



HAL
open science

Origines et évolution de lignées hydrothermales

Perrine Mandon

► **To cite this version:**

Perrine Mandon. Origines et évolution de lignées hydrothermales. Biologie du développement. Sorbonne Université, 2018. Français. NNT : 2018SORUS467 . tel-02924873

HAL Id: tel-02924873

<https://theses.hal.science/tel-02924873v1>

Submitted on 28 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.

Sorbonne Université

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

UMR 7205 Institut de Systématique, Evolution, Biodiversité / Exploration, Espèces et Evolution

UMR 7144 Adaptation et Diversité en Milieu Marin / Adaptation et Biologie des Invertébrés en Conditions Extrêmes

Origines et évolution de lignées hydrothermales

Par Perrine Mandon

Thèse de doctorat de Biologie évolutive, Phylogénie

Dirigée par Sarah Samadi et Stéphane Hourdez

Présentée et soutenue publiquement le 18 septembre 2018

Devant un jury composé de :

M. Aurelle Didier, maître de conférence à l'Université d'Aix-Marseille (Rapporteur)
M. Chen Wei-Jen, professeur à l'Université Nationale de Taiwan (Rapporteur)
M. Higuet Dominique, professeur à Sorbonne Université (Examinateur)
Mme. Matabos Marjolaine, chercheuse à l'IFREMER (Examinatrice)
M. Hourdez Stéphane, chargé de recherche CNRS (co-directeur de thèse)
Mme. Samadi Sarah, professeure au MNHN (directrice de thèse)

A Bergamotte, qui me suis dès le premier jour et me suivra toujours

Remerciements

Parce que cette thèse est un travail que je signe, mais qui en réalité a pu voir le jour, grandir puis finalement s'achever aujourd'hui grâce à la collaboration d'énormément de monde... Cette page est attribuée à l'ensemble de ces personnes.

Pour m'avoir permis de travailler sur ce sujet pendant presque 4 ans, pour m'avoir accompagnée parfois plus que de raison, pour m'avoir secoué les puces ou pris du temps pour me rassurer, chacun à votre façon, pour avoir été là jusqu'au bout et sans défaillir ; pour l'ensemble des efforts que je vous sais avoir fait, Sarah, Stéphane, je vous offre toute ma reconnaissance et ma gratitude. Stéphane plus personnellement pour ton soutien moral, ton tempérament joyeux (tant que ta paillasse de manip est rangée), ta disponibilité dont j'ai ô combien abusé... Sarah, plus personnellement pour ton honnêteté, ta rigueur scientifique, tes remarques constructives et la patience dont tu as fait preuve.

Cette thèse a eu lieu au sein des UMR 7144 Adaptation et Diversité en Milieu Marin et UMR 7205 Institut de Systématique, Ecologie et Biodiversité et ces lignes s'adressent également aux directeurs Philippe Grandcolas et François Lallier.

Aux membres de mon comité de thèse, qui ont accepté les discussions en vidéo conférences, parfois à des heures tardives : Eric Panté, Frédérique Viard, Didier Jollivet. A vous, qui avez été présents lors de ces comités de pilotage et ponctuellement à d'autres moments, en cas de galères de manip, de questionnement durant l'analyse de résultats... A vous qui m'avez suivie et prodigué de sincères encouragements.

Aux membres de mon jury : à Didier Aurelle, Wei-Jen Chen, Dominique Higuët et Marjolaine Matabos pour avoir accepté de relire et d'examiner mon travail, et de vous déplacer pour venir m'écouter.

Egalement, les membres de mes deux équipes, à Roscoff et à Paris, qui m'ont accueillie pendant la durée de ma thèse et m'ont permis d'effectuer ces travaux, de discuter, de souffler, de partager. A Nicolas plus particulièrement pour son coaching rédaction !

A Erwan Corre pour ton accompagnement, ta bienveillance et ta bonne humeur. A toi, qui as eu le courage et la gentillesse de te plonger et de parler passionnément de données techniques et de métriques d'assemblages. A l'ensemble des membres du support ABiMS qui ont supporté mes « erreurs aléatoires » sans rechigner...

A la chaleureuse triade du fond du couloir, Jawad, Régiset Julio. A Jawad, qui n'aime pas mes questions Mais les écoute quand même. A Régis, vif d'or perpétuel, pour ton dynamisme, ton bon entrain, tes conseils avisés que ce soit en manip ou pour des analyses, ou pour tenir debout sur un Mono-roue. A Julio, pour les initiations pseudo-code, les débogages (qui se résolvent souvent avec la même question : *have you tried to turn it off and on again ?!*), mais aussi pour les moments de pertitions où tous les sujets à peu près possibles et imaginables sont abordés.

A la branche féminine du staff : Carole, Céline, Delphine, pour les cafés du matin, les papotages de couloir, pour me rappeler le soir que *pour partir plus tôt là, c'est trop tard*. A Laetitia, pour nos pauses café qui font tourner la montre... A Dame Agnès, pour son soutien, en particulier durant les durs mois d'août, désertés, avec toujours des fraises tagada, un gâteau ou simplement un sourire.

Aux parents, amis, camarades, qui ont suivi cette thèse de plus loin, qui auraient je le sais, bien apprécié de me voir pointer le bout du nez un peu plus souvent ces dernières années. A mes parents qui ont été heureux pour moi, se sont inquiétés parfois, mais qui sont avant tout contents pour moi ! A Fluflu, qui est passé par là et qui, en toute connaissance de cause, posait quand même toujours la question de l'avancement, même en 3^{ème} et PIRE en 4^{ème} année ... Quand bien même, pour nos soirées Whisky - Agnès Obel. A Sophie, toujours là depuis toutes ces années ; pour tes délires, tes joyeuses virevoltes qui me permettent de voyager à ta manière.

A Noémi, pour nos moments et nos discussions phylogénomiques à minuit, de retour de soirée. A toi qui a eu de la patience et de l'attention d'un suricate aux aguets.

A Mathias, Anthony Chloé, Lala, Marie, qui m'ont accompagnée tout le long de cette expérience. Pour ces soirées bars à jeux ou karaoké. A vous qui m'avez épaulée, et ce même avant la thèse. A l'ensemble des autres thésards, avec qui nous sommes ou avons été dans le même bateau : Mathilde, Ninon, Lucie, Morgane, Laetitia, à la bande de FT dans sa plus parfaite entièresité et l'ambiance souvent pailletée qu'ils entraînent avec eux. A Bérèngère, à Claire, pour s'être bien tenu les coudes mutuellement !

A ceux avec qui j'ai partagé mes bureaux, CamTor et Bérénice à Roscoff, la carboglace comme substrat pour mille et une expérimentations. A Paul et Malcolm à la capitale, pour nos soirées Starmania à se demander en brailant, au dépend d'éventuels travailleurs tardifs du couloir, qu'est-ce qu'on pourra bien faire de nos vies demain.... A votre bonne humeur ; à vous qui avez aussi su savamment meubler mon étagère de bureau ! A Sophie, pour les moult discussions, débriefings, nos derniers instants de rédaction.

A toi, que cette thèse m'a permis de rencontrer, contre tellement de probabilité. Toi qui es là jusque dans les tous derniers instants, à me dépanner, m'encourager me soutenir. Qui m'accompagne.

A tous ces gens que j'ai rencontrés, entre ciel et mer : à bord de Belem ou les autres gabiers de l'Hermione à l'occasion d'aventures en mer ; au parachute, où le soleil brille toujours au-dessus des nuages. Et pour tous les autres qui, en ces tous derniers instants sont discrètement dans un coin de ma tête, trop discrets pour être couchés sur ce papier, mais qui savent ce qui leur revient ...

... A vous tous, je dis un très grand merci.

Table des matières :

| | |
|---|-----------|
| I. Introduction générale de la thèse | 15 |
| I.1 Contexte général : | 15 |
| I.1.1 Une histoire de reconsidérations taxonomiques : les vers de la famille Siboglinidae | 17 |
| I.1.2 Etude de la colonisation des sources hydrothermales : les moules de la sous-famille Bathymodiolinae | 19 |
| I.2 Stratégies d'étude de la thèse..... | 21 |
| | |
| II. La phylogénie par une approche multigène : le cas des vers de la famille Polynoidae | 27 |
| II.1 Introduction..... | 27 |
| II.1.1 Les Polynoidae dans la faune des sources hydrothermales | 27 |
| II.1.2 Diversité des Polynoidae | 31 |
| II.1.3 Etat des connaissances de la phylogénie des Polynoidae | 35 |
| II.1.4 But de l'étude | 41 |
| II.2 Matériel et méthodes..... | 43 |
| II.2.1 Sélection des taxons pour l'étude | 43 |
| II.2.2 Obtention et sélection des séquences moléculaire | 45 |
| II.2.3 Analyses phylogénétiques | 48 |
| II.3 Résultats - Discussion | 51 |
| II.3.1 Echantillonnage final | 51 |
| II.3.2 Analyses phylogénétiques | 53 |
| II.4 Conclusions et perspectives | 73 |
| | |
| III. Une approche NGS : le cas des crabes de la famille Bythograeidae | 80 |
| III.1 Introduction..... | 80 |
| III.2 Matériel et méthodes..... | 86 |
| III.2.1 Echantillons biologiques et extraction d'ADN..... | 86 |
| III.2.2 Séquençage du génome mitochondrial en PCR longues..... | 86 |
| III.2.3 Préparation des banques pour le séquençage shotgun | 89 |
| III.2.4 Cartographie des reads obtenus en shotgun | 89 |
| III.2.5 Analyses phylogénétiques | 91 |
| III.3 Résultats - Discussion | 93 |
| III.4 Conclusion et perspectives..... | 107 |

| | |
|---|------------|
| IV. Une approche NGS : le cas des crevettes de la famille Alvinocarididae..... | 111 |
| IV.1 Introduction..... | 111 |
| IV.1.1 Systématique des Alvinocarididae | 111 |
| IV.1.2 Position des Alvinocarididae au sein des Caridea | 113 |
| IV.1.3 Objectif de notre étude | 120 |
| IV.2 Matériel et méthodes..... | 122 |
| IV.2.1 Echantillonnage | 122 |
| IV.2.2 Méthodes | 123 |
| IV.2.3 Recherche et sélection de marqueurs phylogénétiques..... | 135 |
| IV.2.4 Orthologie des marqueurs sélectionnés | 139 |
| IV.2.5 Analyses phylogénétiques..... | 143 |
| IV.3 Résultats - Discussion | 147 |
| IV.3.1 Echantillonnage | 147 |
| IV.3.2 Assemblages des transcriptomes..... | 149 |
| IV.3.3 Recherche et caractéristiques des 425 marqueurs recherchés | 151 |
| IV.3.4 Evaluation de l'intérêt des marqueurs sélectionnés pour des études phylogénétiques | 158 |
| IV.3.5 Analyses phylogénétiques..... | 168 |
| IV.4 Conclusions et perspectives | 171 |
| | |
| V. Conclusions et perspectives générales de la thèse..... | 178 |
| VI. Bibliographie | 186 |
| VII. Annexes..... | 204 |

Lexique

Apomorphie : état dérivé d'un caractère. En opposition à la plésiomorphie, état primitif d'un caractère.

