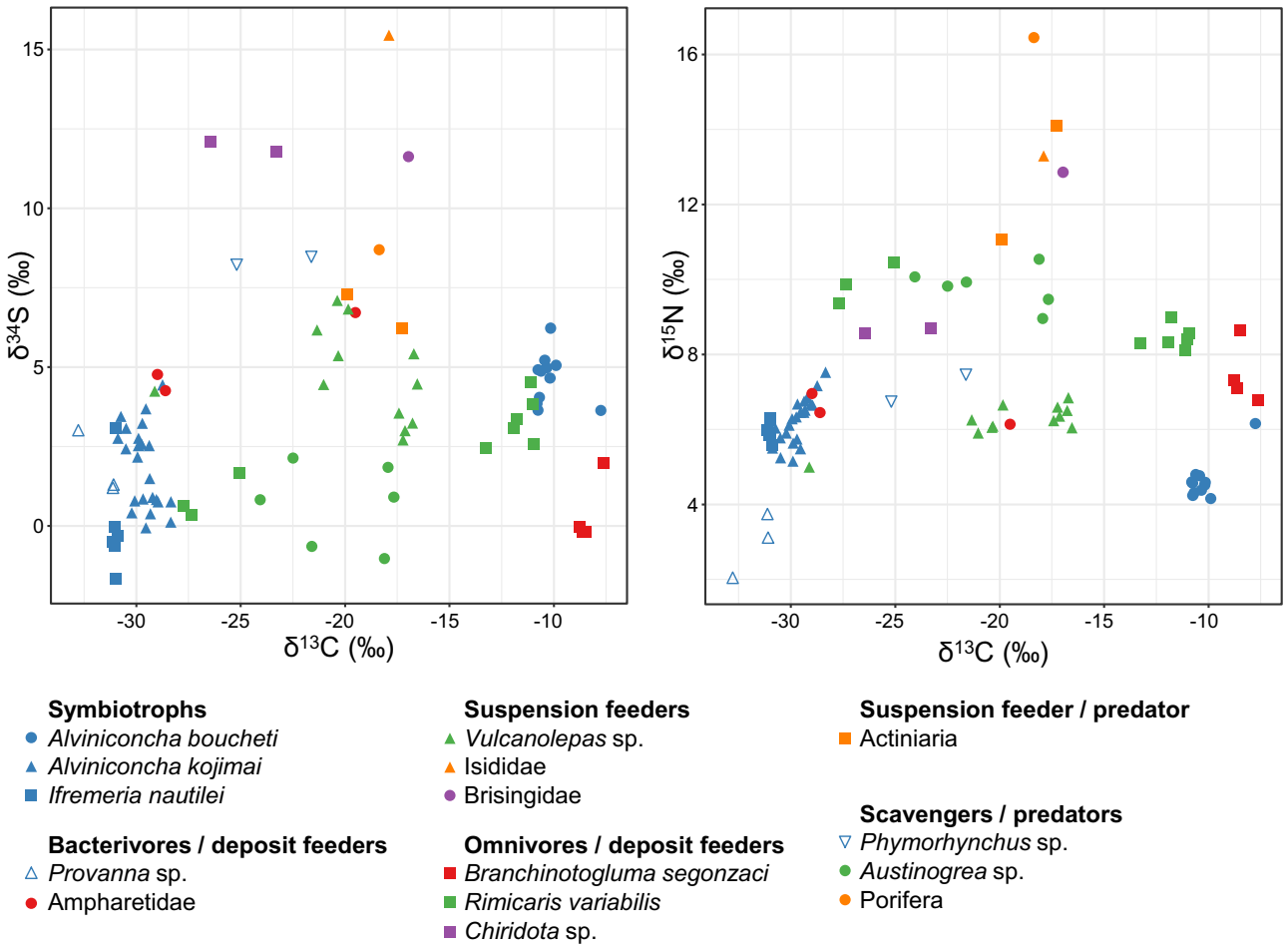
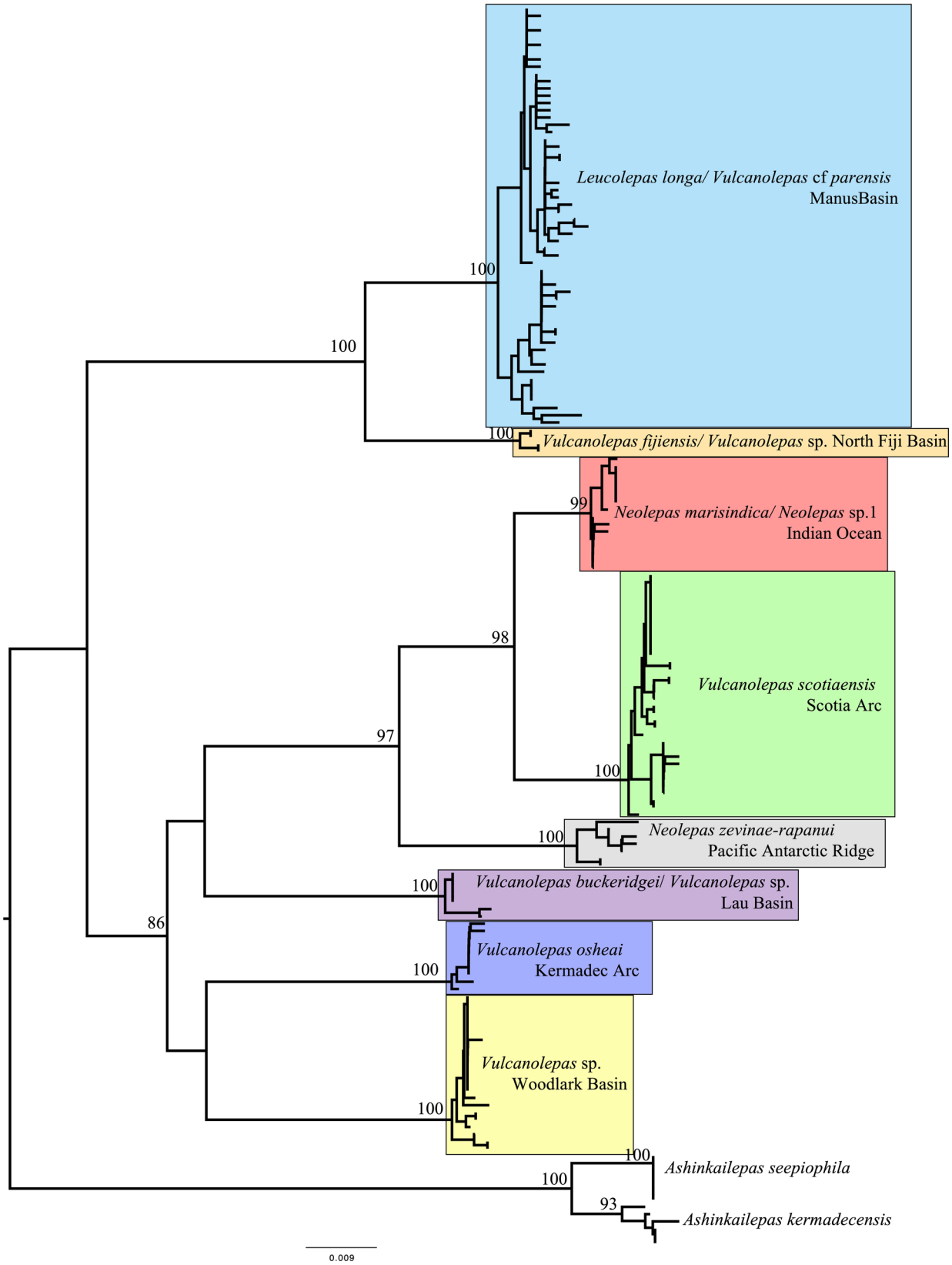


Fig. 6 Stable isotope food web diagram of LSVF animal communities. Stable isotope ratios of carbon vs. sulfur (left) and carbon vs. nitrogen (right) in benthic fauna from La Scala, Woodlark Basin. Blue: molluscs, red: polychaetes, green: crustaceans, orange: sponges and cnidarians, purple: echinoderms. Organisms have been assigned to functional feeding guilds according to literature data.

that have now disappeared by subduction (Fig. 9). The inactive South Fiji or Solomon ridges are relicts of such past ridge systems[17] and may have been transiently linked by the formation of the North Fiji proto-basin after the collision of the Melanesian arc and the Ontong Java plateau about 18 Mya. Amongst back-arc basins, the Woodlark ridge opened (about 6–7 million years[17]) at a time where the northern expansion of the North Fiji proto-basin became extinct, and corresponds to a region that may still contain traces of older dispersal pathways during the Oligocene (i.e., the Solomon Seaway). The Woodlark ridge could thus act as a biodiversity dispersion centre for the modern hydrothermal vent fauna and a crossroad between the Manus, North Fiji, and Lau Basins. To this extent, the discovery of the LSVF on Segment 3 of the Woodlark ridge represents an important opportunity to study the biogeography of the vent fauna associated with back-arc basins. A first examination of our list of species collected at La Scala indicates that the Woodlark assemblage is not noticeably different from other Western Pacific assemblages. It also shows that the basalt-hosted MOR vent system (such as La Scala) and its greater depth do not represent a barrier for most of the vent fauna. Although several studies highlighted genetic differentiation in populations of vent species between the Manus and Lau/Fiji basins[42–44], the addition of populations sampled at Woodlark in genetic analyses will bring further insight into biogeographic patterns and the role of the Woodlark ridge in bridging populations at the regional scale. More population genetics studies using genomic tools are currently underway. So far, the barcoding of symbiotic gastropod species reinforces the view that the Woodlark community, although deep, can be connected to much shallower populations in the Manus Basin (e.g., SuSu volcanoes located nearby, at the opening of the basin) but is also likely to act as a relay for migration between Manus and Fiji/Futuna/Lau, at least for *Ifremeria nautilei* (Fig. 8)[45]. Based on a more general barcoding approach, the Woodlark ridge does not seem to only represent a stepping stone but also a contact zone for some species between these basins (Poitrimol, pers. Comm.). Even if most larval dispersal trajectory modelling in the Southwestern Pacific failed to explain or only weakly explained present-day inter-basin connectivity for the deep-sea fauna even for larvae with a long pelagic duration[46,47], these studies and our own genetic results pointed out the possible role of the Woodlark ridge as a biological cornerstone.

Despite faunal similarities with other Western Pacific communities, communities in the Woodlark Basin may have specific attributes due to the greater depth (3330 mbsl) at which the vent sources are located compared to the shallower vent sites encountered elsewhere in the western Pacific (Manus Pual Ridge, Susu volcanoes, Franklin and Edison seamounts, North Fiji and Lau ridges segments). We found two new species (*Vulcanolepas* sp. nov. and *Austinograea* sp. nov.) at La Scala so far. Further examination of specimens that have been identified at the genus level as the gastropod *Desbruyeresia* or *Provanna* could yield additional new species.



The most striking features of this newly discovered vent assemblage, however, are the scarcity of the symbiotic fauna (no bathymodiolin mussels and small numbers of the large gastropods *Ifremeria nautilei* and *Alviniconcha* spp.) and the astonishing profusion of stalked barnacles (*Vulcanolepas* sp. nov.) that cover most of the vent edifices and their surroundings over nearly 1000 m^2 of basaltic scree. Very extensive dense populations of stalked barnacles have also been observed at Haungaroa (Kermadec Arc, Hourdez, pers. obs.), at deep-sea vents discovered along the flanks of the nearby Edison seamount (near the Lihir island of New Ireland, Papua New Guinea[39]), along the East Scotia Ridge in the Southern Ocean[48] or, along the Southern East

Fig. 7 Phylogenetic network of the LSVF stalked barnacle. Phylogenetic position of the Woodlark stalked barnacle in the genus *Vulcanolepas*. BioNJ tree on Kimura-2-Parameter distances based on a 476 bp alignment of the mitochondrial gene *CoxI*. Numbers above branches are bootstrap values for 100 replicates. Sequences for *Ashinkailepas kermadecensis* and *A. seepiophila* used as outgroup. Accession numbers: *A. kermadecensis* (KP295001, 19, 40, 48, 53, and 61), *A. seepiophila* (KP295022, 28, 31, 46, 69, 90, and 91), *Vulcanolepas* sp. Woodlark (**MW602536-40**, **602552-66**), *Vulcanolepas osheai* (**MW602550-51**, KP295005, 08, 26, 34, 36, 49, 56, and 94), *Vulcanolepas buckeridgei* (KY502196, 97, KP295009, 33, 41, 51, and 80), *Neolepas zevinae rapanui* (KP295007, 55, 60, 63, 67, 84, and 98), *Vulcanolepas scotiaensis* (KP295013, 14, 18, 21, 35, 37, 39, 42, 45, 50, 52, 57, 58, 68, 78, 97, KF739820-38), *Neolepas marisindica/Neolepas* sp. 1 (KP295004, 30, 32, 47, 62, 64, 89, and LC350007-15), *Vulcanolepas fijiensis/Vulcanolepas* sp. (MH636381-83, MN061491), *Leucolepas longa/Vulcanolepas* cf *parensis* (**MW602541-49**, KP295027, 73, 76, 82, 83, 85, and JX036420-64). Accession number that appear in bold were generated for this study.

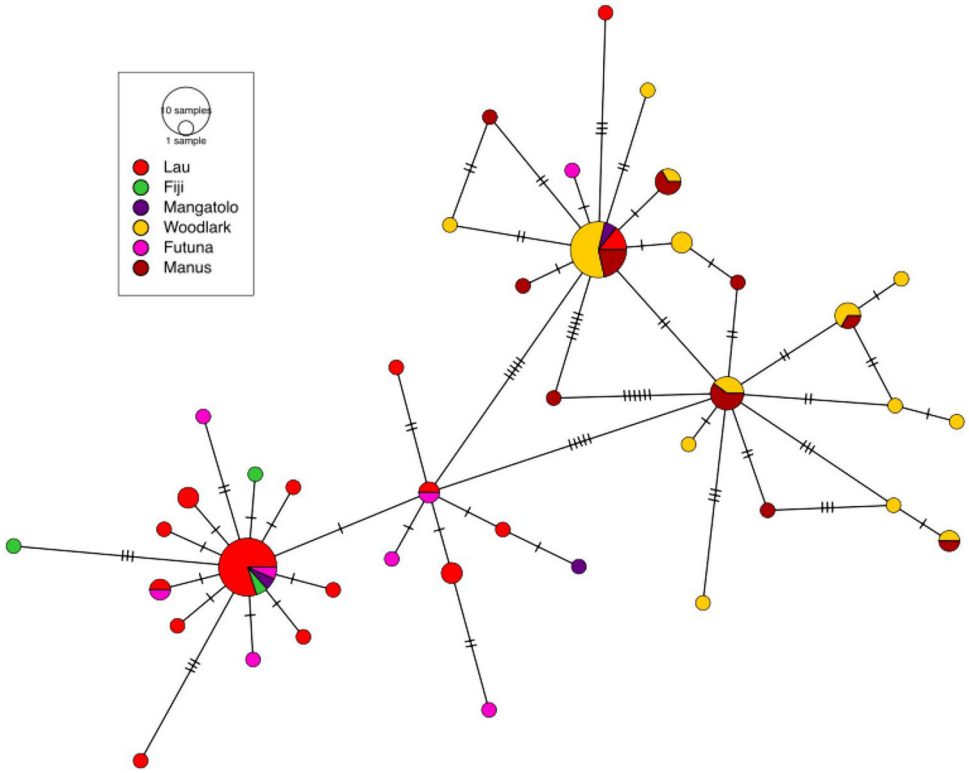


Fig. 8 Haplotype network of the LSVF gastropod *Ifremeria*. Haplotype network of the Woodlark *Ifremeria Cox1* gene (accession numbers: OL448876 - OL448957) with other sequences of geographically referenced individuals using the Median Joining method implemented in the software PopArt[61]. Purple: Woodlark Ridge, pink: Manus basin, Red: Lau basin, Green: North Fiji basin, yellow: Futuna.

Pacific Rise[49]. For each location, however, the stalked barnacle species are clearly different. They probably exploit the suspended particles that can disperse over large distances. This hypothesis is supported by the presence of large bacterial mats at the periphery that most likely also benefit from plume material carried from the main smokers, as observed on the mid-Atlantic ridge[50]. The absence of mussel beds and siboglinid worm agregations, and the fact that symbiotic gastropods are only found in small numbers at the base of the black smoker chimneys do not seem to be attributable to the greater depth of hydrothermal sources. Dense populations of *Alviniconcha* and *Bathymodiolus* have indeed previously been observed at deeper sites (3600 mbsl) on the Mariana Trench[51]. Comparing the vent fauna of Mariana Trench and Mariana Trough (1470 mbsl), Fujikura (1997)[52] concluded that depth had no impact on the species composition of vent communities in the western Pacific. A similar situation was also depicted at the Mid-Caiman Spreading Centre where the community from the Beebe vent field was almost identical to that of the Van Damm vent site, 2000 m shallower[53]. Alternatively, the high number of filter-feeding species (i.e., barnacles) and the reduced, patchy distribution of symbiotic gastropods may represent a transient state of the community where diffuse venting is prevented by focusing the hydrothermal activity into a series of black smokers exporting most of the fluid as high-velocity buoyant plumes that rise to a few tens of metres above the seafloor. This may indicate that the site has recently been reactivated by the tectonic movements that led to the collapse of part of the ridge crest where the emissions have been found, although activity in the general area probably lasted over several thousands of years.

Materials and methods

Bathymetry and acoustic survey. Ship-borne multibeam data were acquired using a Kongsberg EM122 1° × 2° 12 kHz while ROV multibeam data were acquired with a RESON 7125 400 kHz 0.5° × 1°. Both datasets were processed with the GLOBE software (doi.org/10.17882/70460) to provide 20 and 10 m grid (vessel data) and 1 m grid (ROV data) spaced digital terrain models of the surveyed area. The acoustic water column data acquired with the ship-borne 12 kHz MBES at a speed of 8 knots were characterized using the SonarScope (@Ifremer) and GLOBE softwares. Hydrothermal echoes were detected and located by visual inspection of ship-borne water-column polar echograms corresponding to the data of each ping and long-distance echograms associated to the track of the vessel. The most likely position on the seafloor and the maximum height (altitude above the seafloor) of each group of echoes attributed to rising plumes were determined. With this

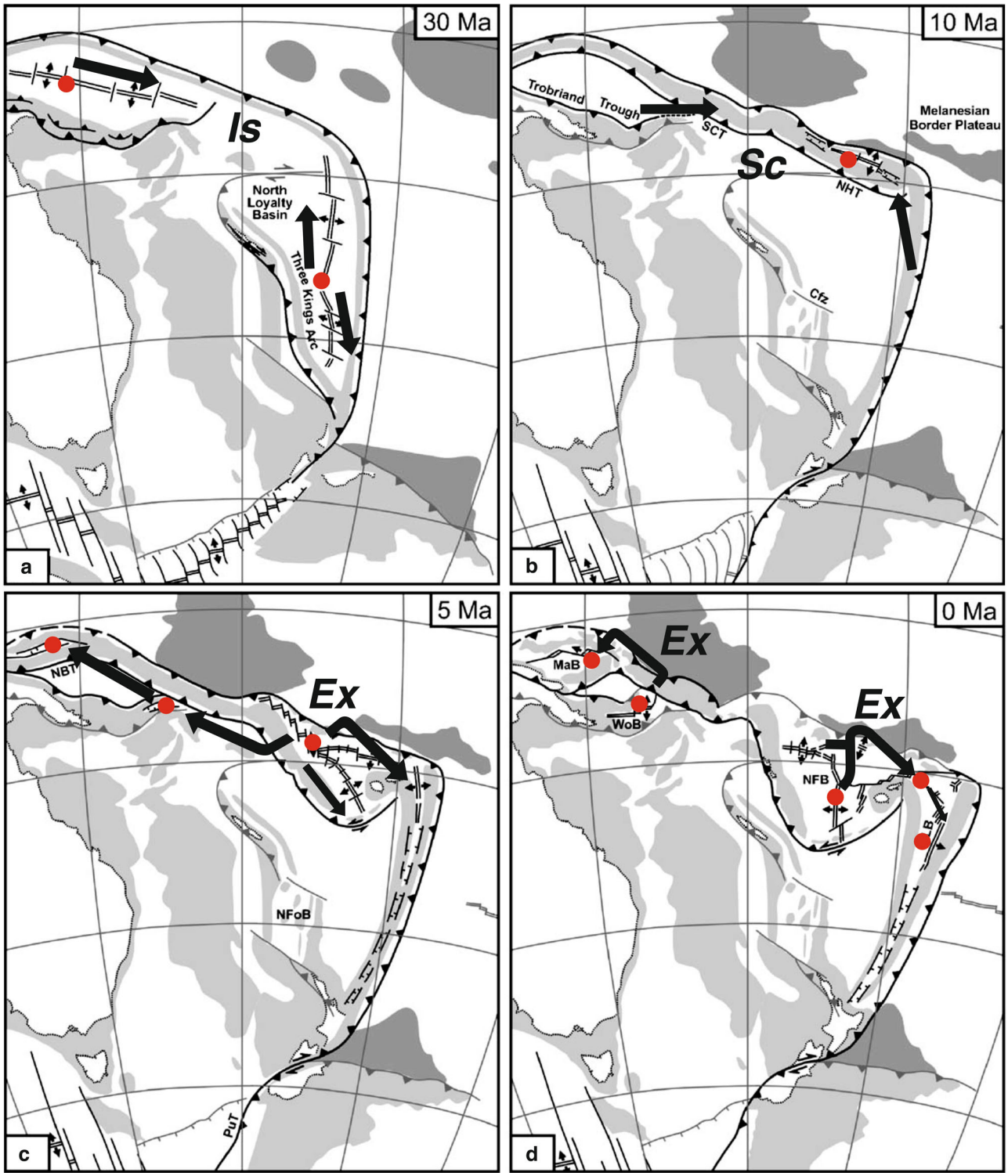


Fig. 9 The Woodlark ridge, a potential biological cornerstone?. Scenario of the Woodlark Ridge's hypothesis as a cornerstone of the vent fauna dispersal in the western Pacific using the geotectonic reconstruction model of the Western back-arc basins adapted from Schellart et al. (2006)[16]. **a** Formation of the Solomon Basin and North Loyalty/South Fiji Basin with two independent ridge systems. **b** Opening of the northern branch of the North Fiji basin after the fossilization of the South Fiji basin and the opening of the Woodlark ridge about 10 Mya. **c** Expansion of North Fiji basin together with the opening of the Lau Basin 5 Mya. **d** Present-days back arc basins with the five hydrothermally active zones, including Futuna. Red dots: active spreading centres, black arrows: putative colonization routes, **Is**: species isolation phase, **Sc**: secondary contact, and **Ex**: population expansion phase.

package, acoustic anomalies were identified, analyzed and attributed to various scatterers (large biological echoes, seafloor reflections…). Anomalies were considered as attributable to hydrothermal plumes if they were connected to the seafloor and reproducible over time, i.e., detected several times at the same location.

Water column operations. All operations were conducted using a 24-Niskin bottle rosette frame onto which were mounted a Seabird CTD 911+, two Turbidimeters (Seapoint Turbidity Metres), a pH sensor (AMT GmBH) and an Eh sensor (AMT GmBH), as well as an altimeter for seafloor detection. The Seasave software provided by Seabird Electronics was used for real-time data acquisition

and display of the down- and up-casts data. Niskin bottles were fired during up-casts at different levels in the water column, whenever an anomaly (T, S, turbidity and Eh) appeared in the real-time data display. The CTD-rosette was deployed in two ways, either as vertical casts or as towed casts ('tow-yos'), which consisted in lowering and raising the CTD-rosette between a constant set depth and ~100 m above the seafloor while the ship moved along a transect at a maximum speed of 1 knot. Vertical casts were stopped 5 m above the seafloor (provided by the altimeter).

Water samples were drawn from the Niskin bottles, fitted with Teflon stopcocks and sealed with Viton O-rings, both compatible with metal analyses. All subsamples were fully processed within 4 h after the CTD-rosette arrived on deck. Several types of samples were collected: (1) unfiltered water used for total dissolvable metal concentrations after acidification to pH 1.8; (2) water filtered directly from the niskin bottles using gravity flow through acid-cleaned Acropak® filters equipped with 0.45 um Supor® Membrane. All fluid transfers from the Niskin bottles were performed under controlled conditions designed to avoid shipborne contaminations, including the use of acid-clean tubing and sampling LDPE bottles placed inside double plastic bags and protected by filling bell outlet. Each precleaned LDPE bottle was rinsed 3 times with seawater sample before final filling. All samples were acidified to pH 1.8 with Optima-grade HCl inside a clean-air flow bench and analyzed back onshore for dissolved and total dissolvable metals (including Mn).

Hydrothermal fluid sampling. The hydrothermal fluid from each black smoker chimney was sampled in duplicate into two 750 mL syringe-style Titanium samplers manipulated by the ROV arm to estimate concentrations of end-members. In total, six samples of fluids from three chimneys were recovered immediately after removing the top of the chimney and measuring temperature with the ROV probe. Temperature was also recorded in real-time during the fluid collection using thermocouple temperature probes (NKE® probes) mounted on the sampler snorkel. Fluids were processed on board straight away after recovery using the following scheme:

i. Aliquots were also transferred directly into evacuated glass bottles that contain a pre-weighed amount of Zn-acetate [$Zn(CH_3COO)_2{\sim}2H_2O$] to precipitate ZnS for $\delta^{34}S$ and H_2S concentration (i.e., total HS^-) measurements.
ii. Aliquots were then collected for shipboard analysis (pH and total HS^- using either titration or Cline, method[54] depending on the concentration).
iii. The remaining fluid solutions in the Ti-samples were then transferred into an acid cleaned 1 L HDPE bottle and homogenized before further separated into aliquots for on shore analyses: ion chromatrographic analysis (major cations/anions; data not reported here); HR-ICPMS (major and trace elements); silica concentration after 100-fold dilution with ultra-pure water to prevent silica precipitation. Due to the high concentrations of metals in the hydrothermal solutions, precipitation often occurs within the titanium samplers as they cool down to ambient temperature. Those particles were found to be nearly entirely transferred into the 1 L HDPE bottle. Precipitated particles remaining in the Ti-samplers were however recovered for chemical analysis when the samplers are disassembled by rinsing with Milli-Q water and ethanol, and filtered through 0.45 µm filters. This fraction is hereafter referred to as "dregs" (not reported here).

On shore fluid analyses. Major and trace elements were measured by high-resolution inductively coupled plasma mass spectrometry (HR-ICPMS) Element XR operated at the French Research Institute for Exploitation of the Sea (IFREMER). Indium solution was added before analysis or mixed on-line at a final concentration of 5 ppb to correct for instrument sensitivity changes. Solutions were diluted 100-fold and introduced into the plasma torch using a quartz spray chamber system equipped with a microconcentric PFA nebulizer operating at a flow rate of about 60 µl/min. For each element, ICPMS sensitivity was calibrated using matrix matched standard solutions corresponding to seawater matrices. Anions (Cl, SO_4) were analyzed by ionic chromatography after appropriate dilutions.

Sulfide analysis (XRF, XRD, HR-ICPMS, $\delta^{34}S$). Rock and sulfide samples collected during the dives were petrographically characterized directly on-board. Two types of subsamples were considered: A representative portion of the sample (referred as bulk) or mineral separates. Both sample types were dried in an air oven at 50 °C and ground to a fine powder using an agate mortar.

Mineralogical and quantitative chemical data were acquired using X-ray techniques (XRD and X-ray fluorescence) at the Laboratoire de Géochimie et Métallogénie at IFREMER. XRD analyses were conducted with a BRUKER AXS D8 Advance diffractometer. Samples were top loaded into 2.5 cm-diameter circular cavity holders, and analyses were run between 5° and 70° 2θ, in steps of 0.01° 2θ at 1 s/step (monochromatic Cu Kα radiation, 40 kV, 30 mA). Minerals were identified using the Diffrac suite EVA software.

X-ray fluorescence analyses were conducted with a wavelength dispersive X-ray fluorescence spectrometer (WD-XRF; BRUKER AXS S8 TIGER) on fusion beads (for major elements) or compressed pellets (for trace elements). After data acquisition, the measured net peak intensities, corrected for inter-element effects, were converted into concentrations using calibration curves generated from analysis of certified geochemical standard powders (measured under identical analytical conditions). Calibrations were established using a set of certified materials obtained mainly from the Canadian Certified reference materials Project (CCRMP) (for example, CCU-1, CZN-1, Fer-1, PTC-1 or PTM-1), Geological Survey of Japan (GSJ)(JP-1), and the Centre de Recherches Pétrochimiques et Géochimiques in France (CRPG)(BE-N).

For sulfur isotopes, sample powder was digested in inverted aqua regia and then purified by elution through a cationic resin to remove any interfering matrix. The solution was analyzed by a Neptune MC-ICP-MS at Ifremer following the procedure of Craddock et al. (2008)[55] and calibrated against a set of internal and external standards (IAEA NZ-1, NZ-2, NZ-4, NBS 123).

All mineral samples were also analyzed for major and trace elements by HR-ICPMS at IFREMER. About 100 mg of dry powder was dissolved in a PTFE beaker on a hot plate using an acid mixture of HCl, and HNO_3. Digested samples were then further diluted 50-fold and analyzed by HR-ICPMs following similar approaches than for fluid samples described above. A set of international georeference materials (e.g., BHVO-2, GH, UB-N, NOD-P-1, IF-G, NIST2711) and internal standard solutions relevant to sulfide-rich samples were used to calibrate the measurements.

Sampling and preservation of animals. Benthic macrofauna was collected using the claw of the hydraulic arm of the ROV Victor 6000 and placed into collection boxes. The remaining fauna on the substrate was then sampled with the suction sampler of the ROV equipped with a 1 mm mesh. Three benthic communities defined by engineer species were specifically sampled over an area of about 0.5 m^2 (i.e., one sample for the *Vulcanolepas* community, one sample for the *Ifremeria* community and two samples for the *Alviniconcha* community). Temperature measurements were made at three locations using the ROV temperature probe prior and after sampling for *Ifremeria* and *Alviniconcha* communities. In addition, opportunistic sampling was carried out to collect organisms found on the black smokers (e.g., polynoid worms, crabs and shrimp) and peripheral area using the robotic arm and claw of the ROV. This sampling was supplemented by observations performed with the high-definition video camera of the ROV for large megafauna. Onboard, samples were sieved on a 250 µm sieve, and all organisms were sorted and preserved in 96° ethanol before their identification at the lowest taxonomic level based on morphological criteria. Some specimens of each species were also frozen at −80 °C in order to perform stable isotopes (^{13}C, ^{15}N, and ^{34}S) analyses.

Stable isotope analyses. On board, animals were dissected to separate soft and non-metabolically active tissues (e.g., muscle, tegument) or, when body size was small, were used whole[56]. All samples were oven-dried at 60 °C for 72 h, then placed in airtight containers and kept at room temperature before further treatment once back from the expedition. They were subsequently ground to a homogeneous powder using mortar and pestle. Samples containing hard inorganic carbon parts that could not be physically removed were acidified by exposing them to HCl vapours for 48 h in an airtight container[57]. Stable isotope ratios measurements were performed via continuous flow—elemental analysis— isotope ratio mass spectrometry (CF-EA-IRMS) at University of Liège (Belgium), using a vario MICRO cube C–N–S elemental analyzer (Elementar Analysensysteme GMBH, Hanau, Germany) coupled to an IsoPrime100 isotope ratio mass spectrometer (Isoprime, Cheadle, United Kingdom). Isotopic ratios were expressed using the widespread δ notation[58], in ‰ and relative to the international references Vienna Pee Dee Belemnite (for carbon), Atmospheric Air (for nitrogen) and Vienna Canyon Diablo Troilite (for sulfur). IAEA (International Atomic Energy Agency, Vienna, Austria) certified reference materials sucrose (IAEA-C-6; $\delta^{13}C = -10.8 \pm 0.5$‰; mean ± SD), ammonium sulfate (IAEA-N-2; $\delta^{15}N = 20.3 \pm 0.2$‰; mean ± SD) and silver sulfide (IAEA-S-1; $\delta^{34}S = -0.3$‰) were used as primary analytical standards. Sulfanilic acid (Sigma-Aldrich; $\delta^{13}C = -25.6 \pm 0.4$‰; $\delta^{15}N = -0.13 \pm 0.4$‰; $\delta^{34}S = 5.9 \pm 0.5$‰; means ± SD) was used as secondary analytical standard. Standard deviations on multi-batch replicate measurements of secondary and internal lab standards (amphipod crustacean muscle) analyzed interspersed with samples (one replicate of each standard every 15 analyses) were 0.2‰ for both $\delta^{13}C$ and $\delta^{15}N$ and 0.5‰ for $\delta^{34}S$.

Genetic analyses

Barcoding analyses. Genomic DNA from gastropod, barnacle and crab specimens collected from the site La Scala were extracted using a CTAB extraction procedure. Tissues were digested overnight in 600 µl of a 1% CTAB buffer solution (1.4-M NaCl, 0.2% 2-mercaptoethanol, 20 mM EDTA, 100-mM Tris-HCl pH 8 and 0.1 mg ml^{-1} proteinase K) also containing 1% PVPP (PolyVinylPolypyrolidone). DNA was then extracted by adding chloroform–isoamyl alcohol (24:1), precipitated with 1 ml of cold 100% ethanol, washed with 70% ethanol and resuspended in 100 µl of sterile solution of 0.1X TE (Tris EDTA pH 8.0). For shrimp, genomic DNA was extracted using E.Z.N.A. Tissue DNA kit (Omega Bio-Tek) according to the manufacturer's recommendations. For the mt *Cox1* gene,

amplifications were performed in 25 µl final volume using the 'universal' applicable primers of Folmer et al. (1994)[59] and the following conditions: 1× reaction buffer), 2.5 mM $MgCl_2$, 0.12 mM of each dNTP, 0.38 µM of each primers, 1U Taq DNA polymerase (Uptitherm), 2.5 µl of template DNA and sterile H_2O. Thermal cycling parameters used an initial denaturation at 94 °C for 2 min, followed by 35 cycles at 94 °C for 30 s, 50 °C for 1 min, 72 °C for 2 min and a final 5 min extension cycle. An alternative protocol was used for shrimp with specific primers for alvinocaridid shrimp[60].

Phylogenetic and haplotype networks. A phylogenetic tree was produced determine the position of the new pedunculate cirriped by a PhyML approach using a general time reversible model of substitutions with a Gamma distribution of the substitution rate. This latter rate, nucleotide equilibrium frequencies, and the proportion of invariant sites were optimized by the search algorithm. Node confidence was evaluated with an αLRT (SH-like) approach. Haplotype networks for the gastropods were produced using the minimum spanning method implemented in PopArt[61] on mt*Cox1* alignments of the three main symbiotic gastropod species.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The datasets generated during the current study are available at https://doi.org/10.6084/m9.figshare.19095659. All bathymetric data and sampling metadata from the cruise are available at https://doi.org/10.17600/18001111.

Received: 27 April 2021; Accepted: 11 February 2022;
Published online: 17 March 2022

References

1. German, C. R. & Von Damm, K. L. *Treatise on Geochemistry* (eds Heinrich, D. H. & Karl, K. T.) 181–222 (Pergamon, 2003).
2. Van Dover, C. *The Ecology of Deep-Sea Hydrothermal Vents* (Princeton University Press, 2000).
3. Spiess, F. N. et al. East Pacific rise: Hot springs and geophysical experiments. *Science* **207**, 1421–1433 (1980).
4. Haymon, R. M. et al. Hydrothermal vent distribution along the East Pacific Rise crest 9° 09'–54' N and its relationship to magmatic and tectonic processes on fast-spreading mid-ocean ridges. *Earth Planetary Sci. Lett.* **104**, 513–534 (1991).
5. Edmonds, H. N. et al. Discovery of abundant hydrothermal venting on the ultraslow-spreading Gakkel Ridge in the Arctic Ocean. *Nature* **421**, 252–256 (2003).
6. German, C. R. et al. Hydrothermal activity and seismicity at Teahitia Seamount: Reactivation of the society islands hotspot? *Front. Mar. Sci* **7**, 73 (2020).
7. de Ronde, C. E. J. et al. Intra-oceanic subduction-related hydrothermal venting, Kermadec volcanic arc, New Zealand. *Earth Planetary Sci. Lett.* **193**, 359–369 (2001).
8. Ishibashi, J. & Urabe, T. *Backarc Basins: Tectonics and Magmatism* (ed Taylor, B.) 451–495 (Springer, 1995).
9. Fouquet, Y. et al. Hydrothermal activity and metallogenesis in the Lau back-arc basin. *Nature* **349**, 778–781 (1991).
10. Boschen, R. E., Rowden, A. A., Clark, M. R. & Gardner, J. P. A. Mining of deep-sea seafloor massive sulfides: A review of the deposits, their benthic communities, impacts from mining, regulatory frameworks, and management strategies. *Ocean Coastal Manage.* **84**, 54–67 (2013).
11. Lisitsyn, A. P. et al. Active hydrothermal activity at Franklin Seamount, Western Woodlark Sea (Papua New Guinea). *Int. Geol. Rev.* **33**, 914–929 (1991).
12. Laurila, T. E. et al. Tectonic and magmatic controls on hydrothermal activity in the Woodlark Basin: Hydrothermalism in the Woodlark Basin. *Geochem. Geophys. Geosyst.* **13**, Q09006 (2012).
13. Goodliffe, A. M. et al. Synchronous reorientation of the Woodlark Basin spreading center. *Earth Planetary Sci. Lett.* **146**, 233–242 (1997).
14. Martínez, F., Taylor, B. & Goodliffe, A. M. Contrasting styles of seafloor spreading in the Woodlark Basin: Indications of rift-induced secondary mantle convection. *J. Geophys. Res.* **104**, 12909–12926 (1999).
15. Taylor, B., Goodliffe, A., Martinez, F. & Hey, R. Continental rifting and initial sea-floor spreading in the Woodlark Basin. *Nature* **374**, 534–537 (1995).
16. Schellart, W. P., Lister, G. S. & Toy, V. G. A Late Cretaceous and Cenozoic reconstruction of the Southwest Pacific region: Tectonics controlled by subduction and slab rollback processes. *Earth-Sci. Rev.* **76**, 191–233 (2006).
17. Hall, R. Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: Computer-based reconstructions, model and animations. *J. Asian Earth Sci.* **20**, 353–431 (2002).
18. Breusing, C. et al. Allopatric and sympatric drivers of speciation in Alviniconcha hydrothermal vent snails. *Mol. Biol. Evol.* **37**, 3469–3484 (2020).
19. Ondréas, H., Scalabrin, C., Fouquet, Y. & Godfroy, A. Recent high-resolution mapping of Guaymas hydrothermal fields (Southern Trough). *BSGF - Earth Sci. Bull.* **189**, 6 (2018).
20. Nakamura, K. et al. Water column imaging with multibeam echo-sounding in the mid-Okinawa Trough: Implications for distribution of deep-sea hydrothermal vent sites and the cause of acoustic water column anomaly. *Geochem. J.* **49**, 579–596 (2015).
21. Xu, G., Jackson, D. R. & Bemis, K. G. The relative effect of particles and turbulence on acoustic scattering from deep sea hydrothermal vent plumes revisited. *J. Acoust. Soc. Am.* **141**, 1446–1458 (2017).
22. Park, S.-H. et al. Petrogenesis of basalts along the eastern Woodlark spreading center, equatorial western Pacific. *Lithos* **316–317**, 122–136 (2018).
23. Chadwick, J. et al. Arc lavas on both sides of a trench: Slab window effects at the Solomon Islands triple junction, SW Pacific. *Earth Planetary Sci. Lett.* **279**, 293–302 (2009).
24. Fouquet, Y. et al. *Geodiversity of Hydrothermal Processes Along the Mid-Atlantic Ridge and Ultramafic-Hosted Mineralization: A New Type of Oceanic Cu-Zn-Co-Au Volcanogenic Massive Sulfide Deposit* (eds Rona, P. A., Devey, C. W., Dyment, J. & Murton, B. J.) Vol. 188, 321–367 (American Geophysical Union, 2010).
25. Von Damm, K. et al. Chemistry of submarine hydrothermal solutions at 21N, East Pacific Rise. *Geochim. Cosmochim. Acta* **49**, 2197–2220 (1985).
26. Seyfried, W. E. & Bischoff, J. L. Experimental seawater-basalt interaction at 300 °C, 500 bars, chemical exchange, secondary mineral formation and implications for the transport of heavy metals. *Geochim. Cosmochim. Acta* **45**, 135–147 (1981).
27. Pester, N. J., Rough, M., Ding, K. & Seyfried, W. E. A new Fe/Mn geothermometer for hydrothermal systems: Implications for high-salinity fluids at 13°N on the East Pacific Rise. *Geochim. Cosmochim. Acta* https://doi.org/10.1016/j.gca.2011.08.043 (2011).
28. Podowski, E. L., Moore, T. S., Zelnio, K. A., Luther, G. W. & Fisher, C. R. Distribution of diffuse flow megafauna in two sites on the Eastern Lau Spreading Center, Tonga. *Deep Sea Res. Part I: Oceanogr. Res. Papers* **56**, 2041–2056 (2009).
29. Collins, P., Kennedy, R. & Van Dover, C. A biological survey method applied to seafloor massive sulphides (SMS) with contagiously distributed hydrothermal-vent fauna. *Mar. Ecol. Prog. Ser.* **452**, 89–107 (2012).
30. Desbruyères, D., Hashimoto, J. & Fabri, M.-C. Composition and biogeography of hydrothermal vent communities in Western Pacific back-arc basins. *Geophys. Monogr. Ser.* **166**, 215–234 (2006).
31. Reid, W. D. K. et al. Spatial differences in East scotia ridge hydrothermal vent food webs: Influences of chemistry, microbiology, and predation on trophodynamics. *PLoS One* **8**, e65553 (2013).
32. Van Audenhaege, L., Fariñas-Bermejo, A., Schultz, T. & Lee Van Dover, C. An environmental baseline for food webs at deep-sea hydrothermal vents in Manus Basin (Papua New Guinea). *Deep Sea Res. Part I: Oceanogr. Res. Papers* https://doi.org/10.1016/j.dsr.2019.04.018 (2019).
33. Erickson, K. L., Macko, S. A. & Van Dover, C. L. Evidence for a chemoautotrophically based food web at inactive hydrothermal vents (Manus Basin). *Deep-Sea Res. Part II: Top. Stud. Oceanogr.* **56**, 1577–1585 (2009).
34. Comeault, A., Stevens, C. J. & Juniper, S. K. Mixed photosynthetic-chemosynthetic diets in vent obligate macroinvertebrates at shallow hydrothermal vents on Volcano 1, South Tonga Arc—evidence from stable isotope and fatty acid analyses. *Cahiers de Biologie Marine* **51**, 351–359 (2010).
35. Bennett, S. A., Dover, C. V., Breier, J. A. & Coleman, M. Effect of depth and vent fluid composition on the carbon sources at two neighboring deep-sea hydrothermal vent fields (Mid-Cayman Rise). *Deep-Sea Res. Part I: Oceanogr. Res. Papers* **104**, 122–133 (2015).
36. Levin, L. A. et al. Hydrothermal vents and methane seeps: Rethinking the sphere of influence. *Front. Marine Sci.* **3**, 1–23 (2016).
37. Hügler, M. & Sievert, S. M. Beyond the Calvin cycle: Autotrophic carbon fixation in the ocean. *Annu. Rev. Mar. Sci.* **3**, 261–289 (2011).
38. Wang, X., Li, C., Wang, M. & Zheng, P. Stable isotope signatures and nutritional sources of some dominant species from the PACManus hydrothermal area and the Desmos caldera. *PLoS One* **13**, e0208887 (2018).
39. Tunnicliffe, V. & Southward, A. J. Growth and breeding of a primitive stalked barnacle Leucolepas longa (Cirripedia: Scalpellomorpha: Eolepadidae: Neolepadinae) inhabiting a volcanic seamount off Papua New Guinea. *J. Mar. Biol. Ass.* **84**, 121–132 (2004).
40. Auzende, J. M., Pelletier, B. & Lafoy, Y. Twin active spreading ridges in the North Fiji Basin (southwest Pacific). *Geology* **22**, 63–66 (1994).
41. Parson, L. M. & Wright, I. C. The Lau-Havre-Taupo back-arc basin: A southward-propagating, multi-stage evolution from rifting to spreading. *Tectonophysics* **263**, 1–22 (1996).

42. Thaler, A. D. et al. 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**, e101345 (2014).
43. Lee, W.-K., Kim, S.-J., Hou, B. K., Van Dover, C. L. & Ju, S.-J. Population genetic differentiation of the hydrothermal vent crab Austinograea alayseae (Crustacea: Bythograeidae) in the Southwest Pacific Ocean. *PLoS One* **14**, e0215829 (2019).
44. Plouviez, S. et al. Amplicon sequencing of 42 nuclear loci supports directional gene flow between South Pacific populations of a hydrothermal vent limpet. *Ecol. Evol.* https://doi.org/10.1002/ece3.5235 (2019).
45. Tran Lu Y, A. et al. Fine-scale genomic patterns of connectivity in the deep sea hydrothermal gastropod Ifremeria nautilei over its species range using outlier scans and demo-genetic inferences. *Mol. Ecol.* (In Revision).
46. Yearsley, J. M. & Sigwart, J. D. Larval transport modeling of deep-sea invertebrates can aid the search for undiscovered populations. *PLoS One* **6**, e23063 (2011).
47. Mitarai, S., Watanabe, H., Nakajima, Y., Shchepetkin, A. F. & McWilliams, J. C. Quantifying dispersal from hydrothermal vent fields in the western Pacific Ocean. *Proc. Natl Acad. Sci. USA* **113**, 2976–2981 (2016).
48. Marsh, L. et al. Microdistribution of faunal assemblages at deep-sea hydrothermal vents in the southern ocean. *PLoS One* **7**, e48348 (2012).
49. Jollivet, D. et al. The Biospeedo cruise: A new survey of hydrothermal vents along the south East Pacific Rise from 7°24' S to 21°33' S. *InterRidge News* **13**, 20–26 (2005).
50. Girard, F. et al. Currents and topography drive assemblage distribution on an active hydrothermal edifice. *Prog. Oceanogr.* **187**, 102397 (2020).
51. Hessler, R. R. & Lonsdale, P. F. Biogeography of Mariana Trough hydrothermal vent communities. *Deep Sea Res. Part A. Oceanogr. Res. Papers* **38**, 185–199 (1991).
52. Fujikura, K. Biology and earth scientific investigation by the submersible 'Shinkai 6500' system of deep-sea hydrothermal and lithosphere in the Mariana back-arc basin. *JAMSTEC J. Deep Sea Res.* **13**, 1–20 (1997).
53. Connelly, D. P. et al. Hydrothermal vent fields and chemosynthetic biota on the world's deepest seafloor spreading centre. *Nat. Commun.* **3**, 620 (2012).
54. Cline, J. D. Spectrophotometric determination of hydrogen sulfide in natural waters. *Limnol. Oceanogr.* **14**, 454–458 (1969).
55. Craddock, P. R., Rouxel, O. J., Ball, L. A. & Bach, W. Sulfur isotope measurement of sulfate and sulfide by high-resolution MC-ICP-MS. *Chem. Geol.* **253**, 102–113 (2008).
56. Mateo, M. A., Serrano, O., Serrano, L. & Michener, R. H. Effects of sample preparation on stable isotope ratios of carbon and nitrogen in marine invertebrates: Implications for food web studies using stable isotopes. *Oecologia* **157**, 105–115 (2008).
57. Hedges, J. I. & Stern, J. H. Carbon and nitrogen determinations of carbonate-containing solids1. *Limnol. Oceanogr.* **29**, 657–663 (1984).
58. Coplen, T. B. Guidelines and recommended terms for expression of stable-isotope-ratio and gas-ratio measurement results: Guidelines and recommended terms for expressing stable isotope results. *Rapid Commun. Mass Spectrom.* **25**, 2538–2560 (2011).
59. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3**, 294–299 (1994).
60. Methou, P., Michel, L. N., Segonzac, M., Cambon-Bonavita, M.-A. & Pradillon, F. Integrative taxonomy revisits the ontogeny and trophic niches of *Rimicaris* vent shrimps. *R. Soc. Open Sci.* **7**, 200837 (2020).
61. Leigh, J. W. & Bryant, D. Popart: Full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**, 1110–1116 (2015).

Acknowledgements

We are deeply grateful to the captain and crew of the French Research Vessel L'Atalante and to the team in charge of the ROV 6000 Victor without whom nothing would have been possible. We also thank Delphine Pierre (Ifremer CTDI) and Emilie Hardouin (Genavir) for their assistance in the production of the High-Resolution Bathymetric Chart of the Woodlark Ridge, as well as Anne-Sophie Alix (Ifremer GM) with the Geographic Information System. We finally thank Claire Daguin-Thiébaut, Stéphanie Ruault and Marion Ballenghien (Station Biologique de Roscoff) for their assistance in the molecular analyses. This work benefited from access to the Biogenouest genomic platform at Station Biologique de Roscoff. The authors would also like to dedicate this article to Jean-Marie Auzende who spent a large part of his life exploring the oceanic ridges of the Western Pacific to better understand the plate dynamics in this region. Finally, we thank the three anonymous reviewers for the useful comments in improving the manuscript. Ship time was supported by the French Oceanographic Fleet while scientific work was funded through the ANR 'CERBERUS' (contract number ANR-17-CE02-0003).

Author contributions

The cruise was led by D.J. and S.H. as co-principal investigators. They also acquired the genetic data on specimens collected during the cruise. C.B. is the coordinator of the present article and was responsible for the water column survey; O.R. was responsible for the geochemical analysis of the rocks and mineralization as well as the geochemical composition of the water column; C.S. was responsible for the acquisition of bathymetry and the processing of the acoustic data (ship and ROV); P.L.M. participated in the pre-treatment of the bathymetry data on board and produced the high-resolution bathymetric map; E.P. (not onboard) helped in the interpretation of the geological structures and geochemical composition of the fluids observed during the surveys; C.C. (not onboard) was responsible for the fluid sampling equipment and helped in the data processing from the CTD; C.P., E.T., and M.M. (not onboard) were responsible for the description of the fauna and the ecology of the sites; J.C. (not onboard) and A.T.L.Y. participated in the genetic analyses of the targeted fauna collected at LSVF. S.C. (not onboard) and A.B. (not onboard) produced the X-ray data; Y.G. carried out the analyses for the isotopic composition of fluids and mineralization; V.G. was in charge of hydrothermal fluids analysis (IC); L.M. investigated the isotopic ratios across the food web; S.A.-H., F.B., T.B., V.C.-.G., V.L.L., S.L'.H., J.M., A.-S.L.P., A.T., and D.C. Kuama participated and contributed to the sampling and data acquisition during the CHUBACARC Cruise (doi: 10.17600/18001111) during which the La Scala Vent field was discovered.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43247-022-00387-9.

Correspondence and requests for materials should be addressed to Cédric Boulart.

Peer review information *Communications Earth & Environment* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Maria-Luce Frezzotti and Clare Davis.

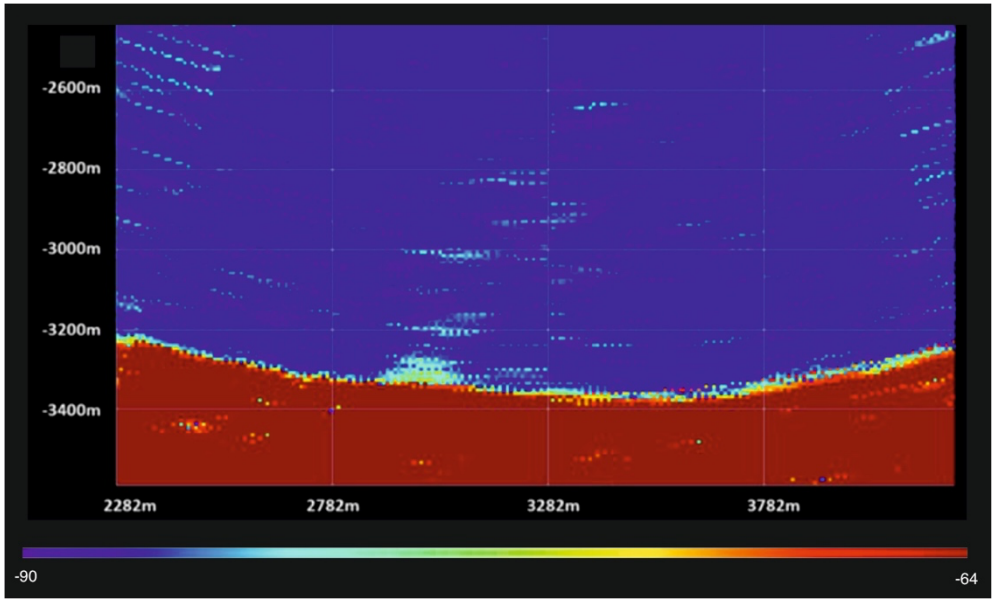
Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

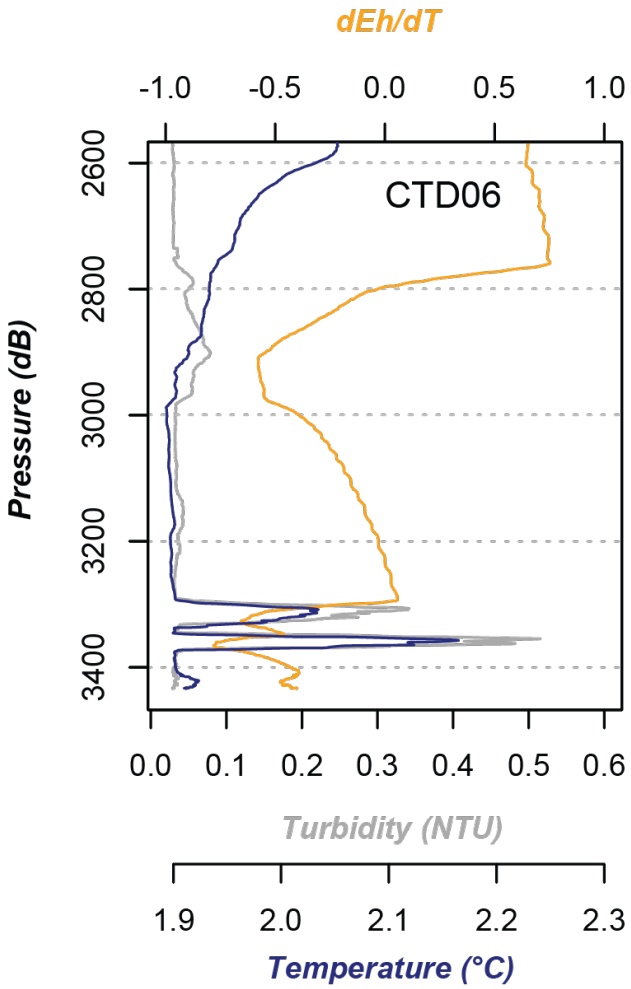
Supplementary information for "Active hydrothermal vents in the Woodlark Basin may act as dispersing centres for hydrothermal fauna"

Boulart, C.[1,*], Rouxel, O.[2], Scalabrin, C.[2], Le Meur, P.[3], Pelleter, E.[2], Poitrimol, C.[1,4], Thiebaut, E.[1], Matabos, M[4]., Castel, J.[1], Tran Lu Y, A.[5,10], Michel, L.N.[4], Cathalot, C.[2], Cheron, S.[2], Boissier, A.[2], Germain, Y.[2], Guyader, V.[2], Arnaud-Haond, S.[6], Bonhomme, F.[5], Broquet, T.[1], Cueff-Gauchard, V.[7], Le Layec, V.[1,10], L'Haridon, S[7]., Mary, J.[1], Le Port, A-S.[1], Tasiemski, A.[8], Kuama, D.C.[9], Hourdez, S.[10], Jollivet, D.[1]

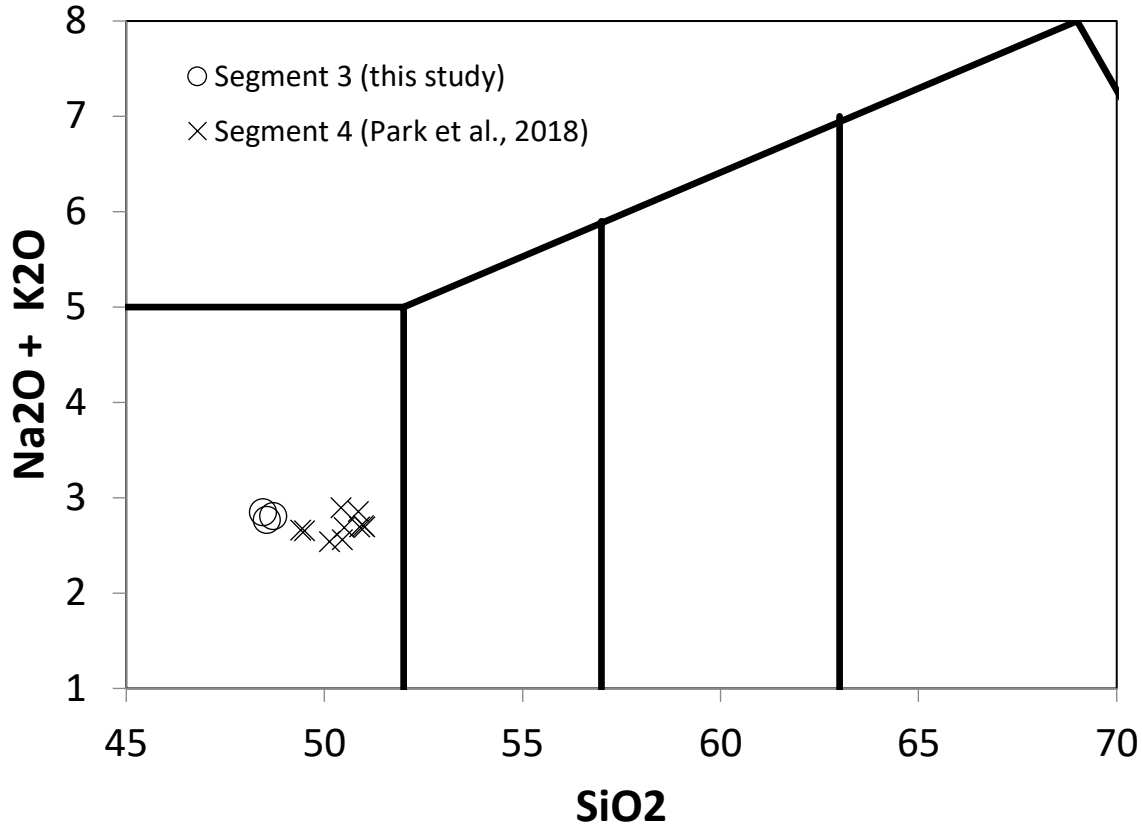
Supplementary Results and Discussion



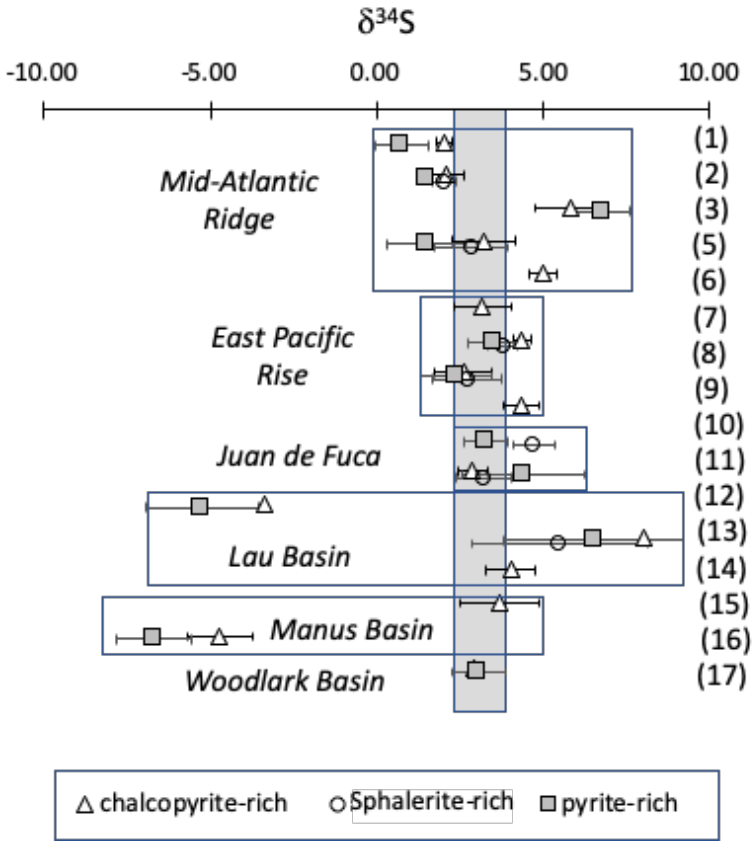
Supplementary Figure 1. Acoustic data of the water column. Expanded view of the post processed echograms of the second profile showing the echo attributed to La Scala hydrothermal vent field above the arc of the specular circle (red color) near the seafloor (colour scale -90 to -64 dB).



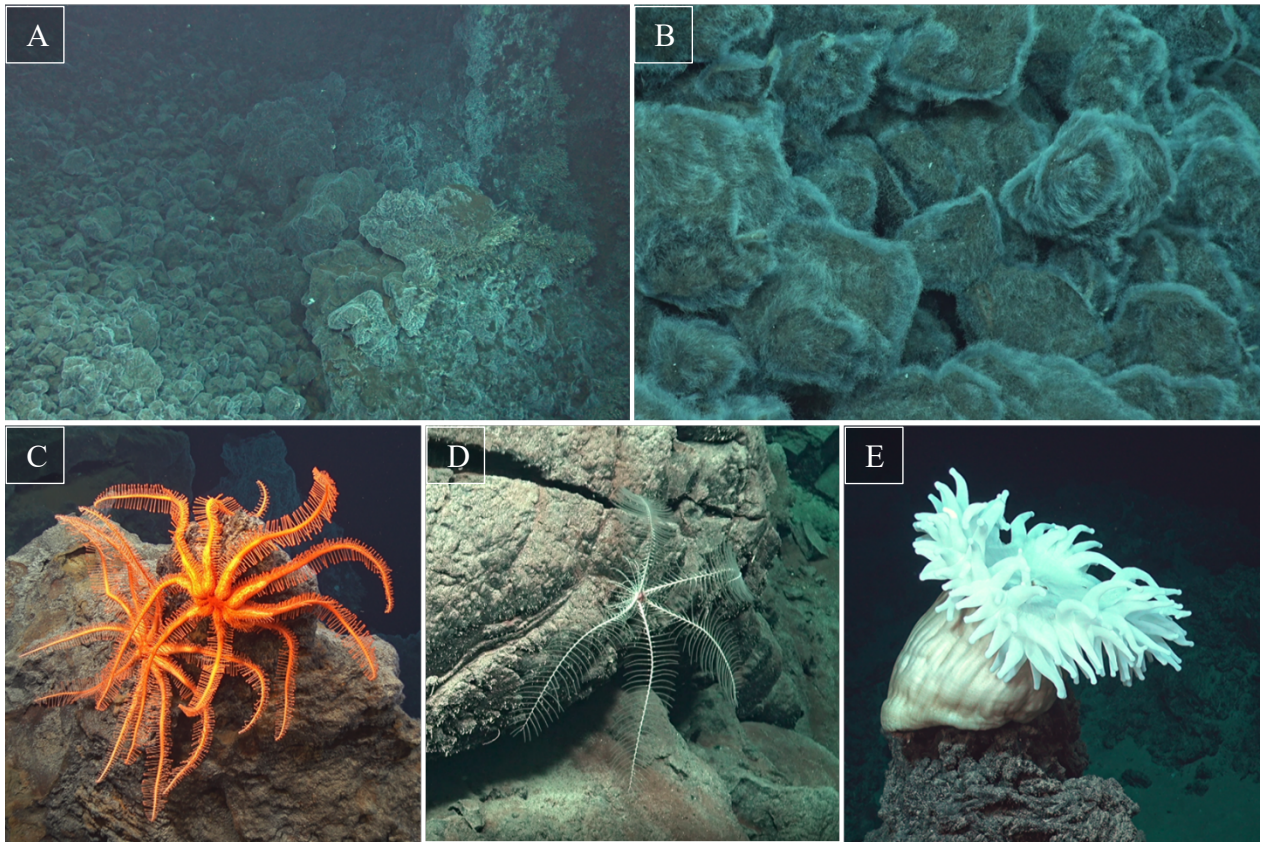
Supplementary Figure 2. CTD06 Vertical profile. Data show an isothermal water column below 3000 mbsl and Eh, temperature, and turbidity anomalies close to the seafloor.



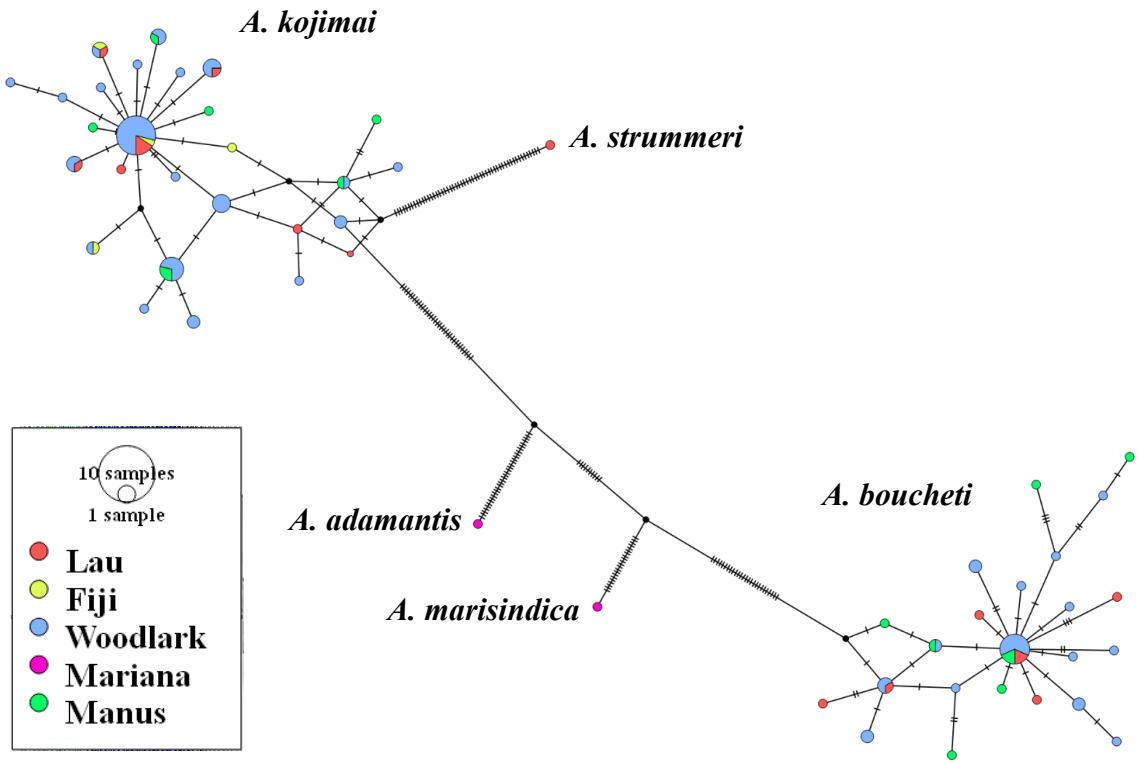
Supplementary Figure 3. Total alkali ($Na_2O + K_2O$) vs. silica (SiO_2) diagram. Data are reported for unaltered basalt (pillow lava) and volcanoclatic rocks from La Scala vent field and its vicinity (segment 3 of the Wooldark Basin, this study). For comparison, data for unaltered glass from basalt recovered from segment 4 of the Woodlark Basin[20]. Field boundaries for igneous rock classification are from Le Maitre et al. (1989)[57].



Supplementary Figure 4. Distribution of $\partial^{34}S$ in sulfides (distinguishing chalcopyrite, pyrite/marcasite, and sphalerite-rich mineral assemblages) at Woodlark Basin (La Scala vent field). The diagram compared LSVF data with literature data for other vent fields (VF) along the Mid-Atlantic Ridge, East Pacific Rise, Juan de Fuca, Lau Basin and Manus Basin. Only average $\partial^{34}S$ values are reported, with error bars corresponding to 1 standard deviation. Data sources: (1) Broken Spur VF[58]; (2) Snake Pit VF[59]; (3) TAG (ODP site); Knott et al. 1998; (4) Lucky Strike VF[60]; (5) Logatchev VF[61]; (6) Rainbow VF[61]; (7) EPR 11-13°N VF[62]; (8) EPR 21°N VF[63–65]; (9) Southern EPR VF[66]; (10) Axial Seamount VF[67]; (11) Southern JdF VF[68]; (12) Hine Hina VF[69]; (13) Vai Lili, White Church VF[70]; (14) ABE, Tui Malila, Mariner VF[66]; (15) Roman and Roger's Ruins VF[66,71]; (16) SuSu Knolls VF[66,71]; (17) La Scala VF, this study.



Supplementary Figure 5. La Scala peripheral communities. A. Field of bacterial mats with Munidopsidae squat lobsters (white dots). B. Bacterial mats. **C**. Brisingidae. D. Crinoidea. E. Actiniaria.

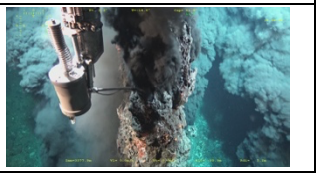
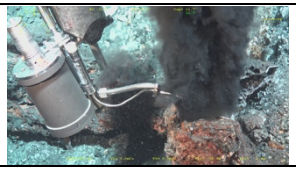
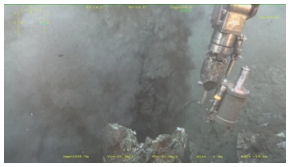
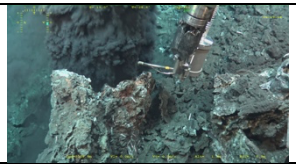


Supplementary Figure 6. Haplotype network. Haplotype network of the Woodlark *Alviniconcha Cox1* gene (accession numbers: OK391249-OK392008) with sequences of the five references species using the minimum spanning method implemented in the software PopArt (Leigh et al. 2015). Woodlark specimens equally fall within two (*A. boucheti* and *A. kojimai*) of the five previously described species of the genus.

Supplementary Table 1. Geochemical composition of volcanic rocks and sediment recovered at La Scala vent field and its vicinity

Name	Lat	Long	LOI	Total	SiO2	TiO2	Al2O3	Fe2O3T	MnO	MgO	CaO	Na2O	K2O	P2O5	Zr
			%	%	%	%	%	%	%	%	%	%	%	%	ppm
CHU-PL20-R2	S 09 47.939	E 155 03.127	0.87	99.6	48.44	1.25	17.36	8.72	0.38	8.58	11.06	2.73	0.12	0.13	90
CHU-PL19-R5	S 09 47.516	E 155 03.511	< 0.2	100.3	48.55	1.27	17.30	9.04	0.15	9.51	11.56	2.69	0.08	0.10	100
CHU-PL19-R6B	S 09 47.924	E 155 03.160	< 0.2	100.0	48.70	1.29	17.12	9.11	0.15	9.14	11.60	2.73	0.08	0.11	100
CHU-PL19-R6S	S 09 47.924	E 155 03.160	30.1	100.3	20.90	0.37	6.62	4.37	0.25	2.18	33.36	1.14	0.15	0.17	50

Supplementary Table 2. Coordinates, depths and preliminary chemical characteristics of the sampled vents.

Sample Name ID	DIVE	Titanium sampler	Lat. S	Long. E	Immersion (m)	Bearing (°)	Max T°C (ROV)	Comments	pH (21°C)	H_2S (mM)	Mg (mM)	Pictures
CHU-PL19-TiG1	PL19	TiG1	9°47.95	155°03.11	3378	9.5	366	Top of 7m tall black smoker (POI smoker #5)	3.58	3.64	3.75	
CHU-PL19-TiG2	PL19	TiG2	9°47.94	155°03.124	3376	9.5	364	Black smoker 6	5.71	0.34	40.95	
CHU-PL19-TiD2	PL19	TiD2	9°47.94	155°03.124	3376	9.5	364	Black smoker 6	4.31	1.17	27.11	
CHU-PL19-TiD3	PL19	TiD3	9°47.95	155°03.11	3380	11.7	367	Base of black smoker 10cm diam.	3.41	4.04	1.94	
CHU-PL20-TiD2	PL20	TiD2	9°47.9315	155°03.1295	3358	262	365	Base of diffuser. Very large conduit	3.64	3.75	4.00	
CHU-PL20-TiG2	PL20	TiG2	9°47.9315	155°03.1295	3358	260	365	Base of diffuser. Very large conduit	3.31	4.08	1.64	

Chapitre 2 :

Sex determination in three deep-sea hydrothermal vent gastropods with different sexual systems using genomic tools

Chapter realized in collaboration with Pradillon F, Cueff V, Léger G, Daguin-Thiébaut C, Ruault S, Mary J, Hourdez S, Jollivet D, and Broquet T.

1- Introduction

Sex determination systems are divided into two main categories (Beukeboom and Perrin, 2014; Bull, 1983). The first category corresponds to genetic sex determinisms (GSD), where we find XX/XY male heterogamy as in humans (Eggers and Sinclair, 2012) or ZZ/ZW female heterogamy as in birds or butterflies (Smith and Sinclair, 2004). GSD also includes systems "derived " from these classical heterogametic schemes, such as the polygenic determinism which involves several genes that can be located on different chromosomes (e.g. Zebrafish; Liew et al., 2012) or the XX/X0 system where males possess only one sex chromosome (e.g. in the nematode *Caenorhabditis elegans*; Ellis, 2008).

In the second category, the determining factor is not genetic but epigenetic (ESD), so the same genotype can produce a male or a female depending on environmental and/or social factors (Beukeboom and Perrin, 2014). Among the epigenetic determinants, we find the environmental determination (e.g. of the effect of temperature in turtles; Bull et al., 1982), social effect and parasitic manipulation (e.g. of feminization caused by *Wolbachia*, endosymbiont of many arthropods; Bouchon et al., 2008). GSD and ESD must be seen as the two ends of a continuum (Sarre et al., 2004). Indeed, in many cases, sex is determined by a combination of genetic and epigenetic factors, as for example in the lizard *Pogona vitticeps*, whose sex is determined by a ZZ/ZW system below 32°C, but ZZ individuals (normally males) tend to develop into females above this thermal limit (Ezaz et al., 2005).

