



Description de la diversité microbienne associée à la crevette *Rimicaris chacei* : une possible double symbiose

Vincent Apremont

► To cite this version:

Vincent Apremont. Description de la diversité microbienne associée à la crevette *Rimicaris chacei* : une possible double symbiose. Microbiologie et Parasitologie. Université de Bretagne occidentale - Brest, 2017. Français. NNT : 2017BRES0107 . tel-02918016

HAL Id: tel-02918016

<https://theses.hal.science/tel-02918016>

Submitted on 20 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse préparée à l'Université de Bretagne Occidentale

pour obtenir le diplôme de DOCTEUR délivré de façon partagée par
L'Université de Bretagne Occidentale et l'Université de Bretagne Loire

présentée par

Vincent Apremont

Spécialité : Microbiologie

Description de la diversité microbienne associée à la crevette *Rimicaris chacei* : une possible double symbiose

Thèse soutenue le 29 Novembre 2017

devant le jury composé de :

Sébastien Duperron (Rapporteur)

Professeur MNHN, Muséum National d'Histoire Naturelle

Eric Bonnivard (Rapporteur)

Maître de conférences UPMC, Université Pierre et Marie Curie

Mohamed Jebbar (Examineur)

Professeur des Universités, Université Bretagne Occidentale

Marie-Anne Cambon-Bonavita (Directrice de thèse)

Chercheur C2B, HDR, Ifremer Centre Bretagne

Magali Zbinden (Directrice de thèse)

Maître de conférences, UPMC, Université Pierre et Marie Curie

Laure Corbari (invité)

Maître de conférences, MNHN, Muséum National d'Histoire Naturelle

Gaëtan Burgaud (invité)

Maître de conférences, ESIAB Technopôle Brest Iroise

Michel Segonzac (Invité)

Attaché honoraire au département Systématique et évolution, Muséum National d'Histoire Naturelle



Remerciements

Faire la liste de toutes les personnes que j'ai rencontrées, avec qui j'ai pu discuter, ou encore que j'aie pu embêter/taquiner/charrier/pranker durant cette thèse serait trop long. Il ne faut donc pas te sentir visé si tu ne fais pas partie de la liste suivante (en plus, les déclarations c'est pas mon truc). Je remercie dans un premier temps tous les membres des laboratoires que sont le LM2E et AMEX pour leur accueil et leur bonne humeur de tous les jours. Je remercie avec un peu plus d'amour les membres de l'équipe symbiose avec qui j'ai pu former une équipe que j'ai trouvé bien soudée. Je remercie mon voisin de bureau, Simon avec qui la cohabitation n'a peut-être pas été symbiotique (pour rester dans la thématique) mais pas loin ; ainsi que Christine, pour ses sourires et ses rires révivificateurs. Je souhaite tout particulièrement remercier mes parents et ma famille pour m'avoir soutenue durant toutes ces années d'études qui m'ont permis d'en arriver où j'en suis aujourd'hui. Mais je ne peux aussi que remercier mes deux directrices de thèse, qui m'ont permis de réaliser ce sujet de thèse, qui ne m'ont pas lâché une seule fois durant ces trois années de thèse, et pour ne pas m'avoir étranglé, égorgé, pendu... pour toutes les fois où j'aurai pu le mériter. Cette thèse, je vous la dois beaucoup, à tous.

Je change volontairement de paragraphe pour remercier une dernière personne sans qui rien n'aurait été possible. On dit que derrière chaque homme, il y a une femme. Je ne fais pas exception à cette règle. Mais cette femme est particulièrement exceptionnelle. Elle m'a soutenu dans les pires moments, a su être présente à chaque fois que j'en ai eu besoin, m'a remis dans le droit chemin quand il le fallait, et a été mon ancre durant ces trois ans : Claire Bourigault (qui changera bientôt pour Apremont). Et c'est à toi que je souhaite dédier cette thèse, à mon amour de toujours.

Table des matières

Remerciements.....	II
Liste des figures	1
Lexique	3
Introduction	4
L'océan profond	5
L'hydrothermalisme océanique profond	6
Répartition et fonctionnement.....	6
Les systèmes hydrothermaux de l'Atlantique	9
Systèmes ultramafiques <i>versus</i> systèmes basaltiques	10
Diversité des types d'émissions de fluide.....	11
Les microorganismes, base du réseau trophique	14
Où trouver des microorganismes en contexte hydrothermal	14
La chimiosynthèse	14
La macrofaune, représentant visible des systèmes hydrothermaux	20
La symbiose, concept important des environnements hydrothermaux	21
Le modèle <i>Rimicaris exoculata</i>	23
Le modèle <i>Rimicaris chacei</i>	31
Le barcoding/metabarcoding dans le monde microbien.....	34
Le sujet de thèse.....	43
Matériels et Méthodes : Les scripts générés pour traiter les données de manière automatique ou semi-automatique	46
Protocole de barcoding sur banque de clones.....	47
Elimination des séquences de mauvaise qualité	48
Tri des séquences	48
« Nettoyage » des séquences	54
Détection de chimères	56
Analyse des données avec Mothur.....	57
Des résultats complets à récupérer.....	65
Workflow Mothur	65
Les protocoles de metabarcoding	67
Pourquoi utiliser différentes architectures informatiques afin de réaliser la même chose ?	67
Les scripts	68
1 ^{er} article : First description of <i>Chorocaris (Rimicaris) chacei</i> symbiont : high morphological and genetic similarity with <i>Rimicaris exoculata</i> (MAR).....	78

Introduction à la problématique	79
Principaux résultats	2
Discussions et perspectives liées à cet article	3
2 ^{ème} Article : Biogéographie comparative des symbiontes de crevettes hydrothermales <i>Rimicaris exoculata</i> et <i>Rimicaris chacei</i> le long de la MAR.....	5
Introduction à la problématique	6
Matériel et méthodes	7
Echantillonnage.....	7
Extraction d'ADN/PCR/Librairies.....	7
Analyses des données	12
Premiers résultats	12
Discussion Générale	21
Discussion scientifique	21
Perspectives scientifiques	25
Discussion personnelle	26
Conclusion générale	28
Bibliographie.....	29
Annexes	44
Bioinformatique associée au barcoding/metabarcoding	45
Le contexte général.....	46

Liste des figures

Figure 1 : Répartition géographique des systèmes hydrothermaux océaniques profonds autour du globe. (Source NOAA).....	6
Figure 2 : Représentation schématique des différents types d'hydrothermalismes (modifiée de l'USGS)	7
Figure 3 : Composition minérale et gazeuse de fluides hydrothermaux de la MAR. Le site MARK-1/2 est aujourd'hui appelé Snake Pit (Charlou et al., 2002).	11
Figure 4 : Diagramme schématique (A) d'un fumeur, (B) d'une ruche, (C) d'un rebord. (Hannington et al., 1995).....	13
Figure 5 : Photographie du modèle <i>Rimicaris exoculata</i>	24
Figure 6 : <i>Rimicaris exoculata</i> , coupe transversale schématique du céphalothorax au niveau du bord antérieur de l'estomac. (modifié de Segonzac et al., 1993).	28
Figure 7: Représentation schématique des différentes parties du système digestif de la crevette <i>Rimicaris exoculata</i>	30
Figure 8 : A gauche, photographie du céphalothorax de <i>Rimicaris chacei</i> . A droite, dessin de description de <i>Rimicaris chacei</i> (Williams and Rona, 1986).	32
Figure 9 : Comparaison schématique de la rapidité et du stockage associé aux différents types de mémoire d'un ordinateur.....	47
Figure 10 : Tableau comparant les avantages et inconvénients des systèmes Windows, Linux, et Mac OS pour faire de la bioinformatique.	51
Figure 11 : Schéma de l'histoire de différents systèmes d'exploitation.....	52
Figure 12 : Représentation schématique des différentes sous unités de l'insuline et de sa transformation post-traduction.....	36
Figure 13 : Représentation schématique de l'horloge évolutive de Fibrinopeptides, de l'Hémoglobine et du cytochrome c au cours du temps.	37
Figure 14 : Tableau récapitulant les différents biais méthodologiques pouvant être rencontrés avant un traitement bioinformatique de données de barcoding et de metabarcoding.	38
Figure 15 : Représentation schématique du vecteur de clonage TOPO TA®.	49
Figure 16 : Capture d'écran d'une portion de séquence de mauvaise qualité au format .AB1.	49
Figure 17 : Principe schématique d'une étape de clonage.....	50
Figure 18 : Capture d'écran d'un alignement réalisé avec l'algorithme Geneious.....	51
Figure 19 : Capture d'écran de la recherche d'une amorce PCR dans un lot de séquences sous Geneious... ..	52
Figure 20 : Capture d'écran de la visualisation d'une amorce dans un lot de séquences sous Geneious.	53
Figure 21 : Capture d'écran montrant la barre d'identité d'un lot de séquences sous Geneious.	54
Figure 22 : Capture d'écran d'une zone d'alignement mettant en évidence un polymère de Guanine devant être corrigé.	55
Figure 23 : Capture d'écran d'une zone d'alignement mettant en évidence une erreur d'identification d'un nucléotide à partir d'un chromatogramme.....	56
Figure 24 : Schéma récapitulatif du protocole de barcoding sous Mothur.....	57
Figure 25 : Ensemble de captures d'écran permettant la création manuelle d'un fichier .name pour Mothur	59
Figure 26 : Capture d'écran d'un résultat de commande summary.seqs sous Mothur.	60
Figure 27 : Représentation de la similarité des échantillons présentés dans le premier article de cette thèse avec l'indice de diversité Bray-Curtis sous la forme d'une heatmap.	64
Figure 28 : Représentation schématique du principe global des scripts générés durant cette thèse.	69
Figure 29 : Représentation schématique d'exemples de données pouvant être reçues d'un prestataire de séquençage haut débit.	70

Figure 30 : Schéma simplifié du protocole de traitement de données de type metabarcoding sous Mothur.	72
Figure 31 : Schéma simplifié du protocole de traitement de données de type metabarcoding sous Qiime..	75
Figure 32 : Schéma simplifié du protocole de traitement de données de type metabarcoding sous FROGS.	76
Figure 33 : Tableau récapitulatif des échantillons ayant été séquencés via la technologie illumina.	11
Figure 34 : Etape de filtrage qualité du workflow FROGS sur l'ensemble du jeu de données NGS (en millions de séquences).	13
Figure 35 : Recherche des chimères avec l'outil SWARM.	16
Figure 36 : Filtrage des données NGS selon un paramètre d'abondance.	17
Figure 37 : Courbes de raréfaction NGS.	17
Figure 38 : Dendrogramme de similarité des échantillons NGS à partir de l'indice de Bray-Curtis. (En bleu sont représentés les échantillons de <i>Rimicaris exoculata</i> , et en rouge les échantillons de <i>Rimicaris chacei</i>)	19

Lexique

ADN : Acide désoxyribonucléotide

AMEX : Adaptations aux Milieux EXtrêmes

BICOSE : Biodiversité, Interactions, COnnectivité et Symbioses en milieux Extrêmes

BioBaz : Bio**b**ologie intégrée de Bathymodiolus azoricus

BLAST : Basic Local Alignment Search Tool

FISH : Fluorescent In Situ Hybridization

FROGS : Find Rapidly OTU with Galaxy Solution

Go : Giga octet

Ifremer : Institut français de recherche et d'exploitation de la mer

LB : Lame Branchiostège

LM2E : Laboratoire de Microbiologie des Environnements Extrêmes

MEB : Microscope Electronique à Balayage

MET : Microscope Electronique à Transmission

NCBI : National Center for Biotechnology Information

OTU : Operationnal taxonomy unit

pb : paire de base

PCR : Polymerase chain reaction

Sc : Scaphognathite

To : Téra octet

UBO : Université de Bretagne Occidentale

UPMC : Université Pierre et Marie Curie

TD : Tube Digestif

EDX : Energy Dispersive X-ray spectrometry

Introduction

L'océan profond

Durant de nombreux siècles, l'Homme a entretenu l'idée que les océans étaient un abîme sans fond. Ferdinand Magellan, grand explorateur du XVI^{ème} siècle, et premier homme (avec son équipage) à faire le tour du globe en 1521, a encouragé cette idée, suite à une expérience menée durant cette expédition. Il réalisa des sondages bathymétriques au milieu du Pacifique à l'aide d'un cordage plombé de 800 m, et fut contraint d'admettre qu'il n'était pas en mesure de mettre en évidence la preuve de l'existence d'un fond de la zone étudiée (Calgano, 2011). Ce que ne savait pas Magellan à cette époque, c'est que l'océan Pacifique présente en très grande majorité une profondeur supérieure à 800 m (Watling *et al.*, 2013). Forbes émit plus tard, suite à ses travaux à bord du HMS Beacon en 1843, la célèbre Théorie Azoïque, dans laquelle il démontra via une étude effectuée sur des végétaux aquatiques que la vie dans les océans ne peut se développer au-delà de 550 m (Forbes, 1843, Forbes, 1844, Forbes, 1859). Le manque de communication scientifique à cette époque ne lui avait pas permis d'apprendre qu'en 1818, le capitaine John Ross, lors d'une mission d'exploration dans les régions arctiques, remonta le premier organisme abyssal découvert par l'Homme. C'est en effectuant un sondage bathymétrique dans la baie de Baffin, avec la même méthode que Magellan, mais avec un cordage plus long, qu'il découvrit un organisme fixé sur le cordage, dont les caractéristiques morphologiques ne permettaient manifestement pas la nage. Il en déduisit que cet organisme, aujourd'hui identifié comme appartenant au groupe des gorgonocéphales était présent sur le fond, dont la profondeur a été estimée à 2 000 m ce jour-là (Ross, 1819, Anderson & Rice, 2006). Cette découverte ne fut cependant dévoilée que plusieurs décennies plus tard, ne permettant donc pas de remettre en cause l'hypothèse de Forbes. C'est en 1861 que les premiers organismes abyssaux furent présentés à l'Académie des Sciences et rendus publics via les travaux d'Alphonse Milne-Edwards, qui étudia un câble téléphonique endommagé tendu entre la Sardaigne et l'Algérie, immergé entre 2 000 m et 2 800 m de fond.

Bien que la vie dans les profondeurs ait été démontrée, il a ensuite été pensé que bien que présente, elle y était rare, notamment à cause de l'absence de lumière nécessaire à la photosynthèse, des faibles températures (environ 2°C), des fortes pressions (200 atmosphères à 2 000 m de profondeur), et de la faible concentration en nutriments (Riley *et al.*, 1965, Prieur, 1998). Il s'avère que cette idée reçue est vraie pour la majorité des zones abyssales en ce qui concerne la macrofaune. Il existe cependant des environnements profonds qui foisonnent de vie. Nous ne développerons dans ce manuscrit que les systèmes hydrothermaux, bien qu'il existe aussi les sources de suintements froids, les carcasses de baleines, les bois coulés (pour ne citer que ceux-là). En 1976 a eu lieu une succession de plongées dans l'océan Pacifique avec l'Alvin au cours d'une expédition franco-américaine, plus précisément sur la dorsale est-Pacifique (EPR, East Pacific Rise en anglais), afin d'observer *de visu* des systèmes hydrothermaux avec un intérêt tourné vers la géologie. La surprise fut de découvrir non seulement les systèmes hydrothermaux en question, mais aussi un écosystème foisonnant de vie (Corliss *et al.*, 1979).

L'hydrothermalisme océanique profond

Il est difficile de comprendre, voire de conceptualiser un écosystème biologique ou microbiologique colonisant les systèmes hydrothermaux, sans comprendre (au moins dans les grandes lignes), le fonctionnement géologique qui génère et régule ce type d'environnement.

Répartition et fonctionnement

Les systèmes hydrothermaux sont des systèmes géochimiques marins rencontrés tout autour du globe. Ils sont tous la résultante d'une activité géologique et tectonique sous-jacente importante. C'est pour cette raison que la majorité des sites hydrothermaux recensés à ce jour sont marins et présents le long des zones de jonctions des plaques tectoniques (Figure 1).

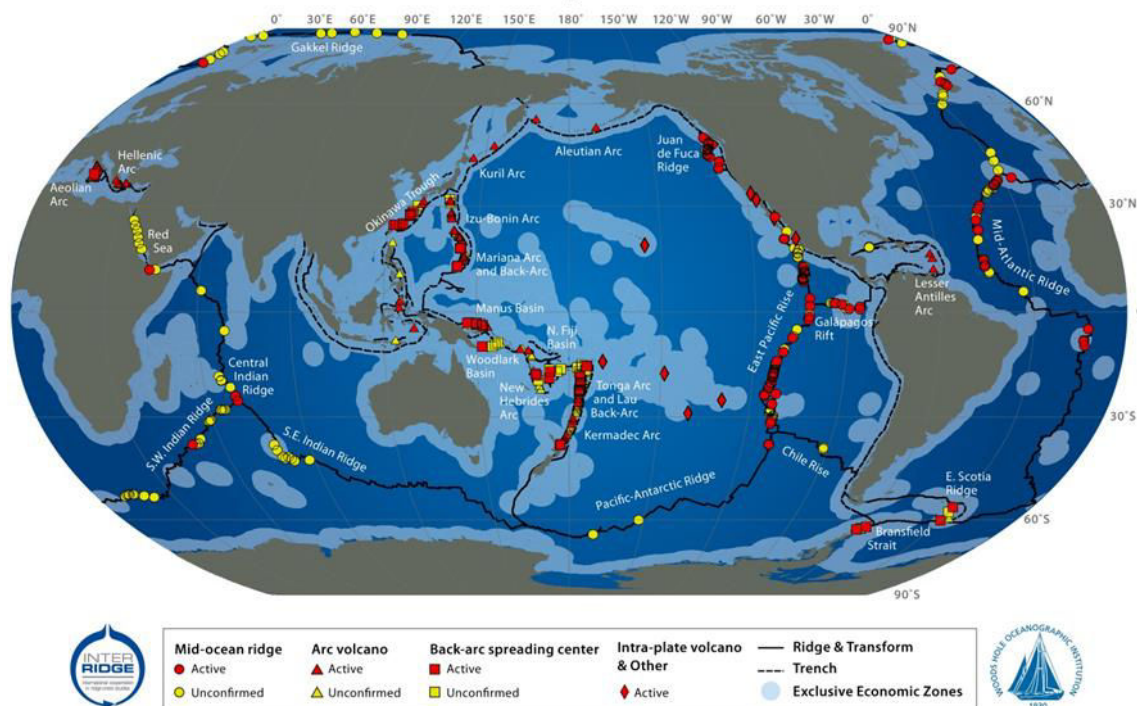


Figure 1 : Répartition géographique des systèmes hydrothermaux océaniques profonds autour du globe. (Source NOAA)

Les autres sites hydrothermaux connus, non marins, ou ne résultant pas de l'activité aux points de jonction de plaques, sont soit des sites hydrothermaux terrestres (aussi appelés site thermaux), ou des sites hydrothermaux issus de points chauds. Un point chaud est une zone de remontée et d'émission locale de magma basaltique, dont l'activité ascendante se traduit en surface de la lithosphère par une activité

volcanique. Pour n'en citer que quelques exemples : l'archipel d'Hawaii, les îles Samoa et Australes, l'île de la Société, ou encore les îles Gambier et Pitcairn (Hékinian & Binard, 2008). Leur activité n'est pas directement liée à la tectonique des plaques, et permet même de l'étudier. En effet, l'activité volcanique générée provient d'un point chaud profond fixe et donc le point d'éruption ne se déplace pas, tandis que la croûte lithosphérique oui (Foucault & Raoult, 2010) (Figure 2), laissant alors des chaînes de volcans retraçant le mouvement de la plaque.

L'autre source d'hydrothermalisme, qui est la principale, est donc la tectonique des plaques, responsable de plus de 75 % du volcanisme planétaire (Van Dover, 2000). On distingue trois types de mouvements, dont deux sont générateurs d'hydrothermalisme : l'accrétion et la subduction (Figure 2).

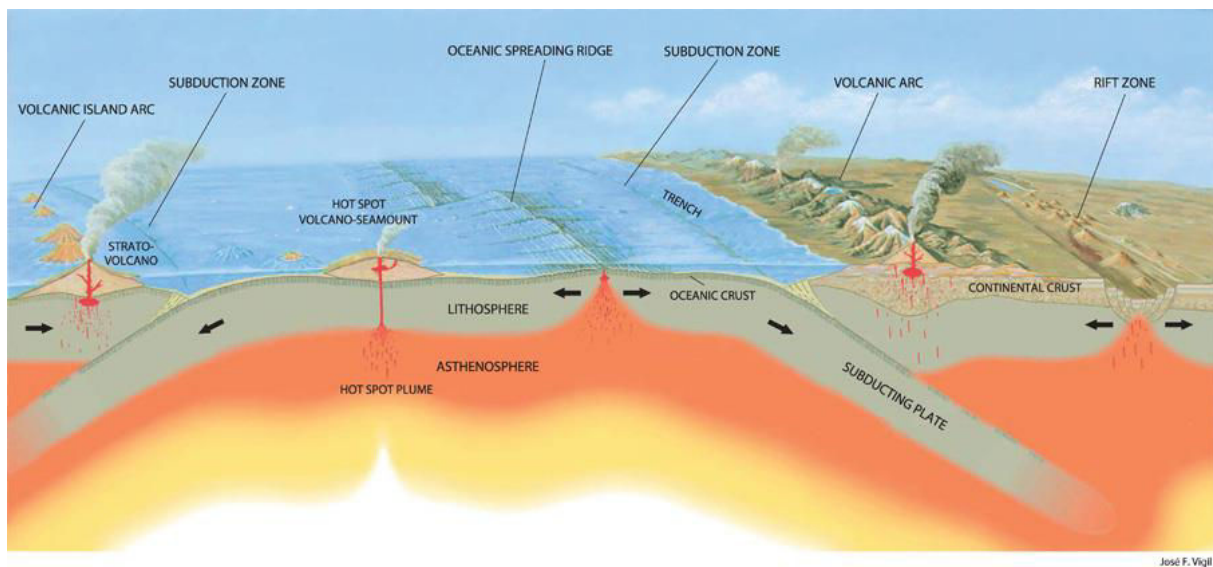


Figure 2 : Représentation schématisée des différents types d'hydrothermalismes (modifiée de l'USGS)

Les zones d'accrétion sont situées au niveau des dorsales océaniques : la dorsale médio-Atlantique (Mid Atlantic Ridge - MAR en anglais), la dorsale est Pacifique, la dorsale Pacifique-Antarctique, ou encore les dorsales sud-est et sud-ouest Indienne. Elles sont toutes marquées par une importante anomalie thermique positive à leur aplomb, ce qui est corrélé avec une remontée de l'asthénosphère au niveau de cette zone (aussi appelé rift). Elles forment ainsi des régions de production de croûte océanique présentant les plus longues et les plus hautes chaînes de montagnes de la planète (Van Dover, 2000). Les nouvelles laves générées (majoritairement basaltiques) au niveau du rift vont s'éloigner géographiquement de leur point d'origine au fur et à mesure que les plaques s'écartent. Ce type de mouvement produit par conséquent une activité sismique régulière (Foucault & Raoult, 2010).

La subduction quant à elle, est un phénomène qui se traduit par la plongée d'une plaque sous une autre. Le plus généralement, c'est la croûte océanique, plus lourde que la croûte continentale, qui plonge sous celle-ci. Ceci se caractérise par la formation de fosses océaniques, ainsi que par de fortes activités sismiques, liées aux frottements des plaques entre elles. Ces frottements sont aussi responsables de volcanisme andésitique situé à la verticale de la plaque plongeante (Foucault & Raoult, 2010), créant ainsi des poches de magma proche de la surface lithosphérique, et donc producteur potentiel d'hydrothermalisme.

Le troisième type de mouvement, qui n'est pas connu pour pouvoir générer d'hydrothermalisme, mais qui présentera un intérêt par la suite, est ce que l'on appelle une faille transformante. Il n'y a ni apport (accrétion), ni absorption (subduction) de matière. Ce sont des failles qui sont parallèles au mouvement des plaques (Fouquet *et al.*, 1994, Foucault & Raoult, 2010), et peuvent présenter des barrières naturelles contre la dispersion biologique (O'Mullan *et al.*, 2001, Guinot & Hurtado, 2003, Guinot *et al.*, 2003, Won *et al.*, 2003, Little *et al.*, 2004, Johnson *et al.*, 2006). Elles peuvent faire plusieurs kilomètres de long pour plusieurs kilomètres de large, et peuvent aller jusqu'à entailler la croûte océanique sur plus de 3000 m d'épaisseur (Hékinian & Binard, 2008). La plus célèbre des failles transformantes est la faille de San Andréas, mais on peut aussi citer la faille transformante de Garrett (Dorsale Pacifique Est), ou encore la faille de Kane (Ride Médio Atlantique ; MAR) (Sykes, 1967, Karson & Dick, 1983, Pockalny *et al.*, 1988).

Si on regarde à une échelle plus locale, la formation d'un système hydrothermal nécessite trois facteurs principaux (Tivey, 2007):

- Une source de chaleur importante, fournie par du magma ou de la lave partiellement solidifiée (voir ci-dessus).
- Un substrat perméable, ou présentant des fissures, comme par exemple un substrat basaltique volcanique régulièrement soumis à une activité tectonique (voir ci-dessus).
- Une source d'un fluide saturant (ou pouvant saturer) le substrat, comme par exemple de l'eau de mer.

On remarque donc que la tectonique des plaques est un moteur puissant dans la génération de systèmes hydrothermaux. Le processus à la base de leur formation et de leur fonctionnement suit les étapes suivantes (on prendra ici l'exemple de systèmes marins) : l'eau de mer s'infiltre dans le plancher océanique fissuré suite à la sismicité locale, au refroidissement du magma en surface, ainsi qu'à l'effondrement du magma sous son propre poids (le magma solidifié étant plus dense que le magma en fusion). Puis se rapprochant d'une source de chaleur (comme une poche de magma), associée à de fortes pressions, l'eau rentre dans un état de surfusion (ou état supercritique) (Norton, 1984, Bischoff & Rosenbauer, 1988), augmentant la solubilité des métaux comme le fer (Fe), le manganèse (Mn), le zinc (Zn) et d'autres composés solides tels que le soufre (S) ou encore le silicate (Si), ainsi que des gaz volatiles provenant du magma comme l'hélium (He), le dioxyde de

carbone (CO₂), le méthane (CH₄) ou le dihydrogène (H₂) (Norton, 1984, Alt, 1995, Tivey, 2007). Tous ces composés sont par la suite lessivés des roches et expulsés sous forme de fluides hydrothermaux formant ensuite un panache, caractéristiques de ces environnements. Ces fluides sont chauds (allant jusqu'à 400 °C), anoxiques, souvent acides (pH 2-3), et chargés en éléments réduits (Charlou *et al.*, 2002). Le contraste important entre le fluide émis (pouvant atteindre jusqu'à 400°C) et l'eau de mer environnant (environ 2-3°C) est tel qu'une partie des composés réduits dissous présents au sein du fluide précipitent, notamment les minéraux, formant les cheminées hydrothermales, ainsi qu'un "nuage" (le panache) pouvant se disperser sur de grandes distances. La composition des fluides varie d'un site à l'autre, mais aussi dans l'espace et le temps (Johnson *et al.*, 1986).

L'étude de ces fluides est très intéressante car elle renseigne sur la composition des roches, et des systèmes géologiques sous-jacents d'un édifice hydrothermal donné. Ces fluides sont source de carbone et d'énergie pour la vie à base de chimiosynthèse (concept sur lequel on reviendra par la suite). Elle peut aussi permettre d'estimer l'activité d'expansion locale de la dorsale (Baker *et al.* 1995; 2004).

Les systèmes hydrothermaux de l'Atlantique

Historiquement, la dorsale médio Atlantique (MAR) a été la première dorsale explorée via le projet FAMOUS (Stakes *et al.*, 1984). Durant ce projet, de nombreuses découvertes ont été réalisées, mais l'observation de systèmes hydrothermaux *de visu* n'a pas été effectuée. Il faudra attendre les années 1985 pour que le site TAG soit découvert au niveau de la MAR à 3670 m de profondeur (Thompson *et al.*, 1988). De nos jours, l'exploration de la MAR se limite principalement à deux zones. La première se situe au niveau du plateau des Açores (38°N). La deuxième se situe plus au sud, il s'agit de la zone « Vema fracture » (11 °N). Quelques zones ponctuelles situées en périphérie de l'Islande ont aussi été explorées (Fricke *et al.*, 1989).

Sur ces deux zones (depuis 1985), une quinzaine de sites hydrothermaux actifs ont été répertoriés. Chacun d'entre eux présente des caractéristiques environnementales qui leur sont propres, les deux principaux facteurs régulant ces caractéristiques étant la profondeur et la nature des roches. Les sites présents à des profondeurs importantes (>3000 m) sont plus stables dans le temps, une activité de 26 000 ans ayant été rapportée sur le site TAG actif actuel (Rona *et al.*, 1993, Lalou *et al.*, 1995, Tivey *et al.*, 1995). A contrario, les zones situées à de plus faibles profondeurs (< 3 000 m), telles que Rainbow et Menez Gwen, sont plus instables dans le temps et l'espace (Desbruyères *et al.*, 2000). Le deuxième facteur impactant ces environnements, c'est-à-dire la nature des roches, permet de classer les systèmes hydrothermaux en deux catégories : les systèmes ultramafiques et les systèmes basaltiques.

Systèmes ultramafiques *versus* systèmes basaltiques

De la même manière que l'on trouve différents types de forêts sur le continent (forêt boréale, humide, méditerranéenne, de montagne...), il existe plusieurs types de systèmes hydrothermaux. Nous n'en présenterons ici que deux (ceux présents sur la MAR) : les sites dits ultramafiques et les sites dits mafiques. Les sites ultramafiques résultent de l'interaction entre l'eau de mer et une remontée du magma profond d'origine mantellique, tandis que les sites basaltiques sont basés sur des remontées de magma basaltique basalte (Van Dover, 2000, Kelley *et al.*, 2002, Fouquet *et al.*, 2013). Cette nuance peut paraître subtile pour une étude de biologie, mais il n'en est rien. Car cette différence a un impact important sur la chimie des fluides émis, notamment sur les fortes concentrations en méthane, fer et hydrogène dans le cas des systèmes mantelliques/ultramafiques (Kelley *et al.*, 2001, Charlou *et al.*, 2002, Schmidt *et al.*, 2007), chimie qui impacte directement le développement microbien (Flores *et al.*, 2011), donc la macrofaune.

La plupart des sites hydrothermaux de la MAR reposent sur un substrat basaltique, à l'exception de trois : Rainbow (Fouquet *et al.*, 1997) ; Logatchev (Bogdanov *et al.*, 1995) et Ashadzé (Fouquet *et al.*, 2008) qui reposent sur un substrat ultramafique. La composition des fluides de ces trois sites présente de forts taux d'hydrogène, de méthane et de fer, mais sont relativement appauvris en sulfures par rapport aux sites basaltiques (Charlou *et al.*, 2002, Douville *et al.*, 2002) (Figure 3).

Mineral and gas endmembers in Rainbow fluids compared to other fluids from the Mid-Atlantic Ridge								
	sw	Logatchev (14°45'N)	MARK-1/2 (23°N)	TAG (26°N)	Broken spur (29°N)	Rainbow (36°14'N)	Lucky strike (37°17'N)	Menez Gwen (37°50'N)
Depth (m)		3000	3460	3670	3200/3300	2300	1700	850
Temperature (°C)	2	347/352	335/350	321/290	356–364	365	170/364	275/284
pH	7.8	3.3	3.9/3.7	3.1	–	2.8	3.5/3.7	4.2/4.3
Si(OH) ₄ (mM)	<0.2	8.2	18.2/18.3	22.0	–	6.9	11.5/16.3	7.7/11.6
Cl (mM)	546	515	559/559	659	469	750	422/533	357/381
Br (μM)	838	818	847	880/1045	749/765	1178	735/924	666/710
SO ₄ (mM)	28.2	0	0	0	0	0	0	0
Na (mM)	464	430	510/509	584	419/422	553	347/446	312/319
Li (μM)	26	245	843/849	411	1006/1033	340	278/357	238/274
K (mM)	9.8	21.9	23.6/23.9	18.0	18.1/19.6	20.4	21.1/26.7	22.1/23.8
Rb (μM)	1.3	27.7	10.5/10.8	10	13.0/13.6	36.9	22.7/39.1	20.3/29.4
Cs (nM)	2.3	385	177/181	110	139/146	333	200/280	330
Mg (mM)	53	0	0	0	0	0	0	0
Ca (mM)	10.2	27.3	9.9/10.5	26.0	11.8/12.8	66.6	31.3/38.2	29.7/33.1
Sr (μM)	87	138	50/51	99	42.9/48.0	200	67/119	100/111
Ba (μM)	0.14	>4.5	>4.3	>19	>12.9/>21.3	>67	10/52	>12
Fe (μM)	<0.001	2500	2180/1832	1640	1684/2156	24050	30/863	<2/18
Mn (μM)	<0.001	330	491/493	1000	250/260	2250	84/446	59/71
Cu (μM)	0.007	15/50	17/10	150	28.3/69.6	121/162	4/26	0.6/3
Zn (μM)	0.01	25/30	50/47	46	40.8/88.0	115/185	5/57	2.4/4.3
<i>Gases</i>								
H ₂ S (mM)	0	0.5/0.8	5.9/5.9	6.7	8.5/11.0	1.2	2.5/3.0	<1.5
CO ₂ (mM)	2.30	10.1	5.2/6.7	2.9/3.4	6.0/7.1	16	13/28	17/20
δ ¹³ C(CO ₂) (‰ PDB)	– 5.1/ – 5.9	– 4.3	–	– 8.4/ – 10.0	– 9.0	– 3.15	– 7.2/ – 10.6	– 6.8/ – 9.1
CH ₄ (mM)	0.0003	2.1	0.023/0.062	0.124/0.147	0.065/0.13	2.5	0.50/0.97	1.35/2.63
δ ¹³ C(CH ₄) (‰ PDB)	–	– 13.6	–	– 8.0/ – 9.5	– 18/ – 19	– 15.8	– 12.7/ – 13.7	– 18.8/ – 19.6
CO (nM)	0.3	–	–	–	–	5000	–	–
Ar (μM)	16	12	17/35	20/40	–	–	11/30	11/38
N ₂ (mM)	0.59	3.0	1.20/3.37	0.90/0.89	–	1.8	0.61/0.97	0.60/1.90
H ₂ (mM)	0.0004	12.0	0.19/0.48	0.15/0.37	0.43/1.03	16	0.02/0.73	0.024/0.048

Figure 3 : Composition minérale et gazeuse de fluides hydrothermaux de la MAR. Le site MARK-1/2 est aujourd'hui appelé Snake Pit (Charlou *et al.*, 2002).

Diversité des types d'émissions de fluide

Au-delà des aspects géologiques conditionnant notamment la chimie des fluides (et par conséquent l'environnement chimique du site en question), certains aspects physiques liés à l'architecture des édifices hydrothermaux sont à prendre en compte pour l'étude biologique de ces environnements. Pour cela, nous allons décrire succinctement trois types d'émission de fluides (qui peuvent d'ailleurs potentiellement tous être retrouvés sur un même édifice) : les fumeurs (smoker), les diffuseurs (diffuser ou beehive) et les rebords (flange).

Les fumeurs (Figure 4a) sont les structures les plus caractéristiques des systèmes hydrothermaux. Comme expliqué précédemment, c'est le mélange entre le fluide chaud, chargé en éléments réduits dissous, et l'eau de mer froide oxygénée qui entraîne la précipitation de ces éléments et crée un panache hydrothermal chargé de particules (Tivey, 2007). Ce sont ces précipités qui forment, à une échelle très locale, les édifices. Les fumeurs noirs sont caractérisés par un conduit central peu perméable, ce qui explique notamment la faible dilution des fluides expulsés, leur température élevée (pouvant atteindre 450°C) ainsi que leur couleur noire (due aux fortes concentrations en soufre et en fer) (Hannington *et al.*, 1995, Van Dover, 2000). Il est

aussi intéressant de noter qu'il existe des fumeurs blancs, tirant leur nom de la couleur du fluide émis, plus dilué, moins chaud et de composition chimique différente (Tivey *et al.*, 1995).

Les diffuseurs (ou ruches) (Figure 4b) sont des structures produisant des excroissances sur le côté ou le dessus des cheminées. Elles sont très poreuses et friables, et diffusent des fluides chauds pouvant aller jusqu'à 100°C. Le débit de ces structures est cependant moindre que celui des fumeurs (Fouquet *et al.*, 1993, Hannington *et al.*, 1995, Van Dover, 2000). Les conditions physico-chimiques à l'échelle de ce type de structures permettent la précipitation des sulfures sous forme de couches, entraînant ainsi une surface extérieure concentrique nervurée faisant penser à des ruches sauvages (Hannington *et al.*, 1995, Van Dover, 2000). A cause des fortes températures à la surface de ces structures, elles sont le plus souvent à nu, sans colonisation microbienne ou animale d'aucune sorte (Haymon, 1983, Fouquet *et al.*, 1993, Van Dover, 2000).

Les rebords (ou flanges en anglais) (Figure 4c) sont des structures latérales pouvant atteindre plus d'un mètre de largeur et plusieurs centimètres d'épaisseur. La surface inférieure présente une forme creuse, servant de réservoir pour le fluide provenant de la base de la structure. La température de sortie du fluide à la base de cette structure est souvent corrélée à celle des sorties de fluide du fumeur auquel elle est reliée. En effet, ces structures sont « greffées » sur les conduits principaux des fumeurs. Il en résulte donc un gradient de température important dans le sens horizontal, ainsi qu'un gradient de température encore plus important dans le sens vertical, notamment au niveau de l'interface fluide - eau de mer. Cependant, le fluide sortant étant moins dense que l'eau de mer, il se retrouve emprisonné sous le réservoir, pour un court laps de temps de l'ordre de 100 secondes (édifice Endeavour sur la ride Juan de Fuca) (Kerr, 1997, Van Dover, 2000). A la surface supérieure de la structure, les conditions sont différentes. Le fluide ayant percolé à travers la matrice présente une température allant de 10 à 80°C, et est plus dilué que le fluide initial, créant ainsi une niche écologique plus propice au développement de la vie microbienne et animale (Lonsdale & Becker, 1985, Lisitsyn *et al.*, 1989, Delaney *et al.*, 1992, Kerr, 1997).

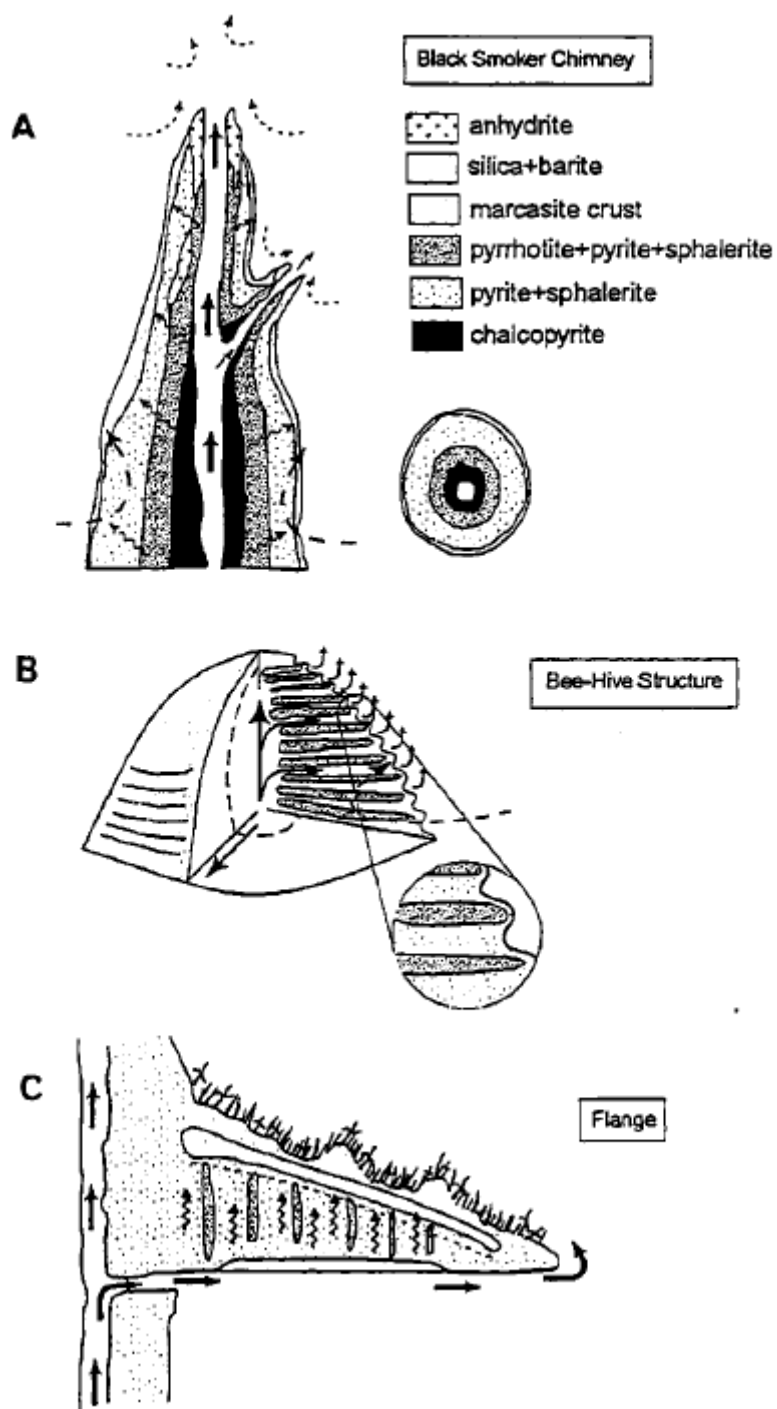


Figure 4 : Diagramme schématique (A) d'un fumeur, (B) d'une ruche, (C) d'un rebord. (Hannington et al., 1995).

Les microorganismes, base du réseau trophique

Où trouver des microorganismes en contexte hydrothermal

Comme expliqué précédemment, les environnements hydrothermaux sont des écosystèmes dynamiques résultant du mélange entre les fluides hydrothermaux et l'eau de mer. De forts gradients physico-chimiques spatio-temporels sont présents. Les microorganismes occupent différentes niches écologiques (Karl, 1995, Van Dover, 1995) qu'il est intéressant de connaître, autant pour pouvoir élaborer des stratégies d'échantillonnages correctes que pour mesurer la complexité phylogénétique, métabolique, biochimique et physico-chimique qui peut résulter de ces microenvironnements :

- Les surfaces des dépôts minéraux, les sédiments ou encore les animaux exposés aux fluides : en fait, toute surface permettant une fixation des microorganismes que ce soit sous forme de biofilm ou de tapis microbien. Il est à noter que les animaux comme support de fixation ne relèvent pas obligatoirement de la symbiose (bien que les frontières puissent être minces).
- Le panache hydrothermal (le fluide hydrothermal dilué, que l'on trouve à plusieurs mètres, centaines de mètres, voire dizaines de kilomètres de la source d'émission). Des formes à la fois libres ou fixées (à des particules minérales par exemple) de microorganismes y sont présentes. Leur pérennité, à savoir leur capacité à se maintenir dans ce type d'environnement reste encore aujourd'hui un débat ouvert.
- La faune hydrothermale. On entend dans ce cas précis la notion d'association symbiotique entre un métazoaire et des bactéries, notion qui sera plus largement développée ultérieurement.
- Les édifices hydrothermaux actifs. En effet, le caractère poreux et stratifié des édifices, associé à un gradient à la fois thermique et chimique, permet le développement de microorganismes aussi bien hyperthermophiles que mésophiles à l'intérieur même de l'édifice.
- Les sédiments hydrothermaux. Bien que l'expulsion des fluides soit très localisée sur les édifices, une fraction moins importante de fluide percole au travers des sédiments environnants, ce qui est source d'énergie pour les microorganismes en présence.

A noter, on ne retrouve pas l'ensemble de ces habitats de manière systématique sur chaque site hydrothermal.

La chimiosynthèse

L'absence de lumière, la pression hydrostatique, les forts gradients physico-chimiques font des systèmes hydrothermaux un milieu qui pour l'homme paraît ou est inhospitalier, et est souvent qualifié d'extrême. La

vie animale dans ce type d'environnements peut paraître incongrue, ce qui n'est pas le cas. Cette vie peut se catégoriser en deux groupes bien distincts : microbienne et animale. Nous traiterons dans cette partie de la vie microbienne, puis de la vie animale.

Pourquoi commencer par la vie microbienne ? Parce qu'elle représente la base trophique de la vie de ces environnements. En effet, l'absence de lumière rend la synthèse de matière organique via la photosynthèse inexistante au-delà de 500 m (Riley *et al.*, 1965), ce qui ne permet a priori pas une vie animale riche en biomasse. En effet, dans les grandes profondeurs, seule la sédimentation de matière organique synthétisée en surface permet un apport de matière organique régulier (en plus de carcasses d'animaux ou de bois...). De plus, les fortes pressions hydrostatiques limitent le développement de la vie à une diversité d'organismes présentant des adaptations à la pression. Des communautés microbiennes chimiosynthétiques, capables d'utiliser les composés réduits des fluides comme source d'énergie et de carbone, se sont développées au niveau des systèmes hydrothermaux. Ces environnements, très riches en composés réduits, sont de vraies oasis de vie pour ces microorganismes chimiosynthétiques. Un certain nombre de métabolismes reste cependant dominant, tels ceux impliqués dans le cycle du carbone, du soufre, du fer et de l'hydrogène.

Le cycle du carbone

La synthèse de carbone organique de *novo* (autotrophie) peut se faire grâce à deux sources d'énergie : la lumière (c'est la photosynthèse), et l'énergie libérée par des réactions chimiques (c'est la chimiosynthèse) (Madigan *et al.*, 2000). Plusieurs voies sont utilisées par les organismes pour fixer le carbone inorganique (CO_2 ou CH_4). Pour le CO_2 ce sont les cycles de Calvin-Benson-Bassham (CBB) et le cycle des acides tricarboxyliques inverses (rTCA) qui sont utilisés. Au cours de ces cycles, le carbone inorganique est fixé sous forme de molécules organiques permettant la croissance des microorganismes autotrophes. Certaines molécules organiques pourront alors être utilisées par les organismes hétérotrophes qui peuvent aussi ensuite consommer les organismes autotrophes.

Le cycle de Calvin-Benson-Bassham

Le mécanisme biochimique impliqué dans la fixation du CO_2 le plus connu est le cycle de Calvin-Benson-Bassham. Il est alimenté par du NAD(P)H et de l'ATP et comprend deux enzymes clés : la ribulose-1,5-biphosphate carboxylase-oxygénase, aussi appelée RuBisCO, et la phosphorybulokinase. La RuBisCO catalyse la première réaction de réduction du CO_2 et est largement répandue dans le monde du vivant. Elle est présente chez les cyanobactéries, les bactéries pourpres, les algues, les plantes vertes, la plupart des bactéries chimiolithotrophes, ainsi que chez certaines archae hyperthermophiles. Il existe en tout 4 formes de RuBisCO : les formes I, II, III et IV (Tabita *et al.*, 2008). Les formes I et II sont fréquemment retrouvées en milieu hydrothermal et diffèrent par leur structure tridimensionnelle et leur signature isotopique (Scott *et al.*, 2004, Scott *et al.*, 2007, Tabita *et al.*, 2008). La RuBisCO de type I est adaptée aux conditions oxygènes

tandis que la RuBisCO de type II est adaptée aux conditions anoxiques (Robinson & Cavanaugh, 1995, Haygood, 1996, Elsaied & Naganuma, 2001, Elsaied *et al.*, 2006). La capacité des bactéries à l'autotrophie est habituellement détectée en recherchant les gènes codant pour les deux formes de RuBisCo : les gènes *cbbI* et *cbbM*, qui codent respectivement pour les formes I et II de la RuBisCO (Shigematsu *et al.*, 1999, Shively *et al.*, 1999).

Le cycle inverse de l'acide citrique (rTCA pour reverse TriCarboxylic Acid)

Le cycle rTCA est une voie métabolique qui a été reconnue en 1990 (Buchanan & Arnon, 1990), bien qu'elle ait été proposée dès 1966 sur un modèle de bactérie verte sulfureuse *Chlorobium thiosulfatophilum* (Evans *et al.*, 1966). Comme son nom l'indique, le rTCA est essentiellement une inversion du cycle TCA (aussi connu sous le nom de cycle de Krebs) et permet la synthèse d'acétyl-CoA à partir de deux molécules de CO₂. La plupart des enzymes du cycle TCA sont réversibles, à l'exception de quatre, qui sont remplacées dans le rTCA par la fumarate réductase, la 2 – oxoglutarate synthase, et les enzymes de clivage du citrate (la pyruvate synthase et l'ATP citrate lyase). Une de ces enzymes réduit le champ d'utilisation de cette voie métabolique. En effet, la 2-oxoglutarate synthase étant sensible à l'oxygène, le cycle rTCA ne semble effectif que dans des conditions anaérobies, ou au moins microaérophiles (Campbell *et al.*, 2006). Bien que les exemples d'utilisation du cycle rTCA de façon exclusive chez les *Aquificales* et les *Epsilonproteobacteria* ne manquent pas (Shiba *et al.*, 1985, Beh *et al.*, 1993, Deckert *et al.*, 1998, Hügler *et al.*, 2005, Campbell *et al.*, 2006, Nakagawa *et al.*, 2007, Sievert *et al.*, 2008, Reysenbach *et al.*, 2009), l'exemple de l'endosymbionte « *Candidatus endoriftia persephone* » du ver *Riftia pachyptila* est très intéressant car une étude de protéomique a permis de mettre en évidence l'utilisation à la fois des cycles CBB et rTCA, ce qui en fait aujourd'hui le seul représentant microbien à posséder ces deux voies de fixation du carbone (Markert *et al.*, 2007).

La méthanotrophie

Dans les systèmes hydrothermaux, le méthane, qu'il soit d'origine biotique ou abiotique, est oxydé par des populations de microorganismes méthanotrophes en un aldehyde, puis en matière organique lorsqu'il atteint les zones oxygénées (Anthony, 1982). Dans la méthanotrophie, on ne dénombre à l'heure actuelle qu'une seule enzyme clé : la méthane monooxygénase. Cependant, elle existe sous deux formes : une forme soluble (sMMo : soluble Methane Monooxygenase), et une forme particulaire associée à la membrane (pMMo : particulate Methane Monooxygenase). Leur abondance dans le monde microbien n'est pas homogène. La forme particulaire est retrouvée chez tous les méthanotrophes à l'exception du genre *Methylocella* (Dedysh *et al.*, 2000), la forme soluble quant à elle est moins représentée, et n'est signalée que chez un nombre restreint de souches (Fuse *et al.*, 1998, Shigematsu *et al.*, 1999). L'utilisation du gène de la

forme particulière (*pmoA*) est un très bon marqueur génétique, et aussi phylogénétique, sa phylogénie étant convergente avec celle du gène codant pour l'ARNr 16S dans certains cas (Duperron *et al.*, 2005).

Des populations de méthanotrophes existent au sein de la classe des *Alphaproteobacteria* et des *Gammaproteobacteria*. On les retrouve notamment dans des tapis microbiens (Crépeau *et al.*, 2011) et il en existe aussi des exemples dans des symbioses entre métazoaires et méthanotrophes, notamment au niveau des systèmes hydrothermaux. C'est le cas pour les bathymodiales, notamment celles du genre *Bathymodiolus* (Duperron *et al.*, 2005, Duperron *et al.*, 2006), ainsi que chez la crevette *Rimicaris exoculata* (Zbinden *et al.*, 2008, Guri *et al.*, 2012).

Le cycle du fer

Chez les procaryotes, le fer joue un rôle clé dans certains métabolismes aérobies et anaérobies. Il est essentiel pour le fonctionnement d'un certain nombre d'enzymes comme les catalases, les peroxydases, les oxygénases, mais aussi pour les cytochromes et les protéines fer-soufre (ferrodoxines).

La ferro-oxydation

L'oxydation abiotique du fer étant très rapide dans des conditions acides et oxydantes, on a longtemps pensé que les microorganismes des systèmes hydrothermaux ne pouvaient pas réaliser la ferro-oxydation (oxydation du fer ferreux : Fe II), leur source d'énergie s'oxydant trop rapidement et le rendement énergétique étant trop faible. Cependant, un certain nombre de bactéries tel que le genre *Acidithiobacillus* (connu pour se développer dans les eaux de drainage des minerais et les stations d'épurations) (Yarzabal *et al.*, 2002, Yarzabal *et al.*, 2004, Quatrini *et al.*, 2007, Valdés *et al.*, 2007, Valdés *et al.*, 2009), ainsi que les genres *Gallionella*, *Leptothrix* et *Sphaerotilus* appartenant à la classe des *Betaproteobacteria* et *Gammaproteobacteria* sont déjà identifiés comme capables de réaliser la ferro-oxydation en condition acide (Nealson, 1997, Emerson & Moyer, 2002, Emerson *et al.*, 2007). Ils sont aussi capables de catalyser la précipitation d'oxydes de fer (Nealson, 1997, Emerson & Moyer, 2002, Emerson *et al.*, 2007). De plus, la souche *Mariprofundus ferroxydans* isolée en 2007 d'un mont sous-marin près d'Hawaï (Emerson *et al.*, 2007), présente la caractéristique d'être ferro-oxydante stricte à pH neutre. C'est la première souche ayant conduit à la création de la récente classe des *Zetaproteobacteria* (Emerson *et al.*, 2007). Cette classe sera ré-abordée plus loin lors de la description de la symbiose chez la crevette *Rimicaris exoculata*.

La ferro-réduction

La ferro-réduction (réduction du fer ferrique : Fe III) est une autre composante du cycle du fer. Elle est réalisée en conditions anoxiques par de nombreux microorganismes organotrophes et lithotrophes. Compte tenu de son potentiel redox, le Fe III est un accepteur d'électrons tout à fait favorable énergétiquement dans un milieu dépourvu d'oxygène (Lovley & Phillips, 1986). Parmi les souches isolées, on retrouve notamment des

bactéries de la classe des *Thermotogales* ou de l'ordre des *Deferribacteres*, qui peuvent être thermophiles, psychrophiles, acidophiles ou alcalinophiles, ainsi que des archae de la sous classe des *Achaeoglobales* et des *Thermococcales* (Jannasch & Mottl, 1985, Lovley *et al.*, 2004). Chez de nombreuses hyperthermophiles cultivées, on retrouve cette capacité à réduire le Fe III, ce qui suggère que ce type métabolique est fortement conservé chez les microorganismes hyperthermophiles (Kashefi *et al.*, 2002, Kashefi & Lovley, 2003).

Le cycle du soufre

Dans un contexte hydrothermal, les composés soufrés sont apportés par le fluide après lessivage des roches. En association avec le fer, le cuivre ou encore le zinc (eux aussi apportés par le fluide), et au contact de l'eau de mer, ces éléments précipitent en sulfures métalliques, et sont en grande partie à l'origine de édifices hydrothermaux (Tivey, 2007). Il existe trois types de métabolismes associés au soufre : la sulfato-réduction, la sulfo-réduction, et la sulfo-oxydation.

La sulfato-réduction

Le sulfate (SO_4^{2-}) étant la forme la plus oxydée du soufre, il est utilisé comme accepteur final d'électrons pour les microorganismes sulfato-réducteurs. Cependant, plusieurs types de donneurs d'électrons sont possiblement rencontrés chez ces organismes, comme l'hydrogène, le lactate, ou encore le pyruvate (Rabus *et al.*, 2006). Le produit final de réaction est le sulfure d'hydrogène (H_2S), qui est rejeté tel quel par la cellule, ou bien utilisé pour former des acides aminés (Rabus *et al.*, 2006). Le fait que le sulfate soit présent en quantité non négligeable dans l'eau de mer (à hauteur de 8 %) permet à une forte diversité de microorganismes d'utiliser cette molécule. On retrouve notamment des bactéries majoritairement affiliées à la classe des *Deltaproteobacteria*, comme avec les genres *Desulfovibrio*, *Desulfurimonas*, *Desulfobacter*..., la classe des Firmicutes avec comme genre principal les *Desulfotomaculum* (Pikuta *et al.*, 2000), ou encore les *Thermodesulfatator* avec les représentants du genre *Thermodesulfobacterium* et *Thermodesulfatator* (Alain *et al.*, 2010). On retrouve aussi des espèces d'archae du genre *Archaeoglobus*, qui ont d'ailleurs été isolées de sources hydrothermales.

La sulfo-réduction

Le soufre élémentaire (S^0) est un composé stable et insoluble. D'un point de vue écologique, la réduction du soufre en sulfure (forme de respiration anaérobie) chez les archae est un processus majeur, notamment au sein du genre *Desulfurococcus*. Dans les systèmes hydrothermaux, un grand nombre d'espèces cultivées sont capables de réduire le soufre élémentaire et présentent des métabolismes divers. En effet, un certain nombre de chimioorganotrophes anaérobies fermentaires telles que les *Thermococcales* et les *Thermotogales* sont capables de réduire le S^0 . Cependant, cette réaction n'est pas indispensable à leur croissance mais la stimule. On trouve aussi un certain nombre de lignées pour lesquelles le soufre élémentaire est l'accepteur final

d'électron. Ces microorganismes thermophiles et hyperthermophiles des genres *Desulfurobacterium* (L'Haridon et al., 1998, Alain et al., 2003, L'Haridon et al., 2006), *Balnearium* (Takai et al., 2003) et *Thermovibrio* (Vetriani et al., 2004), qui forment une lignée distincte au sein des *Aquificales*, ainsi que des *Crenarchaeota*, *Ignicoccus pacificus*, (Huber et al., 2001) sont des chimiolithoautotrophes.

La sulfo-oxydation

L'environnement hydrothermal est un milieu riche en éléments soufrés plus ou moins réduits comme le S^{2-} (ou H_2S , sulfure d'hydrogène), le S^0 (soufre élémentaire), le $S_2O_3^{2-}$ (thiosulfate), ou encore le $S_4O_6^{2-}$ (tétrathionate). Les microorganismes dits sulfo-oxydants présentent la capacité d'utiliser ces composés comme donneurs d'électrons en catalysant leur oxydation. L'accepteur d'électrons est quant à lui généralement l' O_2 . Cependant certains microorganismes utilisent le NO_3^- (nitrate) lorsque l' O_2 est présent en trop faibles concentrations. Cette situation se retrouve préférentiellement au niveau des zones de transitions entre un milieu oxique et anoxique, comme la zone de mélange entre l'eau de mer et le fluide hydrothermal. Du point de vue écologique, les microorganismes sulfo-oxydants jouent un rôle majeur au niveau des systèmes hydrothermaux : l'oxydation des composés soufrés réduits étant une des réactions dominantes dans la production d'énergie nécessaire à la fixation du CO_2 . Une partie non négligeable des bactéries sulfo-oxydantes vit en symbiose avec des représentants de la faune. Nous pouvons citer le vestimentifère *Riftia pachyptila* (Giovannelli et al., 2016), les bivalves du genre *Bathymodiolus* (Duperron et al., 2007), la crevette *Rimicaris exoculata* (Zbinden et al., 2008), ou encore la galathée *Shinkaia crosnieri* (Watsuji et al., 2010). On parle dans ce cas de symbiontes « thiotrophes ». D'autres bactéries sulfo-oxydantes de types filamenteuses se développent sous forme d'épais tapis microbiens dans les zones oxique/anoxique.

Il existe deux voies métaboliques d'oxydation d'éléments réduits contenant du soufre. La voie de l'adénosine phosphosulfate (APS) réductase et la voie de la sulfite oxydase. Bien que la voie de la sulfite oxydase soit souvent majoritaire, la voie de l'APS réductase est utilisée chez certains microorganismes bien connus comme par exemple les *Gammaproteobacteria* symbiotiques des bathymodioles du genre *Idas* ou *Bathymodiolus* (Duperron et al., 2005, Duperron et al., 2006, Meyer & Kuever, 2007, Meyer & Kuever, 2007, Meyer & Kuever, 2007, Duperron et al., 2008). Il est intéressant de noter que l'APS réductase est une enzyme utilisée à la fois durant la sulfo-oxydation et la sulfo-réduction. Elle fait donc partie des marqueurs génétiques du cycle du soufre (Meyer & Kuever, 2007).

Le cycle de l'hydrogène

Le dihydrogène est un donneur d'électron connu, à très haut potentiel énergétique, mais encore peu documenté dans les environnements à base chimiosynthétique, et son rôle est potentiellement sous-estimé (Jannasch & Mottl, 1985). Il est nécessaire aux enzymes de type des monooxygénases (méthane, ammonium) ou de la méthanogénèse (Nealson, 1997). Le dihydrogène est aussi considéré comme un composé

permettant la survie à long terme des formes de résistances bactériennes dans des environnements extrêmes (Morita, 1999). Depuis peu, il est considéré comme source métabolique énergétiquement aussi favorable que le fer en contexte mantellique (Schmidt *et al.*, 2008, Hügler *et al.*, 2011, Petersen *et al.*, 2011). L'enzyme clé de ce métabolisme est une hydrogénase, dont le marqueur génétique est le gène *hynL* (Hügler *et al.*, 2011). Bien qu'il puisse être produit de manière biogénique, en contexte hydrothermal le dihydrogène est majoritairement apporté par le fluide, notamment sur les sites ultramafiques comme le site Rainbow (Charlou *et al.*, 2002).