Assemblage :

- Avec référence : les séquences des reads sont alignées sur une référence (génome de l'espèce considérée ou très proche, ou un transcriptome déjà assemblé).
- *De novo* : les k-mer chevauchants sont fusionnés pour former des séquences plus longues : des contigs.

Contig : Séquence résultant de l'assemblage des reads de séquençage.

Convergence évolutive : réponse adaptative similaire et acquise indépendamment dans plusieurs lignées évolutives.

Couverture de séquençage : nombre de nucléotides séquencés au moins une fois.

Extrémités UTR : régions transcrites en ARNm mais non traduites en protéines. Elles jouent un rôle dans la régulation de la transcription et la traduction. L'extrémité 5' UTR s'étend du début du transcrit jusqu'au codon qui initie la traduction. L'extrémité 3' UTR se situe en aval du codon stop, jusqu'à la fin du transcrit.

FPKM : estimateur de la couverture de séquençage d'un contig, relativement à l'effort de séquençage. Ce chiffre correspond au nombre de Fragments mappés Par Kilobase effectives de transcrit par Million de reads d'effort de séquençage.

Hologénophore : spécimen ou échantillon, dont une partie est déposée comme voucher tandis qu'une autre partie a servi pour obtenir les séquences moléculaires.

Homologie : relation entre deux caractères similaires ayant divergé à partir d'un même ancêtre commun (voir encadré 6, p. 139).

Homoplasie : Relation entre deux caractères dont la similarité ne résulte pas d'un ancêtre commun.

Isoformes : différentes formes d'une même protéine, issues d'épissages alternatifs.

k-mer : sous-chaîne nucléotidique d'une longueur spécifique. Utilisés dans notre étude lors de l'assemblage *de novo* de transcriptome avec $k = 25$.

Monophylie : qualifie un groupe de taxons comprenant un ancêtre commun et tous ses descendants. (groupe monophylétique = groupe naturel = clade = lignée évolutive).

Orthologie : divergence entre des caractères homologues résultant uniquement de la spéciation (voir encadré 6, p. 139).

Paralogie : divergence entre des caractères homologues résultant d'un évènement de duplication chez un ancêtre commun (voir encadré 6, p. 139).

Read : séquence relativement courte obtenue en sortie de séquençage de nouvelle génération. Dans notre étude environ 100 nucléotides de long.

Synapomorphie : état dérivé d'un caractère partagé par plusieurs taxons frères, permettant de caractériser un groupe monophylétique.

Transcrit : séquence résultante de la transcription d'une séquence d'ADN en ARN. Il peut coder pour une protéine (cas des ARN messagers) ou non (cas des ARN ribosomiaux, ARN de transfert, micro ARN, ...).

Transcriptome : l'ensemble des molécules d'ARN (ou transcrits) d'une cellule ou d'un organisme ; cela comprend à la fois les ARN codant pour des protéines et les ARN qui ne codent pas pour une protéine.

Paraphylie : qualifie un groupe qui comprend une espèce ancestrale et seulement une partie de tous ses descendants, généralement défini par une plésiomorphie.

Polyphylie : qualifie un groupe qui contient un certain nombre de taxons, mais ne contient pas l'ancêtre commun à tous, généralement défini par un caractère convergent.

Profondeur de séquençage : nombre de fois où une même base est séquencée.

Voucher : spécimen, échantillon, ses produits et ses données associées qui documentent l'existence physique de cet organisme.

Limites des profondeurs référées dans cette thèse, aux différents étages bathymétriques :

| Etage | Profondeurs (m) |
|--------------|------------------------|
| Sub-littoral | 0 – 200 |
| Bathyal | 200 – 2 000 |
| Abyssal | 2 000 – 6 000 |
| Hadal | 6 000 – 11 000 |

Chapitre I

Introduction générale de la thèse

I. INTRODUCTION GENERALE DE LA THESE

I.1 Contexte général :

A l'époque des Grandes Expéditions du XIX^{ème} siècle, l'environnement abyssal est perçu comme un milieu hostile à la vie, voire azoïque à partir d'une certaine profondeur (hypothèse azoïque, Forbes, 1844). Cette vision est en grande partie liée au fait que la vie dans les profondeurs (à partir de 200 m) est dépendante des apports nutritifs dérivés de la surface où peut avoir lieu la photosynthèse, seule source de production primaire connue à cette époque.

Des spécimens ont pourtant parfois été fortuitement remontés du milieu profond, mais ont été considérés comme de simples curiosités. De grandes expéditions se mettent alors en place afin de tester cette hypothèse azoïque et montrent que, même si la biomasse de la faune profonde est faible, son originalité taxonomique lui confère un attrait scientifique majeur. En particulier, ces expéditions ont permis la découverte d'espèces morphologiquement proches d'espèces connues uniquement du registre fossile, comme le crinoïde *Rhizocrinus lofotensis* Sars 1868. Ces découvertes ont mené à l'idée que le milieu profond, perçu comme stable, monotone et protégé des changements importants impactant la surface (Tunncliffe, 1992), permettrait la survie sur des temps longs de lignées anciennes, sans changement notable des organismes (appelés 'fossiles vivants') (Samadi, 2015). Cependant, cette abondance de lignées anciennes est une perception biaisée de la biodiversité en milieu profond, notamment due à l'étude préférentielle de ces taxons originaux.

En 1976, la découverte des sources hydrothermales profondes (encadré 1) hébergeant une biocénose luxuriante modifie grandement cette perception des écosystèmes profonds. La très forte biomasse rencontrée à cette profondeur a rapidement suggéré un apport nutritif original, déconnecté de celui de la surface. Le réseau trophique de ces oasis repose sur la production primaire par des bactéries chimiosynthétiques qui utilisent l'énergie dégagée par une réaction d'oxydation des composés réduits (typiquement les sulfures mais aussi le méthane, l'hydrogène et le fer ferreux) présents dans les fluides émis (Corliss *et al.*, 1979). Dans cet environnement, certains métazoaires vivent en association symbiotique avec des bactéries chimiosynthétiques (Dubilier *et al.*, 2008 ; Jannasch et Mottl, 1985). Ces associations sont considérées comme une clef adaptative dans la colonisation des sources hydrothermales par ces organismes. Bien que formant la plus grande partie

Encadré 1 : Les sources hydrothermales, un environnement aux conditions extrêmes

La découverte des sources hydrothermales date d'une campagne effectuée en 1976, lors d'une plongée à bords du submersible américain l'*Alvin* (Lonsdale, 1977). Menée par des géologues, cette campagne sur la Ride de Galápagos avait pour but de vérifier l'hypothèse selon laquelle des anomalies détectées dans le bilan thermique seraient dues à l'existence d'une circulation hydrothermale (Lister, 1972).

Aujourd'hui, environ 700 champs hydrothermaux ont été découverts, dont plus d'une cinquantaine sont inactifs. Parmi les 650 restants, l'activité de près de 300 d'entre eux a été confirmée visuellement (InterRidge Vents Database v3.4). Les sources hydrothermales se situent principalement le long des dorsales océaniques ou dans les zones des bassins arrière-arcs (Tyler *et al.*, 2002). Dans ces zones, le plancher océanique est fissuré à cause des mouvements des plaques tectoniques, créant des passages vers la proximité de la chambre magmatique. L'eau de mer s'infiltre dans ces fissures et voit alors sa composition se modifier sous l'effet de la température et de l'interaction avec la roche, avant de ressurgir à 350-400°C dans l'eau de mer froide (figure I.1).

Si les fortes pressions dues à la profondeur et l'absence de lumière sont les mêmes qu'en plaine abyssale, l'émission de fluides et leur mélange avec l'eau de mer froide résulte en un ensemble de conditions physico-chimiques considérées extrêmes : des températures extrêmement fortes et variables, une faible concentration en oxygène et la présence de composés toxiques dans les fluides (sulfures, éléments métalliques, composés contenant de l'arsenic, ...). Les concentrations de ces composés ainsi que leur variabilité spatio-temporelle (à des échelles aussi fines que le cm) rendent cet environnement l'un des plus contraignants sur notre planète.