Co-sexuality (monoecy and hermaphroditism) normally results from epigenetic sex determination but this is not always the case (Beukeboom and Perrin, 2014). Indeed, sometimes co-sexuality results from both genetic and epigenetic factors. For example, in cucumbers, the combination between the genotype and the amount of hormones (including auxin hormone) determines whether an individual is female, hermaphrodite or male/hermaphrodite (andromonoecious: male flowers at the bottom and hermaphroditic flowers at the top of the shoots) (Yamasaki et al., 2001). The amount of hormone secreted is driven by environmental factors which are temperature and day length (Yamasaki et al., 2001). In animals, co-sexuality can mixed GSD and ESD component. This is for example the case in the polychaete *Capitella capitata* where females are genetically heterogamous

(ZZ/ZW system), while homogamy develops females, males, and hermaphrodites depending on the social context (Boidin-Wichlacz et al., 2021; Petraitis, 1991). Sometimes, co-sexuality can also result from only one genotype, as is the case in papaya, where the males are coded XY and the females XX. However, hermaphroditic individuals can emerged from the XX genotype because of a mutation that allowed the loss of the gene responsible for the repression of the male organs (Ming et al., 2007).

Sex determinism is managed by a cascade of genes, the top of which is a trigger (Beukeboom and Perrin, 2014). It can be genetic (e.g. DM-domain in mammals; Matson and Zarkower, 2012), environmental (e.g. temperature) or social (e.g. density of individuals in fish). This cascade leads to the formation of either male or female gonads or both at the same time in hermaphrodites. In gonochoric animal species, the bipotential embryonic gonad, develops as male or female after the trigger signal (genetic or environmental). In simultaneous hermaphrodites, the process of determination relies on intrinsic triggers, such as information about the position of cells in the developing embryo that determines the development of the cell into a spermatozoid or oocyte (Beukeboom and Perrin, 2014). In sequential hermaphrodites, i.e. species where individuals develop first as males (protandry) or females (protogyny) and may later change sex (sometimes even with several back and forth switches between sexes), the trigger will be either intrinsic (e.g. size or age) or extrinsic (social environment; Beukeboom and Perrin, 2014).

In marine gastropods, sex determination includes genetic, epigenetic, and mixed systems (Beukeboom and Perrin, 2014). Unlike terrestrial gastropods, the majority of marine species are gonochoric (as in *Patellogastropoda*, *Vetigastropoda* and *Caenogastropoda*; Castillo and Brown, 2012; Curdia et al., 2005; Kunze et al., 2016). However, there are some hermaphroditic sequential species especially in the superfamily *Calyptraeoidea*, where for example the slipper limpet (*Crepidula fornicata*) is protandrous (individuals mature first as males and they may later switch to become females). Like in other sequential hermaphrodites, sex determination is under environmental and social control (Broquet et al., 2015; Proestou, 2005; Sales and Queiroz, 2021). There are also a few simultaneous hermaphrodites such as the limpet *Patella candei gomesii* or some *Opisthobranchia* (Cunha et al., 2007; Leonard and Lukowiak, 1991; Smolensky et al., 2009). Today there are about 60 genera and more than 100 species of gastropods associated with hydrothermal communities including *Patellogastropoda*, *Vetigastropoda*, *Neomphalina*, *Neritimorphoa*, *Caenogastropoda*, *Heterobranchia*, *Prosobranchia* and *Nudibranchia*. In this particular environment, and

despite the large number of species present, the system of sex determination has to our knowledge never been studied.

This study focuses on three species of *Alviniconcha* (*A. kojimai*, *A. boucheti* and *A. strummeri*) inhabiting hydrothermal vents of the Western Pacific back arc basins. The hydrothermal environment is spatially fragmented and temporally ephemeral, with locally highly dynamic physico-chemical conditions resulting from the chaotic mixing of the hydrothermal fluid (hot, acidic pH=2 and anoxic; Le Bris et al., 2003; Von Damm, 1990) and the surrounding water (cold and well oxygenated). Habitat conditions are heterogeneous at small spatial scales (Du Preez and Fisher, 2018) and constrain the distribution of individuals in space (more or less close to the emission) and in time (variation of the quantity of emission from their birth to their extinction), depending on their physiological requirements to feed and reproduce and fluid tolerance. This led to specific adaptations including the acquisition of chemoautotrophy via symbioses with diverse bacteria (sulfo-oxidative or methanotrophic) that live in the hydrothermal environment. Thus, the hydrothermal ecosystem is a unique setting where fluctuating environmental conditions and highly developed interactions with microorganisms could act as epigenetic factors for sex determination in animals. There are many examples of parasitic manipulation, including some associations in symbiotic organisms (Cordaux et al., 2011), but the best known example is the feminization of isopods by the parasitic bacterium *Wolbachia pipientis* (Bouchon et al., 2008).

The *Alviniconcha* gastropod species complex form dense aggregations in hot (7-42°C), sulphurous (250 µM), and low-oxygen (< 50 µM) vent habitats, indicating that these animals have tolerance to high temperatures and sulphides (Podowski et al., 2009, 2010). In the Western Pacific, *A. boucheti* is found closer to vent emissions when compared to the two other co-occuring species (*A. kojimai* and *A. strummeri*) and seems to have a different ecological niche with higher sulphur and hydrogen concentrations (Beinart et al., 2012; Breusing et al., 2020). These differences in ecological niches would in fact be linked to type of symbionts associated with these species. Previous studies have indeed shown that *A. kojimai* and *A. strummeri* are mostly associated with *gamma-proteobacteria* that tolerate low H_2 and H_2S concentrations, whereas *A. boucheti* possesses *Campylobacteria* which tolerate higher concentrations of H_2 and H_2S (Beinart et al., 2012, 2014; Breusing et al., 2020). As these species are found in different ecological niches with different symbionts, this may influence the system of sex determination and it will be therefore interesting to study how this

mechanism could differ between these large gastropods and its subsequent influence on speication.

During the sampling cruise dedicated to the study of hydrothermal ecosystems in the South West Pacific (CHUBACARC - R/V l'Atalante - Spring 2019) the sex of *Alvinconcha boucheti*, *A. strummeri* and *A. kojimai* individuals was tentatively identified based on observation of the external color of the gonad. This yielded a coarse, sometimes inconclusive, sex identification, which was subsequently complemented with refined dissections and observations of histological sections for a subset of individuals (team of Florence Pradillon, Ifremer Brest). This anatomy study by our colleagues from Brest, which is still in progress, produced a very intriguing result: from the gonad histology, it was clear that *A. boucheti* and *A. strummeri* are gonochoric (separate male and female individuals). By contrast, about half of *A. kojimai* individuals were males while most others presented both male and female gonadal tissues. The proportion of male and female tissues was variable, with a few individuals showing only female tissues although we cannot exclude that very small amounts of male tissues were missed (different parts of the gonad were explored with histology, but not the whole of it because of its large size). Due to this uncertainty, at this stage, *A. kojimai* individuals are either classified as "males" or "hermaphrodite/females". Based on these preliminary observations, it is possible that *A. kojimai* presents a form of androdioecy, where populations consist of males and hermaphrodites' individuals.

Here, a first attempt to infer the sex determination system of each of the *Alviniconcha* target species was done by exploring the genetic structure of these species in the light of the phenotypic sex of the individuals previously observed by F.P.' team in Brest.

2- Materials and Methods

2.1- Sampling

We used a subset of a previously published genomic data set of ddRAD-seq sequences from Castel et al. (submitted: Archive Accession no. PRJNA768636). Sampling conditions are detailed in (Castel PhD thesis, Castel et al., in prep). In short, here we use data from 275 *A. kojimai*, 215 *A. boucheti* and 46 *A. strummeri* individuals sampled from 16 sites in 5 back arc basins of the Western Pacific Ocean during the CHUBACARC expedition (Figure 34; Table S4). The Lau, Futuna and Fiji Basins are distant from the Manus and Woodlark Basins by nearly four thousand kilometres, and with this geographical separation come differences

in gastropod species distribution and a decrease of genetic and demographic connectivity (Castel et al., submitted; Breusing et al., 2020; Tran Lu Y et al., 2022). Therefore, to facilitate discussion, we will refer to these groups as “Region 1” and “Region 2” (Fig. 34).

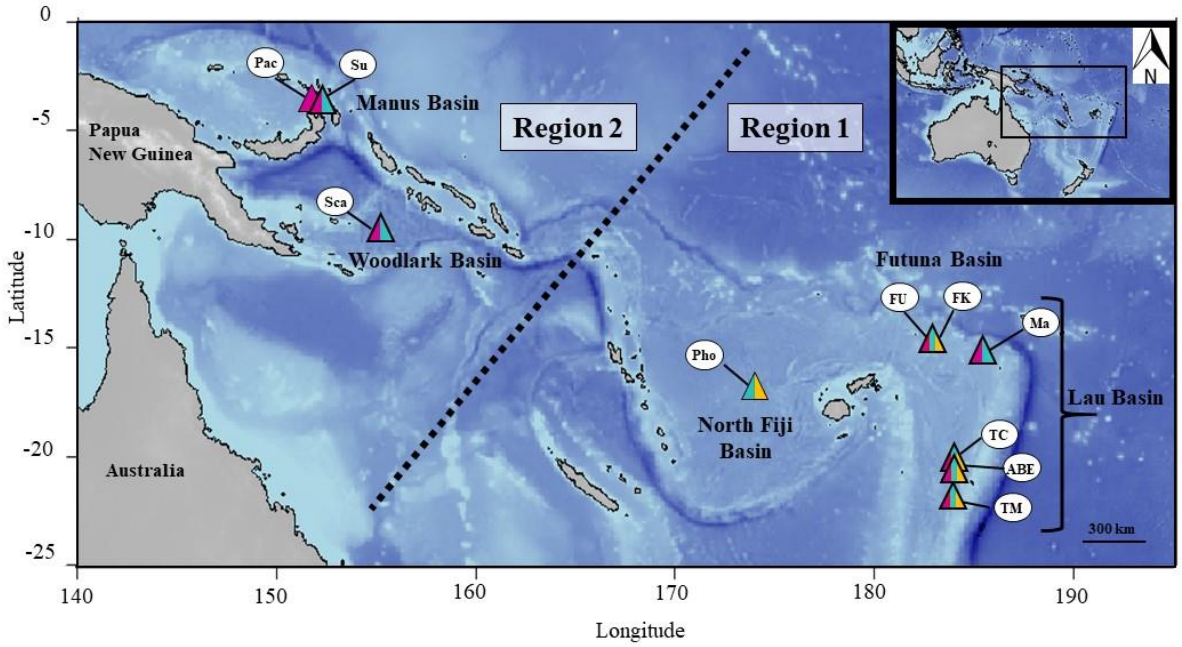


Figure 34: Sampled localities of *Alviniconcha* species during the CHUBACARC expedition (2019). Manus Basin: Susu (Su); Pacmanus (Pac). Woodlark Basin: Scala (Sca). Fiji Basin: Phoenix (Pho). Futuna Basin: Fatu-Kapa (FK); Fati-Ufu (FU). Lau Basin: Mangatolo (Ma); Tow Cam (TC); ABE (ABE); Tui Malila (TM). Colours indicate species occurrence, purple: *A. boucheti*; turquoise: *A. kojimai* and yellow: *A. strummeri*. Lau, Futuna and Fiji basins are grouped in “Region 1”, while the Manus and Woodlark basins are grouped in “Region 2”.

We had information about the gender of individuals (from the observation of gonads or histological sections, F. Pradillon pers. comm.) for half (n=247) of the 495 individuals genotyped. In addition, 23 other individuals had somewhat uncertain sex identification (gonad colour not completely informative, Table S4). The sex of all remaining individuals was unknown.

2.2- Genotyping

After extracting DNA from foot tissues with the NucleoSpin® Tissue kit (Macherey-Nagel) or the CTAB protocol (Doyle and Dickson, 1987 modified in Jolly et al., 2003), individual-based ddRAD libraries were produced (Daguin-Thiebaut, 2021) and sequenced in 150-bp paired-end reads using three lanes of Novaseq 6000 (Details in Castel et al. submitted).

For each species, raw reads were first demultiplexed using the process_radtags module of Stacks software version 2.52 (Catchen et al., 2013; Catchen et al., 2011) to remove adapters and reads with low quality scores (Phred score < 30). We used Kraken v.2 (Wood and Salzberg, 2014) to keep only the eukaryotic reads for the assembly (according to the species, between 0.9% and 1.5% of the initial reads were assigned to prokaryotes and discarded). Average sequence quality per read and GC-content were checked using multiQC version 1.7. Barcode sequences were trimmed and the remaining 144 bp reads were assembled into 'stacks' (equivalent of alleles) if six or more identical reads were found within an individual (*m*: 6).

For *A. kojimai* and *A. boucheti*, assembly parameters were chosen after empirical testing over a range of values on triplicates (within each species, six individuals were triplicated in the RAD-seq library preparation, starting from a unique DNA extract). Stacks core parameters were changed one at a time (by increasing steps of one unit: ustack *m* [2–6], *M* [2–14], and cstack *n* [2–18], while the others were used at their default values, as suggested by Mastretta-Yanes et al. (2015).

The final set of parameters was chosen to maximise the number of single nucleotide polymorphism sites (SNPs) retained while maintaining a minimum genotyping error rate (measured as the frequency of differences found between the three genotypes observed at each SNP of the triplicated individuals). As *A. strummeri* had no replicates in the dataset, parameters were chosen with respect to the other species while retaining the highest number of SNPs (Figures S7, S8 and S9).

As a result of this pilot study, invariant stacks (alleles) in *A. kojimai* were compiled within each individual into sample-specific loci if they differed by less than 10 nucleotides (*M*: 10) and the sample-specific loci were then assembled between individual samples into homologous loci if they differed by less than the same number of nucleotide changes (*n*: 10). The assembly parameters *m*: 6; *M*: 8; *n*: 8 and *m*: 6; *M*: 4; *n*: 4 were used for *A. strummeri* and *A. boucheti* respectively. These high values for *M* and *n* parameters reflect the fact that divergent haplotypes may be present within some individuals due to allele introgression between divergent species, particularly in *A. strummeri* (Castel et al. submitted).

The settings of Stacks included the following conditions for the three species: minimum proportion of individuals sharing a locus $r \geq 0.8$ and minor allele count MAC ≥ 4. The SNPs identified by Stacks were then further filtered in R version (R Core Team, 2020) in order to

only keep SNPs shared by at least 90% of the individuals and, individuals which were genotyped with, at least, 80% of the total number of SNPs except in *A. strummeri*, for which we had less individuals and thus allowed a little more missing data per individual (25%).

In the end, the final data set contained 207 individuals and 163 117 SNPs for *A. boucheti*, 247 individuals and 70 122 SNPs for *A. kojimai* and 41 individuals for 36 568 SNPs for *A. strummeri*. The population structure associated with the SNPs was visualised in R (adegenet package; Jombart, 2008) using a principal component analysis (PCA).

The distribution of genetic diversity was examined using functions from the R package hierfstat (Goudet, 2020). We characterised the global genetic diversity within each species by estimating total gene diversity H_t. The diversity within each sex i was then described by calculating the gene diversity $H_{S,i}$ and observed heterozygosity $H_{O,i}$, and the discrepancy between these parameters (indicating a potential departure from a random association of gametes) was estimated as F_{IS} within each sex. Finally, the genetic differentiation between sexes was estimated by F_{ST} between males and females.

2.3- Identification of genetic sex determination system

To detect genetic systems of sex determination, we focused our attention on SNP markers with alleles differentially fixed on sex chromosomes. If sex was genetically determined and if some SNPs had alleles that were differentially fixed on sex chromosomes, then these SNPs would be strictly homozygous in the homogametic sex (XX females or ZZ males) and heterozygous in the heterogametic sex (XY males or ZW females). We looked for such SNPs by estimating the observed heterozygosity in males ($H_{O,m}$) and females ($H_{O,f}$) at each SNP using functions from the hierfstat R package (Goudet, 2020).

2.4- Identification of sex-linked loci

Besides sex-linked SNPs with fixed, divergent alleles on sex chromosomes, other polymorphic SNPs located on sex chromosomes may have various levels of heterozygosity and sex-linked differentiation depending on the genomic landscape of recombination along these chromosomes. The distribution of genetic differentiation between sexes (F_{ST}) and between individuals within each sex (F_{IS}) is more informative than H_0 to detect sex-linked loci that are not differentially fixed, because sex linkage will tend to result in an excess of heterozygotes in the heterogamous sex (e.g. XY males) and a simultaneous increased differentiation between sexes.

However, we expect some uncertainty associated with these statistics measured using a finite number of male and female individuals (Table 5). Therefore, for each species we computed F_{IS} and F_{ST} values from simulated genotypes randomly created from a range of allelic frequencies assuming autosomal segregation for two groups of individuals with sample sizes equal to the male and female groups in our real dataset (thereby introducing a sampling error variance in the simulations). The simulations were coded in R and consisted of the following steps:

1- Create a number of theoretical bi-allelic SNPs equal to the number of SNPs genotyped in each species and were each SNP as a reference allele with frequency *p* and an alternative allele with frequency 1-*p*, where *p* is sampled randomly from a uniform distribution between 0 and 1.

2- Create a group of "male" and a group of "female" genotypes (with sample sizes equal to the number of males and females in each species) by sampling two alleles randomly from a binomial distribution parameterized using the expected allelic frequencies simulated at step 1.

3- From these simulated genotypes, estimate F_{ST} between the "male" and "female" groups, and F_{IS} within each of these groups at each simulated SNP. These estimates were calculated following Nei (1987) as described in R package hierfstat (Goudet and Jombart, 2015) and simplified for bi-allelic SNPs:

$$F_{ST} = (H_T - H_S)/H_T$$

and

$$F_{IS,i} = \left(H_{S,i} - H_{O,i}\right)/H_{S,i} \text{ for sex } i$$

where:

$$H_T = 2p(1-p) + H_S/\left(n_f + n_m\right) - H_O/2\left(n_f + n_m\right)$$

$$H_S = \left(n_f H_{S,f} + n_m H_{S,m}\right)/\left(n_f + n_m\right)$$

$$H_{S,i} = n_i/(n_i - 1)(2p_i(1-p_i) - H_{0,i}/2\mathrm{n}_i)$$

with *p* is the allelic frequency of the reference allele (recalculated from the simulated genotypes) at the SNP under consideration, ni is the sample size of sex *i*.

In these equations, 2p(1-p) is the gene diversity expected under random mating (Hardy-Weinberg, autosomal segregation at a bi-allelic loci with allelic frequencies *p* and 1-*p*) and all other terms are bias corrections for small sample sizes as described in Nei (1987).

These statistics were computed at each SNP. Thus, after simulation, we can compare the F_{ST} and sex-specific F_{IS} values of the SNPs in our dataset to the simulated values. All SNPs from the real dataset that had combinations of F_{IS} and F_{ST} values outside of a convex envelope encompassing all possible theoretical values obtained from autosomal simulations had properties that were not compatible with autosomal segregation.

2.5- Synteny of sex-linked loci across the three species

As we used different catalogues for each species in Stacks, sex-linked RAD loci shared between species were identified through reciprocal blasts in Geneious (Kearse et al., 2012). Subsequently, a GO (Gene Ontology) annotation of sex-linked loci was performed by looking at each assignment in the Uniprot/Swissprot databases in NCBI.

2.6- Test of selfing in *A. kojimai*

Self-fertilisation is a form of deviation from random association of gametes that results in a rapid increase in the value of F_{IS}. In a panmictic population at Hardy-Weinberg equilibrium, the expected F_{IS} is close to 0. Selfing leads to a deficit of heterozygotes that results in a positive F_{IS}. On the contrary, a negative F_{IS} indicates an excess of heterozygotes. We therefore calculated F_{IS} within *A. kojimai* to assess whether this apparently hermaphroditic species is capable of selfing. As F_{IS} is also strongly impacted by a population substructure with admixture (Wahlund effect) on the one hand, and by the sample size on the other hand, we chose to calculate F_{IS} at each locus at the scale of the smallest spatial sampling unit (the biobox, which corresponds to a sample of individuals very close to each other, over a few tens of square centimetres) and to compare this distribution of F_{IS} to that obtained in the same way in the gonochoric species *A. boucheti* and *A. strummeri*.

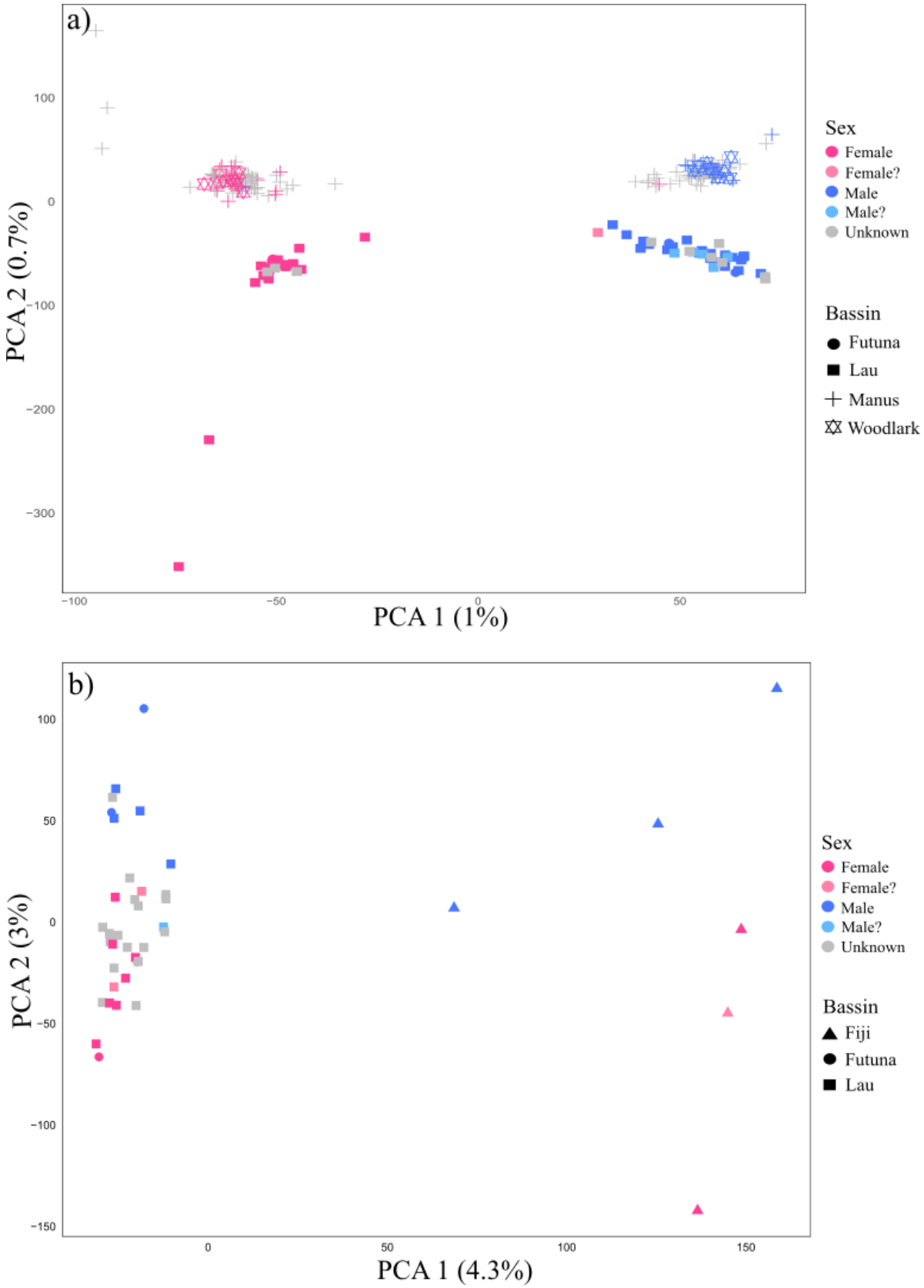
3- Results

3.1- Genetic structure

Results from the PCA are shown in Fig. 35. Remarkably, in species *A. boucheti* (Fig. 35a) the first axis (1% of variance explained) separated individuals according to their sex, and individuals of unknown sex were distributed within the male or female group. The second axis (0.7%) was consistent with geography, separating individuals from the Lau, Futuna, and Fiji basins (Region 1) from individuals sampled in the Manus and Woodlark basins (Region 2). Phenotypic sex did not agree with the genetic clustering observed in the PCA for five individuals only (not all visible in Fig. 35a due to superimposition of points), including one individual for which sex identification based on the color of the gonad had been classified as uncertain (see methods). Since some uncertainties can exist in the phenotypic sex identification (even for the individuals whose sex was not qualified as “uncertain” at first sight), these five individuals were removed from further analysis.

The first axis (4.3 %) of the PCA of *A. strummeri* multilocus genotypes (Fig. 35b) was consistent with geographic origin of the individuals (again clustering individuals from North Fiji vs Lau/Futuna). Interestingly, sex was again mirrored by the clustering of genotypes along axis 2 (3% of variance explained).

Finally, individuals were distributed according to their geographic origin in species *A. kojimai* (Fig. 35c) (first axis of the PCA explaining 1.2% of the variance), while no differentiation by sex was visible.


a)
PCA 2 (0.7%)
PCA 1 (1%)
100
0
−100
−200
−300
−100
−50
0
50
Sex
Female
Female?
Male
Male?
Unknown
Bassin
Futuna
Lau
Manus
Woodlark
b)
PCA 2 (3%)
PCA 1 (4.3%)
100
50
0
−50
−100
−150
0
50
100
150
Sex
Female
Female?
Male
Male?
Unknown
Bassin
Fiji
Futuna
Lau

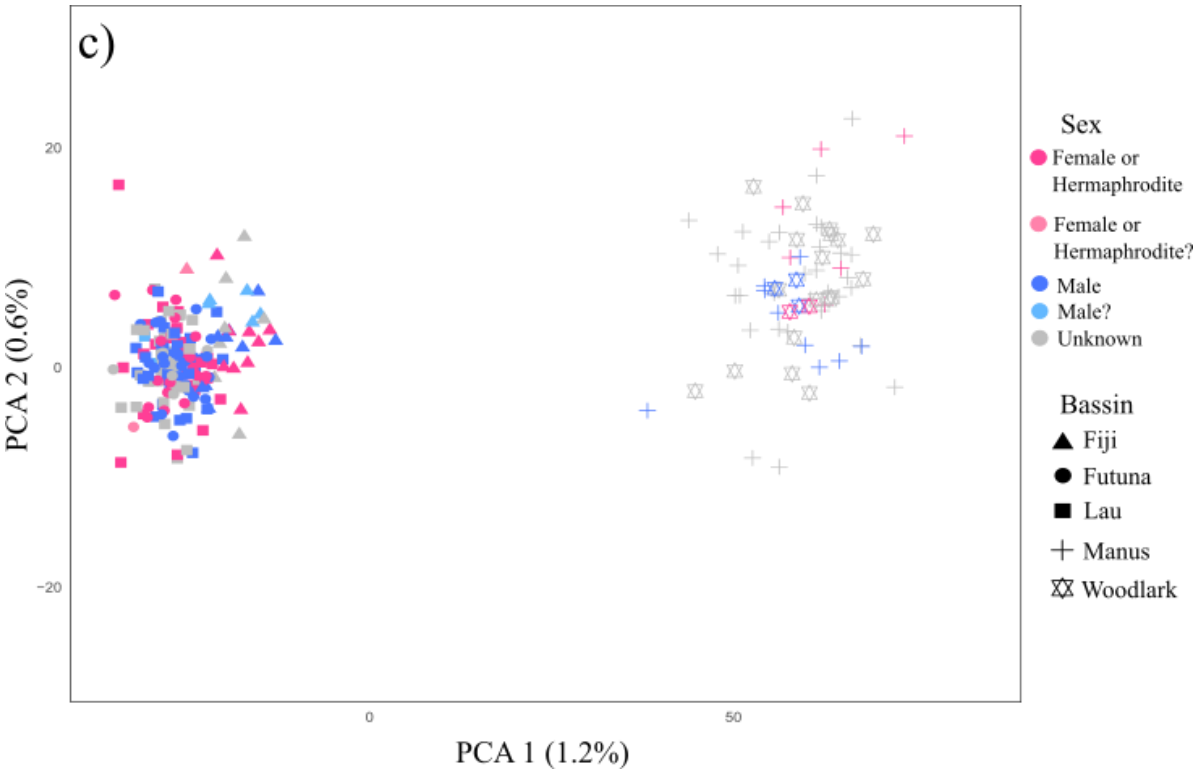


Figure 35: Principal components analyses of multilocus genotypes for the three species of *Alviniconcha* gastropods. The notations “Female?” and “Male?” correspond to individuals for which sex identification based on the gonad colour was uncertain. Square, round and triangle symbols correspond to the Southwestern basins ("Region 1" in Figure 34), while cross and star symbols correspond to the Northwestern basins ("Region 2" in Figure 34). Panel a) *A. boucheti* (207 individuals, 163 117 SNPs). The first axis separates individuals according to their sex, while the second axis is consistent with geography; b) *A. strummeri* (41 individuals, 36 568 SNPs). While the first axis separates individuals according to their geographic origin, the second axis is driven by sex ; c) *A. kojimai* (247 individuals, 70 122 SNPs). The first axis separates individuals according to their geographic origin. The second axis is driven by a single individual (not shown here, see Fig. S10 in supplementary material) without effect of sex.

3.2- Genetic diversity

The diversity indices in Table 5 were calculated only on sexed individuals, thus 89 individuals for *A. boucheti*, 19 individuals for *A. strummeri* and 139 individuals for *A. kojimai* (Table S4). The highest total genetic diversity (H_T) was observed in *A. strummeri* (0.267) and was almost twice as high as in the other two species. In all species, sex-specific genetic diversity ($H_{S,i}$) was very close between sexes but slightly higher in males than females in *A. boucheti* and *A. strummeri*. Differentiation between sexes (assessed by F_{ST}) was very low in *A. kojimai* (0.01%) while it was 0.5% and 0.2% in *A. boucheti* and *A. strummeri*,

respectively. Observed sex-specific heterozygosity (H_O) was higher and F_{IS} lower in males than in females in all species.

Table 5: Diversity indices for the three target species of *Alviniconcha*. H_T: total gene diversity, H_S: sex-specific gene diversity, where m is for male and f/h stands for female in *A. boucheti* and *A. strummeri* and hermaphrodite/female in *A. kojimai* (see main text); F_{ST}: differentiation between sexes; $H_{O,i}$: observed heterozygosity in sex i and $F_{IS,i}$: heterozygosity deficit in sex i. Column "Unsexed" gives the number of individuals with uncertain or unknown phenotypic sex.

	Ind.	Males	Females	Unsexed	SNPs	Loci	H_T	$H_{s,m}$	$H_{s,f/h}$	F_{ST}	$H_{o,m}$	$H_{o,f/h}$	$F_{IS,m}$	$F_{IS,f/h}$
A. boucheti	207	43	46	118	163 117	48 384	0.140	0.140	0.138	0.0053	0.141	0.136	0	0.018
A. strummeri	41	9	10	22	6 724	2864	0.267	0.268	0.265	0.0016	0.254	0.247	0	0
A. kojimai	247	74	65	108	70 122	16788	0.111	0.111	0.111	0.0001	0.110	0.109	0.015	0.016

3.3- Sex determination system

In *A. boucheti*, 595 SNPs on 251 RAD-loci were strictly homozygous in females and heterozygous in males (Figure 36a, top-left corner), pointing towards a XY genetic system of sex determination where each of these SNPs has one allele fixed on the X and another allele fixed on the Y. By contrast, no SNP was found to be indicative of a ZW system. Furthermore, a large number of additional SNPs showed a combination of strong $H_{O,m}$ and low $H_{O,f}$, in agreement with the hypothesis of a XY system with a range of shared genetic diversity on X and Y chromosomes depending on their distance to the no-recombination regions of these chromosomes (see next section below).

For *A. strummeri*, although we had only a few sexed individuals (nine males and ten females) and thus a high error associated with $H_{O,m}$ and $H_{O,f}$ at each locus, we found that 93 SNPs (on 55 RAD-loci) were strictly homozygous in females and heterozygous in males (Figure 36b, top-left corner), again indicating an XY genetic system of sex determination.

Surprisingly, in the species *A. kojimai*, where about half of the individuals have both male and female gonadal tissues, there were still nine SNPs (on seven RAD-loci) with a strong and low H_O in males and hermaphrodites/females, respectively (top-left corner of Figure 36c). So, it suggested the maintenance of a relict XY-like system where males are XY and hermaphrodites/females are XX (see Discussion). But this sex determination was only supported by a very small number of SNPs (9 out of 70 122, that is, less than 0.13 %) with no SNPs strictly homozygous in females and heterozygous in males.

In the three species, a number of SNPs with simultaneously high $H_{O,m}$ and $H_{O,f}$ values were also observed at a low frequency (top-right corner of Figures 36a-c). These SNPs could correspond to paralogous loci that were wrongly assembled, or, perhaps less likely, to nucleotide sites under strong balancing selection.

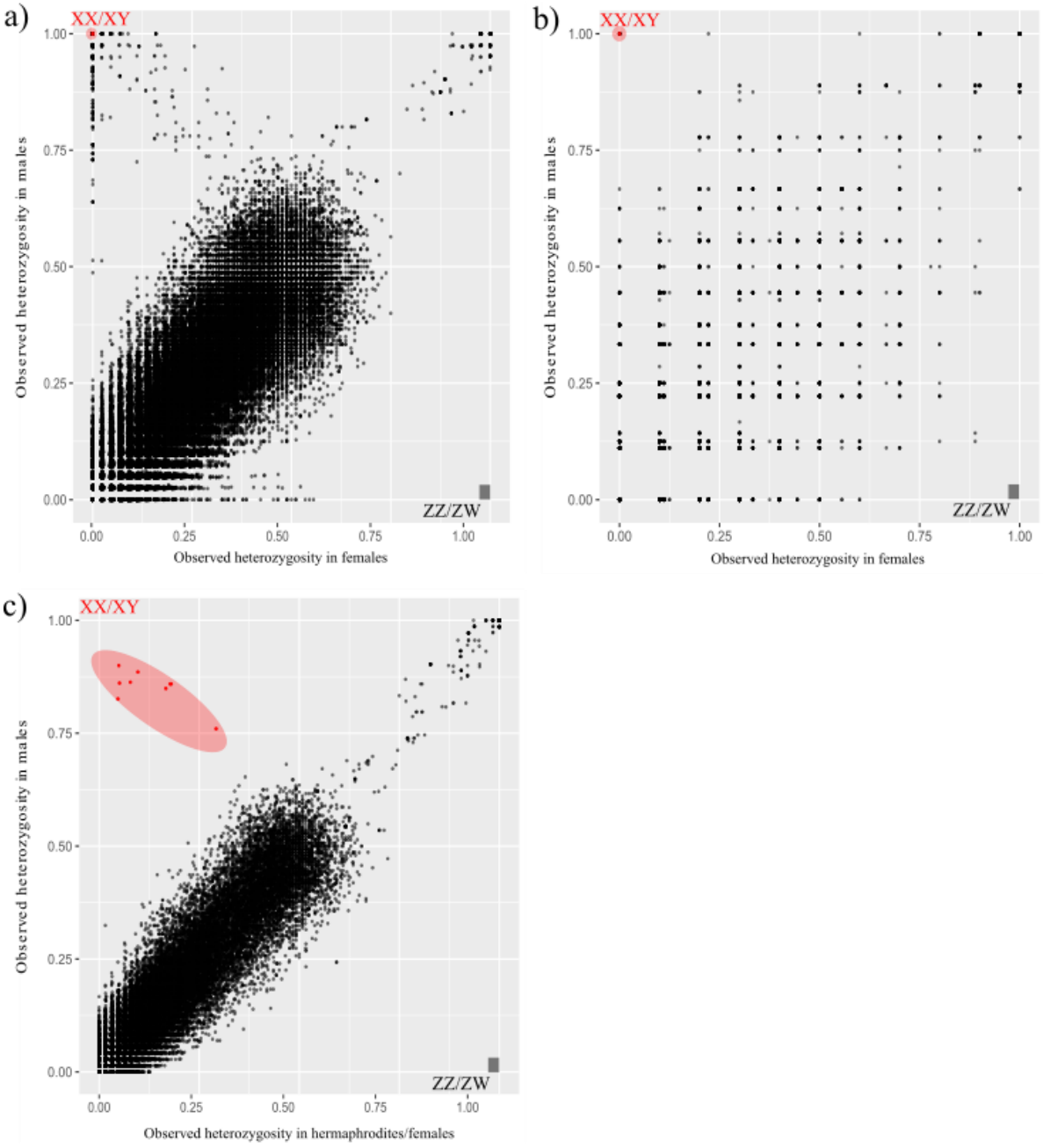


Figure 36: Observed female and male specific heterozygosities $H_{O,f}$ and $H_{O,m}$ in loci of the three *Alviniconcha* species. Panel : (a) *A. boucheti*: 163 117 SNPs (46 females and 43 males). In the top-left corner, 595 SNPs have $H_{O,f}$ = 0 and $H_{O,m}$= 1; (b) *A. strummeri*: 36 568 SNPs (10 females and 9 males). In the top-left corner, 93 SNPs have $H_{O,f}$ = 0 and $H_{O,m}$= 1. Note that sample size was low in this species (ten females and nine males), resulting in noisy estimates of $H_{O,f}$ and $H_{O,m}$; (c) observed

hermaphrodite/female and male specific heterozygosities $H_{O,h}$ and $H_{O,m}$ at 70 122 SNPs in *A. kojimai* (65 hermaphrodites/females and 74 males). In the top-left corner, nine SNPs showed strong heterozygosity in males but low heterozygosity in hermaphrodites/females.

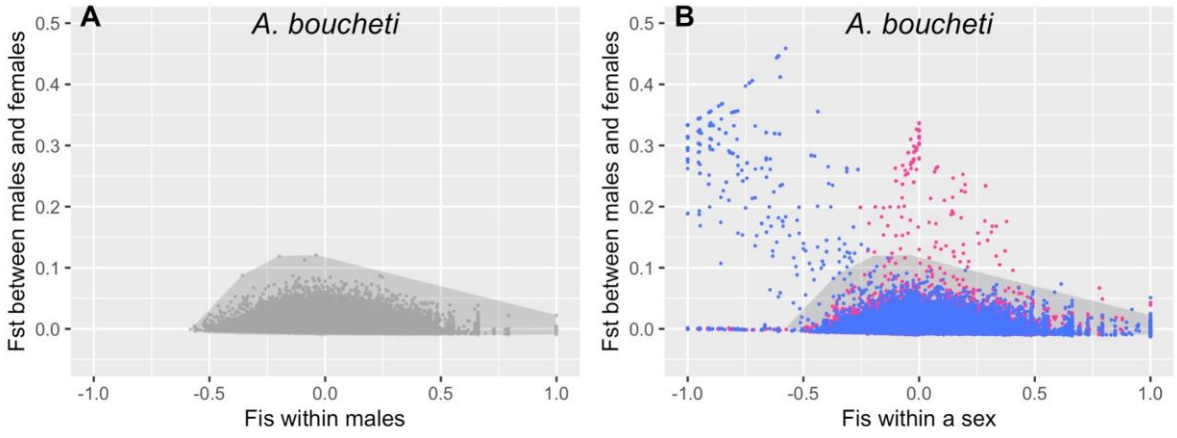
3.4- Identification of sex-linked SNPs

In *A. boucheti*, 1 011 SNPs (on 373 RAD-loci) had a combination of low $F_{IS,m}$ and strong F_{ST} compatible with XY features that were never obtained by simulation of autosomal segregation (Figures 37A-B). This represented 0.19% of the SNPs genotyped in this study (0.14% of the RAD-loci).

Similarly, in *A. strummeri*, 111 SNPs (on 64 RAD-loci) out of 31 790 SNPs had $F_{IS,m}$ and F_{ST} that were never obtained by simulation of autosomal SNPs (Figure 37C-D). These SNPs represented less than 0.03% of the sites genotyped in this study, but the power for detecting sex-linked SNPs is very reduced for *A. strummeri* because only 19 individuals were sexed, and this number is often even slightly lower for each specific locus (due to locus-specific missing data).

Finally, in *A. kojimai*, 20 SNPs (on 17 RAD-loci) had $F_{IS,m}$ and F_{ST} values that do not fit simulations of autosomal SNPs (Figure 37E-F). This represented 0.004% of the SNPs genotyped in this study (0.006% of the RAD-loci).

In line with the expectations for XY sex chromosomes for the three species, all SNPs showing a strong differentiation between sexes (F_{ST}) tended to have negative F_{IS} values in males (as stated above) and null F_{IS} in females (shown in pink in Fig. 37).



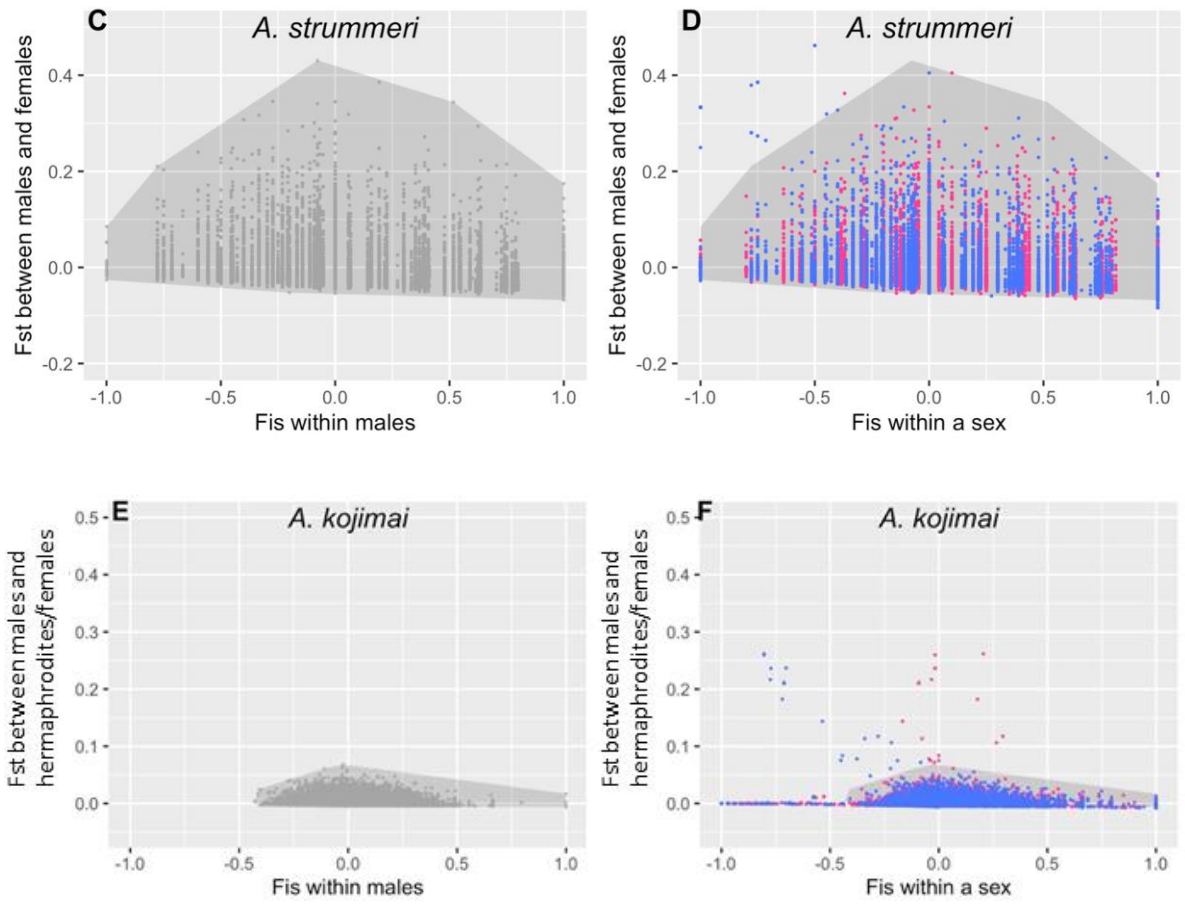


Figure 37: Simulated and observed F_{ST} between males and females and sex-specific F_{IS} for three *Alviniconcha* species. Panels A, C, and E show simulated values obtained for autosomal SNPs with a range of expected allelic frequencies (uniform between 0 and 1) and using real male and female sample sizes from each species (Table 5). The grey polygone is a convex envelope that covers all possible values obtained by simulation for an autosomal locus. Panels B and D show observed values for male F_{IS} (in blue) or female F_{IS} (pink) and male/female F_{ST} (that is, each SNP is represented once in blue and once in pink). Panel F is similar to B and D but, in this specific case, pink symbols refer to hermaphrodites/females, and F_{ST} was calculated between males and hermaphrodites/females. The number of SNPs used was 162 343 for *A. boucheti*, 36 531 for *A. strummeri*, and 70 072 for *A. kojimai* (these are the SNPs for which F_{IS} and F_{ST} values could be calculated). Any SNP outside of the grey polygon represent outliers with properties that were never obtained by neutral autosomal simulations.

3.5- Synteny of sex-linked loci across the three species

Shared sex-linked RAD-loci between species could be identified following a reciprocal blast (Figure 38). Twenty-one loci were identified between *A. boucheti* and *A. strummeri*, 8 between *A. boucheti* and *A. kojimai*, and 3 between *A. kojimai* and *A. strummeri*. Only three RAD-loci were shared by the three species. Note that a large number of loci were found only

in *A. boucheti* (82.2%), which is the species where we had the highest statistical power due to the high number of individuals with a known sex.

Gene annotation was subsequently obtained for 9.5% of the sex-linked loci, all species combined (Table S5). These annotations concerned mainly cell responses to stimulus, signalling pathways, epidermis development, cytoskeleton organisation but also flagellated sperm motility. Of the three loci common to all species, only one could be annotated. This locus is involved in cell adhesion and epidermis development in the scallop *Pecten maximus*.

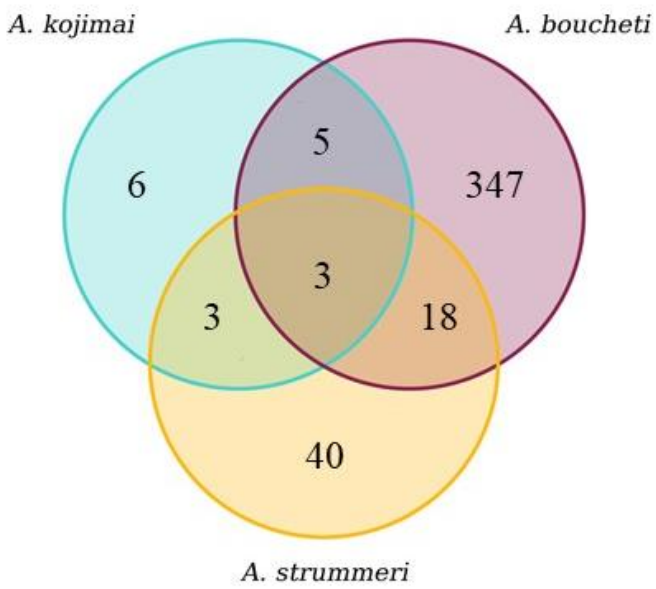


Figure 38: Venn diagram of sex-linked RAD-loci shared between and specific of *A. boucheti* (89 individuals), *A. strummeri* (19 individuals) and *A. kojimai* (139 individuals).

3.6- Test of selfing in *Alviniconcha kojimai*

The distribution of F_{IS} (calculated at each locus using all individuals sampled in the same a biobox, Figure 39) did not show any deviation towards stronger positive values in *A. kojimai* (average F_{IS} = 0.010) vs. *A. boucheti*. (F_{IS} = 0.014) or *A. strummeri* (F_{IS} = 0.024).

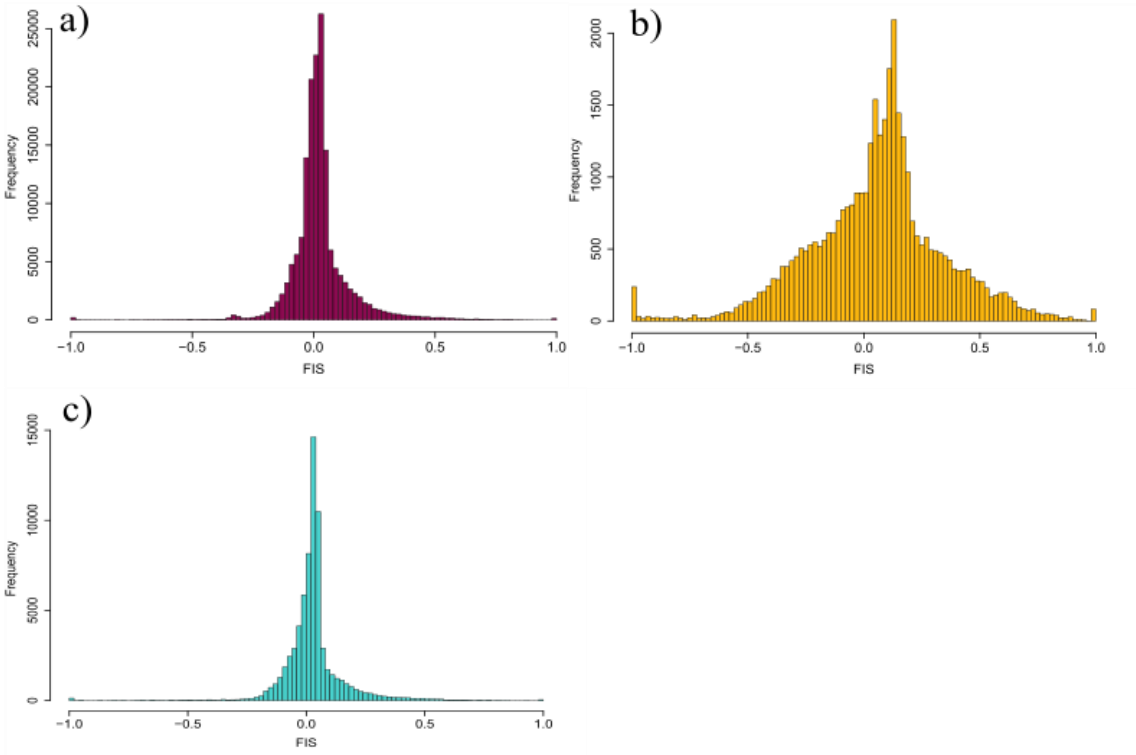


Figure 39: Distribution of F_{IS} for (a) *A. boucheti* in purple (n = 207; 162 343 SNPs), (b) *A. strummeri* in yellow (n= 41; 36 531 SNPs) and (c) *A. kojimai* in turquoise (n= 247, 70 072 SNPs).

4- Discussion

4.1- Sex determination in *Alviniconcha boucheti* and *A. strummeri*

This study based on hydrothermal gastropods from the Western Pacific Ocean showed that in the gonochoric species *Alviniconcha boucheti* and *Alviniconcha strummeri* the sex determinism was controlled by the genotype. Indeed, in these two species, there were a number of strictly differentially fixed SNPs between males and females (93 in *A. strummeri* and 595 in *A. boucheti*; Figure 36). These SNPs were strictly homozygous in females and heterozygous in males, pointing towards a XY genetic system of sex determination where each of these SNPs has one allele fixed on the X and another allele fixed on the Y.

In addition, a number of additional SNPs (416 in *A. boucheti* and 18 in *A. strummeri*) showed a combination of high $H_{O,m}$ / low $H_{O,f}$ in agreement with the hypothesis of a XY system with a range of shared genetic diversity depending on thei position on X and Y chromosomes (Figure 37). Overall, 111 and 1011 SNPs related to sex determinism could be identified in *A. strummeri* and *A. boucheti*, respectively (representing 0.3% and 0.6% of all genotyped SNPs in these two species).

Interestingly, out of the 64 and 373 RAD loci identified as sex-linked respectively in *A. strummeri* and *A. boucheti*, 18 were common to these two species (Figure 38). This proves that the two species share the same (homologous) XY chromosomes. The vast majority (95%) of the sex-linked RAD loci analysed here were found either only in *A. strummeri* or only in *A. boucheti*, but this is large number of unrelated SNPs is not unusual given (i) the very low power of our analysis for *A. strummeri*, for which only 19 sexed individuals were genotyped, and (ii) limitations due to the RADseq technique, which yields a reduced representation of the genome of each individuals with large amounts of data that are not shared among individuals and because a large fraction of enzyme-restriction sites may be different between divergent species according to the mutation-drift equilibrium (see e.g. Fig. 25 in chapter 1).

In gastropods and more particularly in *Caenogastropoda* (a group that includes the genus *Alviniconcha*), the majority of species have an XY system, such as, for example, *Pomacea canaliculata* (Yusa, 2007) or *Littorina saxatilis* (Kozminsky and Serbina, 2020; Rolán-Alvarez et al., 1997). However, this genetic determinism is not always visible on the karyotype (homomorphic sex chromosomes). Thiriot-Quiévreux (2003) showed that out of 230 species of *Caenogastropoda* studied, only 7% of the species had a chromosomal determination system visible using the karyotypic information. In the other species, the sex determination could not be determined.

4.2- How is sex determined in *A. kojimai*?

In *A. kojimai*, no SNP was found to be strictly homozygous in females/hermaphrodites and heterozygous in males (Figure 36) despite the very high number of SNPs screened. However, very interestingly, 20 SNPs showed a combination of high $H_{O,m}$ and low $H_{O,f}$ in agreement with XY sex determination. Morevoer, the eleven RAD loci that bear these 20 SNPs were also found to be sex-linked in at least one of the two other species (Fig. 37). There is thus definitely a genetic part in the determinism of sex in *A. kojimai*. Here, individuals identified as "pure" males would have XY sex chromosomes, while the other individuals presenting a mixture of male and female gonadal tissue would have XX sex chromosomes.

With this observation, *A. kojimai* would be an androdioecious species where the Y chromosome bears an allele that triggers male development, while XX individuals develop both male and female functions. This situation is very interesting because androdioecy is rare, and sex determination of androdiecious species is only know for a little number of species. Androdioecy (populations consisting of males and hermaphrodites) is a scarce mating system

in plants and animals: up to 50 plants and only 36 animals have been described as being androdioecious (Weeks et al., 2006). But androdioecy (and gynodioecy: populations consisting of females and hermaphrodites) are generally viewed as the primary intermediate, transitionary stages between dioecy and hermaphroditism or vice versa and appeared when environmental conditions are changing or when species occupies a new habitat (Charlesworth and Charlesworth, 1978). However, in spite of the energetic cost generated by the simultaneous production of the two types of gametes in one gonad simultaneously, taxa in which this system is maintained exist as for example in the genera *Balanus* where 4 species out of 50 present androdioecy (Weeks et al., 2006). The maintenance of androdioecy has been found in simulated metapopulation models in which species are subdivided into a series of small subpopulations with a recurrent colonization of new sites and a relatively high rate of extinction in existing sites (Pannell, 2002) : a situation which can be easily depicted at vents on fast-spreading ridges (Vrijenhoek, 1997). Sometimes androdioic species, especially in plants, are in fact cryptically dioecious (separate sexes), the "hermaphrodites" being in fact functional females only (Charlesworth, 1984). Thus, it is important to evaluate both the male and female functions in hermaphroditic individuals: the availability and functionality of both gametes in spawning being currently under investigation by F. Pradillon' and colleagues team.

The sex determinism has been studied in few androdioecious animals. For instance, in nematodes of the genus *Caenorhabditis*, hermaphrodites are XX and males X0, so the sex is determined by the ratio between the number of sex chromosomes (X) and autosomes (A) (Nigon and Dougherty, 1949). In crustaceans of the genus *Eulimnadia*, where males coexist with hermaphrodites, sex is initially controlled by a single genetic locus with a recessive allele encoding for males (s) and a dominant allele encoding for hermaphrodites (S). The SS and Ss genotypes therefore produce hermaphroditic individuals while individuals with the ss genotype are male (Sassaman and Weeks, 1993). To our knowledge, no gastropod has yet been shown to be an androdioic species, and XY determinism leading to androdioic species has only been shown in plants (like papaya or spinach; Ming et al., 2007; Wadlington and Ming, 2018). The gastropods of the genus *Alviniconcha* are therefore very interesting to better understand how such reproductive mode could have been selected by habitat, *A. kojimai* being the most widespread over the three species. Further anatomical and histological analyses will help check the functionality of the two types of gametes in this species.

In hermaphrodite individuals, the determinism of each cell is determined by its position in the embryo rather than by environmental effects (Beukeboom and Perrin, 2014). Thus,

despite the variability of the hydrothermal environment (Du Preez and Fisher, 2018), in this study, no evidence for the role of environment on sex determination in *Alviniconcha* was found.

A hypothesis about sex determination in *A. kojimai* is that it purely relies on genetic determinism, as it is the case in the two other studied species (XX in females and XY in males), but with an inactivation of genes that stop the male organ development in females. Following such an assumption would give what we observed here, i.e. a genetic determinism where the XX genotype produced hermaphroditic and female individuals. This is exactly what happens for example in the papaya (Ming et al., 2007). In this species, the XX genotype gives females and the XY genotype gives both males and hermaphrodites by affecting the Y chromosome a region named MSY determining the male sex of the individuals and possessing a gene preventing the development of female organs (the carpels). In hermaphrodite individuals, this repressor gene is inactivated and leads to the development of carpels. A consequence of this determinism system is that the descendants of hermaphrodite/male crosses are not viable (Ming et al., 2007).

Another hypothesis could have been that sex determination is manipulated by endosymbionts. The different types of symbionts present in the different species of *Alviniconcha* could be at the origin of differences in the sex determination system if there is parasitic manipulation. To test this hypothesis, we looked at the symbiotic composition between sexes within each *Alviniconcha* species used in this study (Figure 40). The symbiotic composition was derived from metabarcoding analyses performed in Breusing et al. (2022) and used in Chapter 3 of this manuscript. The symbiotic composition seemed to be identical between males and females for each *Alviniconcha* species (Figure 40). Symbiotic distribution is primarily related to geographic population structure of the hosts (Breusing et al. (2022); Chapter 3). Thus, parasitic manipulation in these species could not be identified and therefore does not seem to manipulate sex determinism in *A. kojimai* when compared to the two others species.

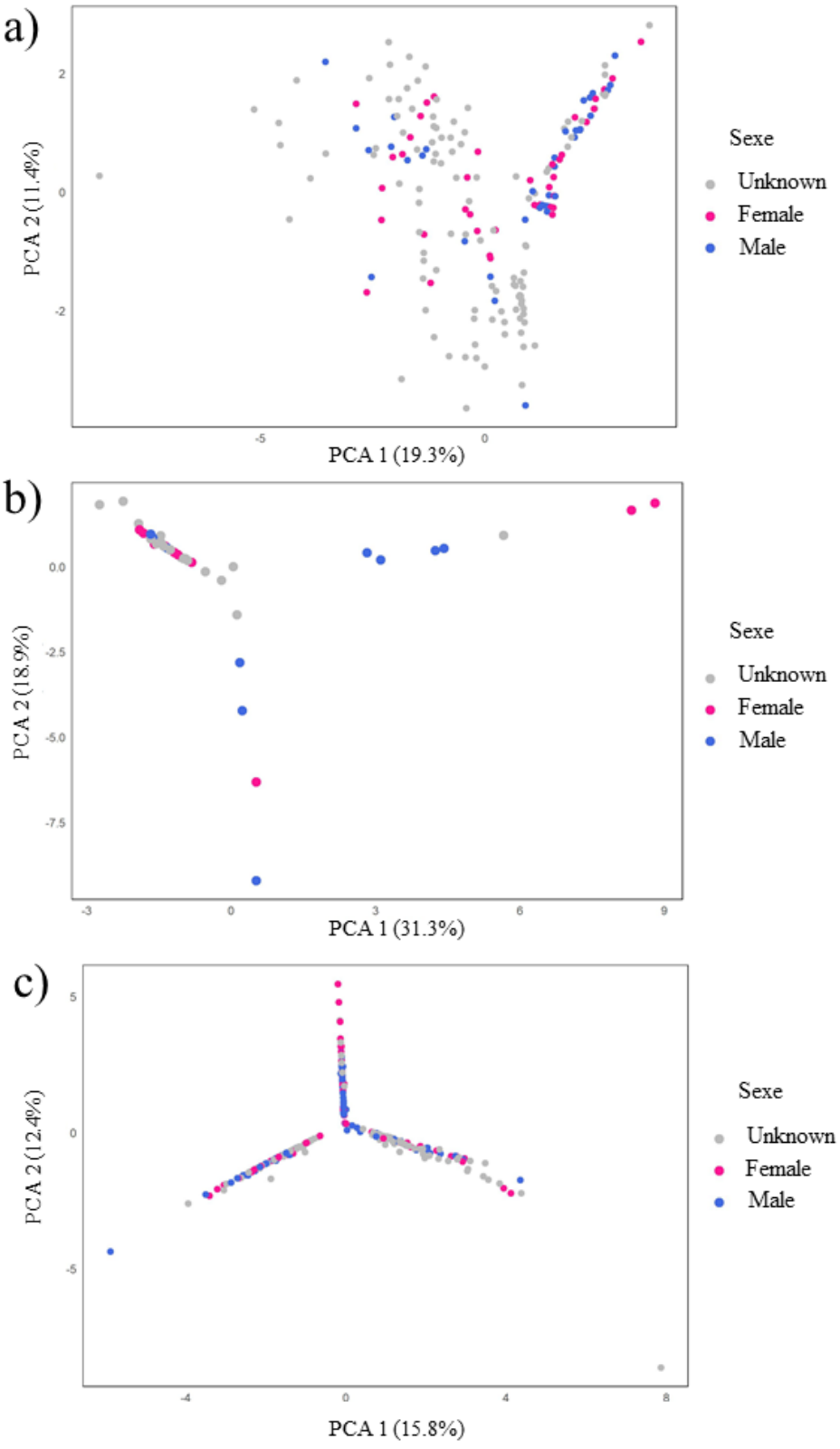


Figure 40: Principal components analysis of : (a) the 18 bacterial strains composing the symbiotic association found in 193 individuals of *A. boucheti* from which 83 individuals could be sexed. (b) the 23 bacterial strains composing the symbiotic association found in 42 individuals of *A. strummeri* of

which 21 individuals could be sexed and (c) the 46 bacterial strains composing the symbiotic association found in 280 individuals of *A. kojimai* of which 160 individuals could be sexed.

4.3- Sex-linked loci synteny across the three species

Within each species, the sex-linked RAD sequences that we identified must belong to a contiguous genomic region where recombination is arrested or restricted in the heterogametic sex (XY males) and where a major gene triggering the sex-determination cascade must be present. We failed to identify all sex-linked RAD-sequences in the three *Alviniconcha* species for three reasons: (1) there is a large component of stochasticity associated with genotyping using the RAD-seq methodology, (2) the power of our analyses to identify sex-linked loci differed markedly between species because of different sample sizes (from 19 individuals in *A. strummeri* to 89 individuals in *A. boucheti*), and (3) the size of the segment where recombination is arrested may vary between species. However, three RAD loci are sex-linked in the three species investigated, and 26 are sex-linked between two species. We conclude from these observations that *A. boucheti*, *A. strummeri*, and *A. kojimai* have homologous sex chromosomes with possibly different sizes.

The statistical power in identifying sex-linked loci was very good for *A. boucheti* and *A. kojimai*, with relatively low F_{ST} (between sexes) and highly negative F_{IS} (within males), as can be seen in Figures. 37B and F. However, the proportion of loci identified as sex-linked was much higher in *A. boucheti* than *A. kojimai* (0.14% vs. 0.006%, $p < 0.001$). Assuming that the proportion of loci identified as sex-linked is a coarse proxy for the proportion of the genome that is composed by a non-recombinant segment in each species, we see that the sex-linked, non-recombinant, region must be much smaller in *A. kojimai*. If the two species have the same genome size, then the sex-linked non-recombinant segment could be more than 20 times smaller in *A. kojimai*. We can therefore speculate that recombination may have been reshuffling genetic variation between X and Y chromosomes in *A. kojimai*. Since it could be a major difference between *A. kojimai* and the two other species studied here, further investigations could try to check whether this could interfere with the developmental program of XX individuals and drive andro-dioecy.

4.4- Evolutionary history of the “hermaphroditic trait” in the genus *Alviniconcha*

The genus *Alviniconcha* is composed of 6 species defined from the mitochondrial gene *Cox1* as well as with some nuclear markers (Breusing et al., 2020; Johnson et al., 2015; Castel

et al. submitted). This genus is sub-divided into two main clades (Figure 41), the first grouping *A. kojimai*, *A. hessleri*, *A. strummeri* and *A. adamantis* and the second grouping *A. boucheti* and *A. marinsidica* (Breusing et al., 2020). Among the species of the first clade, *A. kojimai* and *A. hessleri* are the most closely related and possibly sibling species. As described in this study, the gonochoric character is present in *A. strummeri* and *A. boucheti* (in pink in Figure 41) while the "andro-dioecious character" is present in *A. kojimai* and in *A. hessleri* (in green in Figure 41). Indeed, a recent study on histological sections of gonads conducted on *A. hessleri* by a Canadian team has shown that some individuals of this species are hermaphrodites (Hanson et al., 2021; poster presentation at the 16th Deep Sea Biology Symposium, Brest, France). If this is the case, the most parsimonious view is to consider that the emergence of androdioecy took place in the common ancestor of *A. kojimai* and *A. hessleri* about 10 Mya (green row in Figure 41; Breusing et al., 2020). Although we do not know whether this trait may be found in *A. adamantis* or *A. marisindica*, previous histological examination of the female gonads in the congeneric *Ifremeria nautilei* also supported gonochorism (Reynolds et al., 2010). To complete and clarify the occurrence of androdioecy in *Alviniconcha*, the reproductive mode of more individuals should be analysed among the 6 species of this genus.

The presence of hermaphroditic individuals and therefore possibly androdioic species in *Alviniconcha*, leads to the maintenance of a male/hermaphroditic system for a long time (10 My estimated between *A. hessleri* and *A. kojimai*). However, in most models, this system is transient and could result from habitat change (Charlesworth and Charlesworth, 1978; Charnov, 1982; Weeks, 2012). Here, the system seems to hold, and thus not be disadvantageous to the species. Within a fragmented and locally transient environment, the maintenance of such a system can be explained by the metapopulation model of Pannell (2002), in which a species subdivided into a series of small subpopulations due to the recolonization of new sites because of relatively high extinction rate in existing sites can explain the maintenance of androdioecy. In this model, hermaphroditic individuals are allowed to self-fertilization when the population size becomes too small, but with individuals favouring outcrossing when possible (Pannell, 2002). However, the F_{IS} values observed in this species were not different from other species and did not show any evidence of selfing. Thus, based on this result and the fact that *A. kojimai* displays a very high population density (large number of individuals sampled in the study area), it is perhaps normal that we did not

observe traces of self-fertilization. This view may however change if the species is threatened and undergo strong population bottlenecks with a sufficient decrease in size to favour selfing.

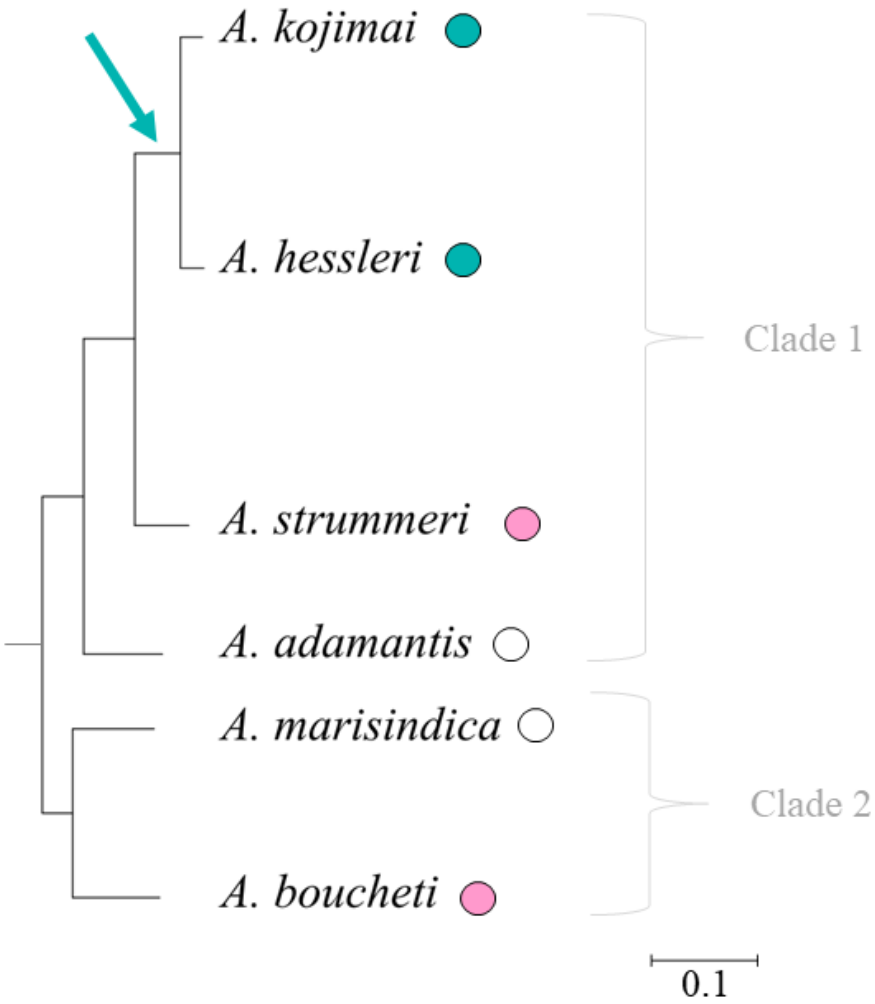


Figure 41: Phylogenetic representation of the genus *Alviniconcha*. In green: hermaphroditic species, in pink: gonochoric species. The green arrow symbolises the potential appearance of hermaphroditism in this genus. Phylogenetic tree made from the *Cox1* mitochondrial gene sequences used in chapter 1 and the tree reconstruction software RAxML (Stamatakis, 2014) under a maximum likelihood model.

4.5- Influence of reproduction modes on the *Alviniconcha* divergence history

Speciation occurs by the evolution of reproductive barriers that ultimately prevent genetic exchanges between previously interbreeding populations. There are many genetic isolation mechanisms, some of which are prezygotic such as gametic mismatches or postzygotic such as inviability of hybrids, sterility or reduction of the hybrid fitness. The

mating system can have a strong effect on pre and postzygotic barriers. Indeed, in species with different mating systems (andro-dioecy/ dioecy), not all possible parental crosses may result in viable zygotes. This is the case for example in *Fucus evanescens* and *F. serratus* or even in nematodes of the genus *Caenorhabditis* where male hybrids are only viable if the mother belongs to the gonochoric species (Coyer et al., 2002; Woodruff et al., 2010). In addition, when hybrids from different reproductive modes are viable, they may be sterile or could have a lower reproductive success compared to the parental species (Coyne, 2007). There is thus selection against the hybrids, which forms a postzygotic barrier.

The demographic history of the three target species of *Alviniconcha* examined clearly showed that they separated in allopatry and followed a long period of isolation preceding a recent recovery of gene flow between them (Castel et al., submitted). Despite this recovery of gene flow, the species are still quite distinct and probably maladapted to cross-mate. Because different mating systems are likely to co-occur in *Alviniconcha* species, with *A. kojimai* and perhaps *A. hessleri* being androdioic while *A. strummeri* and *A. boucheti* are gonochoric, they may have played a role in maintaining genetic divergence between the hybridising *Alviniconcha* species or even at the origin of speciation between the ancestor of *A. kojimai*/*A. hessleri* and *A. strummeri*.

5- Conclusions

The aim of this study based on three species of *Alviniconcha* was to identify the sex determination system. We were able to demonstrate that these species have a XX/XY genetic determinism where males are heterogametic. As opposed to *A. kojimai*, we found a large number of SNPs linked to this determinism in *A. boucheti* and *A. strummeri*. Some of the sex-linked loci were however shared between species, suggesting the presence of a single pair of homologous sex chromosomes. In *A. kojimai*, the presence of hermaphroditic individuals could be related to the presence of a dysfunction in the repressor inhibiting the development of male gonads in XX individuals. We were not able to identify a link between the environment or even symbionts in the determination of sex in these species even if some metapopulation models can predict the maintenance of such systems in natural populations. Thus, this species could be qualified as androdioecious with XY sex determination, and seems to have been maintained for a long time if also present in *A. hessleri*. The maintenance of this type of determinism, usually transient, could be linked to the fragmentation and rapid turn-over of hydrothermal vent sites.