La macrofaune, représentant visible des systèmes hydrothermaux

Lorsque l'on s'approche de systèmes hydrothermaux actifs, l'abondance de vie est le premier élément marquant comparé aux plaines abyssales. Le succès évolutif de ces organismes est indéniable, si on considère leur biomasse, pouvant aller jusqu'à 50 kg par m² pour la crevette *Rimicaris exoculata* (Segonzac *et al.*, 1993) ou encore 70 kg par m² pour la modiole *Bathymodiolus puteoserpentis* (Gebruk *et al.*, 2000). Ces écosystèmes, qualifiés d'oasis de vie, en possèdent toutes les caractéristiques : abondance très localisée d'une ressource, présence d'organismes spécialisés, souvent endémiques (97% appartiennent à de nouveaux taxons (du genre jusqu'à l'ordre)) et persistance de cette faune (Carney, 1994). Si les biomasses et les densités y sont extrêmement élevées, la diversité spécifique en revanche y est beaucoup plus faible que dans l'océan profond avoisinant (Grassle & Maciolek, 1992). Les espèces se distribuent en cercles concentriques autour des zones d'émission de fluides en fonction de leur préférendum physico-chimique (Segonzac, 1992, Van Dover, 2000). La distribution de sortie des fluides, la chimie, la température, la dureté du substrat ainsi que la profondeur des sites hydrothermaux jouent également un rôle important dans la répartition des espèces (MacDonald *et al.*, 1989, Sarrazin *et al.*, 1997, Maas *et al.*, 1999). Il est intéressant d'observer que les organismes les plus abondants au niveau des sites hydrothermaux tels que *Rimicaris exoculata*, *Bathymodiolus azoricus*, *Riftia pachyptila*, *Alvinella pompejana*, *Ifremeria nautilei*, *Shinkaia crosnieri*... sont des organismes présentant tous une ou plusieurs symbioses connues à ce jour. L'association symbiotique avec des bactéries semble donc être une adaptation procurant un certain avantage écologique en milieu hydrothermal. Ces organismes symbiotiques sont considérés comme des consommateurs primaires (Desbruyères & Laubier, 1983), de même que les organismes brouteurs ou filtreurs qui se nourrissent de bactéries libres. Un réseau trophique "classique" vient ensuite avec différents niveaux de prédateurs, des nécrophages et autres décomposeurs.

La symbiose, concept important des environnements hydrothermaux

Lorsqu'on entend parler de symbiose, on pense souvent à la définition suivante : association réciproquement profitable de deux organismes vivants (acception française du terme). Or la symbiose est un concept plus complexe. La définition précédente représente la notion de symbiose au sens le plus strict du terme. D'un point de vue plus généraliste, la symbiose peut être vue comme une association entre plusieurs partenaires (étymologiquement symbiose signifie « vivre avec »). Il est ainsi possible d'identifier d'autres types d'interactions entre des partenaires qui pourraient relever de la symbiose. Nous pouvons citer le neutralisme, l'antagonisme, la compétition, l'amensalisme, le commensalisme, le mutualisme ou symbiose au sens strict du terme, le parasitisme et la prédation (Selosse, 2000). Toutes ces notions font appel à un système de cohabitation (positive ou négative) où chaque partenaire présente des caractéristiques physiques, chimiques et physiologiques qui lui sont propres, découlant de son histoire évolutive, et auxquelles correspondent des besoins et des stratégies spécifiques. Vivre ensemble suppose donc de rendre compatibles, pour un temps plus ou moins long, les modes de vie des partenaires, et permettre leur rencontre, leur reconnaissance, leur communication, et si le jeu s'y prête, établir et maintenir une relation. Pour la suite du sujet, nous choisirons de définir le terme de symbiose comme une association de plusieurs partenaires pour lesquels la relation à plusieurs permet un meilleur *fitness* que s'ils étaient séparés. Le concept de *fitness* étant déjà difficile à estimer sur le plan pratique pour de gros animaux terrestres dans un milieu propice à l'observation, nous le simplifierons ici comme étant un état de développement théorique permettant un meilleur succès de transmission de gènes vers les générations futures. Par souci de commodité, lorsque les partenaires sont de tailles distinctes, nous qualifierons le plus grand des partenaires d'hôte, et le plus petit de symbionte (Gaill *et al.*, 1987, Smith & Douglas, 1987, Selosse, 2000).

Plusieurs critères caractérisent une association symbiotique : sa spécificité, sa durée, le caractère obligatoire ou non, la localisation, le mode de transmission ainsi que les échanges entre partenaires. Concernant la localisation, plusieurs cas sont possibles : la présence du symbionte à l'extérieur ou à l'intérieur de l'hôte. Dans le cas d'un symbionte présent sur la surface externe de l'hôte, celui-ci est défini comme étant un ectosymbionte, tandis qu'un symbionte présent à l'intérieur de son hôte est défini comme étant un endosymbionte. Au sein des endosymbiontes, on distingue les symbiontes intra ou extra-cellulaires (situés à l'intérieur ou l'extérieur des cellules de leur hôte). Les exemples les plus étudiés d'ectosymbiose sont l'annélide *Alvinella pompejana* (Gaill *et al.*, 1987, Desbruyères *et al.*, 1998) et la crevette *Rimicaris exoculata* (Segonzac *et al.*, 1993, Polz *et al.*, 1998, Zbinden *et al.*, 2004, Zbinden *et al.*, 2008, Petersen *et al.*, 2010). Les exemples les plus étudiés d'endosymbiose sont l'annélide *Riftia pachyptila* (Felbeck, 1981) et la moule *Bathymodiolus azorius* (Duperron *et al.*, 2006).

La transmission du/des symbionte(s) d'une génération à l'autre est une autre caractéristique importante, qui a été le premier processus identifié pour comprendre l'évolution des associations symbiotiques (Lipsitch *et al.*, 1995). Trois modes de transmission de symbiontes microbiens sont connus : la transmission horizontale, la transmission verticale et la transmission mixte (Bright & Bulgheresi, 2010). La transmission horizontale implique une production de descendants aposymbiotiques (sans symbiontes) et un apport environnemental de symbiontes, qui vivent aussi sous forme libre. L'hôte est exposé au minimum une fois à son symbionte durant son cycle de vie, de l'état larvaire jusqu'au stade adulte. L'état symbiotique de la bactérie est donc facultatif. Ces microorganismes peuvent être retrouvés à proximité de l'habitat de l'hôte, au niveau des zones benthiques (sédiments) (Costa *et al.*, 2001), pélagiques (Lee & Ruby, 1992, Gros *et al.*, 2003, Aida *et al.*, 2008), et, dans certains cas, l'hôte adulte peut libérer une partie de ses symbiontes dans un environnement proche de sa descendance (Kiers *et al.*, 2003, Salerno *et al.*, 2005). Il s'effectue une sélection environnementale. La transmission verticale quant à elle se réalise de parent à descendant(s) sans sélection environnementale. Elle est réalisée principalement via une transmission du (ou des) symbionte(s) à l'œuf, durant les étapes embryonnaires, jusqu'au stade larvaire, voire pendant le développement à l'intérieur ou à l'extérieur du parent (Bright & Bulgheresi, 2010). Certains cas montrent, via des analyses phylogénétiques, une transmission mixte avec une transmission majoritairement verticale, et occasionnellement une transmission horizontale (Shoemaker *et al.*, 2002).

Le dernier critère caractérisant une association symbiotique que nous développerons ici concerne les échanges entre partenaires. En effet, la symbiose au sens où nous l'entendons a pour principale caractéristique la présence de bénéfices réciproques entre les organismes de l'association, qu'ils soient de nature trophique, ou concernent le *fitness* des partenaires. Le premier avantage pour le symbionte est de pouvoir disposer d'une niche écologique dénuée de prédateurs, stable, et de disposer de substrats métaboliques partagés ou inutiles pour l'hôte (par exemple les déchets). Du point de vue de l'hôte, le symbionte peut apporter une certaine protection contre les pathogènes via compétition (nutriment ou espace) (Selosse, 2000). Il peut aussi produire des composés carbonés assimilables par l'hôte comme c'est le cas dans beaucoup de tubes digestifs, et donc fournir un apport énergétique dans certains cas non négligeable (voire le seul) (Selosse, 2000), ou encore présenter un rôle détoxifiant (Zachary & Colwell, 1979, Zachary *et al.*, 1983). Le symbionte peut encore jouer un rôle dans la mobilité de l'hôte (Cleveland & Grimstone, 1964), interférer sur son sexe ratio (Wernegreen *et al.*, 2003), ou servir de source bioluminescente dans le but de distraire un prédateur de l'hôte, de rechercher une proie ou un partenaire sexuel (Herring, 2002). L'efficacité de la symbiose, illustrée via les exemples ci-dessus, en fait un système très répandu, voire obligatoire, dans la nature, tant dans le monde procaryote qu'eucaryote, animal ou végétal. Elle présente la possibilité de pouvoir intervenir à tous les niveaux de la vie de l'hôte (développement larvaire,

nutrition de l'hôte, détoxification, *fitness*, reproduction...), et influencer au final son évolution, tout comme l'hôte influence celle des symbiontes.

Jusqu'à très récemment, les modèles d'étude des symbioses avaient pour beaucoup un intérêt agronomique ou médical. Les symbioses chez les animaux marins, notamment ceux des profondeurs ont été nettement moins étudiées. Or cet environnement abrite de nombreuses associations originales dont l'impact écologique est encore mal appréhendé. Depuis la découverte dans les années 70 d'oasis de vie dans les grands fonds, la communauté scientifique a renouvelé son intérêt pour ces sujets d'étude (Lonsdale, 1977, Paull *et al.*, 1984). L'existence dans ces environnements extrêmes de telles associations pose de nombreuses questions sur leur apparition, leur évolution, leur fonctionnement ainsi que sur leur rôle écologique dans les océans (Fisher, 1990, Van Dover, 2000, Tunnicliffe *et al.*, 2003).

Durant cette thèse, nous nous focalisons sur l'étude de la symbiose chez les crevettes hydrothermales de la MAR, et plus particulièrement sur l'espèce *Rimicaris chacei*. Jusqu'à ce jour le modèle *Rimicaris exoculata* fait office de référence sur ce sujet. C'est pourquoi nous le développons ci-après.

Le modèle *Rimicaris exoculata*

L'animal

Rimicaris exoculata est une crevette Caridé appartenant à la famille des Alvinocarididae et inféodée aux écosystèmes hydrothermaux atlantiques. Elle a été décrite pour la première fois en 1986 sur le site TAG (Williams & Rona, 1986). Sa classification complète selon la base de données WoRMS est la suivante :

Classification: Biota > Animalia (Kingdom) > Arthropoda (Phylum) > Crustacea (Subphylum) > Multicrustacea (Superclass) > Malacostraca (Class) > Eumalacostraca (Subclass) > Eucarida (Superorder) > Decapoda (Order) > Pleocyemata (Suborder) > Caridea (Infraorder) > Bresilioidea (Superfamily) > Alvinocarididae (Family) > *Rimicaris* (Genus) > *Rimicaris exoculata* (Species).

Cette espèce est retrouvée uniquement sur certains sites hydrothermaux de la MAR : Lucky Strike (Gebruk *et al.*, 2000), Rainbow (Desbruyères *et al.*, 2001), Broken Spur (Copley *et al.*, 1997, Gebruk *et al.*, 1997), TAG (Williams & Rona, 1986), Snake Pit (Segonzac, 1992), Logatchev (Martin *et al.*, 1997), Wideawake, Red Lion (Koschinsky *et al.*, 2006), et Ashadzé (Fabri *et al.*, 2011). Cette crevette vit en agrégats (à l'exception des sites Lucky Strike et Ashadzé où on ne trouve que quelques individus) pouvant atteindre une densité 2 500 individus.m⁻² (Segonzac *et al.*, 1993, Desbruyères *et al.*, 2001), à proximité directe des sorties de fluide. Elle vit à des températures comprises entre 10°C et 30°C (Zbinden *et al.*, 2004, Schmidt *et al.*, 2008). Des observations vidéo font état de crevettes traversant de manière très ponctuelle le fluide chaud en sortie de

cheminée, certaines présentant même des marques de brûlures, voire le céphalothorax percé (Van Dover *et al.*, 1988, Segonzac *et al.*, 1993, Gebruk *et al.*, 2000).

D'autres espèces du genre *Rimicaris* ont été décrites, notamment *Rimicaris kairei*, crevette endémique de la ride indienne (Watabe & Hashimoto, 2002), ou encore *Rimicaris hybisae* provenant de la fosse des Caïmans (Nye *et al.*, 2012). Très récemment, un article (Vereshchaka *et al.*, 2015) a proposé de revoir la taxonomie de la famille des *Alvinocarididae*, incluant dans le genre *Rimicaris* toutes les espèces du genre *Chorocaris*, y compris *Chorocaris chacei*, qui avait été initialement décrite comme *Rimicaris chacei* (Williams & Rona, 1986). Ce sont donc à ce jour 9 espèces de *Rimicaris* qui sont recensées.

Seules 3 espèces possèdent un céphalothorax hypertrophié : *R. exoculata*, *R. kairei* et *R. hybisae*. Les dernières classifications phylogénétiques les donnent comme étant des espèces clairement séparées. Cependant, il est à noter qu'à ce jour, *R. hybisae*, bien que présentant clairement un céphalothorax hypertrophié et étant donc physiquement distincte de *R. chacei*, est très proche de *R. chacei*, objet de la présente thèse, d'un point de vue phylogénie (Vereshchaka *et al.*, 2015). A ce jour, de nouvelles études avec plusieurs marqueurs génétiques montrent une très grande proximité entre *R. hybisae* et *R. chacei* qui pourraient avoir divergé très récemment (Hernandez *et al.*, in prep). De plus, comme *R. exoculata*, *R. hybisae* présente une symbiose au niveau de son céphalothorax avec les mêmes lignées (Assier *et al.*, pers comm in prep.). Ces résultats posent des questions tant d'un point de vue établissement de la symbiose, qu'évolutif (voir article 1).

Comme tous les décapodes, le genre *Rimicaris* (Figure 5) présente un corps découpé en 20 métamères, regroupés en trois tagmes : la tête, le thorax (qui comprend 8 métamères, dont 5 portent des pattes locomotrices) et l'abdomen.



Figure 5 : Photographie du modèle *Rimicaris exoculata*.

Le terme céphalothorax correspond aux tagmes de la tête et du thorax qui sont recouverts par une carapace, qui, lors d'observations dorsales, ne permet pas leur distinction (d'où son nom de céphalothorax). Le genre *Rimicaris* présente cependant la caractéristique singulière de posséder un céphalothorax hypertrophié qui représente près de la moitié de sa longueur totale (Williams & Rona, 1986). Il abrite une communauté de bactéries filamenteuses très importante (Segonzac *et al.*, 1993, Zbinden *et al.*, 2004) . De plus, le genre *Rimicaris* présente la particularité de ne pas posséder de rostre, et d'avoir des yeux fortement modifiés, ne formant pas d'images mais possédant des photorécepteurs à la rhodopsine, sensibles aux très faibles intensités lumineuses par les cheminées hydrothermales. Cet organe pourrait donc être un des systèmes sensoriels lui permettant de se repérer dans son environnement (Pelli & Chamberlain, 1989, Van Dover *et al.*, 1989, O'Neill *et al.*, 1995, Nuckley *et al.*, 1996, Chamberlain, 2000, Komai & Segonzac, 2008).

Les lames branchiostèges (ou branchiostégites) sont les replis latéraux de la carapace du céphalothorax. Elles ont une forme incurvée qui forme un microenvironnement presque clos. Dans cette cavité céphalothoracique se trouvent les scaphognathites (exopodites de la 2^{ème} paire de maxilles), en position dorsale, et les exopodites de la première paire maxillipèdes (que nous appellerons simplement exopodites), en position ventrale. Tous les deux sont hypertrophiés et couverts de soies bactériophores. Les battements des scaphognathites assurent la circulation de l'eau de mer vers les branchies, situées dans la partie postérieure du céphalothorax (Zbinden *et al.*, 2004). La face interne de la lame branchiostège, les scaphognathites et exopodites sont colonisés par de nombreux microorganismes sur lesquels nous reviendrons ci-après (Zbinden *et al.*, 2004, Corbari *et al.*, 2008, Corbari *et al.*, 2008).

Le cycle de vie de cette espèce n'est encore que très partiellement connu, certaines zones d'ombre ayant déjà fait l'objet d'éclaircissements. Les femelles portent des œufs, de couleur orange, protégés dans une gangue de mucus, sous l'abdomen jusqu'à éclosion (Llodra *et al.*, 2000, Guri *et al.*, 2012). Une femelle pourrait porter jusqu'à 800 œufs dont les dimensions peuvent varier de 300 à 400 µm (Tyler & Young, 1999). Aucune publication n'a pu clairement définir la période de reproduction de cette espèce, cependant, les travaux du doctorant Ivan Hernandez tendent à montrer une certaine saisonnalité, avec des pontes et des incubations des œufs durant les mois de décembre à mars (Hernández-Ávila *et al.*, 2016). Les prélèvements de crevettes lors de la campagne BICOSE 2014 (campagne effectuée pendant les mois de janvier et février) ont d'ailleurs montré une forte abondance de femelles gravides (portant des œufs).

Rimicaris exoculata étant présente des sites hydrothermaux de la MAR, répartis sur plusieurs centaines de kilomètres le long de la MAR, une forte capacité de dispersion est soupçonnée. Des études de populations basées sur des données de microsatellites ne montrent pas différenciation génétique entre les populations des différents sites (Teixeira *et al.*, 2011, Teixeira *et al.*, 2012). Les larves de cette espèce n'ayant jamais été observées sous forme libre au niveau des sites hydrothermaux, une capacité de dispersion sur de grandes

distances peut être envisagée, comme c'est le cas pour d'autres espèces d'invertébrés marins profonds (Marsh *et al.*, 2001, Hernández-Avila *et al.*, 2015).

Des regroupements de ce qu'on pensait être des juvéniles de *Rimicaris exoculata* en périphérie des sites hydrothermaux ont été nommés « nurserie » (campagne, BICOSE 2014). Cependant, il semblerait, après analyses moléculaires, que ces regroupements soient en réalité des juvéniles de *Rimicaris chacei*, la différenciation entre les deux espèces, au stade larvaire comme au stade juvénile n'étant pas aisée via des données de vidéos ou même des observations morphologiques (Hernández-Avila *et al.*, 2015). Cependant, des juvéniles de *Rimicaris exoculata* ont déjà été étudiés et présentent une coloration orange (identique chez *Rimicaris chacei*) correspondant à des réserves lipidiques, réserves qui s'amenuisent avec les cycles de mue (et donc la croissance) et du développement, jusqu'à disparaître au stade adulte (Pond *et al.*, 1998, Watling *et al.*, 2013).

Une histoire de mue

Comme tous les arthropodes, la crevette *Rimicaris exoculata* mue à intervalles réguliers afin d'assurer sa croissance, son exosquelette chitineux n'étant pas extensible comme la peau des mammifères. A chaque mue, l'ancienne cuticule (que l'on appelle exuvie) est perdue au profit d'une nouvelle. Cette étape du cycle de vie des arthropodes est coûteuse énergétiquement. Chez *Rimicaris exoculata*, un cycle de mue de 10 jours a été estimé, ce qui est très court par rapport à d'autres espèces de crevette, de même taille, comme l'espèce côtière *Macrobachium rosenbergii* dont le cycle de mue dure entre 41 et 98 jours (Corbari *et al.*, 2008). Un aspect intéressant du cycle de mue chez *Rimicaris exoculata* est qu'il peut être corrélé avec la coloration que prennent ces crevettes, en raison de l'accumulation de minéraux dans le céphalothorax. Lorsqu'un individu vient de muer, il est blanc transparent. Puis, durant son cycle de mue, les branchiostégites se recouvrent progressivement de minéraux (oxydes de fer sur le site Rainbow), donnant une couleur rouille à la crevette (orange puis rouge), ou encore des sulfates de fer comme sur le site de Logatchev ou Ashadze, donnant cette fois-ci une couleur grise, puis noire en fin de cycle de mue (Zbinden *et al.*, 2004, Corbari *et al.*, 2008, Corbari *et al.*, 2008).

Tout comme pour le cycle de vie, le régime trophique chez *Rimicaris exoculata* n'est pas totalement élucidé. Comme tout animal, cette crevette est hétérotrophe, elle a donc besoin de composés organiques afin de pouvoir alimenter son métabolisme. La morphologie de son céphalothorax ne lui permet cependant pas d'être nécrophage, ni prédatrice, car ses appendices buccaux sont trop courts pour pouvoir lui permettre toute préhension en dehors de la cavité céphalothoracique. Différentes hypothèses ont vu le jour quant au régime trophique de *Rimicaris exoculata* (Van Dover *et al.*, 1988). La première était que les microorganismes présents sur les parois des cheminées pouvaient lui servir d'alimentation (Van Dover *et al.*, 1988). Cette hypothèse fut partiellement réfutée par des analyses isotopiques (Polz *et al.*, 1998) qui démontrèrent que

l'ingestion de ces microorganismes comme seule source de matière organique n'était pas envisageable. Une autre hypothèse proposait que les bactéries présentes dans le céphalothorax pouvaient être ingérées (Casanova *et al.*, 1993), ou encore que des molécules organiques produites par ces mêmes bactéries pouvaient être transférées à la crevette via la cuticule interne de la lame branchiostège (Zbinden *et al.*, 2004, Guri *et al.*, 2012). Une dernière hypothèse implique l'utilisation de la communauté microbienne présente au niveau du tube digestive (Polz *et al.*, 1998, Zbinden & Cambon-Bonavita, 2003, Durand *et al.*, 2010). Parmi toutes ces hypothèses, celle concernant le transfert de matière des microorganismes vers la crevette a été approfondie et a mis en évidence, grâce à des incubations sous pression avec des substrats marqués radioactifs (bicarbonate ^{14}C), un passage transcuticulaire de molécules organiques d'origines bactériennes (Ponsard *et al.*, 2013).

Une double symbiose intéressante

Le céphalothorax

La colonisation microbienne des lames branchiostèges, des scaphognathites et des exopodites chez *Rimicaris exoculata* (figure 6) a été démontrée dans de nombreuses études (Casanova *et al.*, 1993, Segonzac *et al.*, 1993, Polz *et al.*, 1998, Zbinden *et al.*, 2004, Zbinden *et al.*, 2008).

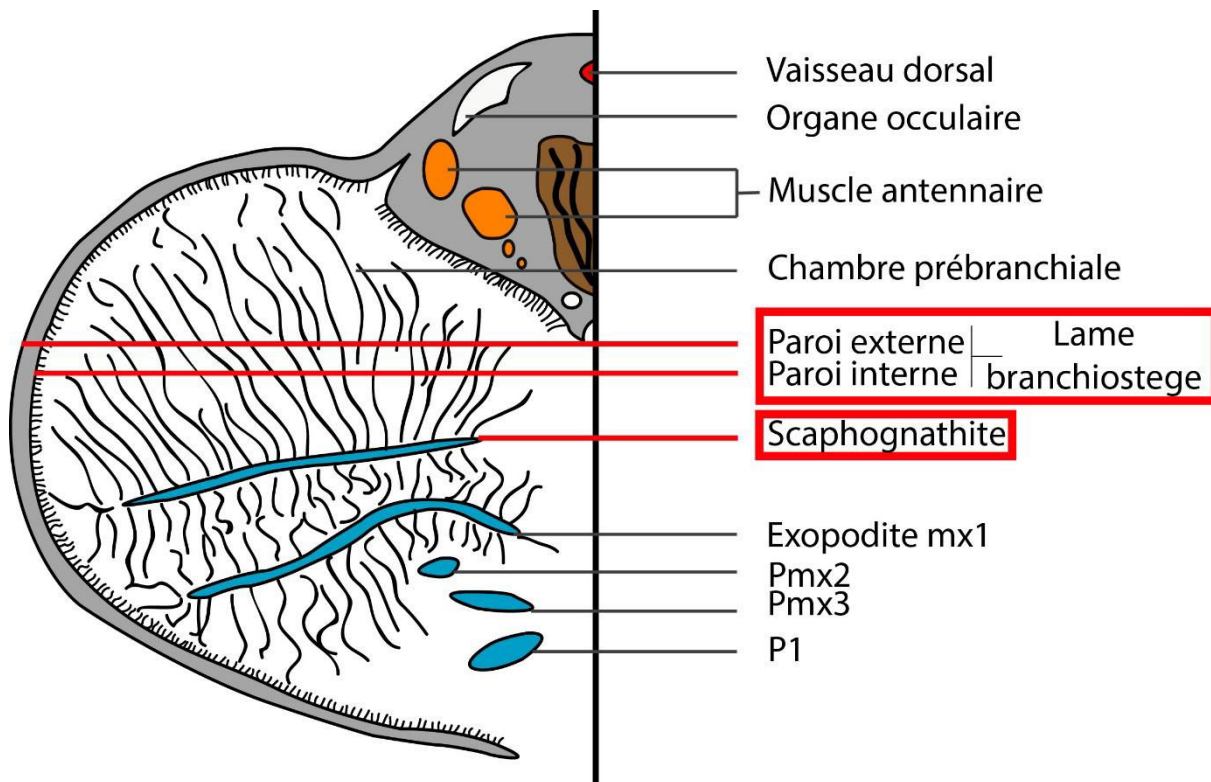


Figure 6 : *Rimicaris exoculata*, coupe transversale schématique du céphalothorax au niveau du bord antérieur de l'estomac. (modifié de Segonzac et al., 1993).

Les symbiontes ont dans un premier temps été décrits comme appartenant tous à un unique phylotype d'*Epsilonproteobacteria* (Polz & Cavanaugh, 1995). Les progrès des techniques de biologie moléculaire ont permis de montrer que cette diversité était plus complexe, présentant d'autres phylotypes associés toujours aux *Epsilonproteobacteria*, mais aussi aux *Gammaproteobacteria*, *Zetaproteobacteria*, *Bacteroidetes*, *Alphaproteobacteria*, et *Deltaproteobacteria*, ces deux derniers phylotypes étant cependant moins représentés (Zbinden et al., 2008, Petersen et al., 2010, Hügler et al., 2011, Guri et al., 2012, Jan et al., 2014). Des observations en microscopie électronique à balayage et transmission ont mis en évidence six morphotypes de bactéries : un morphotype de filaments épais, deux morphotypes de filaments fins, deux morphotypes de bacilles, et un morphotype de coques (ces derniers présentant des membranes internes caractéristiques de bactéries méthanotrophes) (Zbinden et al., 2004, Zbinden et al., 2008). La diversité phylogénétique connue de ces épibiontes étant supérieure à 6 OTUs (Zbinden et al., 2008, Petersen et al., 2010, Guri et al., 2012), ces observations laissent donc penser qu'un même morphotype peut représenter une certaine diversité phylogénétique plus complexe qu'initialement supposée (Zbinden et al., 2008). Lors de chaque mue, le tapis microbien ainsi que les minéraux associés sont éliminés des branchiostégites, scaphognathites et exopodites avec l'ancienne cuticule, laissant place à des surfaces vierges. La recolonisation se fait progressivement sous forme d'îlots (Corbari et al., 2008, Corbari et al., 2008). Les

bacilles semblent être les premiers colonisateurs, suivis par les filaments dont le nombre et la longueur augmentent au cours du cycle de mue. L'accumulation de particules minérales s'agrégeant autour des filaments augmente jusqu'à former une épaisse croûte donnant sa coloration à la crevette. Durant la période précédant la mue, l'épaisseur de l'encroûtement minéral est tel qu'il englobe presque la totalité de la longueur des filaments bactériens, ce qui les contraint dans l'espace, voire les « asphyxie », favorisant ainsi l'apparition de figures de lyses cellulaires (Durand *et al.*, 2010). La méthodologie qui a permis de corréler les approches moléculaires et de microscopie est le FISH (Fluorescente *In Situ* Hybridization). Cette méthode a permis de mettre en évidence la dominance de bactéries filamenteuses associées aux *Epsilon* et *Gammaproteobacteria* (Petersen *et al.*, 2010, Hügler *et al.*, 2011, Guri *et al.*, 2012), ainsi que la présence de bactéries de type méthanotrophes donnant à la bactérie une forme de « doonuts », car le compartiment cellulaire présentant les membranes internes ne contient pas ou très peu de ribosome qui sont répartis le long des membranes externes. Ces dernières sont affiliées aux *Gammaproteobacteria* (Petersen *et al.*, 2010, Hügler *et al.*, 2011, Guri *et al.*, 2012).

L'activité métabolique de ces symbiontes est devenue une question scientifique à part entière. Différentes voies métaboliques ont pu être mises en évidence via l'étude de gènes de fonction, notamment des voies associées à la sulfo-réduction et sulfo-oxydation (via les gènes *AprA*, *APS* et *soxB* (Zbinden *et al.*, 2008, Hügler *et al.*, 2011)), la méthanotrophie (via le gène *pmoA* (Zbinden *et al.*, 2008)), l'utilisation de l'hydrogène (via le gène *hynL* (Hügler *et al.*, 2011)), ainsi que la ferro-oxydation (Corbari *et al.*, 2008, Zbinden *et al.*, 2008). A l'heure actuelle aucun marqueur génétique bactérien de la ferro-oxydation à pH neutre n'est connu. Le métabolisme d'oxydation du fer chez les symbiontes de *Rimicaris exoculata* est supposé en raison des dépôts d'oxydes de fer au niveau du céphalothorax, et de la nature particulière de ces dépôts (Gloter *et al.*, 2004). La détection et l'affiliation de gènes impliqués dans le cycle inverse de l'acide citrique (via le gène *acIB*) ainsi que du cycle de Calvin-Benson-Bassham (via le gène *cbbM*) ont donné les premiers résultats confirmant le caractère chimolithoautotrophe des *Epsilon* et *Gammaproteobacteria* (Hügler *et al.*, 2011). Les auteurs de ces travaux ont posé l'hypothèse d'une possible relation syntrophique parmi les épibiontes. Des données de métagénomiques ont permis par la suite d'affilier certains des différents métabolismes rencontrés aux phylotypes déjà connus, notamment la sulfo oxydation chez les *Epsilon* et *Gammaproteobacteria*, ou encore le cycle rTCA chez les *Epsilonproteobacteria*, et le cycle CBB chez les *Gammaproteobacteria* (Jan *et al.*, 2014). Chez les *Zetaproteobacteria*, les mêmes données de métagénomique ont mis en évidence un métabolisme associé à la ferro-oxydation ainsi que du cycle CBB. Ce phylotype pose encore beaucoup de questions toujours sans réponse.

Le système digestif

La crevette *Rimicaris exoculata* possède un appareil digestif caractéristique. Le stomodeum (partie antérieure = œsophage + estomac) ainsi que le proctodeum (partie postérieure) sont plus réduits que chez la plupart des autres crustacés, favorisant ainsi une taille plus importante du mésentéron (dépourvu de cuticule et siège de l'absorption) (Komai & Segonzac, 2008) (Figure 7).

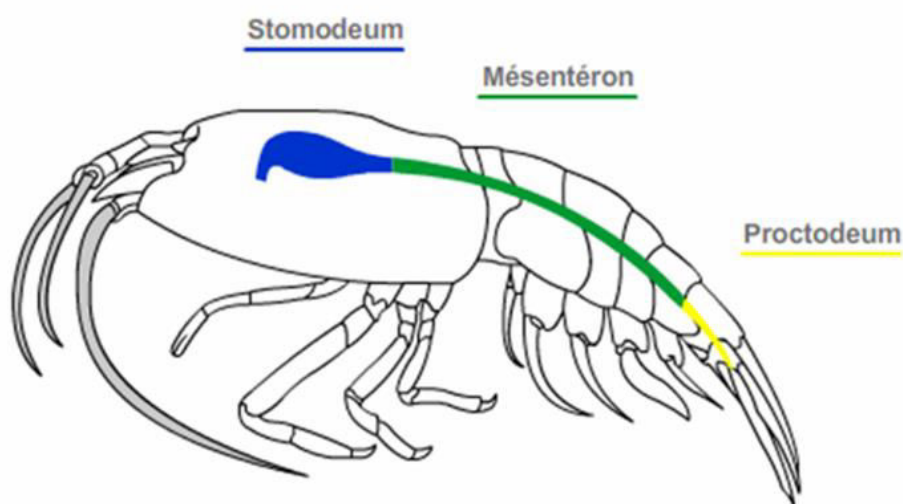


Figure 7: Représentation schématique des différentes parties du système digestif de la crevette *Rimicaris exoculata*.

Le contenu à la fois stomacal et intestinal est majoritairement composé de particules minérales typiques des cheminées hydrothermales (Van Dover *et al.*, 1988, Segonzac *et al.*, 1993). On y trouve également des exuvies de crevettes (Casanova *et al.*, 1993, Segonzac *et al.*, 1993). Les premières explorations du tube digestif chez cette crevette ont été effectuées suite à des travaux utilisant des marqueurs isotopiques qui révélèrent une fixation significative de carbone marqué dans des extraits de tubes digestifs (Polz *et al.*, 1998, Ponsard *et al.*, 2013). Des analyses moléculaires ont été entreprises et ont montré des communautés bactériennes distinctes de celles rencontrées dans le céphalothorax, notamment des *Epsilonproteobacteria* (très proches des lignées retrouvées dans le céphalothorax), des *Entomoplasmatales*, ainsi que des *Deferribacterales* (Zbinden & Cambon-Bonavita, 2003). Une étude plus récente (Durand *et al.*, 2010) a mis en évidence un phylum plus large que celui des *Entomoplasmatales*, les Mollicutes. Un certain nombre des phylotypes rencontrés ne sont pas retrouvés dans l'environnement. Leur origine ainsi que leur rôle restent donc un mystère, mais laissent supposer une communauté microbienne spécifique. De plus, des observations microscopiques ont mis en évidence un seul morphotype de bactéries filamenteuses tapissant les parois du

tube digestif, et s'insérant entre les microvillosités des cellules de l'épithélium intestinal, ainsi que des bactéries et des *archae* dans le bol alimentaire (Durand *et al.*, 2010). L'association durable de ces microorganismes vis-à-vis de leur hôte renforce l'idée d'une association spécifique de par leur localisation. En effet le mésentéron n'est pas soumis à l'exuviation, l'épithélium intestinal n'est donc jamais renouvelé. De plus, les cellules épithéliales ne sont pas en contact direct avec le bol alimentaire grâce à une membrane périthrophique qui englobe ce dernier.

Au sein du bol alimentaire, des morphotypes de coques et de bacilles ont été observés (Zbinden & Cambon-Bonavita, 2003, Durand *et al.*, 2010), dont certains sont affiliés aux *archae*. Ils semblent cependant ingérés de manière non sélective (Durand *et al.* 2010). Toujours dans l'idée de confirmer et de mieux comprendre le caractère spécifique de cette relation au sein du tube digestif, des lots d'individus de *Rimicaris exoculata* ont été soumis à des périodes de jeûnes (de 8, 22 et 72 heures) en aquarium pressurisé (Shillito *et al.*, 2008) afin d'éliminer leur bol alimentaire. Ces expériences ont permis de montrer le maintien après jeûne des *Epsilon*, *Gammaproteobacteria*, et Mollicutes, confirmant ainsi leur caractère spécifique. En revanche l'abondance de *Deferribacterales* diminuait de manière significative après le jeûne (Durand *et al.*, 2010). Trois hypothèses à ceci : les *Deferribacterales* dépendent des apports en minéraux tels que le soufre et le fer dans le bol alimentaire, elles peuvent être présentes de façon libre dans l'estomac et éliminées dans le bol alimentaire ; ou elles pourraient être strictement hétérotrophes et dépendre de la matière organique présente dans le bol alimentaire (Durand *et al.*, 2010). Des analyses via des approches en métagénomiques sont en cours au laboratoire pour essayer de confirmer ou infirmer ces hypothèses.

Le modèle *Rimicaris chacei*

Rimicaris exoculata (figure 8) n'est pas la seule espèce de crevette hydrothermale pouvant présenter un intérêt en termes de symbiose. *Rimicaris chacei*, groupe sœur de *Rimicaris exoculata*, partage des traits phylogénétiques, morphologiques et écologiques avec *Rimicaris exoculata*, en plus d'une colonisation bactérienne potentiellement symbiotique décrite dans la partie qui suit.

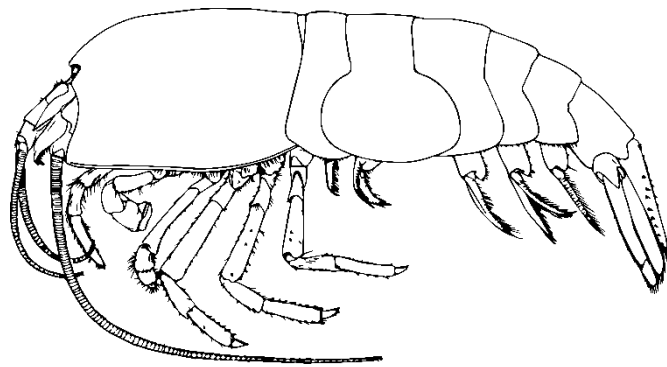


Figure 8 : A gauche, photographie du céphalothorax de *Rimicaris chacei*. A droite, dessin de description de *Rimicaris chacei* (Williams and Rona, 1986).

Le genre *Chorocaris* a été établi en 1990 lors de la découverte de l'espèce *Chorocaris vandoverae* (Martin & Hessler, 1990). L'espèce *Rimicaris chacei* a cependant été décrite en 1986 (en même temps que *Rimicaris exoculata*) (Williams & Rona, 1986). Pour des raisons de différences morphologiques avec le genre *Rimicaris* (notamment l'hypertrophie du céphalothorax), cette espèce a été reclassée au sein du genre *Chorocaris* en 2008 (Komai & Segonzac, 2008), puis de nouveau reclassée au sein du genre *Rimicaris* en 2015 (Vereshchaka *et al.*, 2015).

D'un point de vue anatomique, les trois espèces de crevettes *Rimicaris exoculata*, *Rimicaris chacei*, et *Alvinocaris markensis* présentent des différences / adaptations principalement au niveau du céphalothorax, que l'on pourrait lier à la présence de symbiose (Casanova *et al.*, 1993, Segonzac *et al.*, 1993). En effet, *Rimicaris exoculata* représente l'espèce symbiotique, *Alvinocaris markensis* représente l'espèce non symbiotique (semblable aux espèces côtières), et *Rimicaris chacei* semble être un stade intermédiaire. Là où *Rimicaris exoculata* présente des exopodites et des scaphognathites hypertrophiés qui s'étalent sur les deux tiers de la carapace de l'animal, *Rimicaris chacei* présente des exopodites et des scaphognathites dont la longueur est comparable à celle rencontrée chez *Alvinocaris markensis*, ne représentant qu'un tiers de la taille de l'animal (Casanova *et al.*, 1993). Chez *Rimicaris exoculata*, ces appendices sont densément couverts des longues soies bactériophore, totalement absentes chez *Alvinocaris markensis*, et de taille courte et moins densément réparties chez *Rimicaris chacei* (Casanova *et al.*, 1993). La cavité céphalothoracique presque close chez *Rimicaris exoculata*, ne permet pas aux pinces de sortir, mais reste dans une configuration « traditionnelle » chez *Rimicaris chacei*, à l'image de la configuration retrouvée chez *Alvinocaris markensis*, laissant ainsi les 2 premières paires de chélipèdes apparentes (Casanova *et al.*, 1993, Segonzac *et al.*, 1993). Les volumes stomacaux diffèrent aussi entre ces trois espèces. Ils sont réduits chez *Rimicaris exoculata*, intermédiaire, chez *Rimicaris chacei*, et normalement développé chez *Alvinocaris markensis*, laissant supposer une utilisation différente du système digestif (Casanova *et al.*, 1993) chez ces crevettes.

Ecologiquement, l'espèce *Rimicaris chacei* est proche de *Rimicaris exoculata*. *Rimicaris chacei* est présente sur plusieurs sites hydrothermaux de la MAR : Lucky Strike, Rainbow, TAG, Snake Pit et Logatchev (Gebruk *et al.*, 1997, Komai & Segonzac, 2008). Les deux espèces, en plus d'être présentes sur des sites hydrothermaux communs, occupent des niches écologiques qui se recouvrent. *Rimicaris chacei* vit à proximité des agrégats de *Rimicaris exoculata* mais aussi nettement plus en périphérie (Segonzac *et al.* 1993). Comme décrit dans la partie précédente, *Rimicaris exoculata* vit à proximité directe des sources d'émission de fluides, qu'il s'agisse de fumeurs, de ruches (quand les conditions le permettent), ou de rebords. De plus *Rimicaris exoculata* vit la plupart du temps regroupée en denses agrégats. L'espèce *Rimicaris chacei* quant à elle, ne vit pas en agrégats, mais plutôt en petits groupes, en périphérie des agrégats de *Rimicaris exoculata*, ou autour d'émissions diffuses de fluide. De plus, la densité de *Rimicaris chacei* estimée à 200, voire 300 individus par dm² (Segonzac *et al.*, 1993) ou jusqu'à 50 individus/m² dans d'autres articles (Gebruk *et al.*, 1997) est bien moins importante que celle estimée pour *Rimicaris exoculata* qui est de 2500 individus par m² (Segonzac *et al.* 1993). Cette différence de densité est telle que les individus de *Rimicaris chacei* peuvent être dénombrés sur une surface, tandis que les individus de *Rimicaris exoculata* doivent nécessairement être dénombrés dans un volume. Des stades juvéniles de *Rimicaris chacei* ont été localisés puis identifiés (utilisation de loupe binoculaire) durant la campagne BICOSE 2014 dans une zone appelée « nurserie », en périphérie de la zone active du site TAG. La présence d'individus adultes de *Rimicaris exoculata* sur cette même zone a tout d'abord laissé penser que les juvéniles en présence étaient de l'espèce *Rimicaris exoculata*. Il s'avère que ces juvéniles ont été génétiquement identifiés comme appartenant à l'espèce *Rimicaris chacei* (Hernández-Avila *et al.*, 2015).

Le régime trophique de *Rimicaris chacei* est encore mal défini. Des fragments de cuticule sont retrouvés dans son système digestif (Segonzac *et al.*, 1993). Des analyses isotopiques, notamment celles de l'azote, montrent que les juvéniles de *Rimicaris exoculata* pourraient être des proies de *Rimicaris chacei* adultes (Gebruk *et al.*, 2000). Des comportements de prédation de *Rimicaris chacei* envers *Rimicaris exoculata* ont été observés *in situ*, surtout chez des animaux blessés (communication personnelle Marie-Anne Cambon-Bonavita Campagne BICOSE 2014 et observations personnelles des vidéos de campagne). Tout ceci pourrait laisser présager un régime trophique de type prédateur, voire nécrophage, comme c'est le cas pour l'espèce *Alvinocaris markensis* (Segonzac *et al.*, 1993). Cependant, on retrouve aussi chez *Rimicaris chacei* une colonisation microbienne, notamment au niveau du céphalothorax (Casanova *et al.*, 1993, Segonzac *et al.*, 1993). Les mêmes études indiquent aussi chez cette espèce un régime de type prédateur, voire nécrophage, tendent à montrer qu'une partie de l'apport énergétique pourrait être d'origine symbiotique (Segonzac *et al.*, 1993). En effet, une quantité non négligeable du carbone retrouvé dans les tissus chez cette espèce serait dérivé de bactéries chimiosynthétiques (plus précisément via une RuBiSCO de type II) (Gebruk *et al.*, 2000). Des observations ont montré la présence de bactéries dans des régions protégées du céphalothorax, comme

la membrane interne de la carapace (ce que nous avons appelé jusqu'ici lame branchiostège, ou branchiostégite), sur les soies de la bordure des lames branchiostèges, sur les dents des maxillules, sur les soies des maxilles, ainsi que sur les premiers maxillipèdes (Casanova *et al.*, 1993).

Différents morphotypes de microorganismes ont été observés : un morphotype de bacille, et deux morphotypes filamenteux (Casanova *et al.*, 1993). Les dimensions (Longueur/diamètre) du morphotype de bacille sont de 0,5 à 0,8 μm / 0,2 à 0,4 μm . Les filaments, quant à eux existent sous deux formes : l'une rare, de petit diamètre (0,2 à 0,5 μm) ; l'autre plus commune, de grand diamètre (0,8 à 2,5 μm chez *Rimicaris exoculata* et 0,5 à 1 μm chez *Rimicaris chacei*). Ces gros filaments, constitués par des cellules cylindriques dont la hauteur est sensiblement égale au diamètre, forment des chaînes mesurant 0,5 à 1,5 mm chez *Rimicaris exoculata* et 0,2 à 0,5 mm chez *Rimicaris chacei* (Casanova *et al.*, 1993). L'ensemble de ces observations, bien que similaires à celles rencontrées chez *Rimicaris exoculata*, indiquent une moindre abondance de bactéries chez *Rimicaris chacei* (Casanova *et al.*, 1993, Segonzac *et al.*, 1993). Chez *Rimicaris exoculata*, le céphalothorax a subi une adaptation morphologique se traduisant par hypertrophie du céphalothorax permettant d'accueillir une symbiose développée. Chez *Rimicaris chacei* cette transformation est moins profonde, ce qui indiquerait une adaptation moins poussée, et une symbiose qui semblerait alors moins exclusive que chez *Rimicaris exoculata* (Casanova *et al.*, 1993). L'espèce *Rimicaris chacei* pourrait donc présenter un régime trophique mixte, mêlant symbiose (comme chez *Rimicaris exoculata*) et prédation/nécrophagie (comme chez *Alvinocaris markensis*) (Casanova *et al.*, 1993, Segonzac *et al.*, 1993, Gebruk *et al.*, 1997, Gebruk *et al.*, 2000).

Le barcoding/metabarcoding dans le monde microbien

La philosophie autour du barcoding/metabarcoding

Depuis de nombreux siècles, identifier et catégoriser le monde vivant qui nous entoure est une part non négligeable du travail réalisé dans le monde philosophique/scientifique. Cette identification a d'abord été réalisée sur des facteurs morpho-anatomiques, jusqu'à l'apparition de la génétique, venant compléter / affiner l'identification du monde vivant.

Il devient donc nécessaire de trouver un autre facteur permettant de discriminer des espèces entre elles, autre que la reproduction et la morphologie. L'ADN est un élément constant dans le monde du vivant, ce qui en fait un bon outil. Cependant, on ne peut pas (en 2017 tout du moins) obtenir l'ADN complet d'un organisme aussi facilement qu'une photographie (d'un animal par exemple). La reconstruction de génomes reste encore longue à traiter et coûteuse (bien que tout ceci tende à diminuer). On se contente donc d'étudier des gènes bien précis. Bien sûr, leur choix n'est pas dû au hasard. Dans un premier temps, il doit

être présent chez tous les organismes que l'on souhaite étudier. Ainsi, étudier un gène de fonction tel que *pmoA* (un gène intervenant dans le cycle du méthane) alors que certains organismes de la population étudiée ne sont pas capables de métaboliser ce composé, ne sert à rien si l'on souhaite identifier tous les microorganismes en présence. Cependant, si on souhaite étudier uniquement la population de microorganismes utilisant le méthane car les autres ne nous intéressent pas, le choix de ce gène peut être judicieux.

Un autre facteur à prendre en compte est ce que l'on appelle « l'horloge moléculaire ». Ce phénomène naturel est plus compréhensible en l'expliquant à travers la structure d'une protéine et fait appel à deux notions : la sélection naturelle et les mutations génétiques. Les mutations font partie à part entière du système évolutif du monde vivant. Le but ici n'est pas de rentrer dans le détail des mécanismes liés aux mutations génétiques. Cependant lorsqu'une mutation se produit, c'est sur l'ADN. Ce changement d'acide nucléique, s'il a lieu au sein de la séquence d'un gène, peut avoir des conséquences sur l'acide aminé qui sera traduit lors de la synthèse de la protéine correspondante au gène, et donc potentiellement sur l'activité de la dite protéine. Si ce n'est pas le cas, la mutation a un effet neutre.

Si une mutation entraîne une modification d'un acide aminé lors de la synthèse protéique, et que ce nouvel acide aminé modifie la structure de la protéine (ce qui n'est pas toujours le cas), deux possibilités s'offrent à nous : la modification de structure n'est pas dans une région soumise à une pression de sélection, la mutation n'a donc pas de conséquence notable, elle ne joue pas sur l'horloge moléculaire et est dite « silencieuse ». La deuxième possibilité est que le changement de structure de la protéine se fait dans une région soumise à une pression de sélection. Dans ce cas, la fonction de la protéine peut être compromise, et selon l'importance de celle-ci, le pronostic vital de l'organisme concerné par cette mutation peut être mis en jeu. Un exemple est sûrement plus parlant :

L'insuline est une protéine indispensable à la régulation de la glycémie chez l'homme (et d'autres animaux). Cependant, l'insuline n'est pas directement synthétisée lors de la traduction. Juste après la traduction, l'insuline est sous une forme inactive et nécessite une modification post-traductionnelle pour devenir active. Elle doit pour cela créer des ponts disulfures entre les chaînes A et B (figure 12), et perdre sa chaîne C.

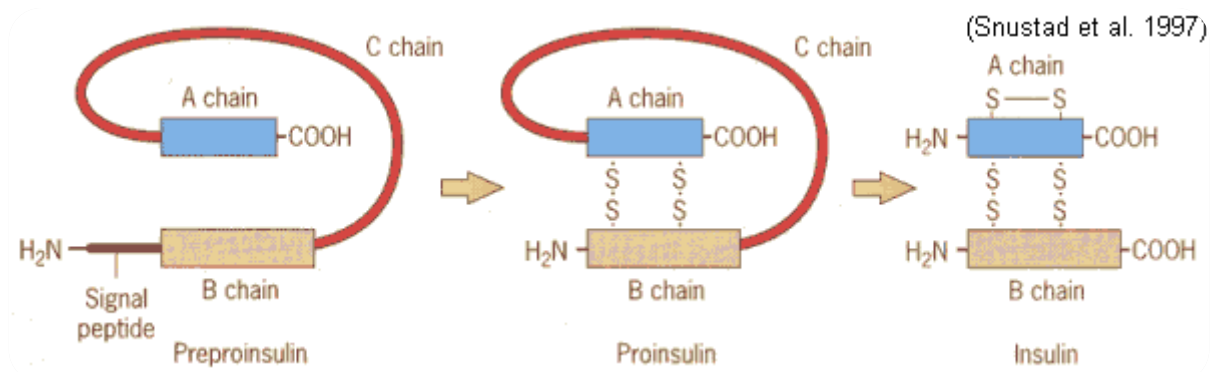


Figure 9 : Représentation schématique des différentes sous unités de l'insuline et de sa transformation post-traduction.

De manière simple, une mutation entraînant une modification de la structure protéique au niveau de la chaîne C aura moins d'impact (puisque cette partie est vouée à disparaître) sur l'activité de l'insuline que si elle a lieu sur la chaîne A ou B. Une modification de structure sur la chaîne A ou B pourra éventuellement donner une insuline moins efficace, ce qui ne sera pas avantageux pour l'organisme. On distingue donc une pression de sélection au sein d'un gène (mutation concernant le site actif ou non), ainsi qu'une pression de sélection sur la fonction même du gène concerné. Le gène codant pour la couleur de vos cheveux est moins vital que celui de l'insuline.

Tout ceci pour dire que plus un gène est soumis à une pression de sélection (selon son rôle), et moins il évoluera vite dans le temps, puisque qu'une mutation peut entraîner dans le pire des cas la mort de l'individu concerné. Donc selon la question évolutive à laquelle on souhaite répondre, il est important de correctement choisir le gène utilisé comme marqueur.

Comme exemple il est possible de comparer l'horloge moléculaire de trois gènes : celui codant pour un Fibrinopeptide, celui codant pour l'hémoglobine, et celui codant pour le cytochrome C, et de les comparer à des événements de l'histoire du vivant relativement parlants : la différenciation entre les grands groupes du monde animal. Le cytochrome C (et donc le gène correspondant) est soumis à une très forte pression de sélection. Il n'évolue donc que très peu dans le temps. On peut donc retracer un événement évolutif très ancien avec ce gène, comme la différenciation entre les plantes et les animaux. Cependant, les mammifères étant un groupe animal très récent dans le monde du vivant, il n'y a pas encore assez de mutations sur le gène du cytochrome C chez ces organismes pour pouvoir les différencier avec ce gène. Pour cela, il faut donc utiliser une protéine (et donc le gène associé) moins soumise à la pression de sélection, comme les fibrinopeptides, qui évolue donc plus vite, et présentera donc plus de différences chez les mammifères que le cytochrome C.

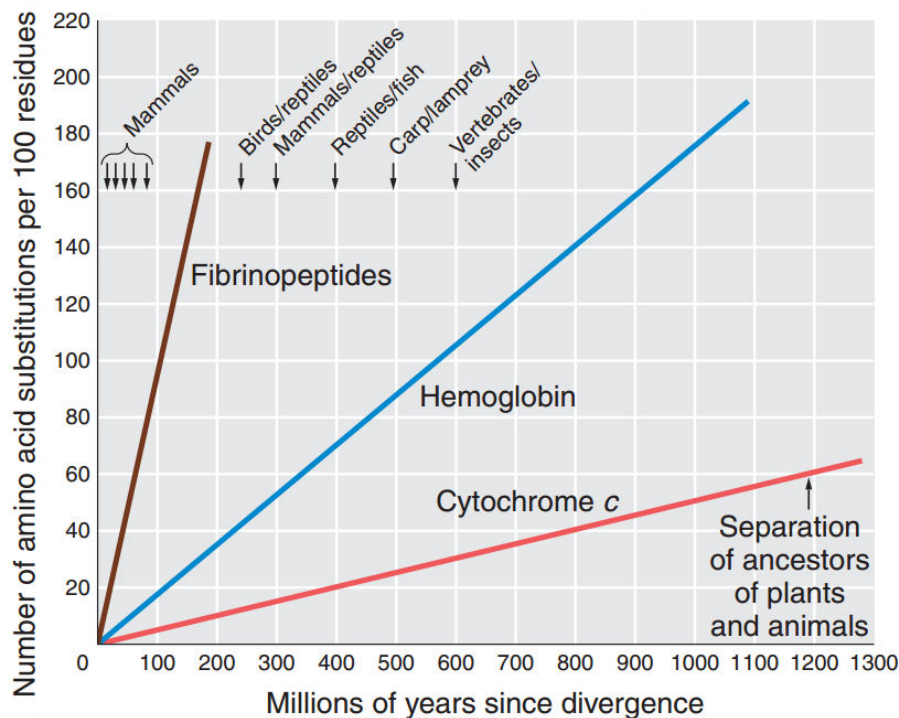


Figure 10 : Représentation schématique de l'horloge évolutive de Fibrinopeptides, de l'Hémoglobine et du cytochrome c au cours du temps.

Donc plus on souhaite remonter loin dans le temps, plus le marqueur génétique doit être soumis à une forte pression de sélection.

La catégorisation et la classification d'individus dans une arborescence est couramment appelée aujourd'hui phylogénie moléculaire par abus de langage, sans pour autant faire valoir d'aspects d'évolutions, notamment en microbiologie (car elle est bien souvent issue de données génétiques). En plus des difficultés décrites précédemment, cette arborescence se construit sous la forme de : Règne, Phylum, Classe, Ordre Espèce. Mais se pose une question qui fait toujours débat dans le monde scientifique : celui de définir ce qu'est une espèce. Nous n'entrerons pas dans ce débat, ni n'en ferons un bilan bibliographique, mais un élément décrit par les biologistes est que deux individus appartenant à la même espèce sont deux individus pouvant se reproduire, et donner une descendance elle-même capable de se reproduire.

Si cet élément de description d'une espèce devient de plus en plus problématique, notamment en botanique, il l'est encore plus dans le monde microbien. L'élément de définition précédent implique un facteur quelque peu problématique ici, dans le monde microbien, où la reproduction est principalement asexuée, basée sur le principe de la mitose. Comment donc définir une espèce ? Comme dans cette thèse nous nous intéressons exclusivement au règne bactérien, nous focaliserons sur ce point. Depuis de nombreuses années le terme d'« Operational Taxonomic Unit » (OTU) est utilisé pour définir une « espèce bactérienne ». Il s'agit à

nouveau d'un abus de langage de la 'microbiologie moléculaire'. Une espèce microbienne par définition est un organisme microbien clonal décrit, et cette description répond à de nombreux critères établis. Tandis que pour une OTU, il s'agit d'un ensemble de séquences nucléotidiques issues du gène codant pour la petite sous unité du ribosome (16S) partageant une similarité de plus de 97% (ce seuil est néanmoins arbitraire et a été établi comme seuil d'espèce dans la description microbiologique pour la partie moléculaire Stakerbrandt et al., 1994). Nous verrons plus tard que bien que ce seuil soit fixe et couramment admis, son utilisation, et les interprétations qui en résultent n'en restent pas moins fluctuantes.

Le barcoding est donc dans le sens large du terme une méthode d'analyse visant à identifier des organismes un à un via l'analyse d'un marqueur génétique, ici le 16S, car c'est un marqueur universel, long, présentant des régions conservées, et variables. Le champ d'application du barcoding est cependant limité à une identification d'un petit nombre d'individus (100 à 500 individus), le nombre de séquences devenant vite trop chronophage à analyser. Le metabarcoding, quant à lui, a le même but, mais pour un nombre d'organismes beaucoup plus important (plusieurs milliers, millions, milliards d'individus).

La pratique

Si en théorie, l'identification d'un organisme dans un écosystème semble simple, il l'est beaucoup moins dans la pratique, qui plus est lorsque l'on souhaite mettre ne place du barcoding pour réaliser des analyses écologiques.

En effet, il faut avoir conscience de certains biais qui peuvent être introduits dans ce type d'analyses, résumés (mais non développés) dans le tableau suivant.

Biais	Barcoding	Metabarcoding
Echantillonnage	OUI	OUI
Extraction d'ADN	OUI	OUI
Choix des amorces PCR utilisées	OUI	OUI
PCR en elle-même	OUI	OUI
Création des librairies	OUI	OUI
Séquençage	NON	OUI

Figure 11 : Tableau récapitulant les différents biais méthodologiques pouvant être rencontrés avant un traitement bioinformatique de données de barcoding et de metabarcoding.

Dans les faits, l'identification manuelle d'une centaine d'individus reste concevable, elle l'est moins pour plusieurs millions.

Un facteur souvent limitant reste les **bases de données**. Il semble important de prendre le temps d'expliquer de manière claire ce biais par rapport aux autres (puisque'il a trait davantage à la bioinformatique). L'utilisation du gène codant pour la petite sous unité 16S est couramment admise pour les études de diversité, mais la grande majorité des procaryotes, que ce soit pour les bactéries, ou les archaea, ne sont pas cultivables. Or, on ne donne un nom de genre et d'espèce à un microorganisme comme on pourrait le faire chez les animaux ou les végétaux qu'à partir du moment où il est cultivable (et décrit selon des critères d'usages fixes – morphologie, mode de culture, optimum de croissance, données moléculaires, activités...). Comment donc faire pour un microorganisme non cultivable, qui représente environ 95% de la diversité d'un écosystème donné ? Il n'aura tout simplement pas de nom, et son identification taxonomique restera sommaire. De plus, on ne peut comparer que ce qui est comparable. Le gène codant pour la petite sous unité 16S, utilisé pour identifier les procaryotes, fait environ 1500pb. A ce jour, seul le séquençage Sanger permet un séquençage complet de ce gène (via les approches de barcoding, *ie* souche par souche). Ceci nécessite l'utilisation de la méthode de clonage pour obtenir des fragments d'ADNr 16S complets. C'est une méthodologie longue et coûteuse. Elle est donc de moins en moins utilisée sauf pour la description de souches. Le problème notable aujourd'hui est que de plus en plus de monde utilise les technologies de séquençage haut débit. Les meilleures technologies en 2017 produisent des fragments de 300pb, qui utilisés en « paired end » (c'est-à-dire un fragment séquencé dans les deux sens mais avec une portion de séquence commune pour les deux, présentant des zones de recouvrement minimum de 10 pb) peuvent donner au maximum des fragments d'ARNr 16S de 590 pb. Nous sommes donc bien loin des 1500 pb et de fait cela est limitant pour les analyses phylogénétiques qui suivent : on ne peut travailler que sur des séquences d'une même longueur et donc les plus courtes limitent les analyses. De plus, différentes régions du 16S peuvent être analysées selon les auteurs. Il en existe 9 décrites chez *E.coli*, appelées couramment régions v1 à v9. Au problème de séquences de gènes d'ARNr 16S incomplets, des régions étudiées pouvant varier en fonction de la question scientifique ou des pratiques de laboratoires, il faut ajouter le nombre de séquences générées par les nouvelles technologies de séquençage, souvent de plusieurs dizaines de millions de séquences, à plusieurs milliards. Il est donc inconcevable d'enrichir les bases de données avec ce type de données et elles sont parfois difficilement comparables.

Il faut donc bien avoir à l'esprit que l'utilisation du barcoding peut éventuellement participer à l'enrichissement des bases de données, avec des séquences cependant incomplètes, mais que le metabarcoding, permettant d'identifier un beaucoup plus grand nombre d'organismes, génère des bases de données beaucoup plus complexes à utiliser et traitées de façon différentes dans les banques de données. L'utilisation des deux approches a donc un but différent : le barcoding pour aider à l'identification

d'organismes isolés, tandis que le metabarcoding permet de répondre à des questions d'ordre écologique globales (c'est-à-dire des études de populations), si le design expérimental est correctement réalisé.

Trois outils différents pour parvenir au même résultat ?

Dans ce chapitre, il est important d'avoir à l'esprit que nous ne parlons que du traitement bioinformatique des données issues du séquençage. Nous ne discutons donc pas des différentes étapes qui pourraient introduire des biais dans une analyse de barcoding / metabarcoding que sont la qualité de l'échantillonnage, l'extraction d'ADN (manuelle, utilisation de kits divers...), le choix des amorces PCR et donc de la zone étudiée, la PCR en elle-même (choix de l'enzyme de réplication, nombre de cycles...), la création des librairies, le séquençage et les biostatistiques. Cela peut sembler être beaucoup d'aspects non abordés, mais relativement bien connus des utilisateurs de ces techniques, et chacun mériterait son propre chapitre. Dans ce travail je vais donc me focaliser sur la partie aval, moins classiquement connue : le traitement bioinformatique des données obtenues.