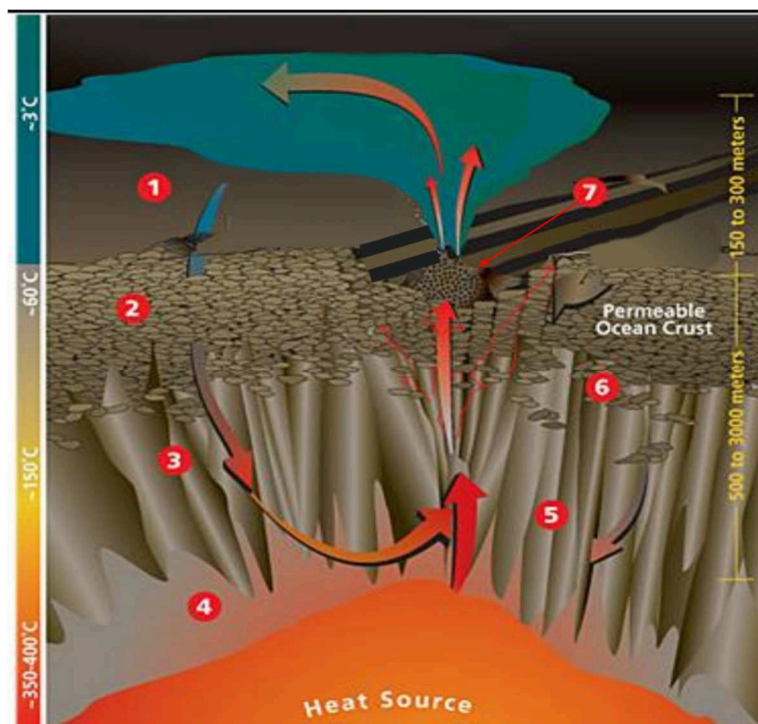


Figure I.1 Schéma de la circulation hydrothermale au niveau des dorsales océaniques. (1) L'eau de mer percole dans les fissures de la croûte océanique ; **(2)** elle s'appauvrit en composés oxydés tels l'oxygène (O_2), les nitrates (NO_3^-) et sulfates (SO_4^{2-}). **(3)** Le fluide (nom donné à l'eau de mer modifiée) perd ses cations calcium (Ca^{2+}) et magnésium (Mg^{2+}) ; **(4)** le lessivage basaltique enrichit le fluide en composés tels que Sodium (Na^+), potassium (K^+), sulfure d'hydrogène (HS^-), manganèse (Mn^{2+}), métaux lourds (Fe, Cu, Zn) et CO_2 ; **(5)** le fluide se rapproche des chambres magmatiques et atteint sa température maximale, ce qui provoque **(6)** sa remontée en surface de la croûte océanique. Lorsque ce fluide chaud entre en contact avec l'eau de mer froide (2-4°C), certains éléments minéraux précipitent et forment des édifices **(7)**. Image modifiée à partir de <https://www.whoi.edu>.

de la biomasse, les espèces symbiotiques sont peu nombreuses et un nombre important d'espèces non symbiotiques sont aussi inféodées aux sources hydrothermales. Celles-ci se nourrissent généralement de tapis bactérien (bactéries autotrophes et hétérotrophes) ou sont des prédateurs d'espèces consommatrices de production primaire bactérienne (Tunnicliffe, 1992).

Dans un premier temps, la particularité de cet environnement, majoritairement indépendant de la production photosynthétique de surface, a suggéré l'hypothèse que la faune hydrothermale a évolué de manière continue sans connaître d'épisode d'extinction massif (Tunnicliffe, 1992). Lors de la description de certains organismes, les auteurs ont ainsi interprété la spécialisation des espèces pour le milieu hydrothermal comme le témoignage de cette hypothèse d'une histoire évolutive longue et isolée des autres environnements marins. Les organismes spécialisés des sources hydrothermales ont donc été interprétés comme des lignées dites « endémiques » de cet environnement, et ont parfois été classés dans des catégories de haut rang taxonomique. Cette classification taxonomique de haut rang reflète l'hypothèse que la lignée est présente depuis longtemps dans le milieu hydrothermal et que la diversification s'est exclusivement faite dans ce milieu (Newman, 1985 ; Wolff, 2005). Les modifications morphologiques sont parfois si extrêmes que certains auteurs ont même créé un nouveau phylum pour accommoder des espèces retrouvées dans ces environnements (par exemple les Vestimentifera ; Jones, 1985). L'avancée des études taxonomiques mais également de l'exploration d'une plus large gamme de milieux marins, notamment dans le domaine profond (au-delà de 200 m), ainsi que des études phylogénétiques ont permis d'invalider la plupart de ces taxons de haut rang (Samadi, 2015).

Les vers de la famille des Siboglinidae et les moules de la sous-famille des Bathymodiolinae sont deux exemples qui illustrent la façon dont les taxons de haut rang initialement proposés ont pu être révisés à l'aide d'approches phylogénétiques fondées sur un échantillonnage taxonomique et écologique élargis autour du groupe hydrothermal d'intérêt.

I.1.1 Une histoire de reconsidérations taxonomiques : les vers de la famille Siboglinidae

Le cas des vers Siboglinidae illustre un changement de rang taxonomique d'une ampleur peu commune : deux phylums avaient été créés pour accommoder des organismes qui sont aujourd'hui rassemblés dans une même famille d'annélides polychètes. Les changements de statut des vers

hydrothermaux symbiotiques ont été revus par Pleijel *et al.* (2009). Cet exemple montre la nécessité d'échantillonner d'autres milieux pour connaître l'origine évolutive d'un groupe et évaluer son éventuelle monophylie.

A l'origine des débats taxonomiques se trouve la découverte, dans les environnements profonds chimiosynthétiques (suintements froids et sources hydrothermales), de grands vers sans tube digestif. Ils ont été décrits sous les noms de *Lamellibrachia barhami* et *Riftia pachyptila*. Leur morphologie peu commune a conduit les auteurs à les placer dans un nouvel ordre : les Vestimentifera (Jones, 1981), rapidement érigé au rang de phylum (Jones, 1985). L'absence de tube digestif est un caractère qui avait été rencontré auparavant pour des organismes provenant de différents milieux profonds (de 200 à 10 000 m) et placés dans un autre phylum, les Pogonophora (Beklemishev, 1944), desquels sont immédiatement rapprochés les Vestimentifères (Jones, 1981). L'absence de tube digestif chez les Pogonophora complique la compréhension du plan d'organisation de ces spécimens de petite taille, et de nombreuses réinterprétations de l'axe dorso-ventral sont proposées, amenant les auteurs à remettre en question leur placement dans les Deutérostomes, proposé lors de la description du groupe (Nørrevang, 1970). La découverte de *Riftia pachyptila* amène également à celle d'un organe particulier, le trophosome, dans lequel sont hébergées des bactéries chimiosynthétiques essentielles pour la nutrition de l'hôte (Felbeck *et al.*, 1981 ; Cavanaugh *et al.*, 1981). A la lumière de la découverte de cette symbiose trophique chez les Vestimentifères, Southward *et al.*, (1981) ont réexaminé les espèces de Pogonophores et ont mis en évidence la présence de cette même symbiose comme source nutritive majeure. Si l'assignation des Pogonophora à un groupe taxonomique connu (Protostome / Deutérostome) était difficile, l'opisthosome (partie postérieure de l'organisme) des Vestimentifera montre une segmentation claire et des soies, permettant de proposer un rapprochement de ces organismes avec les Annélides. En fait, l'opisthosome avait également été observé chez les Pogonophora par Webb (1964), et l'auteur explique que c'est une partie du corps artéfactuellement très peu récoltée lors des prélèvements des spécimens, car elle sert dans l'ancrage des organismes dans leur tube. De plus, la découverte d'une larve de Vestimentifère (dans laquelle le tube digestif est évident) lève le doute sur l'interprétation du plan d'organisation de ces organismes. En effet, la présence d'un cordon nerveux ventral chez la larve appuie l'hypothèse d'un rapprochement avec les Annélides (Jones et Gardiner, 1988). Depuis, plusieurs études phylogénétiques ont rapproché les Vestimentifères et les Pogonophores (appelés Frenulata depuis Webb, 1969) des Annélides (Kojima, 1993 ; Rouse et Fauchald, 1995, 1997). Ils sont aujourd'hui réunis dans une seule et même famille, les Siboglinidae, qui appartient aux Annélides polychètes. Cette famille regroupe tous les vers connus à ce jour qui ne possèdent pas de tube digestif et qui hébergent des bactéries symbiotiques. Elle contient donc les

anciens Vestimentifera, les Frenulata et également des vers retrouvés sur des carcasses de baleines, appartenant au genre *Osedax* (Glover *et al.*, 2005). Ces vers hébergent des bactéries hétérotrophes, permettant la dégradation des os et des lipides qu'ils contiennent (Goffredi *et al.*, 2005). Cette famille inclut également les espèces du genre *Sclerolinum*, retrouvées sur les suintements froids et les carcasses de vertébrés, positionnées en groupe frère des anciens Vestimentifera (Li *et al.*, 2017).

Ainsi, cet exemple montre que des interprétations de caractéristiques morphologiques très inhabituelles peuvent mener à l'érection de regroupements taxonomiques de très haut niveau (deux phylums), alors que les spécimens concernés ne constituent en fait qu'une famille d'annélides polychètes spectaculaires, à la morphologie très modifiée par leur mode de vie symbiotique, et dont toutes les espèces ne sont pas inféodées aux sources hydrothermales. Dans cet exemple, c'est la combinaison de la comparaison d'organismes de différents milieux et à différents stades de leur cycle de vie, avec des approches phylogénétiques s'appuyant sur des caractères à la fois morphologiques et génétiques, qui a permis de revoir l'hypothèse initiale et de corriger le positionnement taxonomique des espèces anciennement placées dans les Vestimentifera.

I.1.2 Etude de la colonisation des sources hydrothermales : les moules de la sous-famille Bathymodiolinae

L'exemple des moules de la sous-famille des Bathymodiolinae montre également l'importance d'échantillonner d'autres environnements que le milieu hydrothermal pour reconstruire l'histoire évolutive des espèces inféodées à ce milieu.