Supplementary material of Chapter 2:

Table S4 : Number of specimens used for determination of sex study

Sample	Site	Field	Basin	*kojimai* male	*kojimai* female/ hermaphrodites	*kojimai* uncertain	*kojimai* unsexed	*boucheti* male	*boucheti* female	*boucheti* uncertain	*boucheti* unsexed	*strummeri* male	*strummeri* female	*strummeri* uncertain	*strummeri* unsexed	Longitude	Latitude	depth (m)
721-GBT1	Tow Cam		Lau	5	3		8									176°08'15,4" W	20°19'04,4" S	2716
721-GBT6	Tow Cam		Lau	1			16	1			1					176°08'15,8" W	20°19'05,1" S	2711
721-GBT7	Tow Cam		Lau						3	5	8					176°08'12,7" W	20°18'59,2" S	2714
722-GBT7	Tui Malila		Lau	4	9	4	2					1	1	2		176°34'04,2" W	21°59'15,2" S	1899
722-GBT1	Tui Malila		Lau	5	2		7					2	3		10	176°34'05,9" W	21°59'21,2" S	1886
722-GBT5	Tui Malila		Lau	3	3		3						3		8	176°34'05,5" W	21°59'21,4" S	1884
731-GBT3	ABE		Lau	1	1			12	7		1	1				176°11'28,9" W	20°45'47,1" S	2149
726-GBT4	Mangatolo		Lau					6	9		3					174°39'12,7" W	15°24'52,8" S	2031
726-GBT3	Mangatolo		Lau													174°39'12,5" W	15°24'52,7" S	2031
726-PBT6	Mangatolo		Lau	6	9											174°39'19,9" W	15°24'57,7" S	2039
726-GBT2	Mangatolo		Lau	3												174°39'12,6" W	15°24'52,5" S	2031
724-GBT4	Phoenix		North Fiji	8	6	1	1									173°55'7,6" E	16°57'0,0" S	1961
724-PBT4	Phoenix		North Fiji	1	3	8	3					3	2	1		173°55'4,7" E	16°56'57,8" S	1973
727-GBT2	AsterX	Fatu Kapa	Futuna	3	1											177°09'07,9" W	14°45'06,5" S	1562
727-GBT4	Stephanie	Fatu Kapa	Futuna	8	5											177°09'57,6" W	14°44'14,7" S	1547
728-GBT2		Fati Ufu	Futuna	7	9	1	1					2	1			177°11'07,0" W	14°45'35,8" S	1519
728-PBT4		Fati Ufu	Futuna	5	5		5								1	177°11'04,9" W	14°45'35,3" S	1519
728-GBT6		Fati Ufu	Futuna	1	1			2	2							177°11'05,9" W	14°45'35,2" S	1518
733-GBT2	Big Papi	Pacmanus	Manus					6	8		7					151°40'20,1" E	3°43'43,9" S	1708
733-GBT8	Fenway	Pacmanus	Manus					3	9		8					151°40'22,4" E	3°43'41,2" S	1696
733-GBT9	Solwara8	Pacmanus	Manus								15					151°40'27,5" E	3°43'49,3" S	1737
733-PBT7	Solwara8	Pacmanus	Manus								5					151°40'26,6" E	3°43'50,1" S	1734

734-GBT9		Pacmanus	Manus							16			152°6'2,8" E	3°43'17,2" S	1659
736-GBT3	Solwara1	Susu	Manus	2	2	14							152°5'47,0" E	3°47'22,1" S	1505
736-GBT10	North Su	Susu	Manus	8	4								152°6'2,8" E	3°47'56,0" S	1218
737-GBT10	South Su	Susu	Manus				2			16			152°6'18,6" E	3°48'35,0" S	1353
737-PBT5	South Su	Susu	Manus			16				12			152°6'17,5" E	3°48'29,8" S	1300
737-GBT7	South Su	Susu	Manus				1			19			152°6'17,9" E	3°48'31,8" S	1343
738-GBT10	Scala		Woodlark	3	2	18							155°03'09,6" E	9°47'56,7" S	3388
739-GBT10	Scala		Woodlark										155°03'07,0' E	9°47'56,3' S	3344
739-PBT5	Scala		Woodlark				10	8	1	1			155°03'08,1" E	9°47'56,0" S	3353

Table S5: Annotation of genes identified as related to sex. In turquoise: in *A. kojimai*; in yellow: in *A. strummeri* and in purple: in *A. boucheti*.

Locus	Description	Go - Molecular function	GO - Biological process	Accession
105604	unconventional myosin-IXb-like isoform X1	Actin, ADP, ATP binding	signal transduction	XP_041362623
29799	sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1-like isoform X3	Calcium ion, chromatin binding	cell adhesion/ epidermis development	XP_033732280
411243	RNA-directed DNA polymerase from mobile element jockey-like isoform X1	RNA-directed DNA polymerase activity		XP_013193561
117457	collagen alpha chain CG42342-like isoform X6	Extracellular matrix structural constituent	Extracellular matrix organization	XP_025084353
59822	sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1-like isoform X3	Calcium ion, chromatin binding	cell adhesion/ epidermis development	XP_033732280
24608	transmembrane protein 26-like	transmembrane protein	transmembrane protein	XP_025088849
60700	*RNA-directed DNA polymerase from mobile element jockey-like isoform X1*	RNA-directed DNA polymerase activity		XP_013193561
89054	nuclear receptor subfamily 4 group A member 2-like	DNA binding	cellular response stimulus	XP_025098363
86749	ubiquitin conjugation factor E4 A-like	Ubiquitin-ubiquitin ligase activity	protein polyubiquitination	XP_025105026
64338	collagen alpha-1(I) chain	extracellular matrix structural constituent	cellular response stimulus/ embryonic skeletal system development/ ossification	XP_035829508
142792	collagen alpha-1(XIII) chain-like isoform X5	extracellular matrix structural constituent	Extracellular matrix organization/ skeletal system development/ ossification	XP_041363379
44670	mitochondrial uncoupling protein 4		response to cold/ mitochondrial transporter proteins	XP_005096635
79842	nuclear factor NF-kappa-B p105 subunit-like	actinin, chromatin, DNAn enzyme binding	cellular response stimulus	XP_025114345

114593	14 kDa phosphohistidine phosphatase-like isoform X5	calcium channel inhibitor activity	signaling pathway/ regulation cell motility	XP_025105068
132160	protein FAM149B1-like isoform X4	cilium assembly		XP_025115323
133109	signal recognition particle subunit SRP68-like	RNA binding	response to drug	XP_025097575
101332	PDZ and LIM domain protein 3-like isoform X1	actin, metal ion binding	cytoskeleton organization/ muscle development/ heart development	XP_025114384
103678	proto-oncogene c-Fos-like isoform X1	DNA binding	regulation of transcription	XP_025105775
10896	polycomb protein suz12-B-like isoform X2	Metal ion binding		XP_025103320
82742	protein phosphatase 1K, mitochondrial-like	Metal ion binding	protein phosphatase	XP_025088483
124722	chitinase-3-like protein 1	Carbohydrate, chitin binding	apoptotic process/ cartilage development/ cellular response stimulus	XP_013090777
39003	unconventional myosin-IXb-like isoform X2	Actin, ADP, ATP binding	signal transduction	XP_041362626
55419	krev interaction trapped protein 1-like	GTPase regulator activity/ microtubule binding	angiogenesis/ regulation of endothelial cell	XP_025100034
39567	epidermal growth factor receptor-like	actin, ATP, cadherin, chromatin, enzyme binding	cellular response stimulus/ cell adhesion	XP_025109037
23354	QRFP-like peptide receptor	Receptor activity	cellular response to peptide/ signaling pathway	XP_041363412
73724	RAD50-interacting protein 1-like		Mitotic DNA damage checkpoint signaling/ protein transport	XP_025106000
75786	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase, mitochondrial-like	ATP binding	Cardiolipin biosynthetic process	XP_025110586
54780	transmembrane protein 26-like	transmembrane protein	transmembrane protein	XP_025088849
105132	rho GTPase-activating protein 24-like	GTPase activator activity	angiogenesis/ cell differentiatio	XP_025086652
67485	actin, cytoplasmic	ATP, protein, kinesin binding	axonogenesis/ cell motility/ chromatin remodeling	XP_796654

119754	ankyrin-1-like	ATPase, enzyme spectrin binding/ cytoskeleton activity	cytoskeleton organization/ exocytosis	XP_025103652
11194	H/ACA ribonucleoprotein complex non-core subunit NAF1, partial	protein, RNA, telomerase RNA binding	telomerase activity/ RNA stabilization	XP_010137290
60906	sodium/hydrogen exchanger 9B2-like isoform X1	protein binding, proton antiporter activity	flagellated sperm motility/ osteoclast development/ clathrin-dependent endocytosis	XP_025095637
34407	protein KTI12 homolog	ATP binding	regulation of transcription	XP_025082952
32431	ankyrin-3-like isoform X1	cadherin, cytoskeletal protein, spectrin binding	Axonogenesis/ membrane assembly	XP_041374460
56568	transmembrane channel-like protein 3	Mechanosensitive ion channel activity		XP_025115665
47642	prosaposin-like isoform X1		signaling pathway/ lysosomal transport	XP_025095361
19214	anoctamin-5-like isoform X2	chloride channel activity	chloride transport/ transmembrane transport	XP_041705744
55033	prosaposin-like isoform X2		signaling pathway/ lysosomal transport	XP_025095369
89419	sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1-like isoform X3	Calcium ion, chromatin binding	cell adhesion/ epidermis development	XP_033732280

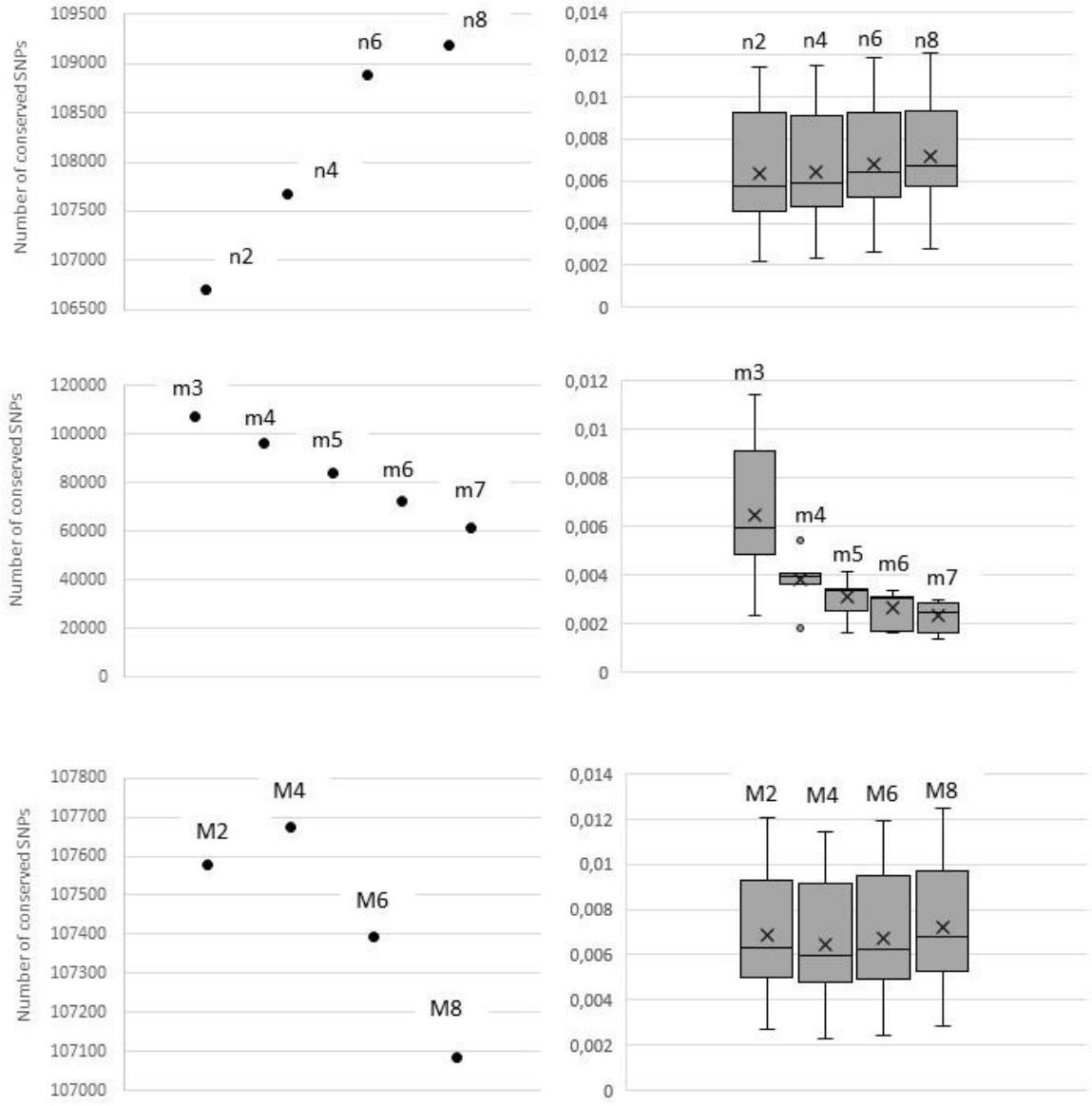


Figure S7 : Number of conserved SNPs and error rate as a function of n, m and M in *A. boucheti* in Stacks' denovo_map.

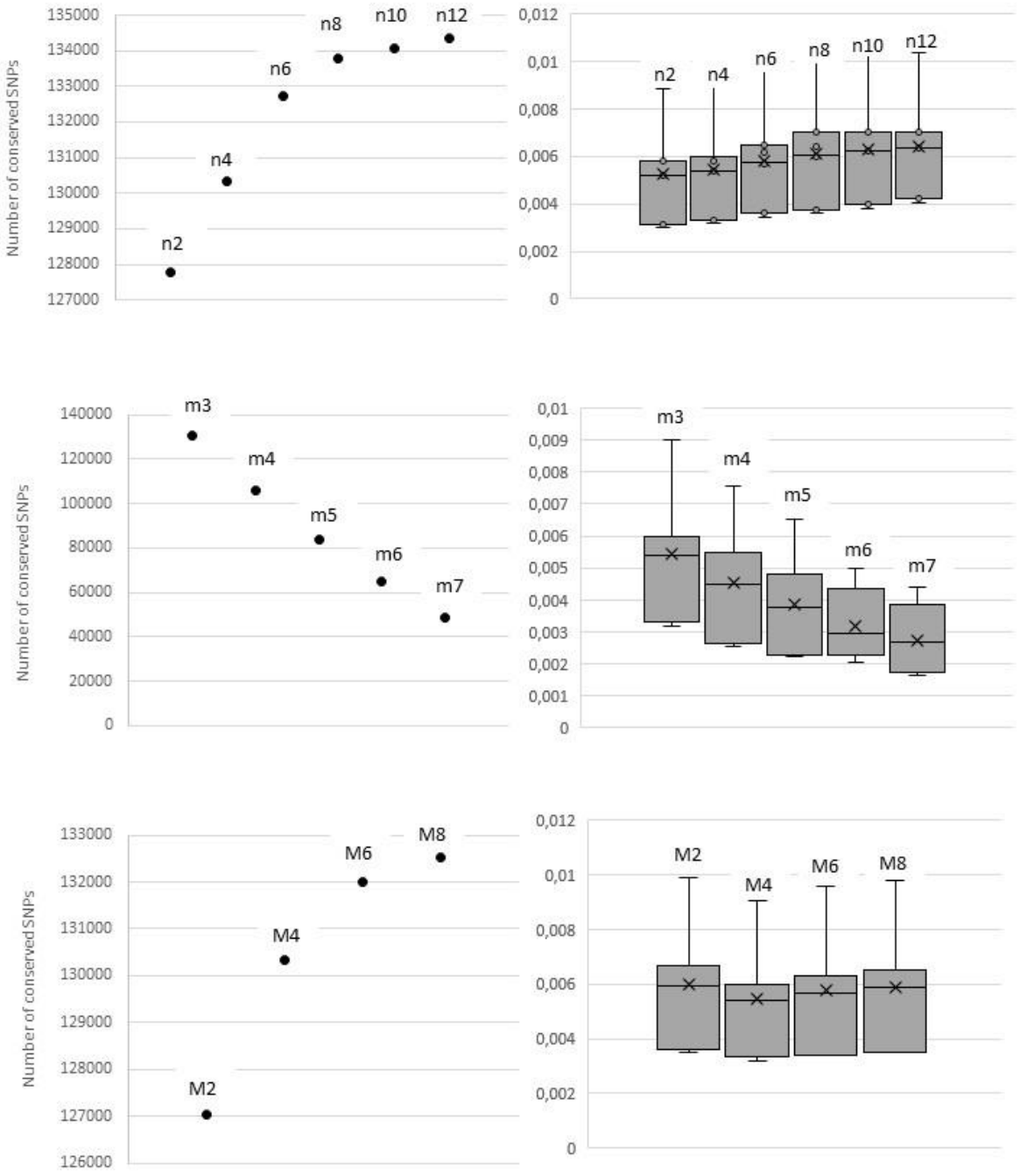


Figure S8: Number of conserved SNPs and error rate as a function of n, m and M in *A. kojimai* in Stacks' denovo_map.

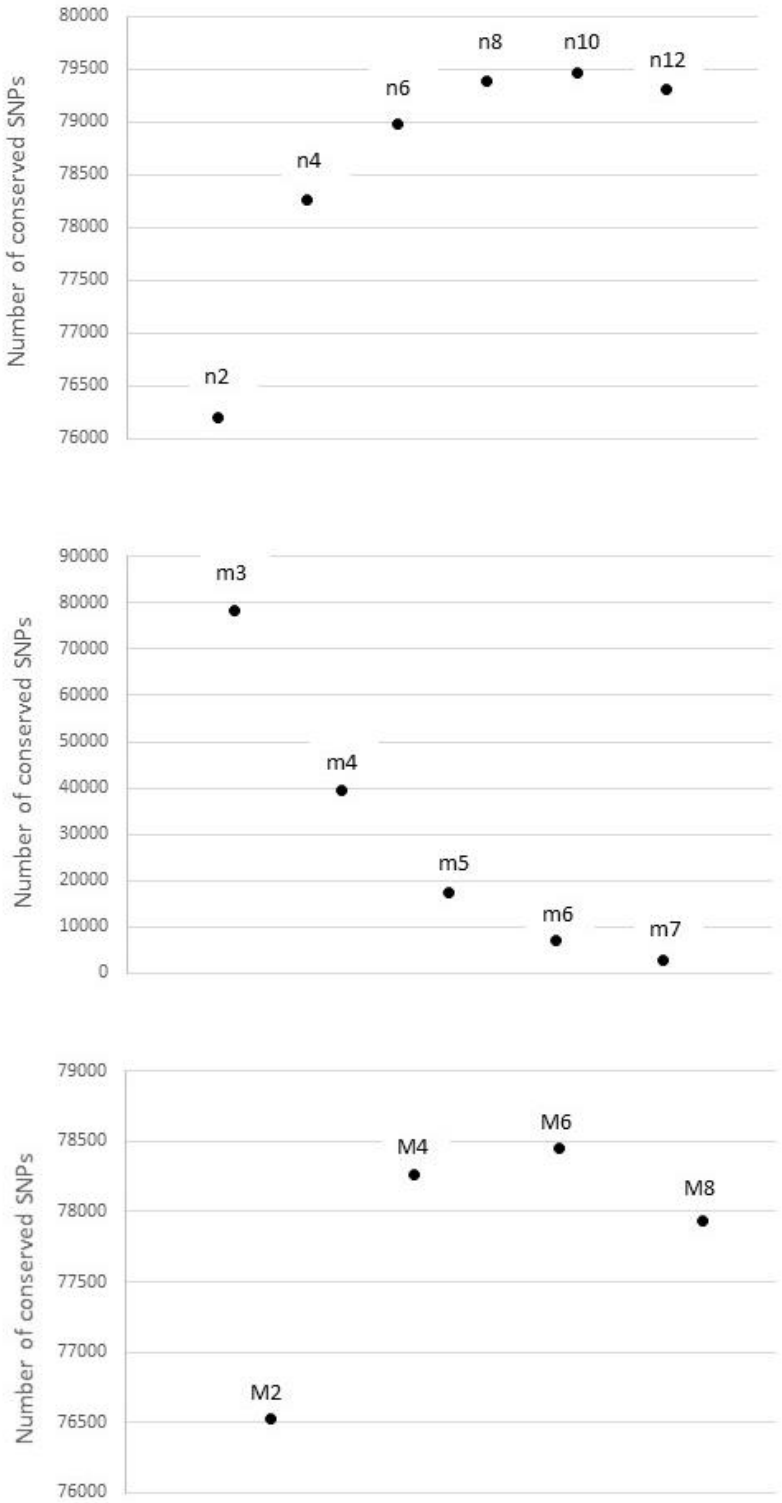


Figure S9: Number of conserved SNPs as a function of n, m and M in *A. strummeri* in Stacks' denovo_map.

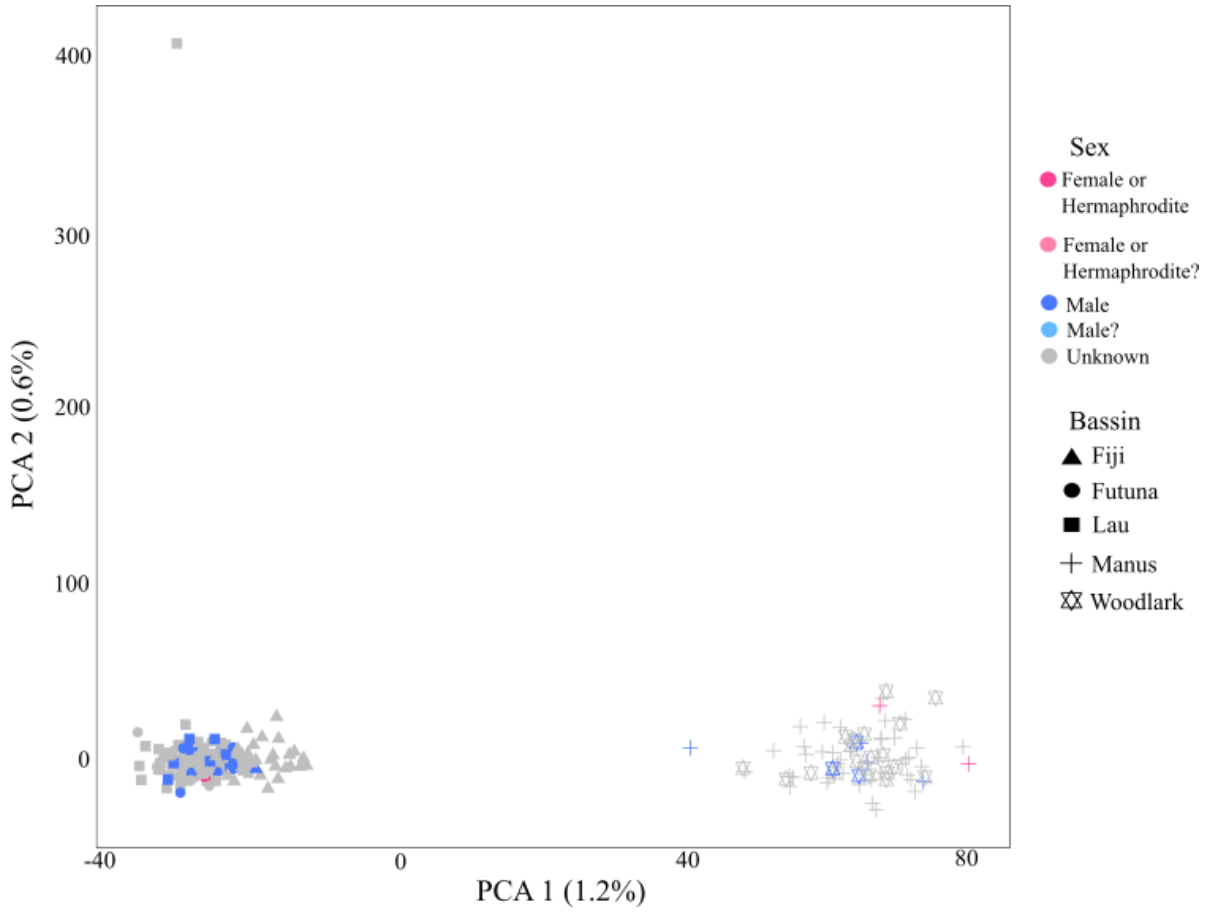


Figure S10: Principal components of multilocus genotypes for *A. kojimai* (247 individuals, 70 122 SNPs). The first axis separates individuals according to their geographic origin: square symbols correspond to the sites from the South-West ("region 1" in Fig. 34 of the thesis), while the dots correspond to the Northern sites ("region 2"). The second axis is driven by a single individual (not shown in Fig. 35 in the main text). Nothing indicates an effect of sex in this species.

Chapitre 3 :

Genomic signatures of potential local adaptation in *Alviniconcha* spp. from the Western Pacific Ocean

Chapter realized in collaboration with Breusing C, Michel L, Cathalot C, Rouxel O, Boulart C, Daguin-Thiébaut C, Ruault S, Mary J, Hourdez S, Gautier M, Beinart R, Jollivet D and Broquet T.

1- Introduction

Understanding how organisms adapt to different environments, or how the environment shapes the genetic diversity of species is a major goal of evolutionary biology (Hoban et al., 2016). Local adaptation occurs when individuals have a better fitness in their local environment (Savolainen et al., 2013). Natural environments are often heterogeneous and variable both in space and time, and local conditions determine which traits will be favoured by natural selection at a given time. When natural selection acts on phenotypic traits, changes in the allele frequencies of loci underlying those traits can move the population toward a local optimum (García-Ramos and Kirkpatrick, 1997). Gradually, natural selection acting on different habitats can lead to adaptive divergence between populations initially adapted to the same environment, providing that selection coefficients are strong enough to counter-balance migration (fitness cost). When ecological divergence between populations also impacts reproductive isolation, identifying the loci underlying local adaptation can help elucidate the process of speciation (Coyne and Orr, 2004). This is what happened for example in cichlid fishes for which divergent selection acted on a gene influencing the fish sensitivity to light and thus their position in the water column and/or specific areas of the lakes. Such a change in phenotypic traits led to the further spatial isolation of populations at different depths/light regimes and also affected the choice of the mating partner which mainly relies on its colouration (Gross, 2006). Conversely, it is important to note that the effects of selection are countered by gene flow homogenising populations. Thus, if dispersal is high among populations, local adaptation will tend to be swamped by the immigration of maladapted individuals from external populations and recombination between local and foreign alleles (Sanford and Kelly, 2011).

The recent development of less expensive and more rapid methods to obtain high-quality genome-wide data allows the identification of potential loci responsible for adaptive differences between populations (genome scans of outlier loci). Two approaches for identifying differentially adapted loci have been described and are commonly used. The first one looks for loci with unexpected levels of high genetic differentiation between populations when compared with the genome-wide average of all loci (aberrant differentiation methods;

i.e. Beaumont and Balding, 2004; Foll and Gaggiotti, 2008). This approach is however not sufficient by itself to certify that the abrupt allele frequency changes at the outlier loci are always linked with diversifying selection as they may be also attributable to hybridization, some chromosomal rearrangements or sexual chromosomes (Bierne et al., 2011). These methods should therefore combine divergence or differentiation (F_{ST}) estimates with additional parameters such as diversity indices or site linkage disequilibrium to discriminate with accuracy adaptive loci (Booker et al., 2021; Nielsen, 2005; Nielsen et al., 2011; Wersebe et al., 2022) but also need to take into account false discoveries in the genome (François et al., 2016). The second set of methods are usually investigating correlations between population allele frequencies and environments (genetic-environment association - GEA; Frichot et al., 2015). However, GEA analyses have also some limitations. The first one is the difficulty of distinguishing patterns associated with neutral demographic processes from those that are the consequence of selection. The second is the difficulty of disentangling the joint effect of geography and habitats when they tightly linked together (i.e. environmental/latitudinal gradients: Nadeau et al., 2016). A third limitation relates to the choice of environmental variables to be included in the analysis. Indeed, the prior selection of the most relevant variables for any particular GEA analysis is complicated because since some previous knowledge about which variables may be relevant in the adaptation process is needed (Bogaerts-Márquez et al., 2021).

Recently developed softwares, such as BayPass (Gautier, 2015), have overcome some of these limitations. This sample-based software identifies SNPs with statistically different allelic frequencies between geographic populations and those associated with environmental variables, while accounting for covariance between allele frequencies due to the shared demographic history of populations and due to local adaptation (Gautier, 2015). Moreover, this software includes a single covariate regression model where the association is estimated for each covariate, which allows the study of adaptation against a large number of ecological factors.

Most analyses dealing with local adaptation in marine environments used few uncorrelated environmental variables (Sanford and Kelly, 2011). Amongst these studies many have sought to understand the impact of abiotic factors such as temperature, wave action, salinity, or chronic pollutions in coastal environments (Daka and Hawkins, 2004; Janson, 1982; Pardo and Johnson, 2005; Rolán-Alvarez et al., 1997; Sokolova and Pörtner, 2001; Trussell, 2002; Yamada, 1989). Others much rarer were more focused on some biotic factors

such as predation and prey availability on local adaptation (Janson, 1983; Johannesson and Johannesson, 1996; Sanford and Worth, 2009; Trussell, 2000). All of these factors play a strong role in local adaptation, as shown for marine species experimented in shared gardens where the individual's fitness is lower to some individuals for which the environment is not conducive (Sanford and Kelly, 2011). In the snail *Littorina saxatilis*, for example, it has been shown that genetic differences occur between the high- and low-shore ecotypes due to the wave action (Galindo et al., 2009). In the hydrothermal environment, other factors in addition to temperature that can play a role in the local adaptation of populations are studied, such as depth and currents (Nukazawa et al., 2015).

The hydrothermal environment despite some convergent specificities limiting life (sulphide-rich, reduced and hypoxic conditions), also represents a mosaic of habitats depending on both the chaotic mixing of the subterranean hot and sulphurous fluid with the surrounding deep water (Le Bris et al., 2003; Von Damm, 1990), and the chemical nature of the underlying oceanic crust. These environmental conditions vary greatly depending on the physical distance of organisms to the fluid emission and the fluid composition which mainly depends on the geographic position of the venting site and the geodynamic features of ridges. These variations of the fluid emissions have therefore a strong impact on species composition and more specifically the composition of microorganism assemblages which exploit the reduced compounds derived from the fluid to produce organic matter and are source of the primary production.

The genus *Alviniconcha* and in particular species *A. kojimai*, *A. boucheti* and *A. strummeri* seem to represent good models for inter- and intra-specific studies of local adaptation. Indeed, these species have colonised a wide-range of hydrothermal vent emissions from diffuse venting sites to chimney walls in the Western Pacific Ocean: *A. boucheti* and *A. kojimai* are found over the ridge system of four back-arc basins (Lau, North Fiji, Manus and Woodlark) as well as in the Futuna volcanic arc, while *A. strummeri* is only present at venting sites of the Futuna arc and the Lau and North fiji basins (Johnson et al., 2015; Castel et al. submitted). For each species studied, subtle habitat differences exist between geographic populations, notably between populations close to chimneys (on polymetallic sulfides) or further away (in diffusion zones on basaltic scree). *A. kojimai* and *A. strummeri* are indeed more frequently found in diffuse venting areas while *A. boucheti* is preferentially found on the wall of chimneys and thus closer to the emissions (Beinart et al., 2012; Breusing et al., 2020; Castel et al. submitted). These differences in ecological niches between species could

be in fact linked to bacterial communities associated with these different substrates and by consequence on the type of symbionts they are able to horizontally acquire. Previous studies have shown that *A. kojimai* and *A. strummeri* are mostly associated with *gamma-proteobacteria* that tolerate low H_2 and H_2S concentrations. *A. boucheti* possesses *Campylobacteria* which, by opposition, can tolerate higher concentrations of H_2 and H_2S, hence the presence of this species closer to the emissions of hydrothermal fluids rich in these compounds (Beinart et al., 2012, 2014; Breusing et al., 2020, 2022). Despite each species possesses a specific type of symbiont, variations in symbiont composition also exist at the intraspecies level as recently shown in Breusing et al. (2022) with bacterial strains of different geographic origins. Thus, geographical distribution of hosts and symbionts, local variations in the habitat chemistry and in symbiont composition between and within back-arc basins makes these symbiotic gastropod species interesting to assess the role of local adaptation on population evolution at the intra-species level and to compare its strength across species.

The three target species of *Alviniconcha* are part of a genus composed of six species that have been previously described using as a proxy their divergence at the mitochondrial gene *Cox1*(Johnson et al., 2015). The demographic study based on these species indicated that on spite of their strong divergence (in average 10% on mtDNA and 3% on nuDNA), these gastropods were still able to sporadically exchange genes due to recent secondary contacts between populations after a long period of isolation (Castel et al. submitted). Such inter-species introgression, although limited, may have a positive effect on the interacting species by promoting new host-symbiont interactions locally and therefore questions the possibility that adaptive introgression locally shapes the distribution of the three species between slightly different ecological niches.

In this study, we combined genome scans of differentiation and genome-wide SNPs GEA analysis using ddRAD-seq data available for 536 samples of *Alviniconcha*, representative of the whole genetic diversity of the species across the Western Pacific Ocean (16 sites). Our aims were to characterise: (1) to which extent environmental variables such as depth, broad habitat type, temperature, chemical conditions, and symbiont composition contributed to adaptive differentiation of local populations in each *Alviniconcha* species, and (2) investigate whether local adaptation may be a factor strengthening interspecies genetic differentiation.

2- Materials and methods

2.1- Sampling collection and sequencing

We used a subset of a previously published genomic data set of ddRAD-seq sequences from Castel et al. (submitted: Archive Accession no. PRJNA768636). In short, 275 individuals of *A. kojimai*, 215 *A. boucheti* and 46 *A. strummeri* were sampled during the CHUBACARC expedition (2019) from 16 sites and 5 back arc basins of the Western Pacific Ocean (Figure 42; Table S6). Individuals were sampled in collection boxes ("Biobox") from gastropod patches on surfaces corresponding to a 0,25 m^2 quadrat, which will be referred to as 'populations' in this study (one population = one sample of 3 to 32 individuals sampled from a unique patch, see general methods in the "data acquisition" chapter of this thesis). Six additional individuals from the Nifonea volcano of the New Hebrides subduction arc collected in 2013 were added to the analyse the genetic structure of *A. boucheti* (in black in Figure 42). gDNA of each specimen was extracted and purified with the NucleoSpin® Tissue kit (Macherey-Nagel) or the CTAB protocol (Doyle and Dickson, 1987 modified in Jolly et al., 2003) to produce individual-based ddRAD libraries following the protocol developed by Daguin-Thiébaut et al. (2021). DNA fragments libraries were Illumina sequenced to generate paired-end 150-bp reads using three lanes of Novaseq 6000 (Details in Echantillonnage et acquisition des données génétique of this thesis).

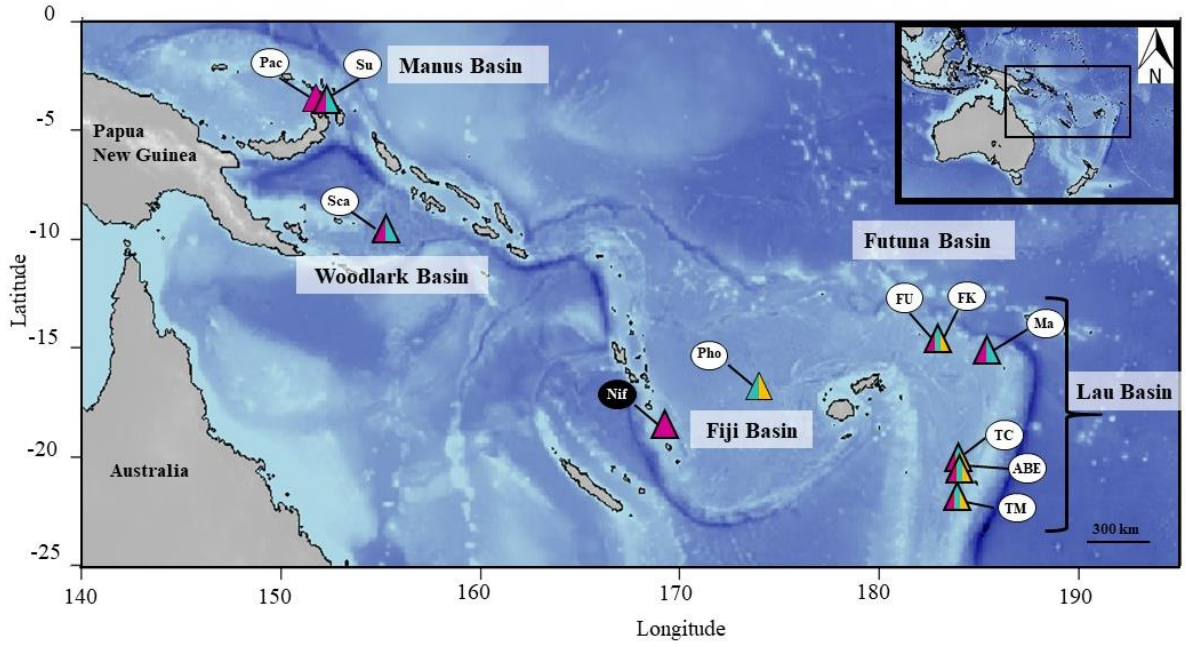


Figure 42 : Sampling location of *Alviniconcha boucheti* (purple), *A. kojimai* (yellow), and *A. strummeri* (turquoise). Most localities (names in white circles) were sampled during the

CHUBACARC expedition in 2019, and one site (Nifonea, black circle) was sampled in 2013. Manus Basin: Susu (Su); Pacmanus (Pac). Woodlark Basin: Scala (Sca). Fiji Basin: Phoenix (Pho); Nifonea (Nif). Futuna Basin: Fatu-Kapa (FK); Fati-Ufu (FU). Lau Basin: Mangatolo (Ma); Tow Cam (TC); ABE (ABE); Tui Malila (TM). The Susu hydrothermal field in the Manus Basin contains the North_Su, South_Su and Suzette sites, not shown on this map. Fenway, Big_Papi and Solwara8 sites were sampled at the Pacmanus field and the Stephanie and AsterX sites at the Fatu-Kapa hydrothermal field of the Futuna arc. For more information on sites at a small-scale geography are presented in Figure 13.

2.2- SNP calling

Individual raw reads obtained after the Illumina sequencing were treated independently within each species in a way similar to the protocol used in Castel et al. (submitted: see chapter 1). These new sets of sequences were then analysed to examine the spatial genetic structure of each species and search for outlier loci candidate for local adaptation (Echantillonnage et acquisition des données génomiques and chapter 2: sex determination). Briefly, RAD data were demultiplexed using the process_radtags module of Stacks v2.52 (Catchen et al., 2013; Catchen et al., 2011). A filtering of non-eukaryotic reads was done with kraken v2 (Wood and Salzberg, 2014) and loci assembly was performed with the denovo_map module of Stacks. The assembly parameters used in this chapter were the same as those used in the sex determination chapter (*A. strummeri*: *m*: 6; *M*: 8; *n*: 8; *A. boucheti: m*: 6; *M*: 4; *n*: 4 and *A. kojimai: m*:6; *M*:10; *n*:10). As a reminder, these parameters were chosen because they maximise the number of SNPs while limiting the error rate on a sub-dataset composed of triplicated individuals. After assembly, the raw SNPs data were filtered in Stacks (module *populations*) against the following thresholds: minimum individuals sharing a locus in a population $r \geq 0.8$ and minor allele count MAC ≥ 4. Then, SNPs identified by Stacks were furthered filtered for missing data using home-made R scripts (R Core Team, 2020) in order to only keep SNPs present in 90% of the individuals and to keep only individuals which presented more than 85% of the total number of SNPs found across populations. An additional filter was also used to remove SNPs present on sex chromosomes for *A. strummeri* and *A. boucheti* only (see chapter 2 on sex determination). The rationale for eliminating these sex-differentiated SNPs is that their allelic frequency differences do not follow the observed geographic structure and thus could be falsely identified by GEA software as a signal of local adaptation.

In the end, data sets of 160 592 SNPs from 211 individuals of *A. boucheti*, 64 049 SNPs from 251 individuals of *A. kojimai*, and 10 608 SNPs from 41 individuals of *A. strummeri* were obtained. The population structure was first visualised in R using a principal component analysis (PCA; adegenet package; Jombart, 2008). Then, F_{ST} between species and between populations of each species were calculated with the R package hierfstat (Goudet, 2020).

2.3- Environmental data acquisition

Up to 76 environmental variables (Table S7) were analysed for each of the 31 samples of *Alviniconcha* species ("Biobox" populations, Table S6). These included, geographic data (depth, habitat type: Table S6), abiotic ecological variables (fluid chemistry: see Table S8) and biotic ecological variables (symbiont and $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ isotopic composition of hosts, see Table S7).

Fluid chemitry of diffuse emissions surrounding gastropods was acquired through collaborations with the geochemists of the CHUBACRAC expedition. (C. Cathalot, O. Rouxel and E. Rinher (Ifremer, Brest) and C. Boulard (Roscoff Biological Station)). The chemical concentrations of 25 metallic elements and gas (detailed in Table S8) were measured from 22 populations of *Alviniconcha* distributed within and between the four back-arc basins and the active zone of Futuna (with the exception of Nifonea). *In situ* measurements of the temperature as well as sulphide and iron concentrations were carried out on each gastropod assemblage sampled using the multi analyzer Chemini (Vuillemin et al., 2009). Diluted fluids were also collected with the PIF multi-sampler (derived from the PEPITO sampler; Sarradin et al., 2008) onto the same assemblages at three distinct spatial spots. The full environmental dataset were partly obtained on board from the probes following standardization with the surrounding water and sulphide and iron standards and metallic elements and gas were subsequently measured by Gas Chromatography-Flame Ionization Detection, high-resolution inductively coupled plasma mass spectrometry (HR-ICPMS) and Gas Chromatograph (Shimadzu GC-2030) fitted with a barrier discharge ionisation detector (BID) and a 30-m SH-Rt-MSieve 5A column. The intra- and inter-species geographic variation of each chemical element was first assessed between bioboxes using both boxplots and PCA and the covariance of the chemical elements was evaluated with the package corrplot (R).

The stable isotopic analysis of $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ were performed by L. Michel (Ifremer, Brest) on 309 gastropod tissues. These individuals were distributed within 29 of the 31 geo-localised populations used for the genetic analysis. Gastropod foots were dried at 60°C, then stored at room temperature before being ground into a homogeneous powder using a mortar and pestle. Stable isotope measurements were performed via a continuous flow-

elemental analysis-isotope ratio mass spectrometry (CF-EA-IRMS) at University of Liège (Belgium), using a vario MICRO cube C-N-S elemental analyzer (Elementar Analyse systeme GMBH, Hanau, Germany) coupled to an IsoPrime100 isotope ratio mass spectrometer (Isoprime, Cheadle, United Kingdom). Isotopic ratios were expressed using the widespread δ notation, in ‰ and relative to the international references Vienna Pee Dee Belemnite (for carbon), Atmospheric Air (for nitrogen) and Vienna Canyon Diablo Troilite (for sulphur). Isotopic ratios were compared between populations and species by performing biplot graphs and PCA (in R). For the GEA analysis, isotopic data ($\delta^{13}C$, $\delta^{15}N$, and $\delta^{34}S$) were used as population variables by averaging individuals sampled within a biobox.

The symbiotic composition of each ddRAD-seq genotyped individual was obtained from the host gill DNA with a metabarcoding approach described in Breusing et al. (2022) (the article is presented as an appendix to this chapter which also contains all the symbiotic composition of our gastropod samples). To summarise how these data were obtained, gDNA was extracted from alcohol-preserved symbiont gill tissue with the Qiagen DNeasy Blood and Tissue kit (Qiagen, Inc.), and a 2x250 bp paired-end amplicon library were constructed for the 16S rRNA V4-V5 region with the 515F/926R primers pair (Walters et al., 2015). This fragment library was then sequenced using a Illumina MiSeq sequencer at the genomic platform of Argonne National Laboratory (Lemont, IL, USA). The USEARCH v11 pipeline (Edgar, 2010) was used to identify the different Amplicon Sequence Variants (ASVs). The taxonomic assignment of each ASV was determined by BLAST+ searches against the NR database (see Breusing et al., 2022 for more details). Phylogenetic relationships among ASVs were determined with the IQTREE (Minh et al., 2020) plugin for QIIME2 based on 10 independent runs with 5000 ultrafast bootstrap resampling of the dataset. The number of reads per ASV depends on the abundance of the bacterial clade in the host as well as on the sequencing effort. For each ASV within each individual, the number of reads was transformed into a frequency per individual and only kept if it represented more than 2.37% of the total number of reads to account for sample cross-contamination (Breusing et al., 2022; Minich et al., 2019). For the GEA analysis, each microbial taxa abundance was coded as presence/absence (0 or 1). The presence of a microbial clade was only considered when it represented more than 2% of the total number of ASVs found within a given host. In the end, a complete information on genetic data and all environmental variables was obtained for 23 bioboxes (containing between 3 and 32 individuals). To optimise the efficiency of the downstream analyses, we considered two separate data sets (Table 6): 23 bioboxes for which

we had the full set of chemical and genetic data and 30 bioboxes for which we only had genetic, habitat, isotopes and microbial composition data.

2.4- Genetic-Environment Association analysis

For GEA analyses, each biotic and abiotic parameters typifying the environment of gastropod samples were averaged within each sampling habitat to be then used as an explanatory variable. Based on SNPs identified within each species separately, genome scans of sample differentiation and associated correlations with environmental variables were performed using BayPass v.2.2 (Gautier, 2015). The model used by BayPass takes into account the correlation structure between allelic frequencies and allows the identification of SNPs subject to sample differentiation unexpectedly greater than expected based on the XtX statistic (Günther and Coop, 2013), which is a measure of the genetic differentiation between samples. BayPass was run under a core model to calculate individual SNP XtX statistics for each species. For further analyses, we focused on SNPs with the highest XtX values (top 5%), following (Bogaerts-Márquez et al., 2021). Usually a threshold of 1% is applied on the highest XtX values to define outliers as they represent loci with a level of differentiation impossible to obtain under neutral assumptions and are therefore necessarily under selection (Gautier, 2015). Here, we chose to use a 5% threshold onto the highest XtX values in order to increase the number of outlier loci for their further annotation and get enough loci to mine possible biological functions. Outlier loci at a 1% threshold were however later used for the associative analysis with ecological variables.

The BayPass STD model extends the previous analysis to determine the association of allele frequencies with population-specific covariates (i.e. environmental variables). This model of covariance is run for each species with each of the 76 previously described environmental variables, and the default options except for the -scalecov option that was used to scale each covariate. We performed two series analyses independently for each species. The first analysis used all populations of a given species and the variables of depth, habitat type, isotope ratios and symbiotic composition of gastropods averaged at the biobox sample level. The second analysis used a genetic dataset with less individuals and the chemical measurements also averaged per biobox (Table 6). Correlations between genetic variants and environmental variables are then assessed using a Bayes factor (BF: empirical Bayesian *P*-values) measured in units of deciban (dB, via the transformation 10log10(BF)) and estimated with an importance sampling algorithm from MCMC samples (Coop et al., 2010; Gautier,

2015). We considered a conservative BF threshold of 20 dB ("decisive evidence" according to Jeffreys' rule; Jeffreys, 1961) as an evidence of association between one environmental variable and one SNP to limit the number of false positives. Finally, significant environmentally-driven outlier loci were retrieved if they jointly fall within the fraction of loci having the highest XtX values at a 1% threshold and a BF equal or greater than 20 dB. A consensus sequence of these loci (obtained with the function –fasta-loci of the population module of Stacks v2.52) was then annotated using blastX against NCBI databases in Geneious (Kearse et al., 2012). Subsequently, a GO (Gene Ontology) annotation was performed for the annotated RAD loci by looking at each blastX assignment in the Uniprot (protein) sequence database.

Table 6: Summary of the three genetic datasets (*A. kojimai*, *A. boucheti* and *A. strummeri*) used in the two BayPass analyses. Nb pop = Number of biosamples used, Nb var = Number of environmental factors, SNPs = number of SNPs retrieved from the reads filtering for each species following Stacks and Kraken. The number of variables for each species is different in analysis 1 because the number of clades identified is changing according to the species.

Analysis	Species	Nb pop	Nb var	SNPs
1st - Depth/Habitat	*A. kojimai*	19	38	37 776
	A. strummeri	7	12	10 060
	A. boucheti	11	19	96 155
2nd - Fluid chemistry	*A. kojimai*	16	25	19 661
	A. strummeri	6	25	19 349
	A. boucheti	6	25	130 039

3- Results

3.1- Geographic structure of the three *Alviniconcha* species

The genetic structure of the three gastropod species has been assessed using species-specific PCAs which are shown in Figure 43. In the *A. boucheti* analysis based on 211 individuals and 160 591 SNPs, the first axis only explained 0.8% of the whole genetic variance and separated individuals according to two distinct geographic metapopulations (see Figure 43a). More precisely, the two groups of differentiated individuals correspond to individuals from the East of the Western Pacific (Lau, Futuna and Fiji basins, which we called “region 2” in other chapters). The second species *A. kojimai* (251 individuals and 64 049 SNPs) displayed the same genetic structure in two distinct East and West metapopulations

(first axis of the PCA explaining 1.2% of the genetic variance) but with also a slight geographic differentiation of individuals from the North Fiji Basin when compared to those from Lau/Futuna. As opposed, Manus and Woodlark individuals were differentiated according to the geography but displayed some within variation in the Western metapopulation. For *A. strummeri* (41 individuals and 10 607 SNPs), the PCA also showed a clear genetic differentiation between individuals according to their geographic origin, but at a more restricted spatial scale which also represents the species range. Specifically, genetic differentiation between populations was more pronounced than the two other *Alviniconcha* species separating the North Fiji and Lau/Futuna basins (first axis of the PCA explains 4.3% of the variance). The total genetic differentiation between populations of each species (as estimated by pairwise F_{ST} between basins) are presented in Table 7. In general, the level of genetic differentiation between populations was weaker for *A. boucheti* (three times less) than for the other two species. As shown by the PCAs, the highest F_{ST} values were obtained between the Eastern and Western basins, while the intra-metapopulation F_{ST} between the Lau and Futuna as well as between Manus and Woodlark Basins was zero. The strongest values of pairwise F_{ST} (0.022 and 0.02) were obtained within *A. strummeri* between Lau/North Fiji and Futuna/North Fiji basins respectively.

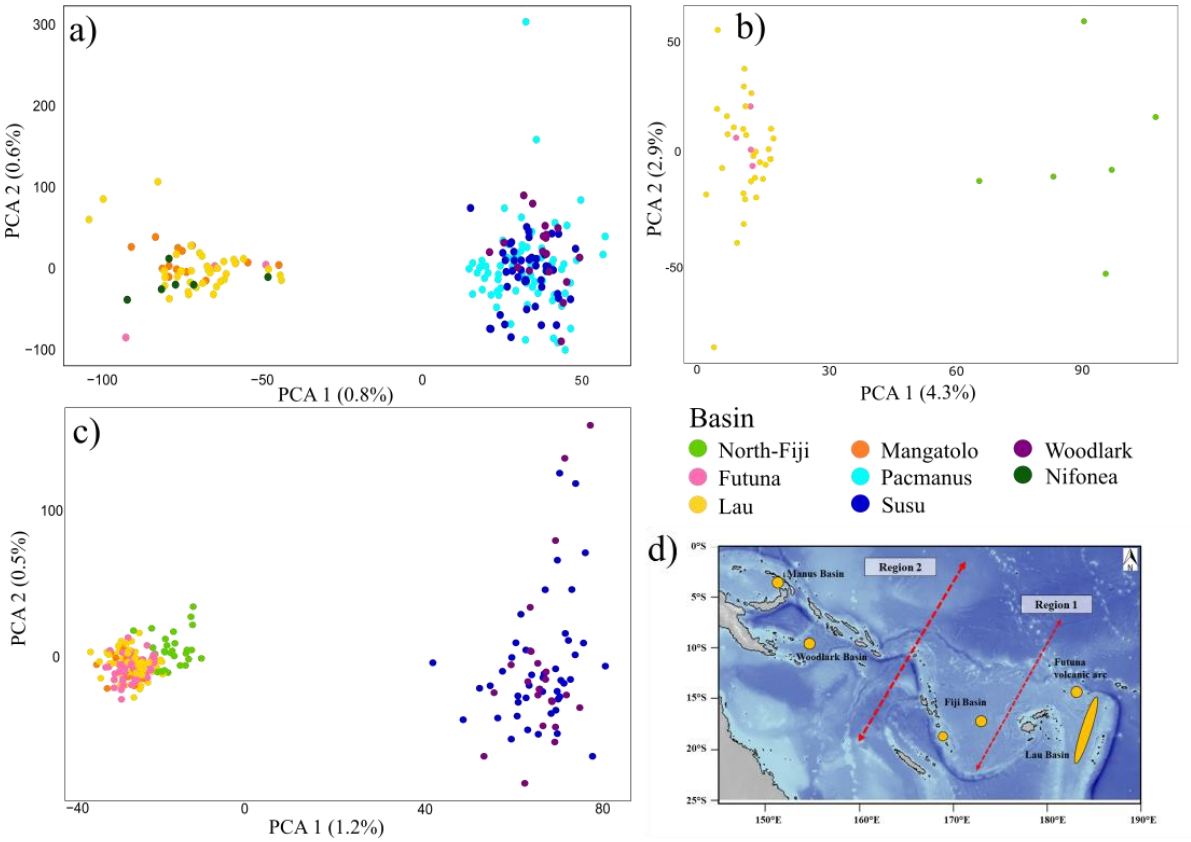


Figure 43 : Principal components analysis on SNP species datasets using the multigenotypic information on individuals for (a) *A. boucheti* using 160 591 SNPs across 211 individuals, (b) *A. strummeri* using 10 607 SNPs and 41 individuals, (c) *A. kojimai* using 64 049 SNPs across 251 individuals with additional information on (d) the sampling design on venting sites where the first axis separated populations of the East (Region 1) and the West (Region 2). In this species, the individuals from the Fiji Basin also seemed to be slightly differentiated from the other populations. The red arrows represent putative genetic breaks between populations of the three *Alviniconcha* species.

Table 7: Pairwise F_{ST} estimates between geographic populations of each species. (a) *A. boucheti*, (b) *A. strummeri* and (c) *A. kojimai*.

a)	Lau Basin	Futuna	Fiji Basin	Manus Basin
Lau Basin				
Futuna	0			
Fiji Basin	0.003	0.003		
Manus Basin	0.01	0.01	0.008	
Woodlark Basin	0.01	0.01	0.008	0

b)	Lau Basin	Futuna
Lau Basin		
Futuna	0	
Fiji Basin	0.022	0.02

c)	Lau Basin	Futuna	Fiji Basin	Manus Basin
Lau Basin				
Futuna	0			
Fiji Basin	0.001	0.001		
Manus Basin	0.002	0.003	0.003	
Woodlark Basin	0.002	0.003	0.003	0

3.2- Symbiotic composition of gastropods

Variations of the symbiotic composition in the gills of individuals from our three *Alviniconcha* species presented here are a subset of the results presented by Breusing et al. (2022). A total of 46 ASVs (bacterial taxa) have been identified in all our genotyped *Alviniconcha* among the 5 back-arc basins of the Western pacific. These ASVs were assigned to two genera of C*ampylobacteria* (*Sulfurovum*, *Sulfurimonas*) and three *gammaproteobacteria* (*Thiolapillus*, *Gamma-Lau* and *Gamma*-1) (Figure 44). Among these bacterial taxa, ASVs (Oligos) appeared to be further subdivided according to geography. This was notably the case for *Gamma-1* where oligos present in the Lau Basin as well as those present in the Futuna were grouped together (Figure 44). *A. boucheti* was dominated (more than 95%) by different basin-specific *Campylobacteria* ASVs of the genera *Sulfurimonas* or *Sulfurovum* (Figure 45) whereas *A. kojimai* and *A. strummeri* were dominated by different Gammaproteobacterial ASVs (more than 80%) from different geographic origins (Figure 45). Individuals of *A. kojimai* (2 populations representing 35 individuals) and *A. strummeri* (1 population - 4 individuals) from the North Fiji and Futuna basins respectively showed a significant presence of *Sulfurovum* (up to 18% of the individual bacterial composition of the gills). It should be noted that the populations of *A. kojimai* found in the Manus and Woodlark basins also exhibited *Thiolapillus* symbionts (up to 5% of the individual bacterial composition of the gills) that are usually typical of *Alviniconcha hessleri*, a sibling species that is present only in the Mariana Trench, 2 000 km North of the Manus Basin (Breusing et al., 2022). While symbiont composition appeared to be generally linked to geography, some variation between sites was unrelated to this structuring factor (such as the presence of *Sulfurovum* at

certain localities) making it interesting to study the influence of symbiont composition on the local adaptation of *Alviniconcha* populations.

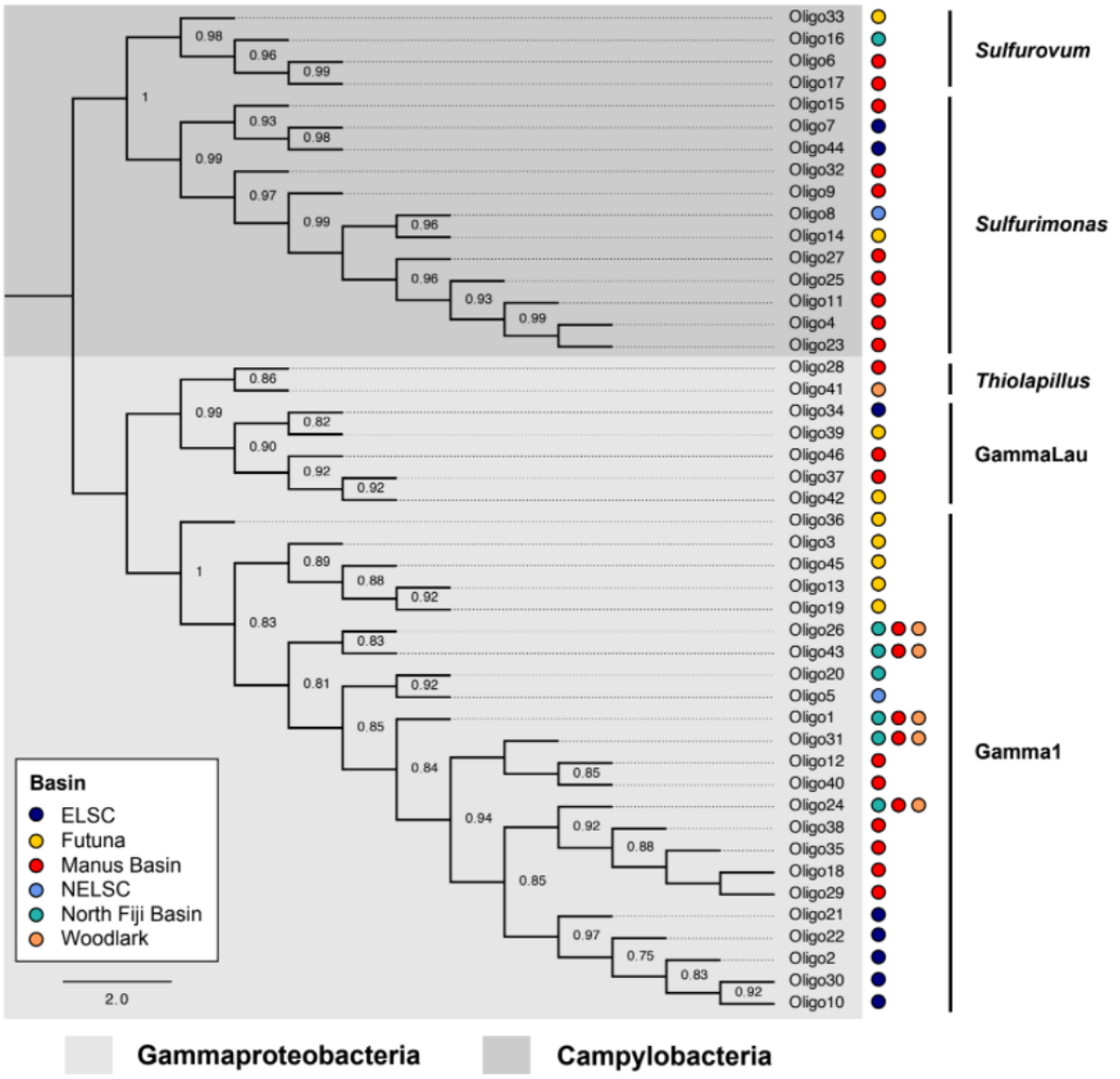


Figure 44: Mid-point rooted IQTree (Minh et al., 2020) phylogeny of ASVs within symbiont genera. Node labels indicate ultra-fast bootstrap support values. This is a subset of the data presented in Breusing et al. (2022), with oligonucleotides renamed because they belong only to the three species targeted in this study.

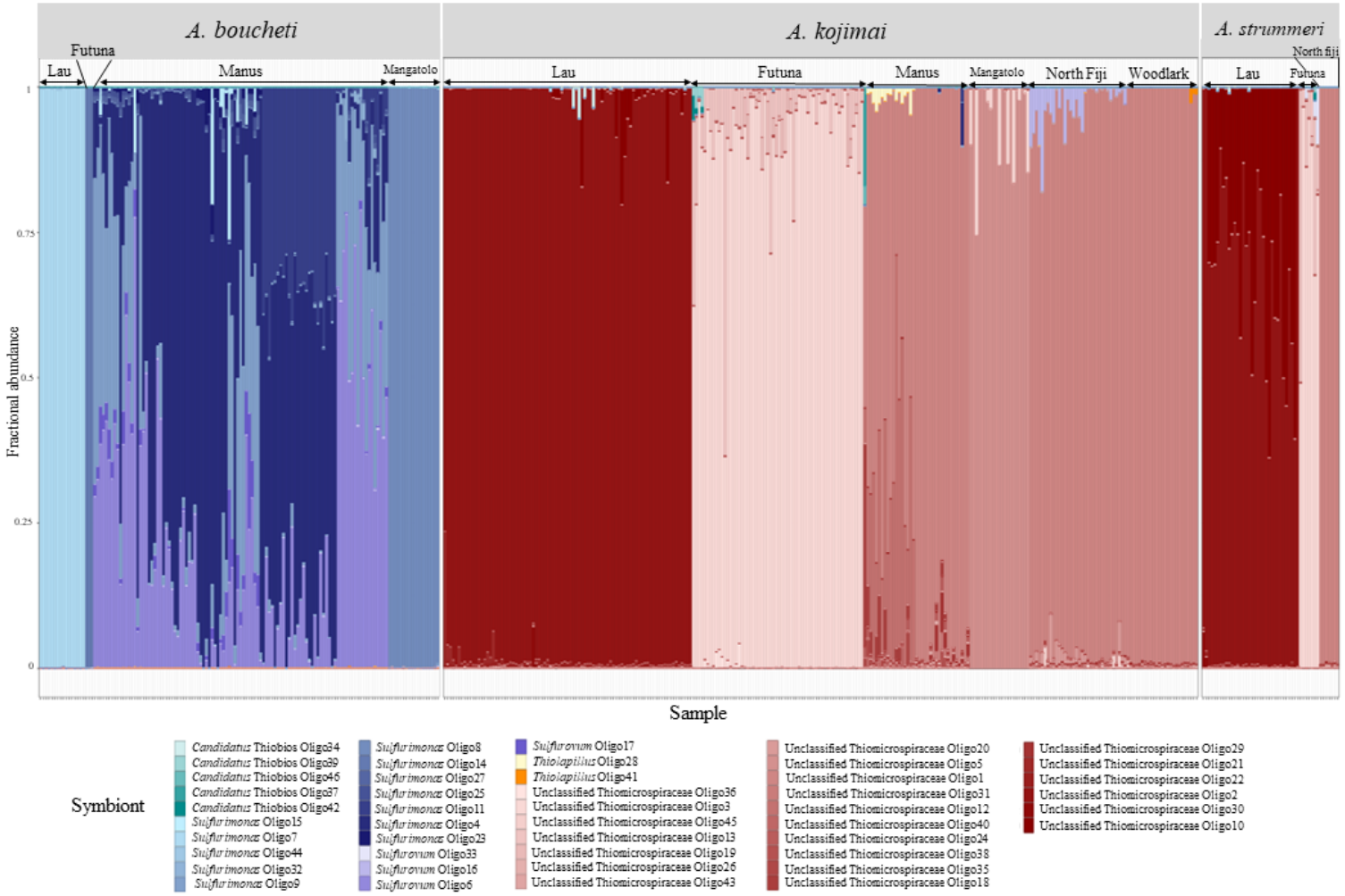


Figure 45: Fractional abundance plot of symbiont ASVs within individuals of *Alviniconcha*. The contrast of colours within a clade follows the genetic distance according to the phylogenetic tree presented in Figure 44.

3.3- Isotopes

Isotope analysis showed that the $\delta^{13}C$ was very variable according to the species, indeed, *A. kojimai* and *A. strummeri* had weak values (between -22 and -32‰) while *A. boucheti* displayed less negative values between -5 and -15‰. Note that whatever the species considered, the lowest values of $\delta^{13}C$ were observed in the Manus Basin and particularly at South_Su and Fenway. $\delta^{15}N$ ranged from 0 to 10‰ for *A. kojimai* and *A. strummeri* while *A. boucheti* had values between 3.5 and 6.5‰. As shown in Figure 46, each species had a very distinct isotopic signal between sites within the same basin. This is particularly the case within the Manus or the Futuna basins, where the $\delta^{13}C$ and $\delta^{15}N$ values clearly discriminate sites. The $\delta^{34}S$ analysis also showed some discrepancies for the three species of *Alviniconcha* with values < 10‰ and negative values between 0 and -10‰ for several sites in the Manus Basin (Figure 46). The variation in the isotopic composition of gastropods was dependent on sampling sites but did not follow any clear geographical pattern (i.e. not correlated with geographical distances of sites), and should be linked with different habitats within and between basins. A PCA was performed on the isotopic composition of all individuals of the three species (Figure 47) and showed that *A. boucheti* (purple envelope) clearly differs from the two other species (turquoise and yellow envelopes) by higher values of $\delta^{13}C$ and $\delta^{15}N$. In this graph, it was worth noting that the isotopic composition of species could differ between *A. kojimai* and *A. strummeri* within the same sample (biobox) at the Futuna vent field Fati-Ufu (arrows on Figure 47). There is thus a strong spatial heterogeneity of the chemical and/or microbial composition which contributes to a segregation between the sites whatever the species considered. This spatial heterogeneity could play a role in local adaptation.

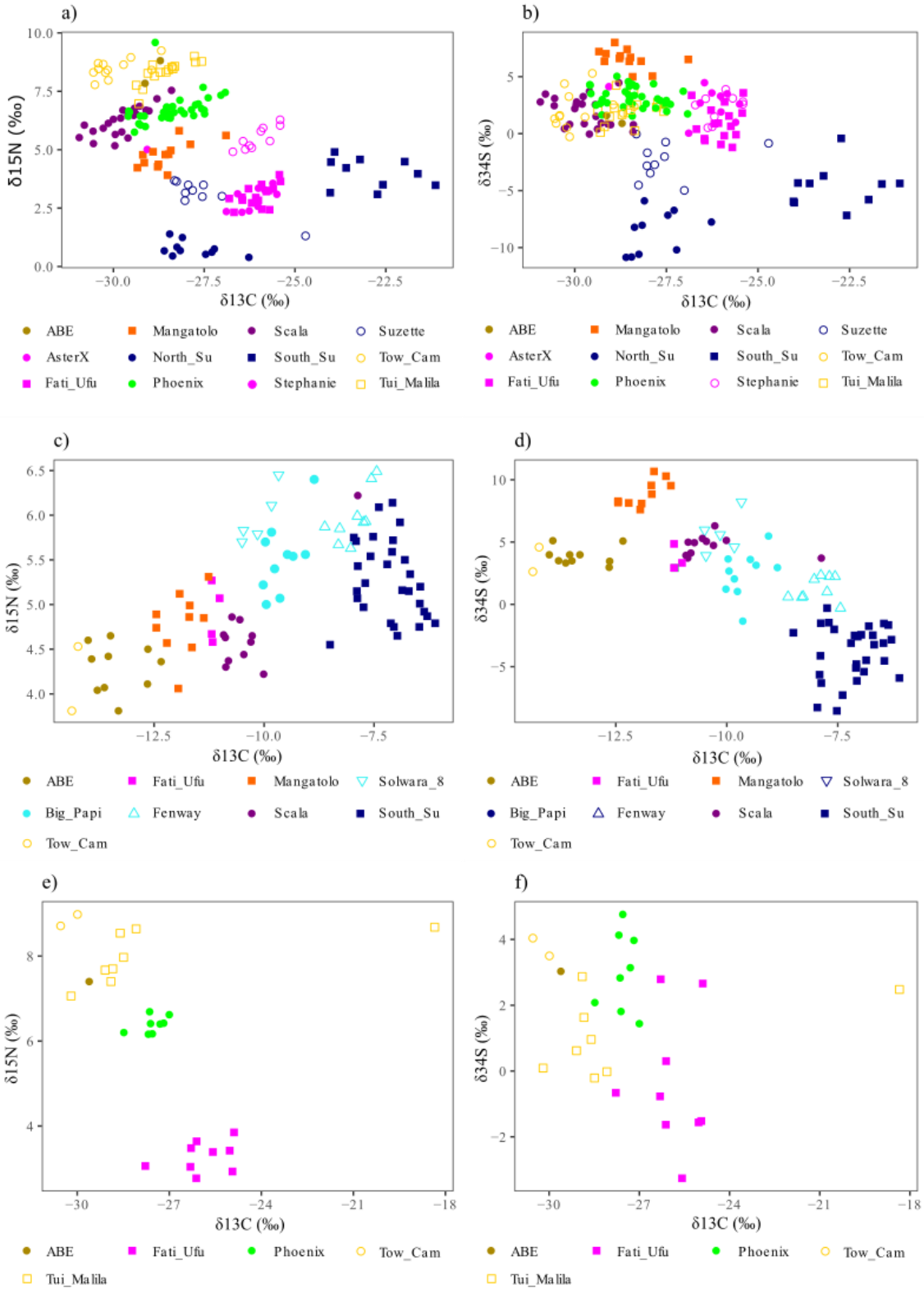


Figure 46: Stable isotope ratios of carbon vs. nitrogen (left) and carbon vs. sulphur (right) for (a-b) *A. kojimai*, (c-d) *A. boucheti* and (e-f) *A. strummeri* collected from five back-arc basins. Each coloured point represents an individual associated with its sampling site.

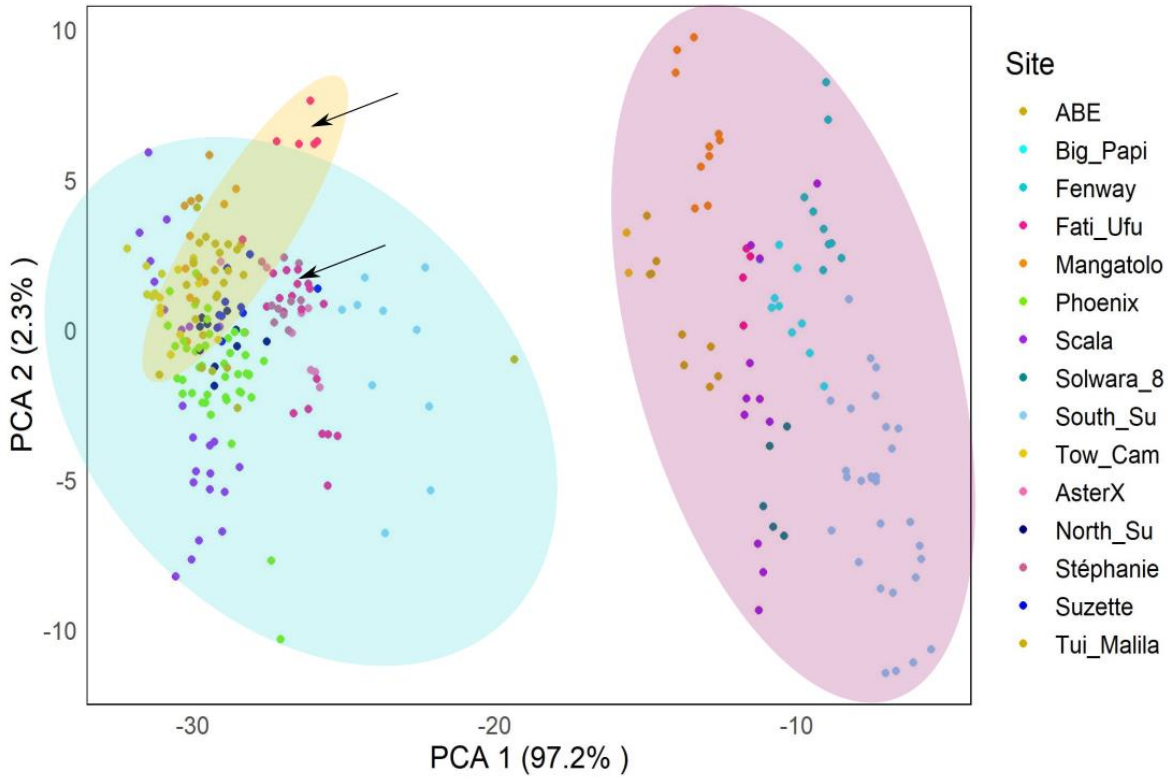


Figure 47: Principal components analysis of the isotopic composition of 287 *Alviniconcha* individuals. Each point corresponds to a single individual, and colours correspond to sampling sites (= 1 or 2 bioboxes). Purple colour groups individuals of *Alviniconcha boucheti* whereas the turquoise colour groups individuals of *A. kojimai* and the yellow one individuals of *A. strummeri*. Arrows underline the population of Fati-Ufu where isotopic composition is different between *A. strummeri* (top arrow) and *A. kojimai* (bottom arrow).

3.4- Fluid chemistry

The fluid chemistry measurements made on 22 samples are summarised in Table S8 and in Figure S11. In each species and for most chemical elements, strong variations between populations were observed, going from simple to double sometimes (as for Magnesium: 47.5-68.9 mM/kg, Lithium: 27.3- 82.6 µM/kg or Molybdenum: 41.9- 138.3 nM/kg) or even being able to vary more than ten times as for Barium: 248.2 – 55863.2 nM/kg, Lead: 2.6 – 18.4 µM/kg, Yttrium: 0.4 – 7.81 nM/kg, Zinc: 0.4 – 20.94 µM/kg, Manganese: 1.1 – 32.1 µM/kg, Iron: 0.5 – 257.9 µM/kg, Methane: 0.5 – 8.5 µM or Hydrogen sulphide: 0.3 – 95.6 µM. Other measures appeared to be more consistent between sites and species such as Vanadium (0.02 -0.06 µM/kg), Copper (0. 02 – 0.08 µM/kg) and Sulphur (24.5 – 33.3 mM/kg). These results showed some notable differences between species especially for iron, Manganese,

Aluminium (higher concentrations in *A. boucheti*) and Barium, Yttrium and Lead where the *A. kojimai* habitat exhibited the highest values.

On each species multifactor PCAs (Figure 48), where each point represents a sampled site characterised by all the chemical measurements made, we can see that microhabitats are partly distributed according to geography, as in the case of *A. boucheti* where all sites from the Manus Basin clustered together. Stronger linkage of fluid chemistry according to basin is usually expected as sites should share the same petrographic composition but this was not always the case (that is, geography did not seem to drive all the chemical variation among sites because of putative differences in how some fluids may precipitate partly in sub-surface). Indeed, in *A. kojimai*, the sample clustering did not respond to geography, and should therefore may serve as a proxy for local adaptation studies.