La grande question reste donc : pourquoi parler de trois outils de traitement des données de séquençage, pourquoi ne pas en prendre un et puis c'est tout ? A ceci je répondrai : pourquoi utiliser tel protocole/kit d'extraction d'ADN, pourquoi choisir telle amorce PCR, pourquoi tel programme PCR ... Chacun d'eux génère un biais et les panacher permet de limiter ces biais. Pour analyser les données de séquençage, il va falloir déterminer certains paramètres de qualité sur vos séquences car elles ne sont pas toutes exploitables. Pour le cas du metabarcoding, il va falloir déterminer des critères de recouvrement, il va aussi falloir aligner vos données, les grouper (clusteriser), les assigner taxonomiquement... Il existe plusieurs outils, algorithmes pour chacune de ces étapes, qui peuvent vous donner des résultats finaux différents. Alors comment choisir ? Comment être sûr ? Essayons-en plusieurs, non pas pour déterminer lequel est le meilleur, aucun n'est faux, sinon ils ne seraient pas publiés, ou même utilisés. Mais essayons-en plusieurs pour vérifier que le résultat obtenu par l'un est conforme à celui des autres. Cela ne revient-il pas à les comparer finalement ? Non, parce que ce qui sera intéressant ici, c'est de vérifier que la question scientifique que vous vous posez, ou vos hypothèses de travail scientifique, sont bien soutenues par l'ensemble de ces outils. Je me répète, mais c'est important : vous n'aurez pas à l'issue de ce chapitre, de réponse claire sur quel outil ou algorithme utiliser, mais des scripts informatiques permettant d'analyser vos données de séquençage même si vous n'êtes pas des professionnels de l'informatique. Ainsi, vous gagnerez du temps, à ne pas apprendre durant des heures à utiliser tel ou tel outils, à tester chaque commande pour l'écrire comme il faut, tout sera déjà fait. Vous n'aurez plus qu'à vous concentrer sur le résultat scientifique. Vous pouvez cependant vous poser la question: mais que font ces scripts, comment être sûr qu'il ne nous embobine pas ? La suite de ce chapitre est là pour ça.

Existe-il d'autres outils pour réaliser des analyses de barcoding/metabarcoding que Mothur, Qiime ou FROGS ? Oui, il en existe, notamment les OBITools, mais le choix a été fait ici d'utiliser Mothur et Qiime pour leur popularité, et FROGS, car il me semble que c'est un Workflow prometteur, qui sera qui plus est sûrement utilisé dans un projet d'envergure, ABYSS, au sein de mon laboratoire d'accueil à l'Ifremer.

Mothur et Qiime

Mothur est l'un des outils de bioinformatique les plus connus pour les analyses écologiques, des données de séquençage issues de technologies Sanger, PacBio, IonTorrent, 454 et Illumina (MiSeq/HiSeq). C'est un outil qui a été mis en ligne en 2009 et qui est la fusion de deux outils antérieurs développés par le même laboratoire : DOTUR et SONS. Mothur est disponible sous Windows en ligne de commande ou avec une interface, ainsi que sous Linux et Mac OS. La version interfacée étant peu ergonomique et peu souple, c'est pour cette raison que les scripts proposés ici concernent la version en ligne de commande. L'inconvénient majeur de Mothur est sa consommation de mémoire RAM. Certaines étapes sont très gourmandes, voire limitantes si l'ordinateur utilisé n'est pas assez puissant (comme par exemple les commandes align.seqs ou encore classify.seqs).

Qiime quant à lui est un outil semblable à Mothur en termes d'objectifs, richesses d'options, souplesse... et inconvénients. Cependant, cet outil est disponible uniquement sur Linux et Mac OS. Les utilisateurs Windows devront donc nécessairement utiliser une machine virtuelle (ou leur environnement Linux sur Windows10).

FROGS

FROGS est un outil différent de Mothur et Qiime, car il s'agit d'un « workflow » (c'est-à-dire qu'il s'agit d'un ensemble d'instructions/commandes d'un ensemble d'outils) développé pour être utilisé sur Galaxy (cf. Matériel et Méthodes). Du fait que ce soit un workflow, il est très peu, voire pas du tout flexible, mais traduit l'esprit de ce chapitre : simplifier la vie des personnes souhaitant analyser des données de metabarcoding de 16S ou 18S de façon ponctuelle. Pour souhaiter l'utiliser, il faut donc être d'accord avec le choix des outils employés.

Oligotyping

Ce quatrième outil, est en fait plutôt une nouvelle méthodologie, une nouvelle manière de voir les études d'écologie basées sur le 16S. Le but de cette méthode est de pouvoir "enquêter" sur la diversité dissimulée au sein d'une OTU. Elle se base sur l'utilisation de positions conservées dans la séquence de l'ADNr 16S afin de chercher une "microdiversité" et créer ainsi des oligotypes. Ainsi, au sein d'une seule OTU, il est possible de définir plusieurs oligotypes (Eren *et al.*, 2013, Eren *et al.*, 2014, Eren *et al.*, 2014).

Cette méthodologie se base sur l'utilisation de l'entropie de Shannon pour sélectionner les positions sur l'ADNr 16S susceptibles d'être intéressantes. Elle permet aussi d'éliminer une partie des biais liés à des erreurs de PCR, ou de séquençage (Eren *et al.*, 2013, Eren *et al.*, 2014, Eren *et al.*, 2014). Le script et les analyses n'ayant pas encore abouti à des résultats au cours de ma thèse faute de temps, il ne sera pas développé dans la suite de manuscrit.

Le sujet de thèse

Comme il a été explicité dans la partie bibliographique, les systèmes hydrothermaux sont des environnements géologiques actifs ponctuels répartis tout autour du globe à la jonction des plaques tectoniques, et en majorité dans les grands fonds marins. La pression hydrostatique, le caractère aphotique, et la toxicité des fluides font des systèmes hydrothermaux des environnements très intéressants, notamment pour l'étude du monde vivant et de ses limites (aussi bien macroscopique que microscopique). En l'absence de lumière, les écosystèmes de ces environnements sont basés sur la chimiosynthèse microbienne, base du réseau trophique. Dans ces environnements, les exemples de symbioses entre microorganismes et invertébrés sont nombreux. A l'heure actuelle, sur la MAR, le seul exemple d'Alvinocarididae où une symbiose trophique est avérée est la crevette *Rimicaris exoculata*. Or d'autres espèces de crevettes vivent dans ces zones (*Rimicaris chacei*, *Alvinocaris markensis*, ou encore *Mirocaris fortunata*). Ces espèces peuvent-elles présenter des interactions durables avec des microorganismes ? Si on part du principe que oui, les conditions environnementales étant différentes selon les sites (composition chimique du fluide, température, pression, pH...), les communautés symbiotiques peuvent-elles varier en fonction des conditions physico-chimiques ? Par conséquent, peut-on mettre en évidence des phénomènes de biogéographie des symbiontes ? Quelles sont les capacités métaboliques des symbiontes ? Quel est le degré de plasticité génétique chez ces symbiontes ? Les questions ne manquent pas, et sont trop nombreuses pour être résolues durant une thèse. C'est pourquoi ce sujet vise à répondre principalement à deux points : 1) Décrire les associations symbiotiques chez la crevette *Rimicaris chacei*, d'un point de vue morphologique et phylogénétique. Cette description permettra une comparaison avec la symbiose connue chez *Rimicaris exoculata*. 2) Mettre en évidence (ou non) un caractère biogéographique des symbiontes sur les sites de la MAR via une approche de séquençage haut débit (et traitement des données via l'oligotyping), pour *Rimicaris chacei* et *Rimicaris exoculata*, avec une comparaison possible entre les deux espèces et entre sites. L'ensemble de ces données permettra de discuter de l'évolution de ces deux espèces sœurs, présentant une niche écologique et des symbiontes proches.

Nous utiliserons différentes approches pour répondre à ces questions : des approches de biologie moléculaire (banques de clones, séquençage haut débit) et de microscopies (loupe binoculaire, microscopie photonique, microscopie électronique à balayage (MEB), microscopie électronique à transmission (MET), et hybridation *in situ* en fluorescence (FISH)).

Cette thèse assez descriptive s'est aussi construite dans le cadre particulier et relativement émergent de la découverte de nombreux minerais et métaux présents dans les dépôts minéraux hydrothermaux (notamment les dépôts de sulfures). Ces composés présentent une valeur commerciale de plus en plus attrayante pour les technologies modernes. L'attrait pour ces minéraux et métaux, ainsi que le

développement des compétences minières nouvelles en environnements profonds font peser sur ces écosystèmes des dangers nouveaux. L'exploitation de ces environnements par l'Homme n'est qu'une question de temps, et présente des risques potentiels de modifications drastiques pour ces écosystèmes suite aux extractions minières. Il est donc important de mieux comprendre ces environnements et l'ensemble des processus biologiques qui les sous-tendent dans leur fonctionnement naturel. D'un point de vue écotoxicologique, cela revient à faire une étude à très grande échelle du temps zéro (état de référence), c'est-à-dire de l'environnement avant sa détérioration éventuelle potentielle. Cette thèse descriptive participe donc à l'établissement de cet état zéro.

Nous présenterons ci-après les deux articles scientifiques prévus lors de l'élaboration du sujet de thèse. Le premier est soumis. Le deuxième article est en cours de rédaction.

Matériels et Méthodes :

Les scripts mis en place pour
traiter les données de manière
automatique ou semi-
automatique

Dans cette partie, seul le volet bioinformatique que j'ai développé au cours de la thèse est développé car il ne pourra pas être explicité dans les différents articles scientifiques. Le volet prélèvement des échantillons, fixations à bord du navire, ainsi que les techniques de microscopie sont décrites dans le matériel et méthodes du premier article.

Protocole de barcoding sur banque de clones

Durant ce travail de thèse, un jeu de données issues de banque de clones et séquençage Sanger a permis l'obtention de 4954 séquences exploitables. C'est un trop grand nombre de séquences pour être analysées manuellement comme cela est classiquement le cas au laboratoire. C'est pour cela qu'un protocole semi-automatique a été créé au cours de mon travail, pour des utilisateurs Windows, qui n'a cependant pas encore son script (Windows étant différent de Linux, le langage de programmation est lui aussi différent).

Quatre outils sont nécessaires pour ce protocole :

- Geneious © (payant)
- Sublime text (gratuit), ou un autre éditeur de texte excepté le bloc note)
- Microsoft office Excel (payant)
- Mothur (gratuit)

Les deux logiciels payants que sont Geneious© et Excel peuvent être remplacés par d'autres outils comme par exemple Benchling pour Geneious© ou encore libre office pour Excel. Bien que les explications soient faites pour les logiciels cités ci-dessus, il est possible de s'en sortir avec d'autres, le seul incontournable étant Mothur. Pour répondre tout de suite à la question de pourquoi utiliser Mothur, c'est parce que c'est le seul des quatre outils cités plus haut disponible sous Windows qui reste le principal système d'exploitation du commerce. De plus, les utilisateurs utilisant d'autres systèmes d'exploitation pourront tout à fait utiliser ce protocole en utilisant leurs propres outils favoris, puisque le seul logiciel incontournable est Mothur, qui est le seul disponible pour Windows, Mac OS et Linux.

Différentes étapes sont nécessaires pour traiter vos données à l'issue du séquençage Sanger.

Elimination des séquences de mauvaise qualité

Lors du séquençage, quatre principaux problèmes peuvent subvenir :

- Le séquençage a tout simplement été raté. C'est-à-dire que pour des raisons techniques et non biologiques, il n'y a pas d'amplification de signal.
- Il n'y a pas assez d'ADN dans le puits à séquencer pour qu'il soit amplifié.
- Deux colonies différentes ont été mises en culture dans un seul puits (pouvant être lié au repiquage des clones). On obtient donc une double séquence.
- Aucun ADN d'intérêt n'a été inséré dans le vecteur, et la colonie a quand même été envoyée à séquencer. On séquence donc du vecteur.

Dans les trois premiers cas, soit le prestataire de services (lors de la récupération des données), soit le %HQ (sur Geneious©) peut renseigner sur la qualité des séquences. J'ai pu remarquer (remarque personnelle) que lorsque le %HQ est inférieur à 30 %, les séquences ne sont pas exploitables (de manière systématique), et qu'entre 30 % et 40 %, il faut regarder chaque séquence indépendamment.

Dans le dernier cas, ni le %HQ ni les procédures du prestataire de services ne peuvent identifier la présence d'une séquence de vecteur. Ce type de séquence est identifiable lors des premiers alignements (car elle déstabilise l'alignement) et une simple identification sur la base de données NCBI (via un BLAST) permet de confirmer s'il s'agit bien de vecteur de clonage ou d'une autre erreur.

Tri des séquences

Dans le cas de l'obtention de séquences issues de banques de clones et donc, par conséquent, d'une étape de clonage, il est important d'avoir à l'esprit que certains vecteurs de clonage, pour des questions de rendement (en étude de diversité tout du moins), ne présentent pas de sélection d'orientation de l'insert lors de l'étape de ligature. C'est le cas notamment des vecteurs TOPO XL® et TOPO TA®, qui présentent une Thymine au niveau de la zone d'insertion du fragment d'ADN à cloner (Figure 15).

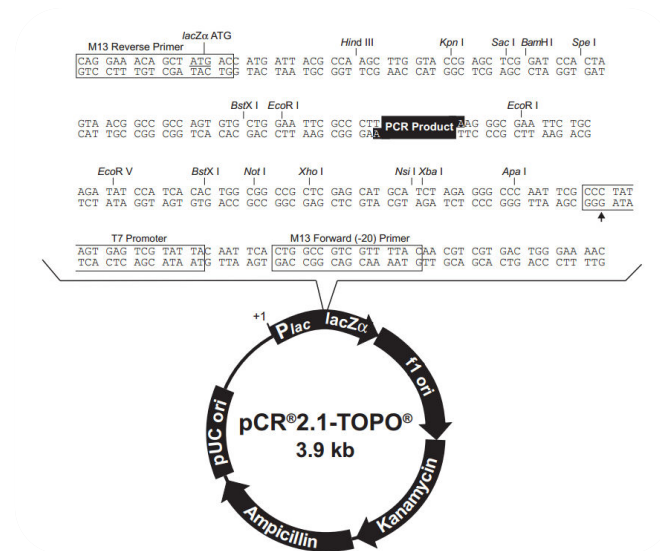


Figure 12 : Représentation schématique du vecteur de clonage TOPO TA®.

En effet, cela se base sur le fait que la polymérase présente la particularité de rajouter une série poly « A » à la fin des étapes d'élongation lors d'une PCR. Le clonage se fait donc par ligature A-T et le sens d'insertion de celle-ci dans le vecteur est donc aléatoire. Par la suite, le séquençage Sanger est initié à partir d'une amorce (ou primer en anglais). Le début du séquençage ne présente en général pas une qualité suffisante pour être utilisé (Figure 16).

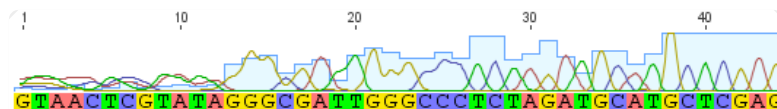


Figure 13 : Capture d'écran d'une portion de séquence de mauvaise qualité au format .AB1.

Afin de pouvoir obtenir le début de séquence de notre gène d'intérêt, il est possible d'utiliser une amorce déjà présente sur le vecteur, comme l'amorce M13 Forward, ou encore l'amorce M13 Reverse, en amont et aval de l'insert. C'est la solution que j'ai personnellement choisie pour exploiter mes données. Cependant, ce choix induit nécessairement un inconvénient dans le traitement des données : les séquences obtenues sont soit séquencées dans le sens direct, que l'on appellera Forward, ou encore dans le sens indirect, que l'on appellera Reverse (Figure 17).

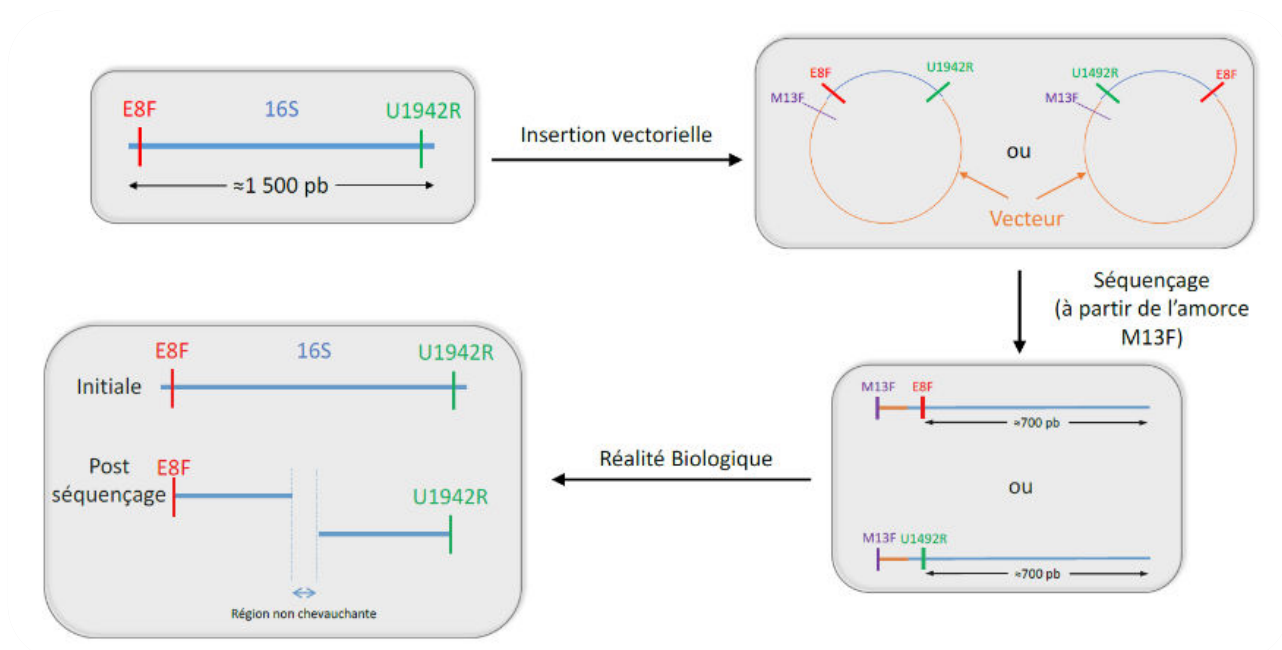


Figure 14 : Principe schématique d'une étape de clonage.

Si on prend l'exemple de l'utilisation des amorces universelles E8F et U1492R, la totalité du gène codant pour la petite sous unité du ribosome (16S) a été amplifiée et clonée, mais le séquençage du gène n'est que partiel. En effet avec une réaction de séquençage, on obtient donc soit le début, soit la fin de la séquence de l'ADNr 16S. Il est théoriquement possible d'avoir un chevauchement de séquences (début et fin du gène) et ainsi avoir un gène d'ARNr 16S complet. Cependant, bien que la longueur maximale de séquençage Sanger atteigne actuellement 900 à 1 000 paires de bases, il est fréquent d'avoir des séquences plus courtes et ne pas pouvoir reconstituer le gène dans sa totalité. Bien évidemment, un choix d'amorces permettant le séquençage d'une zone plus restreinte du gène à étudier reste possible. Il faut cependant tenir compte de la perte d'information que cela engendre, et du fait que les bases de données contiennent principalement des débuts de gènes d'ARNr 16S, souvent autour de 600 pb, qui de facto limite les analyses ultérieures.

Tout ceci pour expliquer que l'on obtient, dans le cas de l'utilisation des amorces E8F et U1492R pour amplifier notre gène d'intérêt, un jeu de données comportant à la fois un lot de séquences « Forward », et un lot de séquences « Reverse ». Or les alignements ne peuvent se faire que sur des séquences comparables, donc sur des zones du gène identiques. Ne pouvant donc être analysées simultanément (notamment d'un point de vue étude de diversité et « phylogénie »), il est nécessaire de séparer les séquences « Forward » des « Reverse ». Comment faire ? Il existe plusieurs méthodes (toutes via bioinformatique bien sûr). Trois seront présentées ici. Bien évidemment il existe sûrement d'autres méthodes. L'exemple sera ici donné avec le logiciel Geneious®.

Quelle que soit la méthode utilisée, créez un dossier *Forward* et un dossier *Reverse* (pour y classer les séquences correspondantes) dans votre jeu de données.

Première méthode

On utilise un alignement pour séparer les séquences « Forwards » des « Reverses ». La méthode d'alignement MUSCLE, par exemple, en plus d'être rapide, permet d'obtenir un résultat où les séquences sont rassemblées par similarité. Concrètement, on observe un jeu de séquences en haut, et un autre en bas de l'alignement, correspondant aux séquences « Forwards » et « Reverses » séparées. Si cela ne fonctionne pas, c'est que certaines séquences peuvent altérer l'alignement, comme par exemple des séquences identifiées comme du vecteur (c'est-à-dire que le plasmide s'est refermé sur lui-même sans avoir inséré de gène d'intérêt). Ces séquences doivent être supprimées du jeu de données. Cette méthode n'est pas celle que je recommande, car cela ne fonctionne pas à chaque fois, et une fois les deux lots de séquences identifiés, le tri restant manuel, et les séquences (dans l'alignement) n'étant pas classées par numérotation mais par similarité, le tri est plus long (et fastidieux).

Seconde méthode

Cette méthode est très semblable à la première. Cette fois ci, on utilise l'alignement Geneious © (lent... très lent... et peu efficace), mais qui présente l'avantage de séparer en deux clusters les séquences « Forward » des séquences « Reverse » comme dans la première méthode (mais de manière plus visuelle) (Figure 18).

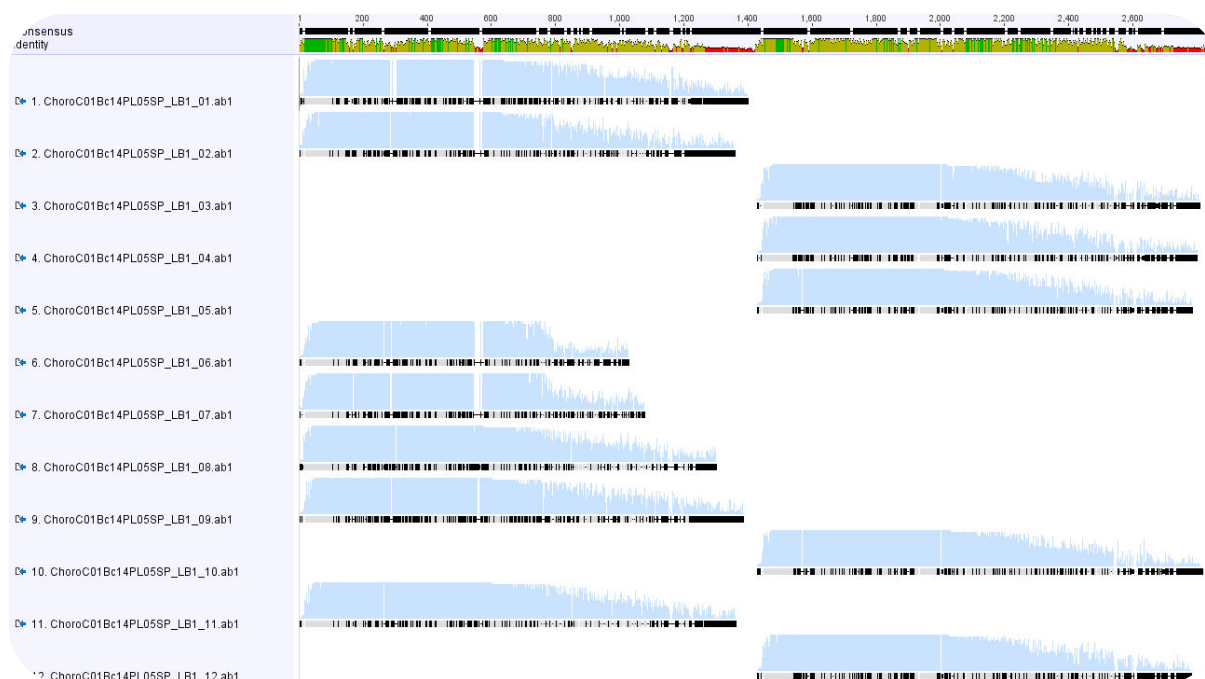


Figure 15 : Capture d'écran d'un alignement réalisé avec l'algorithme Geneious.

Les séquences restent classées dans l'ordre de vos données initiales, mais le tri reste manuel. Cette méthode est cependant moins fastidieuse que la première, mais plus que la troisième.

Troisième méthode

On utilise cette fois-ci la fonction « Recherche » (Ctrl + F) pour rechercher les *primers* dans le jeu de données.

- On sélectionne l'ensemble des séquences à traiter.
- On ouvre la fenêtre de recherche (Ctrl + F)
- On entre la séquence de l'une des deux amorces utilisées en PCR initiale (ici E8F : 5' AGAGTTTGATCATGGCTCAG 3'). Puis on clique sur « rechercher » (Figure 19).

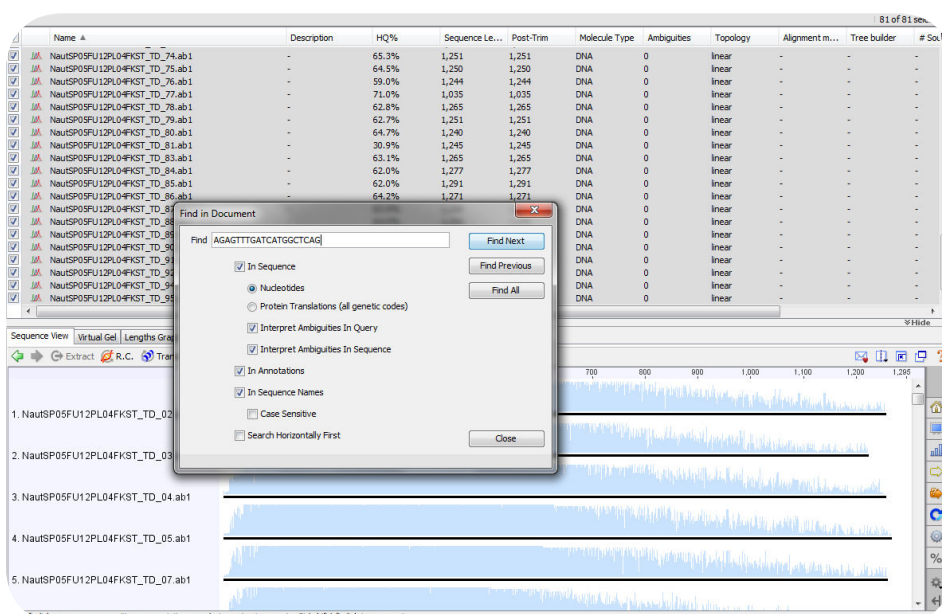


Figure 16 : Capture d'écran de la recherche d'une amorce PCR dans un lot de séquences sous Geneious.

- On observe que l'amorce E8F a bien été trouvée dans la séquence numéro 02 de la banque de clones. Pour la suite, avec un code couleur des nucléotides adéquate, il est facile de noter (sur papier ou dans un fichier texte) quelles séquences sont des « Forwards ».

ATTENTION, avant de noter quoi que ce soit, ne vous amusez pas à réitérer la recherche de l'étape précédente (c'est le coup à faire des erreurs), et vérifiez que vous êtes bien en début de séquences (ici le résultat de recherche est trouvé entre les bases 83 et 102 (ce qui correspond au début de la séquence, puisqu'en amont de l'amorce, on trouve une portion du vecteur). Il vous suffit donc de faire travailler vos yeux (on s'y fait très vite rassurez-vous) et de rechercher le code couleur correspondant à votre amorce. Ne cherchez pas la totalité de l'amorce, ceci ne sert à rien. Pour exemple, quand je cherche l'amorce E8F, je cherche le code couleur **AGAGTTT**. Les séquences 02-03-

04-05-07-08-12 sont des séquences « Forwards ». Par simple présence/absence (de l'amorce E8F), on peut donc en conclure que les séquences 09 et 11 sont des séquences « Reverse ». Il vous suffit donc de placer les bonnes séquences dans le bon dossier.



Figure 17 : Capture d'écran de la visualisation d'une amorce dans un lot de séquences sous Geneious.

- Pour vérifier que vous ne vous êtes pas trompé dans votre tri, un simple alignement des séquences Forwards (uniquement), puis des séquences Reverse (uniquement) vous confirmera votre bon classement ou non. En effet, on observe qu'après recherche de l'amorce E8F dans l'alignement (Figure 20), l'identité (barre verte) sur la zone de l'amorce est verte, ce qui signifie que 100% des séquences présente l'amorce E8F dans le tri réalisé. Dans le cas contraire, où par exemple une séquence « Reverse » se serait glissée parmi les « Forwards », cette barre d'identité serait un mélange de rouge et de marron. La couleur rouge signifiant qu'à une position donnée (par exemple le 50^{ème} nucléotide de l'alignement) est partagée par moins de 30% des séquences. La couleur marron quant à elle indique que la position donnée dans l'alignement est partagé entre 30% et 100% des séquences. A ce moment-là il suffit de rechercher la ou les séquence(s) mal triées pour les remettre dans le bon dossier.



Figure 18 : Capture d'écran montrant la barre d'identité d'un lot de séquences sous Geneious.

« Nettoyage » des séquences

Cette partie du traitement des données est celle qui génère à mon sens le plus de biais lié à l'intervention du manipulateur. Jusqu'ici, le type de fichier utilisé pour analyser les séquences est le format .AB1. Ce format de fichier contient à la fois la séquence texte (composé de A-T-C-G) mais aussi le chromatogramme généré par le séquenceur Sanger. L'inconvénient principal de ce format est qu'il correspond plus ou moins au format brut de sortie du séquenceur. Il contient donc les fins des séquences générées, y compris les nucléotides aberrants dus à un signal trop faible et plus ou moins anarchiques, inhérents aux appareils de séquençage. Il n'y a donc aucun filtrage qualité. A l'inverse, les fichiers .FAS (fasta) fournis par les prestataires de services ont subi un filtrage qualité, mais ces critères ne sont pas fournis par les prestataires de services. Comment donc déterminer quel format est le meilleur ? Le fichier .AB1 avec les informations brutes qui permettent de mieux travailler les séquences, mais sans filtrage qualité ? Ou le format opposé fasta, pré filtré mais sans information ? Quel format donne les séquences les plus longues, tout en gardant une qualité suffisante ? Avec près de 5000 séquences (issues de 57 banques de clones), j'ai effectué un nettoyage manuel de toutes les séquences. A la suite de quoi j'ai comparé statistiquement la longueur des séquences nettoyées manuellement, des séquences filtrées par le prestataire de service via un test de Student. La p value est de $2.2 \cdot 10^{-16}$ avec l'hypothèse H_0 suivante : les jeux de séquences ont une longueur moyenne identique. Le nettoyage manuel permet cependant d'éliminer les nucléotides notés « N » qui sont souvent présents dans les séquences au format fasta.

La première étape du « nettoyage » des séquences est le choix de la méthode d'alignement. En effet, selon vos préférences ou vos objectifs, le résultat peut différer.

La seconde étape consiste à enlever la ou les amorces que vous avez utilisées(s) en PCR, ainsi que la partie du vecteur précédant l'amorce. Pour cela, recherchez l'amorce qui vous intéresse (voir chapitre « tri des séquences »), cliquez sur l'icône « Allow Editing » et supprimez tout ce qui précède le dernier nucléotide de l'amorce. C'est un choix personnel de retirer l'amorce, il vous est possible de la garder, il vous faudra cependant supprimer la portion de séquence correspondant au vecteur de clonage.

La troisième étape (qui n'est pas la plus simple) est de déterminer à quelle position arrêter la séquence (et donc supprimer la fin de la séquence de moins bonne qualité). J'ai pu remarquer que les premières erreurs de séquençages sont dues à des homopolymères, c'est-à-dire des répétitions d'un même nucléotide plusieurs fois (exemple : **AGGGGATCGT** qui présente un homopolymère de 4 G).

C'est l'un des problèmes principaux rencontrés quelle que soit la technologie de séquençage. Il se présente sous la forme suivante dans un jeu de données : Figure 22. Dans un cas comme celui-là, il est intéressant de supprimer la position qui correspond au premier « G », si le nombre de séquences présentant la même erreur à cette position est faible (2 à 3 séquences dans le jeu de données par exemple). Si ce n'est pas le cas, est-ce vraiment une erreur, ou une réalité biologique ? Typiquement, c'est le genre de cas où une erreur peut être générée par l'homme. En cas de doute, je conseille de laisser l'information plutôt que de la supprimer, elle pourra toujours être supprimée ultérieurement.

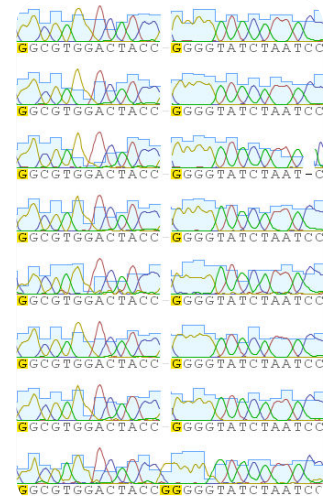


Figure 19 : Capture d'écran d'une zone d'alignement mettant en évidence un polymère de Guanine devant être corrigé.

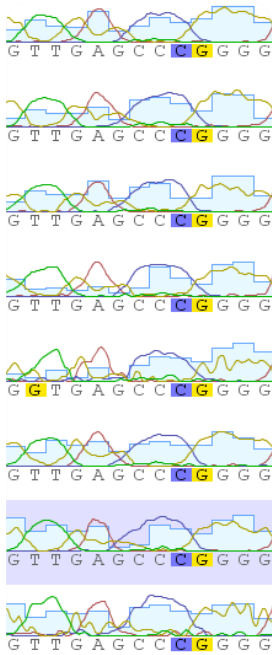


Figure 20 : Capture d'écran d'une zone d'alignement mettant en évidence une erreur d'identification d'un nucléotide à partir d'un chromatogramme.

Comme précisé précédemment, la longueur d'une séquence avoisine généralement les 900 bases. Après suppression des amorces et du vecteur, c'est aux environs de la position 900 (+ ou - 100 bases) dans l'alignement que le chromatogramme présente des erreurs dues à une perte générale du signal. C'est la deuxième source d'erreurs qui peut être introduite par le manipulateur : avoir une séquence longue, mais présentant trop d'erreurs (et qui devient fausse, surtout à la fin). Dans le cas d'étude de diversité, la diversité qui sera calculée ne sera donc pas issue d'une réalité biologique, mais d'erreurs de séquençage.

Une fois les séquences raccourcies, il peut subsister des erreurs de séquençage qui peuvent être corrigées manuellement (notamment en fin de séquence). Les corriger peut être intéressant. Par exemple Figure 23, on observe en deuxième position de la cinquième ligne un G alors que les autres séquences présentent un T. Se pose la question d'une erreur de séquençage. En effet, il s'agit bien d'une erreur de séquençage puisque le chromatogramme affiche deux pics à la même position, un premier correspondant à une guanine, et le second à une thymine. En regardant l'ensemble des nucléotides de cette position, on constate que toutes les autres séquences présentent une thymine, et que la position précédente présente une guanine. Il s'agit ici d'un cas où le chromatogramme du nucléotide qui nous intéresse présente un reste de la position précédente, que l'algorithme d'identification du prestataire de services identifie mal. On peut donc corriger ce « G » en « T ». D'autres types d'erreurs peuvent survenir, et je ne serai pas exhaustif sur ce point, car moi-même je dois passer à côté de certains types d'erreurs. Il faut garder à l'esprit qu'encore une fois, ce type de manipulation peut introduire certaines erreurs liées au manipulateur.

Détection de chimères

L'étape suivante (qui peut être facultative) consiste à chercher des séquences chimériques. C'est-à-dire des séquences qui peuvent paraître propres et intéressantes, mais qui sont en réalité une association artificielle (et donc un biais technique lié à la PCR) de deux morceaux de séquences (ou plus) en une seule. Pour cela, le plus simple est d'utiliser les sites en ligne de détection de chimères comme DECIPHER. Cependant, l'algorithme ressortant dans les forums comme étant le plus performant semble être UCHIME (comme par exemple sur le site drive5.com). L'utilisation la plus simple de cet algorithme se fait via le logiciel Mothur.

Analyse des données avec Mothur

Le graphique suivant reprend les différentes étapes qui seront utilisées, ainsi que les commandes concernées à chaque étape :

- le nombre de fichier(s) à rentrer dans la commande (I=input),
- le nombre de fichier(s) généré(s) par Mothur (O=Output),
- et pour chaque input et output l'extension des fichiers concernés.

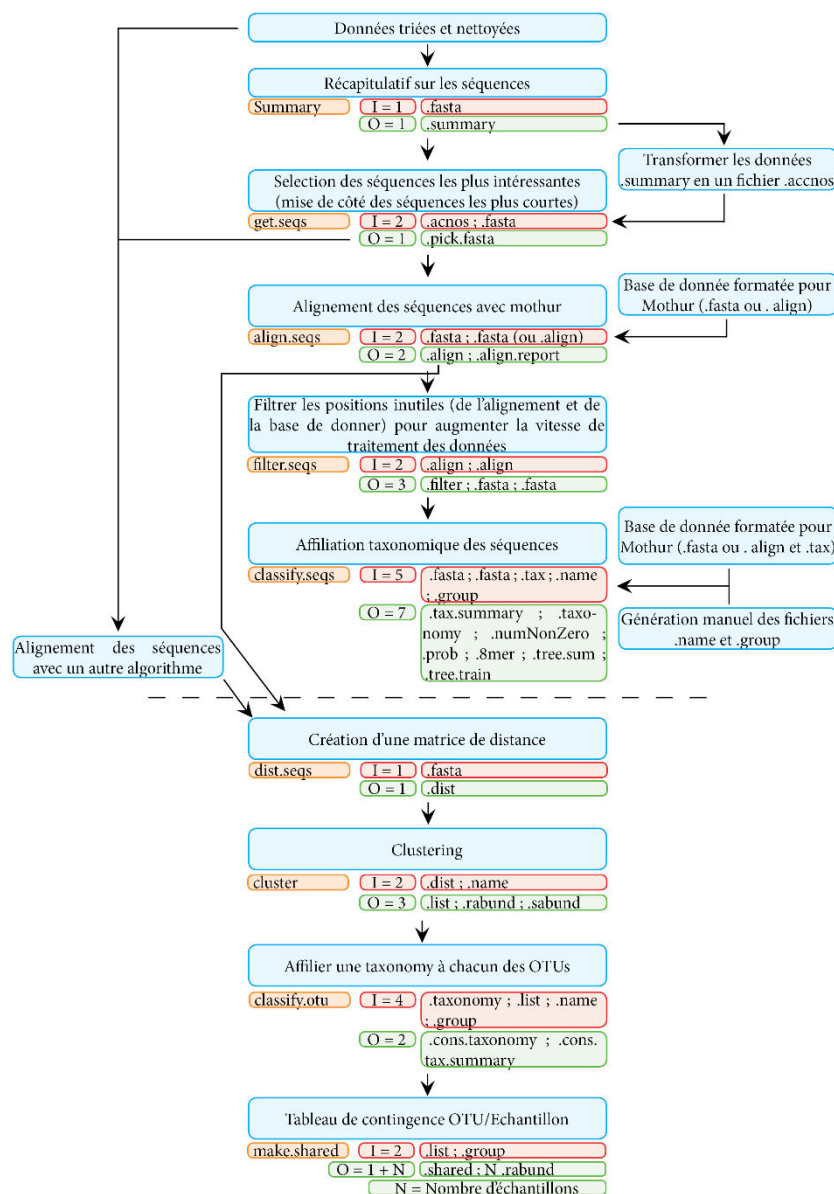


Figure 21 : Schéma récapitulatif du protocole de barcoding sous Mothur.

Il ne sera développé dans cette partie que l'aspect théorique des différentes étapes, et non l'architecture des commandes associées à chacune. Vous pourrez donc appliquer ici les commandes sans réfléchir si vous souhaitez calquer votre traitement des données à celui ci-dessus.

Pour commencer, il faut savoir qu'avec 4 fichiers qui concernent vos séquences, et deux fichiers de bases de données formatées pour Mothur, vous pouvez faire l'ensemble de votre analyse. Les fichiers bases de données peuvent être téléchargés sur le site de Mothur. Si vous choisissez l'option télécharger, il faut être conscient que ces fichiers bases de données sont générés et maintenus par l'équipe de développement de Mothur et ne sont donc pas forcément à jour. Par exemple, au moment où j'écris, la base de données SILVA proposée sur Mothur est la version 128, or Silva va bientôt publier sa prochaine version qui ne sera peut-être pas mise en ligne tout de suite sur le site Mothur. Pour ce qui concerne vos données, il vous faut un fichier .fasta, un fichier .name et un fichier .group et un fichier .align.

Pour le fichier .fasta, il vous suffit, dans Geneious, de sélectionner l'ensemble de vos séquences d'intérêt, puis >File>Export>Selected documents... . Choisissez un nom de fichier sans espace (à remplacer par _ et surtout pas par –) et sans accent, et le format fasta.

Pour obtenir le fichier .name, restez sur Geneious et exportez vos données comme pour le fichier .fasta, mais cette fois changez le format de fichier et choisissez Comma-separated table values (*CSV), puis sélectionnez uniquement la ligne « Name », puis OK. La prochaine étape nécessite d'ouvrir le fichier .CSV généré avec excel, puis de simplement copier la colonne A dans la colonne B, et de supprimer la première ligne. « Sauvegardez sous » ce fichier au format « Texte (séparation : tabulation) (.txt) ». Ceci donne la BD suivante (zoomée pour bien voir).

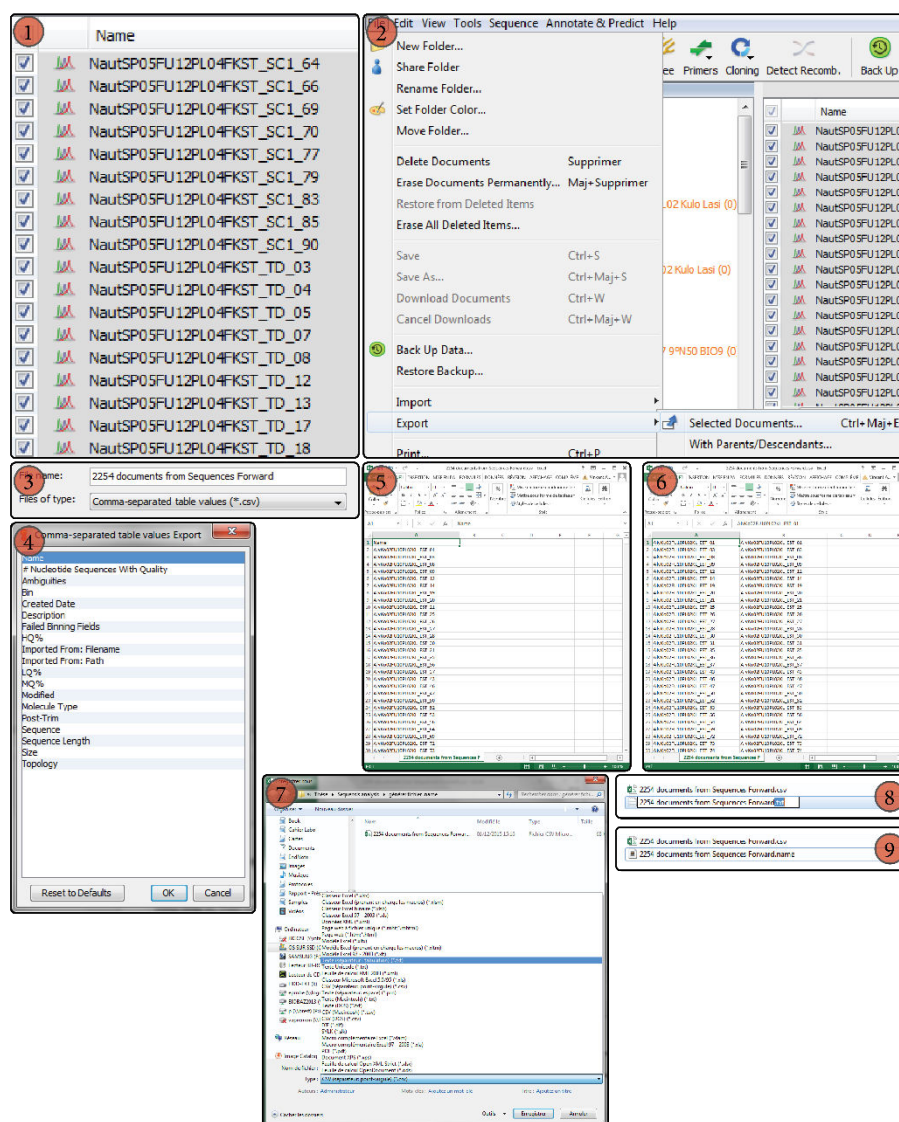


Figure 22 : Ensemble de captures d'écran permettant la création manuelle d'un fichier .name pour Mothur

Afin de créer le fichier .group, utilisez le fichier Comma-separated table values (*.CSV). Dans la colonne B, rentrez le nom de l'échantillon correspondant à votre séquence. Puis « enregistrez sous » le fichier au format « Texte (séparation : tabulation) (.txt) », et renommez le fichier .txt en .group.

Et pour finir, le fichier .align, qui est un fichier .fasta réalisé avec votre méthode d'alignement préférée, toujours avec les mêmes séquences que les fichiers .name et .group. Il faudra cependant changer l'extension de ce fichier de .fasta à .align en sortie d'alignement. Ce fichier n'est à réaliser qu'après la commande `remove.seqs` (voir plus bas).

On peut maintenant attaquer le traitement des données avec Mothur. La première partie va expliciter les commandes, la seconde sera le script direct, où il ne vous faudra plus que lancer des lignes de commandes, en remplaçant WWW par l'adresse de votre dossier de travail, XXX par votre fichier .fasta, YYY par votre fichier .name, ZZZ par votre fichier .group, TTT votre fichier .align (base de données), UUU votre fichier .tax (base de données), tout ceci sans l'extension de fichier. Par exemple, pour la commande `summary.seqs(fasta=XXX.fasta)`, remplacez XXX sans l'extension pour NE PAS obtenir `summary.seqs(fasta=XXX.fasta.fasta)`. Ce fichier XXX.fasta.fasta ne serait pas reconnu par Mothur avec ce script, et votre analyse ne fonctionnera pas.

Summary.seqs

Cette commande permet de récapituler un certain nombre d'informations sur le jeu de données. Un fichier .summary est généré mais ne sera pas utilisé par la suite. L'information intéressante est quant à elle affichée directement sur la fenêtre Mothur comme suit (encadré en rouge) :

Elle nous indique que 2.5 % des séquences ont une longueur inférieure à 628 pb (cette valeur varie selon le jeu de données naturellement). Je vous conseille dans un premier temps d'évincer ces séquences de votre jeu de données car étant trop courtes, elles limiteront la longueur de votre alignement. Par conséquent, vous réaliseriez par la suite un clustering limité par la longueur des séquences de votre jeu de données les plus courtes et ne représentant que 2.5% de votre jeu de donnée. Si jamais ces séquences devaient vous sembler importantes, il vous sera toujours possible de refaire le protocole, sans éliminer ces séquences (une fois le protocole fait une première fois, les fois suivantes vont beaucoup plus vite).

```

mothur.exe - Raccourci
mothur > summary.seqs(fasta=Forward_2254_sequences.fasta)
Unable to open Forward_2254_sequences.fasta. Trying output directory F:\Bioinformatique\Sequences_analysis\Forward_2254_sequences.fasta
Using 1 processors.

```

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1	403	403	0	4	1
2.5%-tile:	1	628	628	0	4	57
25%-tile:	1	817	817	0	5	564
Median:	1	850	850	0	5	1128
75%-tile:	1	875	875	0	6	1691
97.5%-tile:	1	946	946	0	7	2198
Maximum:	1	1191	1191	0	8	2254
Mean:	1	830.776	830.776	0	5.19876	
# of Seqs:						2254

```

Output File Names:
F:\Bioinformatique\Sequences_analysis\Forward_2254_sequences.summary
It took 0 secs to summarize 2254 sequences.
mothur >

```

Figure 23 : Capture d'écran d'un résultat de commande `summary.seqs` sous Mothur.

Screen.seqs

Pour cela, nous allons utiliser la commande `screen.seqs`, qui permet de sélectionner des séquences selon certains critères, et générer un fichier `.accnos` des séquences ne respectant pas les conditions choisies. Ici, nous filtrerons uniquement les séquences en fonction de leur longueur, le filtrage qualité et chimère ayant normalement déjà été fait auparavant. Nous récupérerons ainsi un fichier `.accnos` contenant le nom de toutes les séquences dont la longueur est inférieure à 628 pb.

Remove.seqs

Il faut ensuite supprimer les séquences de votre jeu de données ne correspondant pas à vos attentes (fichier `fasta`), de votre fichier `.name` et de votre fichier `.group`, sinon Mothur cherchera des séquences qui n'existent pas/plus... et affichera des messages d'erreurs, et refusera les commandes que vous lui lancerez.

C'est après cette étape qu'il vous sera nécessaire de réaliser le fichier `.align` en enlevant les séquences contenues dans le fichier dont le nom finit par « `bad.accnos` ».

Classify.seqs

Le but est de créer des OTUs, et de leur assigner une taxonomie. Pour faire ceci, Mothur fonctionne « à l'envers ». C'est-à-dire qu'une fois votre matrice d'OTUs réalisée, Mothur utilise l'affiliation taxonomique de chacune de vos séquences composant une OTU pour déterminer la taxonomie de votre OTU. Il faut donc commencer par affilier une taxonomie à chacune de vos séquences. Pour cela il faut comparer vos séquences à une base de données (que vous avez normalement déjà téléchargée). Il existe plusieurs méthodes d'affiliation taxonomique proposées par Mothur ne nécessitant pas de données alignées : **wang** (par défaut) qui est une méthode probabilistique, **kmn** qui est une méthode par recherche de mot, et **blast**, qui est un blast local comme sur le site du NCBI, sauf que la base de données consultée est celle que vous avez téléchargée (type SILVA).

NB : Cette étape est la plus longue du script en termes de temps de calcul.

Dist.seqs

C'est l'étape avant le clustering. Comme les méthodes de clustering proposées par Mothur sont basées sur les distances génétiques, il est nécessaire que les séquences soient alignées. C'est à cette commande que vos données alignées (`.align`) avec votre algorithme préféré rentrent en jeu. Une simple matrice de distance est générée à partir de ce fichier. Il est donc important que vos données soient parfaitement alignées.

Cluster

Cette étape est celle qui probablement génère le plus de débat dans le monde du barcoding, métabarcoding. En effet, cette étape est celle qui va définir vos OTUs. Mothur propose 6 méthodes de clustering différentes (expliquées ci-après avec le plus de simplicité possible) :

Opticlust (par défaut sur Mothur) et Average neighbor (neighbor) : ce sont deux méthodes étant connues pour n'être ni précises, ni répétables, elles ne seront donc pas décrites car peu utilisées.

Furthest neighbor (furthest) : cette méthode considère comme OTU toutes les séquences ayant moins de 3 % de dissimilarité entre elles.

Nearest neighbor (nearest) : cette méthode considère qu'une séquence appartient à une OTU si celle-ci a moins de 3 % de dissimilarité avec séquence cette OTU en question.

Vsearch agc (agc) : une séquence appartient à une OTU si elle présente moins de 3 % de dissimilarité avec la séquence la plus abondante au sein l'OTU concernée.

Vsearch dgc (dgc) : une séquence appartient à une OTU si elle présente moins de 3 % de dissimilarité avec la séquence représentant le centre (en termes de distance) de l'OTU concernée.

Comme explicité précédemment, il n'existe pas de bon algorithme, il faut tester selon son jeu de données.

Classify.otu

Cette étape est le complément de la fonction `classify.seqs`. Elle permet d'assigner une taxonomie à une OTU en fonction de la taxonomie des séquences qui la compose. C'est une assignation taxonomique consensus qui permet de limiter les assignations à un degré d'arborescence trop haute et qui pourraient se révéler fausses.

Make.shared

Pour pouvoir faire de l'écologie notamment via des calculs d'indice de diversité *alpha* (intra échantillon) ou encore *beta* (inter échantillons), il est nécessaire de disposer d'un tableau de contingence. C'est cette commande qui permet de le réaliser.

Pour continuer sur les calculs d'indices de diversité *alpha* et/ou *beta*, Mothur propose plusieurs commandes:

Rarefaction.single

Cette commande permet de sortir sous forme de tableur, toutes les données permettant de construire des courbes de rarefaction.

Summary.single

Cette commande permet de calculer des indices de diversité alpha via un grand nombre de calculateurs : sobs ; chao ; ace ; jack ; bootstrap ; simpson ; simpson ; shannoneven ; smithwilson ; bergerparker ; shannon ; npshannon ; simpson ; invsimpson ; coverage ; qstat ; boneh ; efron ; shen ; solow ; logseries ; geometric ; bstick. Faire la description de chacun de ces indices ne présente ici aucun intérêt, les plus couramment utilisés sont sobs, chao, simpson, simpson, shannon, shannoneven, d'après la littérature.

Summary.shared

Cette commande permet de calculer des indices de diversité beta sous la forme de tableau à double entrée. Comme pour la commande Summary.single, un grand nombre de calculateurs sont disponibles : sharedsobs ; sharedchao ; sharedace ; anderson ; hamming ; jclass ; jest ; kulczynski ; kulczynskicody ; lennon ; memchi2 ; memchord ; memeuclidean ; mempearson ; ochiai ; sorclass ; soest ; whittaker ; braycurtis ; canberra ; gower ; hellinger ; jabund ; manhattan ; morisita ; odum ; soergel ; sorabund ; spearman ; speciesprofile ; structchi2 ; structchord ; structeuclidean ; structkulczynski ; structpearson ; thetan ; thetayc. Encore une fois, faire le descriptif de chacun de ces calculateurs ne présente pas grand intérêt, le plus communément utilisé est l'indice de Braycurtis dans la littérature.

Tree.shared

Cette commande est plutôt intéressante car elle permet la création d'un dendrogramme représentant la similarité des différents échantillons (relatif au fichier .group) entre eux selon le calculateur choisi (cf commande Summary.shared).

Heatmap.sim

Cette commande est identique à la commande précédente, sauf qu'elle permet de troquer la représentation sous forme de dendrogramme contre une représentation sous forme de heatmap (figure 27).

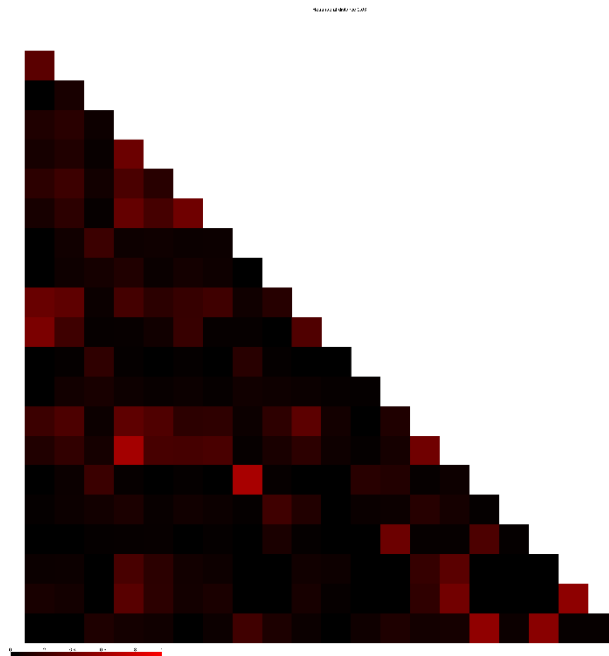


Figure 24 : Représentation de la similarité des échantillons présentés dans le premier article de cette thèse avec l'indice de diversité Bray-Curtis sous la forme d'une heatmap.

Dans le cas où vous souhaiteriez réaliser des analyses « phylogénétiques », il peut être intéressant de récupérer la séquence représentative de chaque OTU. Mothur dispose aussi d'une commande pour cela, la commande :

Get.oturep

Cette commande permet de récupérer les noms et les séquences représentatives de chaque OTU. Deux méthodes sont proposées : la méthode des distances (distance), qui permet de récupérer le nom et la séquence centrale de l'OTU, et la méthode des abondances (abundance), qui permet de récupérer le nom et la séquence la plus abondante de chaque OTU. Dans le cadre de banque de clones, le nombre de séquences est souvent faible, les chances qu'elles soient rigoureusement identiques est faible, le choix de la méthode d'abondance semble donc peu justifiable.

NB : Il peut vous sembler intéressant de récupérer les représentants de chaque OTUs pour les deux méthodes. **Prenez garde** car c'est la même commande qui génère ce type de fichier, les faire en même temps fera que la deuxième commande efface la précédente car les deux fichiers générés porteront le même nom.

Des résultats complets à récupérer

A partir de tous les fichiers générés durant le protocole de la partie précédente, il est possible de générer un fichier tableur contenant toutes les informations de vos rêves. Seulement trois fichiers nous intéressent dont les extensions sont : .shared ; .rep.name ; .cons.taxonomy ; .list.

Ouvrez ces trois fichiers dans votre tableur préféré, et collez leur contenu les uns à côté des autres, à l'exception du fichier .list. Il faut pour ce fichier récupérer la ligne où il est écrit 0.03 dans la première colonne, et effectuer un collage rétro transposé (c'est une option des collages spéciaux) dans le tableur où vous avez réuni l'ensemble des autres fichiers.

Vous voici maintenant avec un fichier tableur dans lequel vous disposez du nom de chaque OTU, sa classification taxonomique, sa séquence représentative, sa répartition dans chaque échantillon (en nombre de séquences) ainsi que toutes les séquences associées à chaque OTU.

Workflow Mothur

Il s'agit ci-dessous (encadré bleu) de l'ensemble des commandes à lancer sous Mothur afin d'obtenir une analyse complète d'étude de diversité microbienne de type barcoding. Comme précisé précédemment, il est nécessaire de remplacer WWW par l'adresse de votre dossier de travail, XXX par votre fichier .fasta, YYY par votre fichier .name, ZZZ par votre fichier .group, TTT votre fichier .align (base de donnée), UUU votre numéro de version de base de données SILVA (ex : 128), tout ceci sans l'extension de fichier. Le texte de couleur verte correspond aux paramètres qu'il vous est possible de changer, tel que la méthode de clustering... (cf. précédemment).

```
set.dir(input=WWW)
set.dir(output=WWW)

summary.seqs(fasta=XXX.fasta)

screen.seqs(fasta=XXX.fasta,minlength=628)

remove.seqs(accnos=XXX.bad.accnos, fasta=XXX.fasta, name=YYY.name, group=ZZZ.group)
```

```
classify.seqs(fasta=XXX.good.filter.fasta,template=silva.nr_vUUU.fasta, taxonomy=silva.nr_vUUU.tax
,name=YYY.name ,group=ZZZ.group)

dist.seqs(fasta=TTT.align, cutoff=0.10)

cluster(column=TTT.dist,name=YYY.pick.name,method=furthest)

classify.otu(taxonomy=XXX.pick.filter.nr_vUUU.wang.taxonomy, list=TTT.fn.list,name=YYY.pick.name,
group=ZZZ.pick.group, label=0.03)

make.shared(list=TTT.fn.list,group=ZZZ.pick.group,label=0.03)

Rarefaction.single(shared=TTT.fn.shared, calc=sobs)

summary.single(shared=TTT.fn.shared,calc=sobs-chao-ace-jack-bootstrap-simpson-even-shannon-even-
smith-wilson-berger-parker-shannon-npshannon-simpson-inv-simpson-coverage-qstat-boneh-efron-shen-
solow-logseries-geometric-bstick)

summary.shared(shared=TTT.fn.shared,calc=braycurtis)

tree.shared(shared=TTT.fn.shared,calc=braycurtis)

heatmap.sim(shared=TTT.fn.shared,calc=braycurtis)
```

Les protocoles de metabarcoding

Pourquoi utiliser différentes architectures informatiques afin de réaliser la même chose ?

L'importance du travail réalisé dans cette partie de metabarcoding a été la rédaction de différents scripts réalisant en une seule commande toute une analyse de metabarcoding avec l'outil concerné.

L'ordinateur de bureau

Si vous avez suivi les conseils donnés en tout fin de chapitre précédent, vous disposez d'une bonne machine capable d'effectuer le protocole de barcoding sur banque de clones sans trop de difficultés. Cependant, les ressources d'une machine de ce type pour réaliser une analyse de metabarcoding complète demanderont toutes les ressources de la machine, pouvant aller jusqu'à vous empêcher d'utiliser des logiciels quotidiens par manque de ressources (ou alors il faut être très très très patient). Alors pourquoi faire un script pour ce type d'architecture puisqu'elle ne sert à rien ? En fait si. Il vous est tout à fait possible de réaliser une analyse partielle de metabarcoding, sur une partie restreinte de votre jeu de données total. Quel intérêt ? Tout simplement parce que ces scripts "vous poseront des questions" à chaque étape importante, concernant des paramètres nécessaires à la suite du traitement, mais venant juste d'être produit. Ceci vous permettra de finaliser une première fois l'une des analyses. Pour passer à l'étape suivante.

Un calculateur type DATARMOR

Ceci est l'étape suivante : utiliser un calculateur. Ici l'exemple est pris avec DATARMOR, calculateur Ifremer, mais le principe est le même sur d'autres types de calculateur. En tant qu'utilisateur, vous disposez de trois espaces qui sont à vous.

Le premier est ce que l'on appelle "home", c'est un espace de stockage sauvegardé où vous pouvez stocker les fichiers que vous désirez, ils sont sauvegardés. Pour DATARMOR, cet espace fait 50 Go par utilisateur.

Votre deuxième espace est un espace temporaire nommé "work". Cet espace est d'une capacité plus importante (500 Go par utilisateur sur DATARMOR), mais les fichiers sont supprimés au bout de 15 jours.

Et enfin, un espace nommé "scratch" qui a la même fonction qu'une poubelle. Tout ce qui y est mis est immédiatement supprimé et non récupérable (autant faire attention à ce dernier).