Dès la description de la première espèce découverte sur les sources hydrothermales, *Bathymodiolus thermophilus*, Kenk et Wilson (1985) établissent la sous-famille Bathymodiolinae pour accommoder ces moules géantes. Comme *B. thermophilus*, les espèces placées par la suite dans cette sous-famille possèdent des bactéries endosymbiotiques chimiosynthétiques (Nelson *et al.*, 1995) et sont dans un premier temps exclusivement celles récoltées dans des environnements hydrothermaux ou de suintements froids (Cosel et Olu, 1998). En se basant sur les moules connues de ces deux types d'environnements chimiosynthétiques, Craddock *et al.* (1995) proposent une hypothèse évolutive sur la colonisation de ces milieux profonds chimiosynthétiques. Ils envisagent un scénario avec une colonisation graduelle des moules depuis les milieux peu profonds jusqu'aux milieux profonds, en colonisant d'abord les suintements froids puis les sources hydrothermales. Dans un second temps, Distel *et al.* (2000) affinent ce scénario en proposant une étape supplémentaire de colonisation de substrats organiques tels que les bois coulés et les carcasses de baleine, avant la colonisation des suintements froids et des sources hydrothermales (the "wooden-step" hypothesis),

en incluant dans leur étude des moules de petite taille trouvées sur ces substrats. Plus tard, Lorion *et al.* (2010) publient une étude phylogénétique basée sur deux marqueurs indépendants (COI et 28S), couplée avec un échantillonnage taxonomique large, ciblant des spécimens récoltés sur ces trois environnements. Elle montre que les moules géantes récoltées dans les milieux hydrothermaux et les suintements froids ont une histoire évolutive entremêlée avec celle des petites moules associées aux substrats organiques coulés dans le milieu profond. La sous-famille décrite pour les moules géantes des sources hydrothermales doit donc intégrer ces petites moules des substrats organiques pour former un groupe monophylétique*. Thubaut *et al.* (2013) améliorent la phylogénie de cette sous-famille en reprenant l'échantillonnage taxonomique de Lorion *et al.* (2010) et en y ajoutant cinq nouveaux marqueurs phylogénétiques (16S, 18S, H3, HSP70 et une translocase adénosine nucléaire - Ant). Cette approche leur permet de résoudre un grand nombre des relations qui ne l'étaient pas dans la précédente étude et de définir des genres dont la plupart comportent non seulement des espèces inféodées à différents environnements, mais aussi des espèces présentes dans plusieurs environnements. Ces travaux montrent qu'au moins quatre lignées ont indépendamment connu une transition des milieux profonds caractérisés par des accumulations de substrats organiques vers les sources hydrothermales ou les suintements froids. Ces transitions s'accompagnent d'une augmentation de la taille des espèces. L'hypothèse du « wooden step » initialement proposée par Distel *et al.* (2010) est donc revue pour un scénario plus complexe, que Thubaut *et al.* (2013) renomment en « multiple wooden steps ». En complément à cette analyse, Lorion *et al.* (2013) proposent des reconstructions d'états ancestraux de plusieurs caractères tels que l'habitat, la présence de symbiotes chimiotrophes et/ou méthanotrophes, ou encore la localisation intracellulaire ou extracellulaire de ces symbiotes. D'une part, leurs résultats confirment la conclusion de Thubaut *et al.* (2013) pour les habitats ancestraux et les multiples colonisations depuis les substrats organiques. D'autre part, ils mettent en évidence que la présence de ces symbiotes chimioautotrophes n'est pas spécifique des milieux hydrothermaux et que ce n'est pas l'acquisition de nouveaux types de symbiotes qui explique la diversification dans les milieux hydrothermaux. En effet, ils montrent que cette acquisition ne se situe pas dans les branches profondes avant une radiation rapide du groupe, mais que celle-ci est observée dans plusieurs lignées indépendantes. Leurs conclusions n'ont été possibles que grâce au large échantillonnage dans les milieux hydrothermaux mais aussi dans les milieux profonds d'accumulation de substrats organiques, ou de suintements froids (Lorion *et al.*, 2013 ; Thubaut *et al.*, 2013), soulignant une fois de plus l'importance d'échantillonner des environnements divers.

* Les mots suivis d'un astérisque dans le texte sont expliqués dans le lexique, pages 10 et 11.

I.2 Stratégies d'étude de la thèse

L'objectif principal de la thèse est de comprendre le contexte évolutif de plusieurs taxons hydrothermaux et de tester, par une approche phylogénétique, les hypothèses évolutives émises lors des descriptions de ces groupes. Une phylogénie basée uniquement sur des caractéristiques morphologiques peut être faussée par l'utilisation de caractères issus de convergences évolutives*. Pour nous affranchir d'un tel biais, nous avons eu recours à des techniques de reconstructions phylogénétiques basées sur des données moléculaires. Nous avons donc essayé, d'une part, de rassembler un échantillonnage représentatif de la biodiversité des groupes d'intérêt, et d'autre part d'utiliser des marqueurs moléculaires permettant de résoudre la profondeur phylogénétique ciblée. Pour qu'un échantillonnage soit représentatif d'un groupe (c'est à dire qu'il couvre sa diversité taxonomique), il faut, dans la mesure du possible, inclure à la phylogénie au moins deux espèces pour chaque genre et au moins deux genres de chaque famille ou sous-famille. Les différents groupes d'études choisis sont ici exposés rapidement. Chacun d'entre eux correspond à un chapitre de cette thèse, et une justification détaillée du choix de ces groupes est fournie dans leur introduction respective.

Les vers de la famille des Polynoidae (**Chapitre II**) représentent une part importante de la biodiversité hydrothermale. Ils y occupent toute la gamme de conditions environnementales dans lesquelles des métazoaires sont rencontrés, et jusqu'à 10 espèces peuvent être retrouvées sur un même site hydrothermal. Cinq sous-familles ont été décrites comme endémiques des sources hydrothermales (Marcus et Hourdez, 2002 ; Pettibone, 1983, 1984, 1985a, 1985b). Depuis, quelques espèces placées dans certaines de ces sous-familles ont été trouvées dans d'autres environnements, tels que bois coulés, carcasses de vertébrés ou suintements froids, où la chimiosynthèse est une source de production primaire. Aucune étude phylogénétique actuelle n'a ciblé les Polynoidae. Afin de répondre à la question de l'origine évolutive (unique ou multiple) des Polynoidae hydrothermaux, il est nécessaire d'éclaircir la définition et les relations des taxons au sein de cette famille des Polynoidae. Les données moléculaires disponibles dans les banques de données et les campagnes réalisées dans différents milieux ont fournis des échantillons permettant, au cours de cette thèse, de proposer une approche phylogénétique moléculaire comportant des espèces provenant d'un grand nombre de sous-familles acceptées aujourd'hui, ainsi que d'une grande variété d'environnements marins. Pour cette famille, l'approche mise en œuvre est la même que celle qui a permis de résoudre la question de l'origine des moules Bathymodiolinae (Thubaut *et al.*, 2013) : un échantillonnage large, mené dans des milieux divers, afin de couvrir une large gamme de conditions écologiques et

des marqueurs génétiques « universels », obtenus par séquençage Sanger (Sanger *et al.*, 1977) de produits de PCR (Polymerase Chain Reaction).

Cependant, ces approches multigènes par PCR suivies de séquençage Sanger ne permettent pas toujours de résoudre les relations phylogénétiques profondes. En effet, l'utilisation de marqueurs « universels » (*i.e.* amplifiables par PCR sur une large gamme taxonomique) n'aboutit pas toujours à l'obtention d'un signal phylogénétique suffisant. Dans des études antérieures, les approches multigènes utilisant le séquençage Sanger de produits de PCR n'a pas permis de résoudre les nœuds profonds permettant de replacer deux autres groupes taxonomiques : la famille de crabes hydrothermaux Bythograeidae appartenant à l'infra-ordre des Brachyura et la famille de crevettes hydrothermales Alvinocarididae appartenant à l'infra-ordre des Caridea. Au cours de cette thèse, nous nous sommes attachés à développer de nouvelles approches dans la définition de marqueurs phylogénétiques et dans la méthode de séquençage pour ces deux groupes taxonomiques.

Les crabes Bythograeidae forment une famille de crabes strictement inféodés aux milieux hydrothermaux (Wolf, 2005). Cette famille s'inscrit dans le très large infra-ordre des Brachyura. Les études publiées à ce jour et incluant quelques espèces de Bythograeidae comprennent soit un échantillonnage centré sur cette famille et un contexte plus large très limité (Mateos *et al.*, 2012), soit un contexte large de Brachyura mais un échantillonnage très limité des Bythograeidae (Shen *et al.*, 2013 ; Yang *et al.*, 2013). L'étude phylogénétique proposant à ce jour l'échantillonnage taxonomique le plus large au sein des Brachyura et couplée à une approche multigène ne résout pas les nœuds profonds, et n'inclut pas la famille des Bythograeidae (Tsang *et al.*, 2014). Ces études ne permettent donc pas de tester la monophylie de cette famille de crabes hydrothermaux. Le **chapitre III** du présent manuscrit expose les résultats de cette étude.

Dans le cas des crevettes Alvinocarididae, Aznar-Cormano *et al.* (2015) ont proposé une phylogénie couplant à la fois un nombre important (cinq) de marqueurs phylogénétiques (obtenus par amplification PCR et séquençage Sanger) et un échantillonnage taxonomique large (28 des 40 familles de l'infra-ordre des Caridea sont couvertes). Cette étude a permis de confirmer la monophylie de la plupart des familles de Caridea, mais les marqueurs phylogénétiques utilisés n'ont là non-plus pas fourni un signal suffisamment informatif pour préciser les relations entre les familles. Parmi les Caridea, les crevettes Alvinocarididae forment une lignée, mais les relations entre celle-ci et les autres familles de Caridea restent non résolues. La résolution des relations phylogénétiques entre les lignées familiales des Caridea, notamment l'identification de son groupe frère, permettra de comprendre l'origine évolutive de la famille des Alvinocarididae. Les résultats de cette étude sont présentés dans le **chapitre IV**.

La différence majeure entre l'étude sur les Bathymodiolinae, pour laquelle l'approche multigène classique a suffi à résoudre les relations permettant de comprendre l'origine évolutive des moules hydrothermales et les études sur les Brachyura et les Caridea est la profondeur phylogénétique ciblée. L'étude de Thubaut *et al.* (2013) sur les Bathymodiolinae cible des espèces ayant divergé entre 40 et 80 Ma (Lorion *et al.*, 2013 ; Miyazaki *et al.*, 2010), tandis que celles de Tsang *et al.* (2014) et d'Aznar-Cormano *et al.* (2015) sur les Brachyura et les Caridea respectivement, ciblent les relations entre espèces au sein d'infra-ordres ayant des divergences estimées à environ 220 Ma pour les Brachyura et 260 Ma pour les Caridea (Bracken *et al.*, 2010). L'objectif de cette thèse est donc d'améliorer le signal phylogénétique dans ces deux groupes en augmentant le nombre de marqueurs partagés à l'échelle des Caridea d'une part et des Brachyura d'autre part. L'approche de l'amplification par PCR se complique lorsque les espèces ciblées sont très divergentes (pour lesquelles les zones d'hybridation des amorces ne sont pas suffisamment conservées à cette échelle ; Aznar-Cormano *et al.*, 2015), et ne permet pas de définir facilement de nouveaux marqueurs. De plus, le séquençage Sanger trouve sa limite dans son coût lorsque le nombre de marqueurs ciblés est trop important. Dans ces cas, des approches génomiques couplées aux techniques de séquençage de nouvelle génération (ou NGS) sont plus appropriées (Metzker, 2005 ; Hutchison, 2007). Plusieurs approches peuvent être mises en œuvre dans ce but d'identifier un plus grand nombre de marqueurs, permettant de résoudre les nœuds profonds. Parmi elles, (i) multiplier les marqueurs en séquençant le génome mitochondrial complet, ce que nous avons fait pendant cette thèse dans le cas des crabes Bythograeidae (**Chapitre III**) ou (ii) déterminer de nouveaux marqueurs à partir de données de transcriptomique*, ce que nous avons fait pendant cette thèse dans le cas des crevettes Caridea (**Chapitre IV**).