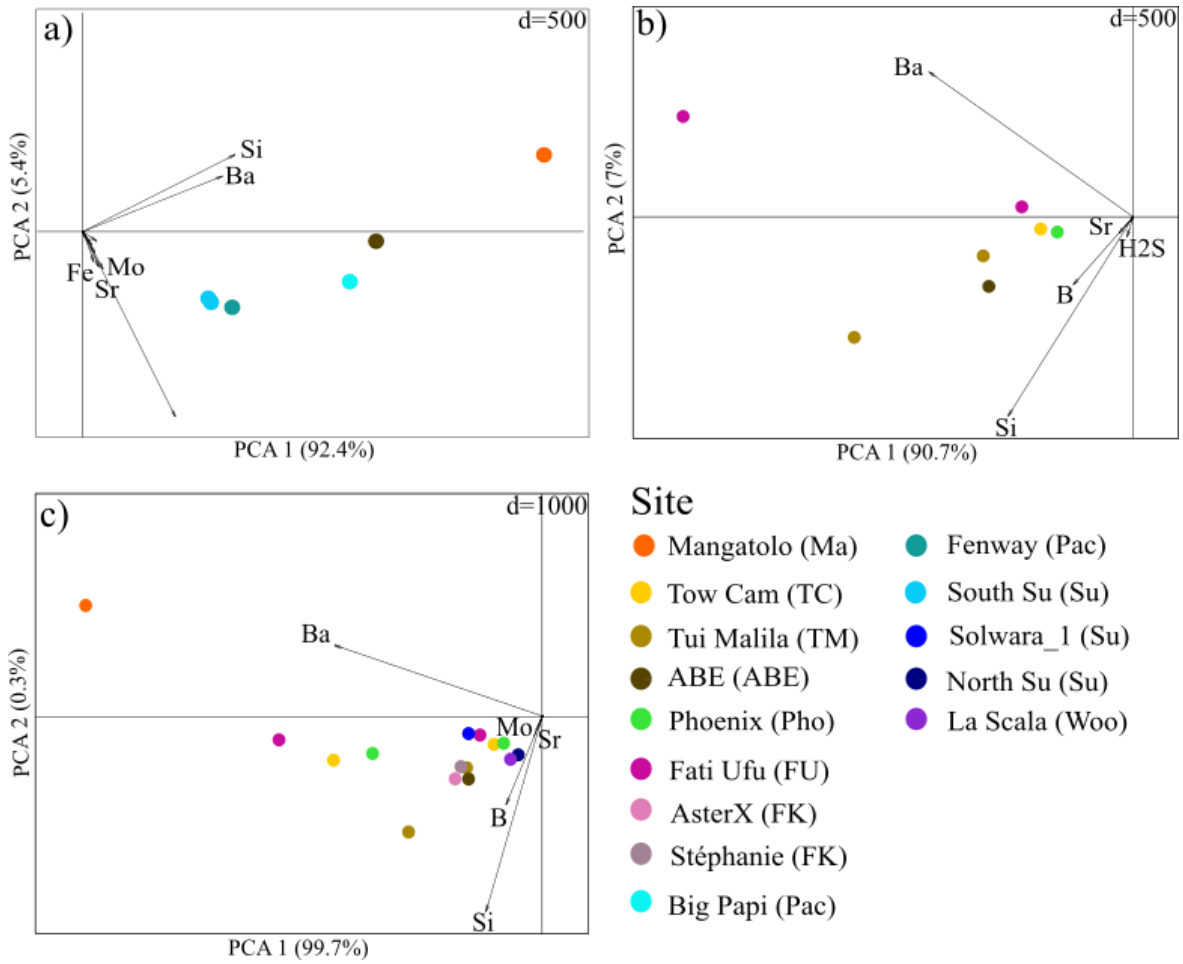


Figure 48: Principal components analysis of the chemical composition of the diluted fluid surrounding animals for (a) *A. boucheti* using 6 populations from 5 vent sites, (b) *A. strummeri* using 7 populations from 5 vent sites and (c) *A. kojimai* using 15 populations from 11 vent sites.

It should be noted that some chemical elements covaried in their geographical distribution (Figure 49). This is notably the case for Copper, Barium and Yttrium or Methane and Zinc. It is also the case of Calcium, Potassium, Silica, Lithium, Rubidium and Strontium.

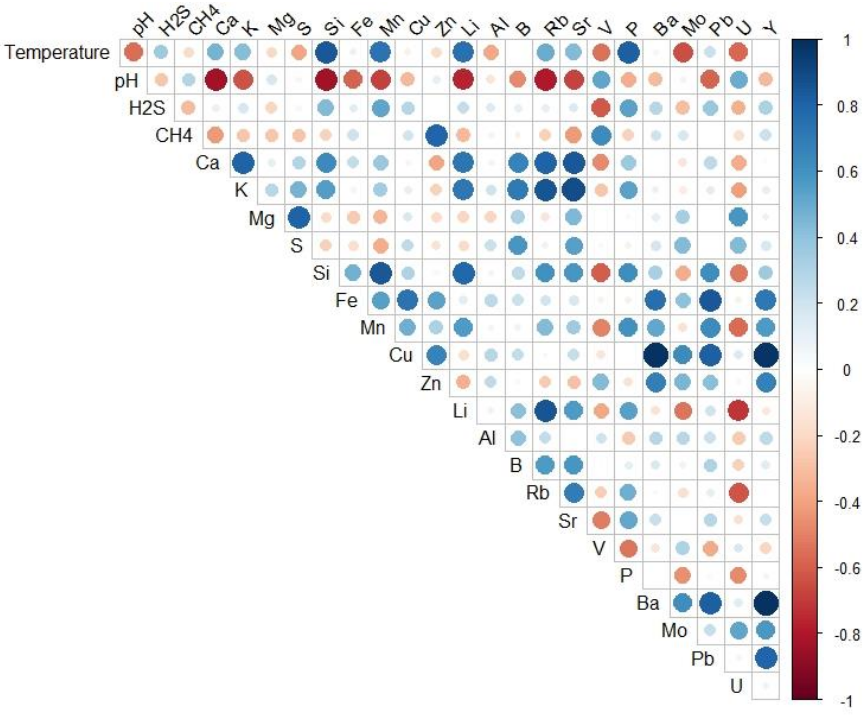


Figure 49: Correlation matrix between the chemical elements analysed in this study.

3.5- Genetic-Environment Association

As expected, the core model used to describe the neutral population structure (**Ω**) of each *Alviniconcha* species showed the same pattern as previously observed (Figure S12) with most sample differentiation between the Eastern (Lau/Futuna/Fiji) and Western (Manus/Woodlark) basins. In agreement with results from the PCA analyses shown above, the finer scale of differentiation observed in *A. strummeri* between Futuna/Lau populations and those from Fiji was also correctly recorded by Baypass's core model (Figure S12). This first step of analysis allowed us to identify SNPs with a strong XtX for each species, i.e. level of genetic differentiation stronger than that observed at most loci (outlier loci at a threshold of 5% - Figure 50). This represented 1889 SNPs in *A. kojimai* (962 locus), 503 in *A. strummeri* (331 locus) and 2885 SNPs in *A. boucheti* (2327 locus).

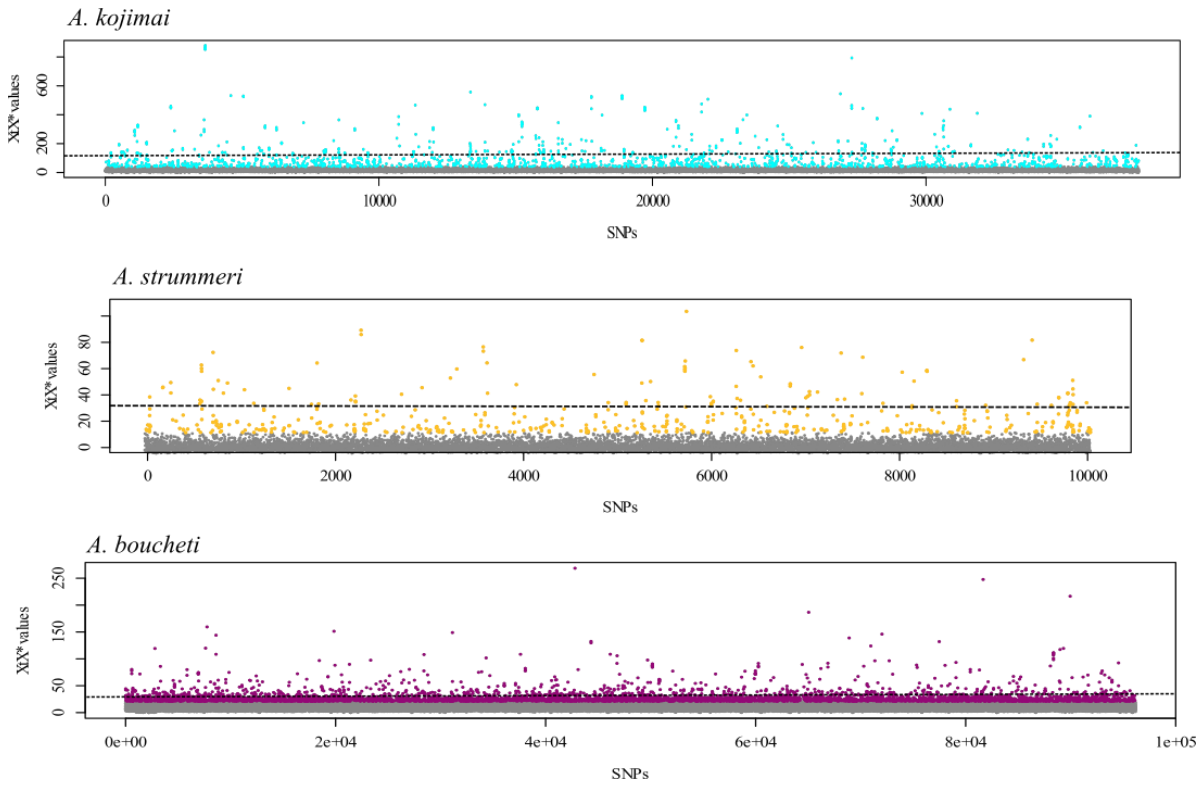


Figure 50: Distribution of the XtX values (monolocus overall sample differentiation) associated with single nucleotide polymorphisms (SNPs) in the three data sets. SNPs with the highest differentiation values (5% stronger) are coloured in turquoise for *A. kojimai*, yellow for *A. strummeri* and purple for *A. boucheti* when compared to the 95% of the other SNPs (in grey). For each graph, the dotted line represents the additional limit of the strongest XtX at the threshold of 1%.

We used BayPass Bayes factors (BF) to associate our 76 environmental (abiotic and biotic) factors with allele frequency differences. Figure 51 and 52 showed the Manhattan plots of the Bayes factors for all variables. SNPs with a "robust" environmental association, i.e. SNPs with both a BF>20 and high XtX values (1%) were then selected and correspond to SNPs coloured in turquoise (*A. kojimai*), yellow (*A. strummeri*) and purple (*A. boucheti*) on Figure 53.

Figure 51: Manhattan plots of SNP Bayes factors obtained using the outlier loci (1%) of the first analysis of different environmental variables. Coloured SNPs represent those with a Bayes Factor greater than 20 (i.e. SNPs whose frequency are reponding more to the environmental factor than the 'neutral' geographic backgound of differentiation).

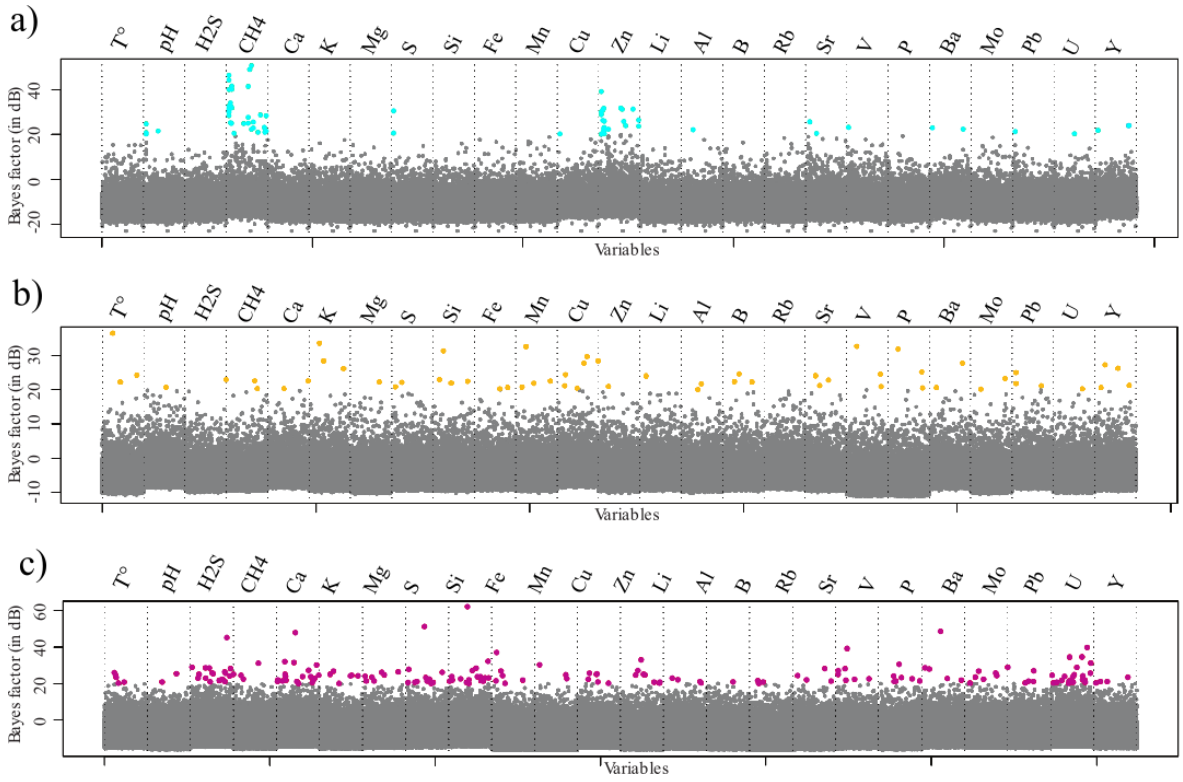


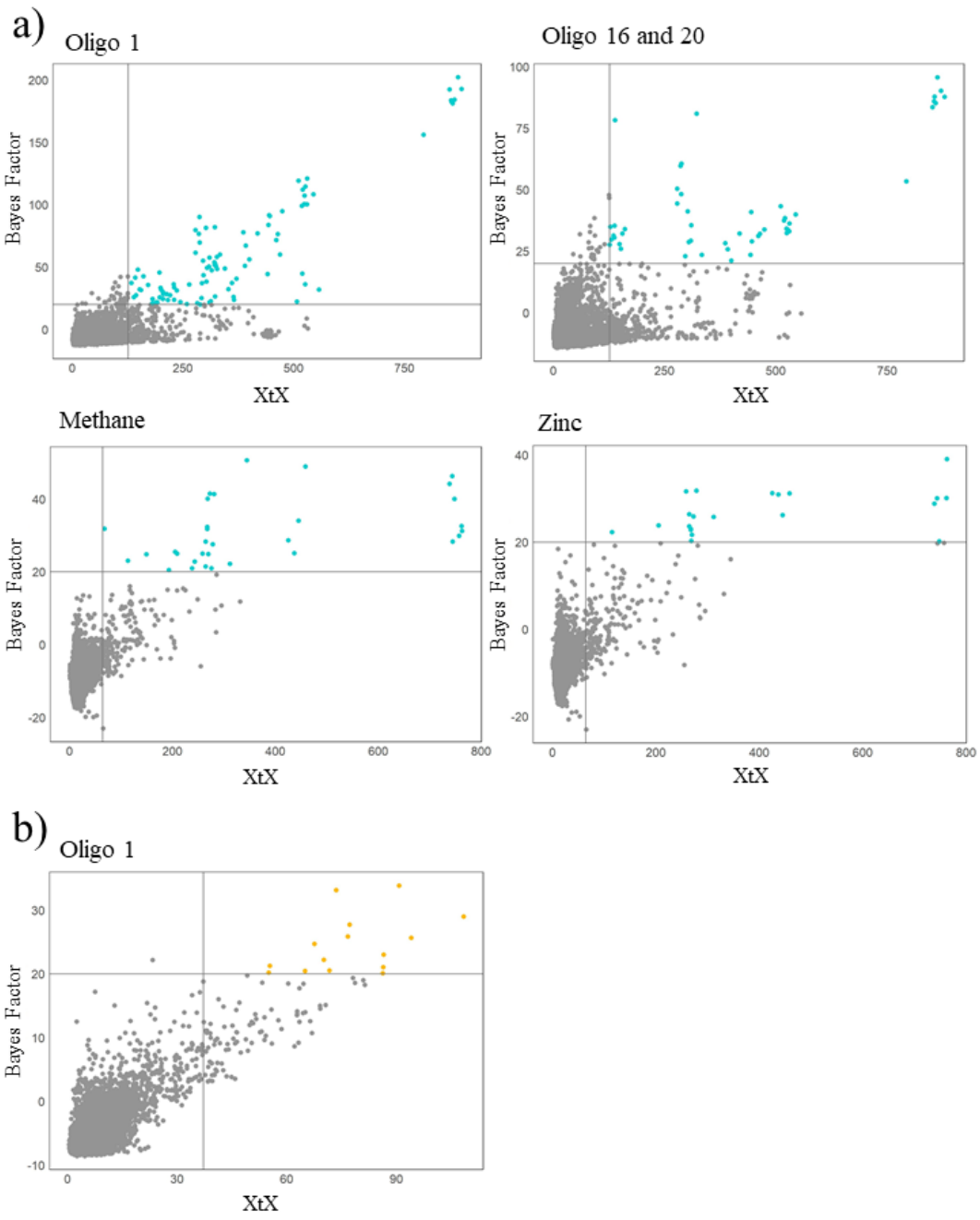
Figure 52: Manhattan plots of SNP Bayes factors obtained using the outlier loci (1%) of the second analysis on gaz and metallic elements (fluid chemistry). Coloured SNPs represent those with Bayes Factor greater than 20.

For *A. kojimai* (Figure 53a), different sets of SNPs had robust associations with Oligo 1 (99 SNPs - 45 RAD loci), and oligos 16 and 20 (50 SNPs - 19 RAD loci) as well as with the composition of the fluid in CH_4 (32 SNPs - 17 RAD loci) and zinc (21 SNPs - 11 RAD loci). Almost all the related-to-environment SNPs identified are shared between these five ecological variables. Oligos 1 and 20 are *Gamma-proteobacteria* while oligo 16 is a *Sulfurovum* (Figures 44). The bacterial strain associated with oligo 1 is present in the Manus, Woodlark and North Fiji basins and those of oligos 16 and 20 are only present in the North Fiji Basin. Among the RAD-loci showing a strong association with oligo 1 only two could be annotated and corresponded to activation of innate immune response and embryonic development of the host. Only one gene could be annotated from the association with oligos 16 and 20 and this gene encodes for animal organ morphogenesis. Concerning the fluid chemistry, concentrations of zinc and CH_4 were highly dependent on the basin in which they are measured. These two compounds have higher concentrations in the Manus and Woodlark basins while they are at low concentrations in the Lau and North Fiji basins and in the Futuna Arc. The annotation of SNPs associated with the fluid chemistry gave information for only

one gene, identical to the one already identified for oligo 1 and coding for innate immune response (stimulator of interferon genes protein-like - XP_0250802890).

For *A. strummeri* (Figure 53b), the signal of genetic-environment association was weaker with very few SNPs showing a strong Bayes factor for a given variable, except for oligo 1 (15 SNPs - 12 RAD-loci). The lack of significant hits in this later species was partly due to the fact that this species was rarer leading to a lower number of environmental replicates among basins. As shown for *A. kojimai*, the oligo1 bacterial strain is a *Gamma-proteobacteria* and is shared between these two species only in the North Fiji Basin (Figure 44). Unfortunately, none of the SNPs showing a significant level of association with oligo 1 could be annotated. No further association have been detected with chemical variables.

For *A. boucheti* (Figure 53c), strong environmental correlations were found with oligos 4 and 6 (82 SNPs - 58 RAD-loci) and with several abiotic factors (H_2S, 13 SNPs on 13 RAD-loci; Calcium, 13 SNPs on 13 RAD-loci; Silica, 10 SNPs on 10 RAD-loci and Uranium, 18 SNPs on 17 RAD-loci). As in *A. kojimai*, RAD-loci associated with these ecological variables were also shared between them. The bacterial strains associated with Oligos 4 and 6 are part of the *Sulfurimonas* and *Sulfurovum Campylobacteria* clades, respectively, and are both only present in the Manus Basin (Figures 44 and 45). These oligos were dominant in abundance in this basin. It should be noted that oligo 6 found in Manus was phylogenetically very close to oligo 16 and 33 identified in the North Fiji Basin and in Futuna in *A. kojimai* and *A. strummeri* respectively. Among the loci associated with these bacterial clades, six have been annotated. The first two have a function of oxidoreductase activity (electron transfer, ATP production), the four others encode for sensory perception of sound and light stimuli, protein localization to organelle, lipid transport and regulation of transcription by the RNA polymerase II. Associated fluid chemistry variables (H_2S, Calcium and Silica) all showed gradual decrease between the Lau and the Manus Basins, with the exception of Uranium which displayed its lowest values in the Lau Basin. Only two loci related to these chemical variables were annotated and correspond to protein localization to organelle and adult locomotory behaviour.


a)
Oligo 1
Bayes Factor
XtX
Oligo 16 and 20
Bayes Factor
XtX
Methane
Bayes Factor
XtX
Zinc
Bayes Factor
XtX
b)
Oligo 1
Bayes Factor
XtX

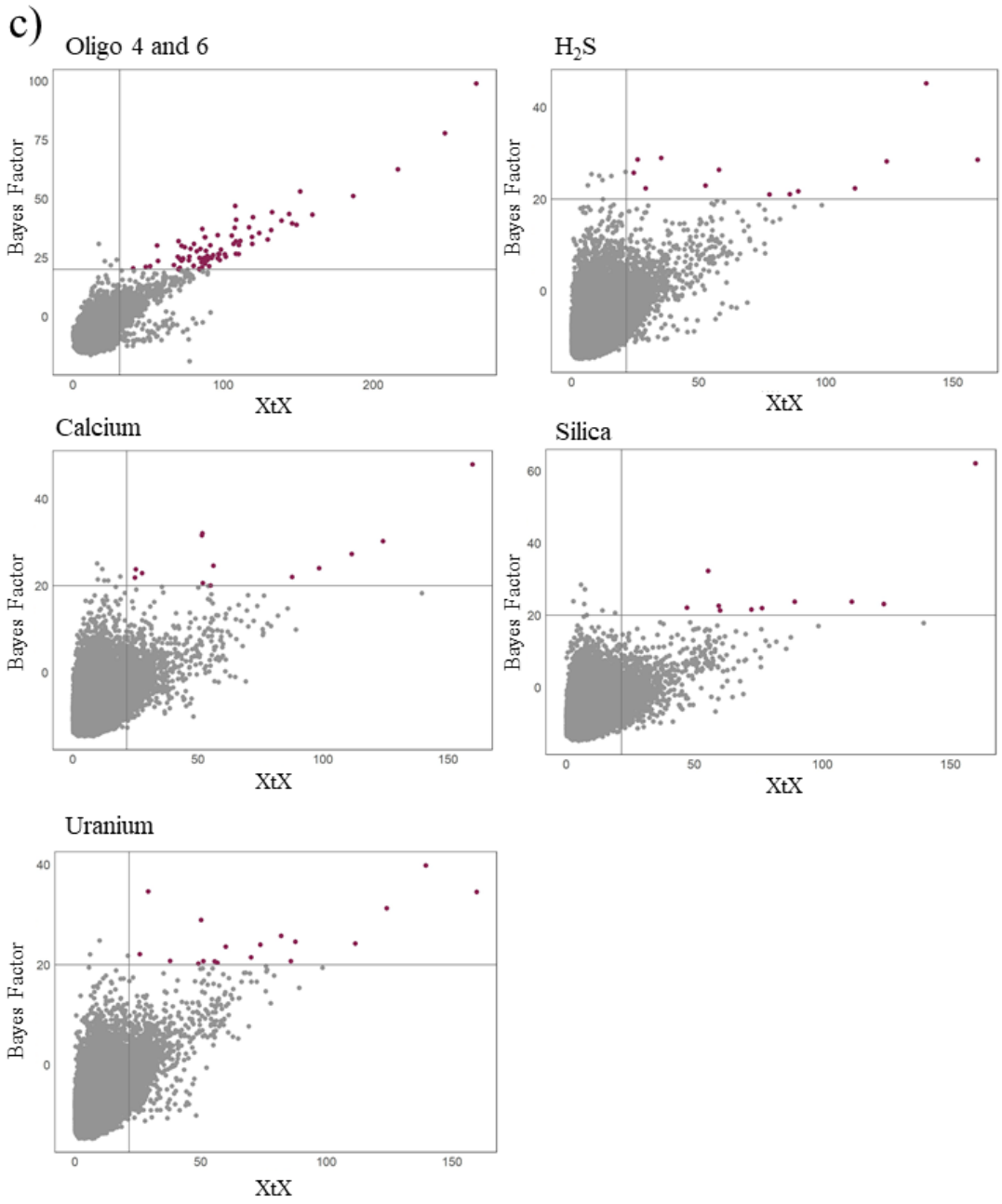


Figure 53: Biplot graphs for SNPs showing a significant level of association with the environmental factors that are the most explanatory. The X axis shows SNPs with the strongest differentiation between samples (XtX values with their 1% threshold: vertical bar) and the Y axis shows the distribution of SNPs according to their Bayes Factor (BF values with their threshold of 20: horizontal bar). The outlier SNPs (highest XtX at 1% with a BF greater than 20) are coloured in (a) turquoise for *A. kojimai*, (b) yellow for *A. strummeri* and (c) purple for *A. boucheti*.

4- Discussion

Alviniconcha gastropods are foundation species and ecosystem engineers at deep-sea hydrothermal vents of the Western Pacific. In the Southwest Pacific, three species (*A. kojimai*, *A. strummeri* and *A. boucheti*) are living in sympatry at vent fields of at least four distinct back-arc basins and the volcanic arc of Futuna but had a long history of separate evolution on different ridge systems. Previous molecular dating studies indeed proposed that *A. boucheti* separated from other species about 48 Mya and that the ancestral population of *A. kojimai* and *A. strummeri* splitted in allopatry 25 Mya (Breusing et al., 2020). It is currently recognized that these species do not exactly share the same ecological niche resulting from host-specific symbiotic composition (Beinart et al., 2012; Breusing et al., 2020). *A. boucheti* harbours symbiotic *Campylobacteria* in its gills and is often found in close association to the wall of vent chimneys whereas *A. kojimai* and *A. strummeri* are more likely associated with diffuse venting areas with lower H_2S and H_2 concentrations and exhibit *Gamma-proteobacteria* (Beinart et al., 2012, 2019). In addition, each of this host species possess variations in the symbiotic composition depending on the sites and in particular along the Lau Basin (Beinart et al., 2012). In this previous study, it also seems that some individuals have non-specific mixtures of symbionts (presence of *Campylobacteria* in *A. kojimai*; Beinart et al., 2012). This suggests the possibility of local adaptation in *Alviniconcha* in the horizontal acquisition of symbionts and questions the specificity of symbiotic associations in each species (differentially fixed or possible mixtures). Local adaptation between populations can only occur when there is variability in environmental conditions that results in genetic variability associated with these environmental differences. Thus, this study aimed to characterize both genetic and environmental variability in *Alviniconcha* species and then to look for traces of local adaptation.

4.1- A series of species with slightly different genetic structure at the scale of the Western Pacific

The three *Alviniconcha* species show remarkably weak genetic structure with average geographic differentiation of 0.003, 0.018 and 0.009 for *A. boucheti*, *A. strummeri* and *A. kojimai* respectively. The first two axes of multilocus PCA of individuals only explains 1.4 to 7.2% of the observed inter-individual variance with a rough genetic differentiation always lower than an F_{ST}: 0.022 between any pairs of basins. This is particularly true for *A. boucheti* where the most distant samples (4 000 km distance between Lau and Manus basins) have an

F_{ST} of only 0.003. This result suggests that the majority of variance is explained by inter-individual variability within local population. In comparison with *Ifremeria nautilei*, another gastropod found in the study area, the *Alviniconcha* populations differentiation (all species studied) is much weaker. *I. nautilei* display strongly differentiated metapopulations between the Eastern (Lau/Futuna/Fiji) and Western (Manus/Woodlark) basins with a F_{ST} of 0.387 (Tran Lu Y et al., 2022).

Ten times highest values of genetic differentiation were however found for *A. strummeri* sampled between populations of Lau/Futuna volcanic arc and the North Fiji Basin. It has been noted that this species seems to be present at a relatively lower abundance (Castel et al. submitted) than the two others. This may be associated with a lower effective population size, which, in turn, could explain the highest genetic differentiation found between the most geographically distant populations. Providing that the three species have the same ability to disperse, populations of smaller size may diverge more rapidly if drift is not totally counter-balance by migration (Kimura and Ohta, 1969). Alternatively, the higher small-scale differentiation of this species can be due to a lower dispersal capacity of the species with planktonic larvae being sent possibly in different water layers of the ocean. In other *Alviniconcha* species such as *A. hessleri* and *A. marisindica*, larvae are planktotrophic and able to reach the surface for dispersing (Kim et al., 2022; Sommer et al., 2017). This fits well with the amount of genetic differentiation estimated the two other species *A. kojimai* and *A. boucheti* with values almost close to zero suggesting that their larvae can cross very long distances between basins.

Differences however exist between these two later species with F_{ST} values three times lower for *A. boucheti* than for *A. kojimai*. Such difference is quite surprising because the two species have nearly the same species range (*A. kojimai* not found at PacManus) and presumably close census population sizes as their local abundances did not seem to be markedly different, although quantitative assessment of their relative abundances remain to be done (but see: Castel et al. submitted; Breusing et al., 2020; Johnson et al., 2015).

A. kojimai may have a lower effective population size (N_e) due to other factors such as demographic variations and episodes of low N_e, stronger variance in reproductive success among individuals (Eldon et al., 2016), different reproductive mode, biaised sex-ratio or overlapping generations (Frankham et al., 2002). As shown in Chapter 2, *A. kojimai* and *A. boucheti* have great differences in their reproductive mode. Indeed, *A. boucheti* is a pure

gonochoristic species whereas *A. kojimai* is an androdioecious species (presenting both male and hermaphroditic individuals). These differences can affect both the sex ratio if the male function is active in hermaphrodite individuals as well as the reproductive success, which can explain a smaller effective size in *A. kojimai* and thus a more pronounced geographical structure.

In the context of habitat fragmentation, population connectivity should benefit from a "long-term larval duration" as it seems to be the case for *A. marisindica* whose larvae have been collected at the surface (Kim et al., 2022). Alternatively, when habitats, even fragmented are more evenly sparsed, species may produce larvae with a more restricted larval duration to colonise the environment from nearby populations following a stepping-stone model (Kimura and Weiss, 1964). This later way of dispersing has been also proposed for a series of vent fauna inhabiting mid-oceanic ridges (France et al., 1992; Jollivet et al., 1995; Matabos et al., 2008). This lack of population differentiation suggests that *A. boucheti* is dispersing better than *A. kojimai* and *A. strummeri* over long distances (maybe with a surface larval phase) or able to colonize more easily intermediate active sites along the volcanic arcs of New Hebrides or Solomon. Although this second hypothesis seems less likely given the very similar geographic distributions of the two species, the hydrothermal sites found in a volcanic caldera (Nifonea) of the New Hebrides arc of Vanuatu only sheltered *A. boucheti* individuals.

Even weak, the three *Alviniconcha* species present an East to West significant global genetic structure. Species indeed exhibit a structure with two differentiated metapopulations between the Eastern (Lau/North Fiji and Futuna) and the Western (Manus and Woodlark) regions of the Western Pacific more or less pronounced depending on the species considered. This geographic structure may not be ideal for GEA analyses by BayPass even though this software considers the geographic differentiation of populations in its neutral model. Indeed, genetic differentiation by habitat may be partly masked if there are some confounding effects between habitat and geography (Gautier, 2015). In the following sections we will examine how the environmental variation is partitioned and how this variation can promote local adaptation via genotype-environment associations.

4.2- Environmental conditions

In this study, based on the CHUBACRAC mission, different ecological factors both abiotic (fluid chemistry and habitat type) and biotic (symbiotic and isotopic composition of the hosts) have been investigated following a complete and standardized environmental

survey of the gastropod assemblages using *in situ* measurements and fluid sampling. Despite the fact that both temperature and depth usually play a strong role in the local adaptation of species (Hallock and Hansen, 1979; Sokolova and Pörtner, 2001), we did not observe any species-specific preferendum triggering the distribution of our three-gastropod species. However, differences in the habitat occupancy can be noted between the species when looking more closely at diffuse venting areas and vent chimneys where the mineral composition of the substratum changes. To this extent, *A. boucheti* is more frequently found on the heights of hydrothermal chimneys (10 of 15 bioboxes) whereas *A. strummeri* and *A. kojimai* are more frequently found in areas of diffuse venting (6 of 8 and 12 of 19 bioboxes respectively).

Chemical conditions can usually vary greatly both spatially and temporally at deep sea hydrothermal vents along oceanic ridges, and even at small spatial scale within a vent field (Chavagnac et al., 2018; Le Bris et al., 2003; Matabos et al., 2008). Conversely, the chemical environment of *Alviniconcha* was quite similar between species (Table S8) despite the great variations of metallic elements found in end-member fluids between back-arc basins vent chimneys (Boulart et al., 2022; Fouquet et al., 1991; Konn et al., 2016). However, it should be noted that values of Mn, Fe and CH_4 were greater on the chimney wall of Big-Papi (Manus Basin) where *A. boucheti* was only present. These results of homogeneity of environmental conditions between *Aliniconcha* species had already been shown by the study of Beinart et al. (2012) where the values of temperature, oxygen and sulfide did not show differences between *Alviniconcha* species.

In this previous study, they also showed that holobionts with *Campylobacteria* symbionts were dominant in hydrothermal sites with higher concentrations of H_2 and H_2S and conversely, holobionts with *Gamma-proteobacterial* symbionts were more abundant in sites with lower in H_2 and H_2S concentrations. Thus, it appears that there is a close relationship between the chemical environment and symbionts in *Alviniconcha* individuals. However, it seems that the symbiosis, even if affected by the chemical environment, remains plastic. This can be shown by the presence of covariant chemical factors as shown in this study for example between methane and zinc demonstrating a tolerance to a wide range of chemical environments by symbionts. Therefore, in this study, we sought to characterize these factors to better understand local adaptation.

As previously shown, symbiosic asssociations are quite different between *Alviniconcha* species, and more specifically with *A. boucheti* which harbour *Campylobacteria* instead of

Gamma-proteobacteria when compared with *A. kojimai* and *A. strummeri* (Beinart et al., 2012, 2014, 2019; Breusing et al., 2020, 2022). In contrast to the previous findings of Beinart et al. (2012), our species are more likely to display specific symbiotic associations with very rare symbiont admixtures at least between *Campylobacteria* and *Gamma-proteobacteria*. These strong symbiotic species differences were reflected on the isotopic composition that was very different between *A. kojimai/A. strummeri* and *A. boucheti*, especially due to the much higher $\delta^{13}C$ values in *A. boucheti* because of its specific association with *Campylobacteria*.

On an intra-specific scale, local variations of temperature and depth differences occur between populations of each species. Indeed, depth greatly vary from 1 218 m in the Manus Basin to 3 388 m on the Woodlark ridge and temperatures could fluctuate between 7 to 23°C both spatially and temporally. It was therefore interesting to study the impact of depth and temperature on the local adaptation of populations, as they can lead to very strong constraints for the maintenance of protein stability in populations present at great depths and/or high temperatures. Although chemical conditions appear to be very similar between sites, they can vary slightly and even within a given site (i.e. between different gastropod aggregations: biobox samples). In *A. boucheti*, these variations are related to geography (with a clear chemical gradient between East and West) where Mangatolo and ABE populations (Lau Basin) are characterised by high concentrations in silica and barium (747.9-1219.8 µM and 557.4-1068.9 nM/kg, respectively: Figure 48). This is not always the case in *A. kojimai* and *A. strummeri* where geography does not fully explain differences between the chemical microenvironment of gastropods (Figure 48). Indeed, Mangatolo displayed a chemical signature that deeply differ from the other populations for which the sites Phoenix, Tow-Cam, Fati-Ufu, La Scala and Solwara_1 are very similar. Despite the small chemical variation observed between sites, the isotopic data shows us a completely different signal.

Indeed, the isotopic composition changed with the sampled vent localities, each gastropod sample (biobox) having its own isotopic signature with very low variations of the carbon, nitrogen and sulfur ratios between the sampled individuals. Such findings are likely due to distinct geographic micro-habitats as the symbiotic composition of each species is almost similar across their geographic range (*Gamma-proteobacteria* for *A. kojimai* and *A. strummeri* and *Campylobacteria* for *A. boucheti*), indicating that the compartmentalization of microhabitats according to geography seems to be more a result of environmental chemistry. As shown with PCAs on the environmental variables, the grouping of micro-habitats however

does not depend on the geographical distance, but the fluid chemistry and thus mainly to the composition of the underlying rocks. Surprisingly, the isotopic composition of gastropods suggest that each assemblage may face distinct micro-habitats both in terms of fluid chemistry and microbial communities. This may be due to the acquisition of locally adapted bacterial strains if the hosts are plastic enough to tolerate a large number of distinct ecological niches or able to diversify in locally adapted ecotypes as a by-product of the host-symbiont co-evolution. To this extent it should be noted that isotopic ratios indeed reflected the bacterial composition of the hosts in populations of *A. kojimai* and *A. strummeri* which exhibited admixtures between *Gamma-proteobacteria* and *Sulfurovum* (this is particularly the case at Fati-Ufu (Futuna)). This discrepancy between chemical and isotopic data may be due to chemical measurements taken at a specific time and therefore does not represent an integrative measure of the environment as is the case with isotopic measurements (Michel et al., 2016). Thus, an integrative measure would also be a better representation of the real habitat of individuals. Isotopic signatures reflect differences in local (fluid) conditions but also probably some differences in symbiotic associations, which still vary somewhat between some sites in some species.

Despite the presence of a single dominant bacterial strain in a given species, additional strains may also be found in some gastropod individuals at a low frequency. Interestingly, *A. kojimai* and *A. strummeri* were able to share a small proportion of *Sulfurovum Campylobacteria* at one location which are usually characteristic of *A. boucheti*. This clade was found mixed (up to 18% of whole symbiont content) with *Gamma-proteobacteria* in 35 *A. kojimai* individuals from the North Fiji Basin, and concerned 4 *A. strummeri* individuals from Futuna (in Fati-Ufu precisely). This mixture of bacterial types was previously noted for six individuals of *A. kojimai* from the Lau Basin in a previous study by Beinart et al. (2012). However, this was the first time it was been observed in *A. strummeri* (Breusing et al., 2022). This observation is very interesting because it means that the immune system of the two species can tolerate *Campylobacteria* to some extent, but at some specific geographic locations. This bacterial mixing are likely to be triggered by environmental factors as *Gamma-proteobacteria* and *Campylobacteria* rely on a very different inorganic carbon fixation pathways and both acquired environmentally (Beinart et al., 2019). *Gamma-proteobacteria* use the Calvin-Benson-Bassham (CBB) cycle while *Campylobacteria* use the tricarboxylic acid (rTCA). Paradoxally, although *A. kojimai* and *A. strummeri* were both present in syntopia (same biobox) at Fati-Ufu (Futuna) and Phoenix (North Fiji) only one of them acquired

Sulfurovum in its gills. Two hypotheses could be evoked to explain why both species did not exhibit the *Campylo*bacteria at the same time. The first explanation is that the two symbiotic species are in competition for the food resources and thus space (physical position of individuals around vents). Because bacteria are acquired horizontally from the surrounding environment (Beinart et al., 2012; Breusing et al., 2020) and possibly live under well-segregated microhabitats, the two *Alviniconcha* species locally acquired one or another type of symbionts during their settlement phase. The symbiont acquisition process may thus rely on the recruitment timing of both species over a newly formed emission and the resultant competition for space with *A. strummeri* preferentially found at the periphery of the patches (Castel et al., submitted). A second alternative hypothesis may also be that each of the two species should be locally adapted to allow and control the bacterial infestation of their gills. To this extent, a prerequisite will be that the gastropod post larvae should be pre-adapted to receive foreign symbionts with adaptive variants emerging locally and not spreading throughout the whole species range due to restricted gene flow between separated basins.

Another remarkable observation is that the clade *Thiolapillus*, previously known to be the main symbiont of *A. hessleri* located in the Mariana Trench (Breusing et al., 2022), was also detected in *A. kojimai* sampled from the Manus and Woodlark basins. This observation raises several questions about the biogeography of *Thiolapillus*, which may be not present in the other basins or the local tolerance of *A. kojimai* in Manus and Woodlark to *Thiolapillus* thank to the emergence of a pre-adapted immune system in Western metapopulation. This latter hypothesis could therefore be a remnant of the common history between *A. kojimai* and *A. hessleri*, which are siblings, or due to local adaptive introgression from *A. hessleri*, depending on whether these two allopatric species have been able to meet secondarily somewhere in the West. To verify these hypotheses, different studies could be conducted. One possibility would be to perform a metabarcoding study in the surrounding water of the populations in order to verify if *Thiolapillus* represents a cosmopolitan genus able to thrive on vents of all the back-arc basins depending locally on the fluid conditions. The second verification would be to perform demographic analyses of ancestry between genomic datasets of *A. kojimai* and *A. hessleri* to search for traces of introgression or shared polymorphisms.

4.3- Genome-environment analysis

The GEA analyses showed two different signals, the first is the presence of outliers related to local differences in symbiotic association. This is the case in *A. kojimai* where 50

SNPs (19 RAD loci) are associated with Oligo 16 and 20 which are symbionts only found in the North Fiji Basin. This finding is very interesting because Oligo 16 is not a *Gamma-Proteobacteria* but a *Sulfurovum Campylobacteria* usually diagnostic of *A. boucheti* symbiosis. Due to the presence of genetically differentiated SNPs in this geographic population, these loci may be the result of coevolutionary processes between the host and its symbionts on targeted genes involved in the defense/control of *Campylobacteria* strains during the symbiont acquisition phase or simply a geographic coincidence with the overall differentiation of the host population over its range. Either way, this observation may represent the very first insight of a population (that of the Northern Basin of Fiji) of *A. kojimai* locally adapted to *Sulfurovum*. The same observation is found in *A. strummeri* where 15 SNPs stand out associated with Oligo 1 which is present in this species only in the North Fiji Basin. This bacterial clade seems to be present in the Manus/Woodlark and North Fiji Basins (Figure 44), but *A. strummeri* is only present in the North Fiji Basin, which could explain this signal. We cannot therefore deduce whether the populations of *A. strummeri* are locally adapted to this symbiont because it is not present at the other sites where this species is present. Note that for this species, the power of the analysis is very low because we have few populations and few individuals per population.

The second signal is the presence of SNPs outliers linked to environmental variables along an East-West geographic gradient. In *A. kojimai*, this signal involves Oligo 1, methane, and zinc, which co-vary together and should reflect a potential adaptive signal between the two major metapopulations in the Eastern region (Lau and North Fiji basins and the Futuna active area) and the Western region (Manus and Woodlark basins). Note that because genetic structure is weak between geographic populations, any SNPs showing very strong differentiation between these regions will appear to be associated with contrasting ecological variables between these regions. Thus, these environmental variables represent potential candidates for population adaptation between two contrasting geographically related hydrothermal systems, but caution is warranted. However, since the outlier loci are identical between the ecological variables and the annotation showed a gene encoding the immune system, we can speculate that this may be related to immune system adaptation between these large regions. In *A. boucheti*, this signal is manifested for a series of portential ecological variables (i.e. Oligo 4 and 6, H_2S, calcium, silica and uranium).

The relationship between ecological factors and the population structure of *Alviniconcha* species, even if in *A. kojimai*, symbiosis seems to play a role in local adaptation,

remain to be defined or confirmed. For this, on way would be to look at local adaptation at an individual rather than population scale by comparing genotypes to the microbial and isotopic composition of individuals. An RDA (Redundancy analysis: Capblancq et al., 2018) approach could be thus considered to avoid to average ecological variables at the sample scale as averaging could mask a signal at the individual scale. Clues of local adaptation in *A. kojimai* and *A. boucheti* were observed at a large geographic scale between their two main metapopulations. Despite the high dispersal capacity of the species as demonstrated by the low F_{ST} retrieved for the three species at a global scale, this slight genetic difference associated with the BayPass outliers may be however due to a less efficient demographic migration between these two geographic metapopulations (Tran Lu Y et al., 2022). Because larval dispersal seems to be high among distant populations, local adaptation tends to be swamped by immigration of maladapted individuals from outside populations, which is why our records on small-scale local adaptation are nearly null (Sanford and Kelly, 2011). Even if some ecological variables appear to be good candidates for local adaptation, we must be cautious because the number of populations analysed remains low.

This study made it possible to show that certain individuals of *A. kojimai* and *A. strummeri* were able very locally to host both *Gamma-proteobacteria* and *Sulfurovom* symbionts and that these capacities were perhaps linked to a local adaptation of their immune system (as seen with outlier SNPs associated with oligo 16 in *A. kojimai*). However, as *Sulfurovom* were mainly present in *A. boucheti*, and as this manuscript made it possible to show that the species of *Alviniconcha* were capable of introgressive hybridization (chapitre 1), the question of adaptive adaptation remains open. Indeed, it could be that the hybridization between the species allowed *A. kojimai* and *A. strummeri* to acquire the capacity to tolerate *Sulfurovom* at least locally. This hypothesis still remains to be demonstrated.

In this study, it was possible to show very similar chemical and thermal environments between the species despite dinstinct habitats (chimney or diffusion zone) and symbiotic compositions. With this finding, we cannot currently assess the impact of the environment on the speciation of *Alviniconcha* species. Indeed, the differences in habitats encountered between *A. boucheti* and the two other species could have been the consequence of specialization after the speciation of the species. However, the ecological niches being different between the species, this could play a current role on the maintenance of these despite the introgressive hybridization by limiting the encounters and therefore the mating between the species. Despite this assumption, the three species of *Alviniconcha* being highly

divergent (as explored in detail in chapter 1) and at the end of the speciation continuum (Roux et al., 2016) with d_A values around 2% on the nuclear genome, the exploration of effects of environmental variables and local adaptation on current ecological divergence and isolation are very difficult. In hydridizing species, studies of local adaptations are usually carried out on hybrid zones in order to assess the role of the environment and adaptation on the ecological and reproductive isolation between species (Rolán-Alvarez et al., 1997).

5- Conclusions

This study, which aimed to study the local adaptation of *Alviniconcha* species and the potential role of the environment on the maintenance of the species, made it possible to show that despite notable differences in the habitat and the symbiontic composition between the species, no Notable differences could only be observed on the chemical conditions. It has been confirmed that *A. boucheti*, present on the walls of hydrothermal chimneys, mainly contain bacteria belonging to the *Campylobacteria* clade, while *A. kojimai* and *A. strummeri*, present mainly on the diffusion zones, mainly harbor *Gamma-proteobacteria*. At present, it is very difficult to say whether these ecological differences could have initiated speciation between *Alviniconcha* species but could currently be a strong barrier against hybridization by limiting the presence of species in syntopia and therefore certainly limiting their reproduction.

This study also made it possible to show strong variations in chimical condiations via the isotopic composition between the populations of each of the species. These ecological variations correlated with genetic variations have made it possible to highlight two possible phenomena of local adaptation. The first at a very small scale in the North Fiji Basin with local adaptation to *Sulfurovom* in *A. kojimai* and the second at a large scale between the Eastern (Lau/Futuna/Fiji) and Western basins (Manus/Woodlark). This second signal is supported in *A. kojimai* and *A. boucheti* and made it possible to show cadidate ecological variables for this adaptation including Oligo1, zinc and methane in *A. kojimai* and Oligo 4 and 6, H_2S, calcium, silicium and uranium in *A. boucheti*. This local adaptation signal still needs to be confirmed because it can be distorted by the initial genetic structure of the populations.

Supplementary material of Chapter 3:

Table S6: Specimens used for genetic-environmental association analysis

Sample	Site	Field	Basin	Ind. *A. kojimai*	Ind. *A. boucheti*	Ind. *A. strummeri*	Longitude	Latitude	depth (m)	habitat
721-GBT1	Tow Cam		Lau	21		1	176°08'15,4" W	20°19'04,4" S	2716	diffuse zone
721-GBT6	Tow Cam		Lau	19	3		176°08'15,8" W	20°19'05,1" S	2711	diffuse zone
721-GBT7	Tow Cam		Lau		19		176°08'12,7" W	20°18'59,2" S	2714	chimney
722-GBT7	Tui Malila		Lau	20		5	176°34'04,2" W	21°59'15,2" S	1899	diffuse zone
722-GBT1	Tui Malila		Lau	18		16	176°34'05,9" W	21°59'21,2" S	1886	diffuse zone
722-GBT5	Tui Malila		Lau	12		12	176°34'05,5" W	21°59'21,4" S	1884	diffuse zone
731-GBT3	ABE		Lau	2	20	1	176°11'28,9" W	20°45'47,1" S	2149	chimney
726-GBT4	Mangatolo		Lau		19		174°39'12,7" W	15°24'52,8" S	2031	diffuse zone
726-PBT6	Mangatolo		Lau	17			174°39'19,9" W	15°24'57,7" S	2039	diffuse zone
726-GBT2	Mangatolo		Lau	3			174°39'12,6" W	15°24'52,5" S	2031	diffuse zone
724-GBT4	Phoenix		North Fiji	18			173°55'7,6" E	16°57'0,0" S	1961	diffuse zone
724-PBT4	Phoenix		North Fiji	16		7	173°55'4,7" E	16°56'57,8" S	1973	diffuse zone
727-GBT2	AsterX	Fatu Kapa	Futuna	4			177°09'07,9" W	14°45'06,5" S	1562	diffuse zone
727-GBT4	Stephanie	Fatu Kapa	Futuna	15			177°09'57,6" W	14°44'14,7" S	1547	chimney
728-GBT2		Fati Ufu	Futuna	19		3	177°11'07,0" W	14°45'35,8" S	1519	chimney
728-PBT4		Fati Ufu	Futuna	19		1	177°11'04,9" W	14°45'35,3" S	1519	diffuse zone
728-GBT6		Fati Ufu	Futuna	3	4		177°11'05,9" W	14°45'35,2" S	1518	chimney
733-GBT2	Big Papi	Pacmanus	Manus		21		151°40'20,1" E	3°43'43,9" S	1708	chimney
733-GBT8	Fenway	Pacmanus	Manus		20		151°40'22,4" E	3°43'41,2" S	1696	diffuse zone
733-GBT9	Solwara8	Pacmanus	Manus		12		151°40'27,5" E	3°43'49,3" S	1737	chimney
733-PBT7	Solwara8	Pacmanus	Manus		5		151°40'26,6" E	3°43'50,1" S	1734	chimney
733-GBT11		Pacmanus	Maus		4		151°40'27,3" E	3°43'49,8" S	1739	chimney
734-GBT9		Pacmanus	Manus		16		152°6'2,8" E	3°43'17,2" S	1659	chimney

736-GBT3	Solwara1	Susu	Manus	18		152°5'47,0" E	3°47'22,1" S	1505	chimney
736-GBT10	North Su	Susu	Manus	12		152°6'2,8" E	3°47'56,0" S	1218	diffuse zone
737-GBT10	South Su	Susu	Manus		18	152°6'18,6" E	3°48'35,0" S	1353	diffuse zone
737-PBT5	South Su	Susu	Manus	16	12	152°6'17,5" E	3°48'29,8" S	1300	chimney
737-GBT7	South Su	Susu	Manus		21	152°6'17,9" E	3°48'31,8" S	1343	diffuse zone
738-GBT10	Scala		Woodlark	23		155°03'09,6" E	9°47'56,7" S	3388	chimney
739-PBT5	Scala		Woodlark		21	155°03'08,1" E	9°47'56,0" S	3353	chimney
	VateTrough	Nifonea	South-Fiji		6				

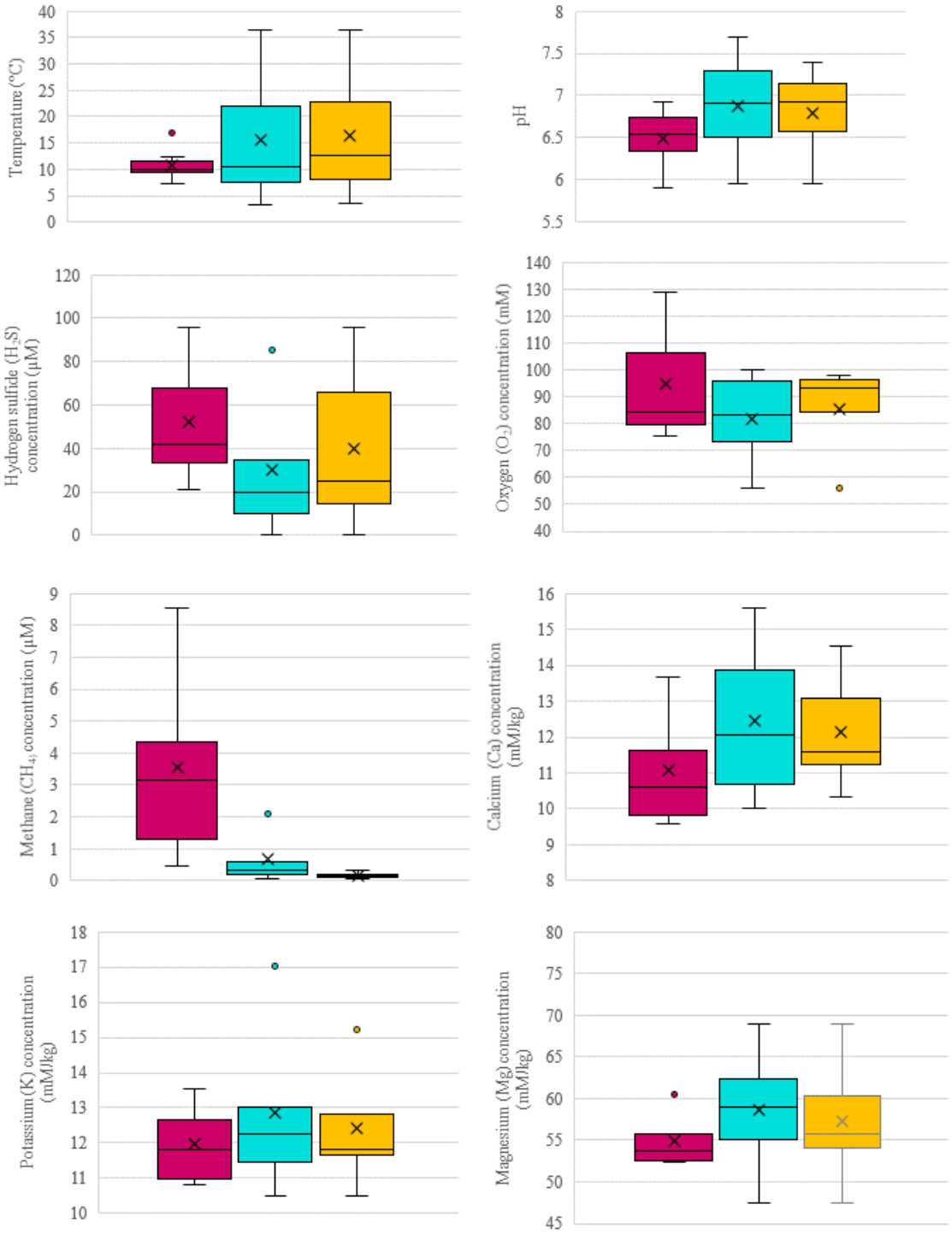
Table S7: List of the 76 environmental variables used in this study

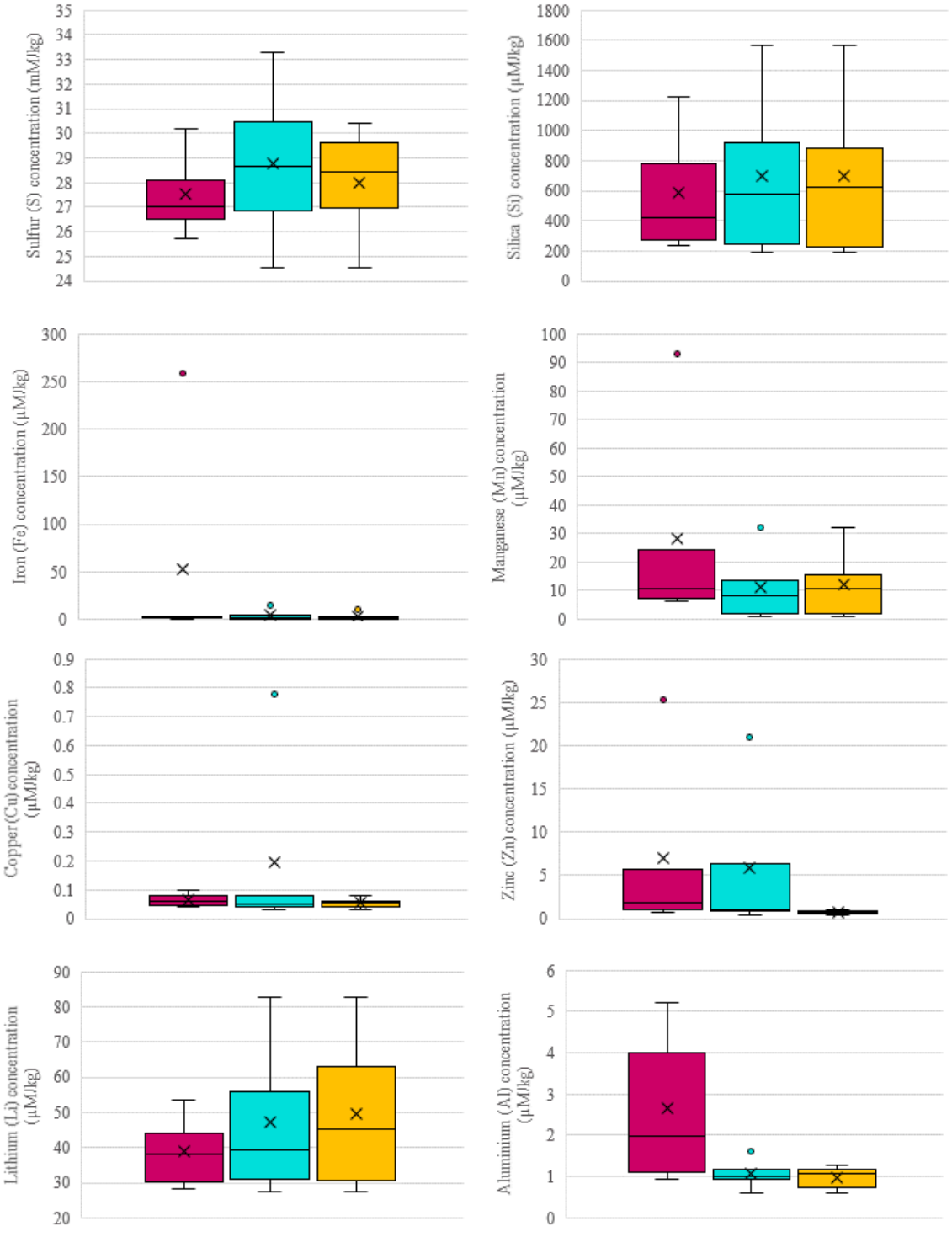
Variables	Description
Depth	Depth (m)
Habitat	Chimney or diffused zone
Temperature	Temperature (°C)
pH	pH
H_2S	Hydrogen sulfide (H_2S) concentration (µM)
CH_4	Mathane (CH_4) concentration (µM)
Ca	Calcium (Ca) concentration (mM/kg)
K	Potassium (K) concentration (mM/kg)
Mg	Magnesium (Mg) concentration (mM/kg)
S	Sulfur (S) concentration (mM/kg)
Si	Silica (Si) concentration (µM/kg)
Fe	Iron (Fe) concentration (µM/kg)
Mn	Manganese (Mn) concentration (µM/kg)
Cu	Copper (Cu) concentration (µM/kg)
Zn	Zinc (Zn) concentration (µM/kg)
Li	Lithium (Li) concentration (µM/kg)
Al	Aluminium (Al) concentration (µM/kg)
B	Boron (B) concentration (µM/kg)
Rb	Rubidium (Rb) concentration (µM/kg)
Sr	Strontium (Sr) concentration (µM/kg)
V	Vanadium (V) concentration (µM/kg)
P	Phosphorus (P) concentration (µM/kg)
Ba	Barium (Ba) concentration (nM/kg)
Cd	Cadmium (Cd) concentration (nM/kg)
Co	Cobalt (Co) concentration (nM/kg)
Pb	Lead (Pb) concentration (nM/kg)
U	Uranium (U) concentration (nM/kg)
Y	Yttrium (Y) concentration (nM/kg)
$\delta^{13}C$	stable carbon isotope
$\delta^{15}N$	stable isotope of nitrogen
$\delta^{34}S$	stable sulfur isotope
Oligo1	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo2	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo3	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo4	ASV of metabarcoding annotated as *Sulfurimonas*
Oligo5	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo6	ASV of metabarcoding annotated as *Sulfurovum*
Oligo7	ASV of metabarcoding annotated as *Sulfurimonas*
Oligo8	ASV of metabarcoding annotated as *Sulfurimonas*
Oligo9	ASV of metabarcoding annotated as *Sulfurimonas*
Oligo10	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo11	ASV of metabarcoding annotated as *Sulfurimonas*
Oligo12	ASV of metabarcoding annotated as *Gamma1-proteobacteria*

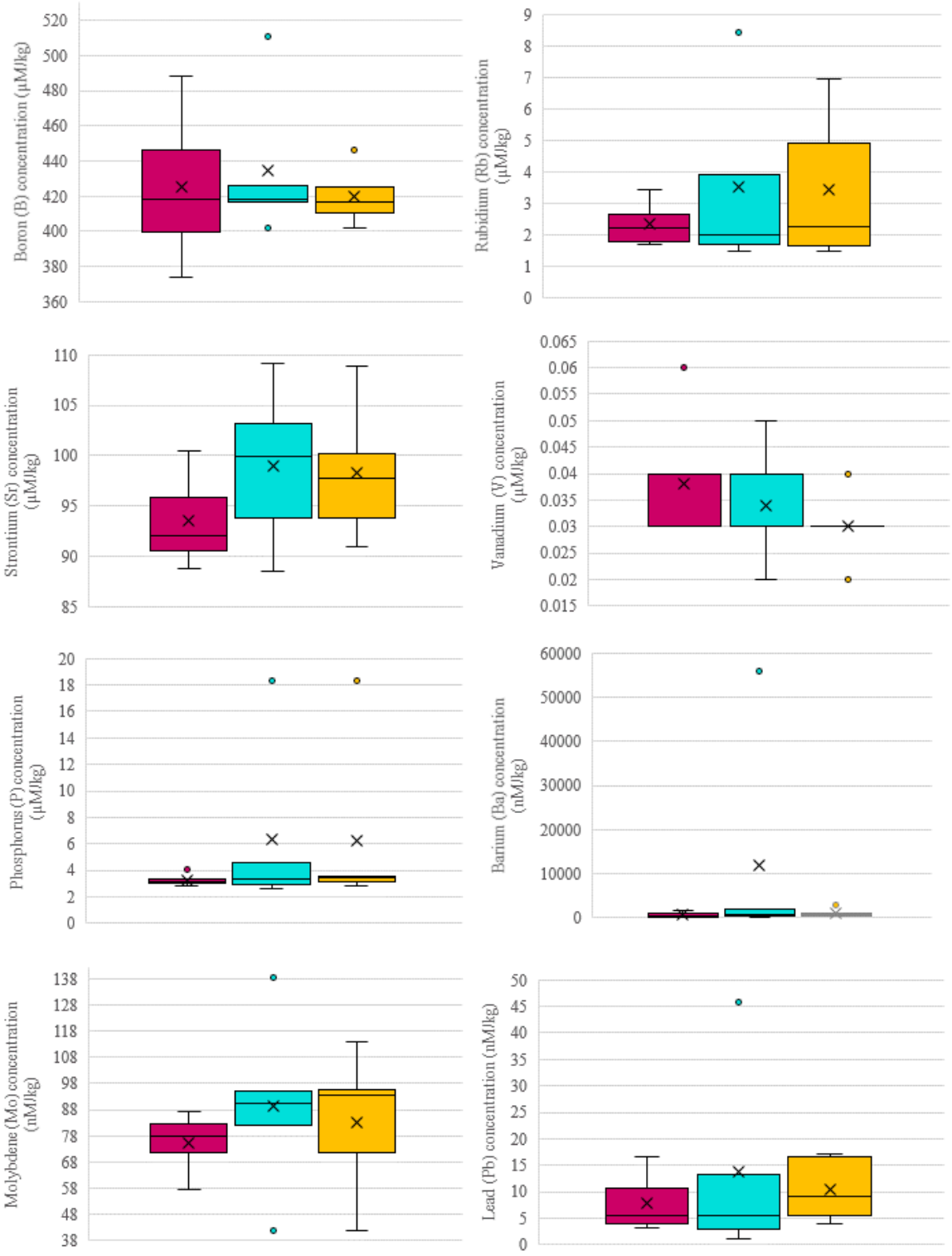
Oligo13	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo14	ASV of metabarcoding annotated as *Sulfurimonas*
Oligo15	ASV of metabarcoding annotated as *Sulfurimonas*
Oligo16	ASV of metabarcoding annotated as *Sulfurovum*
Oligo17	ASV of metabarcoding annotated as *Sulfurovum*
Oligo18	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo19	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo20	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo21	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo22	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo23	ASV of metabarcoding annotated as *Sulfurimonas*
Oligo24	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo25	ASV of metabarcoding annotated as *Sulfurimonas*
Oligo26	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo27	ASV of metabarcoding annotated as *Sulfurimonas*
Oligo28	ASV of metabarcoding annotated as *Thiolapillus*
Oligo29	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo30	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo31	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo32	ASV of metabarcoding annotated as *Sulfurimonas*
Oligo33	ASV of metabarcoding annotated as *Sulfurovum*
Oligo34	ASV of metabarcoding annotated as *GammaLau-proteobacteria*
Oligo35	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo36	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo37	ASV of metabarcoding annotated as *GammaLau-proteobacteria*
Oligo38	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo39	ASV of metabarcoding annotated as *GammaLau-proteobacteria*
Oligo40	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo41	ASV of metabarcoding annotated as *Thiolapillus*
Oligo42	ASV of metabarcoding annotated as *GammaLau-proteobacteria*
Oligo43	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo44	ASV of metabarcoding annotated as *Sulfurimonas*
Oligo45	ASV of metabarcoding annotated as *Gamma1-proteobacteria*
Oligo46	ASV of metabarcoding annotated as *GammaLau-proteobacteria*

Table S8: Ranges in temperature and chemical concentrations between different populations of each *Alviniconcha* species.

	A. kojimai	*A. boucheti*	*A. strummeri*
Temperature (°C)	3.4 - 36,4	7.19 - 12.7	3.6 - 36.4
pH (21°C)	5.95 - 7.7	5.9 - 6.93	5.95 - 7.39
H_2S (µM)	0.29 - 85.58	20.79 - 95.62	0.29 - 95.62
O_2 (mM)	55.8 - 100.17	75.5 - 128.94	55.8 - 98.01
CH_4 (µM)	0.06 - 2.1	0.45 - 8.54	0.06 - 0.32
Ca (mM/kg)	10 - 15.6	9.6 - 13.68	10.34 - 14.52
K (mM/kg)	10.51 - 17.01	10.83 - 13.53	10.51 - 15.22
Mg (mM/kg)	47.52 - 68.95	52.41 - 60.47	47.52 - 68.95
S (mM/kg)	24.55 - 33.26	25.75 - 30.18	24.55 - 30.42
Si (µM/kg)	193.81 - 1562.66	239.49 - 1219.81	193.81 - 1562.66
Fe (µM/kg)	0.18 - 14.14	0.46 - 257.90	0.26 - 10.13
Mn (µM/kg)	1.11 - 32.1	6.09 - 92.84	1.11 - 32.1
Cu (µM/kg)	0.02 - 0.78	0.04 - 0.1	0.03 - 0.08
Zn (µM/kg)	0.42 - 20.94	0.76 - 25.32	0.42 - 0.98
Li (µM/kg)	27.33 - 82.63	28.28 - 53.53	27.33 - 82.63
Al (µM/kg)	0.6 - 1.62	0.94 - 5.21	0.6 - 1.26
B (µM/kg)	401.82 - 450.78	374.26 - 488.56	401.82 - 446.43
Rb (µM/kg)	1.48 - 8.43	1.7 - 3.41	1.48 - 6.94
Sr (µM/kg)	88.5 - 109.2	88.78 - 100.45	91 - 108.89
V (µM/kg)	0.02 - 0.05	0.03 - 0.06	0.02 - 0.04
P (µM/kg)	2.6 - 18.36	2.83 - 4.1	2.79 - 18.36
Ba (nM/kg)	248.19 - 55863.21	157.36 - 1740.95	371.44 - 2783.27
Mo (nM/kg)	41.97 - 138.28	57.43 - 87.15	41.97 - 113.73
Pb (nM/kg)	1.16 - 45.8	3.11 - 16.63	4.05 - 17.19
U (nM/kg)	10.46 - 13.23	9.74 - 12.5	10.46 - 13.23
Y (nM/kg)	0.36 - 7.81	0.93 - 4.71	0.8 - 1.37


Temperature (°C)
pH
Hydrogen sulfide (H_2S) concentration (µM)
Oxygen (O_2) concentration (mM)
Methane (CH_4) concentration (µM)
Calcium (Ca) concentration (mM/kg)
Potassium (K) concentration (mM/kg)
Magnesium (Mg) concentration (mM/kg)


Sulfur (S) concentration (mM/kg)
35
34
33
32
31
30
29
28
27
26
25
24
Silica (Si) concentration (µM/kg)
1800
1600
1400
1200
1000
800
600
400
200
0
Iron (Fe) concentration (µM/kg)
300
250
200
150
100
50
0
Manganese (Mn) concentration (µM/kg)
100
90
80
70
60
50
40
30
20
10
0
Copper (Cu) concentration (µM/kg)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Zinc (Zn) concentration (µM/kg)
30
25
20
15
10
5
0
Lithium (Li) concentration (µM/kg)
90
80
70
60
50
40
30
20
Aluminium (Al) concentration (µM/kg)
6
5
4
3
2
1
0


Boron (B) concentration (µM/kg)
520
500
480
460
440
420
400
380
360
Rubidium (Rb) concentration (µM/kg)
9
8
7
6
5
4
3
2
1
0
Strontium (Sr) concentration (µM/kg)
110
105
100
95
90
85
Vanadium (V) concentration (µM/kg)
0.065
0.06
0.055
0.05
0.045
0.04
0.035
0.03
0.025
0.02
0.015
Phosphorus (P) concentration (µM/kg)
20
18
16
14
12
10
8
6
4
2
0
Barium (Ba) concentration (nM/kg)
60000
50000
40000
30000
20000
10000
0
Molybdene (Mo) concentration (nM/kg)
138
128
118
108
98
88
78
68
58
48
38
Lead (Pb) concentration (nM/kg)
50
45
40
35
30
25
20
15
10
5
0

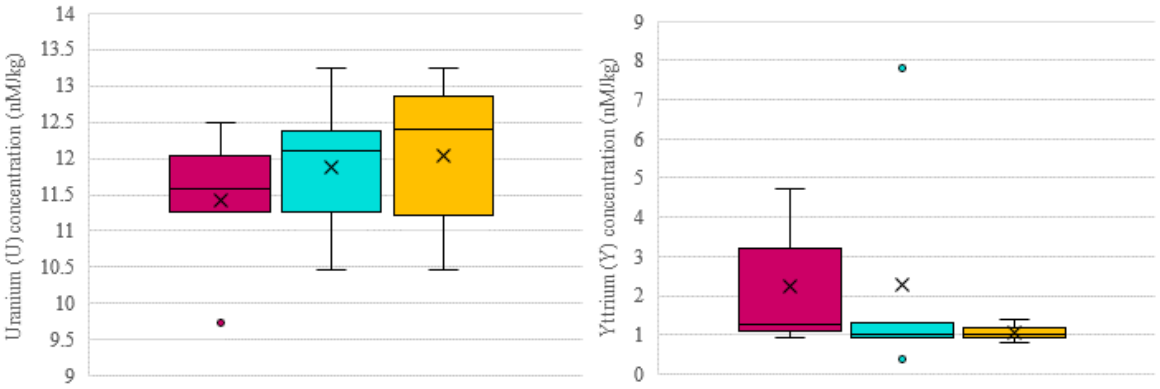
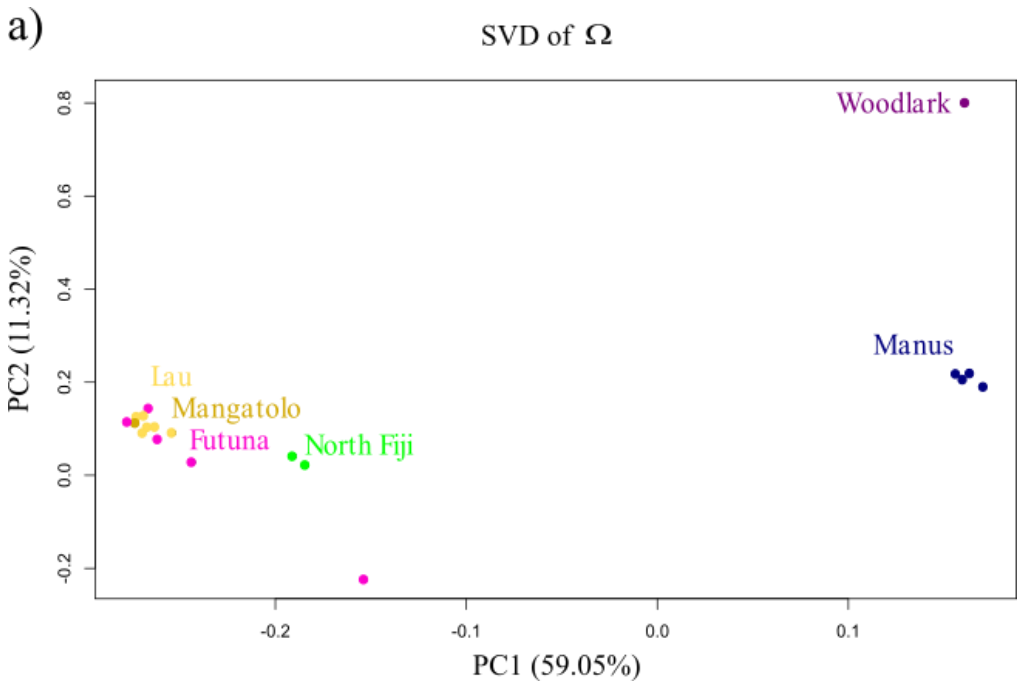


Figure S11: Intra-species distribution of environmental factors (temperature and fluid chemistry) made on gastropod assemblages during CHUBACARC 2019 expedition.