Le gros avantage de ce type de machine, qu'il faut voir un petit peu comme un ordinateur de bureau fonctionnant sous Linux (en ligne de commande), c'est qu'il dispose d'une puissance bien supérieure à une simple machine. Pour DATARMOR, il est possible, le temps d'un calcul, d'avoir 240 cœurs (un ordinateur normal dispose de 4 à 8 cœurs), et 1 To (1000 Go) de RAM. L'inconvénient, c'est que des gros calculs doivent lui être soumis (en ligne de commande). C'est-à-dire que vous lui soumettez un script, et celui-ci s'exécute en arrière-plan. Il ne peut donc pas vous poser de question comme pour l'ordinateur de bureau une fois le script lancé. Tous les paramètres doivent être connus dès le début (d'où la partie amont avec l'ordinateur de bureau pour préparer le travail).

Galaxy

Galaxy est un superbe outil pour tous les allergiques de la ligne de commande. En effet, il s'agit d'une interface web dans laquelle il est possible de faire de la bioinformatique sous forme de « presse bouton ». Aucune ligne de commande à faire. Le rêve me direz-vous. En fait, pas tant que ça. Une fois que l'on s'est résigné à apprendre la ligne de commande, Galaxy peut devenir un peu frustrant. En effet, vous êtes dépendant d'un service dédié à Galaxy. C'est ce service qui met en place les outils dont vous avez besoin (Mothur et Qiime disparaissent des Galaxy, car trop lourds à maintenir et difficiles d'utilisation pour les néophytes), qui règle les problèmes, ou qui corrige les bugs, et fournit les différentes ressources informatiques à chaque logiciel. Pourquoi serait-ce donc un inconvénient ? Principalement car les ressources de Galaxy sont limitées, alors qu'il utilise lui aussi le calculateur. Vous ne pouvez donc pas choisir si vos calculs se feront vite, très vite, ou très très vite, comme sur le calculateur. De plus, Galaxy n'a le droit qu'à une partie de la puissance du calculateur. Une fois que trop de monde sollicite la plateforme Galaxy (quelle qu'elle soit) une file d'attente se met en place, et elle peut être longue. Alors que le calculateur beaucoup moins.

Vous comprendrez donc que votre ordinateur de bureau, bien que pouvant être performant, ne peut pas être votre seul outil de bioinformatique, il n'est qu'un intermédiaire, un lieu de test, et qu'il faudra ensuite passer au calculateur, ou à Galaxy. Ou tout simplement directement aller sur Galaxy, avec les inconvénients qui y sont associés.

Les scripts

L'idée derrière chacun des scripts, qui je le répète, ne fonctionnent que sur Linux, que ce soit sur l'ordinateur, ou sur un calculateur, est toujours la même. L'utilisateur lance un premier script (dans un invité de commande) qui crée tous les dossiers et fichiers nécessaires au bon déroulement du second script, celui d'analyse des données. De manière concrète, le premier script crée un dossier sur le « bureau » de

l'utilisateur (si celui-ci n'existe pas, il est lui aussi créé) appelé soit Mothur, Qiime, ou FROGS (selon l'outil que vous voulez utiliser). Dans ce dossier, deux nouveaux dossiers sont créés : le premier s'appelle « Input », et le second « Output ». Puis dans le dossier « Input », un fichier « propriete.txt » est créé. Celui-ci contient déjà les informations qui vous seront nécessaires à l'analyse. Il vous suffit par la suite de rentrer les paramètres que vous connaissez dans le fichier « propriete.txt ». Dans le cas d'une analyse sur ordinateur, si certains paramètres sont manquants, parce que vous ne les connaissez pas encore, ils vous seront demandés ultérieurement. Dans le cas d'une analyse sur calculateur, ils devront tous être renseignés. Vos données à analyser devront par la suite être mises dans le dossier « Input ». La dernière étape est de lancer le second script, en précisant un nom de projet. Les données à récupérer seront ensuite mises automatiquement dans le dossier « Output » précédemment créé. Pour ce qui est des paramètres qui auraient pu vous manquer, ils sont directement rajoutés dans le fichier « proprietes.txt », qui pourra directement être utilisé avec le script pour le calculateur. L'ensemble de ces informations est représenté graphiquement dans la figure 28.

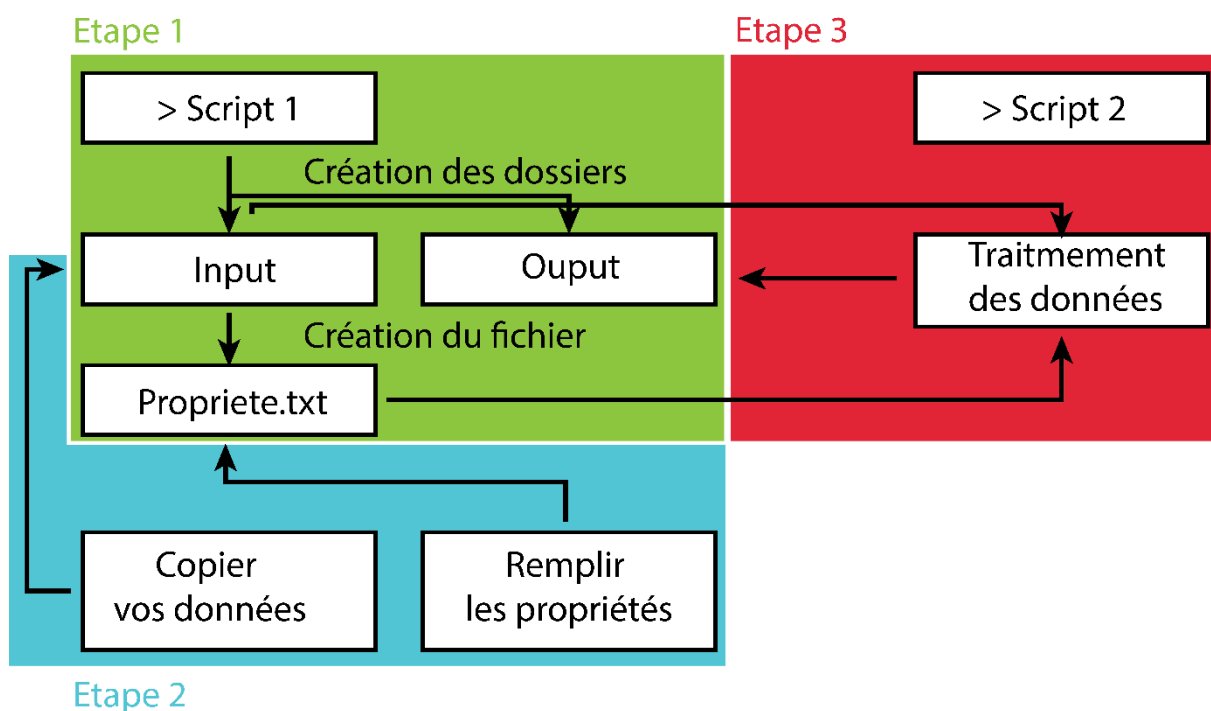


Figure 25 : Représentation schématique du principe global des scripts générés durant cette thèse.

Décrire de manière exhaustive chaque ligne de commande des différents scripts ne présente aucun intérêt pour comprendre ce qui se déroule lors du traitement des données. Il sera donc fait une description des différentes grandes étapes avec les trois outils que sont Mothur, Qiime et FROGS. Bien évidemment, il existe

des workflows pour chacun de ces outils sur internet. Comme il avait été précisé précédemment, aucun nouvel algorithme ou outils n'a été inventé ici (je n'ai pas réinventé la poudre), le travail effectué a été de le rendre plus accessible à quelqu'un qui ne sait pas utiliser de la ligne de commande, ou qui présente une peur phobique de l'informatique.

Lors de la réception des données de séquençages provenant du prestataire de services, les fichiers de données peuvent prendre plusieurs formes, qui sont résumées dans la figure 29 ci-dessous (seulement trois exemples donnés) :

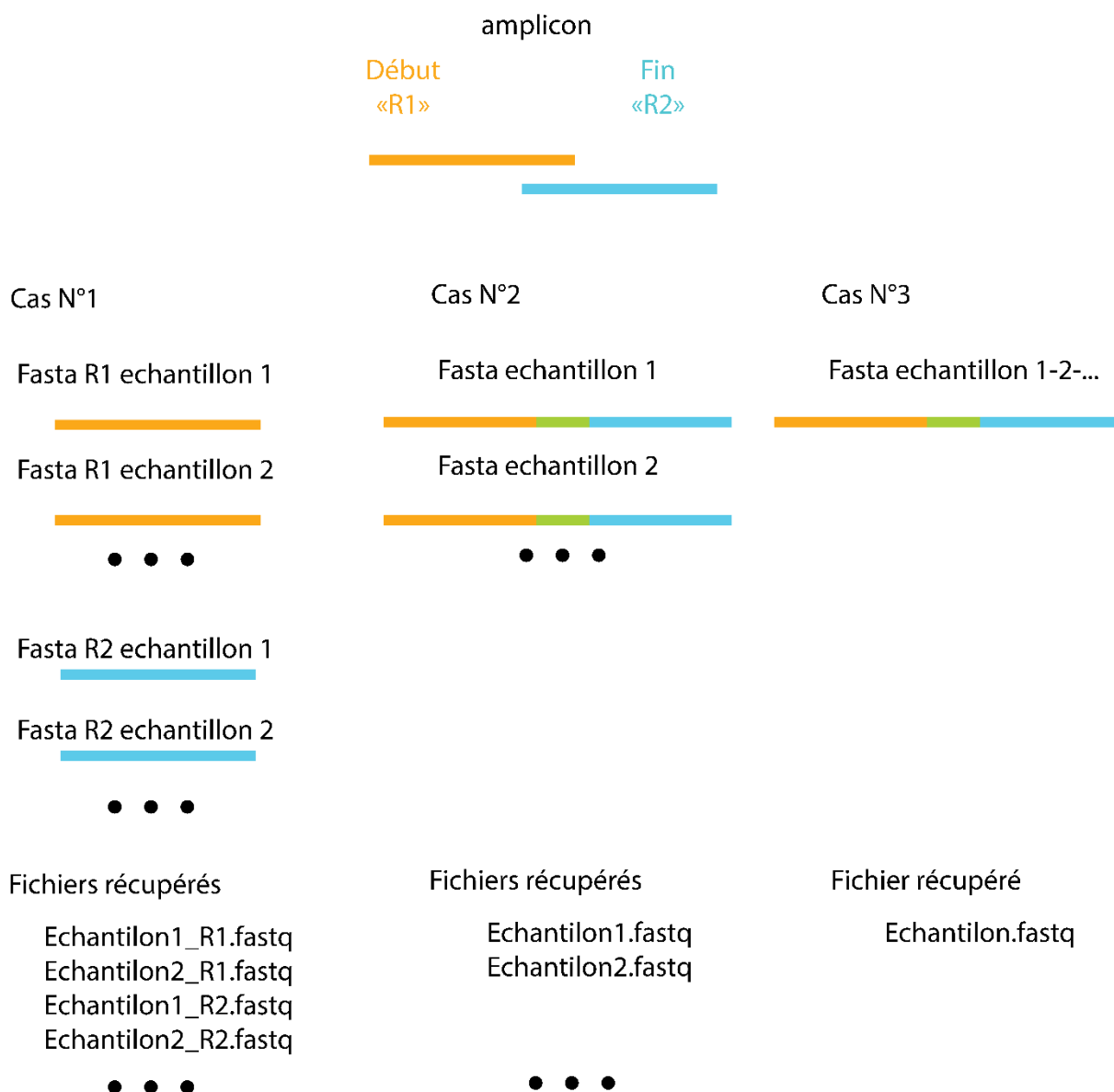


Figure 26 : Représentation schématique d'exemples de données pouvant être reçues d'un prestataire de séquençage haut débit.

Dans le premier cas, les données ressemblent à ce qui se rapproche le plus des données brutes sorties du séquenceur. Nous disposons pour chaque échantillon des données R1 et R2, il faudra donc assembler tous les amplicons, mais la différenciation des différents échantillons est réalisée.

Dans le second cas, les données reçues sont séparées par échantillon, et chaque séquence a été assemblée.

Dans le troisième cas, chaque séquence a été assemblée, mais la différenciation entre les différents échantillons n'est pas faite, il sera donc nécessaire de faire une étape supplémentaire.

Vous comprenez donc que dès le départ, il est possible de recevoir différents types de données formatées de différentes manières. Des étapes de préanalyses sont donc nécessaires pour arriver à un format qui s'approche plus facilement du cas numéro 2.

Durant ce travail de thèse, il a volontairement été demandé au prestataire de services un format de données correspondant au cas numéro 1, afin de disposer des données les plus brutes possible, et ainsi avoir la main sur chaque étape de prétraitement et traitement des données. Par conséquent, au moment de la rédaction du manuscrit, seul le format de données correspondant au cas numéro 1 est reconnu et validé par les différents scripts.

Bien que les différents outils Mothur, Qiime et FROGS permettent de parvenir au même type de données finales, la « philosophie » de traitement des données est différente. C'est pourquoi chaque « philosophie » sera décrite pour chaque outil.

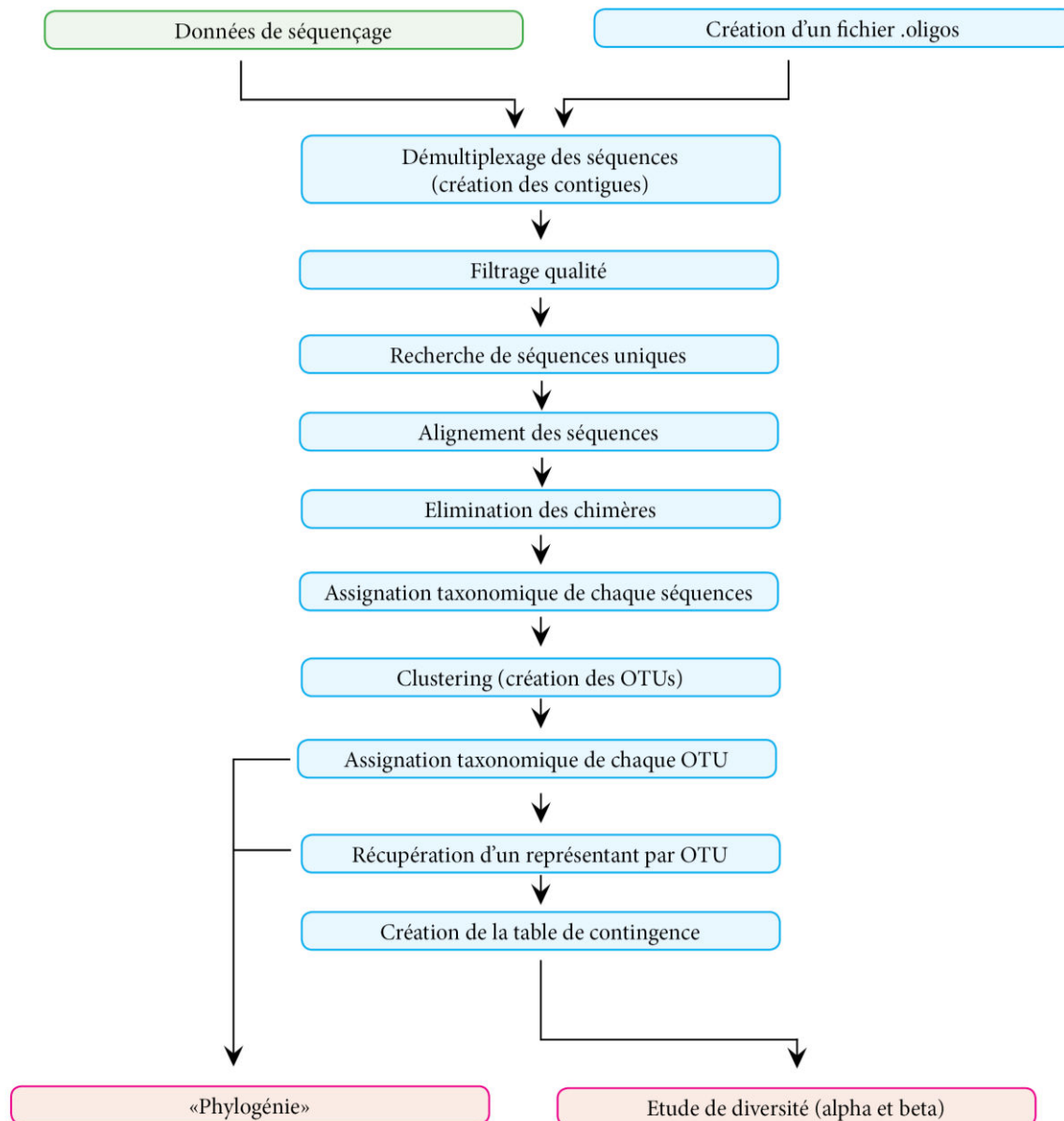


Figure 27 : Schéma simplifié du protocole de traitement de données de type metabarcoding sous Mothur.

Mothur est un outil plutôt gourmand en ressources informatiques pour plusieurs étapes du traitement des données. Après démultiplexage et filtrage des données, Mothur propose de réaliser une étape visant à chercher les séquences dites uniques. Cela consiste à rechercher dans le jeu de données, tous échantillons confondus, des séquences rigoureusement identiques au nucléotide près. Pourquoi une telle étape ? Afin d'éviter dans de futures étapes de répéter une même tâche plusieurs fois inutilement. Prenons l'exemple qui

suit directement la recherche de séquences identiques : l'alignement des séquences. Pour nous humains, il est plutôt simple d'appréhender l'alignement de deux séquences identiques : nous n'avons pas besoin de chercher à les aligner entre elles, elles sont identiques. L'ordinateur ne comprend pas intuitivement ce type de logique, il effectuera un processus d'alignement normal même si les deux séquences sont identiques. Oui, mais puisqu'elles sont identiques, l'alignement ne devrait pas prendre beaucoup de temps ? En effet, individuellement, le temps perdu à effectuer cette tâche est faible, mais rapporté à des millions de séquences, cela peut faire beaucoup de temps. D'où une recherche de séquence unique. Cette perte de temps est ainsi éliminée. D'un point de vue concret, cette étape de recherche de séquence unique est utile principalement pour l'étape d'alignement, de recherche des chimères, et la classification taxonomique. Ce sont les trois étapes les plus gourmandes en ressources informatiques et en temps (plusieurs heures). D'où l'importance d'essayer de réduire les inconvénients précédemment cités avec l'utilisation de séquences uniques.

L'étape d'alignement des séquences est réalisée contre une base de données, en l'occurrence la base de données SILVA. Cette méthode d'alignement est « rapide » en comparaison de ce qui peut être réalisé avec d'autres méthodes telles que MUSCLE, MAFFT, CLUSTALW, CLUSTALX ... En effet, des méthodes comme celles-ci présentent un inconvénient majeur dans le cas d'alignement avec de nombreuses séquences : le temps de calcul augmente de manière exponentielle. En effet, ces méthodes procèdent comme suit : nous prenons les deux premières séquences du fichier qui nous a été soumis, et nous les alignons. Nous prenons ensuite la troisième séquence, que nous alignons avec les deux premières, tout en vérifiant que nous ne pouvons pas réaliser un meilleur alignement global. Puis nous prenons la quatrième séquence.... Plus le nombre de séquences augmente, plus les calculs sont longs à effectuer, d'où un effet exponentiel du temps de calcul en fonction du nombre de séquences à aligner. A l'inverse, aligner des séquences contre une base de données est « plus simple » car l'alignement global optimal est déjà réalisé, surtout avec la base de données SILVA. Donc l'alignement se réalise simplement en cherchant à aligner les séquences soumises contre un alignement déjà fait. Et cela prend moins de temps.

L'étape suivante intéressante à décrire pour comprendre la philosophie Mothur est l'assignation taxonomique de chaque séquence du jeu de données. Pourquoi effectuer cette lourde tâche ? L'intérêt n'est pas immédiat, il est lié notamment à l'affiliation taxonomique des OTUs. Plutôt que d'effectuer l'affiliation taxonomique d'une OTU en fonction d'une séquence (son représentant ou séquence type), Mothur réalise une affiliation taxonomique des OTUs qui est consensus à toutes les séquences qui la composent. Cela permet de limiter les erreurs d'assignation taxonomique en tant que telles, et les doutes quant au niveau d'assignation taxonomique dans l'arbre du vivant.

Le clustering sur Mothur est une étape plutôt limitée d'un point de vue nombre de méthodes proposées. Il n'en existe que 6 en 2017, et elles sont toutes basées sur des méthodes de distance. En plus de la méthode utilisée, il faut donc avoir à l'esprit que le résultat peut aussi être lié au biais d'alignement (effectué ou non dans Mothur pour les plus expérimentés), puisque c'est sur ces résultats que sont basés les calculs de distances.

Il n'est pas fait état ici des méthodes d'analyses phylogénétiques ou de diversité (alpha et beta) car il pourrait en être discuté durant de nombreuses pages d'un point de vue choix de méthode... Or, les différents outils que sont Mothur, Qiime ou FROGS proposent chacun une panoplie (ou non) d'outils permettant d'analyser d'un point de vue scientifique, et non plus technique, les résultats de séquençage obtenus. Ils sortent donc de la thématique abordée ici, au même titre que l'ensemble des biais pré séquençage évoqués précédemment.

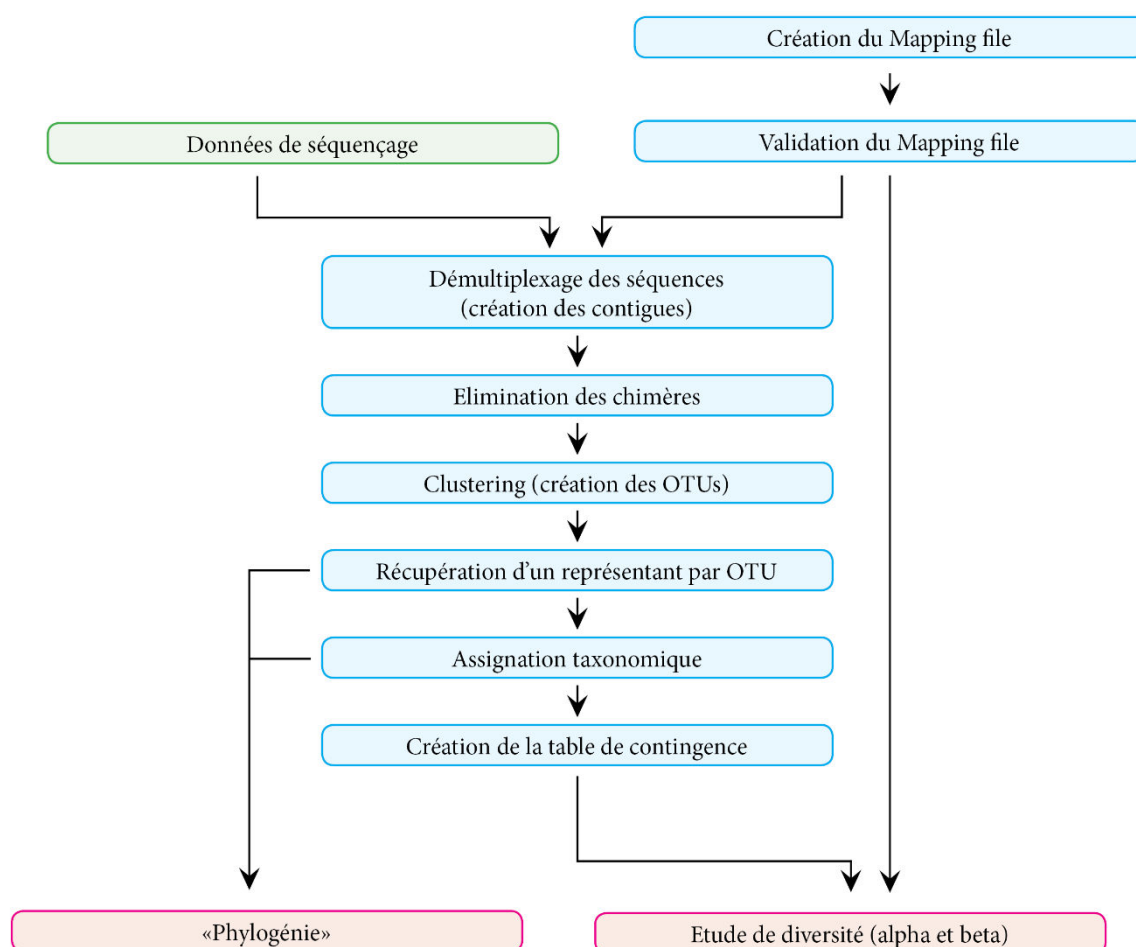


Figure 28 : Schéma simplifié du protocole de traitement de données de type metabarcoding sous Qiime.

La philosophie de l'outil Qiime est plus simple que celle de Mothur : proposer le plus d'outils possible compatibles entre eux à un utilisateur souhaitant analyser des données. Qiime se soucie un peu moins des différentes possibilités qu'il propose, il cherche simplement à en proposer le plus possible (c'est un avis personnel, partagé par de nombreux utilisateurs lorsque l'on lit durant de nombreuses heures les forums de bioinformatique). Il n'en reste pas moins que cela fait de Qiime un outil plus souple que Mothur, mais par conséquent plus dur à prendre en main d'un point de vue scientifique, puisque les choix sont tellement nombreux qu'un néophyte en traitement de metabarcoding peut vite prendre peur. Cependant, le Workflow proposé ici se veut plus simple que celui proposé pour Mothur. En effet, il n'y a pas d'étape de recherche de séquences uniques, d'alignement des séquences, ou de double assignation taxonomique. Comparée à Mothur, l'étape principale ici est celle de clustering, puisque Qiime propose 16 méthodes de clustering

(contre 6 pour Mothur), dont certaines ne sont pas basées sur des méthodes de matrice de distance (comme SWARM, UCLUST, USEARCH ou SUMACLUSt). Au-delà de l'aspect *philosophie*, l'utilisation de Mothur et Qiime dans cette thèse vise aussi à comparer une étape très importante dans l'analyse de données de metarbarcoding : le clustering par méthodes de matrice de distance des autres méthodes de clustering sur notre jeu de données.

FROGS

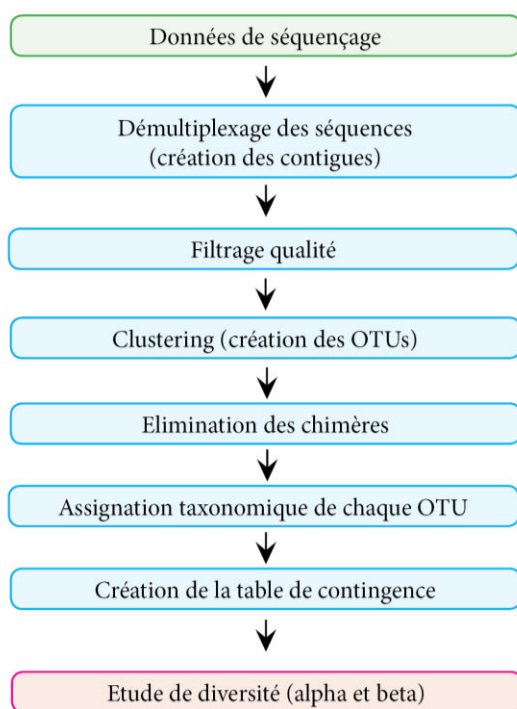


Figure 29 : Schéma simplifié du protocole de traitement de données de type metabarcoding sous FROGS.

La philosophie liée au workflow FROGS est la plus simple des trois outils d'un point de vue de l'utilisateur : proposer une analyse complète de données, qui soit le plus rapide possible, la moins gourmande en ressource informatique possible, et avec un minimum d'options possible. Le gain de temps principal de ce workflow est lié au fait que l'élimination des chimères est effectuée après le clustering. Ainsi, au lieu de chercher si chaque séquence est chimérique ou non, cette étape est effectuée sur les OTUs. Normalement, si tout se passe bien, le traitement des données donne moins d'OTUs qu'il n'y a de séquences. D'où un gain de temps dans la recherche des chimères. Le clustering prend cependant un peu plus de temps (toujours moins long que si la recherche des chimères avait été fait avant le clustering). Si le point fort de FROGS est la faible quantité

d'options proposées, cela représente aussi son principal défaut. Le clustering ne peut être effectué que par la méthode SWARM, et l'affiliation taxonomique des séquences uniquement réalisée par BLAST contre une base de données. Si l'utilisateur n'est pas d'accord avec ces choix méthodologiques, il n'a qu'une seule solution, utiliser un autre outil.

1^{er} article : First description of
Chorocaris (Rimicaris) chacei
symbiont : high morphological
and genetic similarity with
Rimicaris exoculata (MAR)

Introduction à la problématique

A l'heure actuelle, peu de choses sont connues sur l'éventualité d'une symbiose chez la crevette *Rimicaris chacei*. Cette crevette endémique des sources hydrothermales de la ride médio-atlantique (MAR) vit en périphérie des agrégats de *Rimicaris exoculata* (voir partie introduction bibliographique). On l'observe en groupes composés de peu d'individus contrairement à l'autre espèce, et elle est relativement peu mobile. Des études préliminaires ont été menées en 1993 et ont montré la présence de communautés microbiennes dans le céphalothorax mais de moindre mesure que celles observées chez *Rimicaris exoculata*. Ainsi, le peu que l'on connaît de cette espèce amène plusieurs questions. La première est l'identification taxonomique et le positionnement dans l'arbre du vivant de ces microorganismes observés dans la cavité céphalothoracique. La seconde est de savoir si les bactéries observées et décrites en 1993 (sur un seul site, Snake Pit (Casanova *et al.*, 1993, Segonzac *et al.*, 1993)) sont présentes sur tous les sites où *Rimicaris chacei* est rencontrée. La troisième est de savoir s'il y a bien une symbiose, et si oui, est-elle comparable (et donc proche) à celle retrouvée chez *Rimicaris exoculata*, que ce soit d'un point de vue morphologique ou phylogénétique, les deux espèces étant répertoriées sur des sites similaires, et dans des niches écologiques très proches. De plus, l'observation d'un comportement nécrophage chez cette crevette (Segonzac *et al.*, 1993) laisse penser à un mode de nutrition compensatoire à la symbiose chez cette espèce (nutrition mixte, nécrophagie). La question d'une présence de communautés microbiennes dans le tractus digestif se pose donc.

Les approches méthodologiques de cet article sont basées sur des observations macroscopiques, de la microscopie et de la biologie moléculaire. La macroscopie, via des observations vidéo a pour but d'observer le comportement des crevettes *Rimicaris chacei* dans leur environnement. Tandis que des observations d'organes (pré disséqués à bord du bateau durant les campagnes) via loupe binoculaire, ont pour but d'observer d'éventuels signes de colonisation microbienne et de permettre la sélection d'individus pouvant présenter un stade avancé de colonisation.

Les observations en microscopie électronique à balayage (MEB), microscopie électronique à transmission (MET) et observations de coupes histologiques ont pour but de cartographier la répartition des bactéries sur les différents tissus bactériophores et dans le tractus digestif, ainsi que de caractériser les populations microbiennes présentes d'un point de vue morphologique. Ces approches permettront également de décrire et analyser les dépôts minéraux associés aux colonisations microbiennes.

Des microanalyses par rayons X ont été effectuées afin de caractériser d'un point de vue qualitatif la composition atomique de cet encroûtement minéral.

D'un point de vue biologie moléculaire, l'étude du gène codant pour la petite sous-unité du ribosome (16S) a été réalisée par banque de clones afin de : 1) déterminer la diversité bactérienne de nos échantillons par

identification taxonomique des microorganismes identifiés, 2) comparer les populations microbiennes en fonction des sites hydrothermaux étudiés (Lucky Strike, Rainbow, TAG et Snake Pit) 3) rechercher une différence de populations microbiennes entre les différents organes étudiés (lame branchiostège, scaphognathite, tube digestif) 4) réaliser une comparaison « phylogénétique » des OTUs avec celles connues dans la littérature et en particulier celles associées à *Rimicaris exoculata*.

Enfin, l'utilisation de microscope à épifluorescence et de la méthodologie de fluorescence par hybridation *In situ* (FISH) a été réalisée afin d'essayer de corréler les observations microscopiques (notamment la morphologie des microorganismes rencontrés) avec les données moléculaires (taxonomie).

<http://mc.manuscriptcentral.com/fems>

Halfway through symbiosis? High similarity between bacterial communities of the vent shrimp *Rimicaris chacei* (Williams and Rona 1986) and those of its symbiotic counterpart *Rimicaris exoculata* (Williams and Rona 1986).

Journal:	<i>FEMS Microbiology Ecology</i>
Manuscript ID	Draft
Manuscript Type:	Research article
Date Submitted by the Author:	n/a
Complete List of Authors:	Apremont, Vincent; Ifremer, REM Cambon-Bonavita, Marie-Anne; Ifremer, REM Cueff-Gauchard, Valérie; Ifremer, REM David, François; Ifremer, REM Pradillon, Florence; Ifremer, REM Corbari, Laure; Museum national d'Histoire Naturelle, UMR 7138- Equipe Espèce et spéciation Zbinden, Magali; Université Pierre et Marie Curie, UMR CNRS MNHN 7208
Keywords:	Hydrothermal vents, Deep-sea, <i>Rimicaris chacei</i> , Symbiosis, Mid-Atlantic Ridge, <i>Rimicaris exoculata</i>

SCHOLARONE™
Manuscripts

1 1 Halfway through symbiosis? High similarity between bacterial communities of the vent
2 2 shrimp *Rimicaris chacei* (Williams and Rona 1986) and those of its symbiotic
3 3 counterpart *Rimicaris exoculata* (Williams and Rona 1986).
4

5 5 Apremont Vincent ^{1a,b,c-4}, Cambon-Bonavita Marie-Anne ^{1a,b,c}, Cueff-Gauchard Valérie ^{1a,b,c},
6 6 François David ^{1a,b,c}, Florence Pradillon², Laure Corbari³ and Zbinden Magali ^{4*}.
7

8 8 ^{1a} Ifremer, Centre Bretagne, ZI de la Pointe du Diable, Laboratoire de Microbiologie des
9 9 Environnements Extrêmes, REM/EEP/LM2E, UMR 6197 Ifremer-CNRS-UBO, CS 10070,
10 10 29280 Plouzané, France
11

11 11 ^{1b} Université de Brest, Laboratoire de Microbiologie des Environnements Extrêmes, UMR
12 12 6197 Ifremer-CNRS-UBO, Technopôle Iroise, 4 place Nicolas Copernic, 29280 Plouzané,
13 13 France
14

14 14 ^{1c} CNRS, Laboratoire de Microbiologie des Environnements Extrêmes, UMR 6197 Ifremer-
15 15 CNRS-UBO, Technopôle Iroise, 4 place Nicolas Copernic, 29280 Plouzané, France.
16

16 16 ²Ifremer centre Bretagne, ZI de la Pointe du Diable, Laboratoire Environnement Profond,
17 17 REM/EEP/LEP, CS 10070, 29280 Plouzané, France
18

18 18 ³ Muséum national d'Histoire naturelle, Institut de Systématique, Évolution, Biodiversité
19 19 ISYEB - UMR 7205 – CNRS, MNHN, UPMC, EPHE, 57 rue Cuvier, CP 26, F-75005, Paris,
20 20 France
21

21 21 ⁴ Sorbonne Universités, Univ Paris 06, UMR CNRS MNHN 7208 Biologie des Organismes
22 22 Aquatiques et Ecosystèmes (BOREA), Equipe Adaptation aux Milieux Extrêmes, 7 Quai St
23 23 Bernard, 75005 Paris, France
24

* Corresponding author: Magali Zbinden. Email : agali.zbinden@upmc.fr, tel: 00 33 1 44 27 37 93, fax: 00 33 1 44 27 58 01

Keywords: hydrothermal vents; deep sea; *Rimicaris chacei*; symbiosis; Mid-Atlantic Ridge, *Rimicaris exoculata*

Abstract

Rimicaris chacei Williams & Rona 1986, formerly known as *Chorocaris chacei*, is a caridean shrimp living in deep-sea hydrothermal ecosystems. This shrimp is endemic of the Mid Atlantic Ridge (MAR) and live at the periphery of *R. exoculata* Williams and Rona 1986 aggregates. Unlike *R. exoculata*, *R. chacei* lives in small groups of several individuals, rather motionless. Although devoid of the characteristic hypertrophied cephalothorax of *R. exoculata* that harbors the ectosymbionts, a bacterial community has also been observed in the cephalothorax of *R. chacei*. Previous data on morphology, behavior and isotopic values indicated a diet based on a combination of feeding on its epibiotic bacteria and scavenging or occasional predation. In this study, we propose a pluridisciplinary approach to describe the morphology, distribution and phylogeny of the microbial communities associated to *R. chacei* from several MAR sites. Microbial communities have been observed on mouthparts, branchiostegites and digestive tracts. The bacterial distribution pattern reminds, in less developed, that observed on *R. exoculata*. The main lineages are affiliated to *Epsilon* and *Gammaproteobacteria* in the cephalothorax and to *Deferribacteres*, *Mollicutes*, *Epsilon* and *Gammaproteobacteria* in the digestive tract. Hypothesis about the role of these communities and the symbiosis evolution among *Rimicaris* species are discussed.

Introduction

Besides deep abyssal areas showing very low biomass but highly diversified communities

(Desbruyères *et al.* 2001; Baker *et al.* 2010), deep sea hydrothermal vent ecosystems are spots of high endemism harboring high biomasses, but a low biodiversity. Though mainly fueled by chemosynthesis (Jannasch 1985), most trophic behaviors are present in these ecosystems, such as grazing (Van Dover and Fry 1994), filter-feeding (Fiala-Médioni *et al.* 1986), predation and scavenging (Segonzac *et al.* 1993). However, the highest biomass encountered around the vent emissions (for review see Dubilier *et al.* 2008) rely on trophic symbioses, based on chemosynthetic microorganisms (Jannasch 1985). Among the best known examples are essentially endosymbioses, such as the giant tubeworm *Riftia pachyptila* Jones 1981 from the East Pacific Rise (EPR) (Corliss *et al.* 1979; Cavanaugh *et al.* 1981; Felbeck 1981; Felbeck *et al.* 1981) or the *Bathymodiolus* mussels distributed in almost all chemosynthetic based ecosystems (Belkin *et al.* 1986; Childress *et al.* 1986; Duperron *et al.* 2005; Fujiwara *et al.* 2007; Duperron *et al.* 2008; Southward 2008; Duperron *et al.* 2009). Among hydrothermal crustaceans, symbioses are only reported in few species and as epibioses (Dubilier *et al.* 2008; Goffredi 2010; Tsuchida *et al.* 2001; Ponsard *et al.* 2013). The best-studied example is the shrimp *Rimicaris exoculata* from the Mid Atlantic Ridge (MAR) (see references below). The *Rimicaris* genus now comprises 9 species as it has recently been revised by invalidating the genus *Chorocaris* and transferring or resurrecting the species of this genus into the *Rimicaris* one (Vereshchaka *et al.* 2015). Among the *Rimicaris* genus, some species (*R. exoculata*, *R. kairei* Watabe and Hashimoto 2002 and *R. hybisae* Nye, Copley and Plouviez 2011) present inflated branchiosegites, a carapace ovate-oblong, broader than abdomen and about twice as wide as deep in greatest dimensions (i.e. hypertrophied cephalothorax), with exopodites of the second maxilla (scaphognathites) and first maxillipeds greatly expanded anteriorly and conspicuously setose on dorsal and ventral faces. These characters have led to a great confusion in systematics because some *Rimicaris* species (former species belonging to the genus *Chorocaris*) do not harbor these morphological features.

To date, among the 9 species of the genus *Rimicaris*, *R. exoculata* is the most extensively studied, in particular with regard to its dual symbiosis. The first one is located in the gut (Fig. 1A). A dense bacterial population of long thin (0.2 μm x 15 μm) single-celled bacteria is present between the microvilli of the midgut epithelium (Durand *et al.* 2010), whereas rods and cocci are also observed within the gut content (Zbinden and Cambon-Bonavita 2003). This bacterial diversity is dominated by 4 major groups: Deferribacteres, Mollicutes, *Epsilonproteobacteria*, and to a lesser extent *Gammaproteobacteria* (Zbinden and Cambon-Bonavita 2003; Durand *et al.* 2010, 2015). The role of these lineages is still poorly understood, as they seem not to be highly implicated in the host's nutrition, in contrast to the second symbiotic population located in the cephalothorax (Fig. 1B) (Ponsard *et al.* 2013). There, bacteria are mostly present on the bacteriophage setae of the dorsal and ventral sides of hypertrophied exopodites of the second maxilla (scaphognathites) and first maxillipeds, and on the internal tegument of the carapace (branchiostegites) (Casanova *et al.* 1993; Segonzac *et al.* 1993). This community is composed of rod, coccoid, thick and thin filamentous bacteria (Gebruk *et al.* 1993; Casanova *et al.* 1993; Segonzac *et al.* 1993; Zbinden *et al.* 2004). A great phylogenetic diversity has been revealed within this community, bacteria belonging mostly to the Proteobacteria phylum with two dominant groups: *Epsilon*- and *Gammaproteobacteria*, and lower abundance of *Alpha*, *Beta*-, *Delta*-, and *Zetaproteobacteria*. To a lesser extent, Flavobacteria, Bacilli, Clostridia and Aquificae are also present (Zbinden *et al.* 2008, Petersen *et al.* 2010; Hügler *et al.* 2011; Guri *et al.* 2012; Jan *et al.* 2014). Chemoautotrophy, along with sulphide, methane, iron and hydrogen oxidation, were demonstrated as major metabolisms within the epibiosis (Wirsén *et al.* 1993; Polz *et al.* 1998; Zbinden *et al.* 2004; Zbinden *et al.* 2008; Hügler *et al.* 2011; Guri *et al.* 2012; Ponsard *et al.* 2013; Jan *et al.* 2014). The supply of organic carbon to the host by cephalothoracic bacteria has been demonstrated by isotopic analyses (Pond *et al.* 1997; Rieley *et al.* 1999) and a direct nutritional transfer of

soluble bacterial by-products to the host occur by transtegumental absorption across the gill chamber integument (Ponsard *et al.* 2013).

Although *R. hybisae* (Nye, Copley and Plouviez 2011) and *R. kairei* (Watabe and Hasimoto 2002) show similarly an inflated cephalothorax and expanded scaphognathites, with numerous plumose-setae as *R. exoculata*; to date, dense coverage of bacteria was only described from the scaphognathites of *R. hybisae* (Nye, Copley and Plouviez 2011; Assier 2016).

Among species belonging to the invalid genus *Chorocaris*, none has been described with an enlarged cephalothorax and *R. chacei* is the only one in which the presence of epibiotic bacteria has been reported (Casanova *et al.* 1993). *R. chacei* lives along the MAR, in small communities (up to 50 ind m⁻²) in the periphery, when present, of *R. exoculata* aggregates (up to 3000 ind m⁻²) (Gebruk *et al.* 1997; Komai and Segonzac 2008). If this species does not present the cephalothorax hypertrophy, its scaphognathites and exopodites nevertheless bear plumose setae on both sides, although less numerous than in *R. exoculata* (Casanova *et al.* 1993). Dense filamentous microbial communities have been observed on mouthparts and inside the carapace, but the area colonized are smaller than in *R. exoculata* (Casanova *et al.* 1993; Segonzac *et al.* 1993). The occurrence of bacteria in the digestive tract has not been investigated yet. Stable isotope analysis showed that, in small adults of *R. chacei*, much of the carbon is derived from chemosynthetic bacteria, but in larger individuals scavenging or occasional predation on *R. exoculata* or mussels may occur, with still consequent input from the epibiotic bacteria (Gebruk *et al.* 2000).

In the present study, we intended to provide a detailed description and characterization of the association between *R. chacei* and its microbial communities (in both the cephalothorax and the digestive system) and to draw hypothesis about their roles in comparison with *R.*

125 *exoculata* symbiosis. We used both microscopic and molecular approaches to fully localize
126 and describe microbial populations associated with *R. chacei*. The commonalities and the
127 differences between the symbiosis of *R. chacei* and of *R. exoculata* will be discussed by
128 wondering how two species with relatively close symbiosis can have overlapping distribution
129 (with or) without competition, opening questions on host evolution that could be driven by
130 symbiosis development.

131

132 Material and Methods

133 Sampling

134 Specimen of *R. chacei* were collected at several hydrothermal vents along the MAR (Table 1):
135 Lucky Strike (37°17'N; 32°16.3'W; 1730 m depth) and Rainbow (36°13'N; 33°54'W; 2350
136 m depth) during the BioBaz cruise (August 2013) and TAG (26°08'N; 44°49.6'W; 3700 m
137 depth) and Snake Pit (23°23'N; 44°56.1'W; 3480 m depth) during the BICOSE cruise
138 (January 2014). Shrimps were collected using the suction sampler of the Remotely Operated
139 Vehicle "Victor 6000" from the research vessel "Pourquoi pas?". Once on board, individuals
140 (Table 1) were dissected into body parts (branchiostegites (LB), scaphognathites (Sc) and
141 digestive tract (foregut (FG) and midgut (MG)) (Figs 1A and 1B, as defined for *R. exoculata*).
142 For molecular studies, these tissues were directly frozen (-80°C) on board and DNA
143 extractions were performed in the laboratory. For TEM and SEM, samples were fixed in a
144 2.5% glutaraldehyde-seawater solution and later post-fixed in osmium tetroxide. For
145 Fluorescent *in situ* hybridization, samples were immediately fixed for 2h in a formaldehyde
146 3%-sterile seawater solution and further treated as described in Durand *et al.* (2010). In order
147 to observe well developed bacterial community in the cephalothorax, all specimens analyzed
148 in this study were chosen at preecdysial stage (i.e. just before the moult), based on their
149 branchiostegite coloration, as described for *R. exoculata* (Corbari *et al.* 2008a).

150

151 ***DNA extraction and PCR amplification***

152 DNA from all tissues (LB, SC, FG and MG) was extracted and purified using the
153 Nucleospin® soil kit (Macherey-Nagel) following the manufacturer recommendations.
154 Bacterial 16S rRNA gene fragments were PCR-amplified in 30 cycles at an annealing
155 temperature of 48°C, using the general bacterial primer set E8F (5' AGA GTT TGA TCA
156 TGG CTC AG 3') and U1492R (5' GTT ACC TTG TTA CGA CTT 3') (Lane 1991). They
157 were then purified with a Nucleospin® Extract II kit (Macherey-Nagel) following the
158 manufacturer recommendations.

159

160 ***Cloning and sequencing***

161 The amplified PCR products were cloned using the TOPO TA cloning kit (Invitrogen)
162 following the manufacturer's instructions. The plasmid inserts were controlled by
163 amplification using M13F and M13R primers. Positive clones were then sequenced at GATC
164 Biotech (Germany) on a Sanger ABI 3730xl DNA Sequencer.

165

166 ***Phylogenetic analysis***

167 Sequences (16S rDNA) were checked for quality (length, HQ %, and number of repeated
168 nucleotides in sequences fixed under 6), sorted into "Forward" and "Reversed" sequences
169 based on PCR primer identification (E8F and U1492R respectively) and checked for chimera
170 with UCHIME (Edgar *et al.* 2011) and Decipher (Wright *et al.* 2012) algorithms. Sequences
171 were aligned using MAFFT (Kato *et al.* 2002) algorithm and refined manually with
172 Geneious 8.1.8 © (Kearse *et al.* 2012). All phylogenetic trees were built with Geneious 8.1.8
173 ©. Phylogenetic analyses were performed on the basis of evolutionary distance (Neighbor
174 Joining; Saitou and Nei 1987) using general time reversibility (GTR) model for correction

matrix. The robustness of phylogenetic reconstructions was tested by bootstrap re-sampling (1000) (Felsenstein 1985). Sequences showing more than 97% similarity using the Furthest method (Schloss and Handelsman 2005) with MOTHUR software (Schloss *et al.* 2009) were considered to be sufficiently related and were grouped in the same phylotype (OTU: Operational taxonomy units). Rarefaction curves (see Supplementary Figs S1 and S2) were created to estimate sequencing depth. Braycurtis calculator (Bray and Curtis 1957) was used to generate β diversity matrix. Singletons were deleted from data to generate phylogenetic trees but are present in diversity data.

The rarefaction curves and Simpson indices were obtained with MOTHUR (at 97% similarity) for all libraries (Schloss *et al.* 2009). Simpson index was calculated as follows:

$$1 - D_{(Simpson)} = 1 - \left[\frac{\sum_{i=1}^{S_{obs}} n_i (n_i - 1)}{N(N - 1)} \right]$$

where S_{obs} representing the number of OTUs observed, n_i the number of individuals for one OTUs and N the total number of OTUs. Simpson index and related evenness of Simpson index were calculated according to MOTHUR calculator (Schloss and Handelsman 2005; Schloss *et al.* 2009).

Good's coverage was calculated as a percentage, according to the following relation:

$C = [1 - (n/N)] \times 100$, where n represented the number of phylotypes appearing only once in a library and N being the library size (Good 1953).

Fluorescent in situ hybridization (FISH)

The FISH protocol used was described previously (Durand *et al.* 2010). Briefly, 3% formaldehyde fixed dissected tissues were embedded in polyethylene glycol distearate-1-hexadecanol (9: 1) resin (Sigma, St.Louis, MO). Resin blocks were then cut into 6–10 μ m

200 sections using an RM 2165 microtome (Reichert-Jung, Germany). Sections were hybridized
201 using several published probes (Table 2). The hybridization temperature was the same for all
202 samples treated (46°C). Observations and imaging were performed using an ApoTome Axio
203 Imager Z with a COLIBRI system (Zeiss, Jena, Germany) using ZEN software (Zeiss, Jena,
204 Germany).

206 *Scanning electron microscopy (SEM)*

207 Before preparation, samples were observed with a stereomicroscope and photographed
208 (Olympus SZX12 equipped with an Olympus U-CMAD3 camera). Samples were dehydrated
209 through an ethanol series and critical point dried (Emitech K850). Samples were then gold-
210 coated with a JEOL JFC-1200 fine coater. Observations were performed with a scanning
211 electron microscope (Hitachi SU3500), operating between 5 to 25kV according to the sample.

213 *X-ray microanalysis*

214 Elemental energy-dispersive X-ray microanalyses (EDX) were performed on the samples
215 used for SEM. X-ray microanalyses and elemental mappings were carried out using a
216 scanning electron microscope (FEI Quanta 200) operating at 20kV, and acquired with an
217 energy dispersive X-ray detection system (SDD X-Max 80mm detector).

219 *Light microscopy (LM) and Transmission electron microscopy (TEM)*

220 Samples were carefully dissected into small pieces that were dehydrated through an ethanol
221 and propylene oxide series, and then embedded in an epoxy resin (Serlabo). Semi-thin
222 (600nm) and ultra-thin (60nm) sections were obtained from a Reichert-Jung Ultramicrotome
223 (Ultracut E) using a diamond knife. Semi thin sections were stained with toluidine blue for
224 observations by light microscopy (using Zeiss, Jena, Germany microscope). Ultra-thin

sections were laid on copper grids and stained with Urany-less and lead citrate (Delta Microscopies). Observations were performed on ZEISS 912 transmission electron microscope operating at 80 kV.

Nucleotide sequence accession numbers:

The sequences from this study are available through GenBank under the following accession numbers: LT855310 to LT855375 (16S rRNA sequences).

Results

Shrimp morphological observations

Light microscopic observations of the inner branchiostegites revealed two distinct areas (Fig. 1C): (i) the anterior part (corresponding to one third of the branchiostegite surface) showed an orange coloration, due to mineral deposits, (ii) the posterior part facing the gills which was translucent. On the most mineralized specimens, a band of mineral deposits also occurred along the ventral border, which ascend posteriorly, probably beyond the area covered by the sweeping of the long multidenticulate gill-cleaning setae present at the posterior end of the scaphognathites (not shown). Furthermore, the ventral setae and edging close to the epimeral border presented also the orange coloration observed on the branchiostegites. The scaphognathites (Fig. 1D), which bear long bacteriophage setae on their both sides, were also covered by orange mineral deposits. On samples from Lucky Strike (and to a lesser extent those from Snake Pit), stereomicroscopic observations revealed the presence of a few small black/grey mineral particles, on the branchiostegites and scaphognathites. Stomachs and guts of shrimps from the Rainbow site were full of orange/brown content, whereas those from Snake Pit and Lucky Strike were rather full of black/grey content, those from TAG having both orange and black particles in their content (data not shown).

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

Morphology and distribution of bacteria

SEM observations of the branchiostegites (Figs 2A and 2C) and scaphognathites (Fig. 2B) showed a dense microbial coverage. Combining SEM and TEM, a total of 6 bacterial morphotypes were observed: three type of filaments (Figs 2D and 3A-C), two rod-shaped (Fig. 2E), and one coccobacillus morphotype (Fig. 3D-E). All these morphotypes were present on branchiostegites and scaphognathites of shrimps of each studied vent site, except for large filamentous morphotypes that were not observed on TAG site, and methanotrophic like bacteria that were not observed on TAG and Snake Pit sites. It should be noted that the specimen observed for the TAG site exhibited less developed bacterial colonization than others (probably a somewhat earlier stage in the moult cycle), possibly explaining the absence of certain morphotypes. Sizes and abundances of bacteria on vent sites where specimens were collected are summarized in Table 3. Filamentous morphotypes comprised a large type (Lt, Figs 2D and 3A) and two thin types: one with short and thin cells (type 1, Tf1, Figs 2D and 3B) and one with longer and larger cells (type 2, Tf2, Figs 2D and 3C). The length of thin filament cells (type 2) was difficult to measure because of unclear septum identification between each bacterial cell. Among the rods, the first type was longer and thinner than the second type (Fig. 2E). Most of the time, rods were prostrate, resting on their long side. In some areas, rods of type 2 were erected, standing up on their short side. The coccobacillus morphotype (Fig. 3E) presented stacks of internal membranes characteristic of methanotrophic like bacteria type I (Fig. 3D) (Hanson and Hanson 1996). They occurred in small aggregates of 4 to 7 bacteria.

In all SEM observations (n = 4 individuals), bacterial colonization on branchiostegites and scaphognathites was correlated with mineral deposits. Qualitative analyses with X-Ray microanalysis revealed two major types of mineral deposits. The first, and largely dominant

mineral, was mainly composed of iron (major peaks $K\alpha$ at 6.400 keV and $K\beta$ at 7.059 keV) and oxygen, probably as iron oxides (presented here for Rainbow site, Fig. 3F). The second type occurred only punctually and was mainly composed of sulphur (major $K\alpha$ peak at 2.307 keV) and iron, but along with Cu ($K\alpha$ at 8.041 keV and $K\beta$ at 8.907 keV), probably as iron/copper sulphides (Fig. 3G from the Rainbow site). Same profiles were obtained for the minerals associated with shrimp from the other vent sites.

Combining light and electron microscopic observations on shrimps from different vent sites, we identified four distinct bacterial colonization areas on branchiostegites, as summarized in Fig. 4A. The area 1 was characterized by a total absence of bacteria and mineral deposits. The area 2 was characterized by the occurrence of a monolayer of the two rod morphotypes (type 1 colored in red and type 2 in blue on Fig. 4A), and no associated mineral deposits. The areas 1 and 2 corresponded to the translucent part described above with macroscopic observations. The area 3 was characterized by a thick bacterial mat composed of the three filamentous morphotypes (Lf colored in green, Tf1 in orange and Tf2 in purple on Fig. 4A), as well as an increasing amount of mineral deposits towards the fourth area. Above, inside and under this deposit, rods of the two types were also present. The area 4 was similar to the third one, but with a greater thickness of mineral deposits. It was also the only area presenting methanotrophic like bacteria (Fig. 3D), clustering by four or five cells (colored in turquoise on Fig. 4A). The areas 3 and 4 corresponded to the anterior part described above with macroscopic observations.

Scaphognathites were usually heavily mineralized (Fig. 1D), showing occurrence of orange-brown deposits on both sides, as well as on the bacteriophore setae, along with the different morphotypes of rods and filaments (Fig. 2B).

299 Cross-sections of *R. chacei* midgut (deprived of cuticle and where the nutrient absorption
300 occur) revealed large exchange surfaces as the epithelium displayed numerous invaginations
301 (Figs 5A and B). Semi-thin section observations of the gut content, delimited by the
302 peritrophic membrane, revealed that it was mainly composed of black and brown minerals,
303 probably as iron sulfides and oxides, along with organic matter in the form of few cuticle
304 fragments and ingested bacteria (filaments and coccoid cells). Epithelial gut cells were
305 densely colonized by a single morphotype of long (up to 20-25 μm) and thin (about $0.18 \pm$
306 $0.02 \mu\text{m}$ ($n=100$)), unsegmented filamentous bacteria (Figs 5B and C), inserted between the
307 microvilli (Fig. 5D). These filamentous bacteria were observed in dense populations within
308 the gut of all specimens, with no visible difference in their distribution and abundance,
309 regardless of the vent site.

311 *Diversity of Rimicaris chacei bacterial communities*

312 Our results are based on few samples per site due to environmental and technical limitations.
313 Moreover, PCR and cloning approaches are somewhat limited by amplification quality and
314 number of clones treated. Considering all these possible bias, all samples were treated in the
315 same way all along the study combining multi approaches (microscopies, FISH and molecular
316 analyses), which enable us to make some comparisons between specimens and sites treated
317 (Qiu *et al.* 2001).
318 Good's coverages were inferior to those of previous study on *R. exoculata* due to slight
319 differences in technical analysis. In Guri *et al.* (2012), and Durand *et al.* (2010), OTUs
320 (threshold at 97%) were manually inferred from phylogenetic trees analyses. Here we used
321 methods specially developed for clustering, to limit phylogenetic tree possible bias and
322 manual assignment of OTUs. Usually, this method creates more OTUs than manual clustering
323 on phylogenetic tree, which explain slightly lower Good's average of this study. However,

324 rarefaction curves are similar to those from previous studies (Guri *et al.* 2012; Durand *et al.*
325 2010) (see Supplementary Figs S1 and S2).

326 A total set of 1830 clones of 16S rRNA was sequenced (between 41 and 96 sequences per
327 library: i.e. branchiostegites, scaphognathites, foregut and midgut) to study bacterial
328 communities associated with *R. chacei* specimens. Taxonomy was assigned for each sequence
329 and summarized in Table 4. Furthermore, based on 97% of similarity, sequences were
330 clustered in OTUs (the number of OTUs per taxonomy units is also summarized in Table 4).

331 In our data, 12 groups were identified (all samples combined), listed below according to the
332 number of clones: *Epsilonproteobacteria*, *Gammaproteobacteria*, *Deferribacteres*, *Mollicutes*,
333 *Alphaproteobacteria*, *Bacteroidetes*, *Candidate division OD1*, *Deltaproteobacteria*,
334 *Zetaproteobacteria*, *Lentisphaerae*, *Betaproteobacteria* and *Firmicutes*. Studies based on
335 PCR amplifications and clone libraries are known to underestimate genetic diversity and can
336 be biased (Qiu *et al.* 2001). Nevertheless, our results showed that diversity associated with *R.*
337 *chacei* cephalothorax (i.e. branchiostegite and scaphognathite samples) was mostly
338 represented by *Epsilonproteobacteria* and *Gammaproteobacteria* related sequences, while
339 that of the gut (i.e. foregut and midgut samples) was mostly represented by *Deferribacteres*
340 and *Mollicutes* related sequences (Table 4). We then focused our study on these 4 main
341 lineages for phylogenetic analysis, using only forward sequences for quality in the
342 phylogenetic reconstruction, and excluding singleton OTUs.

343 Rarefaction curves (Supplementary Figs S1 and S2) indicated that clone libraries did not
344 exhaust all the diversity but reached the plateau. Good's coverage confirmed rarefaction
345 curves observations, with an average of 80.75% (+/- 9.65%), describing correctly bacterial
346 communities' diversity associated to shrimps. *Alpha* diversity calculated with Simpson index
347 and Simpson evenness highlighted diversity profiles typical of symbiosis: a low observed
348 diversity, with low equitability, reflecting the dominance of some phylotypes. *Alpha* diversity

349 was lower in the gut ($1\text{-Simpson} < 0.7$) than in the cephalothorax ($1\text{-Simpson} > 0.7$). *Beta*
350 diversity based on Bray Curtis analysis confirmed differences between gut and cephalothorax
351 communities' composition (Supplementary Fig. S3).

352

353 ***16S rRNA phylogeny among the Epsilonproteobacterial related sequences:***

354 16S rRNA sequences related to the *Epsilonproteobacteria* within *R. chacei* cephalothorax
355 clone libraries were distributed in 24 OTUs, clustering within *R. exoculata* cephalothoracic
356 (Polz and Cavanaugh 1995; Zbinden *et al.* 2008; Petersen *et al.* 2010; Guri *et al.* 2012) and
357 digestive tract clone sequence groups (Zbinden and Cambon-Bonavita 2003; Durand *et al.*
358 2010; Durand *et al.* 2015). *R. chacei* OTUS were mainly distributed within 2 phylotypes. The
359 first one was related to *Sulfurovum* species, and the second to *Epsilon* 1 to 4 groups according
360 to Petersen and collaborators' nomenclature (Petersen *et al.* 2010) (Fig. 6). However,
361 *Sulfurovum* related OTUs retrieved from *R. chacei* were also closely related to sequences
362 from hydrothermal vent systems worldwide. The closest cultured bacteria related to the most
363 represented phylogenetic group were *Sulfurovum lithotrophicum* (Inagaki *et al.* 2004),
364 *Sulfurovum aggregans* (Mino *et al.* 2014) and *Sulfurovum* sp. strain NBC37-1 (Nakagawa *et*
365 *al.* 2007). These species were all known to be sulfur-oxidizing chemolithoautotrophic
366 bacteria, isolated from hydrothermal vents. The other phylogenetic groups composed by
367 *Epsilon* 1 to 4 were not represented by cultured strains, but were close to *R. exoculata*
368 symbionts clusters (Petersen *et al.* 2010).