Chapitre II

La phylogénie par une approche multigène : le cas des vers de la famille Polynoïdæ

II. LA PHYLOGENIE PAR UNE APPROCHE MULTIGENE : LE CAS DES VERS DE LA FAMILLE POLYNOIDAE

II.1 Introduction

II.1.1 Les Polynoidae dans la faune des sources hydrothermales

La famille des annélides Polynoidae contient une cinquantaine d'espèces strictement inféodées à la proximité des sources hydrothermales. Ces espèces représentent environ 10% des espèces hydrothermales décrites à ce jour (Tunncliffe, 1998 ; Hourdez, *com. pers.*). Pour tirer bénéfice de la production primaire locale importante basée sur la chimiosynthèse, ces espèces doivent supporter les contraintes associées à la présence de fluides hydrothermaux : température parfois élevée et souvent très variable, présence de sulfures, oxygène limité, présence de métaux lourds et autres composés toxiques. Les espèces de Polynoidae hydrothermales sont présentes dans toutes les grandes provinces biogéographiques de la faune hydrothermale (déterminées en se basant sur la composition faunistique ; Moalic *et al.*, 2011) mais chaque espèce, parmi la cinquantaine d'espèces de Polynoidae hydrothermales décrites à ce jour, n'est pas cosmopolite. Sur un site hydrothermal donné, les espèces de Polynoidae se répartissent depuis les zones sans influence du fluide hydrothermal jusqu'à la surface des cheminées (à proximité immédiate des émissions). La diversité sur un site peut atteindre 10 espèces et chacune est inféodée à un habitat particulier : en périphérie, parmi les grands organismes symbiotiques (moules, Vestimentifères), ou encore sur les parois des cheminées (figure II.1). Cette spécialisation par habitat suggère des adaptations différentes aux conditions spécifiques de chaque habitat, en particulier à la température et aux paramètres chimiques qui lui sont associés. Au niveau du genre cependant, à part pour le genre commensal des moules *Branchipolynoe*, aucun n'est spécifiquement associé à un type d'habitat. Par exemple, la paroi des cheminées est habitée par une espèce de *Lepidonotopodium* (*L. fimbriatum*) sur la dorsale du Pacifique Oriental alors que c'est une espèce du genre *Branchinotogluma* (*B. segonzaci*) dans les bassins arrière-arcs du Pacifique Ouest (figure II.1). Ainsi, sur des sites différents, les habitats ne sont pas exclusivement colonisés par une espèce ou un genre en particulier.

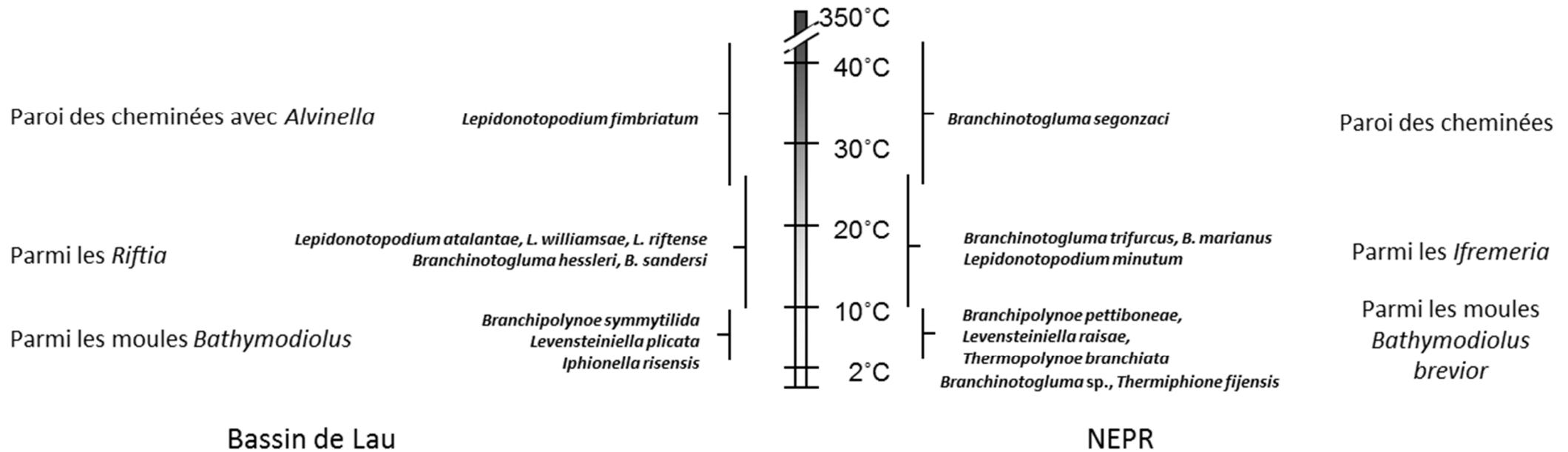


Figure II.1 Distribution des espèces de Polynoidae sur un site du Bassin de Lau (à gauche) et sur le site 9°50'N de la Dorsale du Pacifique Oriental (NEPR, à droite) en fonction de la température (un paramètre corrélé à la chimie dans cet écosystème). Jusqu'à 9 espèces peuvent être trouvées dans ces zones.

Comme la quasi-totalité des Polynoidae connus, les Polynoidae hydrothermaux sont des prédateurs (Voight, 2000 ; Jumars *et al.*, 2015). Les espèces du genre *Branchinotogluma* et quelques espèces de *Lepidonotopodium* semblent également se nourrir de tapis bactérien et alterner entre deux modes de nutrition (Gaudron *et al.*, 2012). *Bathykurila guayamensis* se nourrit également de tapis bactérien de *Beggiatoa* (Glover *et al.*, 2005). La seule exception au régime prédateur est *Vampiropolynoe embleyi*, qui est une espèce dépourvue de mâchoires (apomorphie* unique chez les Polynoidae) (Marcus et Hourdez, 2002). Les espèces du genre *Branchipolynoe* sont commensales des grandes moules Bathymodiolinae et se nourrissent des pseudo-fèces de leur hôte ou même de fragments de leurs branchies (Desbruyères *et al.*, 1985 ; Van Dover, 2002). Ce mode de vie particulier peut mener à des convergences morphologiques, comme par exemple la réduction de la taille des élytres et des parapodes rencontrée chez les *Branchipolynoe*, ce qui est une adaptation commune des espèces de Polynoidae symbiotiques (Martin et Britayev, 1998). Parmi les espèces de Polynoidae hydrothermales, certains traits inhabituels sont rencontrés, tels que : (i) l'absence d'antennes latérales chez toutes les espèces hydrothermales (un caractère partagé avec également les espèces de la sous-famille Macellicephalinae et certaines autres espèces profondes) ; (ii) la présence de branchies chez certaines espèces hydrothermales (toujours absentes chez les espèces non hydrothermales, Hourdez et Jouin-Toulmond, 1998) ; (iii) la présence de papilles ventrales allongées (présentes aussi chez certains Macellicephalinae). On peut se demander si ces traits résultent d'une réponse adaptative similaire et acquise indépendamment dans plusieurs lignées. Selon cette dernière hypothèse, les contraintes environnementales pourraient être à l'origine de convergences morphologiques et physiologiques importantes. Ces caractères, considérés comme inhabituels chez les Polynoidae, sont utilisés pour justifier des rapprochements taxonomiques (définition de genres et de sous-familles notamment). On peut donc se demander si ces rapprochements sont fondés parce qu'ils reposent sur des synapomorphies* ou s'ils sont à réviser parce que les états de caractères utilisés sont homoplasiques et reflètent seulement des convergences.

Ces questions taxonomiques soulèvent celle de l'histoire évolutive à l'origine de la diversité des Polynoidae hydrothermaux. Notamment, la diversité observée à ce jour résulte-telle d'un ou plusieurs événements de colonisation du milieu hydrothermal, accompagné(s) d'une diversification des espèces dans des niches étroites (concept de 'species flock') ? Aux vues des connaissances actuelles, le nombre important de sous-familles et de genres hydrothermaux, au sein desquels les espèces ont des niches thermiques bien différenciées, suggère une dynamique de colonisation suivie d'une radiation adaptative. Afin de tester ceci, il est donc nécessaire de déterminer la position phylogénétique des espèces hydrothermales au sein de l'ensemble des Polynoidae.