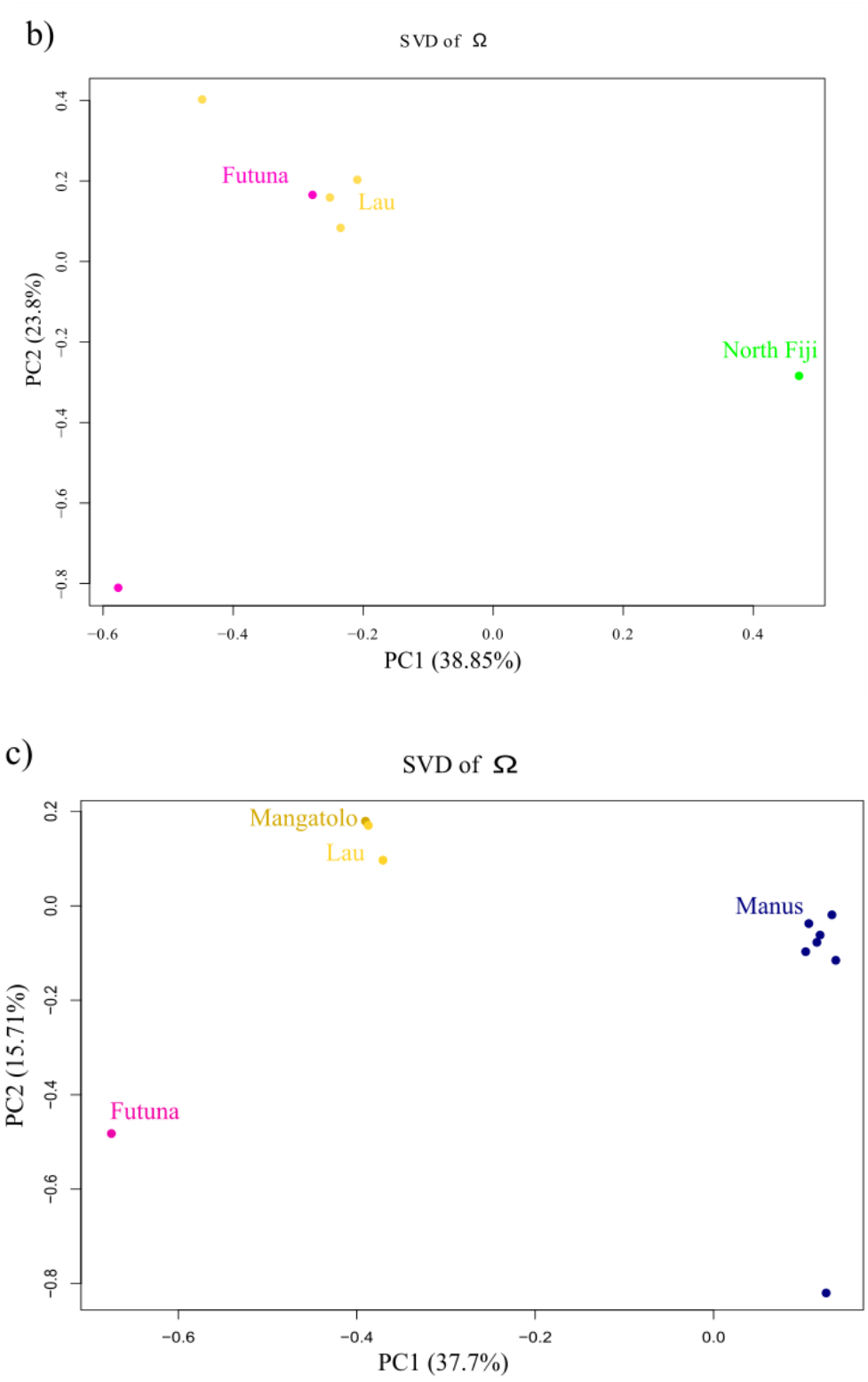


Figure S12: Neutral population structure (**Ω**) described by the core model of spatial differentiation of BayPass for (a) *A. kojimai*, (b) *A. strummeri* and (c) *A. boucheti*. Geographic basins are labelled in different colours: yellow = Lau, rose = Futuna, green = North Fiji, blue = Manus, and violet = Woodlark.

Appendix 3: Global 16S rRNA diversity of provannid snail endosymbionts fro Indo-Pacific deep-sea hydrothermal vents - collaborative article

environmental microbiology reports

Environmental Microbiology Reports (2022) **00**(00), 00–00 doi:10.1111/1758-2229.13051

Brief Report

Global 16S rRNA diversity of provannid snail endosymbionts from Indo-Pacific deep-sea hydrothermal vents

Corinna Breusing,[1] Jade Castel,[2] Yi Yang,[3] Thomas Broquet,[2] Jin Sun,[4] Didier Jollivet,[2] Pei-Yuan Qian[3] and Roxanne A. Beinart[1]*

[1]*Graduate School of Oceanography, University of Rhode Island, Narragansett, RI.*

[2]*CNRS UMR 7144 'Adaptation et Diversité en Milieux Marins' (AD2M), Team 'Dynamique de la Diversité Marine' (DyDiv), Station Biologique de Roscoff, Roscoff, France.*

[3]*Department of Ocean Science, Division of Life Science and Hong Kong Branch of the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), The Hong Kong University of Science and Technology, Hong Kong, China.*

[4]*Institute of Evolution & Marine Biodiversity, Ocean University of China, Qingdao, China.*

Summary

Symbioses between invertebrate animals and chemosynthetic bacteria build the foundation of deep-sea hydrothermal ecosystems worldwide. Despite the importance of these symbioses for ecosystem functioning, the diversity of symbionts within and between host organisms and geographic regions is still poorly understood. In this study we used 16S rRNA amplicon sequencing to determine the diversity of gill endosymbionts in provannid snails of the genera *Alviniconcha* and *Ifremeria*, which are key species at deep-sea hydrothermal vents in the Indo-Pacific Ocean. Our analysis of 761 snail samples across the distributional range of these species confirms previous findings that symbiont lineages are strongly partitioned by host species and broad-scale geography. Less structuring was observed within geographic regions, probably due to insufficient strain resolution of the 16S rRNA gene. Symbiont richness in individual hosts appeared to be unrelated to host size, suggesting that provannid snails might acquire their symbionts only during a permissive time window in early developmental stages in contrast to other vent molluscs that obtain their symbionts throughout their lifetime. Despite the extent of our dataset, symbiont accumulation curves did not reach saturation, highlighting the need for increased sampling efforts to uncover the full diversity of symbionts within these and other hydrothermal vent species.

Introduction

Microbial symbioses are increasingly recognized as universal phenomena that impact virtually all levels of biological organization, from cellular to organismal to ecosystem scale (Bronstein, 2015). Growing evidence from various symbiotic partnerships suggests that microbial symbioses can expand the physiological and ecological capabilities of hosts and symbionts, which are predicted to be critical for ecosystem productivity, stability and biogeochemical cycling (Apprill, 2017; Beinart, 2019; Wilkins *et al*., 2019). Deep-sea hydrothermal vents are probably some of the most enigmatic ecosystems that are sustained by microbial symbioses. In these systems, invertebrate animals live in association with chemoautotrophic bacteria that use chemical energy from venting fluids for the production of organic carbon, thereby providing food for their hosts (Dubilier *et al*., 2008; Sogin *et al*., 2021). Despite decades of research on this topic and the significance of chemosynthetic symbioses for ecosystem processes at hydrothermal vents, the diversity and distribution of symbionts within and across hosts and habitats remain underexplored, especially at large biogeographic scales.

Provannid snails of the sister genera *Alviniconcha* and *Ifremeria* are dominant animals in benthic communities at deep-sea hydrothermal vents in the Indian and Western Pacific Ocean (Van Dover *et al*., 2001; Desbruyères *et al*., 2006). While the Western Pacific genus *Ifremeria* is represented by a single species, *I. nautilei*, that affiliates with methane- and/or sulfide-oxidizing gammaproteobacterial

Received 8 October, 2021; accepted 4 February, 2022. *For correspondence. E-mail rbeinart@uri.edu.

symbionts (Windoffer and Giere, 1997; Borowski *et al*., 2002; Suzuki *et al*., 2006a), the genus *Alviniconcha* comprises five Western Pacific species (*A. adamantis*, *A. boucheti*, *A. hessleri*, *A. kojimai*, *A. strummeri*) and one Indian Ocean species (*A. marisindica*) that live in symbiosis with thiotrophic Gammaproteobacteria or Campylobacteria (Suzuki *et al*., 2006b; Johnson *et al*., 2015; Breusing *et al*., 2020). In both *Alviniconcha* and *Ifremeria*, the symbionts are assumed to be horizontally acquired and are harboured intracellularly within the host's gill tissue (Suzuki *et al*., 2006a, 2006b). Despite an environmental pathway for symbiont transmission, host and symbiont genera or species appear to exhibit a relatively strong selectivity in their partnerships towards each other (Beinart *et al*., 2012; Breusing *et al*., 2020), though host individuals are flexible in recruiting local strains of their specific symbiont phylotype(s) (Breusing *et al*., 2021a; Breusing *et al*., 2021b).

Most current analyses on the variation and structure of microbial symbionts within *Alviniconcha* and *Ifremeria* stem from studies in the Lau Back-Arc Basin, while little is known about these patterns in populations from other spreading systems within the distributional range of these genera. Here, we compiled an extensive dataset of 761 snail samples from 10 geographic regions of the Indo-Pacific Ocean (Fig. 1), some of which were previously unexplored, to assess the global diversity of chemosynthetic gill endosymbionts within *Alviniconcha* and *Ifremeria* through identification of 16S rRNA amplicon sequence variants (ASVs). Using ordination analyses and correlative statistics, we determined the influence of host species, host size, depth and geography on symbiont composition and distribution.

Results and discussion

Symbiont 16S rRNA diversity is partitioned by host species and geography

Our conservative analysis pipeline, which extends a previous study by Breusing *et al*. (2020) to now include seven species and 10 geographic areas, recovered 60 symbiont ASVs that were assigned to two campylobacterial (*Sulfurovum*, *Sulfurimonas*) and four gammaproteobacterial (*Ca*. Thiobios, *Methylomonas*, *Thiolapillus*, unclassified Thiomicrospiraceae) genera of provannid snail endosymbionts (Figs 2 and 3). Average pairwise identities within genera ranged from 95% to 99% (*Sulfurovum*: 95.4%; *Sulfurimonas*: 95.0%; *Ca*. Thiobios: 97.1%; *Methylomonas*: n.a.; *Thiolapillus*: 98.1%; unclassified Thiomicrospiraceae: 99.0%). In agreement with Breusing *et al*. (2020), ASVs were generally segregated by host species and broader geographic region (i.e. back-arc basin, volcanic arc or mid-ocean ridge), except for lineages within the unclassified Thiomicrospiraceae group which were shared between *A. kojimai* and *A. strummeri* (Fig. 2, 4A; Appendix 1: Fig. S1). Based on PERMANOVAs and linear decomposition models the impact of host species and geography superseded the influence of DNA preservation, extraction and sequencing method (81.17% vs. 1.99% explained variation) and was significant even when corrected for confounding technical effects. In addition, there was no evident clustering of samples by methodology in multidimensional scaling, indicating that the observed patterns are true biological signals (Table 1; Appendix 1: Fig. S2).

Like *A. kojimai* and *A. strummeri*, most other host species were associated with particular lineages of thiotrophic Gammaproteobacteria. *Alviniconcha adamantis* was affiliated with symbionts of the genus *Ca*. Thiobios, whereas *A. hessleri* and *I. nautilei* hosted distinct *Thiolapillus* symbiont ASVs. Many *I. nautilei* individuals further harboured a minority methanotrophic symbiont from the genus *Methylomonas*, especially at vent sites within the Eastern Lau Spreading Center (ELSC). Only *Alviniconcha boucheti* and *A. marisindica* were dominated by different region-specific campylobacterial ASVs of the genera *Sulfurimonas* or *Sulfurovum*.

Within geographic area, the gammaproteobacterial symbionts of *A. kojimai* and *A. hessleri* showed evidence

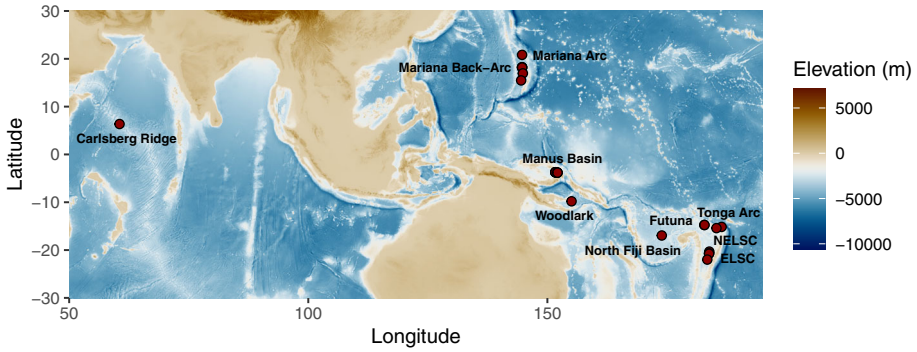


Fig. 1. Locations for *Alviniconcha* and *Ifremeria* species sampled in this study.

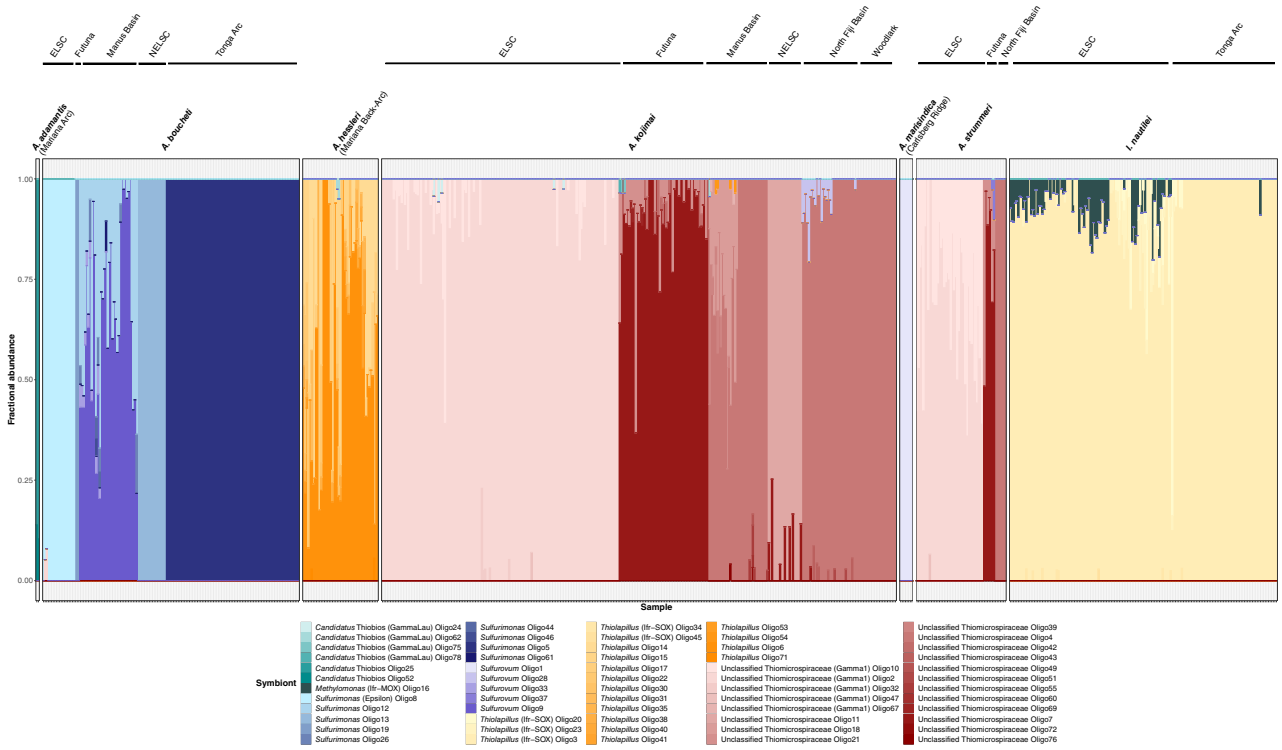


Fig. 2. Fractional abundance plot of symbiont ASVs within individual snails according to *Alviniconcha* and *Ifremeria* species.

for structuring by vent field (Appendix 1: Fig. S3), while no intra-regional differentiation was observed or could be tested in symbionts of any other host species that we sampled from multiple localities (data not shown). However, this finding is likely an artefact of the limited resolution of the 16S rRNA marker gene. For example, recent metagenomic analyses indicate that symbiont populations of all host taxa from the Lau Basin are partitioned between vent sites (Breusing *et al*., 2021a; Breusing *et al*., 2021b). In contrast to the traditional view of microbial biogeography that poses that 'everything is everywhere' (Baas-Becking, 1934), geographic subdivision of microbial symbionts appears to be common in a variety of marine symbioses, often exceeding that of the corresponding host populations (Ho *et al*., 2017; Gould and Dunlap, 2019; Davies *et al*., 2020; Ücker *et al*., 2021; Breusing *et al*., 2021a; Breusing *et al*., 2021b). Depending on the symbiotic system, these patterns might arise from local adaptation, contrasting dispersal limitations between hosts and symbionts, host ecological behaviour and/or differences in environmental transmission mode. Given the strong oceanographic barriers among back-arc basins in the Western Pacific Ocean (Mitarai *et al*., 2016), the observed partitioning of host-specific symbiont ASVs according to broader geographic area might be largely due to decreased symbiont dispersal opportunities (that are likely exacerbated by environmental differences between habitats). By contrast, symbiont structure within regions, where dispersal limitations appear to be weak (Mitarai *et al*., 2016), is probably predominantly driven by ecological factors, such as differences in depth or vent geochemistry (Breusing *et al*., 2021a; Breusing *et al*., 2021b). Indeed, in *A. kojimai* the observed partitioning of symbiont types by vent field was correlated with contrasting depth regimes (Appendix 2), which often aligns with gradients in fluid chemistry (Beinart *et al*., 2012). On the other hand, the strong latitudinal subdivision found for the *Thiolapillus* symbiont of *A. hessleri* might be explained by dispersal limitations as biophysical models indicate that the southern and northern parts of the Mariana Basin are largely isolated (Mitarai *et al*., 2016; Breusing *et al*., 2021a; Breusing *et al*., 2021b).

Our data suggest that other factors, such as host size, have a comparatively small influence on the diversity and composition of symbiont ASVs within host individuals. Despite significant associations of symbiont richness with host size, correlation coefficients were low, suggesting limited biological relevance of this factor on intra-host symbiont diversity (Appendix 1: Fig. S4). These results were consistent independent of whether analyses were carried out across or within individual host species. For intra-species analyses only correlations for *A. kojimai* and *A. boucheti* were significant, though weak ($p \leqq 0.05$;

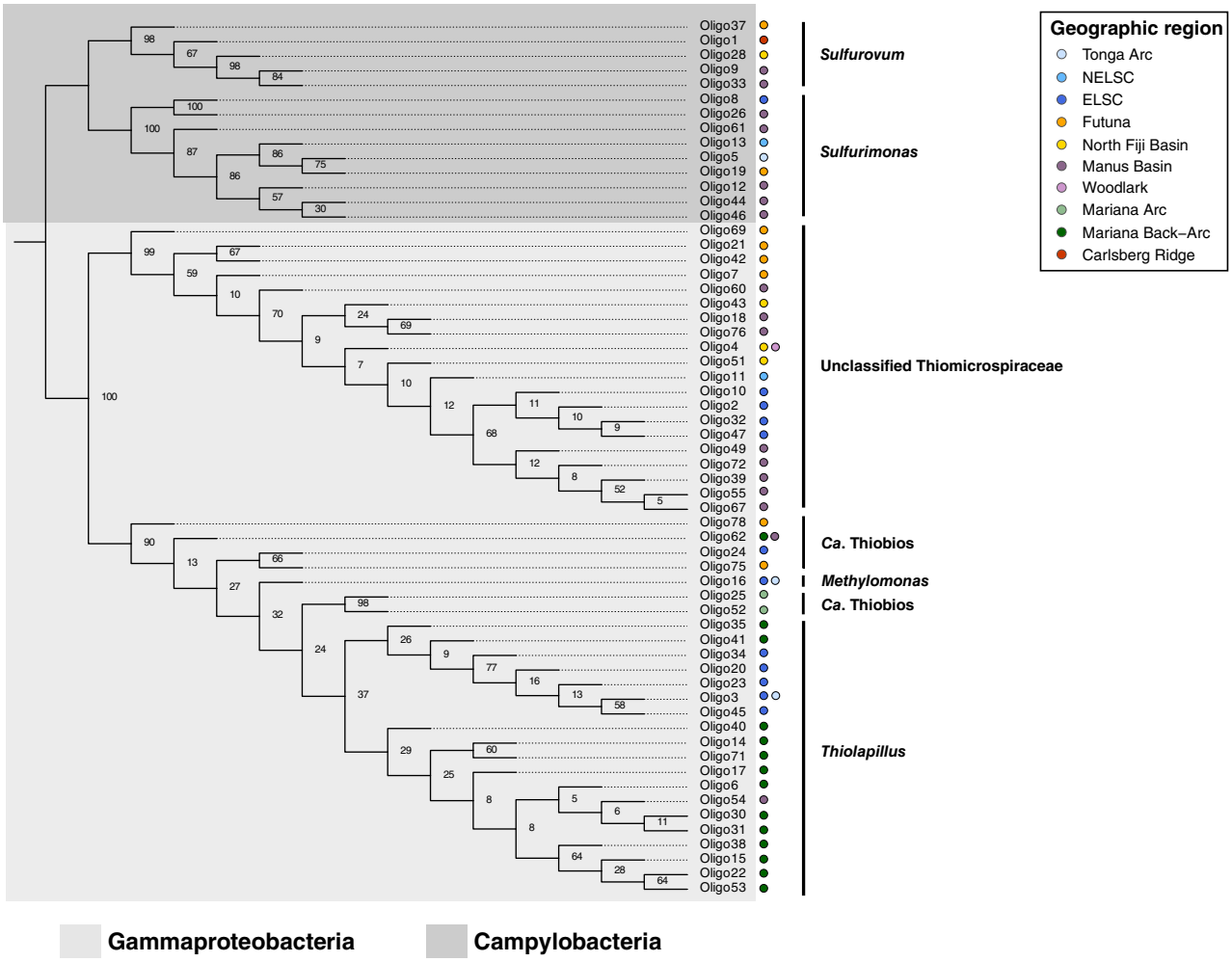


Fig. 3. Mid-point rooted IQTree consensus phylogeny of ASVs within symbiont genera. Node labels indicate ultra-fast bootstrap support values.

$R^2 \leqq 0.09$). In most cases individuals contained only one symbiont ASV in accordance with Sanger sequence analyses (Beinart *et al*., 2012, 2015), though in some individuals up to six ASVs were observed. Although our study lacks data from settling larvae and juveniles, these findings could indicate that symbiont acquisition in provannid snails follows a different process than in bathymodiolin mussels and is more similar to that in vestimentiferan tubeworms. Hydrothermal vent mussels remain competent for symbiont acquisition throughout their lifetime (Wentrup *et al*., 2014; Ansorge *et al*., 2019), which should favour increased symbiont diversity in older individuals as well as newly infected juveniles where symbiont sorting has not yet been completed. By contrast, vestimentiferan tubeworms obtain their symbionts exclusively in a narrow window after settlement during post-larva metamorphosis (Nussbaumer *et al*., 2006). Symbiont diversity can thus be expected to be highest at that developmental stage, with little effect of host size on symbiont richness during later stages. Alternatively, our observations may indicate that 16S rRNA amplicon sequences do not provide enough strain-level resolution to observe shifts in symbiont composition across development stages, and that metagenomic analyses of symbiont populations are necessary instead.

Symbiont richness differs between host species and individuals

Despite low impact of host size, *Alviniconcha* and *Ifremeria* exhibited notable variability in symbiont diversity, both among individuals and species (Fig. 4B). These patterns could result from differences in the availability and composition of free-living symbiont lineages at the time of infection, subsequent mutations inside the host and/or host selection on particular strains. Among host taxa, *A. adamantis* and *A. marisindica* showed the lowest symbiont diversity, which is probably due to the fact that these

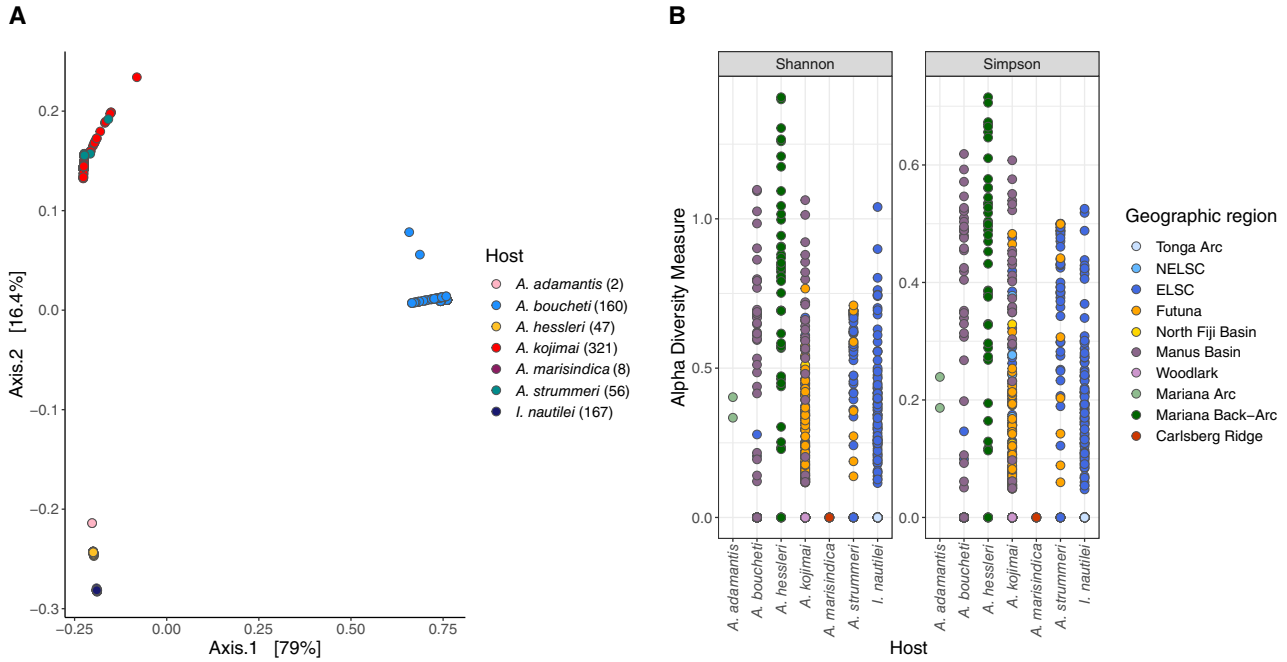


Fig. 4. (A) Principal coordinate analysis plot based on weighted UniFrac distances. Data were normalized to proportions before analysis. Numbers in brackets indicate sample sizes for each host taxon. (B) Alpha diversity within host species based on Shannon's and Simpson's diversity index.

Table 1. Results for linear decomposition models (LDM) and PERMANOVAs based on weighted UniFrac distances.

	LDM				PERMANOVA	
Source of variation	*df*	*F*	VE (%)	*p*	*F*	*p*
Model 1						
Geographic region	8	2.2861	16.49	**0.0001**	324.710	**0.0010**
Host	3	1.3426	25.82	**0.0001**	704.879	**0.0010**
Model 2						
Vent	2	3.5363	30.39	**0.0001**	4959.805	**0.0010**
Host	2	6.1004	52.42	**0.0001**	6424.714	**0.0010**
Model 3						
Methodology	1	0.1179	1.99	**0.0001**	264.664	**0.0010**
Vent	2	3.5363	29.79	**0.0001**	9919.611	**0.0010**
Host	2	6.1004	51.38	**0.0001**	12849.428	**0.0010**

Three different models were run to assess the effects of DNA preservation, extraction and sequencing method on patterns of symbiont diversity: (1) Model including the complete dataset and controlling for effects of methodology, (2) Model restricted to *A. boucheti*, *A. kojimai* and *A. strummeri* from the ELSC and controlling for effects of methodology, (3) Model restricted to *A. boucheti*, *A. kojimai* and *A. strummeri* from the ELSC and including methodology as main explanatory factor. Sources of variation are shown in sequential order tested in the model. Significant sources of variation are indicated in bold. *df* = degrees of freedom, *F* = *F* statistic, VE = explained variation, *p* = *p*-value.

species were each sampled from only a single vent site and were represented by relatively few individuals (Fig. 4B). Interestingly, *A. hessleri* displayed some of the highest alpha diversities, with up to six ASVs within single host individuals, despite its restricted geographic distribution and small sample size compared to some of the other *Alviniconcha* species included in our analyses. Maybe the wide variation of geochemical conditions in the Mariana Back-Arc Basin (Trembath-Reichert *et al*., 2019) allows for a greater range of micro-niches, which could promote diversity in the free-living symbiont pool. In this case, symbionts within this host species might have a higher functional diversity that could favour coexistence of multiple strains, as has recently been reported for bathymodiolin mussels, where hosts can carry up to 16 symbiont strains due to variation in metabolic gene content (Ansorge *et al*., 2019). Alternatively, some of the observed variations might reflect intra-host mutations of a single or a few symbiont phylotypes post-infection. In the absence of genomic data, this explanation seems likely as all *A. hessleri*

symbiont ASVs were very similar to each other, with an average of 99.4% pairwise sequence identity.

Symbiont richness is not saturated

Although we analysed symbiont 16S rRNA composition in over 700 snail individuals, symbiont discovery did not reach saturation in our dataset (Fig. 5). The number of ASVs within *A. hessleri* and *I. nautilei*, which both host symbionts of the genus *Thiolapillus*, was closest to reaching a plateau, while ASV accumulation curves for all other species showed a steady increase (Fig. 5). This is an interesting finding given that *A. hessleri* and *I. nautilei* were sampled across a relatively restricted area compared to some of the other species (Appendix 1: Table S1). For other taxa that were represented by few individuals and geographic locations (e.g. *A. adamantis*, *A. marisindica*), but also those with widespread distributions (e.g. *A. kojimai*, *A. boucheti*), increased sampling efforts will probably reveal a currently hidden diversity of symbiont ASVs in the future. Consequently, while our dataset does not allow comparisons of diversification between symbiont genera or species at this time, more ASVs especially for some of the gammaproteobacterial taxa (e.g. unclassified Thiomicrospiraceae, *Ca*. Thiobios) will likely be recovered given the prevalence of gammaproteobacterial symbioses in provannid snails and other vent invertebrates (Dubilier *et al*., 2008).

Conclusions

Here, we characterized the global diversity of chemosynthetic gill endosymbionts associated with species within the genera *Alviniconcha* and *Ifremeria*. As predicted by previous work, we found that each host species harboured 1–2 species- or genus-level symbiont phylotypes. However, we were able to further assess strain-level symbiont composition and diversity within and between individual snails by employing amplicon analysis of the 16S rRNA gene. In all host species, ASV accumulation curves indicated that the full diversity of symbionts associated with *Alviniconcha* and *Ifremeria* remains to be characterized. In most cases, symbiont ASV composition and richness were related to geographic range, with most ASVs detected in species where we sampled a large number of individuals across >10 geographically distant vent fields (e.g. *A. kojimai* and *A. boucheti*). An exception to this was *A. hessleri*, which had high symbiont richness and inter-region symbiont structure despite a smaller sample size and much more modest geographic range, suggesting that these are not the only factors dictating symbiont composition and diversity. A more complete appraisal of the taxonomic and functional diversity of symbionts associated with *Alviniconcha* and *Ifremeria* will be critical to our understanding of the ecology and evolution of these genera, which have been assessed as 'Endangered' or 'Vulnerable' on the IUCN Red List (https://www.iucnredlist.org) due to imminent risks from deep-seabed mining activities at hydrothermal vents in the Indian and Pacific oceans.

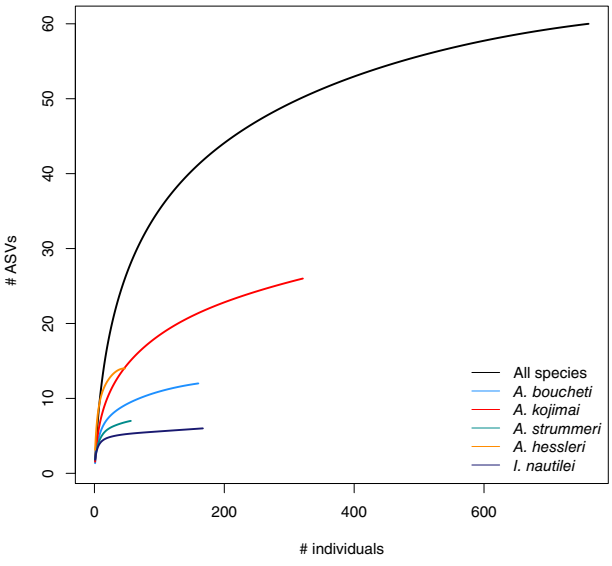


Fig. 5. Symbiont ASV accumulation curves.

Experimental procedures

Sample collection and amplicon library preparation

Animal samples were obtained with remotely or human-operated vehicles from 23 Indo-Pacific vent localities that encompassed the global distributional range of species within the genera *Alviniconcha* and *Ifremeria* (Appendix 2; Fig. 1). Upon recovery of the samples, endosymbiont-bearing gill tissue was dissected and frozen or stored in RNALater™ (Thermo Fisher Scientific, Waltham, MA, USA) at −80°C. DNA was purified with the Zymo Quick DNA 96 Plus and ZR-96 Clean-up kits (Zymo Research, Irvine, CA, USA) or the Qiagen DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany). 2 × 250 bp paired-end amplicon libraries for the 16S rRNA V4–V5 region were constructed with the 515F/926R primer pair (Walters *et al*., 2015) and sequenced to an average of 34 844 total reads on Illumina MiSeq and NovaSeq platforms at the Argonne National Laboratory (Lemont, IL, USA) and Novogene (Beijing, China) respectively (Appendix 2). Host species were identified through shell morphology (Laming *et al*., 2020) and subsequent sequencing of the mitochondrial *COI* gene with universal primers (Folmer *et al*., 1994; Geller *et al*., 2013).

Identification of amplicon sequence variants

We used the USEARCH v11 denoising pipeline (Edgar, 2010) to decompose merged, adapter-clipped paired-end reads into ASVs, imposing a merge length of 300–400 bp, a maximum error rate of 0.001 and a minimum base quality of 20. The taxonomic identity of each variant was determined in QIIME2 (https://qiime2.org) with a Naïve Bayes classifier trained against the V4–V5 region extracted from the SILVA 132 99% reference database as well as through BLAST+ searches against the NR database (Camacho *et al*., 2009). Only ASVs that had a match to a previously verified *Alviniconcha* or *Ifremeria* gill endosymbiont sequence were considered for further analysis. To assess potentially unrecovered variation in the symbiont dataset we applied the OLIGOTYPING v2.0 method (Eren *et al*., 2013). ASVs with less than 2.37% abundance in a sample were excluded to account for sample cross-contamination (Minich *et al*., 2019). Phylogenetic relationships among ASVs were determined with the IQTREE (Minh *et al*., 2020) plugin for QIIME2 based on 10 independent runs with each 5000 ultrafast bootstrap samples. Ultrafast bootstrap trees were optimized through the nearest neighbour interchange procedure with a perturbation strength of 0.2 and a stopping criterion of 200 trees.

16S rRNA diversity analyses

We used the PHYLOSEQ package in R v4.0.3 (McMurdie and Holmes, 2013; R Core Team, 2020) to assess symbiont 16S rRNA variation within and between hosts and geographic regions, excluding samples with less than 1000 reads to ensure statistical robustness. For alpha and beta diversity analyses symbiont abundances were normalized to proportions (McKnight *et al*., 2019). Metric and non-metric multidimensional scaling plots were constructed based on weighted UniFrac distances. To verify that the distribution of ASV diversity is representative of real biological patterns and not technical artefacts from differences in methodology, we performed linear decomposition models (LDMs) and a modified version of PERMANOVA with the LDM package in R, as these methods have been shown to be relatively robust to variance in group dispersion (Hu and Satten, 2020). Analyses were run on both the full dataset and a data subset including only samples of *Alviniconcha* from the ELSC which were processed with a mixture of methods. PERMANOVAs and LDMs were conducted with 1000 and 10 000 maximum permutations respectively, with methodology included as either confounding variable or main explanatory factor. Relationships between number of ASVs and host size were determined based on Spearman rank correlations with the GGPUBR package (Kassambara, 2020).

Acknowledgements

We thank the captains, crews and pilots of the R/V *Falkor* (ROV *Ropos* and *SuBastian*; cruises FK160320, FK160407 and FK161129), R/V *L'Atalante* (ROV *Victor*; cruise Chubacarc 2019) and R/V *Xiangyanghong 9* (HOV *Jiaolong*; cruise DY38) for supporting the sample collections that made this study possible. Verena Tunnicliffe and Amanda Bates are gratefully acknowledged for providing snail specimens from the Mariana Arc and Back-Arc. We further thank the technical staff at the Argonne National Laboratory and Novogene for preparing and sequencing the Illumina 16S rRNA amplicon libraries. This study was supported by the Key Special Project for Introduced Talents Team of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (GML2019ZD0409) (to P.Y.Q.), the General Research Fund of Hong Kong SAR (16101219) (to P.Y.Q.), the Schmidt Ocean Institute, a University of Rhode Island Proposal Development Grant award (to R.A.B.), the U.S. National Science Foundation (grant numbers OCE-1536331, 1819530 and 1736932 to R.A.B. and EPSCoR Cooperative Agreement OIA-#1655221), and the Cerberus project (ANR-17-CE02-0003-01, coord. S. Hourdez).

Data Availability

All bioinformatic scripts and final files for analysis are available on GitHub under https://github.com/cbreusing/Provannid_16S_SSU_meta-analysis. Raw 16S rRNA amplicon reads have been deposited in the Sequence Read Archive under BioProjects PRJNA473256, PRJNA473257, PRJNA610289, PRJNA610290, PRJNA763784 and PRJNA767887, while host *COI* sequences are available in GenBank under accession numbers listed in Appendix 2.

References

Ansorge, R., Romano, S., Sayavedra, L., Porras, M.Á.G., Kupczok, A., Tegetmeyer, H.E., *et al*. (2019) Functional diversity enables multiple symbiont strains to coexist in deep-sea mussels. *Nat Microbiol* **4**: 2487–2497.

Apprill, A. (2017) Marine animal microbiomes: toward understanding host–microbiome interactions in a changing ocean. *Front Mar Sci* **4**: 222.

Baas-Becking, L.G.M. (1934) *Geobiologie of Inleiding Tot De Milieukunde*. The Hague, The Netherlands: W.P. van Stockum & Zoon.

Beinart, R.A. (2019) The significance of microbial symbionts in ecosystem processes. *mSystems* **4**: e00127-19.

Beinart, R.A., Gartman, A., Sanders, J.G., Luther, G.W., and Girguis, P.R. (2015) The uptake and excretion of partially oxidized sulfur expands the repertoire of energy resources metabolized by hydrothermal vent symbioses. *Proc R Soc B* **282**: 20142811.

Beinart, R.A., Sanders, J.G., Faure, B., Sylva, S.P., Lee, R. W., Becker, E.L., *et al*. (2012) Evidence for the role of endosymbionts in regional-scale habitat partitioning by

hydrothermal vent symbioses. *Proc Natl Acad Sci U S A* **109**: E3241–E3250.

Borowski, C., Giere, O., Krieger, J., Amann, R., and Dubilier, N. (2002) New aspects of the symbiosis in the provannid snail *Ifremeria nautilei* from the North Fiji Back Arc Basin. *Cah Biol Mar* **43**: 321–324.

Breusing, C., Genetti, M., Russell, S.L., Corbett-Detig, R.B., and Beinart, R.A. (2021a) Host-symbiont population genomics provide insights into partner fidelity, transmission mode and habitat adaptation in deep-sea hydrothermal vent snails. *bioRxiv* . https://doi.org/10.1101/2021.07.13.452231v1.

Breusing, C., Johnson, S.B., Mitarai, S., Beinart, R.A., and Tunnicliffe, V. (2021b) Differential patterns of connectivity in Western Pacific hydrothermal vent metapopulations: a comparison of biophysical and genetic models. *Evol Appl* . https://doi.org/10.1111/eva.13326.

Breusing, C., Johnson, S.B., Tunnicliffe, V., Clague, D.A., Vrijenhoek, R.C., and Beinart, R.A. (2020) Allopatric and sympatric drivers of speciation in *Alviniconcha* hydrothermal vent snails. *Mol Biol Evol* **37**: 3469–3484.

Bronstein, J.L. (ed). (2015) *Mutualism*, 1st ed. Oxford, United Kingdom: Oxford University Press.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Davies, S.W., Moreland, K.N., Wham, D.C., Kanke, M.R., and Matz, M.V. (2020) *Cladocopium* community divergence in two *Acropora* coral hosts across multiple spatial scales. *Mol Ecol* **29**: 4559–4572.

Desbruyères, D., Hashimoto, J., and Fabri, M.-C. (2006) Composition and biogeography of hydrothermal vent communities in Western Pacific Back-Arc Basins. In *Geophysical Monograph Series*. Christie, D.M., Fisher, C.R., Lee, S.-M., and Givens, S. (eds). Washington, DC: American Geophysical Union, pp. 215–234.

Dubilier, N., Bergin, C., and Lott, C. (2008) Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat Rev Microbiol* **6**: 725–740.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

Eren, A.M., Maignien, L., Sul, W.J., Murphy, L.G., Grim, S.L., Morrison, H.G., and Sogin, M.L. (2013) Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol Evol* **4**: 1111–1119.

Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* **3**: 294–299.

Geller, J., Meyer, C., Parker, M., and Hawk, H. (2013) Redesign of PCR primers for mitochondrial cytochrome *c* oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Mol Ecol Resour* **13**: 851–861.

Gould, A.L., and Dunlap, P.V. (2019) Shedding light on specificity: population genomic structure of a symbiosis between a coral reef fish and luminous bacterium. *Front Microbiol* **10**: 2670.

Ho, P.-T., Park, E., Hong, S.G., Kim, E.-H., Kim, K., Jang, S.-J., *et al*. (2017) Geographical structure of endosymbiotic bacteria hosted by *Bathymodiolus* mussels at eastern Pacific hydrothermal vents. *BMC Evol Biol* **17**: 121.

Hu, Y.J., and Satten, G.A. (2020) Testing hypotheses about the microbiome using the linear decomposition model (LDM). *Bioinformatics* **36**: 4106–4115.

Johnson, S.B., Warén, A., Tunnicliffe, V., Dover, C.V., Wheat, C.G., Schultz, T.F., and Vrijenhoek, R.C. (2015) Molecular taxonomy and naming of five cryptic species of *Alviniconcha* snails (Gastropoda: Abyssochrysoidea) from hydrothermal vents. *Syst Biodivers* **13**: 278–295.

Kassambara, A. (2020) *ggpubr*: *"ggplot2" based publication ready plots*. URL https://CRAN.R-project.org/package=ggpubr.

Laming, S.R., Hourdez, S., Cambon-Bonavita, M.A., and Pradillon, F. (2020) Classical and computed tomographic anatomical analyses in a not-so-cryptic *Alviniconcha* species complex from hydrothermal vents in the SW Pacific. *Front Zool* **17**: 12.

McKnight, D.T., Huerlimann, R., Bower, D.S., Schwarzkopf, L., Alford, R.A., and Zenger, K.R. (2019) Methods for normalizing microbiome data: an ecological perspective. *Methods Ecol Evol* **10**: 389–400.

McMurdie, P.J., and Holmes, S. (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**: e61217.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* **37**: 1530–1534.

Minich, J.J., Sanders, J.G., Amir, A., Humphrey, G., Gilbert, J.A., and Knight, R. (2019) Quantifying and understanding well-to-well contamination in microbiome research. *mSystems* **4**: 4.

Mitarai, S., Watanabe, H., Nakajima, Y., Shchepetkin, A.F., and McWilliams, J.C. (2016) Quantifying dispersal from hydrothermal vent fields in the western Pacific Ocean. *Proc Natl Acad Sci U S A* **113**: 2976–2981.

Nussbaumer, A.D., Fisher, C.R., and Bright, M. (2006) Horizontal endosymbiont transmission in hydrothermal vent tubeworms. *Nature* **441**: 345–348.

R Core Team. (2020) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Sogin, E.M., Kleiner, M., Borowski, C., Gruber-Vodicka, H. R., and Dubilier, N. (2021) Life in the dark: phylogenetic and physiological diversity of chemosynthetic symbioses. *Annu Rev Microbiol* **75**: 1–718.

Suzuki, Y., Kojima, S., Sasaki, T., Suzuki, M., Utsumi, T., Watanabe, H., *et al*. (2006b) Host-symbiont relationships in hydrothermal vent gastropods of the genus *Alviniconcha* from the Southwest Pacific. *Appl Environ Microbiol* **72**: 1388–1393.

Suzuki, Y., Kojima, S., Watanabe, H., Suzuki, M., Tsuchida, S., Nunoura, T., *et al*. (2006a) Single host and symbiont lineages of hydrothermal-vent gastropods *Ifremeria nautilei* (Provannidae): biogeography and evolution. *Mar Ecol Prog Ser* **315**: 167–175.

Trembath-Reichert, E., Butterfield, D.A., and Huber, J.A. (2019) Active subseafloor microbial communities from Mariana back-arc venting fluids share metabolic strategies

across different thermal niches and taxa. *ISME J* **13**: 2264–2279.
Ücker, M., Ansorge, R., Sato, Y., Sayavedra, L., Breusing, C., and Dubilier, N. (2021) Deep-sea mussels from a hybrid zone on the mid-Atlantic ridge host genetically indistinguishable symbionts. *ISME J* **15**: 3076–3083.
Van Dover, C.L., Humphris, S.E., Fornari, D., Cavanaugh, C. M., Collier, R., Goffredi, S.K., *et al*. (2001) Biogeography and ecological setting of Indian Ocean hydrothermal vents. *Science* **294**: 818–823.
Walters, W., Hyde, E.R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., *et al*. (2015) Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *mSystems* **1**: e00009-15.
Wentrup, C., Wendeberg, A., Schimak, M., Borowski, C., and Dubilier, N. (2014) Forever competent: deep-sea bivalves are colonized by their chemosynthetic symbionts throughout their lifetime. *Environ Microbiol* **16**: 3699–3713.
Wilkins, L., Leray, M., O'Dea, A., Yuen, B., Peixoto, R.S., Pereira, T.J., *et al*. (2019) Host-associated microbiomes drive structure and function of marine ecosystems. *PLoS Biol* **17**: e3000533.
Windoffer, R., and Giere, O. (1997) Symbiosis of the hydrothermal vent gastropod *Ifremeria nautilei* (Provannidae) with endobacteria-structural analyses and ecological considerations. *Biol Bull* **193**: 381–392.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1: Supporting information.
Appendix S2: Supporting information.

Supplementary Figures and Tables

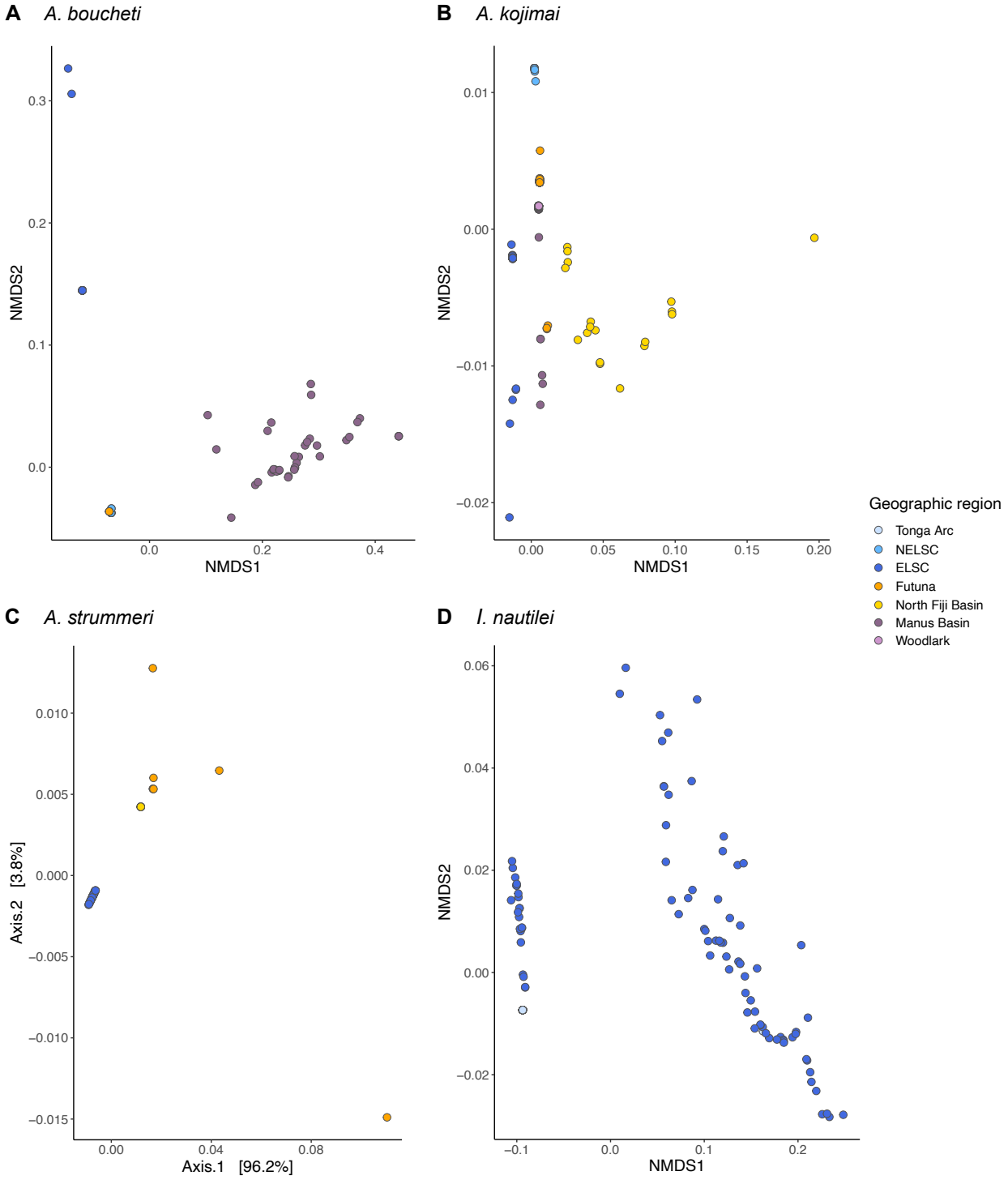


Fig. S1 Metric and non-metric multidimensional scaling plots based on weighted UniFrac distances for host species sampled across broader geographic regions. Proportional ASV abundances for *A. boucheti* and *I. nautilei* were square-root transformed before analysis to improve fit of the data. Metric multidimensional scaling was used in *A. strummeri* due to data limitations. ASVs are notably structured by geographic region in *Alviniconcha*, but not *Ifremeria*.

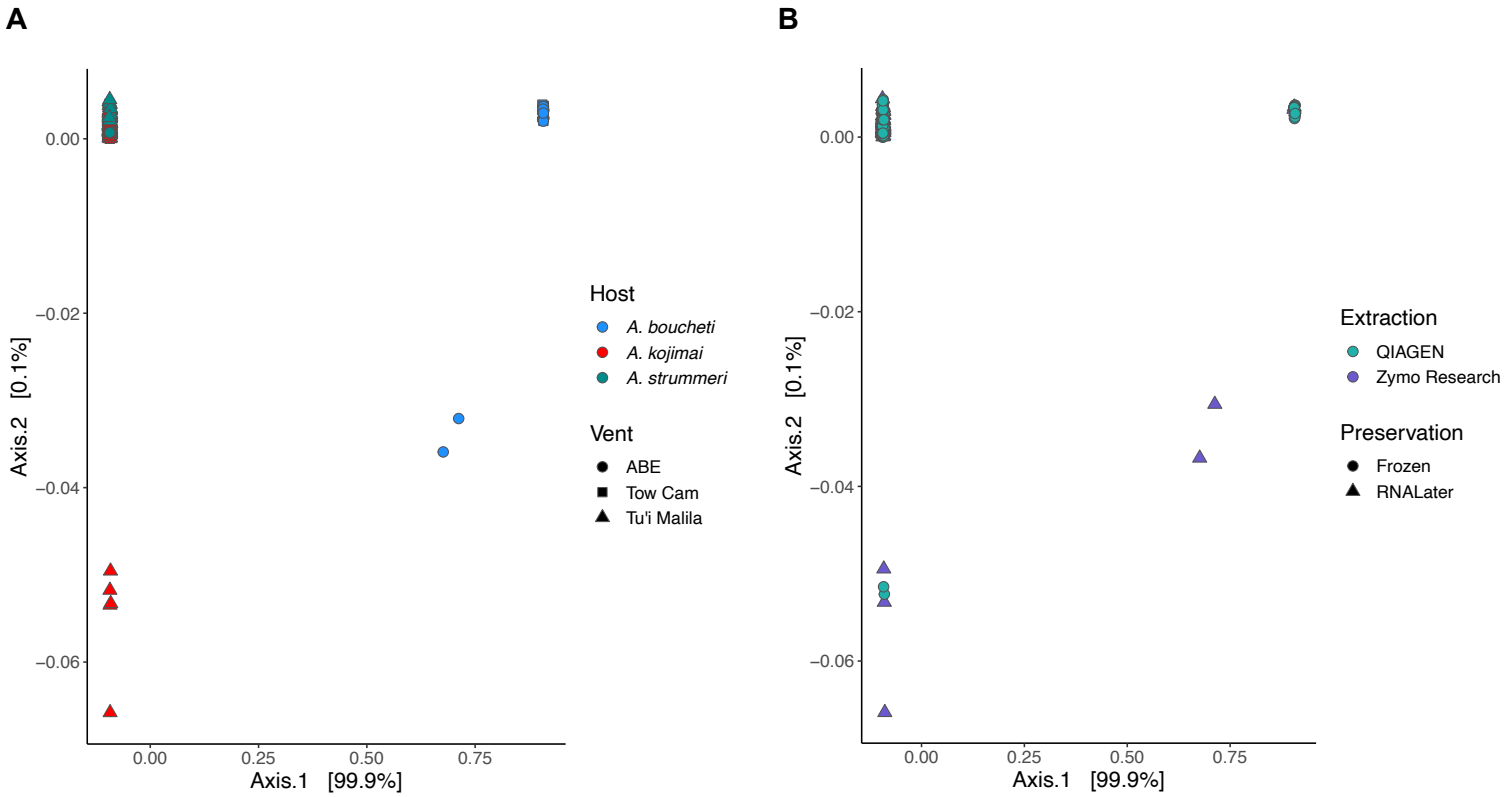


Fig. S2 Metric multidimensional scaling plots based on weighted UniFrac distances for *Alviniconcha* species from the ELSC. Samples from this region were processed with a mixture of DNA preservation and extraction methods and were therefore suitable for assessing the effect of methodology on patterns of symbiont ASV diversity. Proportional abundances were square-root transformed before analysis and points were jittered slightly to improve visualization of overlapping samples. No obvious clustering by method is observed and structuring of samples appears to be better explained by biological factors, such as host species (see also Table 1).

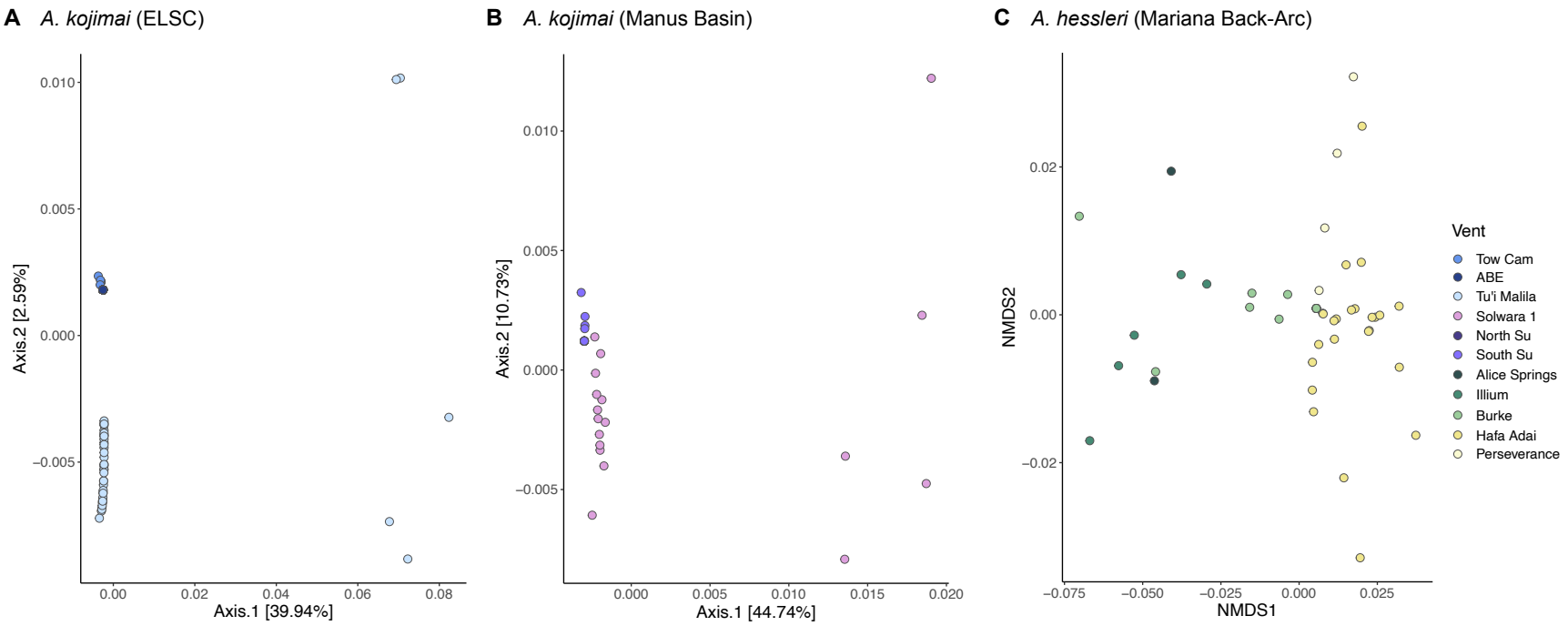


Fig. S3 Metric and non-metric multidimensional scaling plots based on weighted UniFrac distances for *A. kojimai* and *A. hessleri*. Within-basin structure for the other species was either not observed or could not be accurately assessed due to data insufficiency.

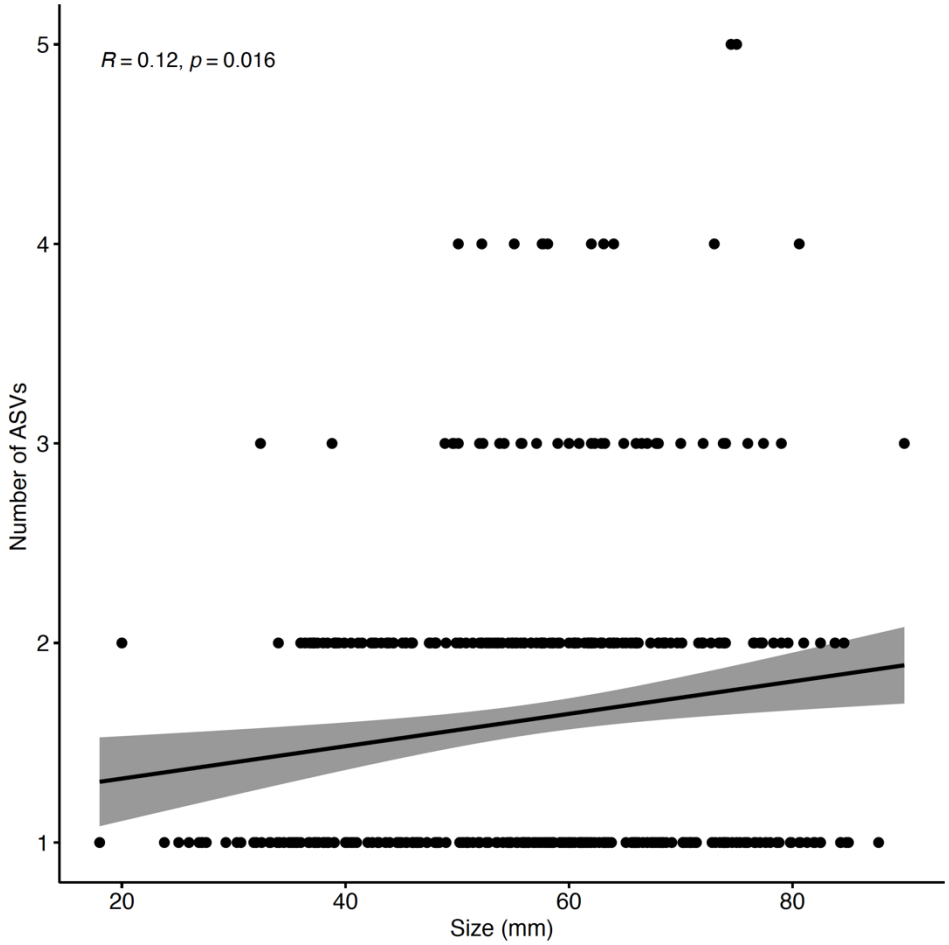


Fig. S4 Correlation of symbiont richness with host size.

Table S1 Number of snails sampled per species and vent field.

Host	Geographic region	Vent	# samples
A. adamantis	Mariana Arc	Chamorro	2
A. boucheti	Tonga Arc	Niua South	83
	NELSC	Mangatolo	18
	ELSC	Tow Cam	7
		ABE	13
	Futuna	Fati Ufu	3
	Manus Basin	Big Papi	8
		Roman Ruins	10
		Fenway	2
		North Su	4
		South Su	12
A. hessleri	Mariana Back-Arc	Alice Springs	2
		Illium	5
		Burke	7
		Hafa Adai	24
		Perseverance	9
A. kojimai	NELSC	Mangatolo	21
	ELSC	Tow Cam	38
		ABE	5
		Tu'i Malila	105
	Futuna	Fati Ufu	41
		Fatu Kapa	15
	North Fiji Basin	Phoenix	35
	Manus Basin	Solwara 1	18
		North Su	4
		South Su	15
	Woodlark	Scala	24
A. marisindica	Carlsberg Ridge	Wocan	8
A. strummeri	ELSC	Tow Cam	2
		ABE	1
		Tu'i Malila	39
	Futuna	Fati Ufu	7
	North Fiji Basin	Phoenix	7
I. nautilei	Tonga Arc	Niua South	59
	ELSC	Tahi Moana	32
		ABE	14
		Tu'i Malila	62

Chapitre 4 :

Role of diversifying selection and reproductive isolation in shaping interspecific divergences in *Alviniconcha*

Chapter realized in collaboration with Hourdez S, Ballenghien M, Corre E, Broquet T and Jollivet D.

1- Introduction

Ecological speciation occurs when two populations adapt to different environments as a consequence of divergent selection on functional traits allowing specialisation to distinct ecological niches. Ecological niches are characterised by both abiotic (physico-chemical conditions of habitats) and biotic factors (competition with neighbouring species, or host-microbe relationships). When conditions, resources, and interactions vary in space and through time, selection can drive populations to adapt to their local conditions, which can fuel divergence and may ultimately contribute to reproductive isolation. This process is independent of geographic context, and can occur in both allopatry and sympatry (Coyne and Orr, 2004).

Due to the movement of the Earth's tectonic plates, the hydrothermal environment is fragmented, often distributed in a linear fashion along ridges, and constantly remodelling. Vent fluid emissions and associated polymetallic chimneys are ephemeral and mostly depend on the nature of the oceanic crust where they stand. Physico-chemical conditions encountered by vent species are therefore very heterogeneous (Du Preez and Fisher, 2018) and highly fluctuating according to the chaotic mixture of the hydrothermal fluid (hot, acidic pH=2, sulphidic and anoxic; Le Bris et al., 2003; Von Damm, 1990) and the surrounding water (cold and well oxygenated). Physico-chemical gradients are thus conditioning the distribution of individuals in space (more or less concentric distribution of species around the vent emissions depending on their nutritional needs and their ability to cope with the fluid toxicy: Desbruyeres, 1982) and in time (evolution of the strength of emissions from their birth to their extinction leading to faunal successions: Shank et al., 1998).

Specific adaptations have allowed species to cope with these ‘extreme’ conditions: development of a very efficient respiratory system with (hypertrophied gills, highly diversified respiratory pigments: Hourdez and Lallier, 2007), mitochondrial respiratory chain adapted to high temperature coupled with H_2S detoxification (Powell and Somero, 1986), specific arsenal to cope with chemical and thermal stresses (oxidative stress: Dilly et al., 2012), and immune systems co-evolving with bacterial symbionts (Tasiemski et al., 2014). Symbiosis in particular relies on the acquisition and evolution of specific

antimicrobial peptides that has allowed some organisms to acquire chemoautotrophy by controlling and cultivating sulfo-oxidative or methanotrophic bacteria living in the hydrothermal environment (Papot et al., 2017).

On the one hand, species associated with hydrothermal vents are subject to a high degree of spatial isolation, strong variations in habitat conditions, and highly dynamic turnover of suitable habitats. It is therefore possible that ecologically-based divergent selection has played a central role in the emergence of new species. On the other hand, the specificity of hydrothermal vent habitats has forced species to adopt highly specialised adaptations that may have constrained the possibility of species diversification, leading instead to some level of morphological and functional stasis (Fontanillas et al., 2017). To address this issue, one needs to assess the role of disruptive selection in local adaptation, niche specialisation, and reproductive isolation, with a particular focus on the evolution of genes that could affect immune, respiratory, and detoxification systems.

Among the species inhabiting the hydrothermal environment, gastropods of the genus *Alviniconcha* form a group of engineer species in the back-arc basins of the Western Pacific as well as on the Indian Ocean rift (Podowski et al., 2009). Six species of *Alviniconcha* have been described based on divergence at the mitochondrial *Cox1* gene (Breusing et al., 2020; Johnson et al., 2015) and some subtle morphological differences for, at least, half of the species (Laming et al., 2020). Assuming that the evolutionary rate of the *Cox1* gene has been constantly low (around 0.0015 substitutions per site per million years, Breusing et al., 2020), the split between the two major *Alviniconcha* clades is estimated at 48 Ma (Breusing et al., 2020). This split separates *A. boucheti* and *A. marisindica* from *A. kojimai*, *A. hessleri*, and *A. strummeri*, while the position of the sixth species *A. adamantis* remains unresolved. This is an important split because it coincides with different symbiotic associations of the gastropods with either *Gamma-proteobacteria* or *Campylobacteria*. In the clade of species bearing *Gamma-proteobacteria*, the separation between *A. strummeri* and *A. kojimai*/*A. hessleri* is more recent and dated at about 25 Ma (Breusing et al., 2020).

Here we focus on three species of *Alviniconcha* (*A. kojimai*, *A. boucheti* and *A. strummeri*) that live sympatrically at vents of the Western Pacific back-arc basins. A first demographic study based on genomic data indicated that despite a strong divergence (in average 10% on mtDNA and 3% on nuDNA), these gastropods were still able to exchange genes due to recent secondary contacts between populations after a long period of isolation

(Castel et al. submitted). Generally, these three species of *Alviniconcha* occupy the hot (7-42°C), sulfidic (250 µM), and low-oxygenated (< 50 µM) habitat of the hydrothermal environment, indicating that these animals have tolerance to high temperatures and high levels of sulphides (Podowski et al., 2009, 2010). But in detail, it seems that these species have different ecological niches. Indeed, *A. kojimai* and *A. strummeri* are more frequently found in diffuse venting zones while *A. boucheti* occupies preferentially the heights of hydrothermal chimneys and is thus closer to the vent emissions enriched in metallic compounds (Beinart et al., 2012; Breusing et al., 2020; Castel et al. submitted). These differences in ecological niches seem to be linked to the type of symbionts associated with these species. Previous studies have shown that *A. kojimai* and *A. strummeri* are associated with *Gamma-proteobacteria* that tolerate low H_2 and H_2S concentrations (Beinart et al., 2012, 2019). *A. boucheti* possesses *Campylobacteria* which tolerate higher concentrations of H_2 and H_2S, hence the presence of this species closer to hydrothermal fluids enriched in these compounds (Beinart et al., 2012, 2014; Breusing et al., 2020). As suggested in some studies (Beinart et al., 2012; Breusing et al., 2020), the host-microbe relationship could have led to differences in ecological niches between these specific gastropods either in response to ecologically-driven selection or in response to selection against hybridization (reinforcement).

In this context, this study aims to better understand the role of adaptive polymorphisms in the divergence and reproductive isolation between the three species of *Alviniconcha* gastropods that inhabit the hydrothermal vents from the South-West Pacific. To this end, we examined the number and the physiological functions associated with both highly divergent and positively-selected genes in the branches leading to the three species.

2- Materials and Methods

2.1- Animal sampling and transcriptome sequencing

We used three *A. boucheti*, two *A. kojimai* and two *A. strummeri* individuals sampled during the TM 235 Lau expedition (2009) on board the R/V Thomas G. Thompson (chief scientist: C.R. Fisher). *A. kojimai* and *A. strummeri* were sampled from the Tui Malila site in the Lau Basin, while two individuals of *A. boucheti* were sampled at Tow Cam and one at ABE (Figure 54 and Table S9).

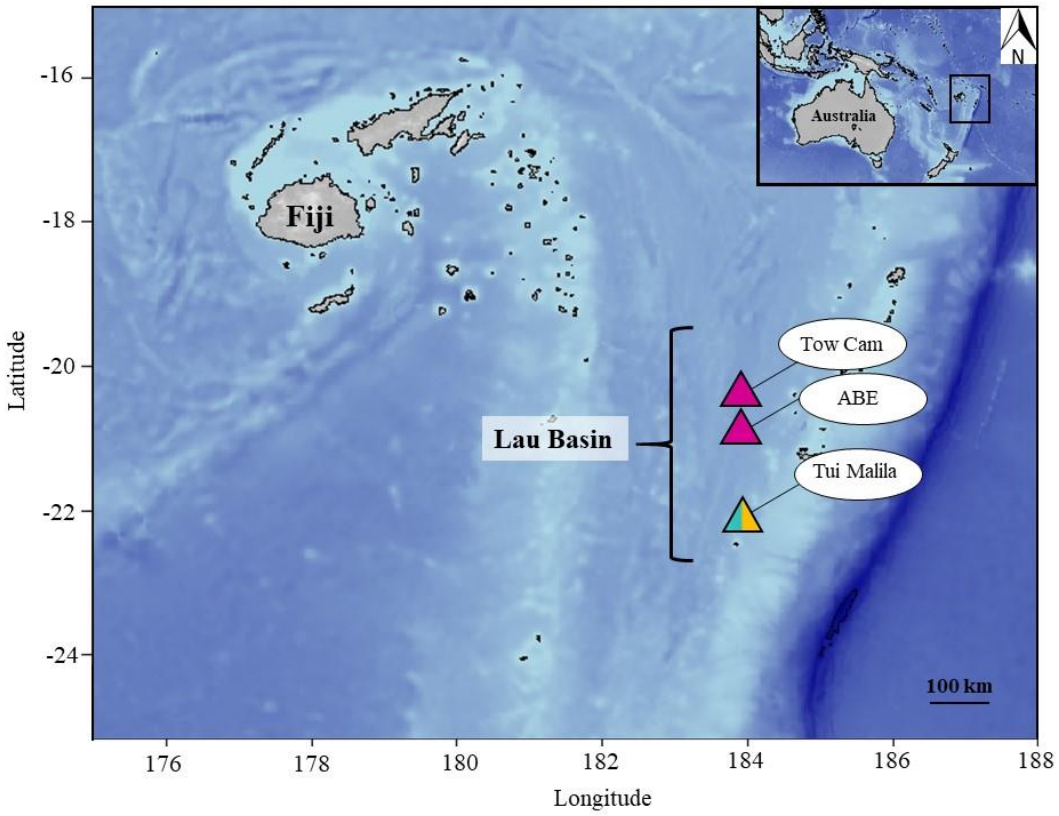


Figure 54: Geographic positions of *Alviniconcha* gastropods. purple: *A. boucheti*; turquoise: *A. kojimai* and yellow: *A. strummeri*) sampled for transcriptome sequencing during the TM 235 Lau expedition (2009) The map background was obtained using the R package marmap (Pante and Simon-Bouhet, 2013).