369

370 ***16SrRNA phylogeny of the Gammaproteobacterial related sequences:***

371 16S rRNA sequences of *R. chacei* clone libraries affiliated to *Gammaproteobacteria* clustered
372 with previously described *R. exoculata* cephalothoracic (Polz and Cavanaugh 1995; Zbinden
373 *et al.* 2008; Petersen *et al.* 2010; Guri *et al.* 2012) and digestive tract clones (Zbinden and

374 Cambon-Bonavita 2003; Durand *et al.* 2010; Durand *et al.* 2015). Two main phylogenetic
375 groups were identified (Fig. 7). The first one was closely related to *Methyloprofundus*
376 *sedimenti* (Tavormina *et al.* 2015), *Methylobacter marinus* (Bowman *et al.* 1993), and
377 *Methylomonas methanica* (Boden *et al.* 2011). These cultivated bacteria are known to be
378 methanotrophic bacteria. The second one was closely related to *Thiothrix flexilis* (Kanagawa
379 *et al.* 2000; Aruga *et al.* 2002; Unz and Head 2015), *Thiomicrospira frisia* (Brinkhoff and
380 Muyzer 1997), *Thiomicrospira psychrophila* (Knittel *et al.* 2005) and *Cocleimonas flava*
381 (Tanaka *et al.* 2011), all known to be chemoautotrophic sulfur-oxydizers. Five OTUs were not
382 related to these phylogenetic groups and were related to methanotrophic or thiotrophic
383 phylogenetic groups, close to environmental sequences retrieved from hydrothermal vents.
384 TAG_SC_05, LS_LB_82 and LS_LB_65 clones represent OTUS with respectively 2, 3 and 7
385 sequences in clone libraries and were close to environmental sequences. The two most
386 represented OTUs in our clone libraries (*i.e.* SP_MG_43 and TAG MG_17 clones) were only
387 retrieved from the midgut of *R. chacei*, respectively from Snake Pit and TAG, and did not
388 cluster with any known sequence related with hydrothermal vents or symbiosis.

389

390 ***16S rRNA phylogeny of the Deferribacteres related sequences:***

391 Compared to *Epsilon* and *Gammaproteobacteria* related OTUs, the diversity of
392 *Deferribacteres* related sequences was very low, even if this taxonomic group was retrieved in
393 apparently high abundance in our samples. Only two OTUs were identified (Fig. 8),
394 exclusively in the midgut samples. The OTU, represented by the clone Rb_MG_71, was
395 present at the 4 hydrothermal vent sites, whereas the other, represented by the clone
396 TAG_MG_78, was absent at Lucky Strike. These 2 OTUS were closely related to previously
397 known *R. exoculata* gut symbionts (Durand *et al.* 2010). The closest cultured bacteria relative
398 to this phylogenetic group were *Deferribacter thermophilus* (Greene *et al.* 1997) and

399 *Deferribacter desulfuricans* (Takai *et al.* 2003), both known to be heterotrophic thermophilic
400 and anaerobic metal-reducers, especially of iron or sulfur.

401

402 ***16S rRNA phylogeny of the Mollicutes related sequences:***

403 As for *Deferribacteres*, the diversity of Mollicute related clones was lower than that of
404 *Epsilon* and *Gammaproteobacteria* ones, but still well represented in our libraries obtained
405 from stomach and midgut samples. Three main (and one singleton) OTUs were identified
406 (Fig. 9) and were closely related to *R. exoculata* gut symbionts, but not to environmental
407 sequences. The OTU represented by the clone Rb_FG_02 was present in the foregut of
408 Rainbow and Lucky Strike shrimps, and in the midgut of Rainbow shrimps. The OTU
409 represented by the clone SP_MG_35 was present in the foregut of Rainbow and Lucky Strike
410 shrimps. The OTU represented by the clone LS_MG_65 was present only in TAG and Snake
411 Pit midgut samples.

412

413 ***Fluorescent in situ hybridization***

414 We used FISH microscopy to merge as best as possible microscopic observations with
415 molecular analysis. Development and tests were done to determine optimal hybridization
416 temperature (here 46°C), stringency (here 40% formamide), and to check specific
417 hybridization for each probe listed in Table 2. Only Eub338 (universal bacterial probe), Epsy
418 549 (dedicated to *Epsilonproteobacteria*) and Gam42a (dedicated to *Gammaproteobacteria*)
419 presented clear hybridization results. Several trials using archaeal probes were carried out but
420 still remained unsuccessful (data not shown).

421 On branchiostegites and scaphognathites, *Epsilonproteobacteria* and *Gammaproteobacteria*
422 were the most abundant hybridized cells. Both lineages were represented by filamentous
423 morphotypes. *Epsilonproteobacteria* were represented by large filaments and thin filaments

type 1 and 2 (in yellow Figs 10A and 10B). *Gammaproteobacteria* were represented by thin filaments type 1 (in red Figs 10A and 10B). Some rods hybridized with the Gam42a probe, but the resolution of the image prevents from confidently associated them with one of the two types described above. Methanotrophic-like bacteria, located on and in the mineral crust, failed to give any conclusive fluorescent signal. In the midgut, none of the probes tested hybridized bacteria, except the general bacterial probe Eub338, as already observed for *R. exoculata* (Durand *et al.* 2010; Durand *et al.* 2015). Furthermore, shrimps' epithelial cells were autofluorescent using Cy3 bandwidth. Using this apparent bias, we were able to observe that bacteria were distributed in high density at the apex of the midgut epithelial cells (Fig. 5B).

Discussion

We conducted here the first comprehensive study of microbial communities associated with the hydrothermal shrimp *Rimicaris chacei*. We used several approaches, including microscopy, molecular biology and FISH analyses that showed a close morphologic and phylogenetic proximity with *R. exoculata* symbiosis.

Similarities between R. chacei and R. exoculata symbiotic communities

Our study confirms and deepens the first observations (Casanova *et al.* 1993; Segonzac *et al.*, 1993) on the localization and morphology of the symbionts of *R. chacei*, and highlights strong similarities with *R. exoculata* symbiosis.

Firstly, although the bacteriophore tissues are less developed in *R. chacei* than in *R. exoculata* (i.e. branchiostegites and scaphognathites are not hypertrophied, bacteriophore seta on the scaphognathites are less numerous, areas covered by bacteria on the inner face of the branchiostegites are less important), the cephalothoracic community of *R. chacei* also forms a

449 dense bacterial mat on the anterior third of the inner side of the branchiostegites and on both
450 sides of the scaphognathites (Figs 1C and D, Zbinden *et al.* 2004). In both species, epibionts
451 of the cephalothorax are associated with iron oxides and sulfides deposits (Figs 1C, 1D, 3F
452 and 3G, Zbinden *et al.* 2004; Corbari *et al.* 2008b).

453 Secondly, associated bacteria of the two species are morphologically similar. Using SEM
454 observations, Casanova *et al.* (1993) described 3 morphotypes (rods, thin and thick filaments)
455 among the cephalothoracic epibionts of *R. chacei* epibionts, the same ones that were also
456 described in *R. exoculata* by these authors, with bacteria of roughly similar sizes. Combining
457 SEM and TEM observations allowed us to refine this first description and revealed a larger
458 diversity with 6 bacterial morphotypes (see Table 3 for measures): i) two rod types : a long
459 thin and a short thick type (Fig. 2E); ii) 3 types of filaments including 2 types of thin
460 filaments, differing by the size and shape of their cells (Figs 2D and 3A-C), and iii) a
461 coccobacillus morphotype, corresponding to methanotrophic like bacteria, never observed
462 before in *R. chacei*. These 6 morphotypes correspond to those described for *R. exoculata* with
463 the same combined approach with SEM and TEM (Zbinden *et al.* 2008)

464 In the digestive tract, only one morphotype was observed on all samples analyzed, located in
465 the mid-gut, inserted between the microvilli of the epithelial gut cells and separated from the
466 bolus by the peritrophic membrane. These bacteria are long thin filaments, made of individual
467 cells without any visible septum. Similar bacteria have been described in *R. exoculata* midgut
468 by Durand *et al.* (2010). Observations of bacteria contained in the bolus are difficult due to
469 the important mineral load. Some rods and cocci were nevertheless observed (not shown),
470 most of them rather undamaged, as already observed in *R. exoculata* (Zbinden and Cambon-
471 Bonavita 2003).

472 Thirdly, associated bacteria of the two species are phylogenetically closely related.
473 Phylogenetic studies on *R. exoculata* cephalothoracic populations revealed a large

1
2
3 474 phylogenetic diversity, dominated by two major groups: *Epsilonproteobacteria* and
4
5 475 *Gammaproteobacteria*, related to chemoautotrophic bacteria (Zbinden *et al.* 2008; Petersen *et*
6
7 476 *al.* 2010; Guri *et al.* 2012; Jan *et al.* 2014). Phylogenetic analyses on *R. exoculata* gut
8
9 477 symbionts have led to the identification of four major clades: Mollicutes, Deferribacteres,
10
11 478 *Gammaproteobacteria* and *Epsilonproteobacteria* (Zbinden and Cambon-Bonavita 2003;
12
13 479 Durand *et al.* 2010; Durand *et al.* 2015; Cowart *et al.* 2017). The same 4 phylotypes are well
14
15 480 represented in our *R. chacei* clones libraries, with *Epsilonproteobacteria* and
16
17 481 *Gammaproteobacteria* more abundant in the cephalothorax as seen using FISH analyses, and
18
19 482 Deferribacteres and Mollicutes only present in the gut.
20
21 483 In this study, methanotrophic-like bacteria were observed on Rainbow and Lucky Strike, but
22
23 484 not on Snake Pit and TAG. In 2012, Guri *et al.* also evidenced methanotrophic-like bacteria
24
25 485 on adult of *R. exoculata* from Rainbow site, but not from Logatchev and Achadze, two
26
27 486 hydrothermal sites of the MAR, further south than our. Although both species host
28
29 487 methanotrophic-like bacteria at the Rainbow site, whose fluids are largely enriched in
30
31 488 methane, correlations with environmental parameters remain uncertain since Logatchev is
32
33 489 also enriched in methane while Lucky Strike is not (Charlou *et al.* 2002). Further analyses are
34
35 490 needed to fully explain these differences.
36
37 491
38
39 492 *Potential trophic role of symbionts and mixotrophy of R. chacei*
40
41
42 493 The similarities, in terms of location, morphology and phylogeny of the symbiotic
43
44 494 communities of these two species of *Rimicaris* could suggest similarities in the trophic diet of
45
46 495 the shrimp. Some morphological characters have to be taken into account when considering
47
48 496 the potential diet of *R. chacei*, compared to *R. exoculata* : 1) the stomach volume is almost
49
50 497 twice as big in *R. chacei* as in *R. exoculata* (Casanova *et al.* 1993), implying probably a more
51
52 498 important digestive function. 2) The mandibles, which are reduced and not used for feeding in
53
54 499 *R. exoculata*, are normally developed in *R. chacei* and functional (Casanova *et al.* 1993). 3)
55
56
57
58
59
60

500 The two pairs of chelipeds are relatively small and cannot extend out of the branchial chamber
 501 in *R. exoculata* (making food collection out of cephalothorax difficult) in contrast to *R. chacei*
 502 (Casanova *et al.* 1993). Taken together, these data could suggest a mixotroph diet in *R.*
 503 *chacei*, with a classical collection, crushing, ingestion, digestion process, as shown by the
 504 presence of organic material in the gut (observed in this study and also reported by Casanova
 505 *et al.* 1993 and Segonzac *et al.* 1993). The relatively developed bacterial communities
 506 observed in *R. chacei* suggest that this diet would be supplemented by an input of organic
 507 carbon from the chemoautotrophic epibiotic bacteria. Carbon stable isotopes indeed show that
 508 *R. chacei* has an intermediate $\delta^{13}\text{C}$ signature (-16.1 to -12.3 ‰, Colaço *et al.* 2002) between
 509 *R. exoculata*, which relies mostly on its epibionts for its diet (Polz *et al.* 1998) and *Mirocaris*
 510 *fortunata*, which is opportunist and feeds upon mussels, shrimps or other invertebrates when
 511 available (Gebruk *et al.* 2000). Transcuticular transfer of bacterial by-product to the shrimp
 512 was shown in *R. exoculata* (Ponsard *et al.* 2013) thanks to *in vivo* incubations with isotope-
 513 labelled inorganic carbon. The hypothesis of this nutrition mode arise from several
 514 observations which are: i) the thinness of the branchiostegite inner cuticle lining the bacteria
 515 (between 0.5 μm and 5 μm) and ii) the absence of areas scrapped by the animal, or free of
 516 bacteria on the inner side of branchiostegites or on scaphognathites (Zbinden *et al.* 2004;
 517 Corbari *et al.* 2008a). Regarding *R. chacei*, no scrapped areas on the inner side of
 518 branchiostegites were ever observed on the samples studied. Furthermore, the thickness of the
 519 branchiostegite inner cuticle is of the same order of magnitude as in *R. exoculata* (5.1 ± 0.9
 520 μm for *R. exoculata*, $n = 16$ and $5.5 \pm 2.2 \mu\text{m}$ for *R. chacei*, $n = 16$, V. Apremont and M.
 521 Zbinden pers. comm.). These morphological features could suggest that a transcuticular
 522 transfer of small organic molecules of bacterial origin also occur in *R. chacei*. *In vivo*
 523 experiments with isotope-labelled inorganic carbon should be carried out to properly prove a
 524 transfer of organic material from epibionts to host and a real trophic symbiosis between *R.*

1
2
3 525 *chacei* and its epibionts, as it has been done for *R. exoculata* (Ponsard *et al.* 2013).
4
5 526 The conservation of a mixed diet with several potential sources of food allows switching
6
7 527 between two food sources in case of shortage of one of the two. The relative contribution of
8
9 528 the different sources could be regulated according to external conditions. Species possessing a
10
11 529 mixotrophic regime based in part on the organic molecules supplied by symbiotic bacteria
12
13 530 combined with an external contribution has already been described in hydrothermal symbiotic
14
15 531 species like the bathymodiolin mussels or some galatheid crabs. For example, *B. azoricus*
16
17 532 obtains energy from both a dual endosymbiosis and filter-feeding (Fiala-Médioni *et al.* 1986).
18
19 533 Martins *et al.* (2008) showed that the relative contribution of filter-feeding and endosymbiosis
20
21 534 varies with *B. azoricus* size, with small mytilids being strongly dependent on filter-feeding,
22
23 535 whilst larger mussels obtain a significant portion of its energy from endosymbiosis. Such
24
25 536 dietary modulations also seem to exist in *R. chacei*, inversely correlated with those of *B.*
26
27 537 *azoricus*, since the smallest adult specimens predominantly rely on their symbionts (Gebruk *et*
28
29 538 *al.* 2000). Among Crustaceans, the squat lobster *K. hirsuta* harbors dense cluster of
30
31 539 filamentous bacteria on the setae that densely covered its pereopods and the ventral surface of
32
33 540 its cephalothorax (Macpherson *et al.* 2005; Goffredi *et al.* 2008) and could serve as a direct
34
35 541 nutritional resource as it has been shown in another squat lobster *Shinkaia crosnieri* (Tsuchida
36
37 542 *et al.* 2011). *K. hirsuta* was observed consuming bathymodioline mussels that had been
38
39 543 damaged by the Alvin submersible, indicative of some degree of omnivory (Macpherson *et al.*
40
41 544 2005). In addition, Martins *et al.* (2008) observed a spatial segregation for *B. azoricus* with
42
43 545 the largest specimens (strongly relying on their symbionts) living at the warmest areas, where
44
45 546 higher concentration of reduced compounds required by bacteria occurred, while the smaller
46
47 547 individuals (less dependent on their symbionts) are found further away from the vent flow.
48
49 548 Such segregation is also observed for the two *Rimicaris* species: *R. exoculata*, which obtains
50
51 549 most of its energy from its epibionts, lives close to the vent emission, while *R. chacei*,
52
53
54
55
56
57
58
59
60

presumably mixotroph and depending only partly on its symbionts, lives further away from fluid outlets (Segonzac *et al.* 1993), where the chemoautotrophic production is lowered.

Mixotrophy in R. chacei: result of a competition or evolutionary stage?

R. chacei and *R. exoculata* live on the same MAR vent sites, on overlapping habitats, the former being restricted to the surroundings of the aggregates of the latter (Segonzac *et al.* 1993; Gebruk *et al.* 1997). *R. exoculata* always dominates numerically these habitats (Segonzac *et al.* 1993), suggesting that its fully symbiotic diet is able to sustain larger biomass than the mixotrophic strategy of *R. chacei*. We can speculate that the situation is due to a competition with *R. exoculata*, that would maintain *R. chacei* at a distance from the nourishing fluid, thus in a less effective symbiosis, requiring complementary contributions of a mixotrophic regime. On the other hand, this state of affairs is a snapshot of the current situation and we can also hypothesize that *R. chacei* is progressively moving from the ancestral scavenger/predator nutrition strategy (found in most basal clades of Alvinocarididae: Mirocaridinae and Alvinocaridinae (Shank 1999 ; Yang *et al.* 2013; Vereshchaka *et al.* 2015)) to the symbiotic nutrition, and has not achieved a full transition.

Interestingly, recent phylogenetic analyses (Teixeira *et al.* 2013; Vereshchaka *et al.* 2015; Hernandez 2016) showed that the closest known relative to *R. chacei* is the Caribbean species *R. hybisae*. Teixeira *et al.* (2013) even suggested that they could be a single species. *R. hybisae* lives in Mid-Cayman Trough on the western Caribbean Sea and presents the same characteristic enlarged cephalothorax than *R. exoculata* (Nye, Copley and Plouviez 2011). Isotopic data showed that it predominantly relied on its gill ectosymbionts for its organic carbon needs (Streit *et al.* 2015), either through nutrients transfer as in *R. exoculata* (Ponsard *et al.* 2013) and/or through grazing them (Assier 2016). The morphological and nutritional differences between *R. chacei* and *R. hybisae*, despite their strong phylogenetic proximity,

could argue that the reason for maintaining a mixed strategy by *R. chacei* is rather linked to a niche competition with *R. exoculata*, keeping *R. chacei* in an intermediate stage, halfway through a complete symbiosis. Nevertheless more data and a robust phylogeny would be necessary to support the hypothetical process of speciation and divergent evolution of these species, possibly with the symbiosis as factor of evolutive success.

580

581 **Funding**

582 This work was supported by the European Union Seventh Framework Programme under the
583 MIDAS project [grant agreement n°603418]. This work was supported by Ifremer and
584 University Pierre and Marie Curie-Sorbonne Universités, for V.A. PhD thesis grant and
585 functional support.

586

587 **Conflict of interest**

588 None of the co-authors have a conflict of interest to declare

589

590 **Acknowledgements**

591 The authors thank all the chief scientists of the BIOBAZ 2013 and BICOSE 2014 (F. Lallier
592 and M.A. Cambon-Bonavita), the captains and crews of the R/V *Pourquoi pas?*, and the ROV
593 Victor6000 submersible teams for their efficiency. SEM observations were performed at the
594 Electronic Microscopy Platform of the National Museum of Natural History, Paris, with the
595 help of Caroline Dalle. TEM observations were performed at the electronic microscopy
596 platform of the Institute of Biology Paris-Seine (IBPS), with the help of G. Toutirais and M.
597 Trichet. X-ray microanalyses were performed at Ifremer Brest with the help of Nicolas Gayet.

598

599 **References**

- 1
- 2
- 3 600 Amann R, Binder B, Olson R *et al.* Combination of 16S rRNA-targeted oligonucleotide
- 4
- 5 601 probes with flow cytometry for analysing mixed microbial populations. *Appl Env*
- 6
- 7 602 *Microbiol* 1990;**56**: 1919-25.
- 8
- 9
- 10 603 Aruga S, Kamagata Y, Kohno T *et al.* Characterization of filamentous Eikelboom type 021N
- 11
- 12 604 bacteria and description of *Thiothrix disciformis* sp. nov. and *Thiothrix flexilis* sp. nov. *Int*
- 13
- 14 605 *J Syst Evol Microbiol* 2002;**52**: 1309-16.
- 15
- 16 606 Assier A. Deep Se(a)quencing: a study of deep sea ectosymbioses using next generation
- 17
- 18 607 sequencingMax-Planck-Institut für marine Mikrobiologie, DOI [https://elib.suub.uni-](https://elib.suub.uni-bremen.de/edocs/00105697-1.pdf)
- 19
- 20 608 [bremen.de/edocs/00105697-1.pdf](https://elib.suub.uni-bremen.de/edocs/00105697-1.pdf). Bremen: Universität Bremen, 2016, 231.
- 21
- 22
- 23 609 Baker MC, Ramirez-Llodra EZ, Tyler PA *et al.* Biogeography, Ecology, and Vulnerability of
- 24
- 25 610 Chemosynthetic Ecosystems in the Deep Sea. In: McIntyre AD (ed.) *Life in the World's*
- 26
- 27 611 *Oceans: Diversity, Distribution, and Abundance*, DOI 10.1002/9781444325508.ch9.
- 28
- 29 612 Oxford, UK: Wiley-Blackwell, 2010, 161-82.
- 30
- 31
- 32 613 Belkin S, Nelson DC, Jannasch HW. Symbiotic assimilation of CO₂ in two hydrothermal vent
- 33
- 34 614 animals, the mussel *Bathymodiolus thermophilus* and the tube worm *Riftia pachyptila*. *Biol*
- 35
- 36 615 *Bull* 1986;**170**: 110-21.
- 37
- 38
- 39 616 Boden R, Cunliffe M, Scanlan J *et al.* Complete genome sequence of the aerobic marine
- 40
- 41 617 methanotroph *Methylomonas methanica* MC09. *J Bacteriol* 2011;**193**: 7001-2.
- 42
- 43 618 Bowman JP, Sly LI, Nichols PD *et al.* Revised taxonomy of the Methanotrophs: Description
- 44
- 45 619 of *Methylobacter* gen. nov., emendation of *Methylococcus*, validation of *Methylosinus* and
- 46
- 47 620 *Methylocystis* species, and a proposal that the family *Methylococcaceae* includes only the
- 48
- 49 621 Group I Methanotrophs. *Int J Syst Evol Microbiol* 1993;**43**: 735-53.
- 50
- 51
- 52 622 Bray JR, Curtis JT. An ordination of the upland forest communities of southern Wisconsin.
- 53
- 54 623 *Ecol Monogr* 1957;**27**: 325-49.
- 55
- 56 624 Brinkhoff T, Muyzer G. Increased species diversity and extended habitat range of sulfur-
- 57
- 58
- 59
- 60

- oxidizing *Thiomicrospira* spp. *Appl Environ Microbiol* 1997;**63**: 3789-96.
- Casanova B, Brunet M, Segonzac M. Impact of bacterial epibiosis on functional-morphology of shrimp associated with the Mid-Atlantic hydrothermal conditions. *Cah Biol Mar* 1993;**34**: 573-88.
- Cavanaugh CM, Gardiner SL, Jones ML *et al*. Prokaryotic Cells in the Hydrothermal Vent Tube Worm *Riftia pachyptila* Jones: Possible Chemoautotrophic Symbionts. *Science* 1981;**213**: 340-2.
- Charlou JL, Donval JP, Fouquet Y *et al*. Geochemistry of high H₂ and CH₄ vent fluids issuing from ultramafic rocks at the Rainbow hydrothermal field (36°14'N, MAR). *Chem Geol* 2002;**191**: 345-59.
- Childress JJ, Fisher C, Brooks J *et al*. A methanotrophic marine molluscan (Bivalvia, Mytilidae) symbiosis: Mussels fueled by gas. *Science* 1986;**233**: 1306-8.
- Colaço A, Dehairs F, Desbruyères D. Nutritional relations of deep-sea hydrothermal fields at the Mid-Atlantic Ridge: a stable isotope approach. *Deep-Sea Res Pt I* 2002;**49**: 395-412.
- Corbari L, Zbinden M, Cambon-Bonavita M-A *et al*. Bacterial symbionts and mineral deposits in the branchial chamber of the hydrothermal vent shrimp *Rimicaris exoculata*: relationship to moult cycle. *Aquat Biol* 2008a;**1**: 225-38.
- Corbari L, Cambon-Bonavita M-A, Long GJ *et al*. Iron oxide deposits associated with the ectosymbiotic bacteria in the hydrothermal vent shrimp *Rimicaris exoculata*. *Biogeosciences* 2008b;**5**: 1825-65.
- Corliss J, Dymond J, Gordon L *et al*. Submarine thermal springs on the Galapagos rift. *Science* 1979;**203**: 1073-83.
- Cowart D, Durand L, Cambon-Bonavita M-A *et al*. Investigation of bacterial communities within the digestive organs of the hydrothermal vent shrimp *Rimicaris exoculata* provide insights into holobiont geographic clustering. *PLoS ONE* 2017;**12**(3): e0172543. doi:10.1371/journal.pone.

- 1
- 2
- 3 650 Desbruyères D, Biscoito M, Caprais J-C *et al.* Variations in deep-sea hydrothermal vent
- 4
- 5 651 communities on the Mid-Atlantic Ridge near the Azores plateau. *Deep-Sea Res Pt 1*
- 6
- 7 652 2001;**48**: 1325-46.
- 8
- 9 653 Dubilier N, Bergin C, Lott C. Symbiotic diversity in marine animals: the art of harnessing
- 10
- 11 654 chemosynthesis. *Nat Rev Micro* 2008;**6**: 725-40.
- 12
- 13 655 Duperron S, Halary S, Lorion J *et al.* Unexpected co-occurrence of six bacterial symbionts in
- 14
- 15 656 the gills of the cold seep mussel *Idas* sp. (Bivalvia: Mytilidae). *Environ Microbiol*
- 16
- 17 657 2008;**10**: 433-45.
- 18
- 19 658 Duperron S, Lorion J, Samadi S *et al.* Symbioses between deep-sea mussels (Mytilidae:
- 20
- 21 659 Bathymodiolineae) and chemosynthetic bacteria: diversity, function and evolution. *CR Biol*
- 22
- 23 660 2009;**332**: 298-310.
- 24
- 25 661 Duperron S, Nadalig T, Caprais J-C *et al.* Dual symbiosis in a *Bathymodiolus* sp. mussel from
- 26
- 27 662 a methane seep on the Gabon continental margin (Southeast Atlantic): 16S rRNA
- 28
- 29 663 phylogeny and distribution of the symbionts in gills. *Appl Environ Microbiol* 2005;**71**:
- 30
- 31 664 1694-700.
- 32
- 33 665 Durand L, Roumagnac M, Cuff-Gauchard V *et al.* Biogeographical distribution of *Rimicaris*
- 34
- 35 666 *exoculata* resident gut epibiont communities along the Mid-Atlantic Ridge hydrothermal
- 36
- 37 667 vent sites. *FEMS Microbiol Ecol* 2015;**91**.
- 38
- 39 668 Durand L, Zbinden M, Cuff-Gauchard V *et al.* Microbial diversity associated with the
- 40
- 41 669 hydrothermal shrimp *Rimicaris exoculata* gut and occurrence of a resident microbial
- 42
- 43 670 community. *FEMS Microbiol Ecol* 2010;**71**: 291-303.
- 44
- 45 671 Edgar RC, Haas BJ, Clemente JC *et al.* UCHIME improves sensitivity and speed of chimera
- 46
- 47 672 detection. *Bioinformatics* 2011;**27**: 2194-200.
- 48
- 49 673 Felbeck H. Chemoautotrophic potential of the hydrothermal vent tube worm, *Riftia pachyptila*
- 50
- 51 674 Jones (Vestimentifera). *Science* 1981;**213**: 336-8.
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

- 675 Felbeck H, Childress JJ, Somero GN. Calvin-Benson cycle and sulphide oxidation enzymes in
676 animals from sulphide-rich habitats. *Nature* 1981;**293**: 291-3.
- 677 Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*
678 1985;**39**: 783-91.
- 679 Fiala-Médioni A, Métivier C, Herry A *et al.* Ultrastructure of the gill of the hydrothermal-
680 vent mytilid *Bathymodiolus* sp. *Mar Biol* 1986;**92**: 65-72.
- 681 Fujiwara Y, Kawato M, Yamamoto T *et al.* Three-year investigations into sperm whale-fall
682 ecosystems in Japan. *Mar Ecol* 2007;**28**: 219-32.
- 683 Gebruk AV, Galkin S, Vereshchaka A *et al.* Ecology and biogeography of the hydrothermal
684 vent fauna of the Mid-Atlantic Ridge. *Adv Mar Biol* 1997;**32**: 93-144.
- 685 Gebruk A, Southward E, Kennedy H *et al.* Food sources, behaviour, and distribution of
686 hydrothermal vent shrimp at the Mid-Atlantic Ridge. *J Mar Biol Assoc UK* 2000;**80**: 485-
687 99.
- 688 Goffredi S, Jones W, Ehrlich H *et al.* Epibiotic bacteria associated with the recently
689 discovered Yeti crab, *Kiwa hirsuta*. *Environ Microbiol* 2008;**10(10)**: 2623-34.
- 690 Goffredi S. Indigenous ectosymbiotic bacteria associated with diverse hydrothermal vent
691 invertebrates. *Env Microbiol Rep* 2010;**2**: 479-88.
- 692 Good IJ. The population frequencies of species and the estimation of population parameters.
693 *Biometrika* 1953;**40**: 237-64.
- 694 Greene AC, Patel BKC, Sheehy AJ. *Deferribacter thermophilus* gen. nov., sp. nov., a novel
695 thermophilic manganese- and iron-reducing bacterium isolated from a petroleum reservoir.
696 *Int J Syst Evol Microbiol* 1997;**47**: 505-9.
- 697 Guri M, Durand L, Cuff-Gauchard V *et al.* Acquisition of epibiotic bacteria along the life
698 cycle of the hydrothermal shrimp *Rimicaris exoculata*. *ISME J* 2012;**6**: 597-609.
- 699 Hanson RS, Hanson TE. Methanotrophic bacteria. *Microbiol Rev* 1996;**60**: 439-71.

- 1
- 2
- 3 700 Hernandez-Avila I. Larval dispersal and life cycle in deep-water hydrothermal vents : the case
- 4
- 5 701 of *Rimicaris exoculata* and related species. PhD Thesis: Université de Bretagne
- 6
- 7 702 Occidentale, 2016, 235p.
- 8
- 9
- 10 703 Hügler M, Petersen JM, Dubilier N *et al.* Pathways of carbon and energy metabolism of the
- 11
- 12 704 epibiotic community associated with the deep-sea hydrothermal vent shrimp *Rimicaris*
- 13
- 14 705 *exoculata*. *PLoS One* 2011;**6**: e16018.
- 15
- 16 706 Inagaki F, Takai K, Nealson KH *et al.* *Sulfurovum lithotrophicum* gen. nov., sp. nov., a novel
- 17
- 18 707 sulfur-oxidizing chemolithoautotroph within the ϵ -Proteobacteria isolated from Okinawa
- 19
- 20 708 Trough hydrothermal sediments. *Int J Syst Evol Microbiol* 2004;**54**: 1477-82.
- 21
- 22
- 23 709 Jan C, Petersen JM, Werner J *et al.* The gill chamber epibiosis of deep sea shrimp *Rimicaris*
- 24
- 25 710 *exoculata*: an in depth metagenomic investigation and discovery of *Zetaproteobacteria*.
- 26
- 27 711 *Environ Microbiol* 2014;**16**: 2723-38.
- 28
- 29
- 30 712 Jannasch H. The chemosynthetic support of life and the microbial diversity at deep-sea
- 31
- 32 713 hydrothermal vents. *P Roy Soc Lond B Bio* 1985;**225**: 277-97.
- 33
- 34 714 Kanagawa T, Kamagata Y, Aruga S *et al.* Phylogenetic analysis of and oligonucleotide probe
- 35
- 36 715 development for Eikelboom type 021N filamentous bacteria isolated from bulking
- 37
- 38 716 activated sludge. *Appl Environ Microbiol* 2000;**66**: 5043-52.
- 39
- 40
- 41 717 Katoh K, Misawa K, Kuma Ki *et al.* MAFFT: a novel method for rapid multiple sequence
- 42
- 43 718 alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;**30**: 3059-66.
- 44
- 45 719 Kearse M, Moir R, Wilson A *et al.* Geneious Basic: an integrated and extendable desktop
- 46
- 47 720 software platform for the organization and analysis of sequence data. *Bioinformatics*
- 48
- 49 721 2012;**28**: 1647-9.
- 50
- 51
- 52 722 Knittel K, Kuever J, Meyerdierks A *et al.* *Thiomicrospira arctica* sp. nov. and *Thiomicrospira*
- 53
- 54 723 *psychrophila* sp. nov., psychrophilic, obligately chemolithoautotrophic, sulfur-oxidizing
- 55
- 56 724 bacteria isolated from marine Arctic sediments. *Int J Syst Evol Microbiol* 2005;**55**: 781-6.
- 57
- 58
- 59
- 60

- 725 Komai T, Segonzac M. Taxonomic review of the hydrothermal vent shrimp genera *Rimicaris*
726 Williams & Rona and *Chorocaris* Martin & Hessler (Crustacea: Decapoda: Caridea:
727 Alvinocarididae). *J Shellfish Res* 2008;**27**: 21-41.
- 728 Lane D. 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (eds.) *Nucleic acid*
729 *techniques in bacterial systematics*. Chichester-New York-Brisbane-Totonto-Singapore:
730 Wiley-Interscience, 1991, 115-48.
- 731 Lin X, Wakeham S, Putnam I *et al.* Comparison of vertical distributions of prokaryotic assemblages
732 in the anoxic Cariaco Basin and Black Sea by use of fluorescence in situ hybridization. *Appl Env*
733 *Microbiol* 2006;**72**: 2679-90.
- 734 MacPherson E, Jones W, Segonzac M. A new lobster family of Galattheoidea (Crustacea,
735 Decapoda, Anomura) from the hydrothermal vents of the Pacific-Antarctic Ridge.
736 *Zoosystema* 2005;**27**(4): 709-22.
- 737 Manz W, Amann R, Wagner M *et al.* Phylogenetic oligodeoxynucleotide probes for the major
738 subclasses of Proteobacteria: problems and solutions. *Syst Appl Microbiol* 1992;**15**: 593-600.
- 739 Martins I, Colaço A, Dando P *et al.* Size-dependent variations on the nutritional pathway of
740 *Bathymodiulus azoricus* demonstrated by a C-flux model. *Ecol Model* 2008;**217**: 59-71.
- 741 Mino S, Kudo H, Arai T *et al.* *Sulfurovum aggregans* sp. nov., a hydrogen-oxidizing,
742 thiosulfate-reducing chemolithoautotroph within the Epsilonproteobacteria isolated from a
743 deep-sea hydrothermal vent chimney, and an emended description of the genus
744 *Sulfurovum*. *Int J Syst Evol Microbiol* 2014;**64**: 3195-201.
- 745 Nakagawa S, Takaki Y, Shimamura S *et al.* Deep-sea vent ϵ -proteobacterial genomes provide
746 insights into emergence of pathogens. *Proc Natl Acad Sci* 2007;**104**: 12146-50.
- 747 Nye V, Copley J, Plouviez S. A new species of *Rimicaris* (Crustacea: Decapoda: Caridea:
748 Alvinocarididae) from hydrothermal vent fields on the Mid-Cayman Spreading Centre,
749 Caribbean. *J Mar Biol Assoc UK* 2011;**92**(5): 1057-72.

- 1
- 2
- 3 750 Petersen JM, Ramette A, Lott C *et al.* Dual symbiosis of the vent shrimp *Rimicaris exoculata*
- 4
- 5 751 with filamentous *gamma* and *epsilon*proteobacteria at four Mid Atlantic Ridge
- 6
- 7 752 hydrothermal vent fields. *Environ Microbiol* 2010;**12**: 2204-18.
- 8
- 9 753 Polz MF, Cavanaugh CM. Dominance of one bacterial phylotype at a Mid-Atlantic Ridge
- 10
- 11 754 hydrothermal vent site. *Proc Natl Acad Sci* 1995;**92**: 7232-6.
- 12
- 13 755 Polz M, Robinson J, Cavanaugh C *et al.* Trophic ecology of massive shrimp aggregations at a
- 14
- 15 756 mid-Atlantic Ridge hydrothermal vent site. *Limnol Oceanogr* 1998;**43**(7): 1631-8.
- 16
- 17 757 Pond D, Dixon D, Bell M *et al.* Occurrence of 16:2(n-4) and 18:2 (n-4) fatty acids in the lipids
- 18
- 19 758 of the hydrothermal vent shrimps *Rimicaris exoculata* and *Alvinocaris markensis*:
- 20
- 21 759 nutritional and trophic implications. *Mar Ecol Progr Ser* 1997;**156**: 167-74.
- 22
- 23 760 Ponsard J, Cambon-Bonavita M-A, Zbinden M *et al.* Inorganic carbon fixation by
- 24
- 25 761 chemosynthetic ectosymbionts and nutritional transfers to the hydrothermal vent host-
- 26
- 27 762 shrimp *Rimicaris exoculata*. *ISME J* 2013;**7**: 96-109.
- 28
- 29
- 30
- 31 763 Qiu X, Wu L, Huang H *et al.* Evaluation of PCR-generated chimeras, mutations, and
- 32
- 33 764 heteroduplexes with 16S rRNA gene-based cloning. *Appl Environ Microbiol* 2001;**67**: 880-
- 34
- 35 765 7.
- 36
- 37
- 38 766 Rieley G, Van Dover CL, Hedrick DB *et al.* Trophic ecology of *Rimicaris exoculata*: a
- 39
- 40 767 combined lipid abundance/stable isotope approach. *Mar Biol* 1999;**133**: 495-9.
- 41
- 42 768 Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing
- 43
- 44 769 phylogenetic trees. *Mol Biol Evol* 1987;**4**: 406-25.
- 45
- 46
- 47 770 Schloss PD, Handelsman J. Introducing DOTUR, a computer program for defining
- 48
- 49 771 operational taxonomic units and estimating species richness. *Appl Environ Microbiol*
- 50
- 51 772 2005;**71**: 1501-6.
- 52
- 53 773 Schloss PD, Westcott SL, Ryabin T *et al.* Introducing mothur: open-source, platform-
- 54
- 55 774 independent, community-supported software for describing and comparing microbial
- 56
- 57
- 58
- 59
- 60

- communities. *Appl Environ Microbiol* 2009;**75**: 7537-41.
- Segonzac M, de Saint Laurent M, Casanova B. Enigma of the trophic adaptation of the shrimp Alvinocarididae in hydrothermal areas along the Mid-Atlantic Ridge. *Cah Biol Mar* 1993;**34**: 535-71.
- Shank T, Black M, Halanich K *et al*. Miocene radiation of deep-sea hydrothermal vent shrimp (Caridea: Bresiliidae): Evidence from mitochondrial cytochrome oxidase subunit I. *Molecular Phylogenetics and Evolution* 1999;**13**: 244-54.
- Southward EC. The morphology of bacterial symbioses in the gills of mussels of the genera *Adipicola* and *Idas* (Bivalvia: Mytilidae). *J Shellfish Res* 2008;**27**: 139-46.
- Stahl D, Amann R. Development and application of nucleic acid probes. In: Stackebrandt E, Goodfellow M (eds.) *Nucleic acid techniques in bacterial systematics*. Chichester - New York - Brisbane - Toronto - Singapore: Wiley-Interscience, 1991.
- Streit K, Bennett SA, Van Dover CL *et al*. Sources of organic carbon for Rimicaris hybisae: Tracing individual fatty acids at two hydrothermal vent fields in the Mid-Cayman rise. *Deep-Sea Res Part I* 2015;**100**: 13-20.
- Takai K, Kobayashi H, Nealson KH *et al*. *Deferribacter desulfuricans* sp. nov., a novel sulfur-, nitrate- and arsenate-reducing thermophile isolated from a deep-sea hydrothermal vent. *Int J Syst Evol Microbiol* 2003;**53**: 839-46.
- Tanaka N, Romanenko LA, Iino T *et al*. *Cocleimonas flava* gen. nov., sp. nov., a gammaproteobacterium isolated from sand snail (*Umbonium costatum*). *Int J Syst Evol Microbiol* 2011;**61**: 412-6.
- Tavormina PL, Hatzepichler R, McGlynn S *et al*. *Methyloprofundus sedimenti* gen. nov., sp. nov., an obligate methanotroph from ocean sediment belonging to the 'deep sea-1' clade of marine methanotrophs. *Int J Syst Evol Microbiol* 2015;**65**: 251-9.
- Teixeira S, Olu K, Decker C *et al*. High connectivity across the fragmented chemosynthetic

- 800 ecosystems of the deep Atlantic Equatorial Belt: efficient dispersal mechanisms or
801 questionable endemism? *Mar Ecol* 2013;**22**: 4663–80.
- 802 Tsuchida S, Suzuki Y, Fujiwara Y *et al*. Epibiotic association between filamentous bacteria
803 and the vent-associated galatheid crab, *Shinkaia crosnieri* (Decapoda: Anomura). *J Mar*
804 *Biol Assoc UK* 2011;**9**(11): 23-32.
- 805 Unz RF, Head IM. *Thiothrix*. In: Whitman B (ed.) *Bergey's Manual of Systematics of Archaea*
806 *and Bacteria*, DOI 10.1002/9781118960608.gbm01229: John Wiley & Sons, Ltd, 2015, 1–
807 16.
- 808 Van Dover CL, Fry B. Microorganisms as food resources at deep-sea hydrothermal vents.
809 *Limnol Oceanogr* 1994;**39**: 51-7.
- 810 Vereshchaka AL, Kulagin DN, Lunina AA. Phylogeny and new classification of
811 hydrothermal vent and seep shrimps of the family Alvinocarididae (Decapoda). *PloS One*
812 2015;**10**: e0129975.
- 813 Watabe H, Hashimoto J. A New Species of the genus *Rimicaris* (Alvinocarididae: Caridea:
814 Decapoda) from the active hydrothermal vent field, “Kairei Field,” on the Central Indian
815 Ridge, the Indian Ocean. *Zool Sci* 2002;**19**: 1167-74.
- 816 Williams AB, Rona PA. Two new caridean shrimps (Bresiliidae) from a hydrothermal field
817 on the Mid-Atlantic Ridge. *J Crustacean Biol* 1986;**6**: 446-62.
- 818 Wirsén C, Jannasch HW, Molyneux S. Chemosynthetic microbial activity at Mid-Atlantic
819 Ridge hydrothermal vent sites. *J Geophys Res* 1993;**98** (B6): 9693-703.
- 820 Wright ES, Yilmaz LS, Noguera DR. DECIPHER, a search-based approach to chimera
821 identification for 16S rRNA sequences. *Appl Environ Microbiol* 2012;**78**: 717-25.
- 822 Yang J-S, Lu B, Chen D-F *et al*. When Did Decapods Invade Hydrothermal Vents? Clues
823 from the Western Pacific and Indian Oceans. *Molecular Biology and Evolution* 2013;**30**:
824 305-9.

- 825 Zbinden M, Cambon-Bonavita M-A. Occurrence of *Deferribacterales* and
826 *Entomoplasmatales* in the deep-sea Alvinocarid shrimp *Rimicaris exoculata* gut. *FEMS*
827 *Microbiol Ecol* 2003;**46**: 23-30.
- 828 Zbinden M, Le Bris N, Gaill F *et al*. Distribution of bacteria and associated minerals in the
829 gill chamber of the vent shrimp *Rimicaris exoculata* and related biogeochemical processes.
830 *Mar Ecol Prog Ser* 2004;**284**: 237-51.
- 831 Zbinden M, Shillito B, Le Bris N *et al*. New insights on the metabolic diversity among the
832 epibiotic microbial community of the hydrothermal shrimp *Rimicaris exoculata*. *J Exp*
833 *Mar Biol Ecol* 2008;**359**: 131-40.
- 834
- 835

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

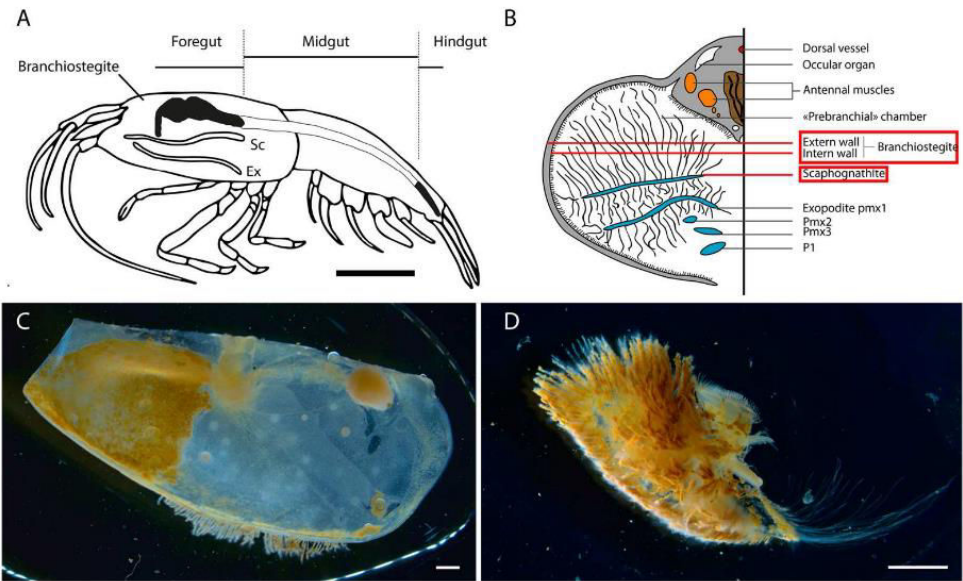


Figure 1: Scheme of *R. exoculata* and SEM micrographs of *R. chacei* showing the location of the two symbiotic bacterial populations. (A) Lateral view of *R. exoculata*, showing the location of scaphognathite (Sc) and exopodite (Ex) of the first maxillipeds, as well as the three part of the gut (foregut, midgut and hindgut) (modified from Williams and Rona 1986 and Ponsard et al. 2013). (B) Cross section of *R. exoculata* cephalothorax (modified from Segonzac et al. 1993). Bacteria occur on the inner side of the branchiostegite and on the bacteriophore setae of the scaphognathites and exopodites of the first maxillipeds. (C) Binocular observation of inner face of a branchiostegite and (D) of a scaphognathite of *Rimicaris chacei*. Scale bars: A = 1 cm, C and D = 1.5 cm.

350x212mm (150 x 150 DPI)

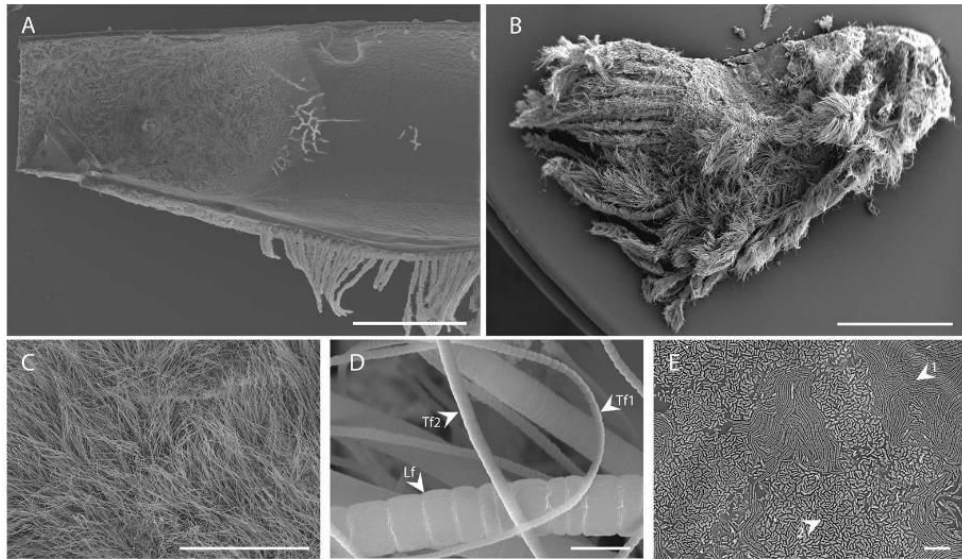


Figure 2: SEM observations of the bacterial colonization in the cephalothorax of *Rimicaris chacei*. A) Overview of the inner side of the branchiostegite showing colonized (on the left) and uncolonized (on the right) area. B) Overview of the dorsal side of the scaphognathite. C) Enlargement on filamentous bacterial mat of the inner side of the branchiostegite. D) High magnification of the three filamentous morphotypes (large filaments (Lf), thin filaments type 1 (Tf1) and type 2 (Tf2)). E) The two bacilli morphotypes (type 1 and 2) on the inner side of the branchiostegite. Scale bars A, B = 1 mm, C = 400 µm, D, E = 5 µm.

350x202mm (150 x 150 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

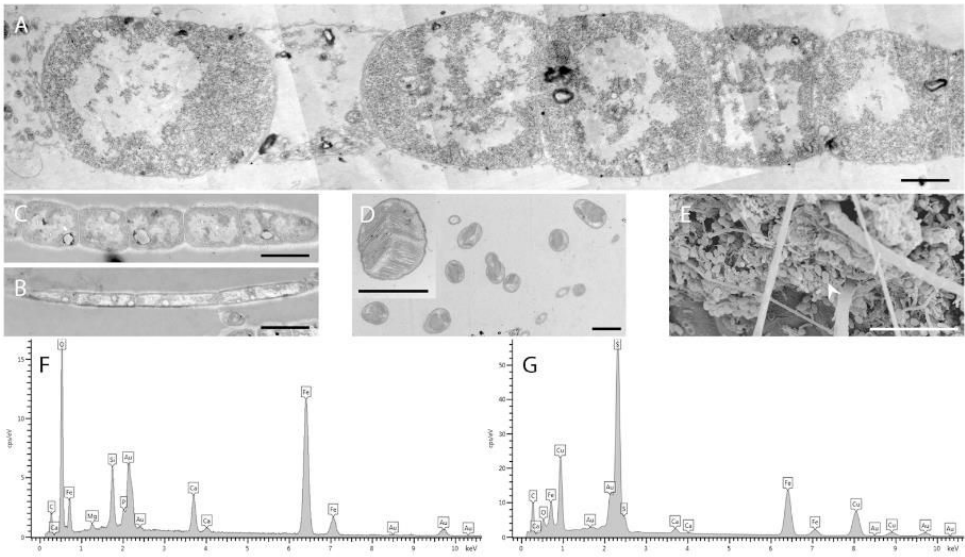


Figure 3: TEM and SEM observations of the different bacterial morphotypes and X-ray analysis of the associated minerals. A) TEM view of a large filament. B) TEM view of a type 1 thin filament. C) TEM view of a type 2 thin filament. D) TEM view of methanotrophic like bacteria. E) SEM observation of methanotrophic like bacteria morphotype (arrow heads). (F-G) Elemental X-ray microanalysis of mineral deposits on branchiostegites and scaphognathites. Most of this rusty mineral deposit is composed of iron and oxygen, supposedly as iron oxide (F). Fe/Cu sulphide particles are also present in places (G). Scale bars A, B, C, D and insert in D = 1 μ m, E = 10 μ m.

330x189mm (150 x 150 DPI)

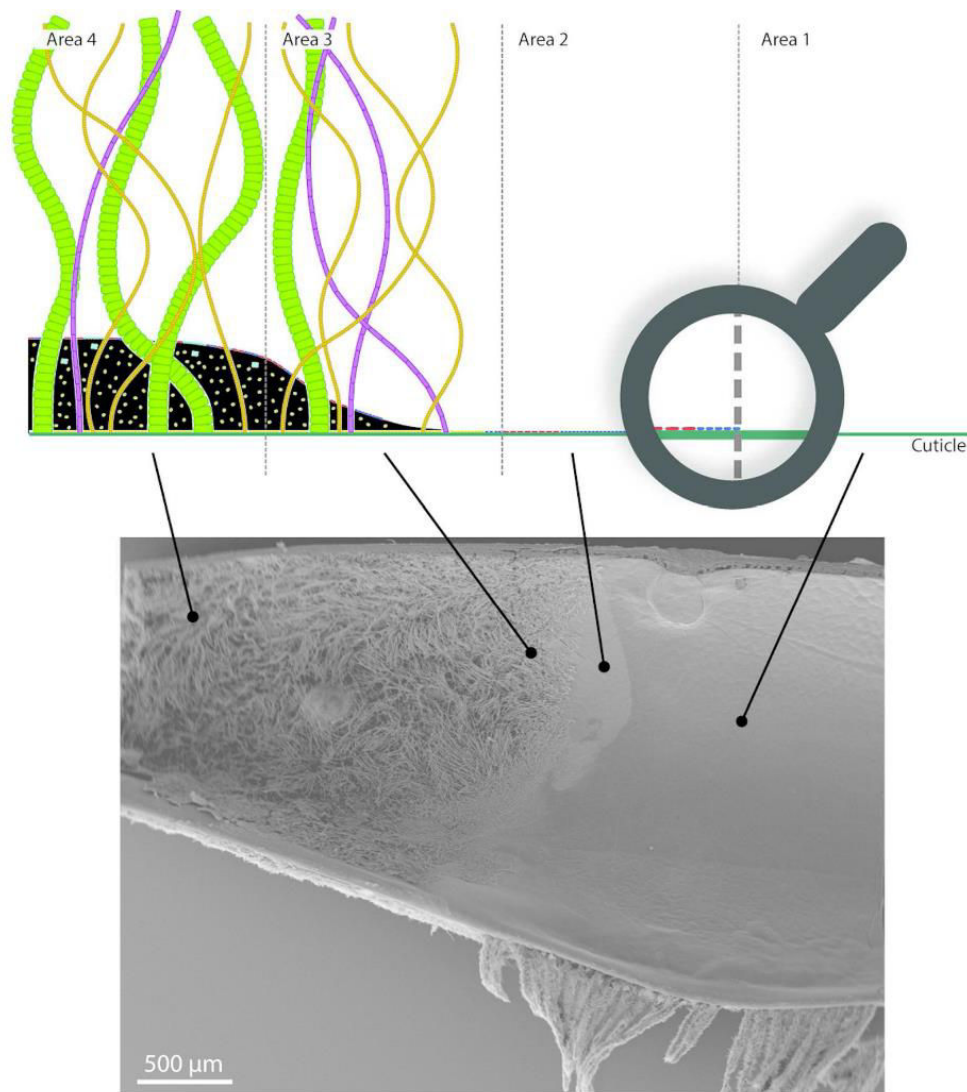


Figure 4: Bacterial colonisation area on the inner side of the branchiostegite. A) Schematic representation of the 4 bacterial colonization areas (proportions of the different bacterial morphotypes are respected). The green line under the drawing is the shrimp cuticle. Rods type 1 are colored in red, rods type 2 in blue, large filaments in green, thin filaments type 1 in orange, thin filaments type 2 in purple and methanotrophic like bacteria in turquoise. B) SEM micrograph showing the inner side of a branchiostegite and the 4 areas described in the text and schematized above.

330x374mm (150 x 150 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

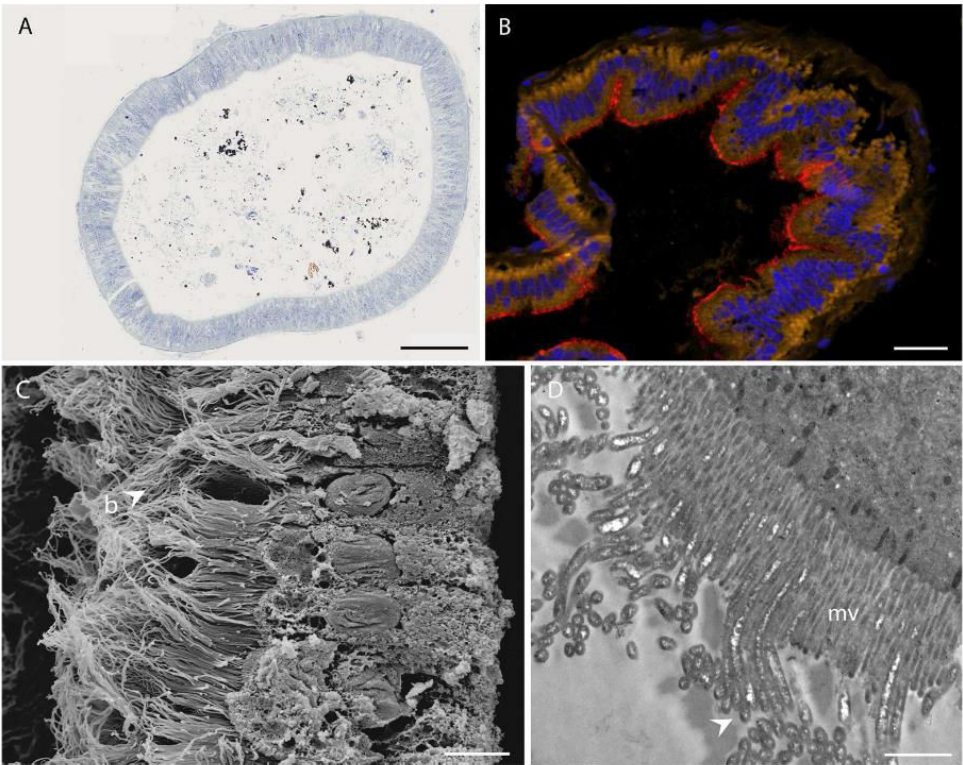


Figure 5: *Rimicaris chacei* midgut. A) Photonic observation of a semi-thin section of the digestive tract (hindgut), showing black and brown mineral particles, as well as organic matter stained by the toluidine blue. B) FISH observations of midgut transversal section stained with DAPI (blue), and hybridised with Eubacteria general probe Eub338-Cy5 (red) and autofluorescence of intestinals cells (yellow). C) SEM observation of bacterial mat (b) on intestinal wall cells. D) TEM observation of filamentous bacteria (arrows) inserted between microvilli (mv) of intestinal cells. Scale bars: A = 10 µm, B = 1 µm, C = 50 µm, D = 10 µm.

330x261mm (150 x 150 DPI)

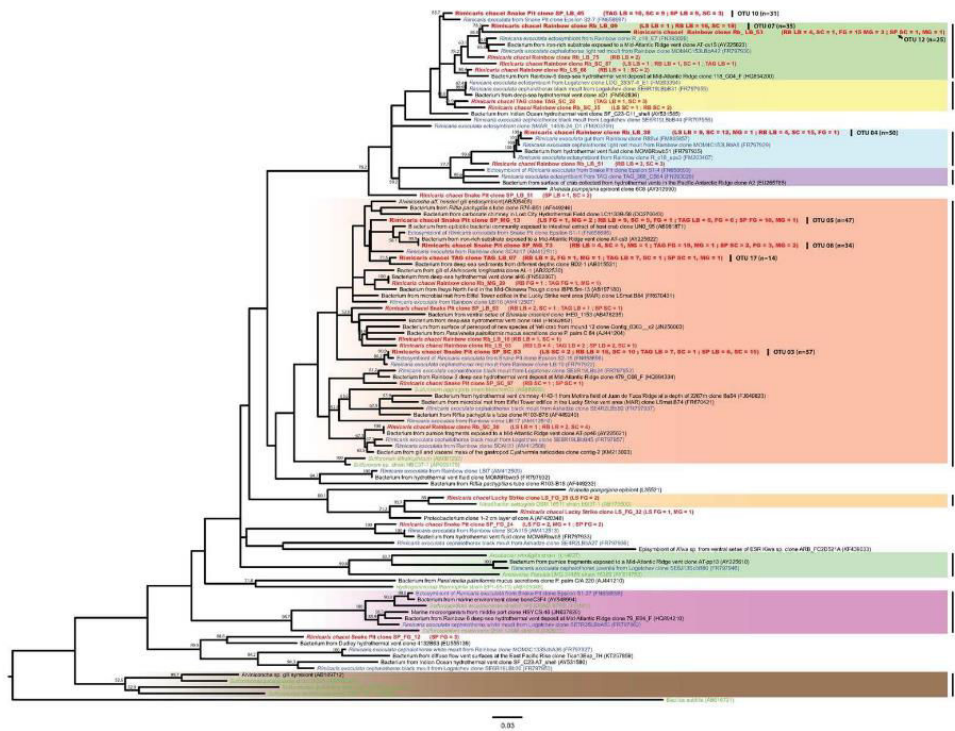


Figure 6: Epsilonproteobacteria 16S rRNA gene phylogenetic tree calculated on 624bp, with neighbor joining method, the GTR model (Lanave et al. 1984) and 1000 bootstrap resampling. Bacterial sequences of *Rimicaris chacei* found in this study are labelled in red, *Rimicaris exoculata* symbionts in blue, and cultivated microbial strains in green. The number of clone sequences per site and per sample are in brackets (Sites: RB: Rainbow, SP: Snake Pit, LS: Lucky Strike. Samples: LB: branchiostegites, SC: scaphognathites, MG: midgut, FG: foregut).