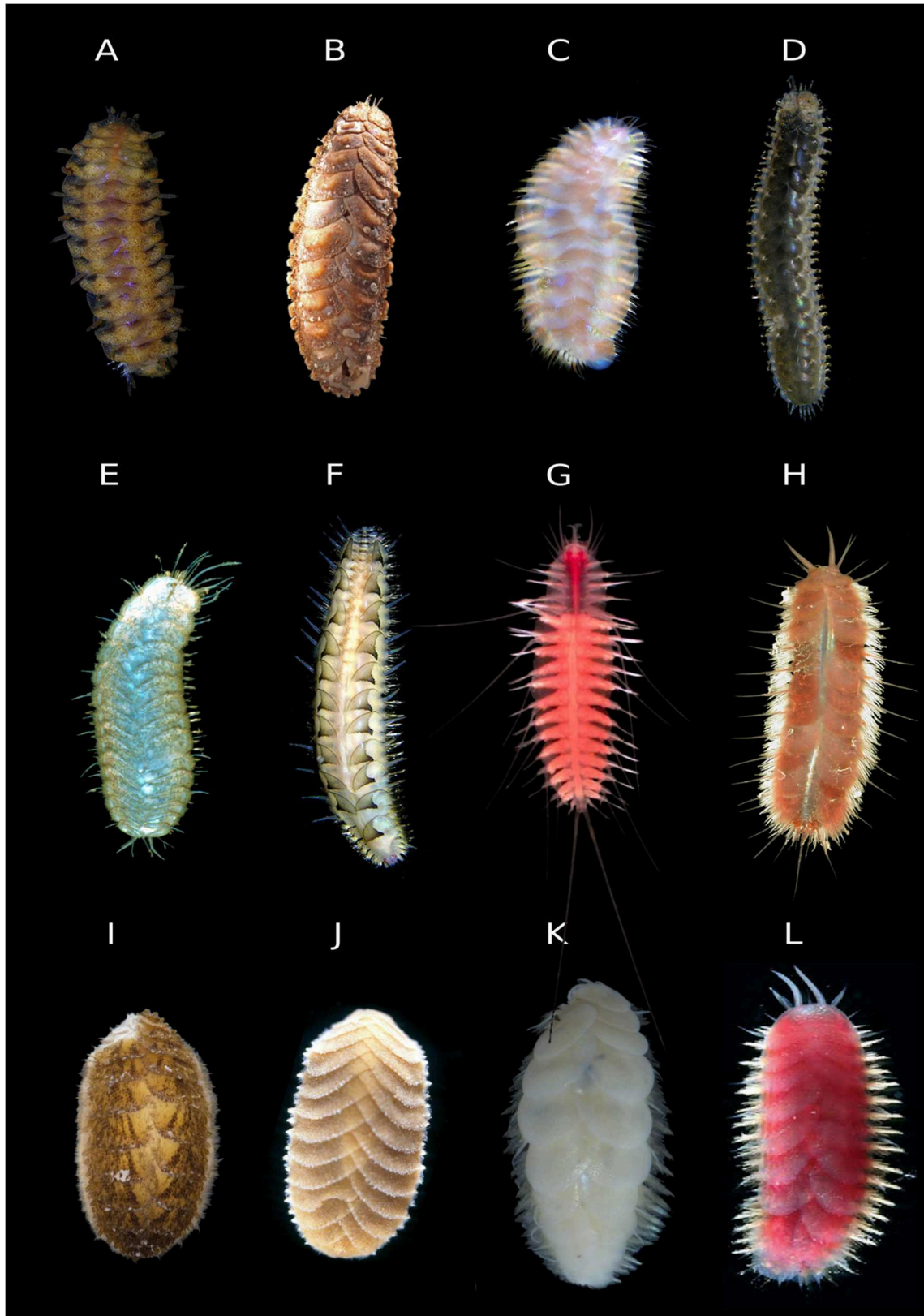


Figure II.2 Quelques espèces de Polynoidae illustrant la diversité morphologique de la famille. Sous-famille Arctonoinae : (A) *Gastrolepidia* sp. ; Lepidonotinae : (B) *Euphione squamosa*, (C) *Halosydna* sp., (D) Lepidonotinae Gen. sp. ; Polynoinae : (E) *Harmothoe fuligineum*, (F) *Malmgreniella* sp. ; Macellicephalinae : (G) *Pelagomacellicephala* sp. ; Branchiplicatinae : (H) *Branchiplicatus cupreus* ; Iphioninae : (I) *Iphione malifera*, (J) *Thermiphione fijensis* ; Lepidonotopodinae : (K) *Lepidonotopodium okinawae* ; Branchinotogluminae : (L) *Branchinotogluma trifurcus*. Pour des raisons de visibilité, les spécimens ne sont pas à l'échelle les uns par rapport aux autres. Crédits photos : Arthur Anker (A, C, D, F), G. Read (B), M. Rauschert (E), T. Iliffe (G), S. Hourdez (H), C. Piotrowski (I), P. Briand (J), J. Sui et X. Li (K), F. Pleijel (L).

II.1.2 Diversité des Polynoidae

Les Polynoidae sont une famille appartenant au sous-ordre des vers à élytres : les Aphroditiformia. Les limites et les relations entre familles de ce sous-ordre ne sont pas encore robustement définies et les changements récents de la classification sont importants. Les Aphroditiformia comptent environ 1 200 espèces réparties dans plus de 200 genres (Norlinder *et al.*, 2012) et sont actuellement généralement divisés en 6 familles (Gonzalez *et al.*, 2017a) : Acoetidae, Aphroditidae, Eulepethidae, Pholoidae, Polynoidae et Sigalionidae. La famille des vers Polynoidae ne possède pas de caractère synapomorphique clair et le placement d'une espèce dans cette famille n'est souvent basé que sur l'absence de certains caractères, caractéristiques des autres familles.

La famille de vers Polynoidae est la plus diversifiée au sein des Aphroditiformes : d'après WoRMS (Read et Fauchald, 2017), elle comporte environ 900 espèces réparties en 177 genres. Ils sont retrouvés dans tous les environnements marins : de la zone intertidale aux grandes profondeurs, depuis les eaux froides de l'Arctique ou de l'Antarctique jusque dans les zones tropicales (Guggolz *et al.*, 2018), dans les environnements profonds et oligotrophes ou dans les milieux enrichis en matière organique comme les sources hydrothermales, les suintements froids, les bois coulés (Amon, 2013 ; Chevaldonné *et al.*, 1998). La morphologie des animaux est très variable au sein de la famille (*cf.* figure II.2) : d'une quinzaine de segments à plus d'une centaine, pour des tailles allant de quelques millimètres à trente centimètres de long (*Eulagisca gigantea*). La plupart des espèces connues sont benthiques au moins à l'état adulte. Cependant, certaines espèces sont holopélagiques (vivent dans la colonne d'eau durant toute leur vie), comme les espèces de la sous-famille Admetellinae, ou encore les espèces des genres *Natopolynoe*, *Drieschia* et *Podarmus* ou l'espèce *Bylgides sarsi* (Jumars *et al.*, 2015 ; Sarvala, 1971). Certaines espèces de Polynoidae vivent cachées sous des pierres, dans des anfractuosités ou des tubes, tandis que d'autres espèces peuvent rester enfouies dans le sédiment (Jumars *et al.*, 2015). De nombreuses espèces vivent en symbiose obligatoire ou facultative (Martin et Britayev, 1998 ; Shields *et al.*, 2013 ; Britayev *et al.*, 2014 ; Serpetti *et al.*, 2017). Par exemple, les espèces placées dans la sous-famille hydrothermale Branchipolynoinae vivent de manière commensale avec des espèces de moules Bathymodiolinae (Pettibone, 1984, 1986 ; Miura, 1991).

La famille des Polynoidae est subdivisée en sous-familles dont la validité est encore discutée. Leur nombre et leurs relations ne sont pas précisément connus et de nombreux remaniements ont été proposés ces 40 dernières années. Dans son livre sur les familles et genres de Polychètes,

Tableau II.1 Sous-familles de Polynoïdæ actuellement reconnues, autorité, répartition bathymétrique et environnements connus. Sub-littoral (0-200 m), bathyal (200-2 000 m), abyssal (2 000-6 000 m), hadal (inférieur à 6 000 m).

| Sous-famille | Autorité | Répartition bathymétrique (Environnement) |
|----------------------|-------------------------|--|
| Admetellinae | Uschakov, 1977 | Bathyal |
| Arctonoinae | Hanley, 1989 | Sub-littoral |
| Bathyedithinae | Pettibone, 1976 | Abyssal |
| Bathymacellinae | Pettibone, 1976 | Abyssal |
| Branchinotogluminae | Pettibone, 1985b | Bathyal, Abyssal (sources hydrothermales et matière organique) |
| Branchiplicatinae | Pettibone, 1985a | Bathyal, Abyssal (sources hydrothermales) |
| Branchipolynoïnae | Pettibone, 1984 | Bathyal, Abyssal (sources hydrothermales, suintements froids) |
| Eulagiscinae | Pettibone, 1997 | Sub-littoral, Bathyal, Abyssal |
| Gesiellinae | Muir, 1982 | Sub-littoral (grottes anchialines) |
| Iphioninae | Baird, 1865 | Sub-littoral, Bathyal, Abyssal (sources hydrothermales et matière organique) |
| Lepidastheniinae | Pettibone, 1989 | Sub-littoral |
| Lepidonotinae | Willey, 1902 | Sub-littoral |
| Lepidonotopodinae | Pettibone, 1983 | Bathyal, Abyssal (sources hydrothermales) |
| Macellicephalinae | Hartmann-Schröder, 1971 | Sub-littoral, Bathyal, Abyssal (sources hydrothermales et grottes anchialines) |
| Macellicephaloidinae | Pettibone, 1976 | Abyssal, Hadal |
| Macelloïdinae | Pettibone, 1976 | Bathyal, Abyssal |
| Polaruschakovinae | Pettibone, 1976 | Abyssal |
| Polynoïnae | Kinberg, 1856 | Sub-littoral, Bathyal, Abyssal |
| Uncopolynoïnae | Wehe, 2006 | Sub-littoral |
| Vampiropolynoïnae | Marcus et Hourdez, 2002 | Bathyal, Abyssal (sources hydrothermales) |

Fauchald (1977) ne reconnaissait que quatre sous-familles en se basant sur la structure de la région antérieure (nombre et position des antennes). Uschakov (1977) reconnaît aussi ces quatre sous-familles, mais utilise en plus d'autres caractères relatifs aux fourreaux antennaires et tentaculaires pour en créer une nouvelle (Admetellinae), le menant ainsi à reconnaître cinq sous-familles. Muir (1982), lors d'une révision des caractères génériques des Polynoidae qu'il étend aux Aphroditiformes, propose douze sous-familles de Polynoidae, y compris une correspondant à la famille actuelle des Acoetidae (Polyodontinae dans l'article). Les caractères morphologiques inhabituels (*e.g.* : absence d'yeux, nombre de segments réduits et les papilles ventrales allongées, etc...) de certaines espèces des milieux profonds (vivant à des profondeurs de plus de 200 m) rendent difficile leur inclusion dans les sous-familles déjà existantes et ont donc mené certains auteurs à décrire de nouvelles sous-familles. Pettibone (1976) a créé notamment cinq nouvelles sous-familles (Macellicephaloidinae, Macelloidinae, Bathyedithinae, Polaruschakovinae et Bathymacellinae) pour placer les espèces profondes décrites plus tôt. De même, la découverte des sources hydrothermales a mis en évidence des Polynoidae morphologiquement originaux et a conduit à proposer cinq sous-familles pour accommoder la diversité morphologique rencontrée dans cet environnement (Branchinotogluminae, Branchipolynoidae, Branchiplicatinae, Lepidonotopodinae et Vampiropolynoidae). La dernière sous-famille en date, Uncopolynoidae Wehe, 2006, a été créée pour une seule espèce, elle-même imparfaitement connue. Au final, 20 sous-familles sont considérées actuellement (tableau II-1), beaucoup ne contenant parfois qu'une seule espèce ou un seul genre.