A sample of gill tissue taken from each individual and preserved dry at -80°C was used for total RNA extraction in order to produce RNAseq datasets and subsequent transcriptome assemblies. All individuals representing *A. kojimai* and *A. strummeri* were sampled from the Tui Malila site in the Lau Basin, while two individuals of *A. boucheti* were sampled at Tow Cam and one at ABE (Figure 54 and Table S9). Total RNA was extracted using NucleoSpin® RNA Plus (Macherey-Nagel) according to the manufacturer's protocol from the frozen gill tissue of each individual. RNA integrity was confirmed using a 2100 Bioanalyzer (Agilent Technologies). The cDNA library for transcriptome sequencing was prepared using NEBNext® mRNA Library Prep Reagent Set for Illumina® (New England Biolabs) which uses a poly(A) processing, cDNA size fragmentation and a pair of illumina i5 and i7 indexes for each library. For each individual's library, the size of the cDNA fragments was selected to range approximately 300-600 bp using a blue Pippin electrophoresis and both ends of these fragments were sequenced over a length of 150 bp using the illumina technology. The indexed

cDNA libraries were sequenced at The McGill Genome Centre (Quebec) on a half line of Novaseq6000.

2.2- Transcriptome analysis

The raw reads were cleaned by removing adaptor sequences, empty/short reads and low-quality sequences (including reads containing 'N') using Trimmomatic (Bolger et al., 2014). Since cDNA libraries were made from gill tissues that contain large amounts of endosymbiotic bacteria, it is possible that bacterial RNA was not totally discarded from the polyA hybridization technique. Hence a taxonomic assignment of reads was performed with Kraken v.2 (Wood and Salzberg, 2014) in order to keep only the eukaryotic reads before assembly. From 5 to 10% of the initial reads per individual were assigned as prokaryotes and removed from the dataset. The remaining reads were then assembled in rnaSPAdes 3.13.1 (Bankevich et al., 2012). Transcripts were then checked with the Busco database in order to assess the level of completeness of the assemblies with reference to the highly conserved 'universal' proteins in eukaryotes (Table S10).

After the finalisation of each assembly, the resultant transcripts were used in the Galaxy pipeline AdaptSearch (Mataigne et al. in prep., repository https://github.com/abims-sbr/adaptsearch) to find orthologous sequences between species, estimate species absolute divergence from the coding sequences and find genes under positive selection. The first step of this pipeline is to filter transcripts to keep only one sequence per gene using both the length (300 pb) and quality score of the transcripts as a proxy, and to perform an additional CAP3 assembly (Filter_assemblies in AdaptSearch; Huang and Madan, 1999). Transcripts of the different species are then put together in orthogroups by using Orthofinder (Emms and Kelly, 2015). The cDNA sequences of orthogroups are then aligned using BlastAlign (Belshaw and Katzourakis, 2005) and the reading frame (partial CDS: coding DNA sequence) is identified from each species alignment with the home-made python script CDSsearch (repository https://github.com/abims-sbr/adaptsearch). At this stage, portions of sequences containing gap or undetermined nucleic acids (N) are removed from each alignment while keeping the right reading frame (i.e. without stop codons). When several alternative CDS were found, each frame was blasted against the SwissProt/Uniprot and RefProtein databases implemented in NCBI. This annotation allowed us to avoid false positives and reduce the risk of later calculating d_N/d_S values from non-coding or incorrectly framed sequences. Sequences with no

annotation from UniProt/RefProtein or with a CDS shorter than 300 nucleotides were removed from the analysis. Sequences with bacterial gene annotation were then eliminated.

2.3- Divergence and d_N/d_S estimation

Estimation of species divergence (*t*), synonymous and non-synonymous sites (*S* and *N*), substitution rates (d_S and d_N), and detection of positive selection in protein-coding DNA sequences were performed using CodeML (Yang and dos Reis, 2011) also implemented in AdaptSearch (Mataigne et al. in prep.).

The rate of non-synonymous mutations (d_N) was compared to the rate of synonymous mutations (d_S) because they can help us to understand the adaptive evolution of genes. Indeed, usually, the synonymous rate is three times greater than the rate of substitutions at non-synonymous sites but should be equal or reversed under gene relaxation or positive evolution. Natural selection may cause a non-synonymous mutation deficiency to eliminate deleterious mutations to maintain 3D structure and thus protein function. Genes can be thus classified in four main categories (1) fast-evolving genes with high d_S and high d_N, (2) positively-selected genes with high d_N and low d_S, (3) relaxed genes with equal d_N and d_S and, (4) highly conserved genes with low d_N and high d_S. A linear regression to account for this very large variance in non-synonymous/synonymous mutation rates was performed for each species branch and compared to the expected line (x=y) representing genes under neutral evolution (Figure 56). Indeed, the further the regression line of observed data from the neutral expectations is, the more positive selection is acting in the diversification of genes.

CodeML uses the method of Yang and Nielsen (2000) to calculate d_S and d_N values from pairwise comparisons of protein-coding DNA sequences. This program analyses the distribution of substitutions along a reference species tree with the branch model (M0) where d_N/d_S is not allowed to vary among the branches and the free-ratio branch model (M1) where d_N/d_S is free to evolve among branches, and using a gene-to-gene approach where the likelihoods of the two models are compared using a likelihood ratio test (LRT). Estimates of divergence (*t*) by the phylogenetic distance between two sequences and the number of synonymous ($S.d_S$) and non-synonymous ($N.d_N$) substitutions were then estimated along the three branches leading to the three *Alviniconcha* species using a trifurcated tree and the M1 model. The outputs of CodeML (M1 model) were analysed and then filtered. When the number of synonymous substitutions was below one, the d_N/d_S obtained for a given branch was not taken into account, however the role of its genes may be important in selection

processes and their function will be discussed in this analysis. Genes under positive selection were then identified from our set of orthologous genes within each species when the d_N/d_S was greater than one in the branch leading to this species for significant values of the LRT between the two models. These genes putatively under divergent selection were then subsequently annotated using a Blastx approach on protein databases implemented in NCBI. Particular attention (annotation) was paid to the 1% most divergent genes, as these are potentially the ones that could have initiated the differentiation between populations.

3- Results

Individual RNAseq data were assembled with rnaSPAdes resulting between 107 573 and 130 205 transcripts for the three individuals of *A. boucheti*, 157 510 and 305 161 transcripts for the two individuals of *A. kojimai* and 149 438 and 244 377 transcripts for the two individuals of *A. strummeri*. After removal of bacterial transcripts (between 0.66% and 1.66%), these individual transcriptome assemblies were merged within each species to produce a species-specific reference transcriptome, resulting in 215 286 transcripts in *A. boucheti*, 254 984 transcripts in *A. strummeri* and 284 840 transcripts in *A. kojimai*. The busco analysis revealed that our species assembly has between 246 and 248 of the 255 genes found in all eukaryotes and between 4285 and 4566 genes out of 5295 common to molluscs, demonstrating good quality assemblies (Table S10). After filtering with Filter_Assemblies, the number of transcripts retained per species was respectively 21 343, 23 071 and 21 404 for *A. boucheti*, *A. strummeri* and *A. kojimai*. Combining these transcripts with OrthoFinder produced 9 201 filtered orthogroups of more than 300 bp length between the three species. Since many of the CDS were short, aligning and finding the correct reading frame only identified 2 676 CDS from which only 1 726 had a conclusive annotation. Checking annotation to remove bacterial genes finally led to 1704 eukaryotic CDS with a correct reading frame. These 1704 annotated CDS were subsequently analysed to examine their d_N/d_S ratio along the branches leading to the three species.

3.1- Distribution of *S*, *N*, d_S, d_N, $S.d_S$ and $N.d_N$

The average value of the number of synonymous sites observed within the 1 704 genes is 212 ($\sigma = 104$), while the average number of non-synonymous sites is 476 ($\sigma = 230$) (Figure 55). The distribution of d_S, $S.d_S$, d_N and $N.d_N$ for each species are shown in Figure 55, they showed higher values of d_S and $S.d_S$ than d_N and $N.d_N$ for each species, with approximately

10-fold higher values of d_S than d_N (For example in *A. kojimai*, d_S : 0.03, $S.d_S$: 5.97, d_N : 0.004 and $N.d_N$: 1.97). Note however, higher values of the estimated number of non-synonymous ($N.d_N$) and synonymous ($S.d_S$) mutations in *A. boucheti* ($N.d_N$: 4.41 and $S.d_S$: 12.82), the species most distant phylogenetically from the other two (*A. kojimai*: $N.d_N$: 1.97 and $S.d_S$: 5.97; *A. strummeri*: $N.d_N$: 1.91 and $S.d_S$: 5.77).

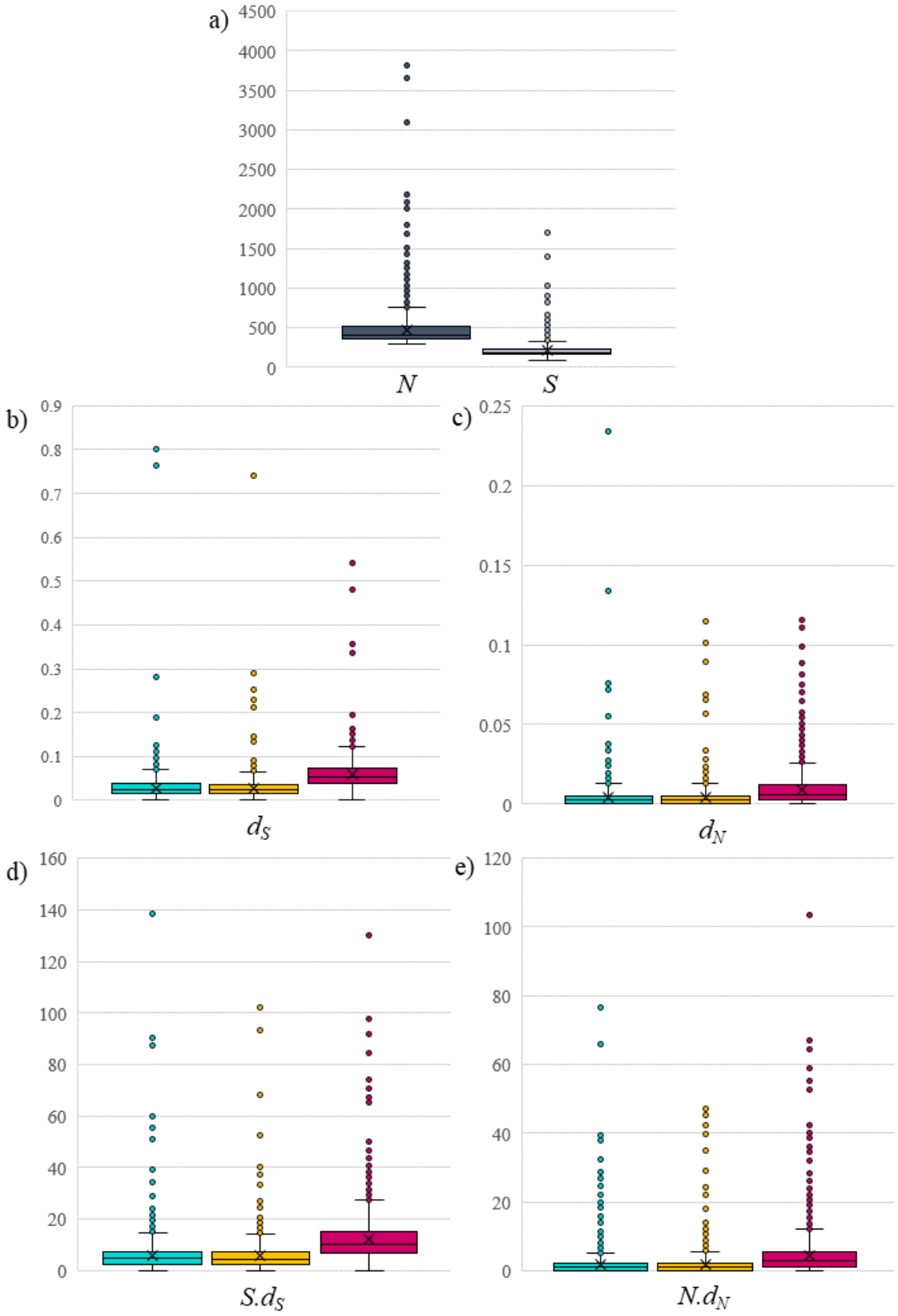
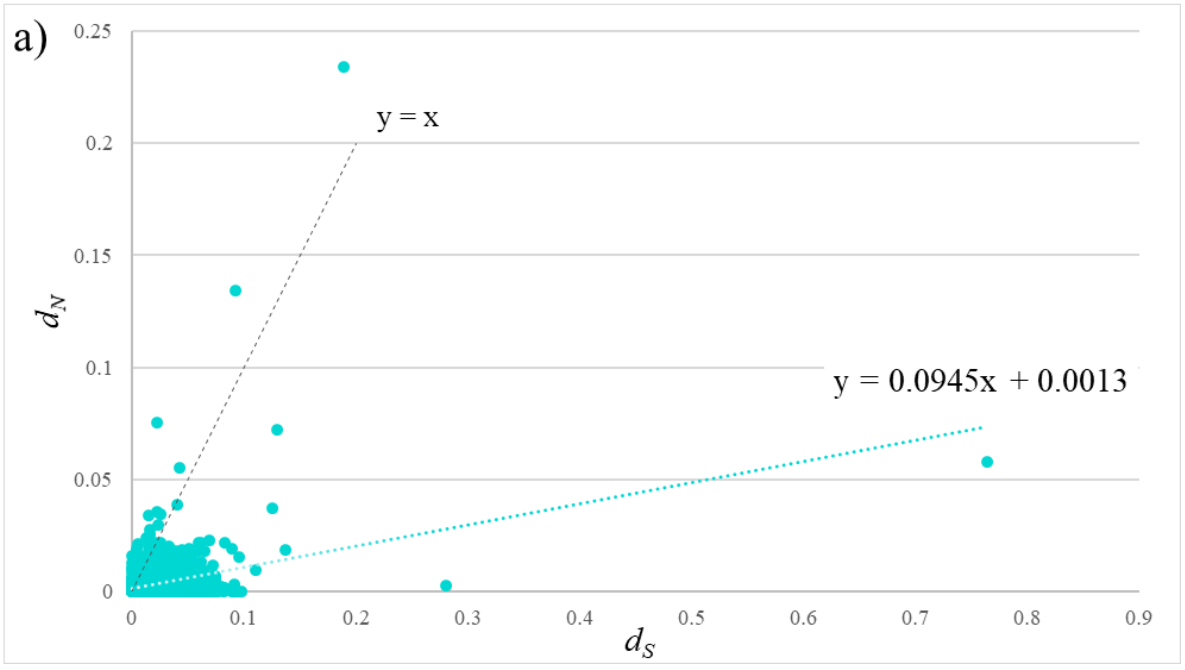
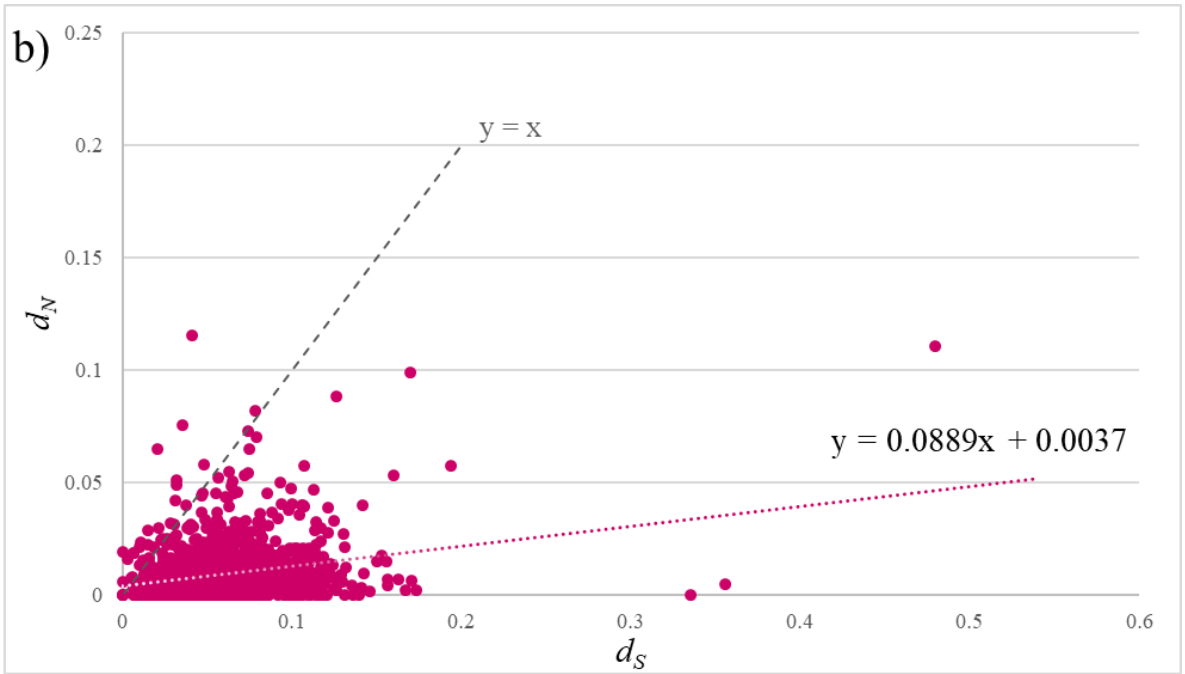


Figure 55: Distribution of non-synonymous and synonymous sites (*N* and *S*) for the 1704 genes in a), substitution rates d_S and d_N in b) and c) and the number of synonymous ($S.d_S$) and non-synonymous

($N.d_N$) substitutions estimates in d) and e) for the *A. kojimai* branch in turquoise, *A. strummeri* branch in yellow and *A. boucheti* branch in purple.

The biplot representing the d_S and d_N are shown in Figure 56. Looking at species branches, the slopes of the linear regression between d_N and d_S were almost similar for the three species with slightly greater coefficients for *A. kojimai* (0.094) and *A. boucheti* (0.089) (Figure 56). Note that these linear regressions are far from the expected value under a neutral model of evolution (x=y).





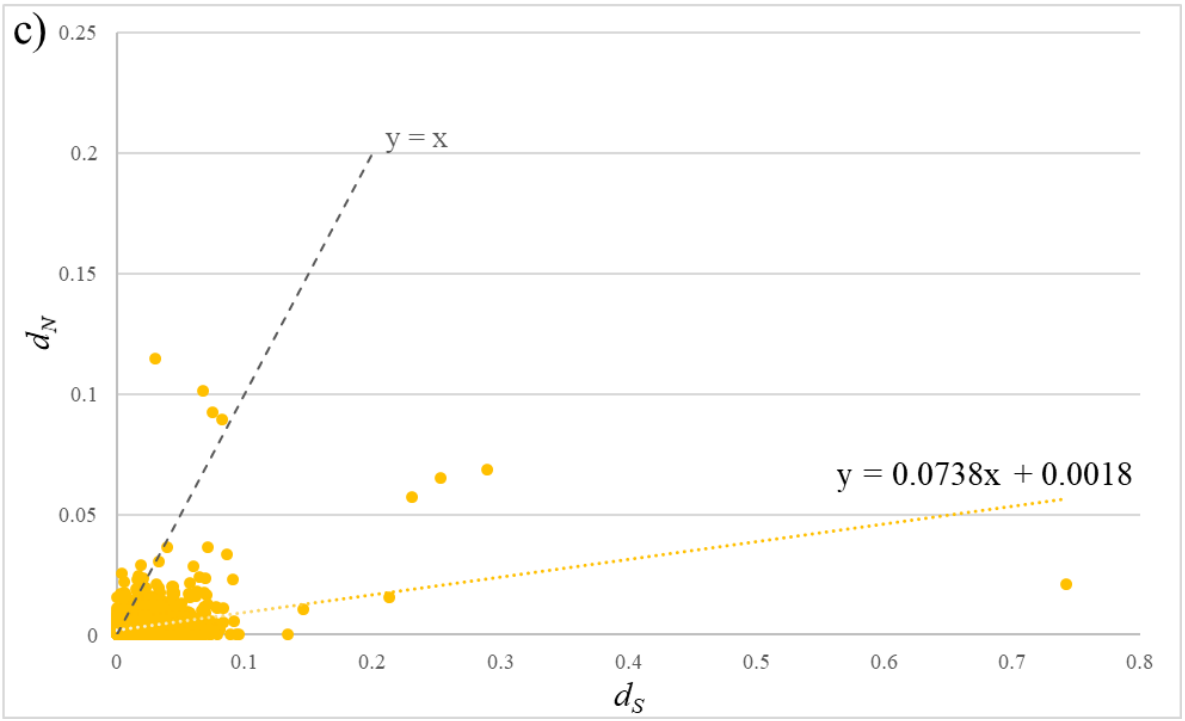


Figure 56: Biplot of synonymous and non-synonymous rates estimated along the branches of a) *A. kojimai*, b) *A. boucheti*, c) and *A. strummeri* using the free ratio model (M1) of codeML. The grey dashed line represents the expected linear relationship between d_S and d_N under neutral evolution with values above the line under positive selection.

3.2- Distribution of gene divergence between species pairs

The number of substitutions accumulated in coding sequences of genes during the speciation process that gave birth to the three Western Pacific species of *Alviniconcha* were estimated by calculating the codon divergence (*t*) along the terminal branches of the mid-rooted tree of the three *Alviniconcha* species from a set of orthologous transcripts. Among the 1 704 orthologous genes studied, the average divergence between species pairs was 0.043 (σ = 0.04), 0.044 (σ = 0.04) and 0.017 (σ = 0.03) for *A. boucheti*/*A. kojimai*, *A. boucheti*/*A. strummeri* and *A. kojimai*/*A. strummeri*, respectively. For each species pair the distributions of divergence were unimodal with *t* values different from zero (Figure 57). For *A. kojimai* and *A. strummeri*, 86% of the genes displayed divergences between 1% and 10%, and 12.8% of the genes diverged by more than 10% ($t > 0.1$) (Figure 57). Of the orthologous genes found for the other species pairs, about half of them had a divergence distributed between 1% and 10% (52.9% in *A. kojimai*/*A. boucheti* and 54.8% in *A. strummeri*/*A. boucheti*), and the other half diverged by more than 10% (46.9% in

A. kojimai/*A. boucheti* and 45.2% in *A. strummeri*/*A. boucheti*) (Figure 57). This difference between *A. kojimai*/*A. strummeri* and the two other pairs confirms that the three species have been separated by two speciation events with different timings, in line with previous results (Breusing et al., 2020; Johnson et al., 2015; Castel et al. submitted).

To test whether highly divergent genes could represent genes associated with the reproductive isolation of these gastropods (e.g. gender attraction, gamete recognition, mito-nuclear interactions), the function of the 1% most divergent genes was examined and reported in Table S11. These fast-evolving 26 genes are mainly encoding proteins involved in (1) cell-cell adhesion, lipid homeostasis and cell response to heat (afadin, obscurin, coadhesin, flocculation protein, SCO-spondin, protein dispatched homolog 3, cadherin), (2) glycolysis process (fructose-bisphosphate aldolase C) and (3) cytoskeleton organisation (tubulin alpha-1 chain, ankyrin-1, tenascin-X). Some of the fast-evolving transcripts (high d_S) involved in the divergence of *A. boucheti* when compared with the two other species also contained numerous non-synonymous mutations in genes coding for proteins involved in DNA/RNA repairs (ubiquitin protein ligase, RNA exonuclease) but also putative positively-selected genes in the two other species such as the mechanosensory protein 2 (mechanosensory behaviour), the RNA pseudouridylate synthase (mito-nuclear interaction) or the Zinc finger protein 708 and the protein RRNAD1 both involved in rRNA methylation and translation regulation. Other fast-evolving genes with both high d_S and d_N found between the two other species also included DNA/RNA repair proteins (helicase, double-strand break repair protein), bacterial defence (phospholipase, sarm1 hydrolase), phospholipid processing (GPCPD1 homolog 2) or mitonuclear interactions (mitochondrial-processing protease).

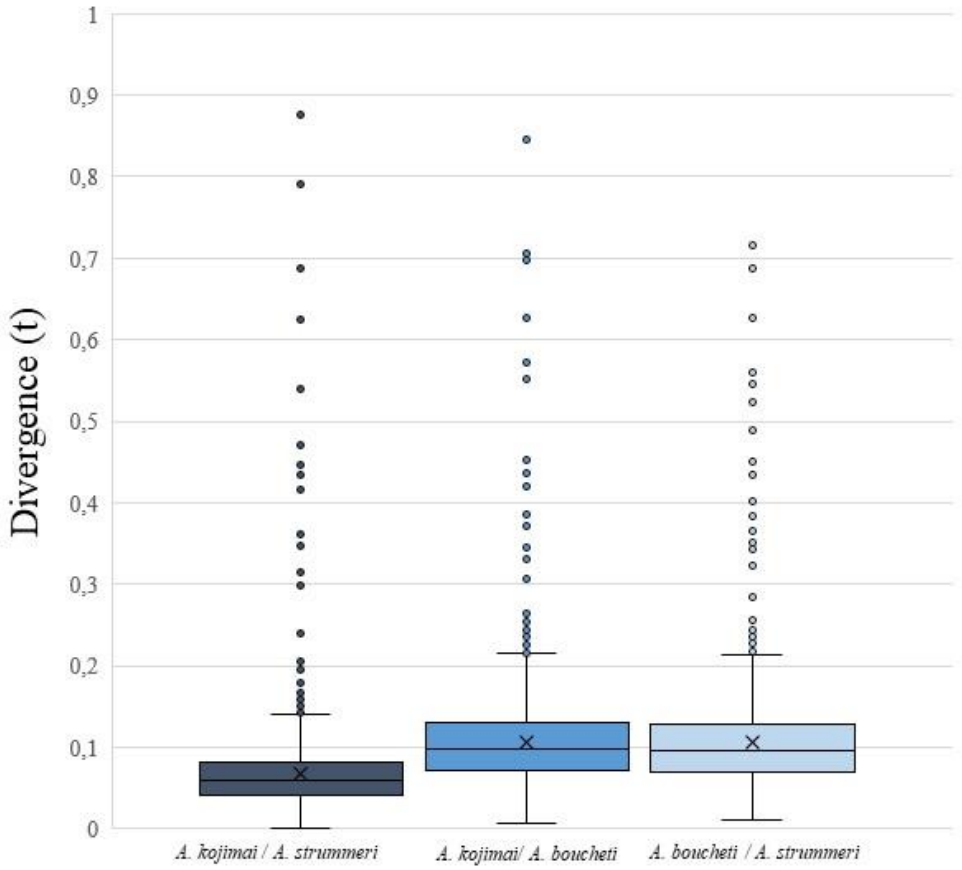


Figure 57: Boxplot summarising the distribution of the divergence (*t*) values (in substitutions per codon) found for 1 704 orthologous genes for the three pairwise comparisons of *Alviniconcha* species.

3.3- Distribution of genes according to their selective signature

The role of positive selection in the sympatric evolution of *Alviniconcha* species was investigated by estimating synonymous to non-synonymous substitution rates along the branches leading to the three *Alviniconcha* species for our set of 1704 orthologous genes. The average d_N/d_S estimated were 0.14, 0.13 and 0.17 for *A. kojimai*, *A. strummeri* and *A. boucheti* respectively (Dotted bars in Figure 58). Regarding the distribution of d_N/d_S values in both the *A. kojimai* and *A. strummeri* branches, a large fraction of genes did not accumulate non-synonymous mutations in at least one of the two branches, with a $N.d_N$ close to zero (35% of genes had $N.d_N < 0.01$, Figure 58). Among genes having accumulated at least one non-synonymous mutation in the branch leading to *A. kojimai* (that is, 1 082 genes), 44% were under strong purifying selection, with d_N/d_S values lower than 0.25, and 18.7% of the d_N/d_S values were comprised between 0.25 and 1 forming a small peak of proteins subject to much more relaxed selective pressure. Only 2.3% of the genes were under positive selection on this

branch with d_N/d_S values higher than 1. The distribution of d_N/d_S for genes having accumulated at least one non-synonymous mutation along the *A. strummeri* branch was almost the same as for *A. kojimai*, with 1.9% of genes under positive selection (d_N/d_S >1).

The distribution of *N.*d_N values in the branch leading to *A. boucheti* showed that only 16.5% of genes did not accumulate non-synonymous mutation along this branch since the ancestral species split. Amongst genes having accumulated at least one non-synonymous mutation in the branch, about 61% of genes were under strong purifying selection (i.e. d_N/d_S) < 0.25), 20.7% had more relaxed proteins with d_N/d_S values distributed between 0.25 and 1.0 and 1.8% of the genes were under positive selection with d_N/d_S values higher than 1. A series of genes (ca. 4% of the analysed genes) exhibiting a few non-synonymous substitutions between species but no synonymous ones and thus infinite d_N/d_S ratios were firstly discarded from the analysis then their annotation was looked at.

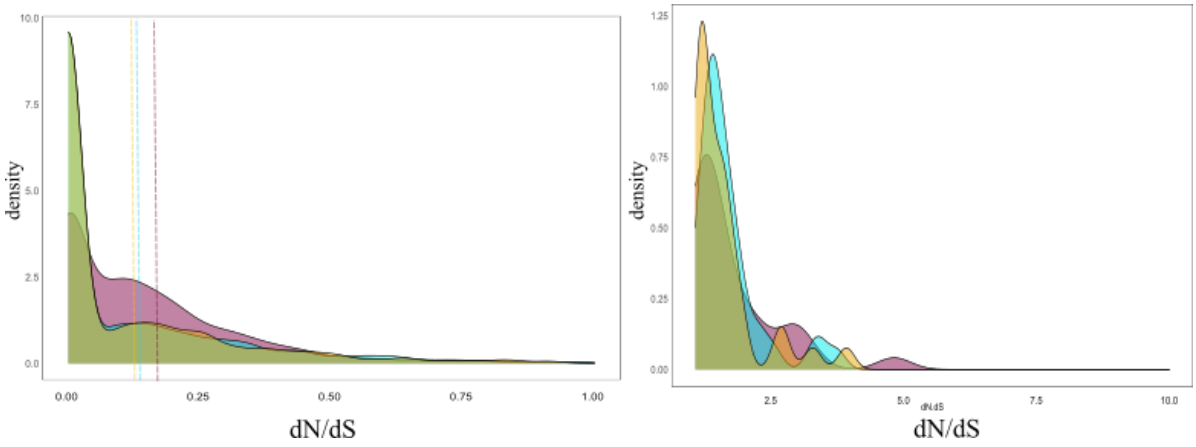


Figure 58: Density distributions of d_N/d_S values estimated for each pairwise alignment of orthologous genes between the three species of *Alviniconcha* implemented in codeML. The purple colour corresponds to substitutions accumulated in the *A. boucheti* branch, the turquoise colour corresponds to those accumulated in the *A. kojimai* branch and yellow to substitutions accumulated in the *A. strummeri* branch. The dotted bars show the average d_N/d_S over all genes. a) Density distribution of d_N/d_S values between 0 and 1 (between 1 665 and 1 674 genes depending on the species) and b) Density distribution of d_N/d_S values between 1 and 10 (between 30 and 39 genes depending on the species).

Among the 83 genes under positive selection (significant values of d_N/d_S >1, Figure 59), three were under positive selection along all three branches leading to *Alviniconcha* species (3.6%). Nine genes were identified as under diversifying selection in branches leading to two of the three species (10.8%). The remaining genes under diversifying selection were all

species-specific, accounting for 84.1% of the total number of positively-selected genes investigated.

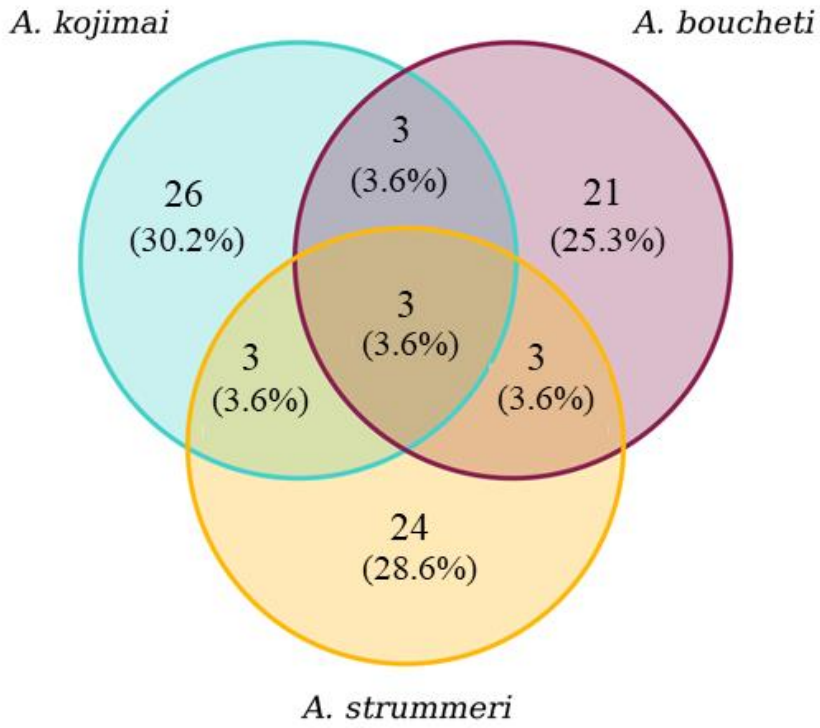


Figure 59: Number and percentage of genes under positive selection (d_N/d_S >1) in each species branch.

Among the 83 genes found to be significantly under positive selection in at least one species, 75 were annotated with a GO Biological function in the UniProt database (Table 8). These genes are encoding proteins involved in (1) transcription/translation/replication regulation and biosynthesis (zinc finger protein CCCH/771/586, smad nuclear-interacting protein 1, ubinuclein-1, adenosine deaminase), (2) endocytosis and the innate immunity response, which may play a crucial role in the gastropod symbioses (adenosine deaminase, caspase-8, WAP four-disulfide protein, Toll-like receptor 4/6, draper protein, cubilin, asialoglycoprotein receptor), (3) embryonic development (maternal protein exuperantia, rapamycin-insensitive companion of mTOR), (4) DNA repairs and methylation (protein fem-1 homolog B, serine/threonine protein kinase), (5) response to salt/ oxygen/ nitrate stresses (aquaporin-11, mitochondrial amidoxime reducing component 2, ankyrin), (6) muscle contraction/ adult walking behaviour (cytosolic carboxypeptidase 1, Titin, bardet-Biedl syndrome 2 protein). Genes under positive selection in each of the six categories described above were identified in all three species, although those for the immune system and in

response to stress appeared to be more present in *A. boucheti* and those related to embryonic development more present in *A. kojimai* and *A. strummeri*.

In addition, when taking into consideration non-synonymous diverging genes with a d_S near to zero, an additional series of genes under positive selection were identified (between 2, 34 and 36 genes in *A. boucheti*, *A. kojimai* and *A. strummeri* respectively, some of which code for neuromedin B receptor, spermine oxidase, cardiolipin hydrolase, galactoside alpha fucosyltransferase, galactosyltransferase, paraplegin, oligoadenylate synthetase or the cilia/flagella associated protein). These genes are likely involved in gonad and gamete formation, and thus possibly gamete recognition, and were mostly found in the branch leading to *A. kojimai*. At least two other genes possibly involved in germ cell development were also found with d_N/d_S greater than one in the branch leading to *A. strummeri* (PR zinc finger protein 4, spindle-like microcephaly protein). By contrast, nearly all positively-selected genes found in the branch leading to *A. boucheti* were related to functions associated with immune and stress responses, protein glycosylation and the formation of peptidoglycans, gene regulation by methylation, and apoptotic processes possibly linked with symbiont digestion. One gene involved in the shell mineralisation (PIF protein) was found to contribute to the diversifying evolution of *A. boucheti*.

Table 8: Annotation of genes under positive selection (d_N/d_S >1) between the three species of *Alviniconcha*. d_N/d_S_K : values of d_N/d_S between *A. kojimai* and other species; d_N/d_S_S : values of d_N/d_S between *A. strummeri* and other species; d_N/d_S_B : values of d_N/d_S between *A. boucheti* and other species.

orthogroup name	d_N/d_S_K	d_N/d_S_S	d_N/d_S_B	Likelihood	Length	Accession	E_value	GO - Biological process	Annotation
orthogroup_1125	1.03	0.39	0.72	2095.32*	582	Q6UXT9	1.86E-13	cellular lipid metabolic process	Protein ABHD15 [Homo sapiens]
orthogroup_157	1.80	0.23	0.40	1169.83*	897	Q08DI8	1.06E-57	mRNA pseudouridine synthesis/ negative regulation of translation	Pseudouridylate synthase 7 homolog [Bos taurus]
orthogroup_1931	1.59	0.24	0.00	326.35*	684	Q5NVF6	8.01E-48	dephosphorylation/ lysosome organization	Lysosomal acid phosphatase [Pongo abelii]
orthogroup_1981	3.70	1.80	1.01	160.26*	585	Q9UPW5	4.13E-18	adult walking behavior/ neuromuscular process/ retina development	Cytosolic carboxypeptidase 1 [Homo sapiens]
orthogroup_2158	1.30	0.59	0.95	7.15*	1044	Q8CHN3	1.97E-06	antibacterial humoral response/ innate immune response/ spermatogenesis	WAP four-disulfide core domain protein 2 [Rattus norvegicus]
orthogroup_2302	0.33	1.01	0.09	255.90*	477	Q28DS0	1.83E-56	positive regulation of protein sumoylation	SUMO-activating enzyme subunit 1 [Xenopus tropicalis]
orthogroup_2395	1.70	0.00	0.53	528.91*	534	Q17DK5	8.34E-73	blue light signaling pathway/ circadian rhythm regulation of gene expression/	Cryptochrome-1 [Aedes aegypti]
orthogroup_2448	0.00	0.00	1.12	4.07	465	O77487	5.34E-07	carbohydrate metabolic process/ protein glycosylation	Galactoside alpha-(1,2)-fucosyltransferase 2
orthogroup_2477	0.17	0.00	1.17	3.08	723	Q9Y3B3	4.96E-30	golgi organization	Transmembrane emp24 domain-containing protein 7
orthogroup_2565	1.29	0.12	0.00	141.02*	624	Q24747	2.63E-05	anterior/posterior axis specification/ embryonic development	Maternal protein exuperantia [Drosophila virilis]

orthogroup_274	1.45	1.24	2.11	1364.79*	615	F1QGD2	3.54E-24	peptidyl-L-cysteine S-palmitoylation	Probable palmitoyltransferase ZDHHC24 [Danio rerio]
orthogroup_2786	0.00	1.22	0.60	810.36*	714	Q8BFR1	3.06E-11	innate immune response/ defense response to virus	Zinc finger CCCH-type antiviral protein 1-like [Mus musculus]
orthogroup_2910	1.24	0.00	0.00	746.29*	870	Q6P1S4	6.18E-133	arginyl-tRNA aminoacylation	Arginine-tRNA ligase, cytoplasmic [Xenopus tropicalis]
orthogroup_2933	0.00	1.55	0.33	698.63*	732	Q8BIZ6	1.84E-53	mRNA splicing/ regulation of transcription by RNA polyperase II	Smad nuclear-interacting protein 1 [Mus musculus]
orthogroup_2985	0.78	1.17	0.43	375.20*	921	Q5F3Z9	7.93E-25	regulation of mRNA stability	Zinc finger CCCH domain-containing protein 14 [Gallus gallus]
orthogroup_3017	0.34	3.23	0.94	415.78*	513	Q2KJ63	1.78E-09	blood coagulation	Plasma kallikrein [Bos taurus]
orthogroup_31	1.58	0.37	2.08	0.89	687	Q6DJM2	1.83E-80	CDP-diacylglycerol biosynthetic process	Phosphatidate cytidylyltransferase, mitochondrial [Xenopus laevis]
orthogroup_3232	0.49	1.06	0.34	65.84*	477	O35403	1.65E-10	sulfation	Amine sulfotransferase [Mus musculus]
orthogroup_3297	1.23	0.32	0.77		507	Q99K82	1.13e-05	spermine catabolic process	Spermine oxidase [Mus musculus]
orthogroup_3345	2.50	2.63	0.98	627.19*	537	P18713	8.94E-34	transcription regulation/ metal ion binding	Gastrula zinc finger protein XlCGF17.1 [Xenopus laevis]
orthogroup_3480	1226	1234	0.78	231.83*	1065	P28336	5.06e-06	phospholipase C-activating G protein-coupled receptor signaling pathway	Neuromedin-B receptor [Homo sapiens]
orthogroup_3738	0.27	0.82	1.02	1.28	489	Q7TS68	2.01e-22	RNA methylation	tRNA (cytosine(72)-C(5))-methyltransferase NSUN6 [Mus musculus]

orthogroup_3834	1.26	0.00	0.00	15.55*	657	Q69ZU8	2.55E-05	signaling pathway	Ankyrin repeat domain-containing protein 6 [Mus musculus]
orthogroup_3993	0.64	1.07	0.31	57.86*	747	A4IG59	2.22E-09	endosomal transport/ exocytosis/ regulation of autophagy	WASH complex subunit 1 [Danio rerio]
orthogroup_4102	0.00	1.17	0.06	373.61*	591	Q5BJS7	1.41E-13	cell differentiation/ cellular response to calcium ion	Copine-9 [Rattus norvegicus]
orthogroup_4174	0.22	2.63	0.17	4311.23*	1692	Q6AYG0	9.66E-10	methylation	Methyltransferase-like protein 25B [Rattus norvegicus]
orthogroup_4319	0.00	0.15	4.77	16.32*	855	Q9BXC9	5.35E-67	adult behavior/ hippocampus development	Bardet-Biedl syndrome 2 protein [Homo sapiens]
orthogroup_4454	1.42	0.00	0.51	200.64*	486	P55265	5.54E-38	innate immune response/ regulation of apoptotic process	Double-stranded RNA-specific adenosine deaminase [Homo sapiens]
orthogroup_4470	1.22	0.32	0.11	1131.41*	498	Q922Q1	1.79E-31	detoxification of nitrogen compound/ nitrate metabolic process	Mitochondrial amidoxime reducing component 2 [Mus musculus]
orthogroup_4581	0.54	0.34	1.20	424.58*	528	Q9W539	4.39E-38	cellular response to hypoxia/ regulation of developmental growth	Hormone receptor 4 [Drosophila melanogaster]
orthogroup_4681	1015.17	0.43	0.19	41.92*	660	A8IW99	1.36e-51	DNA methylation involved in gamete generation/ meiotic cell cycle/ spermatid development	Mitochondrial cardiolipin hydrolase [Chlamydomonas reinhardtii]
orthogroup_4846	0.25	0.21	1.29	3.26	564	Q9Y2L5	3.00e-21	golgi organization	Trafficking protein particle complex subunit 8 [Homo sapiens]
orthogroup_4959	1.45	0.37	0.18	123.34*	567	Q9Y2M2	3.76E-37	odontogenesis	Protein SSUH2 homolog [Homo sapiens]
orthogroup_5180	1.18	0.19	0.82	1128.73*	471	Q99LI2	2.43E-14	chloride transport	Chloride channel CLIC-like protein 1 [Mus musculus]

orthogroup_5255	0.00	0.79	1.05	0.09	501	P02215	9.88e-59	heart development/ oxygen transport	Myoglobin [Cerithidea rhizophorarum]
orthogroup_5332	0.00	0.00	1.43	5.12	600	O15078	6.71e-45	cilium assembly/ eye photoreceptor cell development	Centrosomal protein of 290 kDa [Homo sapiens]
orthogroup_5559	1.27	0.37	0.23	128.99*	555	Q7YT83	2.94E-24	calcium channel impairing toxin/ reurotoxin	Cysteine-rich venom protein [Conus textile]
orthogroup_567	0.00	0.15	1.51	9.04*	591	Q5BK01	5.46E-13	methylation	Methyltransferase-like protein 24 [Rattus norvegicus]
orthogroup_5737	1.06	0.84	0.45	358.17*	969	A6QLE6	1.27E-95	calcium activated phosphatidylcholine scrambling/ chlroride transport	Anoctamin-4 [Bos taurus]
orthogroup_5773	0.00	0.00	1.47	11.21*	543	Q9NPG3	7.01e-24	DNA replication	Ubinuclein-1 [Homo sapiens]
orthogroup_5929	0.00	1.84	0.00	464.98*	579	Q9TU53	4.51E-15	lipoprotein transport/ response to bacterium	Cubilin [Canis lupus familiaris]
orthogroup_5957	1.76	0.00	0.63	1907.74*	1182	P23116	6.69E-40	formation of cytoplasmic translation initiation complex	Eukaryotic translation initiation factor 3 subunit A [Mus musculus]
orthogroup_6131	1.28	0.23	0.05	315.79*	504	Q8WZ42	6.20E-10	cardiac muscle development/ muscle contraction/ skeletal muscle thin filament assembly	Titin [Homo sapiens]
orthogroup_6261	0.53	1.45	0.36	854.70*	480	Q9UHF1	5.31E-15	anatomical structure development/ angiogenesis/ cell adhesion	Epidermal growth factor-like protein 7 [Homo sapiens]
orthogroup_6401	1202	0.14	0.05	79.62*	537	Q17FB8	5.61e-66	positive regulation of collagen fibril organization	Glycosyltransferase 25 family member [Aedes aegypti]
orthogroup_6589	0.29	1.08	2.19	152.30*	693	Q9HB15	2.65E-52	potassium ion transmembrane transport/ stabilization of membrane potential	Potassium channel subfamily K member 12 [Homo sapiens]
orthogroup_6975	0.00	1.85	0.10	6.42*	486	P55265	5.54e-38	innate immune response/ cellular response to virus/	Double-stranded RNA-specific adenosine deaminase [Homo sapiens]

orthogroup_7059	926	1.11	0.31	8.31*	696	O16374	1.07e-10	acute inflammatory response/ development of secondary sexual characteristics/ positive regulation of apoptotic process	Beta-1,4-galactosyltransferase galt-1 [Caenorhabditis elegans]
orthogroup_7152	1.57	0.00	0.32	159.82*	564	Q5SUE7	5.32E-06	adenosine to inosine editing/ RNA processing	Adenosine deaminase domain-containing protein 1 [Mus musculus]
orthogroup_7352	1.13	0.91	1.06	1266.60*	456	Q9Y2C9	1.23E-05	cellular response to bacteria/ immune response/ inflammatory response	Toll-like receptor 6 [Homo sapiens]
orthogroup_7409	0.52	1.13	0.16	289.78*	651	Q8BJ90	4.96E-09	regulation of transcription by RNA polymerase II	Zinc finger protein 771 [Mus musculus]
orthogroup_7498	2.16	0.38	1.59	1769.24*	510	Q554Z5	5.35E-05	ketone body biosynthetic process	Acyl-CoA synthetase short-chain family member B, mitochondrial [Dictyostelium discoideum]
orthogroup_7587	0.37	0.12	1.37	6.35*	525	V5NAL9	2.07e-17	innate immune response/ defense response to bacterium	Toll-like receptor 4 [Pinctada imbricata]
orthogroup_7595	0.31	1.54	0.26	181.48*	645	Q6QI06	2.81E-21	actin cytoskeleton organization/ embryo development/ positive regulation of TOR signaling	Rapamycin-insensitive companion of mTOR [Mus musculus]
orthogroup_779	1.52	0.00	0.14	1320.96*	921	O54904	1.12E-53	galactosylceramide biosynthetic process/ lipid glycosylation	Beta-1,3-galactosyltransferase 1 [Mus musculus]
orthogroup_783	0.24	0.18	2.81	5.62	531	B0G124	7.39E-04	response to salt stress	Ankyrin repeat-containing protein DDB_G0279043 [Dictyostelium discoideum]
orthogroup_7901	0.32	1.42	0.52	4407.34*	453	P24721	4.66E-19	bone mineralization/ cell surface receptor signaling pathway/ glycoprotein metabolic process	Asialoglycoprotein receptor 2 [Mus musculus]

orthogroup_8084	0.21	6.46	0.67	187.90*	816	P62291	2.58e-08	developmental growth/ male gonad development/ oogenesis/ spermatogenesis	Abnormal spindle-like microcephaly-associated protein homolog [Macaca fascicularis]
orthogroup_8179	0.18	1.06	0.18	3297.62*	723	O88643	1.27E-113	cellular response to DNA damage/ chromatin remodeling/ response to hypoxia	Serine/threonine-protein kinase PAK 1 [Mus musculus]
orthogroup_8200	0.31	0.42	1.15	38.32*	1053	C7G0B5	3.01E-06	biomineral-forming organisms/ chitin binding	Protein PIF [Pinctada fucata]
orthogroup_8201	0.00	1.11	0.31	471.60*	534	Q5PQ50	5.76E-28	butyryl-CoA catabolic process/ coenzyme A catabolic process	Nucleoside diphosphate-linked moiety X motif 19 [Xenopus laevis]
orthogroup_8406	0.28	1.57	0.82	394.73*	462	Q3SZ71	2.04E-75	mitochondrial calcium ion transmembrane transport	Mitochondrial-processing peptidase subunit beta [Bos taurus]
orthogroup_8441	0.38	1.13	0.24	810.70*	867	A4FV97	3.44E-37	maturation of LSU-rRNA/ osteoblast differentiation/ regulation of cellular senescence	Ribosomal L1 domain-containing protein 1 [Bos taurus]
orthogroup_8468	0.50	1.05	0.33	1258.57*	684	C4XZ24	5.70E-28	cell-cell adhesion/ filamentous growth/ flocculation	Flocculation protein FLO11 [Clavispora lusitaniae ATCC 42720]
orthogroup_8483	0.60	1.11	0.92	8940.82*	543	Q9W0A0	0.000129	defense response to bacterium/ apoptotic cell clearance/ phagocytosis/ salivary gland cell autophagic cell death	Protein draper [Drosophila melanogaster]
orthogroup_8498	0.17	0.00	1.03	5.27	504	P0C7A1	1.31e-21	protein deglycosylation	Cytosolic endo-beta-N-acetylglucosaminidase [Gallus gallus]
orthogroup_8537	0.00	0.16	1.31	8.39*	582	Q6GPE5	8.12e-05	apoptotic process/ DNA damage checkpoint	Protein fem-1 homolog B [Xenopus laevis]
orthogroup_8610	1.51	1.58	0.89	120.08*	612	Q80VN0	1.17E-24	cilium movement/ regulation of microtubule motor activity	Cilia- and flagella-associated protein 100 [Mus musculus]

orthogroup_8621	0.00	0.00	1.87	10.57*	456	Q5RH51	1.70e-43	proteoglycan biosynthetic process	Glycosaminoglycan xylosylkinase [Danio rerio]
orthogroup_8757	0.32	1.26	0.39	201.99*	594	Q96RW7	0.000006	cell cycle/ cell division/ response to bacterium/ cell-cell adhesion	Hemicentin-1 [Homo sapiens]
orthogroup_8837	0.00	0.00	1.35	5.57	684	G5CTG8	3.01e-17	cellular water homeostasis/ oxygen homeostatis	Aquaporin-11 [Milnesium tardigradum]
orthogroup_8861	0.56	1.50	3.15	1040.80*	582	P18700	3.94E-106	cell division/ microtubule cytoskeleton organization	Tubulin beta chain [Strongylocentrotus purpuratus]
orthogroup_893	3.39	3.87	2.81	373.07*	666	P27393	0.000000119	angiogenesis/ endodermal cell differentiation/ extracellular matrix organization	Collagen alpha-2(IV) chain [Ascaris suum]
orthogroup_9001	1.78	0.48	0.45	277.77*	612	Q9UQ90	8.12E-99	nervous system development/ proteolysis/ cristae mitochondrial formation	Paraplegin [Homo sapiens]
orthogroup_9010	3.24	0.00	0.33	248.92*	771	O61470	1.75E-129	regulation of protein catabolic process/ ubiquitin-dependent protein catabolic process	Probable 26S proteasome non-ATPase regulatory subunit 3 [Anopheles gambiae]
orthogroup_9034	2.26	0.52	0.86	4.59	753	O89110	1.01e-38	regulation of innate immune response/ apoptotic process/ B cell activation	Caspase-8 [Mus musculus]
orthogroup_9230	0.69	1.41	0.00	1821.98*	942	Q2HJ53	1.49E-12	intracellular signal transduction/ regulation of cell growth	Cytokine-inducible SH2-containing protein [Bos taurus]
orthogroup_9292	0.49	1.60	0.44	506.45*	678	P11928	3.33E-14	defense response to bacterium-virus/ toll-like receptor signaling pathway	2'-5'-oligoadenylate synthase 1A [Mus musculus]
orthogroup_9299	1.02	0.56	0.38	1398.24*	525	Q9NXT0	6.69E-32	regulation of transcription by RNA polymerase II	Zinc finger protein 586 [Homo sapiens]
orthogroup_9357	1.84	0.23	0.71	66.59*	456	Q27433	7.84E-48	detection of mechanical stimulus/ mechanosensory behavior/ response to mechanical stimulus	Mechanosensory protein 2 [Caenorhabditis elegans]

orthogroup_9367	1.38	0.00	0.00	499.27*	630	Q7TM99	1.29E-06	monocarboxylic acid transport	Monocarboxylate transporter 9 [Mus musculus]
orthogroup_992	1.30	1.09	0.70	1482.78*	510	P13277	4.02E-43	cysteine-type peptidase activity	Digestive cysteine proteinase 1[Homarus americanus]

4- Discussion

Alviniconcha gastropods are foundation species and ecosystem engineers at deep-sea hydrothermal vents of the Western Pacific. In the Southwest Pacific, three species (*A. kojimai, A. strummeri* and *A. boucheti*) are living in sympatry at vent fields of at least two distinct back-arc basins and the volcanic arc of Futuna but had a long history of separate evolution on different ridge systems. Previous molecular dating studies indeed proposed that *A. boucheti* separated from other species about 48 Mya and that the ancestral population of *A. kojimai* and *A. strummeri* splitted in allopatry 25 Mya (Breusing et al., 2020; Castel et al. submitted). It is currently recognized that these species do not exactly share the same ecological niche resulting from host-specific symbiotic composition (Beinart et al., 2012; Breusing et al., 2020). *A. boucheti* harbours symbiotic *Campylobacteria* in its gills and is often found in close association to the wall of vent chimneys whereas *A. kojimai* and *A. strummeri* are more likely associated with diffuse venting areas with lower H_2S and H_2 concentrations and exhibit *Gamma-proteobacteria* (Beinart et al., 2012, 2019). According to these latter authors, differences in the environmental tolerance of symbiont species could be at the origin of their speciation (disruptive speciation; Breusing et al., 2020).

To test this hypothesis, we evaluated the role of diversifying selection on gene divergence between our three gastropod species. First, we estimated the proportion of fast-evolving genes implicated in the divergence of our three species and looked at their functional annotation in order to see if they could be involved in reproductive isolation (i.e. gamete recognition, partner attraction, mito-nuclear interactions; Palumbi, 2009; Pante et al., 2019). These functions on gamete recognition that act on reproductive isolation within allopatric populations may accelerate their divergence if gamete adaptations in the separated populations reduce cross-fertilization. Shifting reproductive traits by reinforcement may play a diversifying role when previously allopatric populations join (Palumbi, 2009).

In fast-evolving genes, positive selection may be erased when divergence increases with time, at least until genes reach saturation at synonymous sites, both because of progressive purge of deleterious polymorphic non-synonymous mutations during the allelic sorting process and the accompanying divergence formation and the protein-level selection that produces an underestimation of the non-synonymous rate due to the accumulation of very different amino-acid residues between species (i.e amino-acid changes with more than two

substitutions at a codon site are not considered in the rate calculation) (dos Reis and Yang, 2013). Here, the most divergent genes were not those under positive selection and mainly encoded for functions such as DNA/RNA repairs, cell-to-cell recognition/adhesion and cytoskeleton organisation. We were therefore unable to link these genes to an ecological niche specialisation that would have emerged from the gain of specific symbiotic associations. Even if a large proportion of fast-evolving genes can be associated with the DNA repair system and methylation process in *A. boucheti*, which may be a way to account for higher fluid toxicity, temperature or hydrostatic pressure, this does not fit very well with the concept of ecological speciation based on symbiosis. This hypothesis is well supported by demographic models showing an allopatric separation of species followed by a long period of isolation before secondary contact (SC) (Castel et al. submitted). Indeed, in the case of ecological speciation in sympatry, we rather expect a demographic pattern of isolation with migration (IM or AM). This result still needs to be confirmed by studying the intermediate species of *Alviniconcha*. It would certainly be very interesting to carry out similar studies between the sister species *A. kojimai*/*A. hessleri* and *A. boucheti*/*A. marisindica*.

Second, we also investigated the number and function of positively-selected genes within each species to apprehend whether allopatric speciation may have been reinforced by local adaptation issued from the different symbiotic associations and associated fluid chemistries. The idea beyond was indeed to evaluate whether positive selection occurred recently (genes weakly divergent with high d_N) and targeted more specifically genes that could be linked with the symbiotic associations (i.e. innate immunity and bacterial interactions, environmental stress responses).

Only about 2% of the genes were under positive selection (d_N/d_S >1) across all species. However, when taking into consideration non-synonymous diverging genes with a d_S near zero, about 5% of the genes were under selection on the branches leading to *A. kojimai* and *A. strummeri* species. The lower proportion of genes under selection in *A. boucheti* may be explained by the time elapsed since the separation of *A. boucheti* from the two other species, which is much longer than the time since the two later species separated. The number of positively-selected genes between *A. kojimai* and *A. strummeri* is however lower than numbers usually depicted for species that emerged from ecological speciation, such as for instance *Sebastes caurinus* and *S. rastrelliger* (marine fishes), for which 7% of genes have d_N/d_S values greater than 1. However, compared to some studies (for example 0.5% of genes under selection in Tilapia (Xiao et al., 2015) and 1% of genes in copepods (Barreto et al.,

2011) the proportion of genes under positive selection obtained in this study is non-negligible. These genes under positive selection are for the most part potentially involved in reproductive isolation between *A. kojimai* and *A. strummeri*, with annotations relating to spermiogenesis, gamete recognition (sperm-pellucida zona interactions), and gonad development, but also chemosensory by the acyl or acetyl CoA cycle from which by-products are often related to the male-female attraction or related with the immune system. After species met again and were able to hybridise, these specific adaptations are likely to promote pre-zygotic barriers and, together with mitonuclear incompatibilities and cell division/embryo development could act in the maintenance of species by reinforcing post-zygotic barriers. Indeed, post-zygotic barriers due to a long history of divergence in allopatry such as maladaptation (Hauser, 2002) and sterility or inviability of hybrids (Brannock and Hilbish, 2010; Maheshwari and Barbash, 2011) are likely to play a role in maintaining divergence between species. In other organisms, these genes under selection related to spermatogenesis are also found, such as between men and chimpanzees (Nielsen et al., 2005).

For *A. boucheti*, the fact that genes associated with the immune system and stress responses including hypoxia are under selection represent a strong argument towards an 'old' environmental selection in the speciation processes that separated *A. boucheti* from the two other species. Indeed, these species developed specific interactions with different types of symbionts (therefore a specific immune system) which led in turn to possible local adaptations and differing ecological niches. These genes can have contributed in exacerbating population divergence between species during their first separation in allopatry (geographical isolates) and the very long period of isolation. This period allowed species to adapt to their environment (including the acquisition of different symbionts leading to a difference in ecological niche). Moreover, evidence of positive selection on immune-related genes supports the hypothesis that a large proportion of genes under selection in genomes may be due to a coevolutionary arms race between the host immune system and pathogens (Nielsen et al., 2005).

The average values of d_N/d_S found along the branches leading to *Alviniconcha* species ranged from 0.13 to 0.17, and a little lower than the whole genome estimates observed in many taxa around 0.2 (Gibbs and Pachter, 2004; Mikkelsen, 2011; Xiao et al., 2015). Previous studies have shown that on average 70% of mutations are deleterious, 25% are mildly deleterious or neutral and 2-4% of mutations are under positive selection (Eyre-Walker et al.,

2002; Wright et al., 2005). In this study, about 90% of the genes displayed d_N/d_S lower than 0.5 and are therefore under purifying selection. This high proportion of genes under purifying selection is likely the consequence of a primary adaptation to extreme environmental conditions encountered by gastropods, as previously reported by Fontanillas et al., (2017). Indeed, even if *Alviniconcha* gastropods do not live under extreme temperatures (from 7 to 42°C; Podowski et al., 2009, 2010), hydrostatic pressure conditions reflecting depth directly affect protein stability at the scale of the whole proteome (Fontanillas et al., 2017). In addition, living under sulfidic and highly hypoxic conditions have also strong implications on respiration and most aerobic metabolisms, and surely the mitochondrial respiration (O'Brien et al., 1991; Powell and Somero, 1986).

For each species pair the distribution of divergences was clearly unimodal with *t* values significantly different from zero, indicating that the species separation is done. In fact, in species with secondary contacts, we would expect to observe genes with a divergence of 0 between species. This observation confirms the presence of distinct species where lineage sorting is finite (Castel et al. submitted). In this study, compared to the predicted secondary contact demographic model (Castel et al., submitted), we did not identify any introgressed genes between species (which would be characterised by divergence $t = 0$). This may be explained by a low introgression on the nuclear genome between species and the drastic under-sampling of genes analysed in this study (only 1704 genes).

Indeed, in this analysis, only 10% of the genes were analysed on the transcripts identified by species (1 704 orthologous CDS on average 22 000 transcripts per species). This drastic reduction of the dataset may have a strong influence on the sampling of genes with rapid evolutionary rates that are less represented because of their divergence. To overcome this bias, more individuals should be sequenced in order to be able to compare not transcriptomes by species but individuals between them. Moreover, this would allow us to evaluate the polymorphism between individuals of the same species.

5- Conclusions

In this study based on three species of *Alviniconcha* in the Western Pacific Ocean, the objective was to identify the role of divergent selection on interspecies divergence. The study of fast-evolving genes showed molecular functions involved in DNA/RNA repair, cell adhesion and cytoskeleton organisation, thus suggesting no direct impact on recent niche specialisation between species, in line with previous studies supporting allopatric speciation

(Castel et al. submitted). This study subsequently focused on the analysis of genes under positive selection. In *A. boucheti*, the genes under selection are mainly related to the immune system and stress responses (such as hypoxia). It is known that these species have different symbionts that lead to a different ecological niche, closer to the fluid emissions. This suggests an "old" environmental selection in this species, which may have contributed in exacerbating population divergence during their first separation in allopatry. For *A. kojimai* and *A. strummeri*, the majority of genes under selection are immune-related and related to spermatogenesis, gamete recognition, gonad development, or male-female attraction. Thus, after the species have met again (secondary contacts) and may hybridise, these specific adaptations are likely to promote pre-zygotic barriers and, along with mitonuclear incompatibilities and embryonic development, may act in maintaining species by reinforcing post-zygotic barriers.

Supplementary material of Chapter 4:

Table S9: Specimens used for this analysis

Individuals	Species	Site	Bassin	Expedition	Sample collector
SH092-269-L22	*A. boucheti*	ABE	Lau	Lau2009	S. Hourdez
SH092-167-L10	*A. boucheti*	Tow Cam	Lau	Lau2009	S. Hourdez
SH092-172-L12	*A. boucheti*	Tow Cam	Lau	Lau2009	S. Hourdez
SH092-600-L37	*A. kojimai*	Tu'i Malila	Lau	Lau2009	S. Hourdez
SH092-603-L30	*A. kojimai*	Tu'i Malila	Lau	Lau2009	S. Hourdez
SH092-991-L63	*A. strummeri*	Tu'i Malila	Lau	Lau2009	S. Hourdez
SH092-651-L68	*A. strummeri*	Tu'i Malila	Lau	Lau2009	S. Hourdez

Table S10: BUSCO analysis on assembled transcripts on Eukaryota and Mollusc taxa

		A. boucheti	*A. kojimai*	*A. strummeri*
For Eucarya				
	Complete BUSCOs (C)	246	246	248
	Complete and single-copy BUSCOs (S)	142	120	136
	Complete and duplicated BUSCOs (D)	104	126	112
	Fragmented BUSCOs (F)	6	5	4
	Missing BUSCOs (M)	3	4	3
	Total BUSCO groups searched	255	255	255
For Mollusca				
	Complete BUSCOs (C)	4285	4459	4566
	Complete and single-copy BUSCOs (S)	2100	1930	1899
	Complete and duplicated BUSCOs (D)	2185	2529	2667
	Fragmented BUSCOs (F)	123	102	103
	Missing BUSCOs (M)	887	734	626
	Total BUSCO groups searched	5295	5295	5295

Table S11: Annotation of outliers divergent genes (1%) by pair of species of *Alviniconcha*.

orthogroup name	Divergence *kojimai/ strummeri*	Divergence *kojimai/ boucheti*	Divergence *boucheti/ strummeri*	Length	Accession	E-value	GO- Biological process	Annotation
orthogroup_1063	0.470017	0.180052	0.523265	468	Q9V3R8	2.12E-60	larval lymph gland hemopoiesis/ mitochondrion morphogenesis	UbiA prenyltransferase domain-containing protein 1 homolog [Drosophila melanogaster]
orthogroup_1125	0.171488	0.205296	0.323524	1008	Q10003	3.21E-34	glycerophospholipid catabolic process	Putative glycerophosphocholine phosphodiesterase GPCPD1 homolog 2 [Caenorhabditis elegans]
orthogroup_2361	0.018582	0.371513	0.364215	510	B7PRF6	6.33E-75	methionine metabolic process	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase [Ixodes scapularis]
orthogroup_274	0.625004	0.552414	0.449818	711	P16157	3.50E-42	cytoskeleton organization/ exocytosis/ signal transduction	Ankyrin-1 [Homo sapiens]
orthogroup_2910	0.686845	0.704882	0.066445	462	Q3SZ71	2.04E-75	mitochondrial calcium ion transmembrane transport	Mitochondrial-processing peptidase subunit beta [Bos taurus]
orthogroup_30	0.206554	0.385947	0.489315	903	P06603	0	microtubule-based process/ skeleton organization	Tubulin alpha-1 chain [Drosophila melanogaster]
orthogroup_31	0.20291	0.148394	0.161192	594	Q96RW7	6.00E-06	cell cycle/ cell division/ visual perception	Hemicentin-1 [Homo sapiens]
orthogroup_3345	0.071581	0.254413	0.254832	678	P11928	3.33E-14	antiviral innate immune response/ toll-like receptor signaling pathway	2'-5'-oligoadenylate synthase 1A [Mus musculus]

orthogroup_3374	0.132543	0.270687	0.241684	456	P18737	9.07E-14	DNA binding/ metal ion binding	Gastrula zinc finger protein XlCGF8.2DB [Xenopus laevis]
orthogroup_358	0.060709	0.696722	0.687067	456	Q27433	7.84E-48	mechanosensory behavior/ response to mechanical stimulus	Mechanosensory protein 2 [Caenorhabditis elegans]
orthogroup_365	0.155106	0.331632	0.246258	546	P06603	2.72E-116	microtubule-based process/ skeleton organization	Tubulin alpha-1 chain [Drosophila melanogaster]
orthogroup_380	0.204513	0.314382	0.383155	453	P24721	4.66E-19	bone mineralization/ cell surface receptor signaling pathway	Asialoglycoprotein receptor 2 [Mus musculus]
orthogroup_432	0.313542	0.344838	0.28427	681	Q0KL02	1.35E-40	central nervous system development/ axon guidance	Triple functional domain protein [Mus musculus]
orthogroup_4965	0.065545	0.626075	0.626848	510	Q9BVR0	3.17E-12	metal ion binding	Putative HERC2-like protein 3 [Homo sapiens]
orthogroup_5797	0.361434	0.436535	0.492121	1602	A2VEC9	7.34E-04	cell differentiation/ nervous system development	SCO-spondin [Homo sapiens]
orthogroup_6665	0.434399	0.12978	0.401877	741	Q5VST9	3.34E-20	cell-cell adhesion/ ankyrin binding	Obscurin [Homo sapiens]
orthogroup_6835	0.174335	0.264773	0.203846	768	B9U3F2	2.27E-08	cell differentiation/ cholesterol metabolic process/ lipid transport	Protein dispatched homolog 3 [Gallus gallus]
orthogroup_7016	0.239549	0.232909	0.222362	633	Q6NU04	3.63E-07	lens fiber cell differentiation/ lens morphogenesis in camera-type eye/ spermatogenesis/	Tudor domain-containing protein 7 [Xenopus laevis]
orthogroup_735	0.298031	0.572841	0.342118	453	P02556	4.59E-109	cell division/ cellular process/ cytoskeleton organization	Tubulin beta chain [Lytechinus pictus]
orthogroup_77	0.791585	0.075311	0.71629	456	Q9W6K1	4.33E-54	DNA damage stimulus/ DNA repair	Double-strand break repair protein MRE11 [Xenopus laevis]
orthogroup_7843	0.34658	0.060957	0.351115	570	Q498K0	2.76E-55	galactosylceramide process/ myelination	Galactocerebrosidase [Xenopus laevis]

orthogroup_8861	0.539221	0.420457	0.434988	615	P22105	2.52E-08	actin cytoskeleton organization/ cell adhesion/ collagen metabolic process	Tenascin-X [Homo sapiens]
orthogroup_893	0.44541	0.457212	0.545518	528	P16157	7.20E-31	cytoskeleton organization/ exocytosis/ signal transduction	Ankyrin-1 [Homo sapiens]
orthogroup_9133	0.159411	0.306849	0.216732	543	P18713	1.06E-36	DNA binding/ metal ion binding	Gastrula zinc finger protein XlCGF8.2DB [Xenopus laevis]
orthogroup_979	0.875875	0.84551	0.109237	606	Q9GKW3	5.59E-107	epithelial cell differentiation/ glycolytic process	Fructose-bisphosphate aldolase C [Macaca fascicularis]
orthogroup_992	0.416853	0.452652	0.560733	684	C4XZ24	5.70E-28	cell-cell adhesion/ flocculation/ biofilm formation	Flocculation protein FLO11 [Clavispora lusitaniae ATCC 42720]

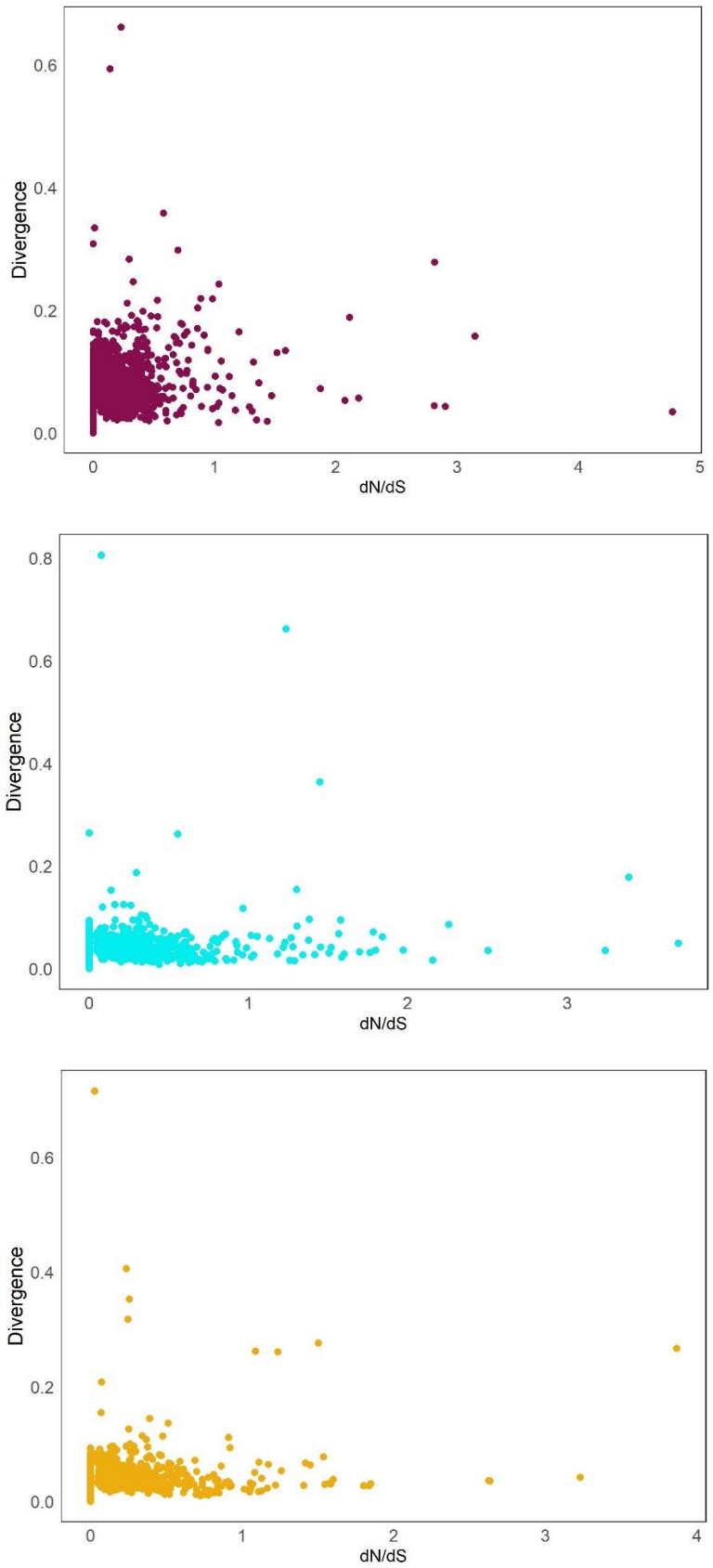


Figure S13 : d_N/d_S and divergence for the 1 704 genes identified for *A. boucheti* (in purple), *A. kojimai* (in turquoise) and *A. strummeri* (in yellow).