155x116mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

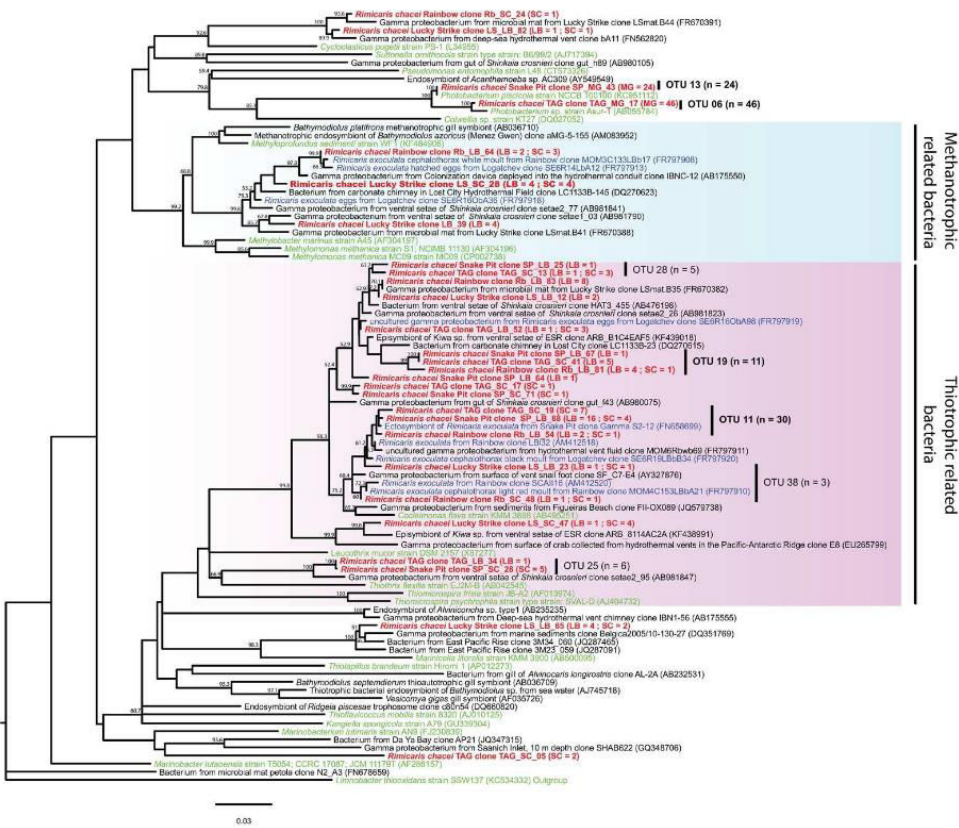


Figure 7: Gammaproteobacteria 16S rRNA gene phylogenetic tree calculated on 685bp, with neighbor joining method, the GTR model (Lanave et al. 1984) and 1000 bootstrap resampling. Bacterial sequences associated with *Rimicaris chacei* are labelled in red, *Rimicaris exoculata* symbionts in blue, and cultivated microbial strains in green. The number of clone sequences per vent site and per sample are in brackets (abbreviations for sites and samples like in Fig. 6).

172x146mm (300 x 300 DPI)

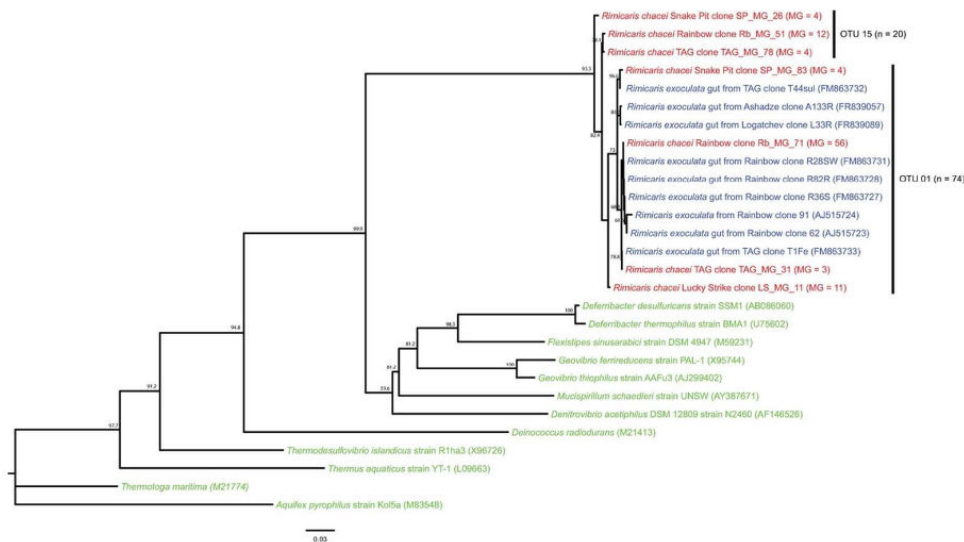


Figure 8: Deferribacteres 16S rRNA gene phylogenetic tree calculated on 764bp, with neighbor joining method, the GTR model (Lanave et al. 1984) and 1000 bootstrap resampling. Bacterial sequences of *Rimicaris chacei* found in this study are labelled in red, *Rimicaris exoculata* symbionts in blue, and cultivated microbial strains in green. The number of clone sequences per vent site and per sample are in brackets (abbreviations for sites and samples like in Fig. 6).

113x63mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

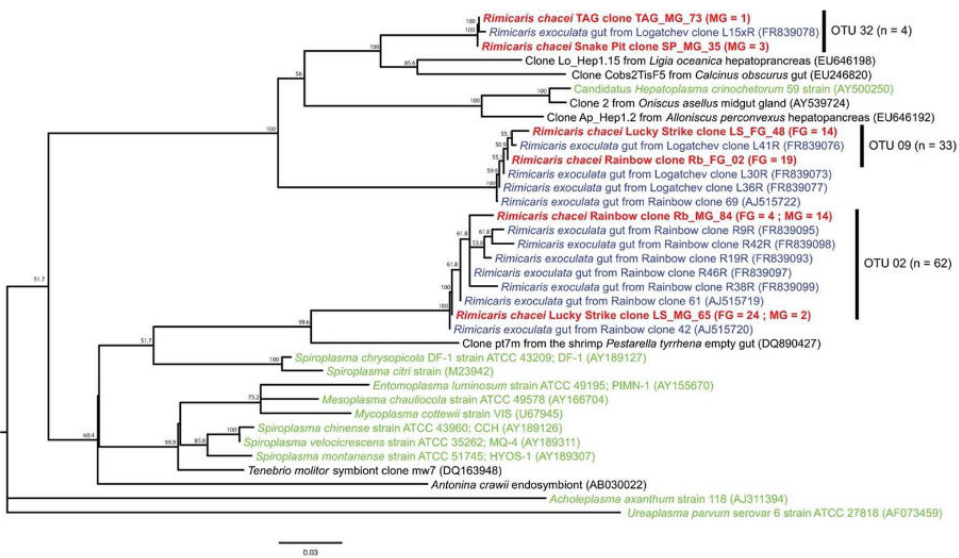


Figure 9: Mollicutes 16S rRNA gene phylogenetic tree calculated on 657bp, with neighbor joining method, the GTR model (Lanave et al. 1984) and 1000 bootstrap resampling. Bacterial sequences of *Rimicaris chacei* found in this study are labelled in red, *Rimicaris exoculata* symbionts in blue, and cultivated microbial strains in green. The number of clone sequences per vent site and per sample are in brackets (abbreviations for sites and samples like in Fig. 6).

117x67mm (300 x 300 DPI)

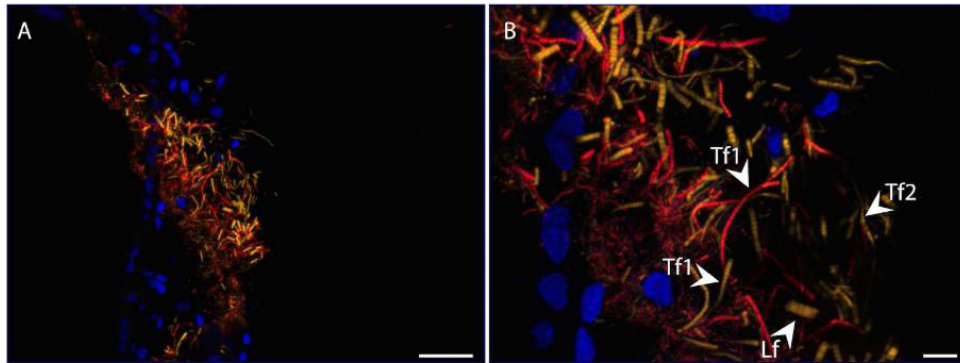


Figure 10: FISH observations of a branchiostegite. A) Overview of branchiostegite stained with DAPI (blue), and hybridised with an Epsilonproteobacteria general probe Epsy549-Cy3 (yellow) and a Gammaproteobacteria general probe Gam42-Cy5 (red) probes. B) Higher magnification of the different morphotypes. Arrow point to filamentous morphotype described previously (Lf: large filaments, Tf1: thin filaments type 1 and Tf2: thin filament type 2). Scale bars: B = 50 μ m, C = 10 μ m.

330x123mm (150 x 150 DPI)

1
2
3 1 **Figure legendes**
4
5 2
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

3 **Figure 1:** Scheme of *R. exoculata* and SEM micrographs of *R. chacei* showing the location of
4 the two symbiotic bacterial populations. (A) Lateral view of *R. exoculata*, showing the
5 location of scaphognathite (Sc) and exopodite (Ex) of the first maxilliped, as well as the three
6 part of the gut (foregut, midgut and hindgut) (modified from Williams and Rona 1986 and
7 Ponsard *et al.* 2013). B) Cross section of *R. exoculata* cephalothorax (modified from
8 Segonzac *et al.* 1993). Bacteria occur on the inner side of the branchiostegites and on the
9 bacteriophore setae of the scaphognathites and exopodites of the first maxillipeds. C)
10 Binocular observation of inner face of a branchiostegite and D) of a scaphognathite of
11 *Rimicaris chacei*. Scale bars: A = 1 cm, C and D = 1.5 cm.
12

13 **Figure 2:** SEM observations of the bacterial colonization in the cephalothorax of *Rimicaris*
14 *chacei*. A) Overview of the inner side of the branchiostegite showing colonized (on the left)
15 and uncolonized (on the right) area. B) Overview of the dorsal side of the scaphognathite. C)
16 Enlargment on filamentous bacterial mat of the inner side of the branchiostegite. D) High
17 magnification of the three filamentous morphotypes (large filaments (Lf), thin filaments type
18 1 (Tf1) and type 2 (Tf2)). E) The two bacilli morphotypes (type 1 and 2) on the inner side of
19 the branchiostegite. Scale bars A, B = 1 mm, C = 400 µm, D, E = 5 µm.
20

21 **Figure 3:** TEM and SEM observations of the different bacterial morphotypes and X-ray
22 analysis of the associated minerals. A) TEM view of a large filament. B) TEM view of a type
23 1 thin filament. C) TEM view of a type 2 thin filament. D) TEM view of methanotrophic like
24 bacteria. E) SEM observation of methanotrophic like bacteria morphotype (arrowheads). (F-
25 G) Elemental X-ray microanalysis of mineral deposits on branchiostegites and

scaphognathites. Most of this rusty mineral deposit is composed of iron and oxygen, supposedly as iron oxide (F). Fe/Cu sulphide particles are also present in places (G). Scale bars A, B, C, D and insert in D = 1 μm , E = 10 μm .

Figure 4: Bacterial colonization area on the inner side of the branchiostegite. A) Schematic representation of the 4 bacterial colonization areas (relative sizes of the different bacterial morphotypes are respected). The green line under the drawing is the shrimp cuticle. Rods type 1 are colored in red, rods type 2 in blue, large filaments in green, thin filaments type 1 in orange, thin filaments type 2 in purple and methanotrophic like bacteria in turquoise. B) SEM micrograph showing the inner side of a branchiostegite and the 4 areas described in the text and schematized above.

Figure 5: *Rimicaris chacei* midgut. A) Photonic observation of a semi-thin section of the digestive tract (hindgut), showing black and brown mineral particles, as well as organic matter stained by the toluidine blue. B) FISH observations of midgut transversal section stained with DAPI (blue), and hybridized with Eubacteria general probe Eub338-Cy5 (red) and autofluorescence of intestinal cells (yellow). C) SEM observation of bacterial mat (b) on intestinal wall cells. D) TEM observation of filamentous bacteria (arrows) inserted between microvilli (mv) of intestinal cells. Scale bars: A = 10 μm , B = 1 μm , C = 50 μm , D = 10 μm .

Figure 6: *Epsilonproteobacteria* 16S rRNA gene phylogenetic tree calculated on 624bp, with neighbor joining method, the GTR model (Lanave *et al.* 1984) and 1000 bootstrap resampling. Bacterial sequences of *Rimicaris chacei* found in this study are labelled in red, *Rimicaris exoculata* symbionts in blue, and cultivated microbial strains in green. The number of clone sequences per site and per sample are in brackets (Sites: RB: Rainbow, SP: Snake

1
2
3 51 Pit, LS: Lucky Strike. Samples: LB: branchiostegites, SC: scaphognathites, MG: midgut, FG:
4
5 52 foregut).
6
7
8 53
9
10 54 **Figure 7:** *Gammaproteobacteria* 16S rRNA gene phylogenetic tree calculated on 685bp,
11
12 55 with neighbor joining method, the GTR model (Lanave *et al.* 1984) and 1000 bootstrap
13
14 56 resampling. Bacterial sequences associated with *Rimicaris chacei* are labelled in red,
15
16 57 *Rimicaris exoculata* symbionts in blue, and cultivated microbial strains in green. The number
17
18 58 of clone sequences per vent site and per sample are in brackets (abbreviations for sites and
19
20 59 samples like in Fig. 6).
21
22
23 60
24
25 61 **Figure 8:** *Deferribacteres* 16S rRNA gene phylogenetic tree calculated on 764bp, with
26
27 62 neighbor joining method, the GTR model (Lanave *et al.* 1984) and 1000 bootstrap
28
29 63 resampling. Bacterial sequences of *Rimicaris chacei* found in this study are labelled in red,
30
31 64 *Rimicaris exoculata* symbionts in blue, and cultivated microbial strains in green. The number
32
33 65 of clone sequences per vent site and per sample are in brackets (abbreviations for sites and
34
35 66 samples like in Fig. 6).
36
37
38 67
39
40 68 **Figure 9:** *Mollicutes* 16S rRNA gene phylogenetic tree calculated on 657bp, with neighbor
41
42 69 joining method, the GTR model (Lanave *et al.* 1984) and 1000 bootstrap resampling.
43
44 70 Bacterial sequences of *Rimicaris chacei* found in this study are labelled in red, *Rimicaris*
45
46 71 *exoculata* symbionts in blue, and cultivated microbial strains in green. The number of clone
47
48 72 sequences per vent site and per sample are in brackets (abbreviations for sites and samples
49
50 73 like in Fig. 6).
51
52
53 74
54
55
56
57
58
59
60

Figure 10: FISH observations of a branchiostegite. A) Overview of branchiostegite stained with DAPI (blue), and hybridized with an *Epsilonproteobacteria* general probe Epsy549-Cy3 (yellow) and a *Gammaproteobacteria* general probe Gam42-Cy5 (red) probes. B) Higher magnification of the different morphotypes. Arrows point to filamentous morphotypes described previously (Lf: large filaments, Tf1: thin filaments type 1 and Tf2: thin filament type 2). Scale bars: B = 50 μ m, C = 10 μ m.

Supplementary data

Figure S1: Rarefaction curves of cephalothoracic clone libraries of *R. chacei* samples.

Figure S2: Rarefaction curves of digestive tract clone libraries of *R. chacei* samples.

Figure S3 : Dendogramme of all analyzed samples based on Bray-Curtis beta diversity on left, and taxonomy of each sequences associated to each samples on right.

Table 1: Number of specimens used per site and for each experiment in this study. LB = Branchiostegites, SC = Scaphognathites, FG = Foregut, MG = Midgut.

	TEM			SEM			16S molecular biology				FISH			
	LB	SC	MG	LB	SC	MG	LB	SC	FG	MG	LB	SC	FG	MG
Lucky Strike	1	1	1	1	1	1	1	1	1	1				
Rainbow	1	1	1	1	1	1	3	3	3	2	2	2	1	1
TAG	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Snake Pit	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 2 : Fluorescent probes (the probe sequences have been compared to our sequences to check their specificity and determine their mismatches)

Specificity	Probe name	Sequence (5'-3')	Fluorescent dye	% Formamide	References
Archaea	Arch915	GTGCTCCCCGCCAATT CCT	Cy3	10 - 20 - 30 - 40	Stahl and Amann 1991
Eubacteria	Eub338	GCTGCCTCCCGTAGGAG T	Cy3 or Cy5	10 - 20 - 30 - 40	Amann <i>et al.</i> 1990
Gamma proteobacteria	GAM42a	GCCTTCCCACATCGTTT	Cy3 or Cy5	10 - 20 - 30 - 40	Manz <i>et al.</i> 1992
Epsilon proteobacteria	EPSY549	CAGTGATTCCGAGTAAC G	Cy3 or Cy5	10 - 20 - 30 - 40	Lin <i>et al.</i> 2006

Table 3: Cell sizes of the various bacterial morphotype observed and they relative abundance (- : absent, ++: abundant, +++ very abundant) for each hydrothermal vent site. Values are given in $\mu\text{m} \pm$ standard deviation

Morphotype	Diameter	Heigh	Lucky Strike	Rainbow	TAG	Snake Pit
Large filament	$2,42 \pm 0,58$ (n=81)	$2,57 \pm 0,75$ (n=57)	+++	+++	-	+++
Thin filament (type 1)	$0,80 \pm 0,07$ (n=114)	$0,70 \pm 0,12$ (n=109)	+++	+++	+++	+++
Thin filament (type 2)	$1,12 \pm 0,13$ (n=27)	$5,61 \pm 1,07$ (n=43)	++	++	++	++
Thick rods (type 1)	$0,26 \pm 0,04$ (n=89)	$1,6 \pm 0,33$ (n=108)	+++	+++	+++	+++
Thin rods (type 2)	$0,28 \pm 0,06$ (n=70)	$0,94 \pm 0,17$ (n=70)	+++	+++	+++	+++
Coccobacilli	$0,66 \pm 0,09$ (n=91)	$1,07 \pm 0,17$ (n=124)	+++	+++	-	-

Table 4 : Clone library results (based on partial 16S rRNA sequences)

taxon	Total clones number per lineage	ChoroC40Bb13PL05LSTEF_TD	ChoroC40Bb13PL05LSTEF_EST	ChoroC08Bb13PL08RBF5_TD	ChoroC04Bb13PL07RBTE_R_TD	ChoroC08Bb13PL08RBF5_EST	ChoroC07Bc14PL10TAG_T_D	ChoroC01Bc14PL05SP_TD	ChoroC09Bc14PL10TAG_EST	ChoroC04Bc14PL05SP_ES_T	ChoroC40Bb13PL05LSTEF_SC1	ChoroC40Bb13PL05LSTEF_LB1	ChoroC08Bb13PL08RBF5_SC1	ChoroC01MOM04PL05RB_LB1	ChoroC08Bb13PL08RBF5_LB1	ChoroC01MOM04PL05RB_SC1	ChoroC04Bb13PL07RBTE_R_SC1	ChoroC04Bb13PL07RBTE_R_LB1	ChoroC07Bc14PL10TAG_SC1	ChoroC01Bc14PL05SP_LB1	ChoroC07Bc14PL10TAG_LB1	ChoroC01Bc14PL05SP_SC1	Total OTUs 97% (« Forwards » data)
<i>Epsilonproteobacteria</i>	890	17	17	2	8	44	4	11	62	87	44	38	90	34	83	36	68	58	28	40	62	57	45
<i>Gammaproteobacteria</i>	352	1					74	45	1		34	35	3	1	4	3	4	27	50	41	7	22	23
<i>Deferribacteres</i>	211	23		70	81		13	24															2
<i>Mollicutes (Tenericutes)</i>	187	43	67	19		43	1	12		2													4
<i>Alphaproteobacteria</i>	79				2	1					6	6		5		4	20	5	3	2	19	6	8
<i>Bacteroidetes</i>	42	3	1		1			1		2	3	5		1	5			1	8	5	5	1	9
<i>Candidate_division_OD1</i>	31							1	27	1										2			3
<i>BD1-5</i>	20								1	1	1				1		2	3	6	5			7
<i>Deltaproteobacteria</i>	9					1				2	3	3											3
<i>Zetaproteobacteria</i>	5								1		1						1				2		1
<i>Lentisphaerae</i>	2								2														1
<i>Betaproteobacteria</i>	1	1																					1
<i>Firmicutes</i>	1					1																	1
Total clone number per cloning experiment	1830	88	85	91	92	90	92	94	94	95	92	87	93	41	93	43	95	94	95	95	95	86	

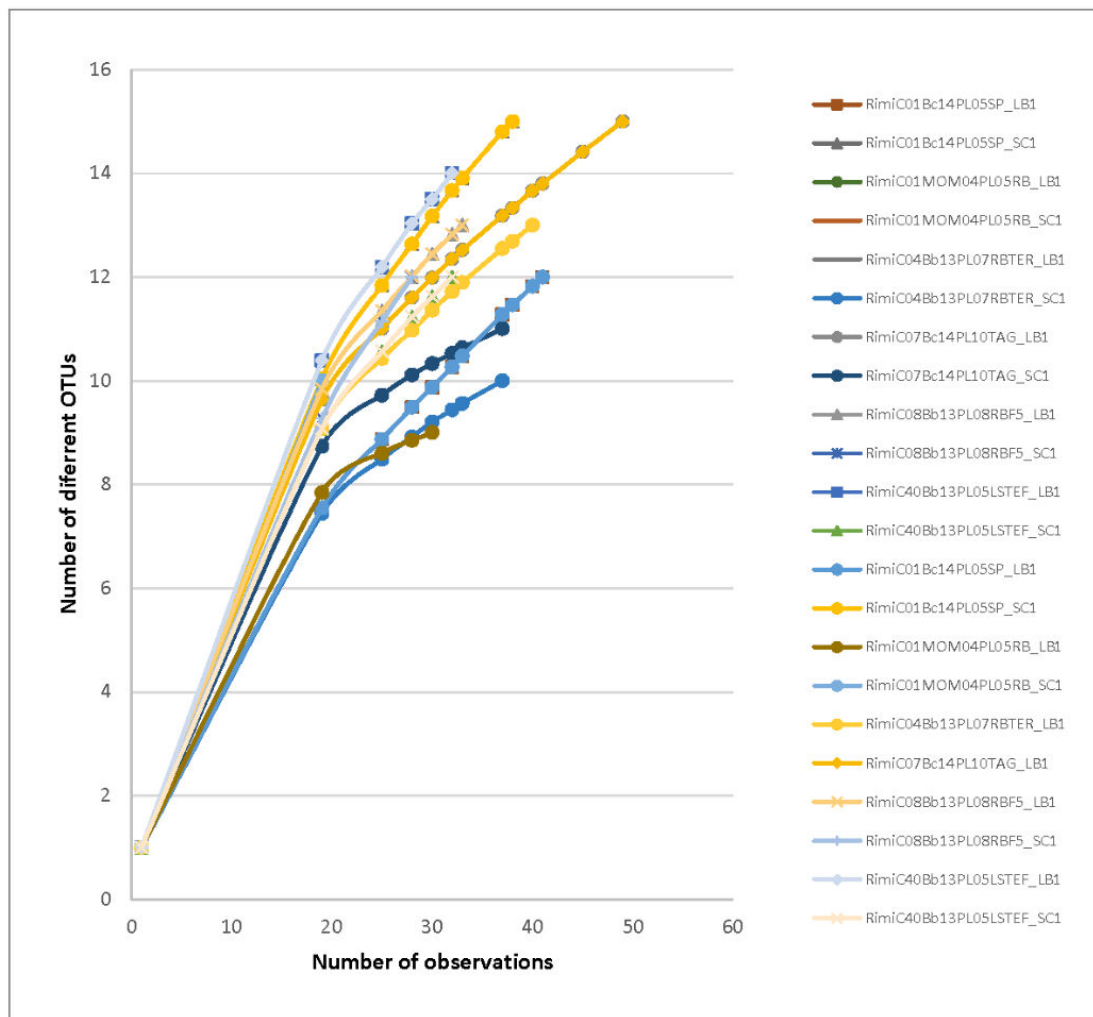
Figure S1: Rarefaction curves of cephalothoracic clone libraries of *R. chacei* samples

Figure S2: Rarefaction curves of digestive tract clone libraries of *R. chacei* samples

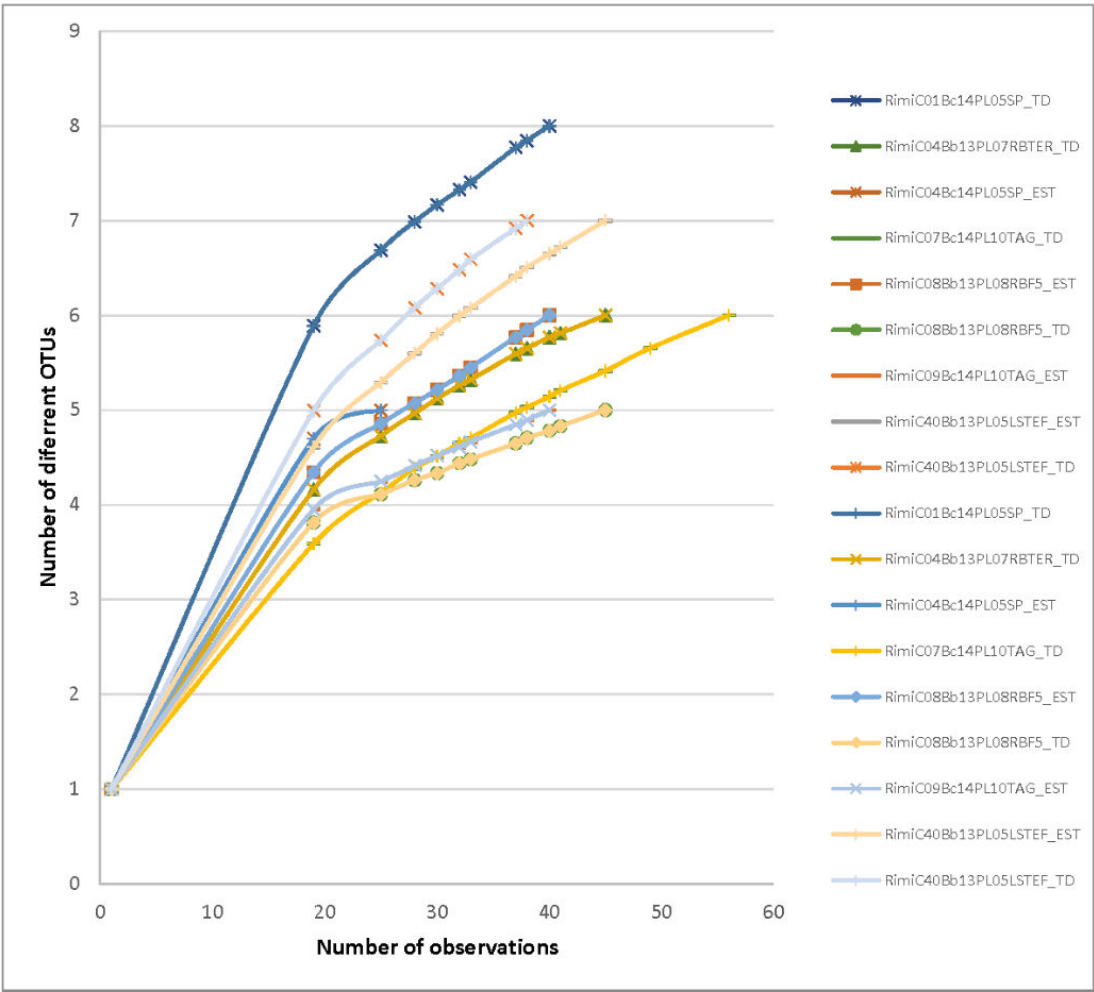


Figure S3 : Dendogramme of all analysed samples based on Bray-Curtis beta diversity on left, and taxonomy of each sequences associated to each samples on right.





29x22mm (600 x 600 DPI)

Principaux résultats

Les principaux résultats concernant cet article sont résumés ci-après.

Les observations vidéo ont montré que les crevettes *Rimicaris chacei* étaient écologiquement distribuées en périphérie des agrégats de *Rimicaris exoculata* sur tous les sites étudiés. Les crevettes *Rimicaris chacei* ont été observées à proximité directe de l'émanation de fluides hydrothermaux à faible activité (cheminées hydrothermales de petite taille situées sur la cheminée principale, zones de diffusion). De manière nette, l'observation sur les mêmes images montre une hypertrophie du céphalothorax nettement moins prononcée chez *Rimicaris chacei* que chez *Rimicaris exoculata*. Enfin, *R. chacei* peut présenter des réactions agressives (d'attaque) à l'encontre d'individus de *Rimicaris exoculata* surtout en cas de blessure montrant une activité à minima de nécrophagie, voire de prédation.

Les différentes approches, qu'elles soient microscopiques ou moléculaires, ont mis en évidence la présence de microorganismes abondants mais différents entre le céphalothorax et le tube digestif. Toutes ces observations, qu'elles soient réalisées via la microscopie ou la biologie moléculaire, mettent en évidence une nette ressemblance avec les observations publiées chez *Rimicaris exoculata*.

Les techniques de microscopies ont permis de mettre en évidence différents modèles de distribution en 3 dimensions des microorganismes au niveau du céphalothorax.

La similarité entre ce travail réalisé chez *Rimicaris chacei* et celui connu chez *Rimicaris exoculata* permet de poser l'hypothèse d'une symbiose trophique chez *Rimicaris chacei* au niveau du céphalothorax, et une colonisation microbienne pérenne dans le système digestif, elle aussi similaire à celle connue chez *Rimicaris exoculata*. Ces hypothèses restent cependant à confirmer via des approches fonctionnelles.

Malgré ces similarités, il faut noter la moindre spécialisation vers la symbiose de *R. chacei* avec un céphalothorax moins hypertrophié (donc une communauté bactérienne moins abondante) chélipèdes dépassant du céphalothorax et permettant donc une prise de nourriture externe (contrairement à *Rimicaris exoculata*) et un estomac plus développé que celui de *R. exoculata*. Ces observations laissent supposer un régime trophique mixte.

Discussions et perspectives liées à cet article

Les principaux résultats concernant cette étude ont permis de compléter un travail initié en 1993 par Michel Segonzac et Bernadette Casanova posant l'hypothèse d'une potentielle symbiose partielle chez *Rimicaris chacei* au niveau du céphalothorax laissant supposer une alimentation mixte de cette crevette, entre nécrophagie et symbiose. Notre article complète des travaux initiaux (Casanova *et al.*, 1993, Segonzac *et al.*, 1993), par une approche en microscopie électronique à transmission, qui a permis de mettre en évidence de nouveaux morphotypes, non décrits en 1993. Notre travail a également permis de définir la composition phylogénétique des communautés bactériennes présentes tant au niveau du céphalothorax qu'au niveau du tractus digestif, et d'émettre des hypothèses quant à leur autotrophie et aux métabolismes utilisés. Bien qu'apportant de nouveaux éléments pour appuyer l'hypothèse d'un régime mixte, cette publication n'apporte cependant toujours pas de preuves directes d'une réelle symbiose trophique, tous les éléments avancés étant des preuves indirectes. Des expérimentations complémentaires seront donc à effectuer, comme des approches d'activités via du marquage isotopique *in vivo* avec du bicarbonate marqué en ^{14}C permettant de suivre l'incorporation de carbone par voie chimiosynthétique comme cela a été fait chez *Rimicaris exoculata* (Ponsard *et al.*, 2013).

Au-delà de cet aspect descriptif d'une symbiose, si on part du principe qu'elle existe, la question se pose du rôle écologique de *Rimicaris chacei* dans l'écosystème où elle est retrouvée, et surtout de son histoire évolutive. D'un point de vue évolutif, *Rimicaris chacei* est une plus proche parente de *Rimicaris hybisae* (présentant elle aussi une « symbiose » développée) que de *Rimicaris exoculata*. La première question est de savoir qui est l'ancêtre commun de *R. chacei* et *R. hybisae*, ou plutôt d'où vient-il ? En effet, *R. hybisae* vit exclusivement dans la fosse des Caïmans tandis que *R. exoculata* n'est présente que sur la MAR avec *R. chacei*. Nous pouvons émettre 3 hypothèses : 1) l'ancêtre commun vient des Caïmans, 2) l'ancêtre commun vient de la MAR, 3) l'ancêtre commun ne vient ni de l'un ni de l'autre (par exemple du Pacifique). Pour une discussion plus fluide, partons du principe que cette dernière possibilité est éliminée pour des raisons géographiques et réalisons deux scénarios hypothétiques de ce qui aurait pu se passer. L'intérêt pour la symbiose peut sembler oublié jusqu'ici, il n'en est rien.

- Hypothèse numéro 1 : L'ancêtre commun de *R. hybisae* et *R. chacei* vient des Caïmans. *R. hybisae* serait donc la forme la plus proche de l'ancêtre commun, donc symbiotique exclusif. *R. chacei* serait donc la forme « qui a dû évoluer » (*R. hybisae* a aussi évolué, mais a gardé des caractères phénotypiques identiques à l'origine à savoir une symbiose développée) lors de sa colonisation vers la MAR. Dans cette hypothèse, *R. exoculata* aurait déjà colonisé la MAR, où elle est bien installée (d'un point de vue écologique). *R. hybisae* en arrivant sur la MAR, se retrouve donc en compétition avec *R. exoculata*, et ne peut pas avoir accès au microenvironnement au plus près du fluide

hydrothermal pour une nutrition optimale. Qui dit peu de nourriture, dit succès reproducteur diminué, donc sélection naturelle des individus les plus aptes. Cette sélection naturelle favoriserait alors au cours du temps les individus présentant la capacité d'avoir une alimentation complémentaire à la symbiose stricte, celle-ci n'étant cependant pas perdue, mais en état de régression partielle. On obtient au court du temps une différenciation de l'espèce jusqu'à *Rimicaris chacei* qui se caractérise par une moindre spécialisation symbiotique et un régime mixte.

- Hypothèse numéro 2 : L'ancêtre commun de *R.hybisae* et *R.chacei* vient de la MAR. *R.chacei* serait donc la forme la plus proche de l'ancêtre commun, donc avec une alimentation mixte et une symbiose intermédiaire (morphologie moins hypertrophiée du céphalothorax, estomac peu réduit). *R.hybisae* serait donc la forme « qui a dû évoluer » se spécialisant vers la symbiose. Dans cette hypothèse cela signifie au préalable que *R.exoculata* et *R.chacei* ont un ancêtre commun (c'est le cas, mais il est antérieur à celui entre *R.hybisae* et *R.chacei* d'après les données de génétiques publiées des hôtes), que la différenciation des deux espèces s'est réalisé de manière sympatrique, et que *R.chacei* a subi lors de cette différenciation une régression de sa symbiose, pour aller vers une alimentation mixte. Lors d'une colonisation des Caïmans par *R.chacei* (on part du principe que la dispersion larvaire de *R. exoculata* ne lui aurait pas permis cette colonisation), les individus ayant le plus d'affinité avec l'alimentation de type symbiotique, ont eu un meilleur succès reproducteur, sur un espace disponible et donc sans compétition, ont donc favorisé ce caractère phénotypique symbiotique, et permis d'arriver à l'espèce *Rimicaris hybisae*. On aurait donc par deux fois dans l'évolution l'obtention de la métamorphose vers des espèces symbiotiques : *R. exoculata* sur la MAR et *R. hybisae* sur les Caïmans.

En posant l'hypothèse d'une bonne résolution phylogénique moléculaire des trois espèces citées (ie d'un bon choix de marqueur évolutif pour les hôtes), ces deux hypothèses me semblent personnellement plausibles pour une histoire évolutive récente, mais les deux hypothèses posent de nombreuses autres questions si on cherche à remonter plus loin dans le temps. Il faudrait alors l'expertise de géologues, physiciens, océanographes, biologistes, et spécialistes de l'évolution et de nombreuses données devraient être acquises dans chaque discipline pour répondre à cette question.

2^{ème} Article : Biogéographie
comparative des symbiontes de
crevettes hydrothermales
Rimicaris exoculata et *Rimicaris*
chacei le long de la MAR



Introduction à la problématique

Dans le premier article, nous avons pu réaliser une étude descriptive d'un point de vue morphologique et phylogénétique) des microorganismes associés à la crevette *Rimicaris chacei*. Une observation est récurrente : les microorganismes observés en microscopie, ou identifiés en biologie moléculaire sont comparables, voire similaires à ceux retrouvés chez *Rimicaris exoculata*, qui partage la même niche écologique. La principale limite de ce premier article est l'absence de réplicats biologiques et techniques, ainsi que la profondeur de séquençage, plus précisément en biologie moléculaire. Il s'avère que le même problème a déjà été mis en évidence chez *Rimicaris exoculata* (Guri *et al.*, 2012) pour des problèmes de profondeur de séquençage (approches par clonage et non NGS). La science a, comme dans la vie courante, ses phénomènes de mode, et le séquençage haut débit (dit NGS) fait partie des modes de la science actuelle. Il faut cependant une question scientifique qui nécessite l'utilisation de cette méthodologie, et ainsi ne pas séquençer pour le plaisir de séquençer (avoir une question, et utiliser le meilleur outil pour y répondre). Or plusieurs questions scientifiques sont venues se poser à la suite des observations issues du premier article qui présente davantage une description qualitative que quantitative. Si symbiose il y a chez *Rimicaris chacei*, et qu'elle est similaire à celle rencontrée chez *Rimicaris exoculata* (postulat du premier article), les symbiontes présentent-ils une biogéographie en fonction des sites hydrothermaux comme décrit chez *R. exoculata* ? Ou, y a-t-il une spécificité de l'espèce hôte plus que de la biogéographie ? Ou alors les effets de la géographie des sites hydrothermaux et de l'espèce hôte peuvent-ils être cumulés ? L'utilisation d'une nouvelle méthodologie pour pouvoir obtenir des données de séquences d'ARNr 16S en grand nombre, sur de nombreux échantillons, à un coût le plus bas possible est nécessaire. Nous avons donc choisi l'utilisation du séquençage haut débit de seconde génération pour répondre de ces questions écologiques. Cependant, dans le cas d'une similarité fine entre les symbiontes des deux espèces de crevettes, l'utilisation des OTUs comme unité écologique est-elle suffisante ? Existe-il une méthodologie pour aller plus loin, ou plus finement tant la diversité au sein des bactéries symbiotiques, qu'au niveau de l'OTU, mais avec des données de gènes d'ARNr 16S, et ainsi descendre dans un seuil de discrimination supérieur et être plus résolutif ? La réponse est oui, grâce à l'oligotyping, qui sera appliqué à la suite d'une étude de métabarcoding classique utilisant l'OTU comme unité écologique, sur les groupes taxonomiques les plus abondants qui seront identifiés.

Matériel et méthodes

Echantillonnage

Les échantillons ont été prélevés sur quatre sites de la ride médio-atlantique (MAR) : Lucky Strike, Rainbow, TAG et Snake Pit durant les campagnes BICOSE (2014) et BioBaz 2013. Les crevettes ont été prélevées avec un aspirateur à faune fixé sur le véhicule téléopéré Victor 6000 à bord du navire *Pourquoi pas ?* Une fois à bord, les individus prélevés ont été disséqués et séparés en différentes pièces anatomiques : branchiostégites (LB) scaphognathite (Sc) et tube digestif (TD). Les différents tissus animaux ont été directement congelés à -80°C et 38 extractions d'ADN ont été réalisées une fois revenus à terre au laboratoire. Ces ADN extraits ont été expédiés à un prestataire de service pour le séquençage.

Extraction d'ADN/PCR/Librairies

L'ADN de l'ensemble des échantillons analysés (LB, Sc, TD) a été extrait et purifié avec le kit Nucleospin soil (Macherey-Nagel) en respectant les consignes d'utilisation du fabricant. Les amplicons bactériens du gène codant pour la petite sous-unité du ribosome (16S) ont été amplifiés par PCR 3 fois pour chaque échantillon (réplicats techniques) avec les amorces « universelles » 357F (V3) 5' CCTACGGGAGGCAGCAG 3' et 926R (v5) 5' CCGTCWATTCMTTTRAGT 3' (modifiées) selon le protocole du prestataire de service.

Des librairies de séquençage ont ensuite été réalisées à partir de ces produits PCR, puis séquencées via la technologie illumina MiSeq 2x300pb en deux séquençages.

Un tableau récapitulatif des différents échantillons est donné ci-dessous :

Espèce	Site	Campagne	Plongée	Edifice	Individu	"Organe"	Nom de l'échantillon	Barcode			Séquençage réussi
<i>Rimicaris chacei</i>											
<i>Rimicaris chacei</i>	Lucky Strike										
<i>Rimicaris chacei</i>	Lucky Strike	Biobaz 2013	PL05	Tour Eiffel	40	SC	RimiC40Bb13PL05LSTE_SC	D59517	D59518	D59519	Oui
<i>Rimicaris chacei</i>	Lucky Strike	Biobaz 2013	PL05	Tour Eiffel	40	TD	RimiC40Bb13PL05LSTE_TD	D74005	D74006	D74007	Oui
<i>Rimicaris chacei</i>	Rainbow										
<i>Rimicaris chacei</i>	Rainbow	Biobaz 2013	PL07	Termitière	04	SC	RimiC04Bb13PL07RbTer_SC	D59520	D59521	D59522	Oui
<i>Rimicaris chacei</i>	Rainbow	Biobaz 2013	PL07	Termitière	04	TD	RimiC04Bb13PL07RbTer_TD	D74008	D74009	D74010	Oui
<i>Rimicaris chacei</i>	Rainbow	Biobaz 2013	PL08	France 5	08	SC	RimiC08Bb13PL08RbF5_SC	D59523	D59524	D59525	Oui
<i>Rimicaris chacei</i>	Rainbow	Biobaz 2013	PL08	France 5	08	TD	RimiC08Bb13PL08RbF5_TD	D74011	D74012	D74013	Non
<i>Rimicaris chacei</i>	Rainbow	Momar04	PL05		02	SC	RimiC02Mom04PL05Rb_SC	D59526	D59527	D59528	Oui
<i>Rimicaris chacei</i>	Rainbow	Momar04	PL05		03	TD	RimiC03Mom04PL05Rb_TD	D74014	D74015	D74016	Non

<i>Rimicaris chacei</i>	TAG										
<i>Rimicaris chacei</i>	TAG	Bicose 2014	PL10		07	SC	RimiC07Bc14PL10TAG_SC	D59529	D59530	D59531	Oui
<i>Rimicaris chacei</i>	TAG	Bicose 2014	PL10		07	TD	RimiC07Bc14PL10TAG_TD	D74017	D74018	D74019	Non
<i>Rimicaris chacei</i>	TAG	Bicose 2014	PL10		09	SC	RimiC09Bc14PL10TAG_SC	D59532	D59533	D59534	Oui
<i>Rimicaris chacei</i>	TAG	Bicose 2014	PL10		09	TD	RimiC09Bc14PL10TAG_TD	D74020	D74021	D74022	Oui
<i>Rimicaris chacei</i>	TAG	Bicose 2014	PL10		11	SC	RimiC11Bc14PL10TAG_SC	D59535	D59536	D59537	Oui
<i>Rimicaris chacei</i>	TAG	Bicose 2014	PL10		11	TD	RimiC11Bc14PL10TAG_TD	D74023	D74024	D74025	Oui
<i>Rimicaris chacei</i>	Snake Pit										
<i>Rimicaris chacei</i>	Snake Pit	Bicose 2014	PL05		01	SC	RimiC01Bc14PL05SP_SC	D59538	D59539	D59540	Oui
<i>Rimicaris chacei</i>	Snake Pit	Bicose 2014	PL05		01	TD	RimiC01Bc14PL05SP_TD	D74026	D74027	D75028	Oui
<i>Rimicaris chacei</i>	Snake Pit	Bicose 2014	PL05		04	SC	RimiC04Bc14PL05SP_SC	D59541	D59542	D59543	Oui
<i>Rimicaris chacei</i>	Snake Pit	Bicose 2014	PL05		04	TD	RimiC04Bc14PL05SP_TD	D75029	D75030	D75031	Oui
<i>Rimicaris chacei</i>	Snake Pit	Bicose 2014	PL05		05	SC	RimiC05Bc14PL05SP_SC	D59544	D59545	D59546	Oui

<i>Rimicaris chacei</i>	Snake Pit	Bicose 2014	PL05		05	TD	RimiC05Bc14PL05SP_TD	D75032	D75033	D75034	Oui
<i>Rimicaris exoculata</i>											
<i>Rimicaris exoculata</i>	Rainbow										
<i>Rimicaris exoculata</i>	Rainbow	Biobaz 2013	PL06	Termitière	I1	SC	RimiEI01Bb14PL06RbTer_SC	D73978	D73979	D73980	Oui
<i>Rimicaris exoculata</i>	Rainbow	Biobaz 2013	PL06	Termitière	I1	TD	RimiEI01Bb14PL06RbTer_TD	D75035	D75036	D75037	Oui
<i>Rimicaris exoculata</i>	Rainbow	Biobaz 2013	PL06	Termitière	I2	SC	RimiEI02Bb14PL06RbTer_SC	D73981	D73982	D73983	Oui
<i>Rimicaris exoculata</i>	Rainbow	Biobaz 2013	PL06	Termitière	I2	TD	RimiEI02Bb14PL06RbTer_TD	D75038	D75039	D75040	Oui
<i>Rimicaris exoculata</i>	Rainbow	Biobaz 2013	PL06	Termitière	I3	SC	RimiEI03Bb14PL06RbTer_SC	D73984	D73985	D73986	Oui
<i>Rimicaris exoculata</i>	Rainbow	Biobaz 2013	PL06	Termitière	I3	TD	RimiEI03Bb14PL06RbTer_TD	D75041	D75042	D75043	Oui
<i>Rimicaris exoculata</i>	TAG										
<i>Rimicaris exoculata</i>	TAG	Bicose 2014	PL08		212	SC	RimiE212Bc14PL08TAG_SC	D73987	D73988	D73989	Oui
<i>Rimicaris exoculata</i>	TAG	Bicose 2014	PL08		212	TD	RimiE212Bc14PL08TAG_TD	D75044	D75045	D75046	Oui

<i>Rimicaris exoculata</i>	TAG	Bicose 2014	PL10		293	SC	RimiE293Bc14PL10TAG_SC	D73990	D73991	D73992	Oui
<i>Rimicaris exoculata</i>	TAG	Bicose 2014	PL10		293	TD	RimiE293Bc14PL10TAG_TD	D75047	D75048	D75049	Oui
<i>Rimicaris exoculata</i>	TAG	Bicose 2014	PL10		298	SC	RimiE298Bc14PL10TAG_SC	D73993	D73994	D73995	Oui
<i>Rimicaris exoculata</i>	TAG	Bicose 2014	PL10		298	TD	RimiE298Bc14PL10TAG_TD	D75050	D75051	D75052	Oui
<i>Rimicaris exoculata</i>	Snake Pit										
<i>Rimicaris exoculata</i>	Snake Pit	Bicose 2014	PL01	Elan	35	SC	RimiE35Bc14PL01SPE_SC	D73996	D73997	D73998	Oui
<i>Rimicaris exoculata</i>	Snake Pit	Bicose 2014	PL01	Elan	35	TD	RimiE35Bc14PL01SPE_TD	D75053	D75054	D75055	Non
<i>Rimicaris exoculata</i>	Snake Pit	Bicose 2014	PL01	Elan	37	SC	RimiE37Bc14PL01SPE_SC	D73999	D74000	D74001	Oui
<i>Rimicaris exoculata</i>	Snake Pit	Bicose 2014	PL01	Elan	37	TD	RimiE37Bc14PL01SPE_TD	D75056	D75057	D75058	Non
<i>Rimicaris exoculata</i>	Snake Pit	Bicose 2014	PL01	Elan	39	SC	RimiE38Bc14PL01SPE_SC	D74002	D74003	D74004	Oui
<i>Rimicaris exoculata</i>	Snake Pit	Bicose 2014	PL01	Elan	39	TD	RimiE38Bc14PL01SPE_TD	D75059	D75060	D75061	Non

Figure 30 : Tableau récapitulatif des échantillons ayant été séquencés via la technologie illumina.

Analyses des données

L'ensemble des données a été analysé via les outils Mothur, Qiime, et le workflow FROGS suivant plusieurs méthodologies.

Premiers résultats

Résultats de qualité :

Dans un premier temps, la qualité des données a été vérifiée en utilisant l'outil FastQC. Les résultats montrent une mauvaise qualité d'une partie des échantillons obtenus à partir du 2^{ème} run de séquençage. Pour autant, la qualité des séquences restantes est suffisamment bonne et les séquences par échantillon en suffisamment grande proportion pour pouvoir être analysées correctement et fournir un signal biologique suffisant.

Par soucis de cohérence, bien que chaque script (Mothur, Qiime et FROGS) soit fonctionnel, les analyses issues uniquement à partir de FROGS seront présentées (car déjà pré-formatées pour être présentées).

Dans un premier temps, on observe qu'une partie du jeu de données est perdue dès l'étape de démultiplexage (Figure 35). Cette perte de données est principalement due au deuxième run de séquençage, corrélé à la mauvaise qualité d'un certain nombre de séquences, un phénomène déjà identifié avec FastQC.

Filtering summary

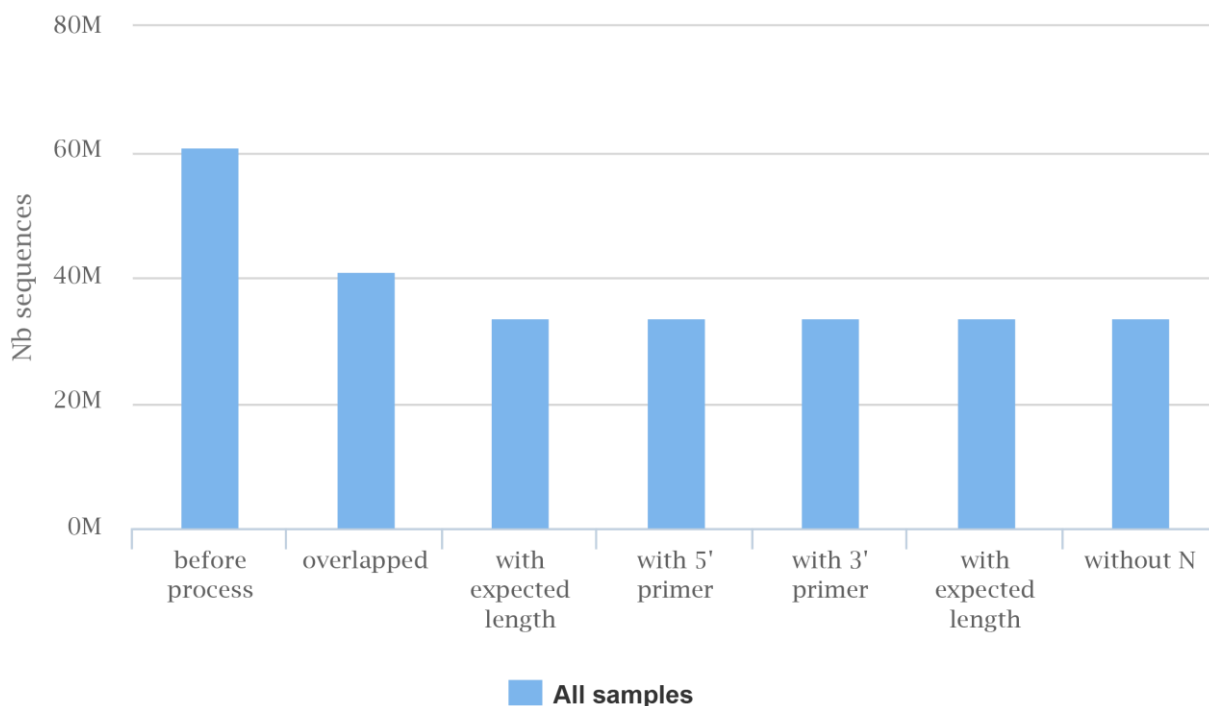


Figure 31 : Etape de filtrage qualité du workflow FROGS sur l'ensemble du jeu de données NGS (en millions de séquences).

La suite des étapes qualité que sont le filtre des amorces, ainsi que la qualité des séquences obtenues induit une faible perte de séquences (Figure 35).

Pour chaque échantillon, le nombre de séquences après chaque étape est donné, et le pourcentage de perte qualité est calculé.

Filtering by sample								
Samples	% kept	before process	overlapped	with expected length	with 5' primer	with 3' primer	with expected length	without N
RimiC01Bc14PL05SP_SC_E1	68.05	642,32	438,819	438,379	438,11	438,047	438,045	437,105
RimiC01Bc14PL05SP_SC_E2	67.85	674,29	459,069	458,774	458,463	458,397	458,39	457,495
RimiC01Bc14PL05SP_SC_E3	68.53	790,901	543,743	543,436	543,101	543,034	543,03	541,965
RimiC01Bc14PL05SP_TD_E1	25.17	666,664	301,584	168,427	168,234	168,163	168,128	167,83
RimiC01Bc14PL05SP_TD_E2	24.31	646,492	294,919	157,752	157,551	157,496	157,474	157,181
RimiC01Bc14PL05SP_TD_E3	25.30	639,297	303,429	162,25	162,07	162,011	161,99	161,723
RimiC02Mom04PL05Rb_SC_E1	76.09	793,724	606,456	605,726	605,34	605,242	605,234	603,931
RimiC02Mom04PL05Rb_SC_E2	74.66	693,32	519,201	519,038	518,73	518,664	518,659	517,631
RimiC02Mom04PL05Rb_SC_E3	75.80	796,579	605,905	605,483	605,134	605,069	605,06	603,8
RimiC04Bb13PL07RbTer_SC_E1	62.03	572,524	356,354	356,092	355,854	355,797	355,791	355,141

RimiC04Bb13PL07RbTer_SC_E2	63.42	610,118	388,388	388,047	387,777	387,721	387,719	386,929
RimiC04Bb13PL07RbTer_SC_E3	63.54	509,53	325,015	324,684	324,497	324,443	324,439	323,768
RimiC04Bb13PL07RbTer_TD_E1	12.20	692,704	482,453	85,122	84,834	84,753	84,691	84,529
RimiC04Bb13PL07RbTer_TD_E2	11.82	629,317	426,662	74,903	74,655	74,588	74,535	74,383
RimiC04Bb13PL07RbTer_TD_E3	12.05	649,385	435,706	78,834	78,549	78,469	78,406	78,269
RimiC04Bc14PL05SP_SC_E1	57.88	699,2	406,62	405,871	405,577	405,508	405,503	404,71
RimiC04Bc14PL05SP_SC_E2	59.46	729,804	436,911	435,125	434,826	434,769	434,764	433,963
RimiC04Bc14PL05SP_SC_E3	58.52	835,805	491,364	490,478	490,201	490,121	490,115	489,139
RimiC04Bc14PL05SP_TD_E1	18.06	720,711	516,155	130,835	130,514	130,42	130,359	130,131
RimiC04Bc14PL05SP_TD_E2	17.31	671,9	470,56	116,945	116,623	116,549	116,498	116,295
RimiC04Bc14PL05SP_TD_E3	16.39	730,386	501,575	120,373	120,046	119,947	119,887	119,679
RimiC05Bc14PL05SP_SC_E1	76.37	607,867	466,043	465,69	465,374	465,284	465,277	464,222
RimiC05Bc14PL05SP_SC_E2	77.58	591,119	460,226	459,963	459,682	459,579	459,57	458,568
RimiC05Bc14PL05SP_SC_E3	77.84	714,778	558,81	558,155	557,751	557,638	557,628	556,398
RimiC05Bc14PL05SP_TD_E1	20.10	1028,774	686,562	207,759	207,358	207,236	207,13	206,74
RimiC05Bc14PL05SP_TD_E3	18.59	627,188	415,087	117,241	116,934	116,842	116,79	116,563
RimiC07Bc14PL10TAG_SC_E1	70.31	757,561	534,424	534,167	533,805	533,704	533,699	532,675
RimiC07Bc14PL10TAG_SC_E2	70.66	812,809	576,841	575,838	575,491	575,38	575,371	574,332
RimiC07Bc14PL10TAG_SC_E3	68.83	786,837	543,306	543,065	542,715	542,606	542,597	541,587
RimiC08Bb13PL08RbF5_SC_E1	64.06	494,214	317,954	317,477	317,269	317,224	317,219	316,581
RimiC08Bb13PL08RbF5_SC_E2	63.90	737,835	473,26	472,91	472,604	472,518	472,514	471,505
RimiC08Bb13PL08RbF5_SC_E3	66.09	667,233	442,557	442,262	441,994	441,916	441,913	440,998
RimiC09Bc14PL10TAG_SC_E1	65.80	764,275	504,43	504,214	503,892	503,783	503,778	502,867
RimiC09Bc14PL10TAG_SC_E2	66.05	771,481	511,206	510,932	510,575	510,486	510,473	509,59
RimiC09Bc14PL10TAG_SC_E3	67.27	683,263	461,142	460,91	460,558	460,474	460,466	459,597
RimiC09Bc14PL10TAG_TD_E1	17.78	493,665	349,606	88,254	88,056	88,003	87,965	87,771
RimiC09Bc14PL10TAG_TD_E2	18.14	524,666	371,21	95,68	95,445	95,387	95,345	95,175
RimiC09Bc14PL10TAG_TD_E3	18.17	517,203	362,078	94,537	94,276	94,226	94,164	93,979
RimiC11Bc14PL10TAG_SC_E1	69.54	707,705	493,952	493,521	493,189	493,119	493,111	492,153
RimiC11Bc14PL10TAG_SC_E2	68.89	681,831	471,506	471,106	470,802	470,705	470,696	469,717
RimiC11Bc14PL10TAG_SC_E3	72.10	801,046	579,328	579,163	578,799	578,692	578,689	577,553
RimiC11Bc14PL10TAG_TD_E1	44.62	308,307	208,95	138,001	137,873	137,829	137,817	137,579
RimiC11Bc14PL10TAG_TD_E2	44.17	416,033	286,819	184,351	184,159	184,104	184,092	183,778
RimiC11Bc14PL10TAG_TD_E3	44.93	631,559	458,721	284,745	284,446	284,346	284,328	283,789
RimiC40Bb13PL05LSTE_SC_E1	61.47	528,007	325,979	325,525	325,319	325,277	325,275	324,587
RimiC40Bb13PL05LSTE_SC_E2	62.56	637,903	400,72	400,266	400,002	399,935	399,93	399,096
RimiC40Bb13PL05LSTE_SC_E3	62.21	605,354	378,107	377,661	377,415	377,364	377,359	376,574
RimiC40Bb13PL05LSTE_TD_E1	46.19	91,462	59,089	42,349	42,312	42,309	42,306	42,249
RimiC40Bb13PL05LSTE_TD_E2	47.63	310,802	204,777	148,402	148,241	148,218	148,211	148,025
RimiC40Bb13PL05LSTE_TD_E3	48.29	310,712	209,298	150,435	150,282	150,257	150,248	150,049
RimiE212Bc14PL08TAG_SC_E1	38.34	688,588	265,416	264,759	264,554	264,51	264,51	264,033
RimiE212Bc14PL08TAG_SC_E2	38.43	730,354	282,134	281,435	281,206	281,168	281,158	280,711
RimiE212Bc14PL08TAG_SC_E3	39.47	751,828	298,503	297,535	297,341	297,277	297,273	296,767
RimiE212Bc14PL08TAG_TD_E1	59.74	400,836	251,441	240,166	239,954	239,883	239,875	239,474
RimiE212Bc14PL08TAG_TD_E2	59.75	295,939	186,92	177,38	177,231	177,158	177,152	176,834

RimiE212Bc14PL08TAG_TD_E3	58.88	321,384	198,127	189,779	189,633	189,547	189,543	189,219
RimiE293Bc14PL10TAG_SC_E1	64.74	716,506	466,411	465,204	464,892	464,841	464,838	463,896
RimiE293Bc14PL10TAG_SC_E2	64.70	493,751	321,25	320,395	320,187	320,159	320,156	319,474
RimiE293Bc14PL10TAG_SC_E3	65.15	707,219	463,458	462,049	461,773	461,716	461,711	460,781
RimiE293Bc14PL10TAG_TD_E1	10.36	570,764	392,996	59,638	59,389	59,269	59,207	59,113
RimiE293Bc14PL10TAG_TD_E2	10.57	710,63	510,847	75,79	75,41	75,289	75,195	75,085
RimiE293Bc14PL10TAG_TD_E3	10.36	642,594	449,29	67,149	66,843	66,726	66,661	66,567
RimiE298Bc14PL10TAG_SC_E1	53.87	1003,545	543,728	542,062	541,663	541,614	541,61	540,585
RimiE298Bc14PL10TAG_SC_E2	53.70	759,256	410,096	408,86	408,582	408,532	408,529	407,741
RimiE298Bc14PL10TAG_SC_E3	54.48	663,573	363,599	362,533	362,333	362,278	362,271	361,542
RimiE298Bc14PL10TAG_TD_E1	32.79	613,477	421,672	202,001	201,685	201,579	201,539	201,163
RimiE298Bc14PL10TAG_TD_E2	34.38	639,175	452,006	220,655	220,312	220,192	220,148	219,73
RimiE298Bc14PL10TAG_TD_E3	33.06	666,403	473,317	221,28	220,912	220,778	220,737	220,325
RimiE35Bc14PL01SPE_SC_E1	70.23	733,462	517,024	516,533	516,224	516,152	516,145	515,105
RimiE35Bc14PL01SPE_SC_E2	68.91	708,481	490,343	489,562	489,263	489,204	489,198	488,189
RimiE35Bc14PL01SPE_SC_E3	69.32	748,117	520,605	520,199	519,856	519,802	519,795	518,622
RimiE37Bc14PL01SPE_SC_E1	75.14	597,617	450,549	450,314	450,059	449,997	449,993	449,045
RimiE37Bc14PL01SPE_SC_E2	77.07	705,946	545,978	545,663	545,299	545,23	545,224	544,068
RimiE37Bc14PL01SPE_SC_E3	75.50	630,431	477,719	477,342	477,032	476,964	476,958	475,992
RimiE38Bc14PL01SPE_SC_E1	73.19	664,851	488,385	488,038	487,729	487,645	487,64	486,635
RimiE38Bc14PL01SPE_SC_E2	74.81	786,025	589,968	589,736	589,377	589,295	589,287	588,03
RimiE38Bc14PL01SPE_SC_E3	77.37	909,507	706,006	705,715	705,288	705,174	705,166	703,694
RimiEI01Bb14PL06RbTer_SC_E1	71.99	770,915	557,295	556,553	556,215	556,143	556,135	554,96
RimiEI01Bb14PL06RbTer_SC_E2	74.15	725,969	540,381	539,875	539,553	539,487	539,483	538,305
RimiEI01Bb14PL06RbTer_SC_E3	75.16	801,205	604,377	603,818	603,469	603,401	603,396	602,19
RimiEI01Bb14PL06RbTer_TD_E1	60.45	362,187	252,992	219,58	219,424	219,376	219,372	218,954
RimiEI01Bb14PL06RbTer_TD_E2	62.43	490,938	354,984	307,408	307,16	307,092	307,073	306,474
RimiEI01Bb14PL06RbTer_TD_E3	61.04	406,88	286,536	249,127	248,94	248,854	248,845	248,366
RimiEI02Bb14PL06RbTer_SC_E1	75.16	629,291	474,639	474,324	474,03	473,952	473,948	472,945
RimiEI02Bb14PL06RbTer_SC_E2	76.04	690,197	526,462	526,35	526,037	525,948	525,942	524,819
RimiEI02Bb14PL06RbTer_SC_E3	76.03	594,817	453,695	453,593	453,294	453,225	453,219	452,26
RimiEI02Bb14PL06RbTer_TD_E1	72.87	649,356	475,839	474,593	474,275	474,129	474,123	473,214
RimiEI02Bb14PL06RbTer_TD_E2	73.36	677,224	499,796	498,284	497,966	497,82	497,812	496,818
RimiEI02Bb14PL06RbTer_TD_E3	73.42	762,045	563,029	561,227	560,833	560,678	560,671	559,528
RimiEI03Bb14PL06RbTer_SC_E1	70.73	706,05	501,937	500,809	500,509	500,447	500,438	499,363
RimiEI03Bb14PL06RbTer_SC_E2	69.26	858,969	598,018	596,551	596,176	596,096	596,094	594,913
RimiEI03Bb14PL06RbTer_SC_E3	70.52	778,795	551,863	550,818	550,494	550,413	550,406	549,213
RimiEI03Bb14PL06RbTer_TD_E1	19.58	451,646	307,573	88,938	88,713	88,632	88,607	88,448
RimiEI03Bb14PL06RbTer_TD_E2	20.20	307,971	212,356	62,598	62,428	62,36	62,339	62,212
RimiEI03Bb14PL06RbTer_TD_E3	20.29	351,957	247,257	71,803	71,668	71,595	71,564	71,417

A l'issue cette étape, le clustering avec l'outil SWARM est effectué sur un jeu de 33 586 713 séquences et génère un total de 3 570 525 OTUs. Ce nombre est trop important, et ne correspond pas à la réalité biologique mais au mode de séparation en OTU des algorithmes. Il faut donc y appliquer des filtres. Le premier est la recherche des chimères qui constituent 2.23 % des OTUs, mais seulement 0.74 % des séquences.