Les derniers changements dans la classification se fondent sur des phylogénies moléculaires et morphologiques (Norlinder *et al.*, 2012 ; Gonzalez *et al.*, 2017a). Ces auteurs proposent des remaniements parfois importants. En particulier, la sous-famille des Iphioninae pourrait être érigée au rang de famille. Cependant les auteurs de ces études soulignent qu'une révision taxonomique formelle incluant des données morphologiques et moléculaires sur un échantillonnage taxonomique couvrant mieux l'ensemble des sous-familles reste nécessaire. Dans notre étude, nous continuerons par conséquent à classer les Iphioninae au rang de sous-famille incluse dans les Polynoidae.

Parmi ces 20 sous-familles, certaines ont des répartitions très larges en terme de profondeur, tandis que d'autres vivent à des profondeurs plus restreintes. Cinq sont composées d'espèces rencontrées exclusivement en milieu littoral (Arctonoidae, Gesiellinae, Lepidastheniinae, Lepidonotinae, Uncopolynoidae) et quatre autres sont composées d'espèces littorales, bathyales (inférieur à 2 000m) et d'espèces abyssales (inférieur à 6 000 m) (Eulagiscinae, Iphioninae, Polynoidae et Macellicephalinae). Ces neuf sous-familles comprennent la grande majorité des espèces décrites à ce jour. Parmi les Macellicephalinae, 18 des 21 genres sont connus en milieu

abyssal (plaine sédimentaire à des profondeurs généralement supérieures à 2 000 m), deux genres sont retrouvés à proximité des sources hydrothermales (*Levensteiniella* et *Bathykurila*) et un genre dans les grottes anchialines (*Pelagomacellicephalo*). Les 11 sous-familles restantes sont exclusivement profondes (plus de 200 m de profondeur) et seulement une centaine d'espèces sont à ce jour connues. Initialement décrites comme restreintes aux sources hydrothermales, des espèces de Branchinotogluminae ont été récoltées sur d'autres milieux enrichis en matière organique (Pettibone, 1993) et des Branchipolynoinae ont été collectés au niveau de suintements froids (Pettibone, 1986). Trois sous-familles, Branchiplicatinae (une seule espèce), Vampiropolynoinae (une seule espèce) et Lepidonotopodinae (neuf espèces), n'ont à ce jour été récoltées que près des sources hydrothermales (Pettibone, 1985 ; Miura, 1994 ; Marcus et Hourdez, 2002 ; Sui et Li, 2017). En plus de ces sous-familles exclusivement associées aux environnements à base chimiosynthétique, certaines espèces décrites de ces environnements appartiennent à la sous-famille des Iphioninae qui comporte aussi des espèces de faible profondeur et d'environnements à base photosynthétique (Pettibone, 1986 ; Imajima, 2005).

Le but de notre étude est de déterminer si les espèces des sous-familles érigées pour accommoder les espèces hydrothermales, proposées sur la base de différences morphologiques, forment des groupes monophylétiques lorsque l'on considère uniquement des caractères moléculaires. L'objectif de ce travail est de replacer les taxons morphologiquement diversifiés et originaux découverts dans les sources hydrothermales au sein d'une phylogénie moléculaire incluant un échantillonnage d'espèces recouvrant la diversité taxonomique et écologique des Polynoidae. Cette approche nous permettra d'inférer l'histoire évolutive des taxons hydrothermaux.

II.1.3 Etat des connaissances de la phylogénie des Polynoidae

Assez peu d'études se sont intéressées aux relations phylogénétiques au sein des Aphroditiformia en général et des Polynoidae en particulier. Uschakov (1977) propose une première étude des relations phylogénétiques chez les Polynoidae, basée sur des observations morphologiques (figure II.3). Il tente de différencier l'état ancestral de l'état dérivé de caractères tels que la forme des élytres ou le nombre de segments au sein des Polynoidae (cinq sous-familles reconnues à ce moment). Les relations entre les genres et les sous-familles restent très floues, les caractères importants principalement utilisés pour définir les sous-familles sont le nombre et la position des antennes.

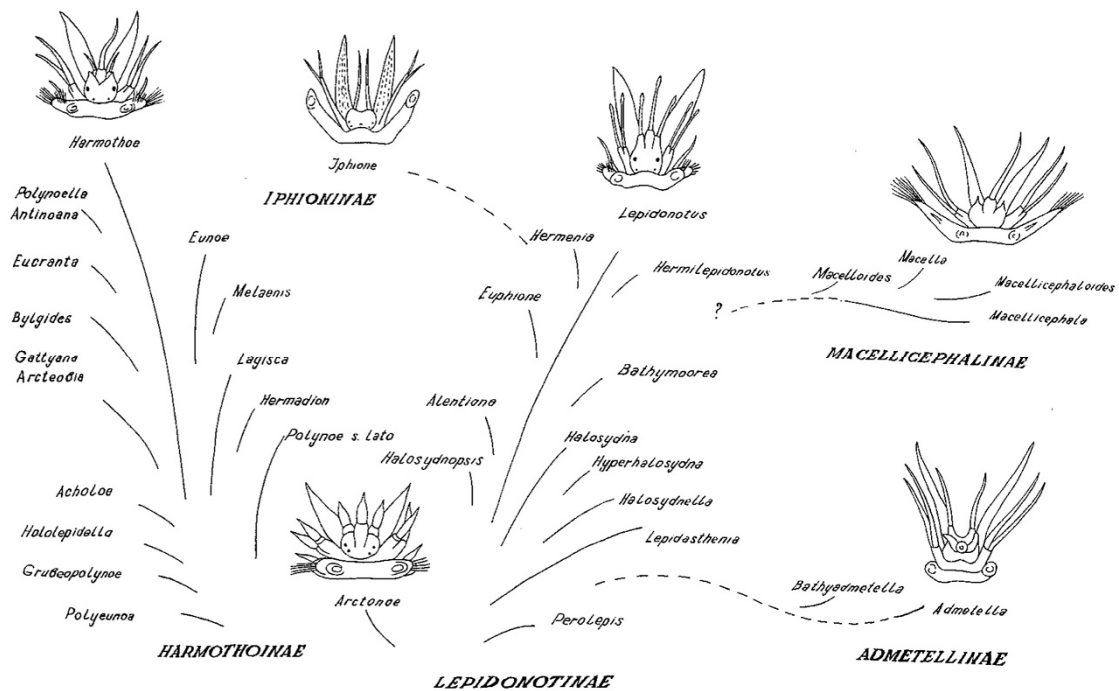


Figure II.3 Relations phylogénétiques entre les genres de Polynoidae selon Uschakov (1977). Uschakov reconnaît cinq sous-familles, essentiellement basées sur la morphologie de la région antérieure (nombre et position des antennes).

Muir (1982) propose ensuite une étude multivariée sur une sélection d'espèces appartenant à divers genres de Polynoidae, basée sur des caractères tels que la taille du spécimen, le nombre de segments, des caractéristiques des neuro- et notosoies, des cirres dorsaux et ventraux, ainsi que des ornementsations des palpes, des antennes et des élytres. Cette étude englobe principalement les Polynoidae et dans une moindre mesure les Aphroditiformia, mais ne propose pas de reconstruction phylogénétique. Avec ce travail, il suggère que la famille des Polyodontidae (aujourd'hui acceptée