Discussion générale et Conclusions

1- L'océan Pacifique Sud-Ouest : un lieu de rencontre entre trois espèces d'*Alviniconcha*

L'océan Pacifique Sud-Ouest est composé entre autres de quatre bassins arrière-arc (bassins de Manus, Woodlark, Nord-Fidjien et Lau) ainsi que de la zone volcanique active de Futuna. De précédentes études ont pu montrer la présence de trois espèces d'*Alviniconcha* dans cette zone géographique, définies principalement sur la base du gène mitochondrial *Cox1* (Breusing et al., 2020 ; Johnson et al., 2015 ; Laming et al., 2020). La campagne d'échantillonnage CHUBACARC (Connectivité et Histoire des commUnautés hydrothermales des BAssins/volCans arrière-ARC du Pacifique) menée en 2019 a permis d'améliorer les connaissances sur la distribution géographique de ces trois espèces par l'échantillonnage de sites actifs connus et la découverte et l'exploration faunistique de trois nouveaux sites hydrothermaux, La Scala dans le bassin de Woodlark (Boulart et al., 2022), Phoenix dans le bassin Nord-Fidjien et Mangatolo dans la partie la plus Nord du bassin de Lau. Avec la découverte de ces nouveaux sites, les aires de distribution de *A. kojimai* et *A. boucheti* ont été précisées par rapport à ce qui était connu précédemment. En effet, *A. kojimai* est distribuée sur l'ensemble des quatre bassins arrière-arc (Fig. 13 du chapitre Echantillonnage et acquisition des données) en étant l'espèce la plus abondante dans les bassins de Lau et Nord-Fidjien ainsi que sur la zone volcanique de Wallis et Futuna. L'espèce *A. boucheti* est présente sur l'ensemble de l'aire d'étude (dont le bassin de Woodlark ; Boulart et al., 2022) même si lors de la campagne CHUBACARC elle n'a pas été retrouvée dans le bassin Nord-Fidjien (Johnson et al., 2015). Les sites habituellement échantillonnés dans ce bassin (White Lady/ Ivory Tower) étant éteints, les prélèvements ont été réalisés sur un nouveau site nommé Phoenix proche d'un champ inactif de cheminées hydrothermales appelé 'Père Lachaise' (Auzende et al., 1989) au point triple ou l'espèce n'a pas été retrouvée. Deux hypothèses sont possibles, la première est que l'espèce était bien présente mais n'a pas été échantillonnée sur les zones d'émissions diffuses visitées ou que cette espèce n'a pas colonisée ce nouveau site hydrothermal. Des prélèvements supplémentaires sont donc nécessaires afin de confirmer la présence actuelle de *A. boucheti* dans le bassin Nord-Fidjien. Cette espèce est la plus abondante dans le bassin de Manus et notamment dans le champ hydrothermal Pacmanus où les autres espèces sont absentes. Ainsi il existe un gradient longitudinal d'abondance inversée entre *A. kojimai* et *A. boucheti*, peut être signe d'une expansion actuelle d'Ouest en Est pour *A. boucheti* alors qu'elle irait dans l'autre sens chez *A. kojimai*. A noter la présence de ces deux espèces en syntopie (au sein d'un même patch de

gastéropodes) dans quatre prélèvements (deux dans le bassin de Lau, un dans le bassin de Manus et un à Futuna). *A. strummeri*, espèce la moins abondante dans cette étude, possède l'aire de distribution la plus petite et est présente uniquement en syntopie avec *A. kojimai* dans les bassins de Lau et Nord-Fidjien ainsi que dans la zone active de Futuna.

2- Une longue période de divergence en allopatrie ...

Ces trois espèces ont donc des aires de distribution qui se recouvrent très largement (sympatrie). Mais elles présentent une forte divergence tant sur le génome mitochondrial (d_A : 8.6% sur le gène *Cox1* entre *A. kojimai* et *A. strummeri* et 12% entre *A. boucheti* et les autres espèces) que sur le génome nucléaire (ddRAD-seq et transcriptomes : d_A : 1.8% entre *A. kojimai* et *A. strummeri* et 3% entre *A. boucheti* et les autres espèces ; Chapitre 1). Ces valeurs concordent avec les niveaux de divergence observés entre des espèces reproductivement isolées. Par exemple, Roux et al. (2016) définissent une "zone grise de spéciation" entre 0.5% et 2% de divergence nucléaire nette (d_A) qui contient à la fois des paires de populations/espèces reproductivement isolées et d'autres qui au contraire sont sujettes à des flux de gènes hétérospécifiques. Au-delà de 2%, toutes les paires d'espèces étudiées montrent un isolement reproductif fort, dans ce cadre-là, les espèces d'*Alviniconcha* sont dans la limite haute de la zone grise de spéciation. Cependant comme les valeurs de divergences sont identiques entre les transcriptomes et le ddRAD-seq, les estimations sur le génome complet sont peut-être sous estimées à cause d'une sélection plus prononcée des SNPs dans les zones codantes du génome en raison de notre choix d'analyse sur les locus communs aux trois espèces. Ainsi, si cette hypothèse est vraie, sur la base d'une divergence supérieure à 2% on s'attend à un isolement reproductif complet entre les espèces d'*Alviniconcha* (Roux et al., 2016).

A ce titre, l'étude des gènes orthologues entre transcriptomes a permis de confirmer l'isolement génétique des 3 espèces avec une distribution gaussienne des divergences entre espèces ne contenant aucun gène avec une divergence proche de zéro, signe d'un tri de lignées complet entre les espèces. Ces résultats suggèrent ainsi une histoire ancienne de spéciation (> 9-10*Ne* generations) entre les trois espèces cibles d'*Alviniconcha* comme suggéré dans l'étude de Breusing et al. (2020). Dans cette étude basée sur une calibration fossile (séparation entre *Neptunea amianta* et *N. antiqua* il y a 33-37 Ma) et un taux de 0.0015 substitutions par million d'année (Ma) sur le génome mitochondrial, la séparation entre *A. boucheti* et les deux autres espèces serait estimée à 48 Ma alors qu'elle ne serait que de 25 Ma entre *A. kojimai* et

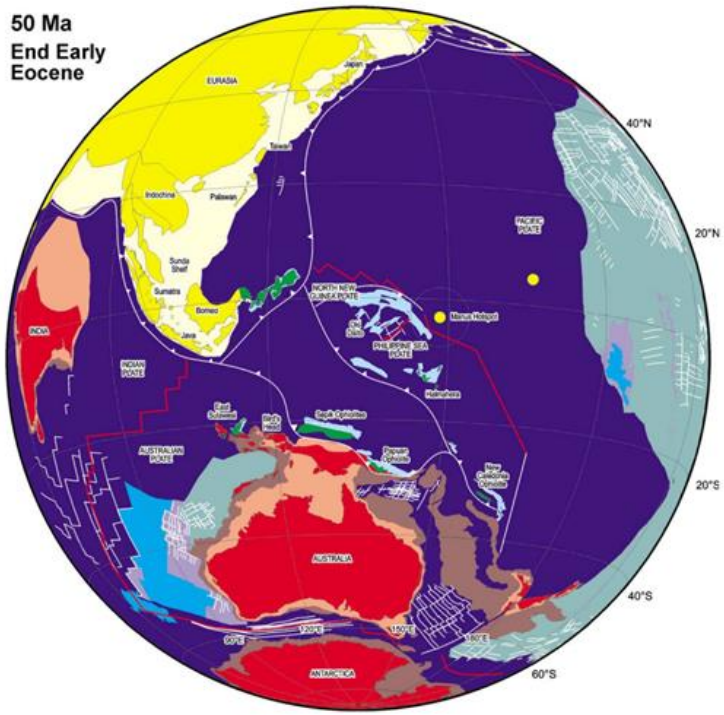
A. strummeri. La séparation entre *A. boucheti* et les deux espèces pourrait être la conséquence de la fermeture du passage entre l'océan Indien et l'océan Pacifique estimée entre 30 et 25 Ma lorsque la plaque australienne s'est subductée sous la plaque des Philippines, ce qui a crée une barrière physique à la dispersion au niveau du plateau Ontong-Java avec la formation et succession de plusieurs bassins arrière-arc (Figure 60 ; Breusing et al., 2020 ; Gaina et Müller, 2007 ; Hall, 2002). Malgré l'ouverture et la subduction/fossilisation de plusieurs dorsales arrière-arc dans cette zone géographique, l'ouverture et l'évolution des bassins arrière-arc actuels ne permettent pas d'expliquer l'origine de la spéciation entre *A. kojimai* et *A. strummeri* il y a 25 Ma (Breusing et al., 2020), *A. kojimai* ayant cependant une histoire récente de vicariance avec *A. hessleri* située dans le bassin des Mariannes.

La divergence génétique observée entre les trois espèces d'*Alviniconcha* se reflète aussi sur certains traits morphologiques qui permettent de les distinguer tout en renforçant l'hypothèse d'une spéciation ancienne. Une précédente étude a pu mettre en évidence des différences anatomiques sur la forme de la radula, du pied, de la columelle ainsi que sur l'ornementation de la coquille entre les trois espèces cibles d'*Alviniconcha* étudiées en sympatrie dans la zone volcanique de Futuna (Laming et al., 2020). Dans cette thèse, les deux derniers critères ont été analysés et complétés sur l'ensemble de l'aire de répartition des espèces et, comme montré en 2020 (Laming et al., 2020), ces critères permettent bien de différencier les espèces. Chez *A. strummeri*, des rangées uniquement de grands poils ont été observées. Ces grands poils sont intercalés par 1 à 3 petits poils chez *A. boucheti* et jusqu'à 6 petits poils chez *A. kojimai*. En ce qui concerne la columelle, elle est simple chez *A. boucheti* alors qu'elle est double chez *A. kojimai* et *A. strummeri*. A partir de ces résultats, des analyses complémentaires sont nécessaires pour chercher notamment s'il existe un patron d'arrangement des poils se répétant au sein des espèces et quel pourrait être le caractère plésiomorphe au genre. Pour cela une analyse sur l'ensemble de la coquille et sur toutes les espèces est nécessaire, contrairement à cette étude qui ne portait uniquement que sur un fragment marginal de la coquille. Les traits phénotypiques observés, comme l'arrangement des poils pourraient agir sur l'isolement reproducteur des espèces s'il existe un système de reconnaissance tactile des mâles et femelles d'une même espèce en l'absence de lumière au sein de ces conglomérats d'espèces. Cette hypothèse est d'autant plus soutenue que lors de l'analyse sur les transcriptomes, des gènes sous sélection ont pu être associé à des comportement mécano-sensoriel (Mechanosensory protein 2 ; Cytosolic carboxypeptidase 1). La reconnaissance pourrait s'accompagner d'autres traits phénotypiques n'ayant pas été

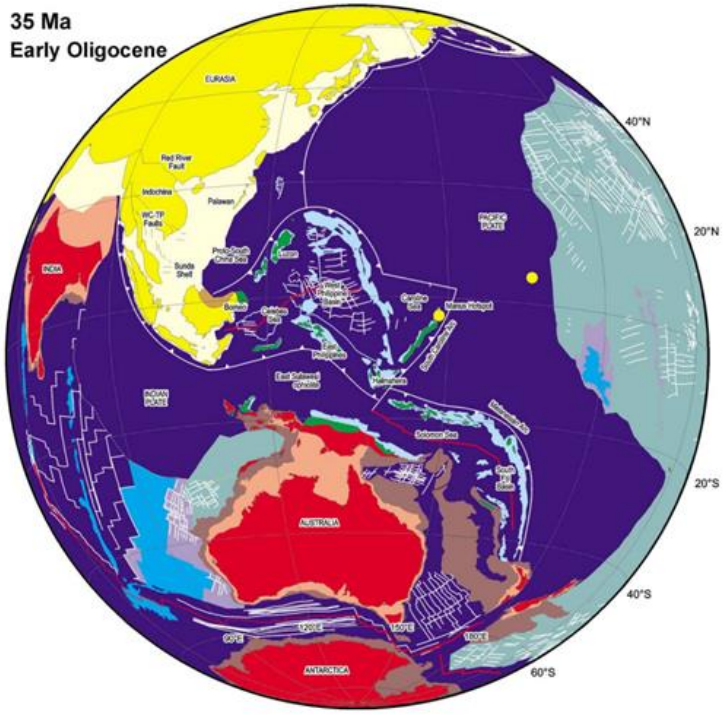
étudiés, comme les signaux olfactifs ou chimiques qui peuvent jouer un rôle très important dans la reconnaissance des espèces (Gabirot et al., 2010 ; Rafferty et Boughman, 2006). Une idée serait d'étudier la composition chimique du mucus libéré par chacune des espèces, en effet il a déjà été montré que cette composition varie d'une espèce à l'autre (Shaheen et al., 2005), ce qui pourrait jouer un rôle dans la reconnaissance des espèces entre elles et donc sur l'isolement reproducteur.

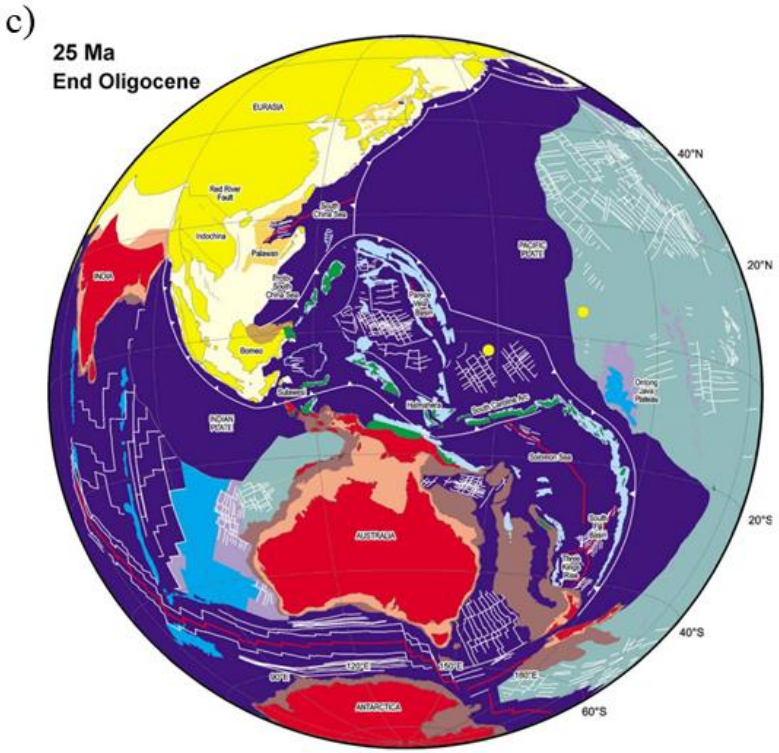
Dans cette thèse, l'étude sur la morphométrie des coquilles n'a pas permis de discriminer avec certitude les espèces, mais il serait sans doute intéressant de poursuivre ce travail avec des mesures complémentaires ou en faisant de la morphométrie 3D (Caro et al., 2019). Ce travail, différent des objectifs de cette thèse, permettrait de renforcer la diagnose entre des espèces encore difficile à différencier à l'œil nu.

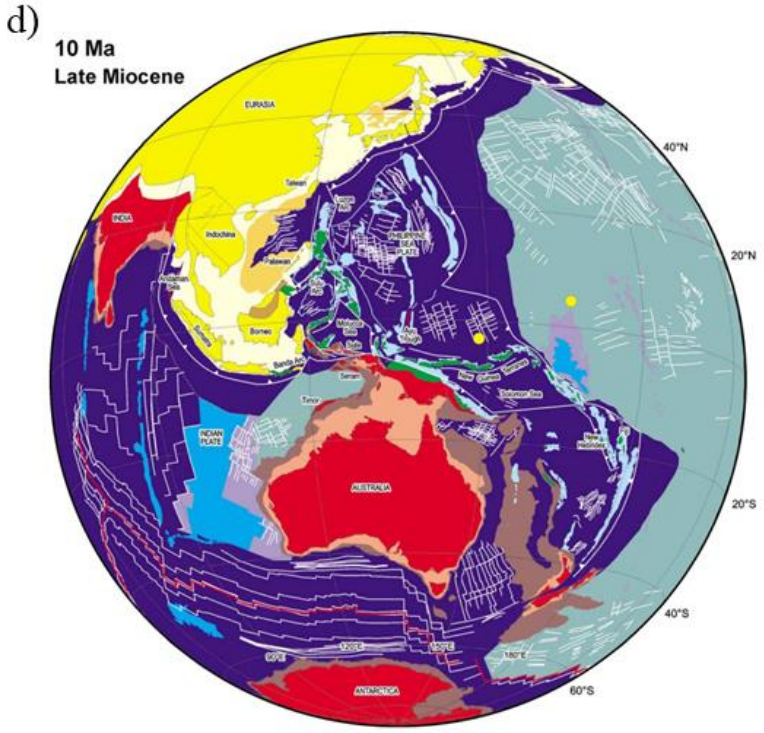
a)


50 Ma
End Early
Eocene
EURASIA
Japan
Taiwan
Indochine
Palawan
Sunda
Shelf
Sumatra
Borneo
Java
INDIA
INDIAN
PLATE
PACIFIC
PLATE
NORTH NEW
GUINEA PLATE
PHILIPPINE SEA
PLATE
Manus Hotspot
Halmahera
East
Sulawesi
AUSTRALIAN
PLATE
AUSTRALIA
ANTARCTICA
40°N
20°N
20°S
40°S
60°S
90°E
120°E
150°E
180°E

b)


35 Ma
Early Oligocene
EURASIA
Red River
Fault
Indochine
Palawan
WC TP
Faults
Sunda
Shelf
Borneo
INDIA
INDIAN
PLATE
PACIFIC
PLATE
Manus Hotspot
Halmahera
East Sulawesi
Celebes Sea
Solomon Sea
AUSTRALIA
ANTARCTICA
40°N
20°N
20°S
40°S
60°S
90°E
120°E
150°E
180°E


c)
25 Ma
End Oligocene
EURASIA
Red River Fault
Indochina
Palawan
INDIA
Borneo
INDIAN PLATE
AUSTRALIA
ANTARCTICA
40°N
20°N
20°S
40°S
60°S
60°E
120°E
180°E


d)
10 Ma
Late Miocene
EURASIA
Taiwan
INDIA
Indochina
Palawan
Borneo
INDIAN PLATE
AUSTRALIA
ANTARCTICA
40°N
20°N
20°S
40°S
60°S
60°E
120°E
180°E

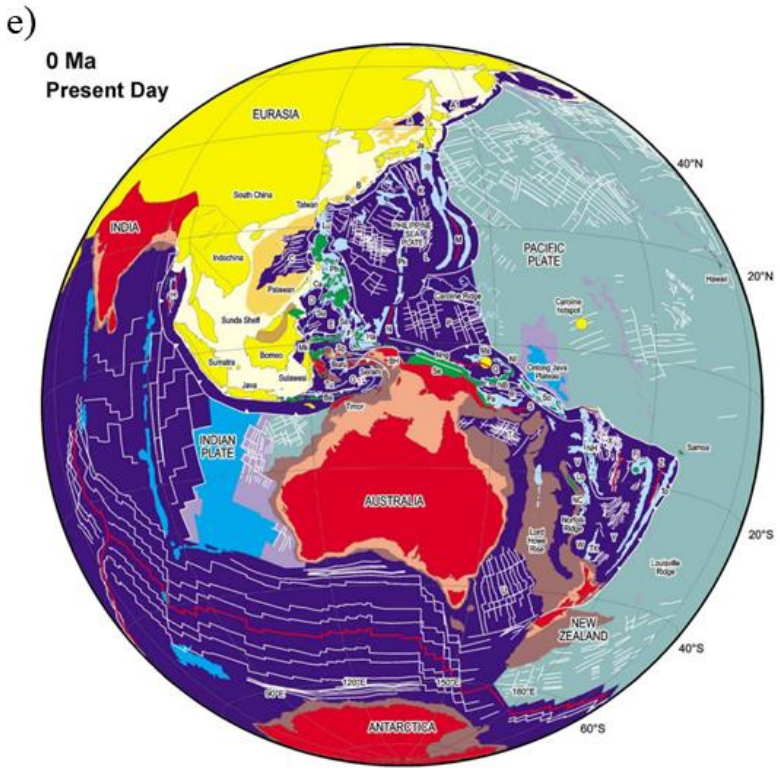


Figure 60: Mouvements tectoniques depuis 50 Ma au sein des océans Pacifique et Indien ayant conduit à la présence actuelle des bassins de Manus, Woodlark, Nord-Fidjien et Lau ainsi que l'arc volcanique de Futuna dans le Pacifique Sud-Ouest (Hall, 2002).

3- … suivie d'une remise en contact avec reprise partielle des flux de gènes entre espèces

Tous les résultats discutés jusqu'ici concordent donc sur l'hypothèse d'une divergence forte héritée d'une longue période de spéciation allopatrique. Cependant, l'utilisation d'une méthode d'inférence démographique (DILS) au cours de mon travail de thèse a produit un résultat surprenant. En effet, les analyses démographiques réalisées sur le polymorphisme partagé entre les 3 espèces ont permis de montrer une faible introgression actuelle entre les espèces et non un reste de polymorphisme ancestral encore non trié. En effet, les modèles démographiques qui s'ajustent le mieux aux données de ddRAD-seq, prédisent une longue période sans échanges génétiques (allopatrie) suivie d'un faible flux de gènes dû à un ou plusieurs contacts secondaires entre espèces (Chapitre 1). Généralement l'introgression concerne moins de 10% du génome des individus sauf très localement comme chez *A. strummeri*, où certains individus présentent jusqu'à 18% d'introgression dans le bassin de Lau. La reprise du flux de gènes est récente et estimée à environ 1 Ma, ce qui pourrait correspondre avec l'ouverture du bassin actuel de Lau estimé entre 1-2 Ma (Hall, 2002). Cette hypothèse est d'autant plus crédible que les individus présentant les plus forts pourcentages d'introgression sont retrouvés dans le bassin de Lau et dans l'arc volcanique de Futuna. Le maintien actuel des espèces avec une divergence élevée nécessite cependant la présence de barrières pré et/ou post-zygotiques fortes afin de contrer les effets homogénéisant de l'hybridation, même si celle-ci semble être sporadique et plutôt rare (aucun hybride de première et seconde génération trouvés).

4- Isolement écologique ?

Les variations de l'environnement hydrothermal ont un rôle structurant dans la distribution des espèces qui lui sont associées (Desbruyeres, 1982 ; Grassle, 1987 ; Hessler et Lonsdale, 1991 ; Van Dover et Hessler, 1990) et auraient pu conduire à des adaptations spécifiques chez *A. boucheti* par rapport aux deux autres espèces. Comme montré par de précédentes études et confirmé dans le chapitre 3 de cette thèse, *A. boucheti* présente majoritairement des symbiontes de type *Campylobacteria* alors que *A. kojimai* et *A. strummeri* présentent un phylotype dominant chez les *Gamma-proteobacteria*. Les *Campylobacteria* tolèrent de plus fortes concentrations en dihydrogène (H_2) et en sulfure d'hydrogène (H_2S) (Beinart et al., 2012), ce qui explique la présence de *A. boucheti* sur les parois des cheminées où les émissions présentent souvent des concentrations fortes en ces composés chimiques.

Les assemblages symbiotiques entre les espèces ont un effet majeur sur la composition isotopique en $\delta^{13}C$ des hôtes, en effet, chez les *Campylobacteria*, la voie métabolique de la fixation du carbone utilise le cycle de Krebs inversé alors que chez les *Gamma-proteobacteria*, la fixation du carbone se fait à partir du cycle de Calvin-Benson-Bassham (CBB) (Beinart et al., 2019). Ces voies métaboliques de biosynthèse du carbone se traduisent par un $\delta^{13}C$ très différent entre les espèces ($\delta^{13}C$ plus petit chez les individus de *A. kojimai* et *A. strummeri* utilisant le cycle CBB). Dans cette thèse, nous n'avons pu montrer de différences marquées entre les environnements chimiques des espèces, même si les individus de *A. boucheti* prélevés dans le bassin de Manus vivent dans un environnement présentant globalement de plus fortes concentrations en fer, méthane et manganèse (Fe : 258 µM/kg ; CH4 : 8.5 µM et Mn : 92.8 µM/kg). Il faut toutefois noter que les mesures chimiques ont été quasi exclusivement réalisées sur dans les zones de diffusion où les espèces étaient souvent mélangées. Il est donc possible que l'environnement chimique associé à *A. boucheti* soit beaucoup plus hétérogène qu'on le pense notamment au niveau des cheminées ou les conditions physico-chimiques ne pouvaient pas être mesurées de façon satisfaisante. Des études supplémentaires sont donc nécessaires afin de mieux caractériser l'environnement chimique des populations vivants sur les cheminées hydrothermales pour évaluer l'impact de l'environnement chimique sur les symbiontes de cette espèce puis sur l'adaptabilité de l'hôte. Ainsi, la présence de symbiontes différents au sein des trois espèces pourrait expliquer par leur partition au sein de niches écologiques distinctes. A plus petite échelle, entre *A. kojimai* et *A. strummeri*, il existe également une différence de niche écologique, en effet, ces deux espèces ne se situe pas au même endroit dans les agrégats de gastéropodes (*A. strummeri* à la périphérie), pouvant être liée à une compétition dans l'espace entre les espèces pour l'accès à la nourriture et donc aux symbiontes qu'ils doivent acquérir horizontalement (Beinart et al., 2012, 2014 ; Breusing et al., 2020, 2022). C'est par exemple le cas à Fati-Ufu (Futuna) et à Phoenix (Fiji), où au sein d'une même parcelle, *A. kojimai* et *A. strummeri* présentent tantôt des *Gamma-proteobacteria* et des *Sulfurovom* (faisant parti des *Campylobacteria*) tantôt exclusivement des *Gamma-proteobacteria* mais pas aux mêmes localités.

L'étude sur les transcriptomes, même si elle ne représente qu'un faible effectif de gènes présents dans les transcriptomes (10%), a permis d'identifier quelques gènes sous sélection positive sur la branche menant à *A. boucheti*, pouvant être liés à des différences d'environnements biotiques et abiotiques. Ces gènes sont notamment impliqués dans le système immunitaire de l'hôte et dans les réponses au stress (notamment l'hypoxie et la

calcification de la coquille). Ce résultat suggère une sélection environnementale ancienne chez cette espèce, qui aurait pu contribuer à l'augmentation de la divergence des espèces lors de leur séparation en allopatrie voire dans une spécialisation vis-à-vis de l'habitat des cheminées. Sur les branches menant aux espèces *A. kojimai* et *A. strummeri*, les gènes sous sélection positive sont également liés au système immunitaire des hôtes mais aussi à la spermatogénèse et la reconnaissance gamétique. Ces processus sont plutôt connus pour être des barrières pré-zygotiques agissant sur le maintien des espèces par renforcement (Nielsen et al., 2005). Pour mieux estimer le rôle de l'environnement sur la divergence entre les espèces, il faudrait donc évaluer la temporalité des événements de sélection ayant agi avant ou après la séparation des espèces. Pour cela, une possibilité est d'utiliser la divergence neutre d_S comme un proxy du temps écoulé depuis la séparation entre deux espèces en la corrigeant de son taux de mutation par l'utilisation d'une divergence plus ancienne avec une troisième espèce. Cela permet de s'affranchir du taux de mutation pour mieux comprendre le rôle des événements de sélection positive dans la chronologie du processus de spéciation en étant soit ancien (et donc potentiellement impliqué dans l'initialisation de la spéciation) ou plus récent (et plus spécifiquement impliqués dans un renforcement après le contact secondaire). Une première étude basée sur ce principe a été réalisée pour *A. kojimai*/*A. strummeri* (Figure 61) et a permis de montrer que les gènes sous sélection positive chez ces espèces sont ceux présentant des d_S les plus petits (donc les plus récents ; rectangle jaune sur la Figure 61). Ce résultat semble donc suggérer que les gènes identifiés dans la spermatogénèse et la reconnaissance gamétique chez ces espèces sont plutôt impliqués dans un processus de renforcement après le contact secondaire.

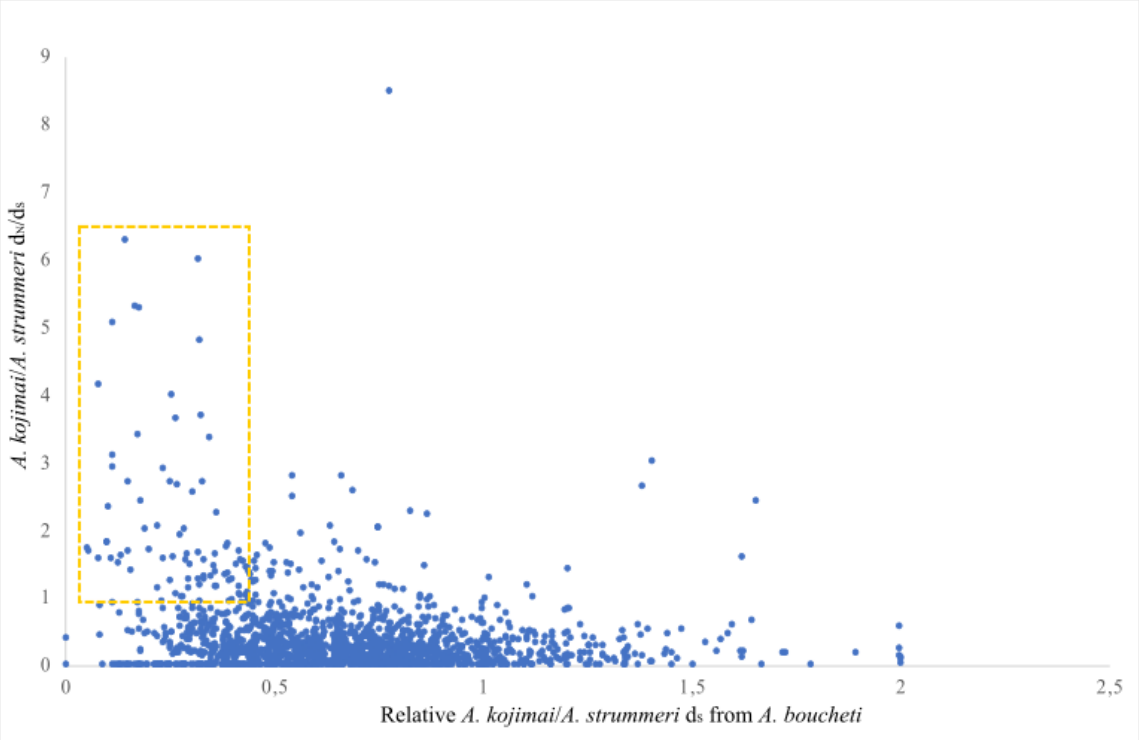


Figure 61: Evolution du d_N/d_S de *A. kojimai*/*A. strummeri* en fonction du temps de divergence.

Cette thèse a permis de mieux caractériser les niches écologiques des différentes espèces d'*Alviniconcha* (Figure 62), ce qui peut être une barrière pré-zygotique entre les espèces si celles-ci ne se rencontrent pas pendant les épisodes de reproduction et peut donc être un frein à l'hybridation actuelle. De plus, lorsque les espèces se retrouvent en mélange, s'il existe des adaptations génétiques nécessaires pour permettre la colonisation des différentes niches écologiques, alors la formation d'hybrides maladaptés (fitness plus faible) peut contrer les effets de l'hybridation et permettre le maintien des espèces. Comme dit précédemment, l'étude sur les transcriptomes de référence semble montrer qu'il existe bien des adaptations génétiques à ces niches écologiques (au moins entre *A. boucheti* et les deux autres espèce), mais ce travail mériterait d'être encore plus approfondi en analysant un plus grand nombre de gènes sur un plus grand nombre d'individus au sein de chacune des espèces. Ceci permettrait d'évaluer la part du polymorphisme intra-espèce et d'avoir une meilleure évaluation plus globale des gènes sous sélection au niveau du transcriptome.

5- Isolement sexuel ?

L'analyse des gènes orthologues entre transcriptomes de référence a montré que la plupart des gènes sous sélection positive ou à fort taux d'évolution retrouvés plus spécifiquement sur les branches menant aux espèces *A. kojimai* et *A. strummeri*, sont liés au

système immunitaire mais aussi à la spermatogénèse, la reconnaissance gamétique ou encore au développement des gonades. Ces processus sont connus pour être des barrières pré-zygotiques qui peuvent agir pour le maintien des espèces par renforcement après des contacts secondaires (Nielsen et al., 2005). Or, un renforcement de ces barrières nécessite au préalable une sélection contre les hybrides, qui s'opère par exemple à cause de mode de reproduction différents entre les espèces, ce qui peut entrainer une baisse de la fitness chez les descendants, comme cela semble le cas chez les *Alviniconcha*.

6- Isolement post-zygotique ?

Le travail effectué au cours de cette thèse a permis de mettre en évidence un mode de reproduction différent entre les espèces d'*Alviniconcha*. En effet, *A. kojimai* semble une espèce androdioïque (individus mâles et hermaphrodites présents) alors que *A. boucheti* et *A. strummeri* sont des espèces gonochoriques (individus mâles et femelles séparés). Ces modes de reproduction sont déterminés par le même système génétique XY qui probablement se localise sur le même chromosome sexuel (Chapitre 2). La présence d'individus hermaphrodites chez *A. kojimai* pourrait cependant découler de l'inactivation d'un gène inhibiteur du développement des gamètes mâles chez les femelles comme c'est par exemple le cas chez la Papaye (Ming et al., 2007). Ce mode de reproduction distinct entre les espèces peut tout de même engendrer des problèmes de viabilité/stérilité/fitness sur les descendants, suivant le croisement des géniteurs, ce qui pourrait représenter une barrière post-zygotique forte entre *A. kojimai* et les deux autres espèces (Coyer et al., 2002 ; Figure 62). L'étude du mode de reproduction chez les autres d'espèces d'*Alviniconcha* et notamment chez *A. hessleri* qui semble être également une espèce hermaphrodite (Hanson et al., 2021) serait intéressant pour mieux évaluer l'impact du système de déterminisme du sexe sur la spéciation de ce genre.

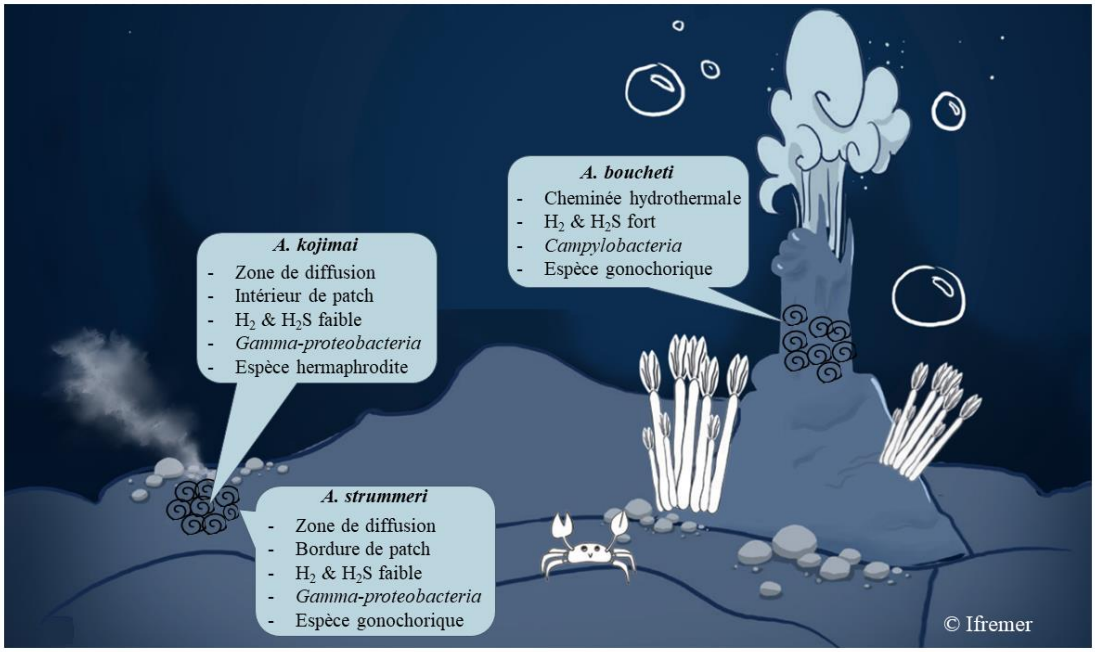


Figure 62: Différences de niches écologiques et du mode de reproduction entre les espèces d'*Alviniconcha* étudiées durant cette thèse. Ces différents éléments peuvent représenter des barrières pré et/ou post-zygotiques au maintien actuel de ces espèces malgré la présence d'hybridation.

7- Quelques perspectives pour l'étude de la spéciation chez *Alviniconcha*

Le travail effectué sur trois espèces d'*Alviniconcha* dans cette thèse doit être élargi à l'ensemble des espèces de ce genre (*A. hessleri* et *A. adamantis* retrouvées dans la fosse des Mariannes ainsi que *A. marisindica* retrouvée sur la ride centrale de l'Océan Indien) afin de reconstruire l'histoire complète et le contexte géographique de la spéciation au sein des *Alviniconcha*. En effet, des analyses sur le partage du polymorphisme ancestral et leurs ajustements à des modèles démographiques sont très certainement biaisées par l'absence des espèces intermédiaires ou jumelles comme ce qui est le cas pour les espèces *A. kojimai* et *A. hessleri*. Ce type d'analyses pourraient compléter les études de phylogénie réalisées par Johnson et al. (2015) et Breusing et al. (2020) en précisant les dates de séparation des espèces et leur histoire démographique. De plus, des études chez les espèces non étudiées dans cette thèse sur l'adaptation locale via des analyses GEA ou l'utilisation de transcriptomes de référence pourraient dans un second temps permettre de comprendre le rôle de l'environnement hydrothermal dans la formation ou le maintien actuel des différentes espèces de ce genre. De plus, avoir les génomes de référence des espèces pourrait être d'une grande aide pour caractériser l'architecture génomique de la divergence entre les espèces,

identifier/quantifier les zones concernées par l'introgression s'il y en a, évaluer si celle-ci peut être adaptative et quels gènes sont affectés.

8- Structure génétique et adaptation locale

En complément de l'étude sur l'origine et le maintien de la divergence entre les trois espèces cibles d'*Alviniconcha*, il est intéressant d'évaluer la différenciation géographique et environnementale au sein de chaque espèce. Cette étude a permis de montrer que la structure génétique des populations de chacune des espèces est très peu marquée, suggérant ainsi une forte capacité dispersive de ces espèces contrairement à une autre espèce de gastéropodes, *Ifremeria nautilei* retrouvée dans la zone d'étude (Tran Lu Y et al., 2022). Ces capacités semblent cependant légèrement différentes entre les trois espèces et pourraient expliquer les disparités trouvées entre espèces au niveau de leur aire de distribution plus ou moins grande. En effet, la différenciation génétique la plus forte est observée chez *A. strummeri* avec un F_{ST} de 0.022 entre les populations du bassin de Lau/Futuna et celles du bassin Nord-Fidjien. Pour *A. kojimai*, une différenciation deux fois plus faible est observée entre des populations pourtant deux fois plus éloignées (F_{ST} de 0.01 entre les populations de Lau/Futuna/Nord-Fidji : région 1 et Manus/Woodlark : région 2). A noter chez cette espèce, une différenciation de 0.003 entre les populations de Lau/Futuna et du bassin Nord-Fidjien. Chez *A. boucheti*, la différenciation entre les populations est encore plus faible (trois fois plus faible que chez *A. kojimai*), puisque des F_{ST} de 0.003 sont retrouvés entre les populations de la région 1 et de la région 2.

Comme précisé précédemment, la structure génétique la plus forte est observée chez *A. strummeri*, or cette espèce semble être localement la moins abondante dans la zone d'étude, sa taille de population pourrait avoir un rôle fort sur la différenciation. En effet, lorsque les tailles efficaces des populations sont faibles, la dérive génétique va avoir tendance à faire augmenter la différenciation génétique entre les populations, ce qui peut expliquer les résultats chez cette espèce. Une autre hypothèse est la dispersion larvaire de l'espèce est plus faible comparée aux deux autres espèces, il serait donc important d'étudier le mode de développement des larves chez cette espèce (notamment à quel moment cette espèce relâche ses larves dans l'environnement). La différence de différenciation génétique observée, même si très faible, entre *A. kojimai* et *A. boucheti* est plus difficile à expliquer, en effet, si on se base sur l'échantillonnage réalisé durant cette thèse, l'aire de distribution et les abondances semblent quasi identiques entre ces deux espèces, ce qui suggère des tailles efficaces proches.

Cependant, différents effets peuvent agir sur la taille efficace malgré une abondance semblable (Frankham et al., 2002), comme, les effets démographiques (réduction de la taille efficace lors d'un goulot d'étranglement par exemple), la variance du succès reproducteur, un biais dans le sexe ratio ou encore des générations chevauchantes (Frankham et al., 2002). Dans le chapitre 2, il a été montré que *A. kojimai* et *A. boucheti* présentent des différences dans leur mode de reproduction. En effet, *A. boucheti* est une espèce à sexes séparés alors que *A. kojimai* est une espèce probablement androdioïque (présentant à la fois des individus mâles et hermaphrodites). Ces différences peuvent affecter à la fois le sexe-ratio si la fonction mâle est active ou pas chez les individus hermaphrodites mais également la distribution du succès reproducteur chez *A. kojimai*. L'androdioécie pourrait expliquer une taille efficace plus faible chez cette espèce et donc une structure géographique plus forte. Une autre hypothèse pouvant expliquer une différenciation inter-bassin plus faible chez *A. boucheti* est une dispersion larvaire plus efficace, peut être avec la présence d'une phase larvaire en surface. L'hypothèse d'une phase larvaire en surface a déjà été validée chez *A. marisindica* dans l'océan Indien (Kim et al., 2022) et pourrait donc être une hypothèse à creuser en réalisant des prélèvements en surface couplés à des analyses de metabarcoding (Kim et al., 2022). Des larves assignées au genre *Alviniconcha* ont également été échantillonnées en surface (entre 0 et 300m) au Nord-Est de Hawaï (à plus de 6 000 km de la fosse des Mariannes ; Sommer et al., 2017). Certaines études sur les espèces hydrothermales suggèrent qu'un flux de gène de proche en proche pourrait être plus efficace pour homogénéiser des populations sur de grandes distance lorsque l'environnement est distribué de façon plus ou moins continue le long d'une dorsale (Audzijonyte et Vrijenhoek, 2010 ; France et al., 1992 ; Jollivet et al., 1995 ; Matabos et al., 2008). Cependant une telle hypothèse est difficile à évaluer dans un système discontinu de dorsales où les populations sont éloignées sur des distances avoisinant 4 000 km (distance entre les bassins de Lau et de Manus). Chez *A. boucheti* cette hypothèse pourrait néanmoins s'appliquer en passant par les zones volcaniques arrière-arc, puisque jusqu'à présent seule cette espèce a été retrouvée sur l'arc des Vanuatu dans la partie Sud-Ouest du bassin Nord-Fidjien. Or cette région géographique permettrait de connecter directement avec les bassins de Manus et de Woodlark à l'ouest et le bassin de Lau à l'est par la dispersion larvaire (Breusing et al., 2021 ; Mitarai et al., 2016).

L'assignation des larves récoltées en surface à Hawaii (Sommer et al., 2017) montre une correspondance à 95% avec *A. hessleri*, ce qui fait penser qu'il existe peut-être une espèce non découverte d'*Alviniconcha* dans l'océan Pacifique Nord. Ce constat démontre que toutes

les espèces d'*Alviniconcha* ne sont à l'heure actuelle pas encore découvertes et que des prospections pour rechercher de nouveaux sites hydrothermaux sont nécessaires. De plus, il serait intéressant d'explorer la dorsale Antarctique/dorsale indienne Sud Est, en effet, des sites hydrothermaux s'ils existent pourraient constituer un trait-d'union entre la dorsale Indienne et les bassins arrière-arc du Pacifique Ouest et donc d'héberger d'autres espèces du genre *Alviniconcha* (German et al., 1998 ; Tunnicliffe et al., 1996).

Durant cette thèse, une recherche de locus 'outliers' pouvant être impliqués dans l'adaptation locale des populations a pu être conduite pour chacune des trois espèces d'*Alviniconcha*. Chez *A. kojimai* et *A. boucheti*, les Analyses Genome-Environnement (GEA) ont pu montrer que les environnements chimiques du milieu hydrothermal auraient pu jouer un rôle dans la différenciation génétique entre les populations de l'Est (Lau/Fidji/Futuna) et de l'Ouest (Manus/Woodlark). Dans ce cas, le niveau de différenciation génétique trouvé à ces locus est significativement plus élevé que celui attendu par le simple éloignement géographique même si ces effets sont confondus. Chez *A. kojimai* les variables candidates expliquant cette différence génétique entre l'Est et l'Ouest sont un clade bactérien (Oligo1) ainsi que les concentrations en zinc et en méthane. Chez *A. boucheti*, les variables associées à un signal d'adaptation entre ces deux grandes régions géographiques sont deux clades bactériens (Oligo 4 et 6) ainsi que les concentrations en calcium, silice et uranium, ce dernier élément étant souvent utilisé comme indicateur du pH de l'environnement (Doubleday et al., 2017). Même s'il est difficile d'affirmer que ces variables écologiques sont à l'origine d'adaptation locale, car, n'importe quel paramètre environnemental non mesuré présentant une variation entre l'Est et Ouest apparaitra corrélée avec la structure génétique des espèces, le signal d'une adaptation bassin-spécifique à grande échelle semble robuste. De plus, le fait que les espèces d'*Alviniconcha* montrent un signal d'adaptation sur certains clades bactériens pourrait signifier que la symbiose joue un rôle dans la différenciation des populations de l'Est et de l'Ouest.

Un deuxième signal d'adaptation locale est également ressorti chez *A. kojimai* et *A. strummeri* pour un site hydrothermal plus spécifiquement. En effet, il existe une corrélation entre des variables bactériennes (Oligo 1 chez *A. strummeri* et Oligo 16 et 20 chez *A. kojimai*) et la différenciation génétique des populations dans le bassin Nord-Fidjien. Chez ces deux espèces, les SNPs outliers sont certainement liés à un ajustement de la symbiose de ces espèces pour prendre en compte une chimie du fluide, potentiellement plus enrichi en H_2/ou H_2S. En effet, il est très intéressant de noter que l'Oligo 16 présent uniquement à Phoenix

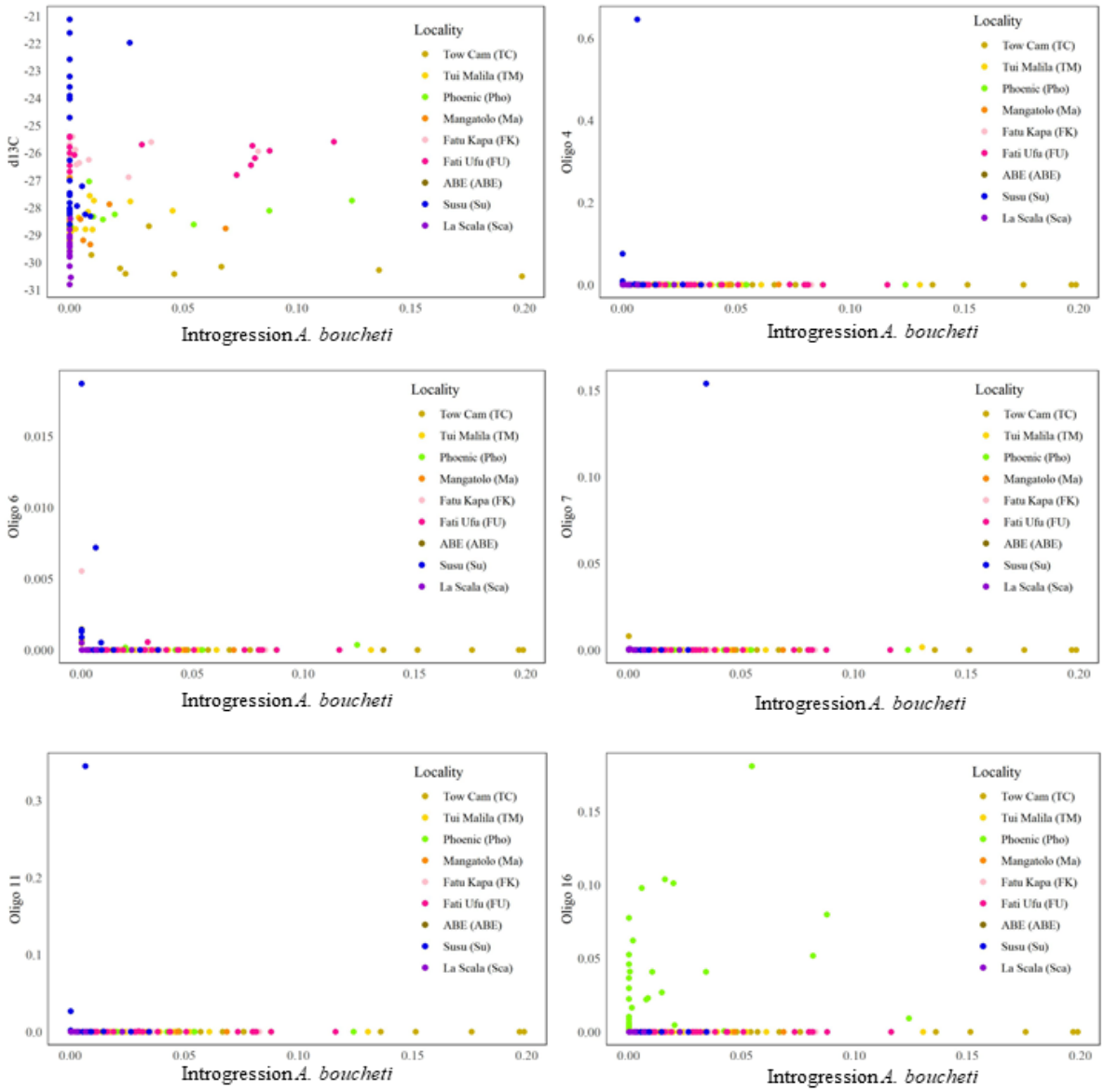
(site de prélèvement dans le bassin Nord-Fidjien) chez *A. kojimai* est une bactérie faisant partie du clade des *Campylobacteria* (*Sulfurovum*) habituellement retrouvée chez *A. boucheti*. Il semblerait donc que chez cette espèce, les individus du bassin Nord-Fidjien sont adaptés localement à l'acquisition de ce symbionte. Chez *A. strummeri*, même s'il est très difficile de tirer des conclusions sur l'adaptation locale du fait du faible nombre de populations analysées, on retrouve cette même association avec *Sulfurovum* à Fati-Ufu. Ainsi, pour établir des parallèles d'adaptation avec un plus grande significativité du signal, un effort d'échantillonnage doit être réalisé chez cette espèce en particulier pour augmenter le nombre de réplicats écologiques ou faire une analyse individuelle d'association telle que l'analyse canonique de redondance (RDA). Pour compléter cette première étude sur l'adaptation locale il serait intéressant d'ajouter des variables écologiques explicatives supplémentaires comme par exemple la concentration en dihydrogène (H_2) qui peut jouer un rôle important sur la présence de certains symbiontes (Beinart et al., 2019). De plus, par manque de temps durant cette thèse, l'annotation de tous les SNPs outliers (5% des XtX les plus forts, cf chapitre 3), n'a pas encore été réalisé mais pourrait être informatif sur les processus à l'origine de l'adaptation locale chez les espèces d'*Alviniconcha* en les comparant notamment aux gènes sous sélection positive obtenus entre espèces lors de la comparaison des transcriptomes. Un moyen de compléter les connaissances sur l'adaptation locale chez ces espèces serait de réaliser également des mesures chimiques individu-ciblées (sondes électrochimiques) ainsi que du metabarcoding des bactéries libres directement sur les communautés d'*Alviniconcha* en prenant soin de cibler les niches écologiques de chacune des espèces (bordure de parcelle pour *A. strummeri*, et hauteurs des cheminées hydrothermales pour *A. boucheti*).

9- Un rôle adaptatif de l'introgression ?

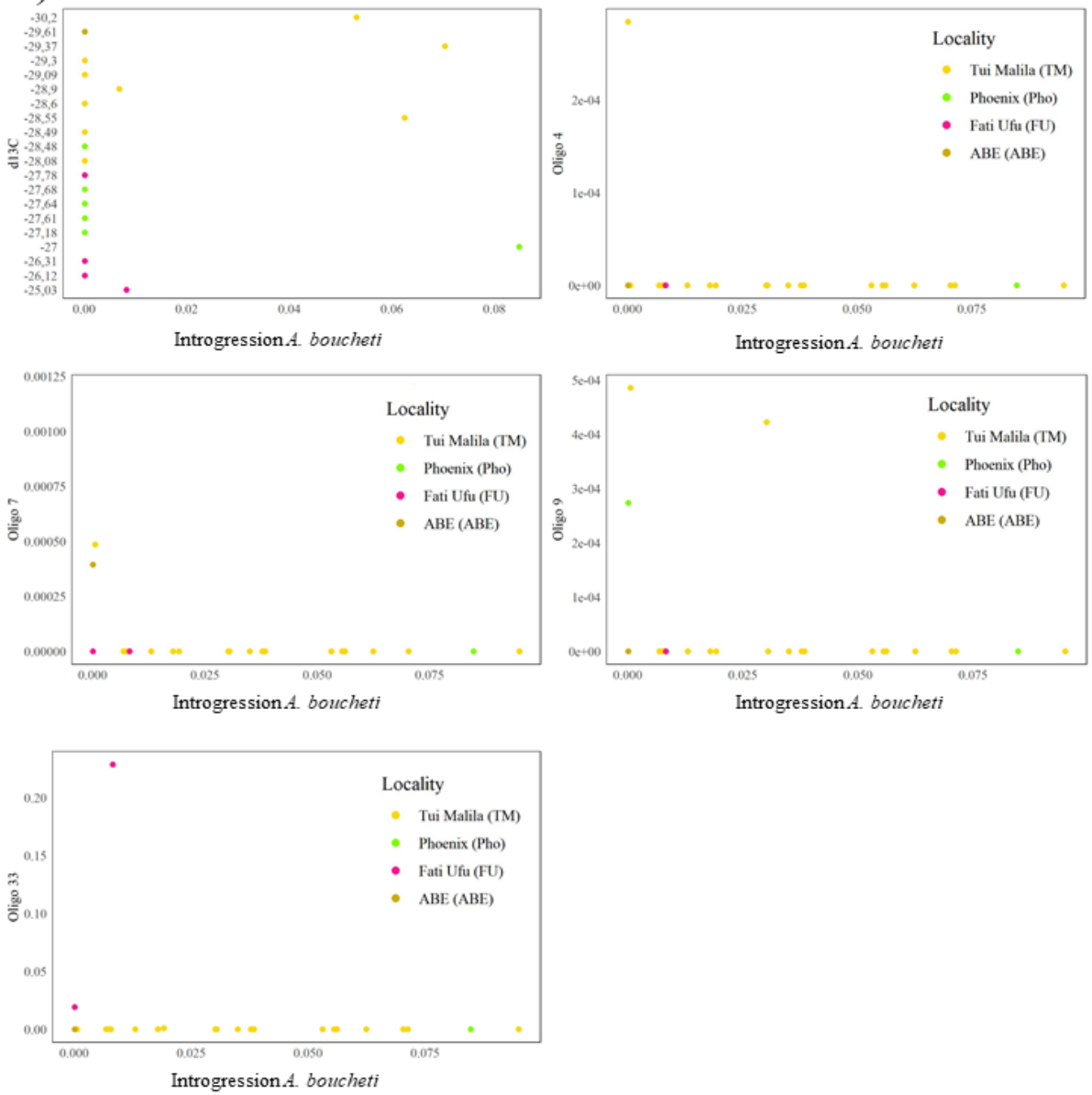
Comme montré au cours de cette thèse, les trois espèces cibles d'*Alviniconcha* montrent des signes récents d'introgression. Il serait donc intéressant de vérifier si l'introgression a permis l'acquisition de variants adaptatifs au sein des espèces, en particulier avec les processus d'adaptation locale liés à la symbiose esquissée ci-dessus. En effet, il a été montré précédemment (Beinart et al., 2012) et confirmé ici que les *Alviniconcha* présentent parfois des mélanges d'endosymbiontes non spécifiques (*Campylobacteria* et *Gamma-proteobacteria*) même si le phylotype non attendu est en faible abondance. C'est ici le cas chez *A. kojimai* et *A. strummeri* qui possèdent respectivement dans le bassin Nord-Fidjien et à Futuna un mélange de *Gamma-proteobacteria* et de *Sulfurovum*. Ainsi, l'acquisition de ce

symbionte toujours retrouvé minoritairement chez *A. kojimai* et *A. strummeri* mais majoritaire chez *A. boucheti* a-t-elle été permise ou facilitée par un processus d'introgression adaptative ? Pour tenter de répondre à cette question, j'ai vérifié si les individus les plus introgressés au niveau de leur génome présentaient aussi une plus forte quantité de bactéries non spécifiques (quantité de *Sulfurovom* chez *A. kojimai* et *A. strummeri* et quantité de *Gamma-proteobacteria* chez *A. boucheti* ; Figure 63). De plus, pour limiter le risque d'avoir rater un clade bactérien avec le metabarcoding, le $\delta^{13}C$ a également été regardé car sa valeur est très influencée par la composition symbiotique en *Gamma-proteobacteria* vs *Campylobacteria* (Voir précédemment). Contrairement à l'hypothèse émise, les individus les plus introgressés de chaque espèce ne présentent pas les plus fortes proportions en clades bactériens exogènes, et n'ont pas de valeurs de $\delta^{13}C$ aberrantes (Figure 63). Chez *A. kojimai* par exemple, les individus les plus introgressés (15-18% du génome) sont trouvés dans le bassin de Lau alors que ceux présentant le plus de bactéries non-spécifiques sont retrouvés dans le bassin Nord-Fidjien (jusqu'à 10% de la composition bactérienne ; Figure 63a). Nous n'avons donc pas observé de signe d'introgression adaptative facilitant l'acquisition de symbiontes exogènes chez ces espèces. Cette conclusion reste cependant à nuancer car nous ne pouvons pas affirmer qu'un tel mécanisme n'a pas eu lieu par le passé sans laisser de signature détectable aujourd'hui en termes de variations de fréquence (par exemple si les variants génétiques permettant la symbiose avec un type bactérien particulier ont été héritées par hybridation introgressive puis redistribués largement entre populations au sein de l'espèce receveuse).

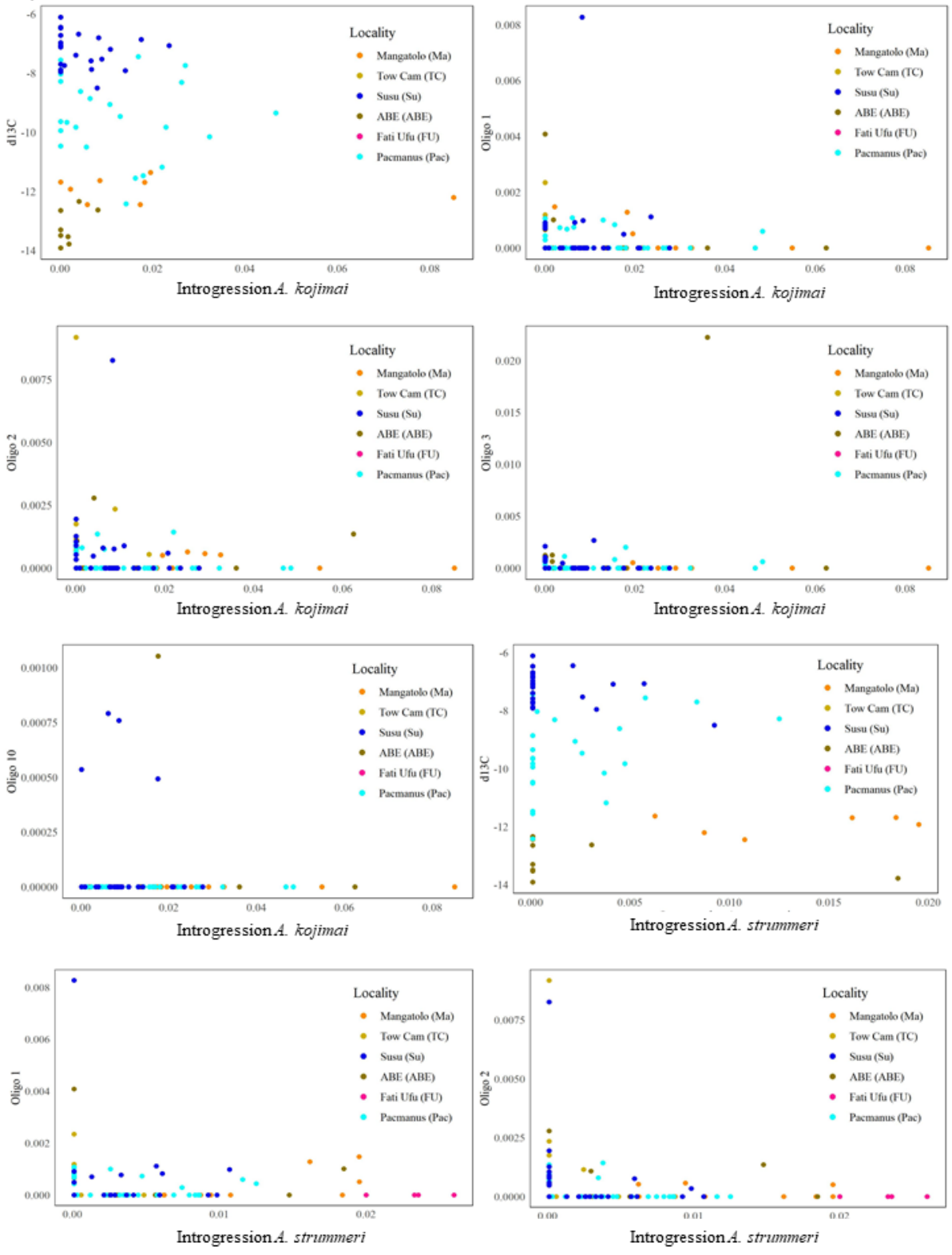
a)


d13C
Oligo 4
Oligo 6
Oligo 7
Oligo 11
Oligo 16
Introgression A. boucheti
Locality
Tow Cam (TC)
Tui Malila (TM)
Phoenic (Pho)
Mangatolo (Ma)
Fatu Kapa (FK)
Fati Ufu (FU)
ABE (ABE)
Susu (Su)
La Scala (Sca)

b)


d13C
-30,2
-29,61
-29,37
-29,3
-29,09
-28,9
-28,6
-28,55
-28,49
-28,48
-28,08
-27,78
-27,68
-27,64
-27,61
-27,18
-27
-26,31
-26,12
-25,03
0.00
0.02
0.04
0.06
0.08
Introgression A. boucheti
Oligo 4
2e-04
1e-04
0e+00
0.000
0.025
0.050
0.075
Introgression A. boucheti
Locality
Tui Malila (TM)
Phoenix (Pho)
Fati Ufu (FU)
ABE (ABE)
Oligo 7
0.00125
0.00100
0.00075
0.00050
0.00025
0.00000
Introgression A. boucheti
Oligo 9
5e-04
4e-04
3e-04
2e-04
1e-04
0e+00
Introgression A. boucheti
Oligo 33
0.20
0.15
0.10
0.05
0.00
Introgression A. boucheti

c)


d13C
Oligo 1
Oligo 2
Oligo 3
Oligo 10
Locality
Mangatolo (Ma)
Tow Cam (TC)
Susu (Su)
ABE (ABE)
Fati Ufu (FU)
Pacmanus (Pac)
Introgression A. kojimai
Introgression A. strummeri

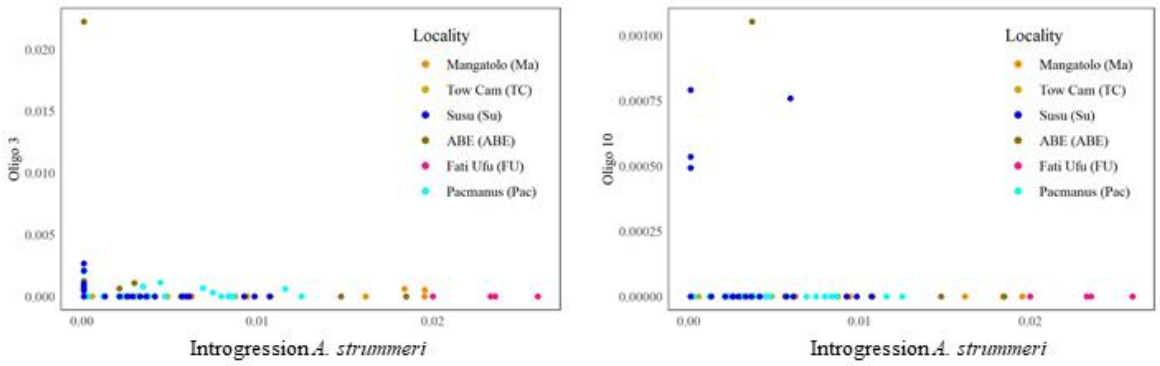


Figure 63: Graphes en 'biplot' représentant le niveau d'introgression dans chacune des espèce d'*Alviniconcha* en fonction de la fraction des symbiotes non spécifiques (propre à une autre espèce) et le $\delta^{13}C$ chez (a) *A. kojimai*, (b) *A. strummeri* et (c) *A. boucheti*.

10- Conclusions

Dans cette thèse, le but était de mieux comprendre le rôle de l'environnement hydrothermal profond et de la géographie sur la distribution et la spéciation chez le genre *Alviniconcha* dans le Pacifique Sud-Ouest. Ainsi, ce travail a permis de confirmer la présence de trois espèces sympatriques (*A. kojimai*, *A. strummeri* et *A. boucheti*) présentant une divergence forte tant sur le génome mitochondrial que sur le génome nucléaire. Cette divergence est certainement issue d'une spéciation allopatrique initiée lors de la formation des bassins arrière-arc et renforcée chez *A. boucheti* par une adaptation environnementale pour s'associer aux *Campylobacteria* menant à un isolement écologique partiel de celui-ci. Malgré cette divergence élevée, les espèces sont encore capables de s'hybrider localement mais rarement (pas de F1/F2 trouvés), ce qui sous-entend la présence de barrières pré et/ou post-zygotiques contrant les effets de l'hybridation pour maintenir l'intégrité génétique des espèces. Dans cette thèse, certaines barrières ont pu être identifiées, avec un rôle certain de la niche écologique (dont dépend l'association symbiotique de l'hôte) ou encore du mode de reproduction pouvant limiter les flux de gènes entre espèces. Ainsi, il semblerait que la géographie ai joué un rôle fort à l'origine de la séparation des espèces puis que l'adaptation à des niches écologiques distinctes ait permis d'accentuer la divergence entre les espèces et notamment entre *A. boucheti* et les deux autres. Ainsi, l'environnement hydrothermal, de part sa nature fragmentée et instable à permis l'émergence et le maintien actuel de trois espèces d'*Alviniconcha* dans le Pacifique Sud-Ouest.

Références bibliographiques

A

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229-246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Alfaro-Lucas, J. M., Shimabukuro, M., Ogata, I. V., Fujiwara, Y., and Sumida, P. Y. G. (2018). Trophic structure and chemosynthesis contributions to heterotrophic fauna inhabiting an abyssal whale carcass. *Marine Ecology Progress Series*, *596*, 1-12. https://doi.org/10.3354/meps12617

Allio, R., Donega, S., Galtier, N., and Nabholz, B. (2017). Large variation in the ratio of mitochondrial to nuclear mutation rate across animals : Implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Molecular Biology and Evolution*, *34*(11), 2762-2772. https://doi.org/10.1093/molbev/msx197

Anderson, T. R., and Rice, T. (2006). Deserts on the sea floor : Edward Forbes and his azoic hypothesis for a lifeless deep ocean. *Endeavour*, *30*(4), 131-137. https://doi.org/10.1016/j.endeavour.2006.10.003

Arbiza, L., Dopazo, J., and Dopazo, H. (2006). Positive selection, relaxation, and acceleration in the evolution of the human and chimp genome. *Plos Computational Biology*, *2*(4), e38. https://doi.org/10.1371/journal.pcbi.0020038

Arellano, S. M., Van Gaest, A. L., Johnson, S. B., Vrijenhoek, R. C., and 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., and 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

Auzende, J.-M., Ishibashi, J., Beaudoin, Y., Charlou, J.-L., Delteil, J., Donval, J.-P., Fouquet, Y., Gouillou, J.-P., Ildefonse, B., Kimura, H., Nishio, Y., Radford-Knoery, J., and Ruellan, É. (2000). Les extrémités orientale et occidentale du bassin de Manus, Papouasie-Nouvelle-Guinée, explorées par submersible : La campagne Manaute. *Comptes Rendus de l'Académie des Sciences - Series IIA - Earth and Planetary Science*, *331*(2), 119-126. https://doi.org/10.1016/S1251-8050(00)01387-2

Auzende, J.-M., Lafoy, Y., and Marsset, B. (1988). Recent geodynamic evolution of the North Fiji basin (Southwest Pacific). *Geology*, *16*(10), 925-929.

Auzende, J.-M., Pelletier, B., and Lafoy, Y. (1994). Twin active spreading ridges in the North Fiji Basin (Southwest Pacific). *Geology*, *22*(1), 63-66.

Auzende, J.-M., Urabe, T., Deplus, C., Eissen, J.-P., Grimaud, D., Huchon, P., Ishibashi, J., Joshima, M., Lagabrielle, Y., and Mevel, C. (1989). Le cadre géologique d'un site hydrothermal actif : La campagne STARMER 1 du submersible Nautile dans le Bassin Nord-Fidjien. *CR Acad. Sci. Paris*, *309*(2), 1787-1795.

B

Baker, E. T., Cormier, M.-H., Langmuir, C. H., and Zavala, K. (2001). Hydrothermal plumes along segments of contrasting magmatic influence, 15°20′–18°30′N, East Pacific Rise : Influence of axial faulting. *Geochemistry, Geophysics, Geosystems*, *2*(9). https://doi.org/10.1029/2000GC000165

Bandelt, H. J., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*(1), 37-48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

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., and 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

Barreto, F. S., Moy, G. W., and Burton, R. S. (2011). Interpopulation patterns of divergence and selection across the transcriptome of the copepod *Tigriopus californicus*. *Molecular Ecology*, *20*(3), 560-572. https://doi.org/10.1111/j.1365-294X.2010.04963.x

Barton, N., and Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity*, *57*(3), 357-376. https://doi.org/10.1038/hdy.1986.135

Barton, N. H., and Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, *16*(1), 113-148. https://doi.org/10.1146/annurev.es.16.110185.000553

Bateson, W. (1909). Heredity and variation in modern lights. *Darwin and modern science*.

Beaumont, M. A., and Balding, D. J. (2004). Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, *13*(4), 969-980. https://doi.org/10.1111/j.1365-294X.2004.02125.x

Beinart, R. A., Luo, C., Konstantinidis, K. T., Stewart, F. J., and Girguis, P. R. (2019). The bacterial symbionts of closely related hydrothermal vent snails with distinct geochemical habitats show broad similarity in chemoautotrophic gene content. *Frontiers in Microbiology*, *10*. https://www.frontiersin.org/article/10.3389/fmicb.2019.01818

Beinart, R. A., Nyholm, S. V., Dubilier, N., and 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

Beinart, R. A., Sanders, J. G., Faure, B., Sylva, S. P., Lee, R. W., Becker, E. L., Gartman, A., Luther, G. W., Seewald, J. S., Fisher, C. R., and Girguis, P. R. (2012). Evidence for the role of endosymbionts in regional-scale habitat partitioning by hydrothermal vent symbioses. *Proceedings of the National Academy of Sciences*, *109*(47), E3241-E3250. https://doi.org/10.1073/pnas.1202690109

Belshaw, R., and Katzourakis, A. (2005). BlastAlign : A program that uses blast to align problematic nucleotide sequences. *Bioinformatics*, *21*(1), 122-123. https://doi.org/10.1093/bioinformatics/bth459

Beukeboom, L. W., and Perrin, N. (2014). The evolution of sex determination. *Oxford University Press*.

Bierne, N., Gagnaire, P.-A., and David, P. (2013). The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. *Current Zoology*, *59*(1), 72-86. https://doi.org/10.1093/czoolo/59.1.72

Bierne, N., Welch, J., Loire, E., Bonhomme, F., and David, P. (2011). The coupling hypothesis : Why genome scans may fail to map local adaptation genes. *Molecular Ecology*, *20*(10), 2044-2072. https://doi.org/10.1111/j.1365-294X.2011.05080.x

Bogaerts-Márquez, M., Guirao-Rico, S., Gautier, M., and González, J. (2021). Temperature, rainfall and wind variables underlie environmental adaptation in natural populations of *Drosophila melanogaster*. *Molecular Ecology*, *30*(4), 938-954. https://doi.org/10.1111/mec.15783

Boidin-Wichlacz, C., Jollivet, D., Papot, C., Roisin, L., Massol, F., and Tasiemski, A. (2021). Genetic diversification and life-cycle of the polychaete *Capitella* spp. from the English Channel : Evidence for sympatric cryptic species and alternative reproductive strategies. *Marine Biology*, *168*(12), 176. https://doi.org/10.1007/s00227-021-03972-2

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic : A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Booker, T. R., Yeaman, S., and Whitlock, M. C. (2021). Global adaptation complicates the interpretation of genome scans for local adaptation. *Evolution Letters*, *5*(1), 4-15. https://doi.org/10.1002/evl3.208

Borda, E., Kudenov, J. D., Chevaldonné, P., Blake, J. A., Desbruyères, D., Fabri, M.-C., Hourdez, S., Pleijel, F., Shank, T. M., Wilson, N. G., Schulze, A., and Rouse, G. W. (2013). Cryptic species of *Archinome* (Annelida : Amphinomida) from vents and seeps. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1770), 20131876. https://doi.org/10.1098/rspb.2013.1876

Bouchemousse, S., Liautard-Haag, C., Bierne, N., and Viard, F. (2016). Distinguishing contemporary hybridization from past introgression with postgenomic ancestry-informative SNPs in strongly differentiated *Ciona* species. *Molecular Ecology*, *25*(21), 5527-5542. https://doi.org/10.1111/mec.13854

Bouchon, D., Cordaux, R., and Grève, P. (2008). Feminizing *Wolbachia* and the evolution of sex determination in isopods. *Insect symbiosis*, *3*, 273-294.