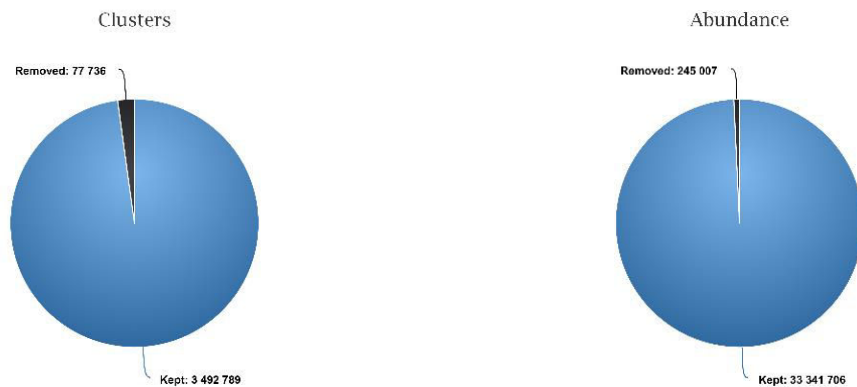


Figure 32 : Recherche des chimères avec l'outil SWARM.

Le nombre d'OTUs reste cependant bien trop important. En effet, SWARM est un outil qui génère beaucoup de singleton (1 séquence par OTU). Pour supprimer uniquement les singletons, on peut appliquer un filtre de 2 (il faut au moins deux séquences dans une OTU pour que celle-ci soit conservée). Ils sont au nombre de 3 278 172 OTUs singleton (9.8 % des données après détection des chimères). Un filtre d'abondance est donc effectué et fixé à 10, c'est-à-dire que chaque OTU présentant moins de 10 séquences est éliminée. Ce choix 10 est totalement arbitraire mais couramment utilisé dans la littérature (figure 37). Après un filtrage de 10, il ne reste plus que 143 OTUs.

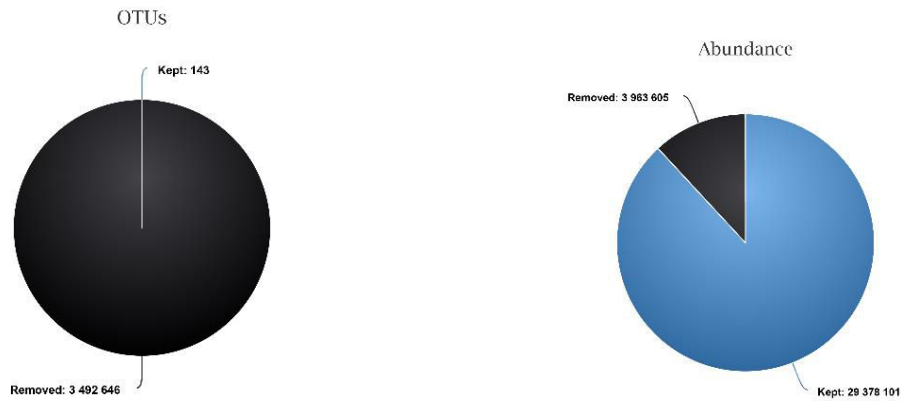


Figure 33 : Filtrage des données NGS selon un paramètre d'abondance.

Cette réduction du nombre d'OTUs peut paraître drastique. D'une part les séquences uniques sont très souvent dues à des séquences de moindre qualité ou présentant des erreurs de séquençage ne permettant plus de les attribuer aux OTU d'origine, créant ainsi une 'fausse' diversité. De plus, dans la recherche d'espèces symbiotiques, le nombre de microorganismes symbiotiques sont de prime abord abondants, et dominants. Eliminer les OTUs minoritaires, qui peuvent potentiellement être des microorganismes transitoires (de l'eau filtrée par la crevette par exemple) est donc à prendre en compte. De plus, l'obtention de 143 OTUs reste similaire en ordre de grandeur par rapport aux banques de clones du premier article qui atteignaient presque le plateau (dans les courbes de raréfactions), donc l'épuisement de la diversité. C'est le cas aussi ici en approche NGS (Figure 38).

Rarefaction curves

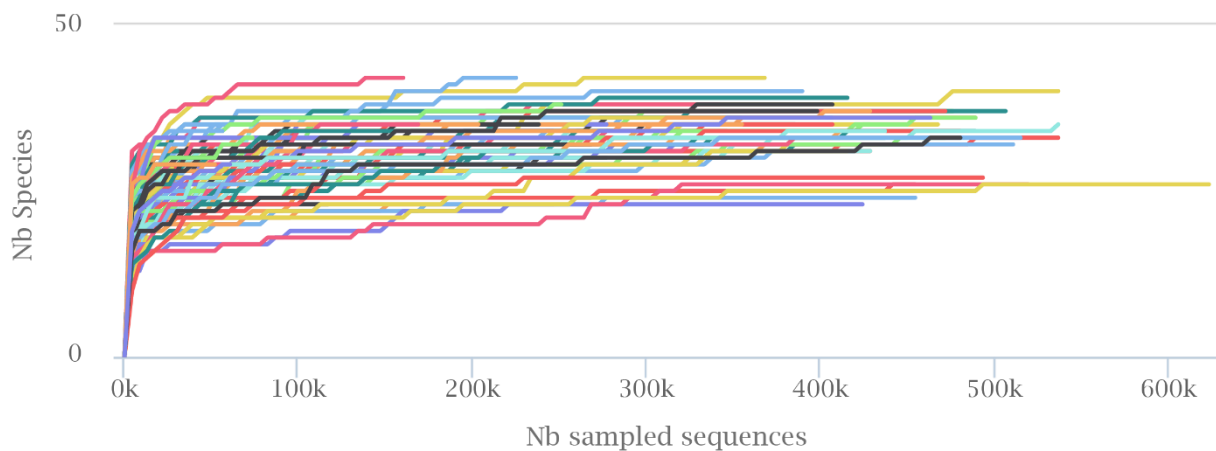


Figure 34 : Courbes de raréfaction NGS.

Cependant ces courbes de raréfaction restent conditionnées par le filtre d'abondance. Il est donc nécessaire de poursuivre les investigations sur cette étape pour réussir à trouver le meilleur compromis biologique entre séquences à éliminer pour ne pas avoir un trop grand jeu de données et ne pas éliminer des séquences qui auraient une importance écologique.

Après normalisation des données, un dendrogramme de similarité basé sur l'indice de Bray-Curtis est obtenu et montre que les réplicats techniques sont tous cohérents, c'est-à-dire que les triplicats de PCR d'un même échantillon sont regroupés dans le dendrogramme (Figure 39). On observe également qu'une distinction est faite entre les échantillons des scaphognathites et ceux du système digestif. De plus, d'après ce même dendrogramme, il ne semble pas y avoir de similarité entre les échantillons provenant de la même espèce, mais une similarité des échantillons en fonction du site hydrothermal, laissant supposer une biogéographie.

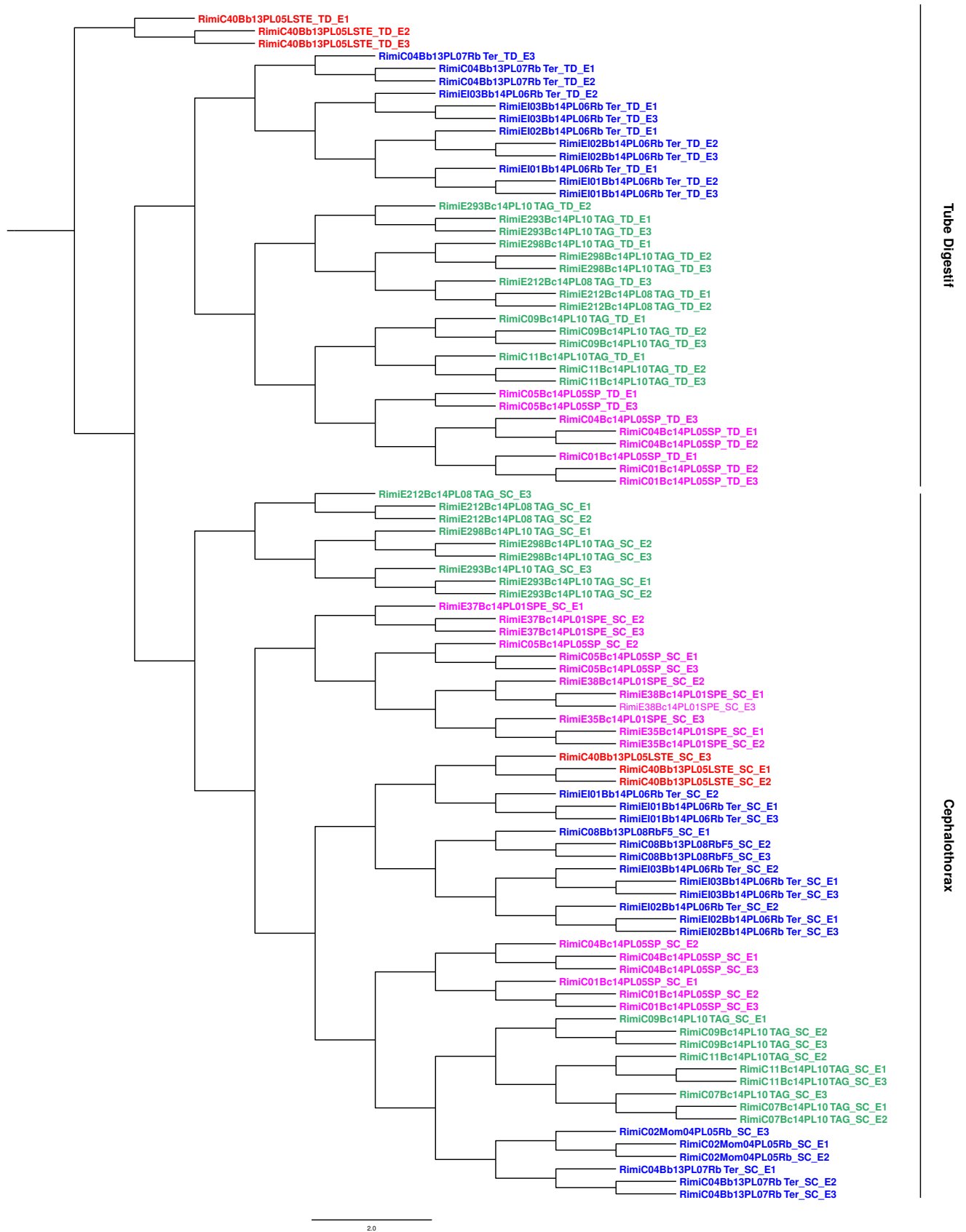


Figure 35 : Dendrogramme de similarité des échantillons NGS à partir de l'indice de Bray-Curtis. (En bleu sont représentés les échantillons de *Rimicaris exoculata*, et en rouge les échantillons de *Rimicaris chacei*)

Ce dendrogramme représente une première interprétation des résultats par approche classique d'analyse des OTUs, mais reste préliminaire. Si la biogéographie n'est pas nettement distinguée, il est possible que ce soit parce que la discrimination des séquences similaires entre elles ne soit potentiellement pas suffisante. Autrement dit, le seuil de l'OTU à 97 % ne serait pas suffisamment résolutif. Pour aller plus loin, il pourrait être intéressant de diminuer le seuil de l'OTU, pour aller à 98 % de similarité, ou 99 %. La limite technique qui apparaît ici sont les erreurs de séquençage. En effet, en augmentant le seuil de l'OTU, on augmente les probabilités de générer des OTUs artefacts, qui seront par la suite traduits en facteurs biologiques lors de l'interprétation des résultats. L'oligotyping est une méthode alternative qui a pour but de dépasser le seuil de 97 % de l'OTU tout en faisant abstraction des erreurs de PCR/séquençage (Eren *et al.*, 2011, Eren *et al.*, 2013, Eren *et al.*, 2014). Pour cette étude nous utilisons le logiciel oligotyping. Pour commencer, les OTUs ou groupes taxonomiques dominants seront extraits du jeu de données, toujours dans l'hypothèse où les symbiontes sont les microorganismes les plus abondants. Puis des oligotypes seront générés pour chaque OTU ou groupes taxonomiques et comparés en fonction des échantillons. Cette définition plus fine de la diversité des symbiontes pourra permettre de mettre en évidence une potentielle différenciation biogéographique plus fine (effet site d'origine), ou une sélection des symbiontes en fonction de l'espèce de crevette hôte. Ou encore une sélection des symbiontes en fonction du site hydrothermal d'origine et de l'espèce de crevette hôte. Ces résultats ne sont pas encore générés. Cependant, les données acquises sont de qualité suffisante pour être analysées, et répondent aux critères de l'oligotyping, c'est-à-dire au moins 1000 séquences par OTUs d'intérêts et par échantillons. C'est donc cette approche qui sera appliquée à la suite de l'analyse écologique « traditionnelle ».

D'un point de vue taxonomique, les différentes OTUs retenues précédemment sont affiliées aux *Epsilonproteobacteria*, *Gammaproteobacteria*, *Alphaproteobacteria*, *Bacteroidetes*, *Deferribacteres*, *Mollicutes*. L'ensemble de cette taxonomie est semblable à celle retrouvée dans le premier article, venant ainsi conforter la cohérence du traitement des données d'un point de vue biologique. Malheureusement la suite du traitement des données est relativement longue et n'a pu être traitée au cours de cette fin de thèse.

Discussion Générale

Discussion scientifique

De nombreux éléments chimiques présents dans les dépôts minéraux hydrothermaux (notamment les dépôts de sulfures) présentent une valeur commerciale de plus en plus attrayante pour les hautes technologies modernes. Cet attrait pour les minéraux et métaux et le développement des compétences minières en environnements profonds font peser sur ces écosystèmes des dangers nouveaux (Glover & Smith, 2003, Ramirez-Llodra *et al.*, 2011). L'exploitation de ces environnements par l'Homme n'est qu'une question de temps, et présente des risques potentiels de modifications drastiques de ces écosystèmes suite aux extractions minières. Il est donc important de mieux comprendre ces environnements et l'ensemble des processus biologiques qui les sous-tendent en fonctionnement naturel. D'un point de vue écotoxicologique, cela revient à faire une étude à très grande échelle du temps zéro (état de référence), c'est-à-dire de l'environnement avant sa détérioration éventuelle.

Jusqu'à ce travail de thèse, les seules études portant sur *Rimicaris chacei* l'étaient en comparaison avec d'autres espèces (Casanova *et al.*, 1993, Segonzac *et al.*, 1993, Gebruk *et al.*, 1997, Gebruk *et al.*, 2000). Ces travaux ont montré qu'elle était peu abondante lorsqu'elle était rencontrée, qu'elle abritait des bactéries dans sa cavité céphalothoracique et son tractus digestif, et que son régime trophique était mixte, alliant symbiose et nécrophagie.

Récemment (durant la durée de cette thèse) le genre *Chorocaris* fut réaffilié au genre *Rimicaris*, et *Chorocaris chacei* est donc devenue *Rimicaris chacei*. Bien que présentant de nettes différences morphologiques avec les trois espèces initialement connues du genre *Rimicaris*, *R. chacei* possède également un certain nombre de points communs, dont la présence de bactéries au niveau du céphalothorax et dans le tractus digestif.

Rimicaris chacei est donc un modèle biologique encore mal connu qui commence à soulever de nombreuses questions. Il faut cependant bien commencer quelque part, et en tant que microbiologiste, nous nous sommes tout naturellement intéressés à l'aspect symbiose.

Si un travail avait déjà été effectué sur l'observation des régions céphalothoraciques colonisées par des microorganismes chez cette crevette, ces observations étaient effectuées uniquement par des approches de microscopie électronique à balayage, sur des individus ne provenant que d'un seul site (Snake Pit). Nous avons donc décidé de compléter cette description, en refaisant, il faut bien le

reconnaître, la microscopie effectuée au siècle dernier (Casanova *et al.*, 1993, Segonzac *et al.*, 1993). Nous avons donc décidé de compléter ces observations avec des individus de différents sites. De plus, une approche par la microscopie électronique à transmission a été réalisée et a permis d'affiner la diversité morphologique établie en microscopie électronique à balayage, et de voir ce qui se déroule dans et sous l'encroûtement minéral des lames branchiostèges et des scaphognathites. Nous nous sommes également penchés sur les bactéries associées au tube digestif pour lesquelles aucune donnée n'existait jusqu'à présent.

Un volet totalement nouveau chez cette espèce, et néanmoins indispensable à l'étude d'une éventuelle symbiose, a également été entrepris : l'analyse de la composition et de la diversité phylogénétique des bactéries présentes, dans le céphalothorax et dans le tube digestif.

L'ensemble de cette étude a souffert de la faible densité à laquelle on trouve cette crevette sur les sites hydrothermaux, en comparaison avec l'espèce *R. exoculata* (2500 vs. 50 individus.m⁻², Segonzac *et al.*, 1993). Les faibles possibilités d'échantillonnage, dus à cette faible densité de population, et à la difficulté de capture des spécimens, ont été une problématique constante durant la thèse : pas assez d'individus pour faire suffisamment de réplicats biologiques, que ce soit en microscopie ou en biologie moléculaire. Des échantillons précieux à ne pas abîmer, surtout en microscopie, d'où des microanalyses par rayon X difficiles à optimiser car les échantillons avaient une structure 3D, ce qui est loin d'être optimal pour cette méthodologie.

Durant cette thèse, de nouveaux morphotypes bactériens ont été identifiés au niveau du céphalothorax et du tube digestif de *Rimicaris chacei*. En tenant compte du faible nombre de spécimens observés, des présence/absence de certains morphotypes en fonction de sites ont été notées, qui pourraient être expliquées par des variations dans la composition chimique des fluides hydrothermaux des sites étudiés (ultramafiques vs basaltiques). Cela laisse déjà présager une possible forme de biogéographie des symbiontes. Est-elle liée à une absence, ou une sous-représentation de certaines lignées dans l'environnement ? ou d'une sélection de *Rimicaris chacei* envers certaines lignées microbiennes ? L'hôte pourrait-il être capable de sélectionner ses symbiontes afin de maximiser son *fitness* ? A ces questions, l'approche moléculaire haut débit pourrait apporter une réponse.

L'identification taxonomique des microorganismes a révélé la présence de groupes bactériens affiliés aux *Epsilonproteobacteria*, *Gammaproteobacteria*, *Alphaproteobacteria*, *Bacteroidetes*, *Deferribacteres* et *Mollicutes*. Outre une taxonomie commune en fonction des sites, ces différents phylums sont rencontrés chez le modèle *Rimicaris exoculata*, laissant supposer une similarité symbiotique chez ces deux espèces de crevettes. L'étude écologique par banques de clones via la

création d'OTU montre un nombre limité de clusters, semblables à ce qui peut être attendu lors d'une symbiose. De plus la « phylogénie » effectuée sur les différentes OTUs identifiées montre une proximité certaine avec les lignées rencontrées chez la crevette *Rimicaris exoculata*. A partir de cette même phylogénie, une inférence métabolique putative a été proposée à partir du positionnement des OTUs des *Epsilonproteobacteria* et des *Gammaproteobacteria* dans l'arbre du vivant. Elle tient compte de la littérature concernant les lignées cultivées les plus proches, ou les activités métaboliques connues des représentants incultivés les plus proches. Ces inférences métaboliques restent cependant à nuancer, car la phylogénie, surtout lorsqu'elle a trait au monde microbien, n'est pas reconnue pour être robuste pour ce type de conclusion. Cela reste donc des hypothèses de travail. Une biogéographie des symbiontes en fonction des sites étudiés a aussi été proposée, mais reste spéculative. En effet, le nombre de séquences (et non la profondeur de séquençage) par site n'est pas suffisant via les approches de clonages. Il n'est donc pas possible de conclure de manière efficace sur cette hypothèse. Cependant, de nouvelles méthodologies comme le séquençage NGS permettent à la fois d'obtenir un nombre suffisant de séquences, ainsi que la possibilité de réplicats biologiques et techniques à moindre coût. C'est ce que nous avons donc entrepris.

L'ensemble des résultats laissent supposer de manière indirecte une symbiose au niveau du céphalothorax et du tube digestif chez *Rimicaris chacei*, avec une forte similarité partagée avec *Rimicaris exoculata*.

Par comparaison avec ce qui est connu chez *Rimicaris exoculata*, tant morphologiquement que microbiologiquement, le régime trophique de *Rimicaris chacei* semble donc au moins partiellement reposer sur la symbiose, complété par la nécrophagie (premier article) et pose la question du rôle écologique et de la place de cette crevette dans les écosystèmes hydrothermaux.

La similarité des symbiontes partagés avec *Rimicaris exoculata* laisse supposer une histoire commune ou un partage de symbiontes lié à l'environnement (acquisition horizontale). Le régime trophique mixte de *R. chacei* la rapproche des crevettes 'classiques' (nécrophages ou prédatrices) alors que *R. exoculata* semble avoir évolué vers une spécialisation symbiotique typique des systèmes hydrothermaux comme c'est le cas par exemple pour les modioles. Cependant, la phylogénie des hôtes rapproche *Rimicaris chacei*, présente sur la MAR, de *Rimicaris hybisae*, présente dans la fosse des Caïmans. Cela pose la question au cours de l'évolution d'une spécialisation vers la symbiose ayant conduit à *R. exoculata* dans un premier temps, puis à *R. hybisae* dans un second temps, tout en maintenant *R. chacei* dans une évolution intermédiaire.

Quelle est donc l'histoire évolutive de cette crevette : deux cousines avec un céphalothorax hypertrophié et des colonisations abondantes (et même une symbiose trophique démontrée chez *Rimicaris exoculata*), et *Rimicaris chacei* sans cette hypertrophie caractéristique ? Cette symbiose est-elle liée à la niche écologique occupée ? *Rimicaris chacei* est-elle en train de perdre sa symbiose ? Est-elle en train de simplement régresser sans pour autant perdre cette symbiose ? Est-elle en train d'évoluer vers une symbiose ? L'aspect symbiotique est-il le seul facteur pouvant expliquer l'histoire de *Rimicaris chacei* ? La symbiose, qui semble conférer un meilleur *fitness* aux espèces hydrothermales étant semble-t-il réduite ici, ceci explique-t-il aussi le shift entre le nombre de juvéniles qui sont retrouvés en périphérie de édifices hydrothermaux (succès reproductif) par rapport au nombre d'adultes matures retrouvés (moindre succès de développement/croissance) ? Le débat reste ouvert, et les idées nombreuses.

Pour commencer, il faudrait confirmer de manière directe la symbiose chez *Rimicaris chacei*. Pour cela, des incubations sous pression, et en fournissant aux crevettes un substrat carboné chargé en carbone C¹⁴ que seuls les microorganismes pourraient assimiler, et transférer ensuite à la crevette, seront nécessaires. Cette expérience n'a pas pu être réalisée durant cette thèse du simple fait de l'absence d'échantillons ayant subi ce type d'incubation. La symbiose trophique reste donc hypothétique.

L'étude d'une possible biogéographie des symbiontes a été approfondie. En effet, le nombre d'échantillons conditionnés pour des analyses moléculaires a été suffisant pour initier une analyse de métabarcoding sur différentes parties de crevettes (scaphognathite et tube digestif) de différents individus, provenant de différents sites hydrothermaux, et ceci pour les espèces de crevette *Rimicaris chacei* et *Rimicaris exoculata*. La question scientifique relative à cette ségrégation biogéographique est la suivante : si une biogéographie des symbiontes existe chez les crevettes *Rimicaris chacei* et *Rimicaris exoculata*, est-elle due aux variations physico-chimiques existant entre les différents sites hydrothermaux de la MAR, ou en fonction de l'espèce de crevette hôte, ou ces deux effets peuvent-ils être cumulatifs ? Un lourd travail de traitement bioinformatique (et non de biostatistique) a été réalisé pour cette analyse (développé tout au long du manuscrit) toujours en cours. Les analyses sont cependant assez avancées pour confirmer que le jeu de données est exploitable d'un point de vue biologique et technique. Cependant, certaines limites ont été mises en évidence, comme le fait que les tubes digestifs de *Rimicaris exoculata* sont globalement trop petits pour permettre l'extraction de suffisamment d'ADN bactérien pour que les spécimens soient analysés individuellement. Bien que préliminaire, les premiers résultats montrent une distinction nette des populations résidant dans le céphalothorax de celle du tube digestif chez *Rimicaris chacei*. Ceci confirme les résultats obtenus dans

le premier article. De plus, une certaine biogéographie semble exister mais reste à approfondir. La question du seuil de l'OTU se pose. 97% de similarité est-il un seuil suffisant pour répondre à notre question. N'est-il pas intéressant d'aller au-delà l'OTU, avec la méthodologie d'oligotyping (cf. article 2).

Perspectives scientifiques

Plusieurs perspectives à ce travail peuvent être envisagées (en faisant abstraction des coûts que cela représente et des campagnes nécessaires pour un échantillonnage suffisant) :

- Etudier le cycle de vie de *Rimicaris chacei* afin de déterminer par exemple le temps nécessaire à la réalisation d'un cycle de mue complet. Cela permettrait de comprendre le « rythme » de vie de cette crevette par rapport à celui de *Rimicaris exoculata*.
- Etudier au court du temps, de manière saisonnière (par exemple), les variations de populations microbiennes en fonction des saisons, ou des périodes de reproduction, afin de déterminer si un régime trophique, ou une modification de la symbiose est favorisé/entre en jeu en fonction de certaines contraintes environnementales ou du cycle de vie.
- Etudier les communautés microbiennes présentes chez le mâle et chez la femelle pour déterminer si le sexe est un facteur affectant la symbiose (cf. le travail fait chez *Rimicaris exoculata*).
- Réaliser une étude de métagénomique comparative des symbiontes de *Rimicaris chacei* et *Rimicaris exoculata* sur différents sites hydrothermaux afin de décrire les potentiels génétiques des symbiontes. Il serait même possible d'aller jusqu'à de la métatranscriptomique comparative.
- Capturer sur un site des spécimens de *Rimicaris chacei* pour les éloigner des édifices hydrothermaux et déterminer si le régime trophique subit un phénomène d'acclimatation en privilégiant la nécrophagie à la symbiose, ou vice versa. Ceci permettrait aussi de voir si la crevette maintient une forme de relation pérenne avec les microorganismes la colonisant, ou si elle éliminerait les symbiontes tout simplement. Sera-t-elle aussi par la suite capable de rétablir une relation aussi efficace avec des microorganismes ?
- Capturer des individus adultes maintenus en vie à bord du bateau et les mettre en contact avec les fluides provenant d'un autre site hydrothermal. Ceci permettrait de savoir si 1) Les symbiontes sont capables de s'acclimater 2) Si la crevette est capable d'accepter un

changement d'activité de ses symbiontes, voire un changement dans la composition de ses communautés symbiotiques.

- Etudier l'acquisition de la symbiose chez *Rimicaris chacei*, à l'image du travail déjà en cours au LEP/LM2E (thèse en cours sur ce sujet).

Au-delà du simple fait de mieux connaître la symbiose chez l'espèce *Rimicaris chacei*, l'ensemble de ces perspectives permettrait d'appréhender les capacités de résilience de cette espèce. Il ne faut pas oublier le contexte général dans lequel se trouve ce sujet de thèse. Celui de faire un état des lieux, état de référence, des environnements hydrothermaux dans le cas d'exploitation minière potentielle. Dans le cas de *Rimicaris chacei*, son régime trophique mixte serait-il un avantage pour recoloniser le milieu en cas de détérioration/destruction ? Serait-elle capable de remplacer *Rimicaris exoculata* et devenir la nouvelle *Rimicaris hybisae* version MAR ? Serait-il possible de réintroduire des individus adultes de *Rimicaris chacei* sur un site détérioré, mais provenant d'un autre site hydrothermal voisin, ou serait-il impératif d'utiliser (d'une manière encore inconnue) les stades larvaires de crevette ?...

Discussion personnelle

Il est donc venu le moment d'avoir une discussion plus personnelle du travail effectué. Durant cette thèse, j'ai voulu raconter le début d'une histoire (qui je l'espère sera longue), celle de *Rimicaris chacei*. Si les conventions m'avaient permis de choisir l'intitulé de mon sujet de thèse, je l'aurais aujourd'hui appelé : « Il était une fois *Rimicaris chacei* ».

D'un point de vue méthodologique, cette thèse a aussi permis de générer beaucoup de données. Le barcoding effectué sur *Rimicaris chacei* ne représente qu'une fraction de ce qui a été généré durant la thèse (en plus des données de microscopie). D'autres espèces comme *Alvinocaris markensis*, *Mirocaris fortunata*, ou *Rimicaris sp.* ont été étudiées, mais les résultats n'ont pour l'instant pas été totalement analysés. Les informations préliminaires (phylogénétiques) montrent très nettement que *Rimicaris chacei* était un bon modèle d'étude (c'est-à-dire possédant les communautés symbiotiques les plus développées). L'analyse de l'ensemble de ce jeu de données toutes espèces confondues a nécessité quelques bouleversements au sein du laboratoire, notamment dans le traitement des données de séquençage obtenues par PCR clonage. En effet, l'analyse manuelle de plus de 5 000 séquences avec les méthodes présentes au laboratoire n'étaient tout simplement pas envisageable. C'est de cette constatation qu'est née l'idée d'un protocole de barcoding largement décrit dans le chapitre matériel et méthode. Mais si j'en avais besoin pourquoi pas d'autres ? J'ai donc proposé ce projet à différents

membres du laboratoire, qui se sont trouvés intéressés. Je n'ai donc pas cherché à développer ce protocole que pour ma propre utilisation, mais aussi pour autrui, un autrui qui n'est d'ailleurs pas forcément un adepte de l'informatique. Et je suis heureux de pouvoir écrire qu'en 1 an, ce protocole a été utilisé à ma connaissance pour 5 autres projets que le mien dont deux qui ne sont pas du laboratoire.

Pour l'analyse de données de métabarcoding, le constat a été le même que pour le barcoding : il n'y avait pas de protocole disponible au laboratoire pour analyser les données 'en routine', chaque projet ayant tenté de développer avec plus ou moins de succès, ses propres protocoles d'analyses. Etant un grand passionné de technologie et d'informatique, après quelques recherches sur internet, je suis tombé sur l'article qui a lâchement été recopié en introduction (article tiré de bioinfo-fr). Ce n'était pas pour rajouter du contenu à mon manuscrit, mais pour montrer à ceux qui liraient cette thèse que deux profils sur les 6 énoncés font entrer en jeu des biologistes. Certains, armés de volonté, peuvent y arriver. C'est de cette lecture qu'est née l'envie de créer les scripts « tout faits » pour l'utilisation de Mothur, Qiime et FROGS, que ce soit pour simple machine de bureau (un minimum puissante tout de même) ou pour ordinateur. Encore une fois j'ai fait le choix personnel de développer ce travail pour que d'autres l'utilisent. Pour que ceux qui préfèrent s'attarder sur les problématiques scientifiques plutôt que techniques gagnent un temps précieux. Ainsi, peut-être, le temps gagné permettra-t-il de comprendre plus vite les environnements profonds, avant leur destruction potentielle.

Conclusion générale

A l'instar du modèle phare qu'est *Rimicaris exoculata*, crevette au succès écologique (et médiatique) à la fois par son régime alimentaire et la densité des populations rencontrées au niveau des systèmes hydrothermaux, *Rimicaris chacei* est restée jusqu'à ce jour dans l'ombre de sa cousine. Une crevette qui ne prend pas la « grosse tête », semblant disposer du même potentiel symbiotique avec les mêmes lignées microbiennes que sa voisine, à la fois discrète et disposant d'un atout non négligeable, un régime trophique mixte, entre symbiose et nécrophagie. Dans un contexte écologique et économique où l'homme détruit l'environnement avec comme future cible les systèmes hydrothermaux, la plasticité que semble présenter *Rimicaris chacei* lui permettra-t-elle de devenir la future star de ces environnements de l'océan Atlantique, ou la bataille pour le rôle de crevette en titre sera-t-il plus difficile à obtenir que prévu ? L'analyse de la biogéographie des symbiontes en fonction des sites hydrothermaux partagés entre ces deux espèces scellera-t-il le destin de *Rimicaris chacei* ?...

La suite dans une prochaine thèse ?

Bibliographie

Aida M, Kanemori M, Kubota N, Matada M, Sasayama Y & Fukumori Y Distribution and Population of Free-Living Cells Related to Endosymbiont A Harbored in *Oligobranchia mashikoi* (a Siboglinid Polychaete) Inhabiting Tsukumo Bay. *Microbes and Environments* (2008);**23**:81-88.

Alain K, Postec A, Grinsard E, Lesongeur F, Prieur D & Godfroy A *Thermodesulfatator atlanticus* sp. nov., a thermophilic, chemolithoautotrophic, sulfate-reducing bacterium isolated from a Mid-Atlantic Ridge hydrothermal vent. *International Journal of Systematic and Evolutionary Microbiology* (2010);**60**:33-38.

Alain K, Rolland S, Crassous P, et al. *Desulfurobacterium crinifex* sp. nov., a novel thermophilic, pinkish-streamer forming, chemolithoautotrophic bacterium isolated from a Juan de Fuca Ridge hydrothermal vent and amendment of the genus *Desulfurobacterium*. *Extremophiles* (2003);**7**:361-370.

Alt JC (1995) Subseafloor Processes in Mid-Ocean Ridge Hydrothermal Systems. p. 85-114.

Anderson TR & Rice T Deserts on the sea floor: Edward Forbes and his azoic hypothesis for a lifeless deep ocean. *Endeavour* (2006);**30**:131-137.

Anthony C *Biochemistry of methylotrophs*. Academic Press, (1982).

Beh M, Strauss G, Huber R, Stetter KO & Fuchs G Enzymes of the reductive citric acid cycle in the autotrophic eubacterium *Aquifex pyrophilus* and in the archaebacterium *Thermoproteus neutrophilus*. *Archives of Microbiology* (1993);**160**:306.

Bischoff JL & Rosenbauer RJ Liquid-vapor relations in the critical region of the system NaCl-H₂O from 380 to 415°C: A refined determination of the critical point and two-phase boundary of seawater. *Geochimica et Cosmochimica Acta* (1988);**52**:2121-2126.

Bogdanov Y, Sagalevitch A, Chernyaev E, Ashadze A, Gurvich E, Lukashin V, Ivanov G & Peresyphkin V A study of the hydrothermal field at 14°45' N on the Mid-Atlantic Ridge using the "MIR" submersibles. *BRIDGE newsletter*, 1995 (1995);**9**:9-13.

Bright M & Bulgheresi S A complex journey: transmission of microbial symbionts. *Nature Reviews Microbiology* (2010);**8**:218-230.

Buchanan BB & Arnon DI A reverse KREBS cycle in photosynthesis: consensus at last. *Photosynthesis Research* (1990);**24**:47-53.

Calgano R *Les grands fonds marins*. (2011).

Campbell BJ, Engel AS, Porter ML & Takai K The versatile *epsilon*-proteobacteria: key players in sulphidic habitats. *Nature Reviews Microbiology* (2006);**4**:458-468.

Carney RS Consideration of the oasis analogy for chemosynthetic communities at Gulf of Mexico hydrocarbon vents. *Geo-Marine Letters* (1994);**14**:149-159.

Casanova B, Brunet M & Segonzac M Impact of bacterial epibiosis on functional-morphology of shrimp associated with the Mid-Atlantic hydrothermal conditions. *Cahier de Biologie Marine* (1993);**34**:573-588.

Chamberlain SC Vision in hydrothermal vent shrimp. *Philosophical transactions of the Royal Society of London B* (2000);**355**:1151-1154.

Charlou JL, Donval JP, Fouquet Y, Jean-Baptiste P & Holm H Geochemistry of high H₂ and CH₄ vent fluids issuing from ultramafic rocks at the Rainbow hydrothermal field (36°14' N, MAR). *Chemical Geology* (2002);**191**:345-359.

Cleveland LR & Grimstone AV The fine structure of the flagellate *Mixotricha paradoxa* and its associated micro-organisms. *Proceedings of the Royal society of London B : Biological Sciences* (1964);**159**:668-686.

Copley J, Tyler P, Murton B & Van Dover C Spatial and interannual variation in the faunal distribution at Broken Spur vent field (29°N, Mid-Atlantic Ridge). *Marine Biology* (1997);**129**:723-733.

Corbari L, Zbinden M, Cambon-Bonavita M-A, Gaill F & Compère P Bacterial symbionts and mineral deposits in the branchial chamber of the hydrothermal vent shrimp *Rimicaris exoculata*: relationship to moult cycle. *Aquat Biol* (2008);**1**:225-238.

Corbari L, Cambon-Bonavita M-A, Long GJ, Grandjean F, Zbinden M, Gaill F & Compère P Iron oxide deposits associated with the ectosymbiotic bacteria in the hydrothermal vent shrimp *Rimicaris exoculata*. *Biogeosciences Discussions* (2008);**5**:1825-1865.

Corliss JB, Dymond J, Gordon LI, *et al.* Submarine thermal springs on the galapagos rift. *Science* (1979);**203**:1073-1083.

Costa JL, Paulsrud P, Rikkinen J & Linblad P Genetic diversity of *Nostoc* symbionts endophytically associated with two bryophyte species. *Applied and Environmental Microbiology* (2001);**67**:4393-4396.

Crépeau V, Cambon-Bonavita M-A, Lesongeur F, Randrianalivelo H, Sarradin P-M, Sarrazin J & Godfroy A Diversity and function in microbial mats from the Lucky Strike hydrothermal vent field. *FEMS microbiology ecology* (2011);**76**:524-540.

Deckert G, Warren PV, Gaasterland T, *et al.* The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* (1998);**392**:353-358.

Dedysh SN, Liesack W, Khmelenina VN, Suzina NE, Trotsenko YA, Semrau JD, Bares AM, Panikov NS & Tiedje JM *Methylocella palustris* gen. nov., sp. nov., a new methane-oxidizing acidophilic bacterium from peat bogs, representing a novel subtype of serine-pathway methanotrophs. *International Journal of Systematic and Evolutionary Microbiology* (2000);**50**:955-969.

Delaney JR, Robigou V, McDuff RE & Tivey MK Geology of a vigorous hydrothermal system on the Endeavour Segment, Juan de Fuca Ridge. *Journal of Geophysical Research: Solid Earth* (1992);**97**:19663-19682.

Desbruyères D & Laubier L (1983) Primary consumers from hydrothermal vents animal communities. In *Hydrothermal processes at seafloor spreading centers*, Springer, 711-734.

Desbruyères D & Segonzac M *Handbook of deep-sea hydrothermal vent fauna*. Editions Quae, (1997).

Desbruyères D, Almeida A, Biscoito M, Comtet T, Khripounoff A, Le Bris N, Sarradin P & Segonzac M (2000) A review of the distribution of hydrothermal vent communities along the northern Mid-

Atlantic Ridge: dispersal vs. environmental controls. In *Island, Ocean and Deep-Sea Biology*, Vol. 152 Springer, 201-216.

Desbruyères D, Biscoito M, Caprais J-C, Colaço A, Comtet T, Crassous P, Fouquet Y, Khripounoff A, Le Bris N & Olu K Variations in deep-sea hydrothermal vent communities on the Mid-Atlantic Ridge near the Azores plateau. *Deep Sea Research Part I: Oceanographic Research Papers* (2001);**48**:1325-1346.

Desbruyères D, Chevaldonné P, Alayse AM, *et al.* Biology and ecology of the “Pompeii worm” (*Alvinella pompejana* Desbruyères and Laubier), a normal dweller of an extreme deep-sea environment: A synthesis of current knowledge and recent developments. *Deep Sea Research Part II: Topical Studies in Oceanography* (1998);**45**:383-422.

Douville E, Charlou JL, Oelkers EH, Bienvenu P, Jove Colon CF, Donval JP, Fouquet Y, Prieur D & Appriou P The rainbow vent fluids (36°14' N, MAR): the influence of ultramafic rocks and phase separation on trace metal content in Mid-Atlantic Ridge hydrothermal fluids. *Chemical Geology* (2002);**184**:37-48.

Duperron S, Halary S, Lorion J, Sibuet M & Gaill F Unexpected co-occurrence of six bacterial symbionts in the gills of the cold seep mussel *Idas sp.* (Bivalvia: Mytilidae). *Environmental Microbiology* (2008);**10**:433-445.

Duperron S, Sibuet M, MacGregor BJ, Kuypers MMM, Fisher CR & Dubilier N Diversity, relative abundance and metabolic potential of bacterial endosymbionts in three *Bathymodiolus* mussel species from cold seeps in the Gulf of Mexico. *Environmental Microbiology* (2007);**9**:1423-1438.

Duperron S, Nadalig T, Caprais JC, Sibuet M, Fiala-Médioni A, Amann R & Dubilier N Dual symbiosis in a *Bathymodiolus sp.* mussel from a methane seep on the Gabon continental margin (Southeast Atlantic): 16S rRNA phylogeny and distribution of the Symbiont in Gills. *Applied and Environmental Microbiology* (2005);**71**:1694-1700.

Duperron S, Bergin C, Zielinski F, Blazejak A, Pernthaler A, McKiness ZP, DeChaine E, Cavanaugh CM & Dubilier N A dual symbiosis shared by two mussel species, *Bathymodiolus azoricus* and *Bathymodiolus puteoserpentis* (Bivalvia: Mytilidae), from hydrothermal vents along the northern Mid-Atlantic Ridge. *Environmental Microbiology* (2006);**8**:1441-1447.

Durand L, Zbinden M, Cueff-Gauchard V, Duperron S, Roussel EG, Shillito B & Cambon-Bonavita M-A Microbial diversity associated with the hydrothermal shrimp *Rimicaris exoculata* gut and occurrence of a resident microbial community. *FEMS microbiology ecology* (2010);**71**:291-303.

Elsaied H & Naganuma T Phylogenetic diversity of ribulose-1,5-bisphosphate carboxylase/oxygenase large-subunit genes from deep-sea microorganisms. *Applied and Environmental Microbiology* (2001);**67**:1751-1765.

Elsaied HE, Kimura H & Naganuma T Composition of archaeal, bacterial, and eukaryal RuBisCO genotypes in three Western Pacific arc hydrothermal vent systems. *Extremophiles* (2006);**11**:191-202.

Emerson D & Moyer CL Neutrophilic Fe-oxidizing bacteria are abundant at the Loihi Seamount hydrothermal vents and play a major role in Fe oxide deposition. *Applied and Environmental Microbiology* (2002);**68**:3085-3093.

Emerson D, Rentz JA, Lilburn TG, Davis RE, Aldrich H, Chan C & Moyer CL A novel lineage of *proteobacteria* involved in formation of marine Fe-oxidizing microbial mat communities. *PLoS ONE* (2007);**2**:e667.

Eren AM, Borisy GG, Huse SM & Welch JLM Oligotyping analysis of the human oral microbiome. *Proceedings of the National Academy of Sciences* (2014);**111**:E2875-E2884.

Eren AM, Zozaya M, Taylor CM, Dowd SE, Martin DH & Ferris MJ Exploring the diversity of *Gardnerella vaginalis* in the genitourinary tract microbiota of monogamous couples through subtle nucleotide variation. *PloS one* (2011);**6**:e26732.

Eren AM, Morrison HG, Lescault PJ, Reveillaud J, Vineis JH & Sogin ML Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *The ISME Journal* (2014);**9**:968-979.

Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG & Sogin ML Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods in Ecology and Evolution* (2013);**4**:1111-1119.

Evans MC, Buchanan BB & Arnon DI A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. *Proceedings of the National Academy of Sciences* (1966);**55**:928-934.

Fabri M-C, Bargain A, Briand P, Gebruk A, Fouquet Y, Morineaux M & Desbruyeres D The hydrothermal vent community of a new deep-sea field, Ashadze-1, 12° 58' N on the Mid-Atlantic Ridge. *Journal of the Marine Biological Association of the United Kingdom* (2011);**91**:1-13.

Felbeck H Chemoautotrophic potential of the hydrothermal vent tube worm, *Riftia pachyptila* Jones (Vestimentifera). *Science* (1981);**213**:336-338.

Fisher CR Chemoautotrophic and methanotrophic symbioses in marine-invertebrates. *Reviews in Aquatic Sciences* (1990);**2**:399-436.

Flores GE, Campbell JH, Kirshtein JD, *et al.* Microbial community structure of hydrothermal deposits from geochemically different vent fields along the Mid-Atlantic Ridge. *Environmental Microbiology* (2011);**13**:2158-2171.

Forbes E *Report on the Mollusca and Radiata of the Aegean Sea: And on Their Distribution, Considered as Bearing on Geology.* (1843).

Forbes E *On the Light Thrown on Geology by Submarine Researches: Being the Substance of a Communication Made to the Royal Institution of Great Britain, friday evening, the 23d february 1844.* (1844).

Forbes E *The Natural History of the European Seas.* J. Van Voorst, (1859).

Foucault A & Raoult J-F *Dictionnaire de Géologie - 7e édition.* Dunod, (2010).

Fouquet Y, Wafik A, Cambon P, Mevel C, Meyer G & Gente P Tectonic setting and mineralogical and geochemical zonation in the Snake Pit sulfide deposit (Mid-Atlantic Ridge at 23 degrees N). *Economic geology* (1993);**88**:2018-2036.

Fouquet Y, Charlou JL, Ondreas H, Radford-Knoery J, Donval JP, Douville E & Normand A Discovery and first submersible investigations on the Rainbow hydrothermal field on the MAR (36°14' N). *Eos Trans. AGU*, 1997 (1997);**78**:832.

Fouquet Y, Charlou J-L, Donval J-P, Radford-Knoery J, Pelle P, Ondreas H, Lourenco N, Segonzac M & Tivey M A detailed study of the Lucky-Strike hydrothermal site and discovery of a new hydrothermal site:«Menez-Gwen». Preliminary results of DIVA 1 cruise (5-29 May, 1994). *Inter-Ridge News* (1994);**3**:14-17.

Fouquet Y, Cherkashov G, Charlou J, Ondréas H, Birot D, Cannat M, Bortnikov N, Silantyev S, Sudarikov S & Cambon-Bonavita M Serpentine cruise-ultramafic hosted hydrothermal deposits on the Mid-At-lantic Ridge: First submersible studies on Ashadze 1 and 2, Logatchev 2 and Krasnov vent fields. *InterRidge News* (2008);**17**:16-21.

Fouquet Y, Cambon P, Etoubleau J, *et al.* (2013) 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. In *Diversity Of Hydrothermal Systems On Slow Spreading Ocean Ridges*, American Geophysical Union, 321-367.

Fricke H, Giere O, Stetter K, Alfredsson GA, Kristjansson JK, Stoffers P & Svavarsson J Hydrothermal vent communities at the shallow subpolar Mid-Atlantic ridge. *Marine Biology* (1989);**102**:425-429.

Fuse H, Ohta M, Takimura O, Murakami K, Inoue H, Yamaoka Y, Oclarit JM & Omori T Oxidation of trichloroethylene and dimethyl sulfide by a marine *Methylobacterium* strain containing soluble methane monooxygenase. *Bioscience, biotechnology, and biochemistry* (1998);**62**:1925-1931.

Gaill F, Desbruyères D & Prieur D Bacterial Communities Associated with "Pompei Worms" from the East Pacific Rise Hydrothermal Vents: SEM, TEM Observations. *Microbial Ecology* (1987);**13**:129-139.

Gebruk A, Southward E, Kennedy H & Southward A Food sources, behaviour, and distribution of hydrothermal vent shrimps at the Mid-Atlantic Ridge. *Journal of the Marine Biological Association of the UK* (2000);**80**:485-499.

Gebruk A, Galkin S, Miller Y, Pimenov N, Moskalev L & Ivanov M Trophic structure of the Broken Spur hydrothermal community shown by carbon stable isotope and C: H: N: S data. *Bridge News* (1997);**12**:40-44.

Gebruk AV, Galkin S, Vereshchaka A, Moskalev L & Southward A Ecology and biogeography of the hydrothermal vent fauna of the Mid-Atlantic Ridge. *Advances in Marine Biology* (1997);**32**:93-144.

Gebruk AV, Chevaldonné P, Shank T, Lutz RA & Vrijenhoek RC Deep-sea hydrothermal vent communities of the Logatchev area (14°45'N, Mid-Atlantic Ridge): diverse biotopes and high biomass. *Journal of the Marine Biological Association of the United Kingdom* (2000);**80**:383-393.

Giovannelli D, Chung M, Staley J, Starovoytov V, Le Bris N & Vetriani C *Sulfurovum riftiae* sp. nov., a mesophilic, thiosulfate-oxidizing, nitrate-reducing chemolithoautotrophic Epsilonproteobacterium isolated from the tube of the deep-sea hydrothermal vent polychaete, *Riftia pachyptila*. *International Journal of Systematic and Evolutionary Microbiology* (2016).

Gloter A, Zbinden M, Guyot F, Gaill F & Colliex C TEM-EELS study of natural ferrihydrite from geological–biological interactions in hydrothermal systems. *Earth and Planetary Science Letters* (2004);**222**:947-957.

Glover AG & Smith CR The deep-sea floor ecosystem: current status and prospects of anthropogenic change by the year 2025. *Environmental Conservation* (2003);**30**:219-241.

Grassle JF & Maciolek NJ Deep-sea species richness: regional and local diversity estimates from quantitative bottom samples. *American naturalist* (1992);**139**:313-341.

Gros O, Liberge M, Heddi A, Khatchadourian C & Felbeck H Detection of the free-living forms of sulfide-oxidizing gill endosymbionts in the lucinid habitat (*Thalassia testudinum* environment). *Applied and Environmental Microbiology* (2003);**69**:6264-6267.

Guinot D & Hurtado LA Two new species of hydrothermal vent crabs of the genus *Bythograea* from the southern East Pacific Rise and from the Galapagos Rift (Crustacea Decapoda *Brachyura Bythograeidae*). *Comptes Rendus Biologies* (2003);**326**:423-439.

Guinot D, Hurtado LA & Vrijenhoek R New genus and species of brachyuran crab from the southern East Pacific Rise (Crustacea Decapoda *Brachyura Bythograeidae*). *Comptes rendus biologies* (2003);**325**:1143-1152.

Guri M, Durand L, Cueff-Gauchard V, Zbinden M, Crassous P, Shillito B & Cambon-Bonavita M-A Acquisition of epibiotic bacteria along the life cycle of the hydrothermal shrimp *Rimicaris exoculata*. *The ISME journal* (2012);**6**:597-609.

Hannington MD, Jonasson IR, Herzig PM & Petersen S Physical and chemical processes of seafloor mineralization at mid-ocean ridges. *Washington DC American Geophysical Union Geophysical Monograph Series* (1995);**91**:115-157.

Haygood MG The potential role of functional differences between Rubisco forms in governing expression in chemoautotrophic symbioses. *Limnology and oceanography* (1996);**41**:370-371.

Haymon RM Growth history of hydrothermal black smoker chimneys. *Nature* (1983);**301**:695-698.

Hékinian R & Binard N *Le feu des abysses*. Editions Quae, (2008).

Hernández-Avila I, Cambon-Bonavita MA & Pradillon F Morphology of First Zoeal Stage of Four Genera of Alvinocaridid Shrimps from Hydrothermal Vents and Cold Seeps: Implications for Ecology, Larval Biology and Phylogeny. *PloS one* (2015);**10**:e0144657.

Hernández-Ávila I, Cambon-Bonavita M-A & Pradillon F Morphology of First Zoeal Stage of Four Genera of Alvinocaridid Shrimps from Hydrothermal Vents and Cold Seeps: Implications for Ecology, Larval Biology and Phylogeny. *PLOS ONE* (2016);**10**:e0144657.

Herring P *The Biology of the Deep Ocean*. Oxford University Press, (2002).

Huber H, Burggraf S, Mayer T, Wyschkony I, Rachel R & Stetter KO *Ignicoccus* gen. nov., a novel genus of hyperthermophilic, chemolithoautotrophic Archaea, represented by two new species, *Ignicoccus islandicus* sp nov and *Ignicoccus pacificus* sp nov. and *Ignicoccus pacificus* sp. nov. *International Journal of Systematic and Evolutionary Microbiology* (2001);**50** 2093-2100.

Hügler M, Wirsén CO, Fuchs G, Taylor CD & Sievert SM Evidence for autotrophic CO₂ fixation via the reductive tricarboxylic acid cycle by members of the epsilon subdivision of *proteobacteria*. *Journal of Bacteriology* (2005);**187**:3020-3027.

Hügler M, Petersen JM, Dubilier N, Imhoff JF & Sievert SM Pathways of carbon and energy metabolism of the epibiotic community associated with the deep-sea hydrothermal vent shrimp *Rimicaris exoculata*. *PLoS One* (2011);**6**:e16018.

Jan C, Petersen JM, Werner J, Teeling H, Huang S, Glöckner FO, Golyshina OV, Dubilier N, Golyshin PN & Jebbar M The gill chamber epibiosis of deep-sea shrimp *Rimicaris exoculata*: an in-depth metagenomic investigation and discovery of *Zetaproteobacteria*. *Environmental microbiology* (2014);**16**:2723-2738.

Jannasch HW & Mottl MJ Geomicrobiology of deep-sea hydrothermal vents. *Science* (1985);**229**:717-725.

Johnson KS, Beehler CL, Sakamoto-Arnold CM & Childress JJ In situ measurements of chemical distributions in a deep-sea hydrothermal vent field. *Science* (1986);**231**:1139-1141.

Johnson SB, Young CR, Jones WJ, Warén A & Vrijenhoek RC Migration, isolation, and speciation of hydrothermal vent limpets (Gastropoda; *Lepetodrilidae*) across the Blanco Transform Fault. *The Biological Bulletin* (2006);**210**:140-157.

Karl DM *Ecology of free-living, hydrothermal vent microbial communities*. (1995).

Karson JA & Dick H Tectonics of ridge-transform intersections at the Kane Fracture Zone. *Marine Geophysical Researches* (1983);**6**:51-98.

Kashefi K & Lovley DR Extending the upper temperature limit for life. *Science* (2003);**301**:934.

Kashefi K, Holmes DE, Reysenbach AL & Lovley DR Use of Fe (III) as an electron acceptor to recover previously uncultured hyperthermophiles: isolation and characterization of *Geothermobacterium ferrireducens* gen. nov., sp. nov. *Applied and Environmental Microbiology* (2002);**68**:1735-1742.

Kelley DS, Baross JA & Delaney JR Volcanoes, fluids, and life at mid-ocean ridge spreading centers. *Annual Review of Earth and Planetary Sciences* (2002);**30**:385-491.

Kelley DS, Karson JA, Blackman DK, *et al.* An off-axis hydrothermal vent field near the Mid-Atlantic Ridge at 30°N. *Nature* (2001);**412**:145-149.

Kerr RC Heat transfer and hydrothermal fluid flow at flanges on large seafloor sulphide structures. *Earth and Planetary Science Letters* (1997);**152**:93-99.

Kiers ET, Rousseau RA, West SA & Denison RF Host sanctions and the legume-rhizobium mutualism. *Nature* (2003);**425**:78-81.

Komai T & Segonzac M Taxonomic review of the hydrothermal vent shrimp genera *Rimicaris* Williams & Rona and *Chorocaris* Martin & Hessler (Crustacea: Decapoda: Caridea: Alvinocarididae). *Journal of Shellfish Research* (2008);**27**:21-41.

Koschinsky A, Billings A, Devey C, Dubilier N, Duester A, Edge D, Garbe-Schönberg D, German C, Giere O & Keir R Discovery of new hydrothermal vents on the southern Mid-Atlantic Ridge (4°S–10°S) during cruise M68/1. *InterRidge News* (2006);**15**:9-15.

L'Haridon S, Cilia V, Messner P, Ragúenès G, Gambacorta A, Sleytr UB, Prieur D & Jeanthon C *Desulfurobacterium thermolithotrophum* gen. nov., sp. nov., a novel autotrophic, sulphur-reducing bacterium isolated from a deep-sea hydrothermal vent. *International journal of systematic bacteriology* (1998);**48** 701-711.

L'Haridon S, Reysenbach AL, Tindall BJ, Schönheit P, Banta A, Johnsen U, Schumann P, Gambacorta A, Stackebrandt E & Jeanthon C *Desulfurobacterium atlanticum* sp. nov., *Desulfurobacterium pacificum* sp. nov. and *Thermovibrio guaymasensis* sp. nov., three thermophilic members of the *Desulfurobacteriaceae* fam. nov., a deep branching lineage within the Bacteria. *International Journal of Systematic and Evolutionary Microbiology* (2006);**56**:2843-2852.

Lalou C, Reyss JL, Brichet E, Rona PA & Thompson G Hydrothermal activity on a 105-year scale at a slow-spreading ridge, TAG hydrothermal field, Mid-Atlantic Ridge 26° N. *Journal of Geophysical Research: Solid Earth* (1995);**100**:17855-17862.

Lee KH & Ruby EG Detection of the light organ symbiont, *Vibrio fischeri*, in Hawaiian seawater by using lux gene probes. *Applied and environmental microbiology* (1992);**58**:942-947.

Lipsitch M, Nowak MA, Ebert D & May RM The population dynamics of vertically and horizontally transmitted parasites. *Proceedings of the Royal Society of London B: Biological Sciences* (1995);**260**:321-327.

Lisitsyn AP, Bogdanov YA, Zonenshayn LP, Kuz'min MI & Podrazhanskiy AM "Black smokers" in the gulf of California. *International Geology Review* (1989);**31**:541-555.

Little CT, Danelian T, Herrington RJ & Haymon RM Early Jurassic hydrothermal vent community from the Franciscan Complex, California. *Journal of Paleontology* (2004);**78**:542-559.

Llodra ER, Tyler PA & Copley JT Reproductive biology of three caridean shrimp, *Rimicaris exoculata*, *Chorocaris chacei* and *Mirocaris fortunata* (Caridea: Decapoda), from hydrothermal vents. *Journal of the Marine Biological Association of the United Kingdom* (2000);**80**:473-484.

Lonsdale P Clustering of suspension-feeding macrobenthos near abyssal hydrothermal vents at oceanic spreading centers. *Deep Sea Research* (1977);**24**:857-863.

Lonsdale P & Becker K Hydrothermal plumes, hot springs, and conductive heat flow in the Southern Trough of Guaymas Basin. *Earth and Planetary Science Letters* (1985);**73**:211-225.

Lovley DR & Phillips EJ Availability of ferric iron for microbial reduction in bottom sediments of the freshwater tidal potomac river. *Applied and Environmental Microbiology* (1986);**52**:751-757.

Lovley DR, Holmes DE & Nevin KP Dissimilatory Fe(III) and Mn(IV) Reduction. *Advances in Microbial Physiology* (2004);**49**:219-286.

Maas PA, O'Mullan GD, Lutz RA & Vrijenhoek RC Genetic and morphometric characterization of mussels (Bivalvia: Mytilidae) from mid-atlantic hydrothermal vents. *The Biological Bulletin* (1999);**196**:265-272.

MacDonald IR, Boland GS, Baker JS, Brooks JM, Kennicutt II MC & Bidigare RR Gulf of Mexico hydrocarbon seep communities. *Marine Biology* (1989);**101**:235-247.

Madigan MT, Martinko JM & Parker J *Brock Biology of Microorganisms*. Upper Saddle River, NJ: prentice hall, (2000).

Markert S, Arndt C, Felbeck H, *et al.* Physiological proteomics of the uncultured endosymbiont of *Riftia pachyptila*. *Science* (2007);**315**:247-250.

Marsh AG, Mullineaux LS, Young CM & Manahan DT Larval dispersal potential of the tubeworm *Riftia pachyptila* at deep-sea hydrothermal vents. *Nature* (2001);**411**:77-80.

Martin JW & Hessler RR *Chorocaris vandoverae*, a new genus and species of hydrothermal vent shrimp (Crustacea, Decapoda, Bresiliidae) from the western Pacific. *Natural History Museum of Los Angeles County* (1990);**417**:1-11.

Martin JW, Signorovitch J & Patel H A new species of *Rimicaris* (Crustacea: Decapoda: Bresiliidae) from the Snake Pit hydrothermal vent field on the Mid-Atlantic Ridge. *Proceedings-Biological society of Washington* (1997);**110**:399-411.

Meyer B & Kuever J Molecular analysis of the diversity of sulfate-reducing and sulfur-oxidizing prokaryotes in the environment, using *aprA* as functional marker gene. *Applied and Environmental Microbiology* (2007);**73**:7664-7679.

Meyer B & Kuever J Phylogeny of the alpha and beta subunits of the dissimilatory adenosine-5'-phosphosulfate (APS) reductase from sulfate-reducing prokaryotes—origin and evolution of the dissimilatory sulfate-reduction pathway. *Microbiology* (2007);**153**:2026-2044.

Meyer B & Kuever J Molecular analysis of the distribution and phylogeny of dissimilatory adenosine-5'-phosphosulfate reductase-encoding genes (*aprBA*) among sulfur-oxidizing prokaryotes. *Microbiology* (2007);**153**:3478-3498.

Morita RY Is H₂ the universal energy source for long-term survival? *Microbial ecology* (1999);**38**:307-320.

Nakagawa S, Takaki Y, Shimamura S, Reysenbach A-L, Takai K & Horikoshi K Deep-sea vent epsilon-proteobacterial genomes provide insights into emergence of pathogens. *Proceedings of the National Academy of Sciences* (2007);**104**:12146-12150.

Nealson KH Sediment bacteria: who's there, what are they doing, and what's new? *Annual Review of Earth and Planetary Sciences* (1997);**25**:403-434.

Norton DL Theory of hydrothermal systems. *Annual Review of Earth and Planetary Sciences* (1984);**12**:155.

Nuckley DJ, Jinks RN, Battelle B-A, Herzog ED, Kass L, Renninger GH & Chamberlain SC Retinal Anatomy of a New Species of Bresiliid Shrimp from a Hydrothermal Vent Field on the Mid-Atlantic Ridge. *Biological Bulletin* (1996);**190**:98-110.

Nye V, Copley J & Plouviez S A new species of *Rimicaris* (Crustacea: Decapoda: Caridea: Alvinocarididae) from hydrothermal vent fields on the Mid-Cayman Spreading Centre, Caribbean. *Journal of the Marine Biological Association of the United Kingdom* (2012);**92**:1057-1072.

O'Mullan GD, Maas P, Lutz RA & Vrijenhoek RC A hybrid zone between hydrothermal vent mussels (Bivalvia: Mytilidae) from the Mid-Atlantic Ridge. *Molecular Ecology* (2001);**10**:2819-2831.

O'Neill PJ, Jinks RN, Herzog ED, Battelle BA, Kass L, Renninger GH & Chamberlain SC The morphology of the dorsal eye of the hydrothermal vent shrimp, *Rimicaris exoculata*. *Visual Neurosci* (1995);**12**:861-875.

Paull CK, Hecker B, Commeau R, Freeman-Lynde RP, Neumann C, Corso WP, Golubic S, Hook JE, Sikes E & Curray J Biological communities at the Florida escarpment resemble hydrothermal vent taxa. *Science* (1984);**226**:965-967.

Pelli DG & Chamberlain SC The visibility of 350 C black-body radiation by the shrimp *Rimicaris exoculata* and man. *Nature* (1989);**337**:460-461.

Petersen JM, Ramette A, Lott C, Cambon-Bonavita MA, Zbinden M & Dubilier N Dual symbiosis of the vent shrimp *Rimicaris exoculata* with filamentous *gamma*- and *epsilon*-proteobacteria at four Mid-Atlantic Ridge hydrothermal vent fields. *Environmental microbiology* (2010);**12**:2204-2218.

Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R, Hourdez S, Girguis PR, Wankel SD & Barbe V Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* (2011);**476**:176-180.

Pikuta E, Lysenko A, Suzina N, Osipov G, Kuznetsov B, Tourova T, Akimenko V & Laurinavichius K *Desulfotomaculum alkaliphilum* sp. nov., a new alkaliphilic, moderately thermophilic, sulfate-reducing bacterium. *International Journal of Systematic and Evolutionary Microbiology* (2000);**50** 25-33.

Pockalny RA, Detrick RS & Fox PJ Morphology and tectonics of the Kane transform from Sea Beam bathymetry data. *Journal of Geophysical Research* (1988);**93**:3179-3193.

Polz MF & Cavanaugh CM Dominance of one bacterial phylotype at a Mid-Atlantic Ridge hydrothermal vent site. *Proceedings of the National Academy of Sciences* (1995);**92**:7232-7236.

Polz MF, Robinson JJ, Cavanaugh CM & Van Dover CL Trophic Ecology of Massive Shrimp Aggregations at a Mid-Atlantic Ridge Hydrothermal Vent Site. *Limnology and Oceanography* (1998);**43**:1631-1638.

Pond DW, Dixon DR, Bell MV, Fallick AE & Sargent JR Occurrence of 16: 2 (n-4) and 18: 2 (n-4) fatty acids in the lipids of the hydrothermal vent shrimps *Rimicaris exoculata* and *Alvinocaris markensis*: nutritional and trophic implications. *Mar Ecol Prog Ser* (1998);**156**:167-174.

Ponsard J, Cambon-Bonavita MA, Zbinden M, Lepoint G, Joassin A, Corbari L, Shillito B, Durand L, Cuff-Gauchard V & Compère P Inorganic carbon fixation by chemosynthetic ectosymbionts and nutritional transfers to the hydrothermal vent host-shrimp *Rimicaris exoculata*. *Isme J* (2013);**7**:96-109.

Ponsard J, Cambon-Bonavita M-A, Zbinden M, Lepoint G, Joassin A, Corbari L, Shillito B, Durand L, Cuff-Gauchard V & Compère P Inorganic carbon fixation by chemosynthetic ectosymbionts and nutritional transfers to the hydrothermal vent host-shrimp *Rimicaris exoculata*. *The ISME journal* (2013);**7**:96-109.

Prieur D Un enfer très fécond. *Biofutur* (1998);**1998**:25-27.

Quatrini R, Martínez V & Osorio H Iron homeostasis strategies in acidophilic iron oxidizers: comparative genomic analyses. *Advanced Materials Research* (2007);**20-21**:531.

Rabus R, Hansen TA & Widdel F Dissimilatory sulfate-and sulfur-reducing prokaryotes. *The Prokaryotes* (2006);**2**:659-768.

Ramirez-Llodra E, Tyler PA, Baker MC, Bergstad OA, Clark MR, Escobar E, Levin LA, Menot L, Rowden AA & Smith CR Man and the last great wilderness: human impact on the deep sea. *PLoS One* (2011);**6**:e22588.

Reysenbach A-L, Hamamura N, Podar M, *et al.* Complete and draft genome sequences of six members of the *Aquificales*. *Journal of Bacteriology* (2009);**191**:1992-1993.

Riley GA, Van Hemert D & Wangersky PJ Organic aggregates in surface and deep waters of the Sargasso Sea. *Limnol. Oceanogr* (1965);**10**:345-363.

Robinson JJ & Cavanaugh CM Expression of form I and form II Rubisco in chemoautotrophic symbioses: implications for the interpretation of stable carbon isotope values. *Limnology and Oceanography* (1995);**40**:1496-1502.

Rona PA, Hannington MD, Raman C, Thompson G, Tivey MK, Humphris SE, Lalou C & Petersen S Active and Relict Sea-Floor Hydrothermal Mineralization at the TAG Hydrothermal Field, Mid-Atlantic Ridge. *Economic Geology* (1993);**88**:1989-1989.

Ross J *A voyage of discovery: made under the orders of the Admiralty, in His Majesty's ships Isabella and Alexander, for the purpose of exploring Baffin's Bay, and inquiring into the probability of a north-west passage.* J. Murray, (1819).

Salerno JL, Macko SA, Hallam SJ, Bright M, Won Y-J, McKiness Z & Van Dover CL Characterization of Symbiont Populations in Life-History Stages of Mussels from Chemosynthetic Environments. *Biological Bulletin* (2005);**208**:145-155.

Sarrazin J, Robigou V, Juniper SK & Delaney JR Biological and geological dynamics over four years on a high-temperature sulfide structure at the Juan de Fuca Ridge hydrothermal observatory. *Mar Ecol Prog Ser* (1997);**153**:5-24.

Schmidt C, Le Bris N & Gaill F Interactions of deep-sea vent invertebrates with their environment: the case of *Rimicaris exoculata*. *Journal of Shellfish Research* (2008);**27**:79-90.

Schmidt K, Koschinsky A, Garbe-Schönberg D, De Carvalho LM & Seifert R Geochemistry of hydrothermal fluids from the ultramafic-hosted Logatchev hydrothermal field, 15°N on the Mid-Atlantic Ridge: Temporal and spatial investigation. *Chemical Geology* (2007);**242**:1-21.

Scott KM, Schwedock J, Schrag DP & Cavanaugh CM Influence of form IA RubisCO and environmental dissolved inorganic carbon on the $\delta^{13}\text{C}$ of the clam-chemoautotroph symbiosis *Solemya velum*. *Environmental Microbiology* (2004);**6**:1210-1219.

Scott KM, Henn-Sax M, Harmer TL, Longo DL, Frame CH & Cavanaugh CM Kinetic isotope effect and biochemical characterization of form IA RubisCO from the marine cyanobacterium *Prochlorococcus marinus* MIT9313. *Limnology and Oceanography* (2007);**52**:2199-2204.

Segonzac M Les peuplements associés à l'hydrothermalisme océanique du Snake Pit (dorsale médio-atlantique; 23 N, 3480 m): composition et microdistribution de la mégafaune. *Comptes rendus de l'Académie des sciences. Série 3, Sciences de la vie* (1992);**314**:593-600.

Segonzac M, Desaintlaurent M & Casanova B Enigma of the trophic adaptation of the shrimp Alvinocarididae in hydrothermal areas along the Mid-Atlantic Ridge. *Cah Biol Mar* (1993);**34**:535-571.

Selosse MA *La symbiose: structures et fonctions, rôle écologique et évolutif*. Vuibert, (2000).

Shiba H, Kawasumi T, Igarashi Y, Kodama T & Minoda Y The CO₂ assimilation via the reductive tricarboxylic acid cycle in an obligately autotrophic, aerobic hydrogen-oxidizing bacterium, *Hydrogenobacter thermophilus*. *Archives of Microbiology* (1985);**141**:198-203.

Shigematsu T, Hanada S, Eguchi M, Kamagata Y, Kanagawa T & Kurane R Soluble methane monooxygenase gene clusters from trichloroethylene-degrading *Methylomonas* sp. strains and detection of methanotrophs during in situ Bioremediation. *Applied and Environmental Microbiology* (1999);**65**:5198-5206.

Shillito B, Hamel G, Duchi C, Cottin D, Sarrazin J, Sarradin P-M, Ravaux J & Gaill F Live capture of megafauna from 2300m depth, using a newly designed pressurized recovery device. *Deep Sea Research Part I: Oceanographic Research Papers* (2008);**55**:881-889.

Shively JM, van Keulen G & Meijer WG Something from almost nothing: carbon dioxide fixation in chemoautotrophs. *Annual Review of Microbiology* (1999);**52**:191-230.

Shoemaker DD, Machado CA, Molbo D, Werren JH, Windsor DM & Herre EA The distribution of *Wolbachia* in fig wasps: correlations with host phylogeny, ecology and population structure. *Proceedings of the Royal Society B: Biological Sciences* (2002);**269**:2257-2267.

Sievert SM, Scott KM, Klotz MG, *et al.* Genome of the epsilonproteobacterial chemolithoautotroph *Sulfurimonas denitrificans*. *Applied and Environmental Microbiology* (2008);**74**:1145-1156.

Smith DC & Douglas AE *The biology of symbiosis*. Edward Arnold (Publishers) Ltd., (1987).

Stakes DS, Shervais JW & Hopson CA The volcanic-tectonic cycle of the FAMOUS and AMAR Valleys, Mid-Atlantic Ridge (36° 47' N): Evidence from basalt glass and phenocryst compositional variations for a Steady State MagmaChamber Beneath the Valley Midsections, AMAR 3 *Journal of Geophysical Research: Solid Earth* (1984);**89**:6995-7028.

Sykes LR Mechanism of earthquakes and nature of faulting on the mid-oceanic ridges. *Journal of Geophysical Research* (1967);**72**:2132-2153.

Tabita FR, Satagopan S, Hanson TE, Kreel NE & Scott SS Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *Journal of Experimental Botany* (2008);**59**:1515-1524.

Takai K, Nakagawa S, Sako Y & Horikoshi K *Balnarium lithotrophicum* gen. nov., sp. nov., a novel thermophilic, strictly anaerobic, hydrogen-oxidizing chemolithoautotroph isolated from a black smoker chimney in the Suiyo Seamount hydrothermal system. *International Journal of Systematic and Evolutionary Microbiology* (2003);**53**:1947-1954.

Teixeira S, Serrao EA & Arnaud-Haond S Panmixia in a fragmented and unstable environment: the hydrothermal shrimp *Rimicaris exoculata* disperses extensively along the Mid-Atlantic Ridge. *PLoS One* (2012);**7**:e38521.

Teixeira S, Cambon-Bonavita MA, Serrão EA, Desbruyeres D & Arnaud-Haond S Recent population expansion and connectivity in the hydrothermal shrimp *Rimicaris exoculata* along the Mid-Atlantic Ridge. *J Biogeogr* (2011);**38**:564-574.

Thompson G, Humphris SE, Schroeder B, Sulanowska M & Rona PA Active vents and massive sulfides at 26 degrees N (TAG) and 23 degrees N (Snakepit) on the Mid-Atlantic Ridge. *The Canadian Mineralogist* (1988);**26**:697-711.

Tivey MK Generation of seafloor hydrothermal vent fluids and associated mineral deposits. *Oceanography* (2007);**20**:50.

Tivey MK, Humphris SE, Thompson G, Hannington MD & Rona PA Deducing patterns of fluid flow and mixing within the TAG active hydrothermal mound using mineralogical and geochemical data. *Journal of Geophysical Research: Solid Earth* (1995);**100**:12527-12555.

Tunncliffe V, Juniper SK & Sibuet M Reducing environments of the deep-sea floor. *Ecosystems of the World* (2003);81-110.

Tyler PA & Young CM Reproduction and dispersal at vents and cold seeps. *Journal of the Marine Biological Association of the United Kingdom* (1999);**79**:193-208.

Valdés J, Quatrini R, Hallberg K, Dopson M, Valenzuela PDT & Holmes DS Draft genome sequence of the extremely acidophilic bacterium *Acidithiobacillus caldus* ATCC 51756 reveals metabolic versatility in the genus *Acidithiobacillus*. *Journal of Bacteriology* (2009);**191**:5877-5878.

Valdés JH, Pedroso I, Quatrini R, Hallberg KB, Valenzuela PDT & Holmes DS Insights into the metabolism and ecophysiology of three *Acidithiobacilli* by comparative genome analysis. *Advanced Materials Research* (2007);**20**:439-442.

Van Dover C *The Ecology of Deep-sea Hydrothermal Vents*. Princeton University Press, (2000).

Van Dover CL Ecology of mid-Atlantic ridge hydrothermal vents. *Geological Society, London, Special Publications* (1995);**87**:257-294.

Van Dover CL, Szuts EZ, Chamberlain SC & Cann JR A novel eye in 'eyeless' shrimp from hydrothermal vents of the Mid-Atlantic Ridge. *Nature* (1989);**337**:458-460.

Van Dover CL, Fry B, Grassle JF, Humphris S & Rona PA Feeding biology of the shrimp *Rimicaris exoculata* at hydrothermal vents on the Mid-Atlantic Ridge. *Marine Biology* (1988);**98**:209-216.

Vereshchaka AL, Kulagin DN & Lunina AA Phylogeny and new classification of hydrothermal vent and seep shrimps of the family Alvinocarididae (Decapoda). *PloS one* (2015);**10**:e0129975.

Vetriani C, Speck MD, Ellor SV, Lutz RA & Starovoytov V *Thermovibrio ammonificans* sp. nov., a thermophilic, chemolithotrophic, nitrate-ammonifying bacterium from deep-sea hydrothermal vents. *International Journal of Systematic and Evolutionary Microbiology* (2004);**54**:175-181.

Watabe H & Hashimoto J A new species of the genus *Rimicaris* (Alvinocarididae: Caridea: Decapoda) from the active hydrothermal vent field, "Kairei Field," on the Central Indian Ridge, the Indian Ocean. *Zool Sci* (2002);**19**:1167-1174.

Watling L, Guinotte J, Clark MR & Smith CR A proposed biogeography of the deep ocean floor. *Progress in Oceanography* (2013);**111**:91-112.

Watsuji T-O, Nakagawa S, Tsuchida S, Toki T, Hirota A, Tsunogai U & Takai K Diversity and Function of Epibiotic Microbial Communities on the Galatheid Crab, *Shinkaia crosnieri*. *Microbes and Environments* (2010);**25**:288-294.

Wernegreen JJ, Degnan PH, Lazarus AB, Palacios C & Bordenstein SR Genome evolution in an insect cell: distinct features of an ant-bacterial partnership. *The Biological Bulletin* (2003);**204**:221-231.

Williams AB & Rona PA Two new caridean shrimps (Bresiliidae) from a hydrothermal field on the Mid-Atlantic Ridge. *Journal of Crustacean Biology* (1986);**6**:446-462.

Won Y, Young CR & Lutz RA Dispersal barriers and isolation among deep-sea mussel populations (Mytilidae: *Bathymodiolus*) from eastern Pacific hydrothermal vents. *Molecular Ecology* (2003);**12**:169-184.

Yarzabal A, Brasseur G & Bonnefoy V Cytochromes c of *Acidithiobacillus ferrooxidans*. *FEMS Microbiology Letters* (2002);**209**:189-195.

Yarzabal A, Appia-Ayme C, Ratouchniak J & Bonnefoy V Regulation of the expression of the *Acidithiobacillus ferrooxidans* rus operon encoding two cytochromes c, a cytochrome oxidase and rusticyanin. *Microbiology* (2004);**150**:2113-2123.

Zachary A & Colwell RR Gut-associated microflora of *Limnoria tripunctata* in marine creosote-treated wood pilings. *Nature* (1979);**282**:716-717.

Zachary A, Parrish KK & Bultman JD Possible role of marine bacteria in providing the creosote-resistance of *Limnoria tripunctata*. *Marine Biology* (1983);**75**:1-8.

Zbinden M & Cambon-Bonavita M-A Occurrence of *Deferribacterales* and *Entomoplasmatales* in the deep-sea Alvinocarid shrimp *Rimicaris exoculata* gut. *FEMS microbiology ecology* (2003);**46**:23-30.

Zbinden M, Le Bris N, Gaill F & Compère P Distribution of bacteria and associated minerals in the gill chamber of the vent shrimp *Rimicaris exoculata* and related biogeochemical processes. *Mar Ecol Prog Ser* (2004);**284**:237-251.

Zbinden M, Shillito B, Le Bris N, de Montlaur CDV, Roussel E, Guyot F, Gaill F & Cambon-Bonavita M-A New insights on the metabolic diversity among the epibiotic microbial community of the hydrothermal shrimp *Rimicaris exoculata*. *J Exp Mar Biol Ecol* (2008);**359**:131-140.

Annexes

Bioinformatique associée au barcoding/metabarcoding

ATTENTION : A partir de ce paragraphe (jusqu'au premier article), les thématiques abordées sont l'informatique et la bioinformatique. Ces domaines d'applications pouvant vite devenir difficiles à comprendre ou à lire, la rédaction, le vocabulaire, ainsi que les tournures de phrases sont volontairement plus « légères », de manière à désacraliser cette boîte noire qu'est la bioinformatique, dans le cas d'une application au barcoding/métabarcoding. Les laboratoires que sont le LM2E (Ifremer/UBO) et AMEX (UPMC) n'ayant pas pour domaine de compétence particulier l'informatique et la bioinformatique, cet aspect de ma thèse a donc été particulièrement intéressant, même passionnant, et m'a donné l'idée de cette partie de manuscrit dans le but d'aider a minima mes collègues ou futurs arrivants des laboratoires que sont le LM2E et AMEX.

L'informatique est une technologie aujourd'hui présente dans tous les domaines de la vie courante, et les sciences en font partie. La bioinformatique, quant à elle, permet l'analyse de données issues du vivant, et est utilisée dans diverses disciplines. Son utilisation, bien que très intéressante, reste plutôt difficile d'accès pour un néophyte. Comme n'importe quel dosage, révélation, ou autre manipulation réalisable à la pailasse, l'utilisation de la bioinformatique nécessite un certain nombre de compétences. Les outils d'analyses bioinformatiques sont nombreux, souvent libres de droits, et gratuits, il n'en reste pas moins que la plupart sont utilisables uniquement en ligne de commandes, sous différents langages de programmation (python, perl, java...), et sous différents systèmes d'exploitation tels que Windows, Mac ou encore Linux. Cependant, l'utilisation de ce type d'outils nécessite bien souvent plusieurs heures, voire jours. Pour être étudiés et testés, il faut les choisir et évaluer leur pertinence... Toutes ces informations sont très chronophages à analyser, et réparties un petit peu partout sur le net. Et si on gagnait du temps avec ce chapitre ? Où trouver de l'information, les outils, comment les utiliser... Ce chapitre ne sera bien évidemment pas exhaustif, mais il a pour but d'être simple, didactique, et alimenté de diverses sources (notamment scientifiques), afin de permettre à n'importe quelle personne, du stagiaire au chercheur, de pouvoir se lancer dans la bioinformatique, et découvrir un nouveau monde. Aucun nouvel algorithme (d'alignement, de clustering...) ne sera donc créé ici, mais des scripts d'analyses, permettant l'utilisation d'une sélection d'outils, et l'analyse complète de jeux de données (workflow) de barcoding et metabarcoding d'origine microbienne. De cette dernière partie de phrase, il faut comprendre qu'une toute petite partie du potentiel de la bioinformatique sera développée ici.

Trois buts principaux concernant ce chapitre sont à retenir : 1) faire office de « tutoriel » pour toute personne lisant ce manuscrit (du stagiaire au chercheur) souhaitant essayer et/ou utiliser la

bioinformatique pour traiter ses données, 2) être vulgarisé pour qu'aucune formation particulière ne soit prérequis (à part savoir utiliser un clavier, une souris, internet, et un éditeur de texte), 3) le tout appuyé par des sources scientifiques et/ou communautaires.

Tout lecteur ne sera pas concerné de la même manière par toutes les parties de ce chapitre.

Le contexte général

L'informatique au sens large occupe une part non négligeable dans bien des domaines, que ce soit politique, économique, social, scientifique... Bien évidemment, c'est l'aspect scientifique qui nous intéresse ici. Que ce soit pour écrire des projets, gérer son emploi du temps, rédiger un article, modéliser les courants océaniques, ou reconstruire le génome de votre animal, végétal ou microorganisme préféré, il vous faut un ordinateur. Mais de la même manière qu'il est difficile de faire du béton avec une fourchette, ou peu utile de couper un concombre avec une tronçonneuse, il existe différents types d'ordinateurs avec des utilités qui leurs sont propres.

Comment choisir sa machine ?

Pour bien comprendre comment choisir l'ordinateur adéquat pour faire de la bioinformatique, il faut faire un tour sous le clavier de votre ordinateur portable, ou mettre la tête à l'intérieur de votre tour.

On distinguera ici trois pièces principales pour faire les bons choix (il en existe bien d'autres).

- Le processeur : c'est lui qui est en charge de faire "les calculs", de traiter l'information, et de la répartir dans le reste de l'ordinateur, où il faut, quand il faut, mais il ne mémorise pas grand-chose.
- La carte de mémoire vive (RAM) : c'est elle qui est en charge de garder l'information à court terme pour le processeur. Pour faire un peu plus compliqué, il faut savoir que l'ordinateur comporte plusieurs mémoires différentes : 1) la mémoire cache. Elle est présente dans le processeur, elle est très petite (moins de 10 Mo le plus souvent), mais très rapide, et très temporaire. 2) la mémoire vive, moins rapide que la mémoire cache, mais plus grande et tout aussi temporaire (qui va de plusieurs centaines de Mo, à plusieurs Go). 3) La mémoire de masse. C'est le disque dur traditionnel, sur lequel sont stockées à long terme les données, comme vos photos, vidéos, documents, logiciels... Cette mémoire est encore moins rapide que les précédentes, mais bien plus importante en termes de capacité de stockage (de plusieurs Go à plusieurs To).

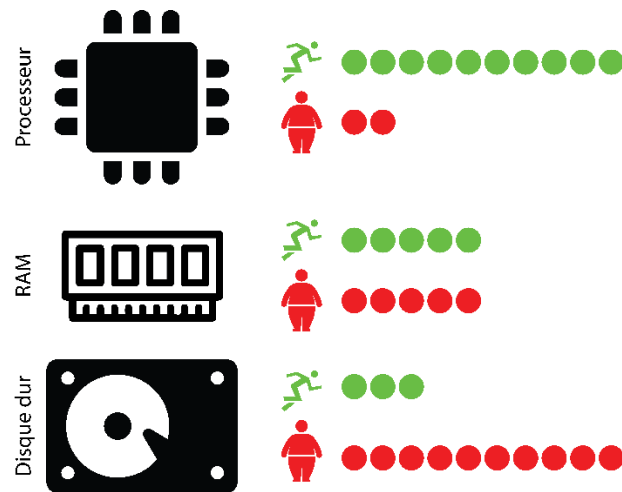


Figure 36 : Comparaison schématique de la rapidité et du stockage associé aux différents types de mémoire d'un ordinateur

- La carte graphique. C'est une forme de processeur dont l'architecture est prévue pour traiter de l'information graphique, comme des vidéos, des modèles 3D, des images complexes (de plusieurs Go)... Elle permet de lire photos et vidéos, films ou graphiques de travail et disposent de capacités plus ou moins importantes selon le modèle choisi. Un processeur normal peut effectuer les calculs d'une carte graphique, il sera cependant bien plus long, chauffera d'avantage... C'est pour cette raison que certaines applications nécessitent une carte graphique, les processeurs traditionnels sont trop long à répondre.

Dans le monde des sciences, il existe quatre grandes classes d'ordinateurs que l'on peut associer à ces trois pièces différentes de l'ordinateur. Pour faire de la bureautique, un ordinateur tout ce qu'il y a de plus basique est suffisant. Dès qu'il s'agit de faire autre chose, il faut adapter l'ordinateur à l'usage qui en sera fait. Pour analyser ou générer des modèles graphiques, il faut à votre ordinateur une carte graphique (qui n'est pas présente sur les ordinateurs de bureautique). Plus le type d'analyse sera poussé, plus la/les cartes graphiques devront être puissantes. Dans le cadre de modélisation mathématique, il faut davantage de calcul brut, donc un/des processeurs de bonne(s) qualité(s).

La dernière classe, qui nous intéresse ici, est une classe d'ordinateur gourmande en mémoire vive, notamment pour faire de la bioinformatique.

Pour être plus précis, certaines analyses, ou manipulations des données sont bien trop volumineuses pour être stockées dans la mémoire cache, et inutile à stocker sur la mémoire de masse. Il est impératif que ces informations soient stockées dans la mémoire vive. Cependant, le calcul peut être fait avec n'importe quel type de processeur, même celui de votre smartphone par exemple, le temps de calcul sera simplement plus long, voire beaucoup beaucoup plus long....

A ce stade, la question qui doit brûler vos lèvres est finalement : que dois-je choisir comme ordinateur pour faire de la bioinformatique à mon bureau ? Quel processeur utiliser, quelle quantité de mémoire vive est nécessaire ? Ai-je besoin d'une carte graphique ?

Pour faire de la bioinformatique, aucune ressource graphique n'est nécessaire, pas besoin de faire de 3D, de traitement d'image (comme des images satellites), donc pas besoin de carte graphique. Avec le recul, les besoins en mémoire vive minimum sont de 8Go, idéalement 16Go (ceci est un avis personnel). Il va de soi que choisir 8Go RAM sera plus limitant en termes d'outils utilisables qu'avec 16Go de RAM. Pour donner un exemple qui sera développé plus tard dans ce chapitre, 8Go de RAM vous permettra d'utiliser l'outil FROGS, mais pas Mothur. Alors que les 16Go vous permettront d'utiliser les deux. Pour finir, le processeur. Le leader du marché reste la marque Intel, le plus souvent retrouvé dans nos machines. En termes de qualité, l'idéal sont les processeurs de type i : les i3, les i5 et les i7. Les processeurs de type i3 feront de votre ordinateur de bureau une machine fiable par rapport à vos attentes. Les processeurs de type i5 sont destinés aux calculs importants. Et les

processeurs type i7 sont destinés aux calculs et présentent une architecture différente des i5 leur permettant de prendre en charge une partie des calculs effectués par une carte graphique (d'où un prix plus élevé). L'idéal pour la bioinformatique est un processeur de type i5. Il faut savoir que même si les processeurs i7 sont considérés comme meilleurs, il existe certains processeurs de type i5 plus performants d'un point de vue calcul brut que leur équivalent i7.

Vous avez maintenant toutes les cartes pour pouvoir choisir astucieusement votre matériel de travail quotidien.

Quel système d'exploitation choisir ?

Une fois votre machine en vue ou achetée, se pose à un moment la question du système d'exploitation idéal pour faire de la bioinformatique : Windows, Mac, "Linux" ... Il faut tout d'abord savoir que le nombre de systèmes d'exploitation à travers le monde ne se limite pas à trois. Il en existe des dizaines. Les systèmes Windows, Mac ou encore ce que l'on appelle couramment "Linux" sont les principaux utilisés. Sur le marché des ordinateurs portables et de bureau, Windows représente environ 90% du marché, Mac moins de 10% et Linux moins de 5%. Ces nombres restent imprécis en raison des différentes sources consultées, il n'en reste pas moins que la majorité des parts de marché sont incontestablement détenues par Windows. Pour autant, Windows n'est pas le système le plus adéquat pour faire de la bioinformatique. Contre toute attente ce sont les systèmes basés sur le noyau UNIX qui sont privilégiés par les bioinformaticiens. Pour mieux comprendre tout ceci, il est nécessaire de faire un peu d'histoire.

Pour cela remontons, en 1984, lorsque Microsoft sort son premier système d'exploitation MS-DOS (ancêtre de Windows, mais complètement différent de celui-ci). Ce dernier est commercialisé, mais loin d'être abouti, surtout que d'autres systèmes d'exploitation existent, notamment le système UNIX, le plus connu de ses concurrents, qui est bien plus abouti, plus puissant, mais difficile à utiliser, d'où un manque d'intérêt de la part du grand public. Pourtant à cette époque, les deux systèmes ont la même ergonomie, une console noire avec du texte blanc. La même année, le projet GNU fut initié par Richard Stallman, dans le but de créer une "copie" (code source différent, mais les noms de commandes restent les mêmes) d'UNIX, et de le distribuer gratuitement. Car à cette époque, UNIX est aussi payant, et cher, et ne peut être acheté que par des grosses sociétés ou des universités. En plus de fournir un système d'exploitation gratuit, le projet GNU vise à créer un système libre de droit (souvent aussi appelé "Open Source"). Un système (ou logiciel) libre de droit est un système dont on peut avoir le code source. A l'inverse des systèmes propriétaires, comme Windows, dont vous n'aurez jamais la recette. Pour faire une analogie simple, Windows est comme Coca-Cola, jamais vous n'aurez la recette. Par conséquent un logiciel libre de droit est le plus souvent gratuit (pas toujours, mais cela

reste rare), il est donc possible de le copier, le modifier, et le redistribuer à votre convenance, sans aucun souci.

GNU prône ainsi une philosophie informatique nouvelle visant à donner le code source au plus grand nombre, et laisser la communauté s’y intéressant améliorer les choses.

C’est à ce moment-là qu’une nouvelle personne entre en jeu, Linus Torvalds, qui crée en 1991 son propre système d’exploitation. Quel rapport avec le projet GNU ? Il s’avère en fait que les deux projets étaient complémentaires. Le projet GNU visait à créer ce que l’on appelle des programmes de base (comme le copier/coller, la suppression de fichier, l’éditeur de texte...), tandis que Linus Torvald créait le “cœur” du système d’exploitation, que l’on appelle le noyau. La collaboration de ces deux projets finit par donner le système d’exploitation GNU/Linux (Linux étant la contraction de Linus et UNIX), qui, par abus de langage, deviendra ce que l’on appelle aujourd’hui : Linux.

Et Mac OS dans tout ça ? C’est un système d’exploitation avec une interface graphique qui est développé et commercialisé depuis 1984, dont le noyau est directement basé sur UNIX.

Alors pourquoi faire autant d’histoire pour Linux, et si peu pour Mac OS et Windows. Tout simplement parce que le système Windows est le plus commercialisé (et que le lecteur que vous êtes l’a sûrement déjà utilisé), et que le système Mac OS est le plus onéreux à l’achat à ce jour (donc “peu” abordable). L’utilisation de Linux présente donc un certain nombre d’avantages : il est gratuit, utilisé pour la majorité des outils de bioinformatiques développés à ce jour, et qu’il peut très simplement être installé sur un disque dur externe, et que la base UNIX est utilisée sur la majorité des calculateurs (décrit plus tard). Il sera donc plus facile de communiquer avec une machine dont vous connaissez ou comprenez à priori peu de choses.

NB : Depuis Windows 10, Linux (plus précisément Ubuntu) a été directement incorporé au système. Il vous est donc possible d’utiliser directement Linux sur votre ordinateur Windows 10, tout en pouvant continuer à travailler sur vos logiciels Windows préférés. Pour les plus connaisseurs, la différence avec les machines virtuelles est que l’incorporation directe de Linux à Windows 10 donne des résultats de performances bien meilleurs qu’une machine virtuelle, presque aussi rapide qu’une machine nativement démarrée sous Linux.

Pour récapituler, les systèmes Mac OS et Linux sont les descendants du système UNIX (ils sont donc “similaires” d’un point de vue ligne de commande), tandis que Windows est un système à part.

	Windows	Linux	Mac OS
Descendant de	MC-OS	UNIX	UNIX
Interface graphique	+++	+	+++
Disponibilité de logiciels interfacés	+++	+	+++
Disponibilité de logiciels en lignes de commandes	+	+++	+++
Utilisé en bioinformatique	+	+++	+++
Prix	++	0	+++

Figure 37 : Tableau comparant les avantages et inconvénients des systèmes Windows, Linux, et Mac OS pour faire de la bioinformatique.

Quelle distribution Linux utiliser ?

Dans la partie précédente, le système d'exploitation Linux a été décrit, et c'est le système choisi pour effectuer la bioinformatique de ce chapitre, en association avec Windows. Il reste tout de même un problème avec Linux qui peut être perturbant : le choix de la distribution. Même si ici le but premier de Linux est de pouvoir faire de la bioinformatique, tous les utilisateurs Linux ne font pas de bioinformatique. De plus, comme précisé précédemment, le système Linux est libre de droit, il peut donc être copié, MODIFIÉ, et REDISTRIBUÉ. Ce sont ces deux derniers termes qui sont importants, car certains utilisateurs, sociétés, communautés informatiques ont modifié le Linux originel pour l'enrichir, le rendre plus esthétique, plus simple d'utilisation... Ceci a donné lieu à de nombreuses distributions. Dans un terme plus commun, si vous achetez une voiture "ordinaire", il peut y avoir les options sport, familiales, finitions cuir, bois... C'est ici la même chose avec les distributions Linux, à vous de choisir votre "modèle". Pour simplifier encore les choses, le graphique suivant résume comment fonctionnent les distributions.

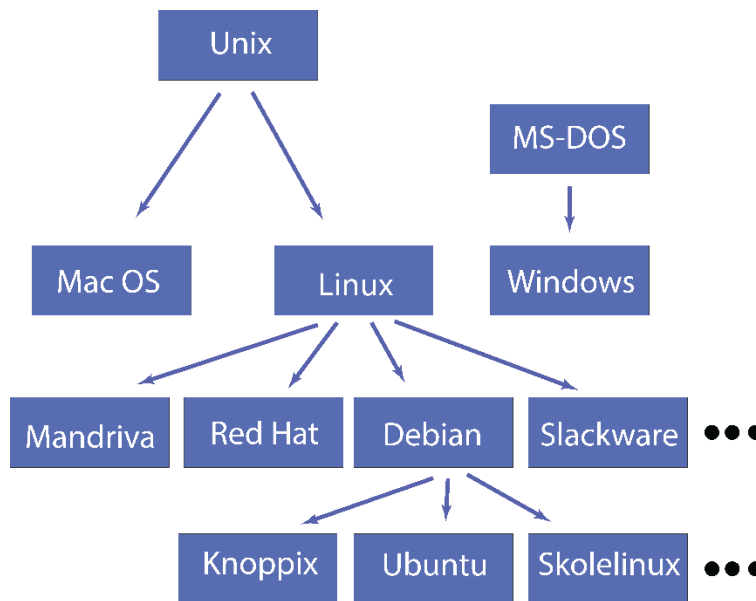


Figure 38 : Schéma de l'histoire de différents systèmes d'exploitation.

Le terme Ubuntu vous est sans doute familier, et c'est normal, car c'est la distribution de Linux la plus connue du grand public. Si vous souhaitez vous initier à l'environnement de travail Linux, la distribution Ubuntu est un choix intéressant. Le slogan de la distribution Ubuntu est d'ailleurs "Linux pour des êtres humains". De plus, les mises à jour sont fréquentes (tous les 6 mois environ), et facteur non négligeable, comme il y a beaucoup d'utilisateurs de par le monde, il y a aussi beaucoup d'aide et de documentation sur internet. Si vous vous posez une question, quelqu'un y a sûrement déjà pensé, et une réponse y a été apportée.

NB : Dans le cas d'utilisateurs Ifremer, la distribution conseillée est Fedora. Le réseau informatique, ainsi que le calculateur (développé ultérieurement) DATARMOR, sont configurés pour être plus simples d'utilisation avec cette distribution. Il n'en reste pas moins que pour toute aide recherchée sur internet, les réponses Ubuntu sont souvent adaptables sur Fedora. Après tout, toutes ces distributions sont basées sur le noyau Linux..., La base reste donc la même.

La bioinformatique en tant que telle

Il faut savoir que la bio-informatique est un domaine non pas pluridisciplinaire mais présentant plusieurs domaines de compétences spécifiques, tout comme la microbiologie est représentée par de la biologie moléculaire, de la culture ... Seulement, les différents aspects de la bio-informatique sont

représentés par des genres de bioinformaticiens ayant des qualifications qui leur sont propres. On peut en citer 7 (tiré du site bioinfo-fr.net) :

« Le scripteur

Nous commençons très fort en entrant dans un monde dangereux : le monde des trolls. En effet, qui dit scripteur dit langage de programmation de prédilection. Vous aurez droit à tout : le python, c'est le langage le plus puissant du monde, le perl, c'est juste bon à avoir mal à la tête, le ruby est sans contestation possible le langage du futur... Bref je ne vais pas tous vous les faire, on y passerait la journée (voire plus). Généralement donc, le scripteur porte un langage dans son cœur et s'y tient. Souvent incollable sur toutes les petites astuces de ce dernier, il affectionnera également de maîtriser ses différentes librairies exotiques pour sublimer ses résultats.

Le scripteur est utilisé bien souvent pour l'analyse de gros jeux de données biologiques. Il aimera s'aider des diverses banques de données biologiques mises à sa disposition sur le web ou en interne dans son laboratoire.

Le bash lover

Ceux-là n'ont besoin que d'une seule chose pour pouvoir travailler : un terminal. La puissance de ce dernier fait que, quand on maîtrise son langage, tout devient possible. Ces bioinformaticiens sont souvent assimilés à des "maîtres jedis" ou sont généralement les grands "gourous" des labos. C'est en effet toujours assez impressionnant pour nos amis biologistes de voir un homme "parler" à un ordinateur. Pour peu que le bash soit en fond noir avec le texte en vert (Matrix) et là vous aurez tout le respect de vos collègues...

Blague à part, ces bioinformaticiens sont souvent assimilés aux scripteurs. Mais n'allez surtout pas leur dire, car dans leur tête c'est bien deux mondes différents ! On ne mélange pas les torchons avec les serviettes ;-). Sans parti pris, je pense pouvoir dire que les deux se valent à niveau égal. L'un aimera partir de zéro, l'autre aimera avoir des petites fonctions magiques déjà toutes faites à appeler de temps en temps.

Le codeur fou

Ce genre-là est assez paradoxal. Il aurait très bien pu s'en tenir à l'informatique pure et dure, des 1 et des 0 avec un poil d'algorithmique. Mais ces gens-là aiment l'idée de pouvoir combiner les choses, et allier la biologie à l'informatique les a sans doute séduits. Quoique vous leur demanderez, bizarrement,

leur solution ne tendra jamais vers la facilité. Leur esprit complexe et tordu les amène souvent à réinventer la roue. Certaines rumeurs vont jusqu'à dire que ce sont des informaticiens qui cherchent l'amour, mais nous n'irons pas jusque-là. Ils aiment manier plusieurs langages à la fois, connaître leurs forces et leurs faiblesses et utiliser le bon langage au moment propice tout en justifiant leur choix grâce à une multitude d'arguments qui tiennent la route mais qui peuvent malheureusement bien souvent être assimilés à du troll.

Le plus souvent ce bioinformaticien est nécessaire pour construire des bases de données, implémenter des logiciels ou encore concevoir des interfaces web pour des biologistes.

L'élus (50% bio - 50% info)

Une espèce assez rare mais qui mérite d'être citée. Être bioinformaticien, c'est arriver à allier la biologie et l'informatique, mais la difficulté c'est de faire ça de manière équitable. En effet, il est rare de croiser un bioinformaticien une pipette à la main, et pourtant ça paraîtrait presque logique ! Ce genre-là est donc capable à la fois de réaliser ses expériences sur la paillasse et de les analyser ensuite grâce à l'outil informatique.

C'est la définition la plus parfaite de ce que peut être un bioinformaticien, mais c'est pourtant la moins répandue. Peut-être parce qu'à un moment donné il faut quand même choisir entre biologie ou informatique car les deux cases ne rentrent pas dans les papiers de l'administration ?

Ces bioinformaticiens sont donc rares, mais existent. D'un côté, on se surprend à les envier de par leur pluridisciplinarité quasi parfaite, et d'un autre non car on se dit qu'il n'est pas possible de s'impliquer réellement dans les deux à la fois vraiment à fond. Il doit rester un goût d'inachevé...ou pas.

Le pousse-boutons

Ils sont souvent comparés au "bas niveau" de la bioinformatique, quand bien même ils sont déjà acceptés dans la caste très noble du bioinformaticien. Leur fonction est en effet d'utiliser les scripts, logiciels ou bases de données existantes pour en tirer des informations ou des résultats. En gros et en exagéré, ce bioinformaticien là est une sorte de technicien à qui on donnera un jeu de données et qui maîtrisera un/plusieurs outils pour en extraire ce qui lui sera demandé. En théorie, il faut donc "juste" être capable de savoir lire un/plusieurs manuel et de savoir interpréter des données biologiques.

Ça peut être marrant/reposant un moment d'appartenir à ce genre de bioinformaticien là, mais n'y restez pas trop longtemps sans quoi vous risqueriez d'y perdre votre flamme.

Le matheux

Une espèce intéressante ! En effet, souvent lorsque qu'une équation apparaît dans une présentation, la plupart de l'auditoire (composé de biologistes et/ou de bioinformaticiens) prend peur ou bien tout

simplement décroche d'un seul coup d'un seul. Alors que notre matheux, lui, sera complètement fasciné par cette petite équation toute bête qui, pour lui, sera là simplement par évidence. C'est un fait : les maths sont aussi présentes dans le monde de la bioinformatique. On peut les côtoyer à petites doses (ou je pourrais plutôt dire les éviter, pour certains), ou bien alors chercher à utiliser leur force pour arriver à nos fins. Mais tout le monde n'a pas l'esprit des mathématiques. Ce qui explique sûrement que cette catégorie de personnes reste appréciée dans notre milieu, bien que souvent encore incomprises. On parle en effet d'une science théorique mise en relation avec une science appliquée. Pas toujours évident, mais notre matheux lui, il y croit dur comme fer et vous prouvera par " $a + b$ " que les maths sont l'avenir de la bioinformatique.

Le padawan

Et pour finir en beauté, j'ai choisi de vous parler de ceux qui trouvent la voie, la lumière, osons même jusqu'à dire la sagesse. Il n'est en effet plus très rare de voir des biologistes de formation faire les yeux doux à la bioinformatique qui les aide de plus en plus dans leurs travaux de tous les jours. Bien que la discipline se répande de plus en plus de nos jours, tous les laboratoires n'ont pas encore leur bioinformaticien, ni plateforme de bioinformatique. Il faut alors que les pauvres biologistes se débrouillent par leurs propres moyens. Les plus courageux s'aventureront à la découverte de notre magnifique discipline. La motivation sera leur seule alliée. Après de longues heures à lire les articles de bioinfo-fr.net, à faire et refaire les exercices de [rosalind](#) et à arpenter les divers tutoriels à droite et à gauche, généralement ils donneront des bioinformaticiens d'une efficacité spectaculaire.

À n'en pas douter, ils seront un jour le départ d'une nouvelle race de bioinformaticiens. »

D'après ces différentes descriptions, la suite de ce travail est davantage orientée vers un travail de padawan. En effet, les scripts qui seront accessibles à tous sous peu sur le site GitHub, sous mon patronyme vincentapremont, n'ont pas pour but de révolutionner la science, mais de simplifier des analyses qui peuvent apparaître complexes aux non utilisateurs courants. Alors pourquoi réaliser des scripts accessibles à tous ? Pour que mon travail perdure durant les quelques années à venir au sein de mon laboratoire d'accueil, pour tout type d'utilisateur, mais aussi car de nombreux tests ont été réalisés pour analyser les données selon différents paramètres. Réaliser des scripts automatisant une analyse complète est donc devenue nécessaire. Si l'aspect « programmation » des scripts n'est pas optimisé tel qu'un vrai informaticien aurait pu le faire, ils n'en restent pas moins efficaces et cohérents d'un point de vue biologique, à condition de les utiliser correctement.

Encadrement stagiaire

Encadrement de David François durant son stage de M1 (2mois) sur de la microscopie FISH (Fluorescent *In situ* hybridization) pour un sujet intitulé: Recherche d'une potentielle symbiose bactérienne chez *Rimicaris chacei*.

Formations suivies.

- Imagerie scientifique (MNHN)
- Phylogénie : support à l'analyse de l'évolution (MNHN)
- Workshop Métabarcoding (Ifremer)
- Metarbarschool (université de Grenoble)
- Anglais intensif (UBO)
- Aide à la publication en anglais (UBO)
- Fortran 95 (UBO)
- LaTeX (UBO)
- Prévention des risques radiologiques (Ifremer)
- Initiation à R (MOOC internet)

Formation donnée.

- L'enfer de l'imagerie scientifique (formation interne au laboratoire)

Participations grand public.

- Forum des étudiant Azimut 2016 et 2017.
- Fêtes de la science 2017.
- Nuit européenne des chercheurs 2016. (Production d'un poster double)

Participation à des colloques :

International :

- Congrès international Deep Sea Biology Symposium session 2015 (poster)

National :

- CONNECT 2015 (poster)
- CONNECT 2016 (oral)

Autres oraux hors colloques :

- Journée UMR BOREA 2014 et 2015
- Forum des doctorant LM2E 2015 et 2016
- Journée LEP 2016
- Journée de l'EDSM 2017

Autres :

- Création de la Newsletter du LM2E version web et participation à la rédaction de la Newsletter durant les années 2015 et 2016.

Preliminary data on the investigation of hydrothermal shrimp *Chorocaris chacei* symbiosis.

Vincent Appriou¹, Valérie Guezennec², Marie-Alice Gaudin³, Magali Chateau⁴

¹UMR 1102, IFREMER, Centre de Recherche Halieutique Atlantique, 17000 La Rochelle, France; ²UMR 1102, IFREMER, Centre de Recherche Halieutique Atlantique, 17000 La Rochelle, France; ³UMR 1102, IFREMER, Centre de Recherche Halieutique Atlantique, 17000 La Rochelle, France; ⁴UMR 1102, IFREMER, Centre de Recherche Halieutique Atlantique, 17000 La Rochelle, France

I Introduction

Chemosynthetic symbioses in marine environment are currently known in different organisms. Figure 1A (Dubilier *et al.* 2008). In deep-sea ecosystems, chemosynthetic endosymbioses were first described in the tube worm *Riftia pachytila*, Figure 1B (Felbeck, 1981). Other associations between animals and bacteria have since been described. For example, ectosymbioses of the annelida *Alvinella pompejana*, Figure 1C (Gall *et al.* 1988), and gill endosymbionts in the bivalvia, e.g. *Bythotrephes cederstroemi* and gastropoda, e.g. *Alviniconcha* sp. Figure 1D, E (Vidal *et al.* 2002 ; Beintart *et al.* 2012).

Another example is the Alvinocarididae shrimp *Rimicaris exoculata* Figure 1F, which harbours a complex ectosymbiosis in both its

II Sampling

Shrimps were sampled from 4 MAR hydrothermal vent sites (one shrimp per site) : Lucky Strike, Rainbow (BIOBAZ cruise), TAG, Snake Pit (BICOSE cruise) (Figure 2). We also analysed one post-moult shrimp, and two early juveniles from TAG. For each individual, observations on cephalothorax parts (Branchiostegite and Scaphognathite Figure 3) and gut (Midgut) were performed.

III First results

Chorocaris chacei

Rimicaris exoculata

Cephalothorax

We observed mineralisation zones under binocular Figure 4A and B. These areas correlated with the presence of bacteria, which was confirmed by Scanning Electronic Microscopy (SEM) Figure 4C and D. Five main morphotypes were identified, present on both Branchiostegites and Scaphognathites at each site. Two are bacilli, the remaining being filamentous bacteria. The mean size (diameter x length) of first bacillus type was about 0.22 x 2.12 µm (Orange), the second about 0.26 x 0.91 µm (Purple), whereas the mean size for the first filamentous bacterium was about 2.91 x 3.11 µm (Green), 0.71 x 0.52 µm for the second (Orange), and 0.76 x 5.47 µm for the third (Pink) (Figure 7E and F).

Midgut

We observed only one morphotype of bacteria. Two cases were identified : one appearing as a bacterial mat (Figure 9A), and a second with less abundant filaments, which permit to see the host microvilli (Figure 9C). This was confirmed with TEM observation Figure 9E. Average diameter was measured : 0.18 µm. All these data match with those obtained for *Rimicaris exoculata*.

Moreover, we identified methanotrophic like bacteria Figure 6A for Lucky Strike and Rainbow specimens, in high abundance. They were not observed on Snake Pit and TAG specimens. On SEM observations (Figure 6D), coccobacilli (colored in blue) have the same morphology and size (0.5 to 1.3 µm) as those identified on TEM (Figure 6A and B). This could be correlated to CH₄ measurements in vent fluid (mmol.kg⁻¹) : Lucky strike : 0.5-0.97 ; Rainbow : 2.5 ; TAG : 0.124-0.147 ; Snake Pit : 0.023-0.062 (Charlou, *et al.* 2002).

These results raise the question of the coexistence in the same habitat of two shrimp species with very close symbionts, morphologically and phylogenetically (data not shown)?

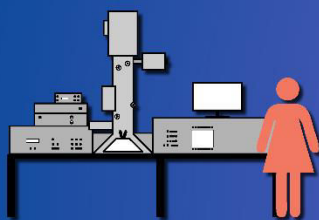
IV Perspectives

- Using X-Ray elemental analysis to compare minerals in *Chorocaris chacei* gill chamber with those in *Rimicaris exoculata* (Zbinden *et al.* 2004) ?
- Using correlative microscopy to associate molecular data with microscopic observations directly (Laming *et al.* 2015).

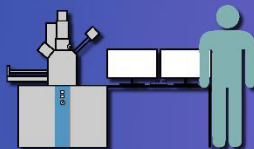


Les abysses : le petit

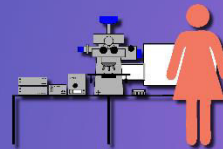
Comment les microorganismes des milieux extrêmes contribuent au bon fonctionnement de ces écosystèmes originaux et assurent l'intégrité de leur matériel génétique?



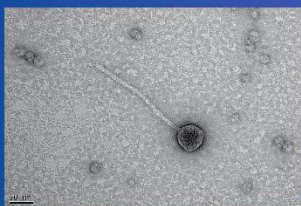
Microscope électronique à transmission



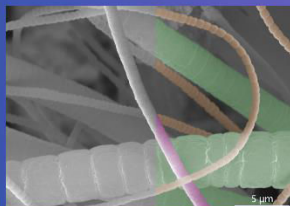
Microscope électronique à balayage



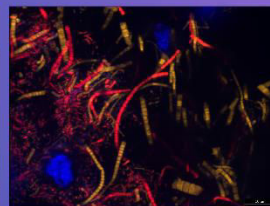
Microscope à épifluorescence



Virus / Bactéries / Archées / Cellules / Tissus



Bactéries / Archées / Organes / Animaux / Minéraux



Bactéries / Archées / Cellules / Tissus



◀ Fermenteur sous pression qui permet de cultiver les microorganismes dans des conditions de pressions identiques à leur milieu naturel.



◀ Flasks de culture pure. Elles permettent d'avoir de petites quantités de culture, pour de grandes répétitions de manipulations, ou pour ensemencer des fermenteurs plus gros.



◀ Flasks de culture de sédiments. Elles permettent d'avoir de petites quantités de culture, pour de grandes répétitions de manipulations, car les sédiments ne sont pas des substrats de culture à grande échelle.



◀ Enceinte à atmosphère modifiée. Elle permet de travailler dans une atmosphère contrôlée, le plus souvent pour obtenir un environnement anaérobie (sans oxygène).

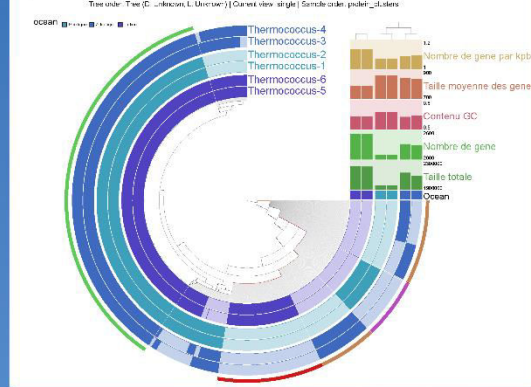
Fermenteur pour des microorganismes cultivables en condition de pression atmosphérique, pour des manipulations génétiques (OGM), ou bien des recherches de molécules actives présentant un intérêt biotechnologique.



monde des fausses idées

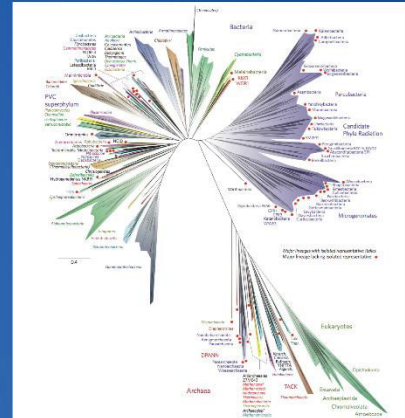


Répartition des groupes de protéines parmi 6 génomes de *Thermococcus*

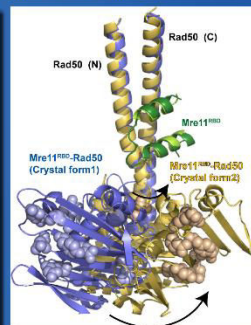
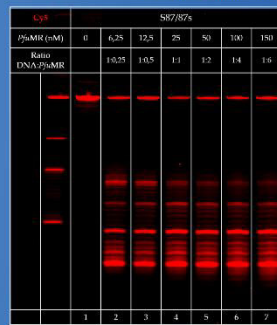


« Aujourd'hui, en plus de savoir reconstruire des génomes procaryotes, il nous est aussi possible, grâce à l'augmentation constante des capacités informatiques, de les comparer entre eux pour faire par exemple de l'écogénomique.

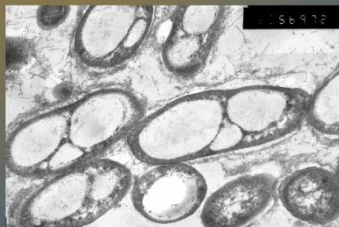
Arbre phylogénétique du monde vivant. Le monde Eukaryote (animaux, végétaux, protistes) ne représente pas une diversité si importante quand on la compare au monde Procaryote (Bactéries et Archées).



Les techniques de biochimie permettent de visualiser les différentes étapes de la réplication et de la réparation *in vitro* de l'ADN chez les Archées.



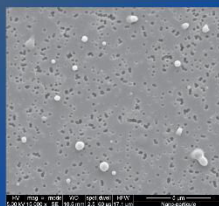
Boîte de clonage présentant un criblage blanc/bleu, qui permet de repérer les microorganismes génétiquement modifiés (en blanc) de ceux qui ne le sont pas (en bleu).



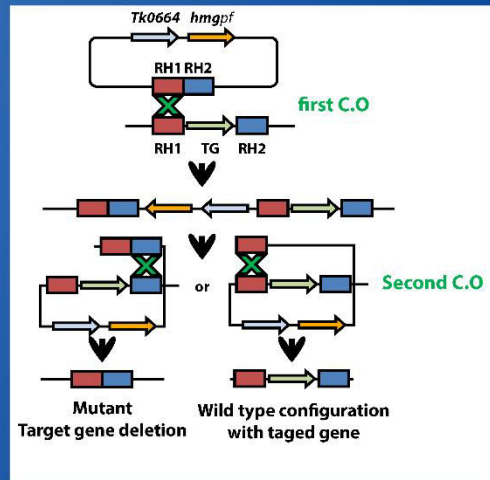
▲ Vacuoles (cercles blancs) de polymère intracellulaires.



▲ Bioplastiques produits par des bactéries.



▲ Nanoparticules de PolyHydroxyAlcanates (PHA).



▲ Représentation schématique permettant de créer des mutants (OGM) de *Thermococcus barophilus* afin de pouvoir comprendre les systèmes d'adaptation à la pression.

**Description de la diversité microbienne associée à *Chorocaris chacei* :
une possible double symbiose.**

Bien que très peu explorés, les fond des océans pourraient représenter le futur de l'industrie minière mondiale. L'impact écologique de ces activités reste encore inconnu, notamment sur les cibles privilégiées que sont les systèmes hydrothermaux et leurs écosystèmes. Ce travail s'intéresse à la crevette *Rimicaris chacei* et à sa potentielle symbiose avec des bactéries hydrothermales. *R. chacei* présente une alimentation mixte, nécrophagie et symbiose, plasticité alimentaire qui pourrait être un atout en cas de modification de son environnement. La proximité phylogénétique et écologique de *C. chacei* avec une autre crevette hydrothermale, *Rimicaris exoculata*, laisse supposer une possible histoire commune de leurs symbioses.

Deux compartiments biologiques des crevettes ont été étudiés au cours de ce travail, le céphalothorax et le tube digestif, via deux approches complémentaires: l'imagerie et la biologie moléculaire. Nous souhaitons répondre à deux questions principales (1) Décrire le type d'association entre la crevette et ses « symbiontes » d'un point de vue morpho-anatomique et phylogénétique. (2) Evaluer le degré de similarité entre les symbiontes de *C. chacei*, et *R. exoculata* via une approche de métabarcoding. Notre but est de déterminer si une possible variation de la distribution des symbiontes existe chez ces deux espèces en fonction du site hydrothermal d'origine (facteur environnemental), en fonction de l'espèce de crevette étudiée (facteur hôte), ou une association de ces deux facteurs. La bioinformatique ayant une part importante dans l'analyse des données de barcoding/metabarcoding, une partie du manuscrit lui est dédiée pour « désacraliser » ce type d'analyse.

Mots clés : *Rimicaris (Chorocaris) chacei* ; *Rimicaris exoculata* ; hydrothermal ; symbiose, métabarcoding.

Description of the microbial diversity associated with *Chorocaris chacei*:

A possible double symbiosis.

Deep-sea ocean may soon represent deep-sea mining industry future. However, these environments are still poorly explored. Therefore, mining issues on deep-sea ecosystems are not yet evaluated, mostly on their main target: the deep-sea hydrothermal vents. Our study focused on a shrimp, *Rimicaris chacei* and to its potential symbioses with chemoautotrophic microbial communities. This shrimp have a mixotrophic behaviour, mixed between necrophagy and symbiosis. It could therefore have a potential trophic plasticity in case of steep environmental modifications. Moreover, *C. chacei* is closely related to *Rimicaris exoculata*, in terms of both phylogeny and ecology. This could let suppose a common symbiosis history, presenting nowadays two different levels of association.

Two potential symbiotic microbial communities have been studied here, one located in the cephalothorax and the other in the digestive tract, using two complementary approaches: microscopy and molecular analyses. Two main points have been focused in our work: (1) Describing the shrimp and its associated microbial communities in terms of morphology, repartition and phylogeny. (2) Using a metabarcoding approach to evaluate the similarity level shared between *C. chacei* and *R. exoculata* associated microbial communities. We intend to analyse a possible genetic variation among symbionts of the two hosts, whether it would be linked to the hydrothermal vent origin (geography), or to the studied shrimp (host), or both of them. As bioinformatics was an important part of my work to analyse barcoding/metabarcoding data, a part of my thesis is dedicated to explain these analyses as a tutorial for all future users.

Keywords : *Rimicaris (Chorocaris) chacei* ; *Rimicaris exoculata* ; hydrothermal ; symbioses, metabarcoding.