Boulart, C., Rouxel, O., Scalabrin, C., Le Meur, P., Pelleter, E., Poitrimol, C., Thiebaut, E., Matabos, M., Castel, J., Tran Lu Y, A., Michel, L., Cathalot, C., Cheron, S., Boissier, A., Germain, Y., Guyader, V., Arnaud-Haon, S., Tasiemski, A., Kuama, D. C., … Jollivet, D. (2022). Active hydrothermal vents in the Woodlark Basin may act as dispersing centres for hydrothermal fauna. *Communications Earth and Environment*. https://orbi.uliege.be/handle/2268/288581

Brannock, P. M., and Hilbish, T. J. (2010). Hybridization results in high levels of sterility and restricted introgression between invasive and endemic marine blue mussels. *Marine Ecology Progress Series*, *406*, 161-171. https://doi.org/10.3354/meps08522

Brelsford, A., Dufresnes, C., and 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., Castel, J., Yang, Y., Broquet, T., Sun, J., Jollivet, D., Qian, P.-Y., and Beinart, R. A. (2022). Global 16S rRNA diversity of provannid snail endosymbionts from Indo-Pacific deep-sea hydrothermal vents. *Environmental Microbiology Report*. https://doi.org/10.1111/1758-2229.13051

Breusing, C., Johnson, S. B., Mitarai, S., Beinart, R. A., and Tunnicliffe, V. (2019). 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., and 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

Broquet, T., Barranger, A., Billard, E., Bestin, A., Berger, R., Honnaert, G., and Viard, F. (2015). The size advantage model of sex allocation in the protandrous sex-changer *Crepidula fornicata* : Role of the mating system, sperm storage, and male mobility. *The American Naturalist*, *186*(3), 404-420. https://doi.org/10.1086/682361

Brucker, R. M., and Bordenstein, S. R. (2012). Speciation by symbiosis. *Trends in Ecology and Evolution*, *27*(8), 443-451. https://doi.org/10.1016/j.tree.2012.03.011

Bull, J. J. (1983). Evolution of sex determining mechanisms (Vol. 1). *Benjamin-Cummings Publishing Company*.

Bull, J. J., Vogt, R. C., and Bulmer, M. G. (1982). Heritability of sex ratio in turtles with environmental sex determination. *Evolution*, *36*(2), 333-341. https://doi.org/10.2307/2408052

Burton, R. S., and Barreto, F. S. (2012). A disproportionate role for mtDNA in Dobzhansky–Muller incompatibilities? *Molecular Ecology*, *21*(20), 4942-4957. https://doi.org/10.1111/mec.12006

Butlin, R. K., and Smadja, C. M. (2018). Coupling, reinforcement, and speciation. *The American Naturalist*, *191*(2), 155-172. https://doi.org/10.1086/695136

C

Calvignac, S., Konecny, L., Malard, F., and Douady, C. J. (2011). Preventing the pollution of mitochondrial datasets with nuclear mitochondrial paralogs (numts). *Mitochondrion*, *11*(2), 246-254. https://doi.org/10.1016/j.mito.2010.10.004

Capblancq, T., Luu, K., Blum, M. G. B., and Bazin, E. (2018). Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources*, *18*(6), 1223-1233. https://doi.org/10.1111/1755-0998.12906

Caro, A., Gómez-Moliner, B. J., and Madeira, M. J. (2019). Integrating multilocus DNA data and 3D geometric morphometrics to elucidate species boundaries in the case of *Pyrenaearia* (Pulmonata : Hygromiidae). *Molecular Phylogenetics and Evolution*, *132*, 194-206. https://doi.org/10.1016/j.ympev.2018.12.007

Castillo, V. M., and Brown, D. I. (2012). Un Caso de Trifalia en el Caracol Marino *Echinolittorina peruviana* (Caenogastropoda : Littorinidae). *International Journal of Morphology*, *30*(3), 791-796. https://doi.org/10.4067/S0717-95022012000300003

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks : An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124-3140. https://doi.org/10.1111/mec.12354

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. (2011). Stacks : building and genotyping loci *denovo* from short-read sequences. *G3 Genes|Genomes|Genetics*, *1*(3), 171-182. https://doi.org/10.1534/g3.111.000240

Charlesworth, B., and Charlesworth, D. (1978). A model for the evolution of dioecy and gynodioecy. *The American Naturalist*, *112*(988), 975-997. https://doi.org/10.1086/283342

Charlesworth, D. (1984). Androdioecy and the evolution of dioecy. *Biological Journal of the Linnean Society*, *22*(4), 333-348. https://doi.org/10.1111/j.1095-8312.1984.tb01683.x

Charnov, E. L. (1982). Alternative life-histories in protogynous fishes : A general evolutionary theory. *Marine Ecology Progress Series*, *9*(3), 305-307.

Chavagnac, V., Leleu, T., Fontaine, F., Cannat, M., Ceuleneer, G., and Castillo, A. (2018). Spatial variations in vent chemistry at the Lucky Strike hydrothermal field, Mid-Atlantic Ridge (37°N) : Updates for subseafloor flow geometry from the newly discovered Capelinhos vent. *Geochemistry, Geophysics, Geosystems*, *19*(11), 4444-4458. https://doi.org/10.1029/2018GC007765

Chevaldonne, P., Jollivet, D., Desbruyeres, D., Lutz, R., and 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.

Chiu, Y.-W., Chen, H.-C., Lee, S.-C., and Chen, C. A. (2002). Morphometric analysis of shell and operculum variations in the Viviparid snail, *Cipangopaludina chinensis* (Mollusca : Gastropoda), in Taiwan. *Zoological Studies*, 11.

Coop, G., Witonsky, D., Di Rienzo, A., and Pritchard, J. K. (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics*, *185*(4), 1411-1423. https://doi.org/10.1534/genetics.110.114819

Cordaux, R., Bouchon, D., and Grève, P. (2011). The impact of endosymbionts on the evolution of host sex-determination mechanisms. *Trends in Genetics*, *27*(8), 332-341. https://doi.org/10.1016/j.tig.2011.05.002

Coyer, J. A., Peters, A. F., Hoarau, G., Stam, W. T., and Olsen, J. L. (2002). Hybridization of the marine seaweeds, *Fucus serratus* and *Fucus evanescens* (Heterokontophyta : Phaeophyceae) in a 100-year-old zone of secondary contact. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *269*(1502), 1829-1834. https://doi.org/10.1098/rspb.2002.2093

Coyne, J. A. (2007). Sympatric speciation. *Current Biology*, *17*(18), R787-R788.

Coyne, J. A., and Orr, H. A. (2004). Speciation, volume 37 Sunderland. *MA: Sinauer Associates*.

Cunha, L., Martins, G. M., Amaral, A., and Rodrigues, A. (2007). A case of simultaneous hermaphroditism in the Azorean endemic limpet *Patella candei* gomesii (Mollusca : Patellogastropoda), a gonochoristic species. *Invertebrate Reproduction and Development*, *50*(4), 203-205. https://doi.org/10.1080/07924259.2007.9652247

Curdia, J., Rodrigues, A. S., Martins, A. M. F., and Costa, M. J. (2005). The reproductive cycle of *Patella candei* gomesii Drouët, 1858 (Mollusca : Patellogastropoda), an Azorean endemic subspecies. *Invertebrate Reproduction and Development*, *48*(1-3), 137-145. https://doi.org/10.1080/07924259.2005.9652180

D

Daguin-Thiebaut, C. (2021). Construction of individual ddRAD libraries V.1.

Daka, E. R., and Hawkins, S. J. (2004). Tolerance to heavy metals in *Littorina saxatilis* from a metal contaminated estuary in the Isle of Man. *Journal of the Marine Biological Association of the United Kingdom*, *84*(2), 393-400. https://doi.org/10.1017/S0025315404009336h

Darwin, C. (1859). On the origin of species, 1859. Routledge.

Decker, S. K., and Ammerman, L. K. (2020). Phylogeographic analysis reveals mito-nuclear discordance in *Dasypterus intermedius*. *Journal of Mammalogy*, *101*(5), 1400-1409. https://doi.org/10.1093/jmammal/gyaa106

Denis, F., Jollivet, D., and Moraga, D. (1993). Genetic separation of two allopatric populations of hydrothermal snails *Alviniconcha* spp. (Gastropoda) from two South Western Pacific back-arc basins. *Biochemical Systematics and Ecology*, *21*(4), 431-440. https://doi.org/10.1016/0305-1978(93)90101-V

Desbruyeres, D. (1982). Donnes ecologiques sur un nouveau site d'hydrothermalisme actif de la ride du Pacifique oriental. *Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences, Paris*, *295*, 489-494.

Desbruyères, D., Alayse-Danet, A.-M., Ohta, S., and the Scientific Parties of biolauand starmerCruises. (1994). Deep-sea hydrothermal communities in Southwestern Pacific back-arc basins (the North Fiji and Lau Basins) : Composition, microdistribution and food web. *Marine Geology*, *116*(1), 227-242. https://doi.org/10.1016/0025-3227(94)90178-3

Desbruyeres, and Laubier. (1982). *Paralvinella grasslei*, new genus, new species of *Alvinellinae* (Polychaete: Ampharetidae) from the Galapagos rift geothermal vents. *Proceedings of the Biological Society of Washington*, *95*(3), 11.

Dieckmann, U., and Doebeli, M. (1999). On the origin of species by sympatric speciation. *Nature*, *400*(6742), 354-357. https://doi.org/10.1038/22521

Dilly, G. F., Young, C. R., Lane, W. S., Pangilinan, J., and Girguis, P. R. (2012). Exploring the limit of metazoan thermal tolerance via comparative proteomics : Thermally induced changes in protein abundance by two hydrothermal vent polychaetes. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1741), 3347-3356. https://doi.org/10.1098/rspb.2012.0098

Dobzhansky, T. (1937). Genetics and the origin of species. *Columbia university press*.

Doebeli, M., and Dieckmann, U. (2003). Speciation along environmental gradients. *Nature*, *421*(6920), 259-264. https://doi.org/10.1038/nature01274

Dos Reis, M., and Yang, Z. (2013). Why do more divergent sequences produce smaller nonsynonymous/synonymous rate ratios in pairwise sequence comparisons? *Genetics*, *195*(1), 195-204. https://doi.org/10.1534/genetics.113.152025

Doubleday, Z. A., Nagelkerken, I., and Connell, S. D. (2017). Ocean life breaking rules by building shells in acidic extremes. *Current Biology*, *27*(20), R1104-R1106. https://doi.org/10.1016/j.cub.2017.08.057

Doyle, J. J., and Dickson, E. E. (1987). Preservation of Plant Samples for DNA Restriction Endonuclease Analysis. *Taxon*, *36*(4), 715-722. https://doi.org/10.2307/1221122

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*.

Du Preez, C., and Fisher, C. R. (2018). Long-term stability of back-arc basin hydrothermal vents. *Frontiers in Marine Science*, *5*. https://www.frontiersin.org/article/10.3389/fmars.2018.00054

Duda, T. F. (2021). Patterns of variation of mutation rates of mitochondrial and nuclear genes of gastropods. *BMC Ecology and Evolution*, *21*(1), 13. https://doi.org/10.1186/s12862-021-01748-2

Duranton, M. (2019). Intégrer des approches expérimentales et d'évolution moléculaire en génomique de la spéciation afin d'identifier les mécanismes impliqués dans la divergence entre bar atlantique et loup méditerranéen [These de doctorat, Montpellier]. https://www.theses.fr/2019MONTG031

E

Edgar, R. (2010). Usearch. https://www.osti.gov/sciencecinema/biblio/1137186

Edwards, ScottV., and Beerli, P. (2000). Perspective : gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, *54*(6), 1839-1854. https://doi.org/10.1111/j.0014-3820.2000.tb01231.x

Eggers, S., and Sinclair, A. (2012). Mammalian sex determination—Insights from humans and mice. *Chromosome Research*, *20*(1), 215-238. https://doi.org/10.1007/s10577-012-9274-3

Ekimova, I., Valdés, Á., Chichvarkhin, A., Antokhina, T., Lindsay, T., and Schepetov, D. (2019). Diet-driven ecological radiation and allopatric speciation result in high species diversity in a temperate-cold water marine genus *Dendronotus* (Gastropoda : Nudibranchia). *Molecular Phylogenetics and Evolution*, *141*, 106609. https://doi.org/10.1016/j.ympev.2019.106609

Eldon, B., Riquet, F., Yearsley, J., Jollivet, D., and Broquet, T. (2016). Current hypotheses to explain genetic chaos under the sea. *Current Zoology*, *62*(6), 551-566. https://doi.org/10.1093/cz/zow094

Ellis, R. E. (2008). Chapter 2 Sex determination in the *Caenorhabditis elegans* germ line. In *Current Topics in Developmental Biology* (Vol. 83, p. 41-64). Academic Press. https://doi.org/10.1016/S0070-2153(08)00402-X

Emms, D. M., and Kelly, S. (2015). OrthoFinder : Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, *16*(1), 157. https://doi.org/10.1186/s13059-015-0721-2

Excoffier, L., and 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

Eyre-Walker, A., Keightley, P. D., Smith, N. G. C., and Gaffney, D. (2002). Quantifying the slightly deleterious mutation model of molecular evolution. *Molecular Biology and Evolution*, *19*(12), 2142-2149. https://doi.org/10.1093/oxfordjournals.molbev.a004039

Ezaz, T., Quinn, A. E., Miura, I., Sarre, S. D., Georges, A., and Marshall Graves, J. A. (2005). The dragon lizard *Pogona vitticeps* has ZZ/ZW micro-sex chromosomes. *Chromosome Research*, *13*(8), 763-776. https://doi.org/10.1007/s10577-005-1010-9

Ezenwa, V. O., Gerardo, N. M., Inouye, D. W., Medina, M., and Xavier, J. B. (2012). Animal behavior and the microbiome. *Science*, *338*(6104), 198-199. https://doi.org/10.1126/science.1227412

F

Feder, J. L., Egan, S. P., and Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, *28*(7), 342-350. https://doi.org/10.1016/j.tig.2012.03.009

Feder, J. L., and Nosil, P. (2010). The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*, *64*(6), 1729-1747. https://doi.org/10.1111/j.1558-5646.2009.00943.x

Feder, J. L., Nosil, P., Wacholder, A. C., Egan, S. P., Berlocher, S. H., and Flaxman, S. M. (2014). Genome-wide congealing and rapid transitions across the speciation continuum during speciation with Gene Flow. *Journal of Heredity*, *105*(S1), 810-820. https://doi.org/10.1093/jhered/esu038

Fiala-Médioni, A. (1984). Mise en évidence par microscopie électronique à transmission de l'abondance de bactéries symbiotiques dans la branchie de Mollusques bivalves de sources hydrothermales profondes. *Comptes rendus des séances de l'Académie des sciences. Série 3, Sciences de la vie*, *298*(17), 487-492.

Flaxman, S. M., Wacholder, A. C., Feder, J. L., and Nosil, P. (2014). Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Molecular Ecology*, *23*(16), 4074-4088. https://doi.org/10.1111/mec.12750

Foll, M., and 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

Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial.

Fontanillas, E., Galzitskaya, O. V., Lecompte, O., Lobanov, M. Y., Tanguy, A., Mary, J., Girguis, P. R., Hourdez, S., and Jollivet, D. (2017). Proteome evolution of deep-sea hydrothermal vent *Alvinellid* polychaetes supports the ancestry of thermophily and subsequent adaptation to cold in some lineages. *Genome Biology and Evolution*, *9*(2), 279-296. https://doi.org/10.1093/gbe/evw298

Fouquet, Y., Pelleter, E., Konn, C., Chazot, G., Dupré, S., Alix, A. S., Chéron, S., Donval, J. P., Guyader, V., Etoubleau, J., Charlou, J. L., Labanieh, S., and Scalabrin, C. (2018). Volcanic and

hydrothermal processes in submarine calderas : The Kulo Lasi example (SW Pacific). *Ore Geology Reviews*, *99*, 314-343. https://doi.org/10.1016/j.oregeorev.2018.06.006

Fouquet, Y., Stackelberg, U. V., Charlou, J. L., Donval, J. P., Erzinger, J., Foucher, J. P., Herzig, P., Mühe, R., Soakai, S., Wiedicke, M., and Whitechurch, H. (1991). Hydrothermal activity and metallogenesis in the Lau back-arc basin. *Nature*, *349*(6312), 778-781. https://doi.org/10.1038/349778a0

Fraïsse, C., Popovic, I., Mazoyer, C., Spataro, B., Delmotte, S., Romiguier, J., Loire, É., Simon, A., Galtier, N., Duret, L., Bierne, N., Vekemans, X., and Roux, C. (2021). DILS : Demographic inferences with linked selection by using ABC. *Molecular Ecology Resources*, *21*(8), 2629-2644. https://doi.org/10.1111/1755-0998.13323

France, S. C., Hessler, R. R., and Vrijenhoek, R. C. (1992). Genetic differentiation between spatially-disjunct populations of the deep-sea, hydrothermal vent-endemic amphipod *Ventiella sulfuris*. *Marine Biology*, *114*(4), 551-559. https://doi.org/10.1007/BF00357252

François, O., Martins, H., Caye, K., and Schoville, S. D. (2016). Controlling false discoveries in genome scans for selection. *Molecular Ecology*, *25*(2), 454-469. https://doi.org/10.1111/mec.13513

Frankham, R., Ballou, S. E. J. D., Briscoe, D. A., and Ballou, J. D. (2002). Introduction to Conservation Genetics. *Cambridge University Press*.

Frichot, E., and François, O. (2015). LEA : An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925-929. https://doi.org/10.1111/2041-210X.12382

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., and François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, *196*(4), 973-983. https://doi.org/10.1534/genetics.113.160572

Frichot, E., Schoville, S. D., de Villemereuil, P., Gaggiotti, O. E., and François, O. (2015). Detecting adaptive evolution based on association with ecological gradients : Orientation matters! *Heredity*, *115*(1), 22-28. https://doi.org/10.1038/hdy.2015.7

G

Gabirot, M., Castilla, A. M., López, P., and Martín, J. (2010). Differences in chemical signals may explain species recognition between an island lizard, *Podarcis atrata*, and related mainland lizards, *P. hispanica*. *Biochemical Systematics and Ecology*, *38*(4), 521-528. https://doi.org/10.1016/j.bse.2010.05.008

Gage, J. D., and Tyler, P. A. (1991). Deep-sea biology : a natural history of organisms at the deep-sea floor. *Cambridge University Press*.

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., Pavey, S. A., Normandeau, E., and Bernatchez, L. (2013). The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish wpecies pairs assessed by rad Sequencing. *Evolution*, *67*(9), 2483-2497. https://doi.org/10.1111/evo.12075

Gaina, C., and Müller, D. (2007). Cenozoic tectonic and depth/age evolution of the Indonesian gateway and associated back-arc basins. *Earth-Science Reviews*, *83*(3), 177-203. https://doi.org/10.1016/j.earscirev.2007.04.004

Galindo, J., Morán, P., and Rolán-Alvarez, E. (2009). Comparing geographical genetic differentiation between candidate and noncandidate loci for adaptation strengthens support for parallel ecological

divergence in the marine snail *Littorina saxatilis*. *Molecular Ecology*, *18*(5), 919-930. https://doi.org/10.1111/j.1365-294X.2008.04076.x

García-Ramos, G., and Kirkpatrick, M. (1997). Genetic models of adaptation and gene flow in peripheral populations. *Evolution*, *51*(1), 21-28. https://doi.org/10.1111/j.1558-5646.1997.tb02384.x

Gautier, M. (2015). BayPass genome-wide scan for adaptive differentiation and association analysis with population-specific covariables (en lien avec la publication Gautier M. 2015. Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics 201 (4) : 1555-1579* (10.1534/genetics.115.181453). https://hal.inrae.fr/hal-02798150

Gavrilets, S. (2004). Fitness landscapes and the origin of species (MPB-41). *Princeton University Press*.

Gavrilets, S., Li, H., and Vose, M. D. (2000). Patterns of parapatric speciation. *Evolution*, *54*(4), 1126-1134. https://doi.org/10.1111/j.0014-3820.2000.tb00548.x

Gavrilets, S., and Waxman, D. (2002). Sympatric speciation by sexual conflict. *Proceedings of the National Academy of Sciences*, *99*(16), 10533-10538. https://doi.org/10.1073/pnas.152011499

Gayral, P., Melo-Ferreira, J., Glémin, S., Bierne, N., Carneiro, M., Nabholz, B., Lourenco, J. M., Alves, P. C., Ballenghien, M., Faivre, N., Belkhir, K., Cahais, V., Loire, E., Bernard, A., and Galtier, N. (2013). Reference-free population genomics from next-generation transcriptome data and the vertebrate–invertebrate gap. *Plos Genetics*, *9*(4), e1003457. https://doi.org/10.1371/journal.pgen.1003457

German, C. R., Baker, E. T., Mevel, C., Tamaki, K., and the FUJI Science Team. (1998). Hydrothermal activity along the Southwest Indian ridge. *Nature*, *395*(6701), 490-493. https://doi.org/10.1038/26730

Gibbs, R. A., and Pachter, L. (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, *428*(6982), 493-521. https://doi.org/10/nature02426-s9.pdf

Goodliffe, A. M., Taylor, B., Martinez, F., Hey, R., Maeda, K., and Ohno, K. (1997). Synchronous reorientation of the Woodlark Basin spreading center. *Earth and Planetary Science Letters*, *146*(1), 233-242. https://doi.org/10.1016/S0012-821X(96)00227-0

Goudet, J. (2020). hierfstat : Estimation and tests of hierarchical F-statistics. R package version 0.04-22. 2015.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Trinity : Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature biotechnology*, *29*(7), 644-652. https://doi.org/10.1038/nbt.1883

Grassle, J. F. (1987). The ecology of deep-sea hydrothermal vent communities. In J. H. S. Blaxter and A. J. Southward (Éds.), *Advances in Marine Biology* (Vol. 23, p. 301-362). Academic Press. https://doi.org/10.1016/S0065-2881(08)60110-8

Gross, L. (2006). Demonstrating the theory of ecological speciation in Cichlids. *PLOS Biology*, *4*(12), e449. https://doi.org/10.1371/journal.pbio.0040449

Günther, T., and Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics*, *195*(1), 205-220. https://doi.org/10.1534/genetics.113.152462

H

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

Hall, T. (1999). BioEdit : A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.*, *41*, 95-98.

Hallock, P., and Hansen, H. J. (1979). Depth adaptation in *Amphistegina* : Change in lamellar thickness. *Bulletin of the Geological Society of Denmark*, *27*, 99-104.

Hamilton, Z. R., and Johnson, M. S. (2015). Hybridization between genetically and morphologically divergent forms of *Rhagada* (Gastropoda : Camaenidae) snails at a zone of secondary contact. *Biological Journal of the Linnean Society*, *114*(2), 348-362. https://doi.org/10.1111/bij.12410

Hannington, M. D., Jonasson, I. R., Herzig, P. M., and Petersen, S. (1995). Physical and chemical processes of seafloor mineralization at mid-ocean ridges. *Washington DC American Geophysical Union Geophysical Monograph Series*, *91*, 115-157. https://doi.org/10.1029/GM091p0115

Hanson et al. (2021). Poster presentation at the 16th Deep Sea Biology Symposium, Brest, France.

Harr, B. (2006). Genomic islands of differentiation between house mouse subspecies. *Genome Research*, *16*(6), 730-737. https://doi.org/10.1101/gr.5045006

Harrison, R. G. (1993). Hybrid zones and the evolutionary process. *Oxford University Press.*

Hauser, T. P. (2002). Frost sensitivity of hybrids between wild and cultivated carrots. *Conservation Genetics*, *3*(1), 73-76. https://doi.org/10.1023/A:1014256302971

Havird, J. C., and Sloan, D. B. (2016). The roles of mutation, selection, and expression in determining relative rates of evolution in mitochondrial versus nuclear genomes. *Molecular Biology and Evolution*, *33*(12), 3042-3053. https://doi.org/10.1093/molbev/msw185

Hein, R. J., and Mizell, K. (2013). Hydrothermal systems and mineralization of volcanic arcs : Comparison of West Pacific and Mediterranean arcs. In *Geological Setting, Mineral Resources and Ancient Works of Samos and Adjacent Islands of the Aegean Sea, 26–30 August 2013, Karlovassi, Samos Island, Greece* (p. 216-225). Citeseer.

Hendry, A. P., and Day, T. (2005). Population structure attributable to reproductive time : Isolation by time and adaptation by time. *Molecular Ecology*, *14*(4), 901-916. https://doi.org/10.1111/j.1365-294X.2005.02480.x

Henry, M. S., Childress, J. J., and Figueroa, D. (2008). Metabolic rates and thermal tolerances of chemoautotrophic symbioses from Lau Basin hydrothermal vents and their implications for species distributions. *Deep Sea Research Part I: Oceanographic Research Papers*, *55*(5), 679-695. https://doi.org/10.1016/j.dsr.2008.02.001

Hessler, R. R., and Lonsdale, P. F. (1991). Biogeography of Mariana Trough hydrothermal vent communities. *Deep Sea Research Part A. Oceanographic Research Papers*, *38*(2), 185-199. https://doi.org/10.1016/0198-0149(91)90079-U

Hessler, R. R., Smithey, W. M., Boudrias, M. A., Keller, C. H., Lutz, R. A., and Childress, J. J. (1988). Temporal change in megafauna at the Rose Garden hydrothermal vent (Galapagos Rift; Eastern tropical Pacific). *Deep Sea Research Part A. Oceanographic Research Papers*, *35*(10), 1681-1709. https://doi.org/10.1016/0198-0149(88)90044-1

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

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., Poss, M. L., Reed, L. K., Storfer, A., and Whitlock, M. C. (2016). Finding the genomic basis of local adaptation :

pitfalls, practical solutions, and future directions. *The American Naturalist*, *188*(4), 379-397. https://doi.org/10.1086/688018

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., and Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine Stickleback using sequenced RAD Tags. *PLOS Genetics*, *6*(2), e1000862. https://doi.org/10.1371/journal.pgen.1000862

Hourdez, S., and Lallier, F. H. (2007). Adaptations to hypoxia in hydrothermal-vent and cold-seep invertebrates. In R. Amils, C. Ellis-Evans, and H. Hinghofer-Szalkay (Éds.), *Life in Extreme Environments* (p. 297-313). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6285-8_19

Huang, X., and Madan, A. (1999). CAP3 : A DNA Sequence Assembly Program. *Genome Research*, *9*(9), 868-877. https://doi.org/10.1101/gr.9.9.868

Hudson, R. R., and Turelli, M. (2003). Stochasticity overrules the "Three-Times Rule" : genetic drift, Ggenetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution*, *57*(1), 182-190. https://doi.org/10.1111/j.0014-3820.2003.tb00229.x

Huxley, J. (1942). Evolution. The Modern Synthesis. *Evolution. The Modern Synthesis.* https://www.cabdirect.org/cabdirect/abstract/19432202794

J

Janson, K. (1982). Genetic and environmental effects on the growth rate of *Littorina saxatilis*. *Marine Biology*, *69*(1), 73-78. https://doi.org/10.1007/BF00396963

Janson, K. (1983). Selection and migration in two distinct phenotypes of *Littorina saxatilis* in Sweden. *Oecologia*, *59*(1), 58-61. https://doi.org/10.1007/BF00388072

Jeffreys, H. (1961). Theory of probability : Oxford university press. *New York*, 472.

Johannesson, B., and J.ohannesson, K. (1996). Population differences in behaviour and morphology in the snail *Littorina saxatilis* : Phenotypic plasticity or genetic differentiation? *Journal of Zoology*, *240*(3), 475-493. https://doi.org/10.1111/j.1469-7998.1996.tb05299.x

Johnson, S. B., Warén, A., Tunnicliffe, V., Dover, C. V., Wheat, C. G., Schultz, T. F., and Vrijenhoek, R. C. (2015). Molecular taxonomy and naming of five cryptic species of *Alviniconcha* snails (Gastropoda : Abyssochrysoidea) from hydrothermal vents. *Systematics and Biodiversity*, *13*(3), 278-295. https://doi.org/10.1080/14772000.2014.970673

Johnson, S. B., Young, C. R., Jones, W. J., Warén, A., and Vrijenhoek, R. C. (2006). Migration, isolation, and speciation of hydrothermal vent *Limpet* (Gastropoda; Lepetodrilidae) across the blanco transform fault. *The Biological Bulletin*, *210*(2), 140-157. https://doi.org/10.2307/4134603

Jollivet, D., Chevaldonne, P., and Planque, B. (1995). 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., Hashimoto, J., Auzende, J.-M., Honza, E., Ruellan, E., Dutt, S., Iwabuchi, Y., Jarvis, P., and Joshima, M. (1989). First observations of faunal assemblages associated with hydrothermalism in the North Fiji back-arc basin. *C. R. HEBD. SEANCES ACAD. SCI.(III), PARIS.*, *309*(8), 301-308.

Jolly, M. T., Jollivet, D., Gentil, F., Thiébaut, E., and Viard, F. (2005). Sharp genetic break between Atlantic and English Channel populations of the polychaete *Pectinaria koreni*, along the North coast of France. *Heredity*, *94*(1), 23-32. https://doi.org/10.1038/sj.hdy.6800543

Jolly, M., Viard, F., Weinmayr, G., Gentil, F., Thiébaut, E., and Jollivet, D. (2003). Does the genetic structure of *Pectinaria koreni* (Polychaeta : Pectinariidae) conform to a source–sink

metapopulation model at the scale of the Baie de Seine? *Helgoland Marine Research*, *56*(4), 238-246. https://doi.org/10.1007/s10152-002-0123-1

Jombart, T. (2008). adegenet : A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403-1405. https://doi.org/10.1093/bioinformatics/btn129

K

Kawecki, T. J. (1997). Sympatric speciation via habitat specialization driven by deleterious mutations. *Evolution*, *51*(6), 1751-1763. https://doi.org/10.1111/j.1558-5646.1997.tb05099.x

Kawecki, T. J., and Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*(12), 1225-1241. https://doi.org/10.1111/j.1461-0248.2004.00684.x

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A. (2012). Geneious Basic : An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647-1649. https://doi.org/10.1093/bioinformatics/bts199

Kim, M., Kang, J.-H., and Kim, D. (2022). Holoplanktonic and meroplanktonic larvae in the surface waters of the Onnuri vent field in the Central Indian Ridge. *Journal of Marine Science and Engineering*, *10*(2), 158. https://doi.org/10.3390/jmse10020158

Kimura, M. (1983). The neutral theory of molecular evolution. *Cambridge University Press.*

Kimura, M. (1985). The role of compensatory neutral mutations in molecular evolution. *Journal of Genetics*, *64*(1), 7. https://doi.org/10.1007/BF02923549

Kimura, M., and Ohta, T. (1969). The average number of generations until fixation of a mutant gene in a finite population. *Genetics*, *61*(3), 763-771.

Kimura, M., and Weiss, G. H. (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, *49*(4), 561-576.

Kmiec, B., Woloszynska, M., and Janska, H. (2006). Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. *Current Genetics*, *50*(3), 149-159. https://doi.org/10.1007/s00294-006-0082-1

Kojima, S., Segawa, R., Fijiwara, Y., Fujikura, K., Ohta, S., and 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

Kondrashov, A. S., and Kondrashov, F. A. (1999). Interactions among quantitative traits in the course of sympatric speciation. *Nature*, *400*(6742), 351-354. https://doi.org/10.1038/22514

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., and 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

Kozminsky, E. V., and Serbina, E. A. (2020). Mechanisms of sex determination in mollusks *Littorina saxatilis* and *L. obtusata* (Gastropoda: Littorinidae). *Science Education Practive,* 160.

Kunze, T., Heß, M., and Haszprunar, G. (2016). 3D-interactive microanatomy of *Ventsia tricarinata* Warén and Bouchet, 1993 (Vetigastropoda : Seguenzioidea) from Pacific hydrothermal vents. *Journal of Molluscan Studies*, *82*(3), 366-377. https://doi.org/10.1093/mollus/eyw002

L

Lafoy, Y. (1989). Evolution géodynamique des bassins marginaux Nord-Fidjien et de Lau (Sud-Ouest Pacifique) [These de l' Université de Bretagne]. https://archimer.ifremer.fr/doc/00034/14529/

Lalou, C. (1991). Deep-sea hydrothermal venting : A recently discovered marine system. *Journal of Marine Systems*, *1*(4), 403-440. https://doi.org/10.1016/0924-7963(91)90007-H

Laming, S. R., Hourdez, S., Cambon-Bonavita, M.-A., and Pradillon, F. (2020). Classical and computed tomographic anatomical analyses in a not-so-cryptic *Alviniconcha* species complex from hydrothermal vents in the SW Pacific. *Frontiers in Zoology*, *17*(1), 12. https://doi.org/10.1186/s12983-020-00357-x

Langmuir, C., Humphris, S., Fornari, D., Van Dover, C., Von Damm, K., Tivey, M. K., Colodner, D., Charlou, J.-L., Desonie, D., Wilson, C., Fouquet, Y., Klinkhammer, G., and Bougault, H. (1997). Hydrothermal vents near a mantle hot spot : The Lucky Strike vent field at 37°N on the Mid-Atlantic Ridge. *Earth and Planetary Science Letters*, *148*(1), 69-91. https://doi.org/10.1016/S0012-821X(97)00027-7

Laubier, L. (1992). Vingt Mille Vies sous la mer. Odile Jacob.

Le Bris, N., Sarradin, P.-M., and Caprais, J.-C. (2003). Contrasted sulphide chemistries in the environment of 13°N EPR vent fauna. *Deep Sea Research Part I: Oceanographic Research Papers*, *50*(6), 737-747. https://doi.org/10.1016/S0967-0637(03)00051-7

Leigh, J. W., and Bryant, D. (2015). popart : Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110-1116. https://doi.org/10.1111/2041-210X.12410

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology and Evolution*, *17*(4), 183-189. https://doi.org/10.1016/S0169-5347(02)02497-7

Leonard, J. L., and Lukowiak, K. (1991). Sex and the simultaneous hermaphrodite : Testing models of male-female conflict in a sea slug, *Navanax intermis* (Opisthobranchia). *Animal Behaviour*, *41*(2), 255-266. https://doi.org/10.1016/S0003-3472(05)80477-4

Liew, W. C., Bartfai, R., Lim, Z., Sreenivasan, R., Siegfried, K. R., and Orban, L. (2012). Polygenic sex determination system in Zebrafish. *Plos One*, *7*(4), e34397. https://doi.org/10.1371/journal.pone.0034397

Lima, T. G., Burton, R. S., and Willett, C. S. (2019). Genomic scans reveal multiple mito-nuclear incompatibilities in population crosses of the copepod *Tigriopus californicus*. *Evolution*, *73*(3), 609-620. https://doi.org/10.1111/evo.13690

Lindholm, M. (2014). Morphologically conservative but physiologically diverse : the mode of stasis in *Anostraca* (Crustacea: Branchiopoda). *Evolutionary Biology*, *41*(3), 503-507. https://doi.org/10.1007/s11692-014-9283-6

Lonsdale, P. (1977). Clustering of suspension-feeding macrobenthos near abyssal hydrothermal vents at oceanic spreading centers. *Deep Sea Research*, *24*(9), 857-863. https://doi.org/10.1016/0146-6291(77)90478-7

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

M

Maheshwari, S., and Barbash, D. A. (2011). The genetics of hybrid incompatibilities. *Annual Review of Genetics*, *45*(1), 331-355. https://doi.org/10.1146/annurev-genet-110410-132514

Marsh, A. G., Mullineaux, L. S., Young, C. M., and Manahan, D. T. (2001). Larval dispersal potential of the tubeworm *Riftia pachyptila* at deep-sea hydrothermal vents. *Nature*, *411*(6833), 77-80. https://doi.org/10.1038/35075063

Marsh, L., Copley, J. T., Huvenne, V. A. I., Linse, K., Reid, W. D. K., Rogers, A. D., Sweeting, C. J., and Tyler, P. A. (2012). Microdistribution of faunal assemblages at deep-sea hydrothermal vents in the Southern Ocean. *Plos One*, *7*(10), e48348. https://doi.org/10.1371/journal.pone.0048348

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., and 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

Matabos, M., and 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., Plouviez, S., Hourdez, S., Desbruyères, D., Legendre, P., Warén, A., Jollivet, D., and Thiébaut, E. (2011). Faunal changes and geographic crypticism indicate the occurrence of a biogeographic transition zone along the Southern East Pacific Rise. *Journal of Biogeography*, *38*(3), 575-594. https://doi.org/10.1111/j.1365-2699.2010.02418.x

Matabos, M., Thiébaut, E., Le Guen, D., Sadosky, F., Jollivet, D., and 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

Matson, C. K., and Zarkower, D. (2012). Sex and the singular DM domain : Insights into sexual regulation, evolution and plasticity. *Nature Reviews Genetics*, *13*(3), 163-174. https://doi.org/10.1038/nrg3161

Mayr, E. (1942). Birds collected during the Whitney South Sea expedition. Notes on the Polynesian species of *Aplonis*. *American Museum novitates*; no. 1166.

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., and Tevi, T. (2005). New hydrothermal activity and alkalic volcanism in the backarc Coriolis Troughs, Vanuatu. *Geology*, *33*(1), 61-64. https://doi.org/10.1130/G20870.1

Michel, L. N., David, B., Dubois, P., Lepoint, G., and De Ridder, C. (2016). Trophic plasticity of Antarctic echinoids under contrasted environmental conditions. *Polar Biology*, *39*(5), 913-923. https://doi.org/10.1007/s00300-015-1873-y

Mikkelsen, P. M. (2011). Speciation in modern marine bivalves (Mollusca : Bivalvia): Insights from the published record. *American Malacological Bulletin*, *29*(1/2), 217-245. https://doi.org/10.4003/006.029.0212

Ming, R., Yu, Q., and Moore, P. H. (2007). Sex determination in papaya. *Seminars in Cell and Developmental Biology*, *18*(3), 401-408. https://doi.org/10.1016/j.semcdb.2006.11.013

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2 : New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, *37*(5), 1530-1534. https://doi.org/10.1093/molbev/msaa015

Minich, J. J., Sanders, J. G., Amir, A., Humphrey, G., Gilbert, J. A., and Knight, R. (2019). Quantifying and understanding well-to-well contamination in microbiome research. *mSystems*, *4*(4), e00186-19. https://doi.org/10.1128/mSystems.00186-19

Mitarai, S., Watanabe, H., Nakajima, Y., Shchepetkin, A. F., and 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

Moalic, Y., Desbruyères, D., Duarte, C. M., Rozenfeld, A. F., Bachraty, C., and Arnaud-Haond, S. (2012). Biogeography revisited with network theory : retracing the history of hydrothermal vent communities. *Systematic Biology*, *61*(1), 127. https://doi.org/10.1093/sysbio/syr088

Muller, H. (1942). Isolating mechanisms, evolution, and temperature. *Biol. Symp.*, *6*, 71-125.

Murchison, E. P. (2008). Clonally transmissible cancers in dogs and Tasmanian devils. *Oncogene*, *27*(2), S19-S30. https://doi.org/10.1038/onc.2009.350

Muths, D., Davoult, D., Gentil, F., and Jollivet, D. (2006). Incomplete cryptic speciation between intertidal and subtidal morphs of *Acrocnida brachiata* (Echinodermata : Ophiuroidea) in the Northeast Atlantic. *Molecular Ecology*, *15*(11), 3303-3318. https://doi.org/10.1111/j.1365-294X.2006.03000.x

Muths, D., Jollivet, D., Gentil, F., and Davoult, D. (2009). Large-scale genetic patchiness among NE Atlantic populations of the brittle star *Ophiothrix fragilis*. *Aquatic Biology*, *5*(2), 117-132. https://doi.org/10.3354/ab00138

N

Nadeau, S., Meirmans, P. G., Aitken, S. N., Ritland, K., and Isabel, N. (2016). The challenge of separating signatures of local adaptation from those of isolation by distance and colonization history : The case of two white pines. *Ecology and Evolution*, *6*(24), 8649-8664. https://doi.org/10.1002/ece3.2550

Nei, and Gojobori. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*. https://doi.org/10.1093/oxfordjournals.molbev.a040410

Nei, M. (1987). Chapter 11 : Phylogenetic Trees. *Columbia University Press*. https://doi.org/10.7312/nei-92038-012

Nei, M., and Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, *76*(10), 5269-5273. https://doi.org/10.1073/pnas.76.10.5269

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, *39*(1), 197-218. https://doi.org/10.1146/annurev.genet.39.073003.112420

Nielsen, R., Bustamante, C., Clark, A. G., Glanowski, S., Sackton, T. B., Hubisz, M. J., Fledel-Alon, A., Tanenbaum, D. M., Civello, D., White, T. J., Sninsky, J. J., Adams, M. D., and Cargill, M. (2005). A scan for positively selected genes in the genomes of Humans and Chimpanzees. *Plos Biology*, *3*(6), e170. https://doi.org/10.1371/journal.pbio.0030170

Nielsen, R., Paul, J. S., Albrechtsen, A., and Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, *12*(6), 443-451. https://doi.org/10.1038/nrg2986

Nigon, V., and Dougherty, E. C. (1949). Reproductive patterns and attempts at reciprocal crossing of *Rhabditis elegans* Maupas, 1900, and *Rhabditis briggsae* Dougherty and Nigon, 1949 (Nematoda : Rhabditidae). *Journal of Experimental Zoology*, *112*(3), 485-503.

Nosil, P., Funk, D. J., and Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, *18*(3), 375-402. https://doi.org/10.1111/j.1365-294X.2008.03946.x

Nukazawa, K., Kazama, S., and Watanabe, K. (2015). A hydrothermal simulation approach to modelling spatial patterns of adaptive genetic variation in four stream insects. *Journal of Biogeography*, *42*(1), 103-113. https://doi.org/10.1111/jbi.12392

O

O'Brien, J., Dahlhoff, E., and Somero, G. N. (1991). Thermal resistance of mitochondrial respiration : Hydrophobic interactions of membrane proteins may limit thermal resistance. *Physiological Zoology*, *64*(6), 1509-1526. https://doi.org/10.1086/physzool.64.6.30158227

Ockelmann, K. W., and Dinesen, G. E. (2011). Life on wood – the carnivorous deep-sea mussel *Idas argenteus* (Bathymodiolinae, Mytilidae, Bivalvia). *Marine Biology Research*, *7*(1), 71-84. https://doi.org/10.1080/17451001003714504

Ohta, T., and Ina, Y. (1995). Variation in synonymous substitution rates among mammalian genes and the correlation between synonymous and nonsynonymous divergences. *Journal of Molecular Evolution*, *41*(6), 717-720. https://doi.org/10.1007/BF00173150

Okutani, T., and Ohta, S. (1988). A new gastropod mollusk associated with hydrothermal vents in the Mariana Back-Arc Basin, Western Pacific. *Venus (Japanese Journal of Malacology)*, *47*(1), 1-9.

O'Mullan, G. D., Maas, P. a. Y., Lutz, R. A., and 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

Ondréas, H., Cannat, M., Fouquet, Y., and Normand, A. (2012). Geological context and vents morphology of the ultramafic-hosted Ashadze hydrothermal areas (Mid-Atlantic Ridge 13°N). *Geochemistry, Geophysics, Geosystems*, *13*(11). https://doi.org/10.1029/2012GC004433

Orr, H. A., and Turelli, M. (2001). The evolution of postzygotic isolation : accumulating Dobzhansky-Muller incompatibilities. *Evolution*, *55*(6), 1085-1094. https://doi.org/10.1111/j.0014-3820.2001.tb00628.x

P

Palumbi, S. R. (1994). Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics*, *25*(1), 547-572. https://doi.org/10.1146/annurev.es.25.110194.002555

Palumbi, S. R. (2009). Speciation and the evolution of gamete recognition genes : Pattern and process. *Heredity*, *102*(1), 66-76. https://doi.org/10.1038/hdy.2008.104

Pannell, J. R. (2002). The evolution and maintenance of androdioecy. *Annual Review of Ecology and Systematics*, *33*(1), 397-425. https://doi.org/10.1146/annurev.ecolsys.33.010802.150419

Pante, E., Becquet, V., Viricel, A., and Garcia, P. (2019). Investigation of the molecular signatures of selection on ATP synthase genes in the marine bivalve *Limecola balthica*. *Aquatic Living Resources*, *32*, 3. https://doi.org/10.1051/alr/2019001

Pante, E., Rohfritsch, A., Becquet, V., Belkhir, K., Bierne, N., and Garcia, P. (2012). SNP detection from *denovo* transcriptome sequencing in the Bivalve *Macoma balthica* : marker development for evolutionary studies. *Plos One*, *7*(12), e52302. https://doi.org/10.1371/journal.pone.0052302

Pante, E., and Simon-Bouhet, B. (2013). marmap : A package for importing, plotting and analyzing bathymetric and topographic data in R. *PLOS ONE*, *8*(9), e73051. https://doi.org/10.1371/journal.pone.0073051

Papot, C., Massol, F., Jollivet, D., and Tasiemski, A. (2017). Antagonistic evolution of an antibiotic and its molecular chaperone : How to maintain a vital ectosymbiosis in a highly fluctuating habitat. *Scientific Reports*, *7*(1), 1454. https://doi.org/10.1038/s41598-017-01626-2

Pardo, L. M., and Johnson, L. E. (2005). Explaining variation in life-history traits : Growth rate, size, and fecundity in a marine snail across an environmental gradient lacking predators. *Marine Ecology Progress Series*, *296*, 229-239. https://doi.org/10.3354/meps296229

Parsons, P. A. (1994). Morphological stasis : An energetic and ecological perspective incorporating stress. *Journal of Theoretical Biology*, *171*(4), 409-414. https://doi.org/10.1006/jtbi.1994.1244

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double digest RADseq : An inexpensive method for *denovo* SNP discovery and genotyping in model and non-model species. *Plos One*, *7*(5), e37135. https://doi.org/10.1371/journal.pone.0037135

Petraitis, P. S. (1991). The effects of sex ratio and density on the expression of gender in the polychaete *Capitella capitata*. *Evolutionary Ecology*, *5*(4), 393-404. https://doi.org/10.1007/BF02214156

Pinceel, J., Jordaens, K., and Backeljau, T. (2005). Extreme mtDNA divergences in a terrestrial slug (Gastropoda, Pulmonata, Arionidae) : Accelerated evolution, allopatric divergence and secondary contact. *Journal of Evolutionary Biology*, *18*(5), 1264-1280. https://doi.org/10.1111/j.1420-9101.2005.00932.x

Podowski, E. L., Ma, S., Iii, G. W. L., Wardrop, D., and Fisher, C. R. (2010). Biotic and abiotic factors affecting distributions of megafauna in diffuse flow on andesite and basalt along the Eastern Lau spreading center, Tonga. *Marine Ecology Progress Series*, *418*, 25-45. https://doi.org/10.3354/meps08797

Podowski, E. L., Moore, T. S., Zelnio, K. A., Luther, G. W., and Fisher, C. R. (2009). Distribution of diffuse flow megafauna in two sites on the Eastern Lau spreading center, Tonga. *Deep Sea Research Part I: Oceanographic Research Papers*, *56*(11), 2041-2056. https://doi.org/10.1016/j.dsr.2009.07.002

Powell, M. A., and Somero, G. N. (1986). Adaptations to sulfide by hydrothermal vent animals : Sites and mechanisms of detoxification and metabolism. *The Biological Bulletin*, *171*(1), 274-290. https://doi.org/10.2307/1541923

Proestou. (2005). Sex change in *Crepidula fornicata :* Influence of environmental factors on reproductive success and the timing of sex change. https://www.proquest.com/openview/8d735a4383f151a0752771fb9901eb19/1?pq-origsite=gscholarandcbl=18750anddiss=y

R

Racimo, F., Marnetto, D., and Huerta-Sánchez, E. (2017). Signatures of archaic adaptive introgression in present-day human populations. *Molecular Biology and Evolution*, *34*(2), 296-317. https://doi.org/10.1093/molbev/msw216

Rafferty, N. E., and Boughman, J. W. (2006). Olfactory mate recognition in a sympatric species pair of three-spined sticklebacks. *Behavioral Ecology*, *17*(6), 965-970. https://doi.org/10.1093/beheco/arl030

Reeves, E. P., Seewald, J. S., Saccocia, P., Bach, W., Craddock, P. R., Shanks, W. C., Sylva, S. P., Walsh, E., Pichler, T., and Rosner, M. (2011). Geochemistry of hydrothermal fluids from the PacManus, Northeast Pual and Vienna Woods hydrothermal fields, Manus Basin, Papua New Guinea. *Geochimica et Cosmochimica Acta*, *75*(4), 1088-1123. https://doi.org/10.1016/j.gca.2010.11.008

Reynolds, K. C., Watanabe, H., Strong, E. E., Sasaki, T., Uematsu, K., Miyake, H., Kojima, S., Suzuki, Y., Fujikura, K., Kim, S., and 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

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., and Ripley, M. B. (2013). Package 'mass'. *Cran r*, *538*, 113-120.

Riquet, F., Simon, A., and Bierne, N. (2017). Weird genotypes? Don't discard them, transmissible cancer could be an explanation. *Evolutionary Applications*, *10*(2), 140-145. https://doi.org/10.1111/eva.12439

Rocha, L. A., and Bowen, B. W. (2008). Speciation in coral-reef fishes. *Journal of Fish Biology*, *72*(5), 1101-1121. https://doi.org/10.1111/j.1095-8649.2007.01770.x

Rogers, A. D., Tyler, P. A., Connelly, D. P., Copley, J. T., James, R., Larter, R. D., Linse, K., Mills, R. A., Garabato, A. N., Pancost, R. D., Pearce, D. A., Polunin, N. V. C., German, C. R., Shank, T., Boersch-Supan, P. H., Alker, B. J., Aquilina, A., Bennett, S. A., Clarke, A., … Zwirglmaier, K. (2012). The discovery of new deep-sea hydrothermal vent communities in the Southern Ocean and complications for biogeography. *Plos Biology*, *10*(1), e1001234. https://doi.org/10.1371/journal.pbio.1001234

Rolán-Alvarez, E. (2007). Sympatric speciation as a by-product of ecological adaptation in the Galician *Littorina saxatilis* hybrid zone. *Journal of Molluscan Studies*, *73*(1), 1-10. https://doi.org/10.1093/mollus/eyl023

Rolán-Alvarez, E., Johannesson, K., and Erlandsson, J. (1997). The maintenance of a cline in the marine snail *Littorina saxatilis*: The role of home site advantage and hybrid fitness. *Evolution*, *51*(6), 1838-1847. https://doi.org/10.1111/j.1558-5646.1997.tb05107.x

Rosenberg, N. A. (2003). The shapes of neutral gene genealogies in two Species : Probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. *Evolution*, *57*(7), 1465-1477. https://doi.org/10.1111/j.0014-3820.2003.tb00355.x

Roux, C., Fraïsse, C., Romiguier, J., Anciaux, Y., Galtier, N., and Bierne, N. (2016). Shedding light on the grey zone of speciation along a continuum of genomic divergence. *Plos Biology*, *14*(12), e2000234. https://doi.org/10.1371/journal.pbio.2000234

Roux, C., Tsagkogeorga, G., Bierne, N., and Galtier, N. (2013). Crossing the species barrier : Genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. *Molecular Biology and Evolution*, *30*(7), 1574-1587. https://doi.org/10.1093/molbev/mst066

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., and Sánchez-Gracia, A. (2017). DnaSP 6 : DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, *34*(12), 3299-3302. https://doi.org/10.1093/molbev/msx248

S

Sales, L., and Queiroz, V. (2021). Sexual dimorphism in the parasitic snail *Nanobalcis worsfoldi* : A histological and morphometric approach with insights for the family Eulimidae. *Canadian Journal of Zoology*, *99*(11), 995-1001. https://doi.org/10.1139/cjz-2021-0118

Sanford, E., and Kelly, M. W. (2011). Local adaptation in marine invertebrates. *Annual Review of Marine Science*, *3*(1), 509-535. https://doi.org/10.1146/annurev-marine-120709-142756

Sanford, E., and Worth, D. J. (2009). Genetic differences among populations of a marine snail drive geographic variation in predation. *Ecology*, *90*(11), 3108-3118. https://doi.org/10.1890/08-2055.1

Sarradin, P.-M., Lannuzel, D., Waeles, M., Crassous, P., Le Bris, N., Caprais, J. C., Fouquet, Y., Fabri, M. C., and Riso, R. (2008). Dissolved and particulate metals (Fe, Zn, Cu, Cd, Pb) in two habitats from an active hydrothermal field on the EPR at 13°N. *Science of The Total Environment*, *392*(1), 119-129. https://doi.org/10.1016/j.scitotenv.2007.11.015

Sarre, S. D., Georges, A., and Quinn, A. (2004). The ends of a continuum : Genetic and temperature-dependent sex determination in reptiles. *BioEssays*, *26*(6), 639-645. https://doi.org/10.1002/bies.20050

Sassaman, C., and Weeks, S. C. (1993). The genetic mechanism of sex determination in the Conchostracan shrimp *Eulimnadia texana*. *The American Naturalist*, *141*(2), 314-328. https://doi.org/10.1086/285475

Savolainen, O., Lascoux, M., and Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, *14*(11), 807-820. https://doi.org/10.1038/nrg3522

Scally, A., and Durbin, R. (2012). Revising the human mutation rate : Implications for understanding human evolution. *Nature Reviews Genetics*, *13*(10), 745-753. https://doi.org/10.1038/nrg3295

Schellart, W. P., Lister, G. S., and Toy, V. G. (2006). A late Cretaceous and Cenozoic reconstruction of the Southwest Pacific region : Tectonics controlled by subduction and slab rollback processes. *Earth-Science Reviews*, *76*(3), 191-233. https://doi.org/10.1016/j.earscirev.2006.01.002

Schliewen, U. K., Tautz, D., and Pääbo, S. (1994). Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature*, *368*(6472), 629-632. https://doi.org/10.1038/368629a0

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, Å., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., … Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, *15*(3), 176-192. https://doi.org/10.1038/nrg3644

Segonzac, M. (1992). Les peuplements associés a l'hydrothermalisme oceanique du Snake Pit (dorsale medio-atlantique ; 23°N, 3480 m) : Composition et microdistribution de la megafaune. *C R Acad Sci Paris Ser III*, *314*, 593-600.

Sell, A. F. (2000). Life in the extreme environment at a hydrothermal vent : Haemoglobin in a deep-sea copepod. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *267*(1459), 2323-2326. https://doi.org/10.1098/rspb.2000.1286

Selz, O. M., Pierotti, M. E. R., Maan, M. E., Schmid, C., and Seehausen, O. (2014). Female preference for male color is necessary and sufficient for assortative mating in two cichlid sister species. *Behavioral Ecology*, *25*(3), 612-626. https://doi.org/10.1093/beheco/aru024

Sen, A., Podowski, E. L., Becker, E. L., Shearer, E. A., Gartman, A., Yücel, M., Hourdez, S., Luther, I., George W., and Fisher, C. R. (2014). Community succession in hydrothermal vent habitats of the Eastern Lau spreading center and Valu Fa ridge, Tonga. *Limnology and Oceanography*, *59*(5), 1510-1528. https://doi.org/10.4319/lo.2014.59.5.1510

Shaheen, N., Patel, K., Patel, P., Moore, M., and Harrington, M. A. (2005). A predatory snail distinguishes between conspecific and heterospecific snails and trails based on chemical cues in slime. *Animal Behaviour*, *70*(5), 1067-1077. https://doi.org/10.1016/j.anbehav.2005.02.017

Shank, T. M., Fornari, D. J., Von Damm, K. L., Lilley, M. D., Haymon, R. M., and Lutz, R. A. (1998). Temporal and spatial patterns of biological community development at nascent deep-sea hydrothermal vents (9°50′N, East Pacific Rise). *Deep Sea Research Part II: Topical Studies in Oceanography*, *45*(1), 465-515. https://doi.org/10.1016/S0967-0645(97)00089-1

Shropshire, J. D., and Bordenstein, S. R. (2016). Speciation by symbiosis : The microbiome and behavior. *mBio*, *7*(2), e01785-15. https://doi.org/10.1128/mBio.01785-15

Smith, C. A., and Sinclair, A. H. (2004). Sex determination : Insights from the chicken. *BioEssays*, *26*(2), 120-132. https://doi.org/10.1002/bies.10400

Smolensky, N., Romero, M. R., and Krug, P. J. (2009). Evidence for costs of mating and self-fertilization in a simultaneous hermaphrodite with hypodermic insemination, the Opisthobranch *Alderia willowi*. *The Biological Bulletin*, *216*(2), 188-199. https://doi.org/10.1086/BBLv216n2p188

Sokolova, I., and Pörtner, H. (2001). Temperature effects on key metabolic enzymes in *Littorina saxatilis* and *L. obtusata* from different latitudes and shore levels. *Marine Biology*, *139*(1), 113-126. https://doi.org/10.1007/s002270100557

Sommer, S. A., Van Woudenberg, L., Lenz, P. H., Cepeda, G., and 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

Spencer, P. S., and Barral, J. M. (2012). Genetic code redundancy and its influence on the encoded polypeptides. *Computational and Structural Biotechnology Journal*, *1*(1), e201204006. https://doi.org/10.5936/csbj.201204006

Springer, S. A., and Crespi, B. J. (2007). Adaptive gamete-recognition divergence in a hybridizing *Mytilus* Population. *Evolution*, *61*(4), 772-783. https://doi.org/10.1111/j.1558-5646.2007.00073.x

Stamatakis, A. (2014). RAxML version 8 : A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312-1313. https://doi.org/10.1093/bioinformatics/btu033

Stelkens, R., and Seehausen, O. (2009). Genetic distance between species predicts novel trait expression in their hybrids. *Evolution*, *63*(4), 884-897. https://doi.org/10.1111/j.1558-5646.2008.00599.x

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., and 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

Suzuki, Y., Sasaki, T., Suzuki, M., Nogi, Y., Miwa, T., Takai, K., Nealson, K. H., and Horikoshi, K. (2005). Novel chemoautotrophic endosymbiosis between a member of the *Epsilonproteobacteria* and the hydrothermal-vent gastropod *Alviniconcha hessleri* (Gastropoda : Provannidae) from the Indian Ocean. *Applied and Environmental Microbiology*, *71*(9), 5440-5450. https://doi.org/10.1128/AEM.71.9.5440-5450.2005

T

Tasiemski, A., Jung, S., Boidin-Wichlacz, C., Jollivet, D., Cuvillier-Hot, V., Pradillon, F., Vetriani, C., Hecht, O., Sönnichsen, F. D., Gelhaus, C., Hung, C.-W., Tholey, A., Leippe, M., Grötzinger, J., and Gaill, F. (2014). Characterization and function of the first antibiotic isolated from a vent organism : The extremophile metazoan *Alvinella pompejana*. *Plos One*, *9*(4), e95737. https://doi.org/10.1371/journal.pone.0095737

Taylor, B., Goodliffe, A., Martinez, F., and Hey, R. (1995). Continental rifting and initial sea-floor spreading in the Woodlark Basin. *Nature*, *374*(6522), 534-537. https://doi.org/10.1038/374534a0

Telschow, A., Gadau, J., Werren, J. H., and Kobayashi, Y. (2019). Genetic incompatibilities between mitochondria and nuclear genes : Effect on gene flow and speciation. *Frontiers in genetics*, *10*, 62.

Thioulouse, J., Dray, S., Dufour, A.-B., Siberchicot, A., Jombart, T., and Pavoine, S. (2018). Multivariate analysis of ecological data with ade4. *Springer New York*. https://doi.org/10.1007/978-1-4939-8850-1

Thiriot-Quiévreux, C. (2003). Advances in chromosomal studies of gastropod molluscs. *Journal of Molluscan Studies*, *69*(3), 187-202. https://doi.org/10.1093/mollus/69.3.187

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface : Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, *25*(24), 4876-4882. https://doi.org/10.1093/nar/25.24.4876

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., and 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

Tran Lu Y, A., Ruault, S., Daguin-Thiébaut, C., Castel, J., Bierne, N., Broquet, T., Wincker, P., Perdereau, A., Arnaud-Haond, S., Gagnaire, P.-A., Jollivet, D., Hourdez, S., and Bonhomme, F. (2022). Subtle limits to connectivity revealed by outlier loci within two divergent metapopulations of the deep-sea hydrothermal gastropod *Ifremeria nautilei*. *Molecular Ecology*. https://doi.org/10.1111/mec.16430

Trussell, G. C. (2000). Phenotypic clines, plasticity, and morphological trade-offs in an intertidal snail. *Evolution*, *54*(1), 151-166. https://doi.org/10.1111/j.0014-3820.2000.tb00016.x

Trussell, G. C. (2002). Evidence of countergradient variation in the growth of an intertidal snail in response to water velocity. *Marine Ecology Progress Series*, *243*, 123-131. https://doi.org/10.3354/meps243123

Tsuchida, T., Koga, R., and Fukatsu, T. (2004). Host plant specialization governed by facultative symbiont. *Science*, *303*(5666), 1989-1989. https://doi.org/10.1126/science.1094611

Tunnicliffe, V. (1991). The biology of hydrothermal vents : Ecology and evolution. *Oceanogr Mar Biol Annu Rev*, *29*, 319-407.

Tunnicliffe, V., Botros, M., De Burgh, M. E., Dinet, A., Johnson, H. P., Juniper, S. K., and McDuff, R. E. (1986). Hydrothermal vents of explorer ridge, Northeast Pacific. *Deep Sea Research Part A. Oceanographic Research Papers*, *33*(3), 401-412. https://doi.org/10.1016/0198-0149(86)90100-7

Tunnicliffe, V., Fowler, C. M. R., and Mcarthur, A. G. (1996). Plate tectonic history and hot vent biogeography. *Geological Society, London, Special Publications*, *118*(1), 225-238. https://doi.org/10.1144/GSL.SP.1996.118.01.14

Turner, T. L., Hahn, M. W., and Nuzhdin, S. V. (2005). Genomic islands of speciation in *Anopheles gambiae*. *Plos Biology*, *3*(9), e285. https://doi.org/10.1371/journal.pbio.0030285

Tyler, P. A., and 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

U/V

Urakawa, H., Dubilier, N., Fujiwara, Y., Cunningham, D. E., Kojima, S., and Stahl, D. A. (2005). Hydrothermal vent gastropods from the same family (Provannidae) harbour ε- and γ-

proteobacterial endosymbionts. *Environmental Microbiology*, *7*(5), 750-754. https://doi.org/10.1111/j.1462-2920.2005.00753.x

Vacquier, V. D., Carner, K. R., and Stout, C. D. (1990). Species-specific sequences of abalone lysin, the sperm protein that creates a hole in the egg envelope. *Proceedings of the National Academy of Sciences*, *87*(15), 5792-5796. https://doi.org/10.1073/pnas.87.15.5792

Van Dover, C. L., and Hessler, R. R. (1990). Spatial variation in faunal composition of hydrothermal vent communities on the East Pacific Rise and Galapagos spreading center. In G. R. McMurray (Éd.), *Gorda Ridge* (p. 253-264). Springer. https://doi.org/10.1007/978-1-4612-3258-2_18

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 Dover, C. L. V. (2000). The ecology of deep-sea hydrothermal vents. *Princeton University Press*. https://doi.org/10.1515/9780691239477

Via, S. (2009). Natural selection in action during speciation. *Proceedings of the National Academy of Sciences*, *106*(supplement_1), 9939-9946. https://doi.org/10.1073/pnas.0901397106

Von Damm, K. L. (1990). Seafloor hydrothermal activity : Black smoker chemistry and chimneys. *Annual Review of Earth and Planetary Sciences*, *18*(1), 173-204. https://doi.org/10.1146/annurev.ea.18.050190.001133

Von Cosel, R., and Olu, K. (1998). Gigantism in *Mytilidae*. A new *Bathymodiolus* from cold seep areas on the Barbados accretionary Prism. *Comptes Rendus de l'Académie Des Sciences - Series III - Sciences de La Vie*, *321*(8), 655-663. https://doi.org/10.1016/S0764-4469(98)80005-X

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. (2009). Cryptic species, phenotypic plasticity, and complex life histories : Assessing deep-sea faunal diversity with molecular markers. *Deep Sea Research Part II: Topical Studies in Oceanography*, *56*(19), 1713-1723. https://doi.org/10.1016/j.dsr2.2009.05.016

Vuillemin, R., Le Roux, D., Dorval, P., Bucas, K., Sudreau, J. P., Hamon, M., Le Gall, C., and Sarradin, P. M. (2009). CHEMINI : A new in situ CHEmical MINIaturized analyzer. *Deep Sea Research Part I: Oceanographic Research Papers*, *56*(8), 1391-1399. https://doi.org/10.1016/j.dsr.2009.02.002

W

Wadlington, W. H., and Ming, R. (2018). Development of an X-specific marker and identification of YY individuals in spinach. *Theoretical and Applied Genetics*, *131*(9), 1987-1994. https://doi.org/10.1007/s00122-018-3127-1

Walters, W., Hyde, E. R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., Gilbert, J. A., Jansson, J. K., Caporaso, J. G., Fuhrman, J. A., Apprill, A., and Knight, R. (s. d.). Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *mSystems*, *1*(1), e00009-15. https://doi.org/10.1128/mSystems.00009-15

Warèn, A., and 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

Wares, J. P., and Cunningham, C. W. (2001). Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution*, *55*(12), 2455-2469. https://doi.org/10.1111/j.0014-3820.2001.tb00760.x

Watremez, P., and Kervevan, C. (1990). Origine des variations de l'activité hydrothermale : Premiers éléments de réponse d'un modèle numérique simple. *Comptes rendus de l'Académie des sciences. Série 2, Mécanique, Physique, Chimie, Sciences de l'univers, Sciences de la Terre*, *311*(1), 153-158.

Weeks, S. C. (2012). The role of androdioecy and gynodioecy in mediating evolutionary transitions between dioecy and hermaphroditism in the Animalia. *Evolution*, *66*(12), 3670-3686. https://doi.org/10.1111/j.1558-5646.2012.01714.x

Weeks, S. C., Benvenuto, C., and Reed, S. K. (2006). When males and hermaphrodites coexist : A review of androdioecy in animals. *Integrative and Comparative Biology*, *46*(4), 449-464. https://doi.org/10.1093/icb/icj048

Welschmeyer, N. (2009). Reproductive strategies of the dominant gastropods of the Lau Basin hydrothermal vent system: *Alviniconcha hessleri* and *Ifremeria nautilei* [PhD Thesis]. California State University Monterey Bay.

Wersebe, M. J., Sherman, R. E., Jeyasingh, P. D., and Weider, L. J. (2022). The roles of recombination and selection in shaping genomic divergence in an incipient ecological species complex. *Molecular Ecology*. https://doi.org/10.1111/mec.16383

Williams, D. L., Green, K., van Andel, T. H., von Herzen, R. P., Dymond, J. R., and Crane, K. (1979). The hydrothermal mounds of the Galapagos Rift : Observations with DSRV Alvin and detailed heat flow studies. *Journal of Geophysical Research: Solid Earth*, *84*(B13), 7467-7484. https://doi.org/10.1029/JB084iB13p07467

Wilson, G. D. F., and Hessler, R. R. (1987). Speciation in the Deep Sea. *Annual Review of Ecology and Systematics*, *18*(1), 185-207. https://doi.org/10.1146/annurev.es.18.110187.001153

Wood, D. E., and Salzberg, S. L. (2014). Kraken : Ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, *15*(3), R46. https://doi.org/10.1186/gb-2014-15-3-r46

Woodruff, G. C., Eke, O., Baird, S. E., Félix, M.-A., and Haag, E. S. (2010). Insights into species divergence and the evolution of hermaphroditism from fertile interspecies hybrids of *Caenorhabditis* nematodes. *Genetics*, *186*(3), 997-1012. https://doi.org/10.1534/genetics.110.120550

Wright, S. (1951). The genetical structure of species. *Ann. Eugen*, *15*(323-354), 747.

Wright, S. I., Bi, I. V., Schroeder, S. G., Yamasaki, M., Doebley, J. F., McMullen, M. D., and Gaut, B. S. (2005). The effects of artificial selection on the Maize genome. *Science*, *308*(5726), 1310-1314. https://doi.org/10.1126/science.1107891

Wu, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, *14*(6), 851-865. https://doi.org/10.1046/j.1420-9101.2001.00335.x

X/Y/Z

Xiao, J., Zhong, H., Liu, Z., Yu, F., Luo, Y., Gan, X., and Zhou, Y. (2015). Transcriptome analysis revealed positive selection of immune-related genes in *Tilapia*. *Fish and Shellfish Immunology*, *44*(1), 60-65. https://doi.org/10.1016/j.fsi.2015.01.022

Yamada, S. B. (1989). Are direct developers more locally adapted than planktonic developers? *Marine Biology*, *103*(3), 403-411. https://doi.org/10.1007/BF00397275

Yamasaki, S., Fujii, N., Matsuura, S., Mizusawa, H., and Takahashi, H. (2001). The M locus and ethylene-controlled sex determination in andromonoecious cucumber plants. *Plant and Cell Physiology*, *42*(6), 608-619. https://doi.org/10.1093/pcp/pce076

Yang, Z., and dos Reis, M. (2011). Statistical properties of the branch-site test of positive selection. *Molecular Biology and Evolution*, *28*(3), 1217-1228. https://doi.org/10.1093/molbev/msq303

Yang, Z., and Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution*, *17*(1), 32-43. https://doi.org/10.1093/oxfordjournals.molbev.a026236

Yusa, Y. (2007). Nuclear sex-determining genes cause large sex-ratio variation in the apple snail *Pomacea canaliculata*. *Genetics*, *175*(1), 179-184. https://doi.org/10.1534/genetics.106.060400

Zielske, S., and Haase, M. (2014). When snails inform about geology : Pliocene emergence of islands of Vanuatu indicated by a radiation of truncatelloidean freshwater gastropods (Caenogastropoda: Tateidae). *Journal of Zoological Systematics and Evolutionary Research*, *52*(3), 217-236. https://doi.org/10.1111/jzs.12053

Zierenberg, R. A., Adams, M. W. W., and Arp, A. J. (2000). Life in extreme environments : Hydrothermal vents. *Proceedings of the National Academy of Sciences*, *97*(24), 12961-12962. https://doi.org/10.1073/pnas.210395997

Résumé

Les sources hydrothermales ne sont pas seulement des environnements qui imposent des conditions de vie extrêmes aux espèces qui leurs sont inféodées. Ils sont également fragmentés, localement éphémères et hétérogènes, ce qui en fait un système d'étude intéressant pour étudier les rôles relatifs de l'isolement géographique et de la variabilité environnementale sur l'évolution des espèces et les mécanismes de spéciation. Dans le Pacifique Ouest, parmi les espèces peuplant l'environnement hydrothermal, les *Alviniconcha* (Gastropoda : Abyssochrysoida) forment un complexe de 5 espèces architectes abondantes. Ces gros gastéropodes symbiotiques, occupent le pôle chaud (7 – 42°C), soufré (250µM) et peu oxygéné (< 50µM) de cet écosystème particulier. Au cours de ma thèse, j'ai étudié l'impact de la géographie et de l'hétérogénéité spatiale de l'environnement hydrothermal sur la distribution, la divergence, l'adaptation locale et les conditions de la spéciation de trois espèces d'*Alviniconcha* (*A. kojimai*, *A. strummeri* et *A. boucheti*) dans cinq bassins arrière-arc du Pacifique sud-ouest. L'étude du transcriptome, du polymorphisme génomique nucléaire, et d'un fragment de gène du génome mitochondrial (*Cox1*) a montré une divergence élevée entre les espèces ciblées qui renforce l'hypothèse d'une origine allopatrique de ces espèces. Mais l'inférence génétique du scénario démographique le plus probable suggère également que la divergence en allopatrie a ensuite été suivie d'une reprise partielle des flux de gènes par hybridation introgressive lors de la colonisation de ce milieu. Plusieurs mécanismes potentiellement impliqués dans les barrières à la reproduction entre espèces ont été étudiés. En premier lieu, nous avons mis en évidence un système de déterminisme génétique du sexe hétérogamétique mâle (XY) chez *A. boucheti* et *A. strummeri*, qui sont toutes les deux gonochoriques, mais ce déterminisme est moins clair chez *A. kojimai*, qui semble être une espèce androdioïque. Par ailleurs, l'adaptation locale des espèces à différentes niches écologiques aurait pu renforcer leur isolement géographique lors de leur remise en contact. En effet, chez *A. boucheti* les gènes identifiés sous sélection positive sont plus spécifiquement liés à la symbiose et aux stress environnementaux. Cela pourrait être lié à la mise en place d'une symbiose particulière avec des *Campylobacteria* qui tolèrent de plus fortes concentrations en H_2 et en H_2S et expliquerait pourquoi cette espèce est présente au plus proche des émissions contrairement aux 2 autres espèces qui, elles, partagent le même habitat avec une évolution positive récente de leurs mécanismes pré-zygotiques.

Mots clés: spéciation, divergence, introgression, adaptation locale, environnement hydrothermal, *Alviniconcha*

Abstract

Hydrothermal vents are not only environments that impose extreme living conditions on the species. They are also fragmented, locally ephemeral, and heterogeneous, making them an interesting study system to investigate the relative roles of geographic isolation and environmental variability on species evolution and speciation mechanisms. In the western Pacific, among the species inhabiting the hydrothermal environment, *Alviniconcha* (Gastropoda: Abyschrysoida) form a complex of 5 abundant architectural species. These large symbiotic gastropods, occupy the warm (7 - 42°C), sulfurous (250µM) and low oxygened (< 50µM) pole of this particular ecosystem. In my thesis, I investigated the impact of geography and spatial heterogeneity of the hydrothermal environment on the distribution, divergence, local adaptation, and speciation conditions of three *Alviniconcha* species (*A. kojimai*, *A. strummeri*, and *A. boucheti*) in five back-arc basins of the Southwest Pacific. Studies of the transcriptome, nuclear genomic polymorphism, and a gene fragment of the mitochondrial genome (*Cox1*) showed high divergence among the targeted species that reinforces the hypothesis of an allopatric origin of these species. But genetic inference of the most likely demographic scenario also suggests that divergence into allopatry was subsequently followed by partial recovery of gene flow by introgressive hybridization during the colonization of this environment. Several mechanisms potentially involved in interspecies reproductive barriers have been investigated. First, we found a system of genetic determinism of heterogametic male sex (XY) in *A. boucheti* and *A. strummeri*, which are both gonochoric, but this determinism is less clear in *A. kojimai*, which appears to be an androdiotic species. Furthermore, the local adaptation of the species to different ecological niches could have reinforced their geographical isolation when they were recontacted. Indeed, in *A. boucheti* the genes identified under positive selection are more specifically related to symbiosis and environmental stress. This could be related to the establishment of a particular symbiosis with *Campylobacteria* that tolerate higher concentrations of H_2 and H_2S and would explain why this species is present closer to the emissions contrary the other two species that share the same habitat with a recent positive evolution of their prezygotic mechanisms.

Key words: speciation, divergence, introgression, local adaptation, hydrothermal environment, *Alviniconcha*