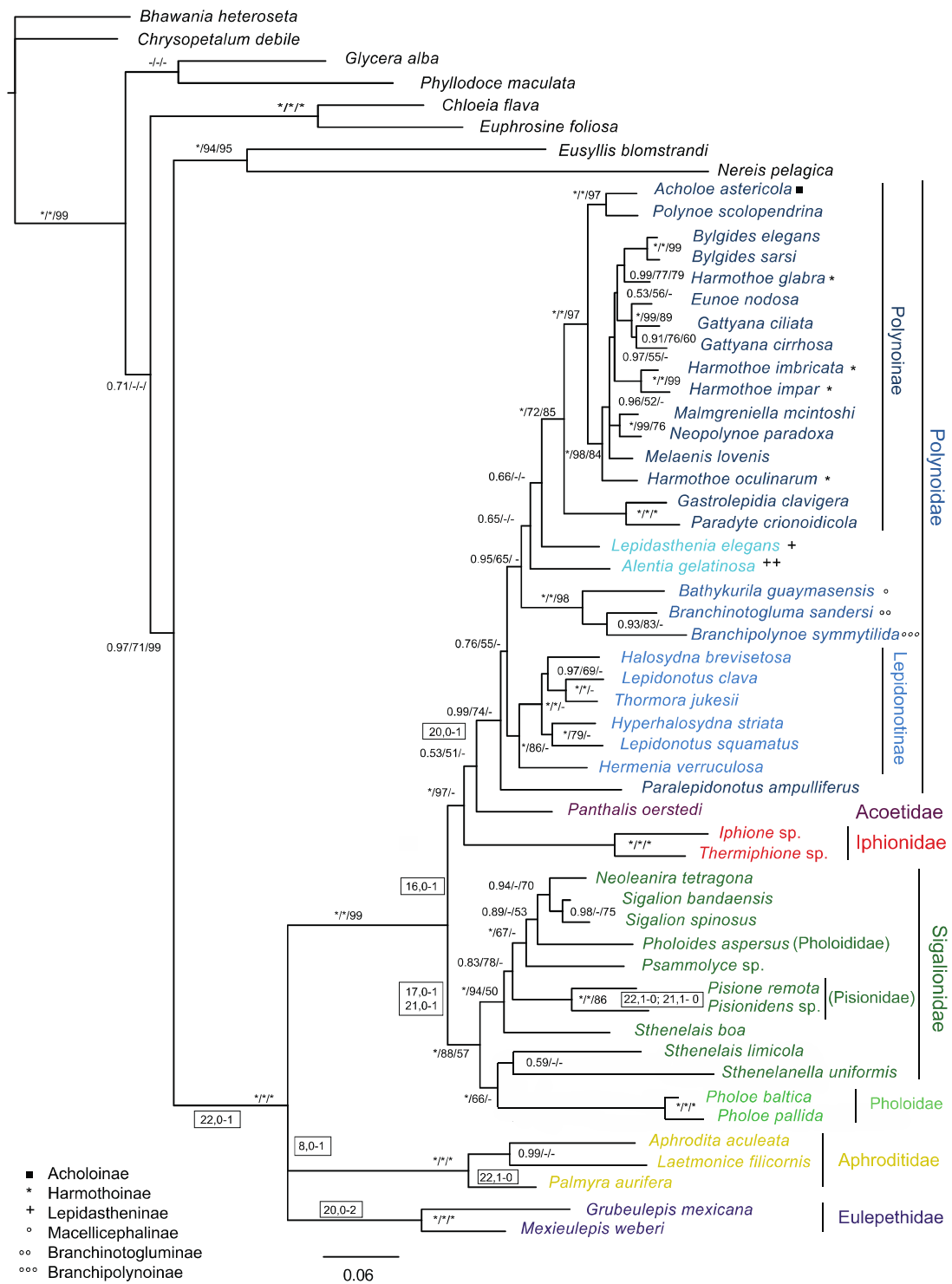


Figure II.4 Phylogénie des Aphroditiformia selon Norlinder *et al.* (2012). Les nombres aux nœuds correspondent aux probabilités postérieures pour l'analyse Bayésienne, bootstraps pour l'analyse de maximum de vraisemblance puis bootstraps pour l'analyse de maximum de parcimonie (effectuée sur les données morphologiques). Les nombres encadrés correspondent à des caractères morphologiques.

comme la famille des Acoetidae (Pettibone, 1989) n'est pas valide car sa définition est basée sur un unique caractère synapomorphique : la présence de glandes spinnigères dans les parapodes.

Après ces études initiales basées sur des caractères morphologiques, des auteurs ont commencé à utiliser des séquences d'ADN pour produire des phylogénies au niveau des Aphroditiformia. De ces analyses ressortent certaines observations intéressantes pour les Polynoidae. Avec des données moléculaires (marqueurs mtCOI et 18S) et morphologiques sur 14 espèces d'Aphroditiformia (5 familles), Wiklund *et al.* (2005) proposent une première phylogénie. Cette première étude confirme la monophylie des Polynoidae, mais comme elle ne comprend que cinq espèces de Polynoidae (appartenant à deux sous-familles) elle ne permet pas d'éclaircir les relations au sein de la famille. Dans une étude similaire, Norlinder *et al.* (2012) combinent des caractères moléculaires (18S, 28S, mtCOI, mt16S) et morphologiques, avec un échantillonnage permettant de mieux définir les limites de certaines familles et sous-familles (48 espèces d'Aphroditiformia, dont 30 Polynoidae représentant 7 sous-familles ; Norlinder *et al.*, (2012), voir figure II.4). Dans leur étude, comme présenté en figure II.4, les relations récentes (c'est à dire entre espèces d'un même genre ou d'une même sous-famille) sont généralement soutenues, ce qui n'est pas le cas pour la plupart des relations entre les sous-familles échantillonnées. Cependant, les marqueurs qu'ils utilisent leur permettent de résoudre des relations plus profondes, comme par exemple entre les familles. Leur échantillonnage leur permet de proposer la synonymie des sous-familles Harmothoinae (4 représentants), Acholoinae (1 représentant correspondant à l'espèce type et le genre type de la sous-famille) et Polynoinae (12 représentants) de manière soutenue. Toutefois, une espèce de Polynoinae (*Paralepidonotus ampulliferus*) ne se positionne pas avec les autres. Leur échantillonnage des Lepidonotinae (six espèces) permet également de dire que cette sous-famille est monophylétique. Par ailleurs, le faible échantillonnage au sein des Polynoidae ne leur permet pas de tester la monophylie des autres sous-familles (12 sous-familles manquantes dans leur échantillonnage). Cette étude inclut deux espèces hydrothermales (*Branchinotogluma sandersi* et *Branchipolynoe symmytilida*) qui appartiennent à des sous-familles différentes, et une espèce collectée sur une carcasse de baleine (*Bathykurila guaymasensis*). Comme ces données ne comprennent qu'une seule espèce par sous-famille (respectivement sous-familles Branchinotogluminae, Branchipolynoinae et Macellicephalinae), il leur est impossible de tester la monophylie de ces dernières. Ces trois espèces forment un clade* soutenu dans leur analyse et constitue le seul clade qui reconstruit des relations entre sous-familles de manière soutenue.

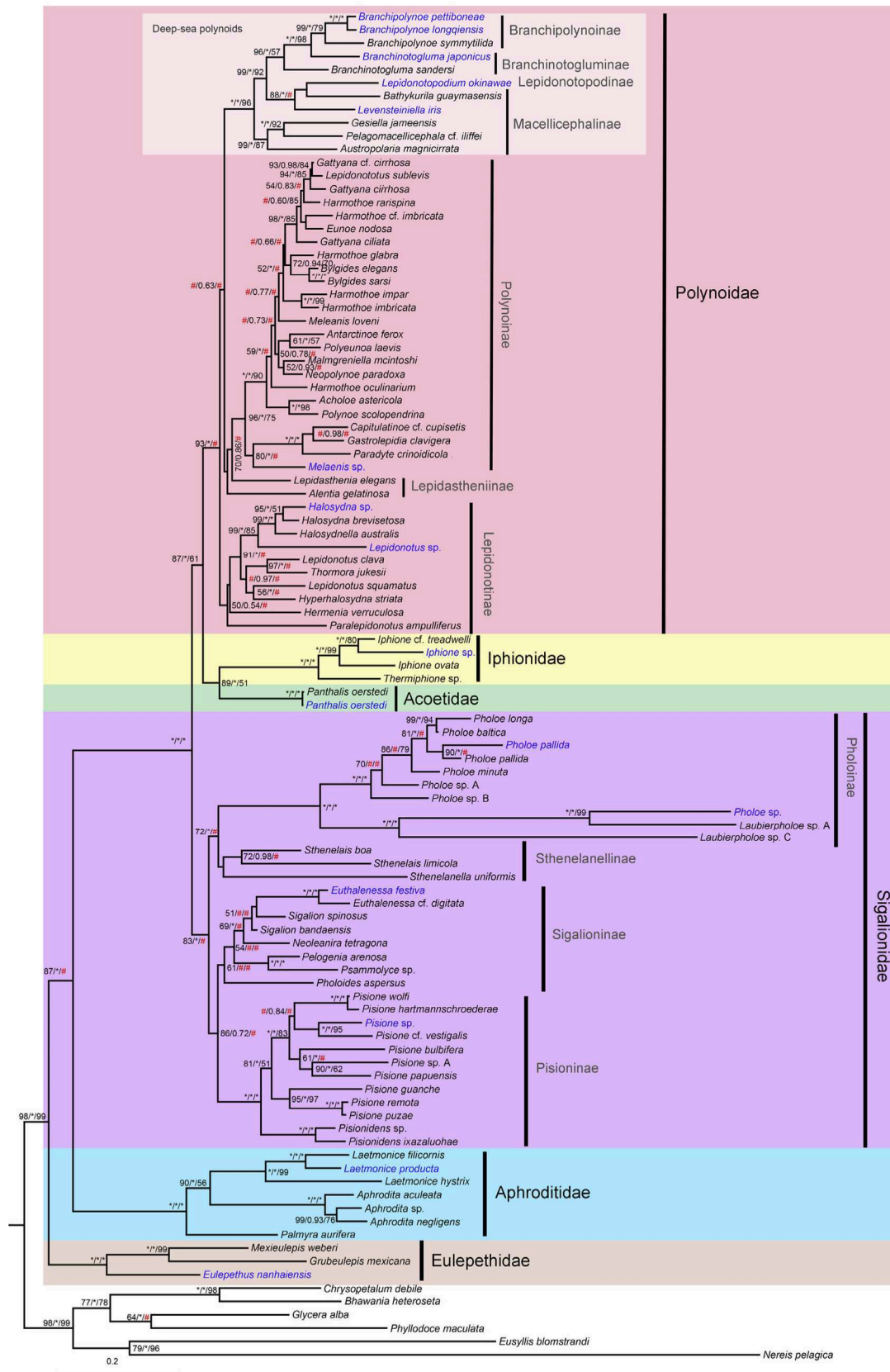


Figure II.5 Phylogénie des Aphroditiformia selon Zhang *et al.* (2018). Topologie ML obtenue à partir de la concaténation des gènes COX1, 16S, 18S et 28S rRNA. Les valeurs aux nœuds correspondent aux soutiens en Maximum de vraisemblance/Inférence bayésienne/Maximum de parcimonie. Les astérisques indiquent un soutien maximal. Les dièses indiquent des soutiens < 0,50 ou < 50.

Norlinder *et al.* (2012) montrent que les deux espèces d'Iphioninae incluses dans l'étude (*Iphione* sp. et *Thermiphione* sp.) forment un groupe monophylétique, frère du regroupement Polynoïdæ et Acoetidae. Les auteurs suggèrent en conséquence l'érection de la famille des Iphionidae. Cette suggestion est rapidement acceptée par la communauté des taxonomistes spécialistes des Aphroditiformia, étant donné que les espèces de cette lignée partagent des états de caractères morphologiques qui les distinguent clairement des autres Polynoïdæ, notamment l'absence d'une antenne médiane ou la présence de structures polygonales sur les élytres. Cependant, la famille des Acoetidae n'est représentée dans cette étude que par une seule espèce, *Panthalis oerstedii*, avec un soutien faible pour le nœud groupant les Acoetidae et les Polynoïdæ non Iphioninae (0,53 de probabilité bayésienne postérieure et 51% de valeur de bootstrap). L'inclusion d'un plus grand nombre d'espèces attribuées à la famille des Acoetidae pourrait remettre en cause sa monophylie et donc la validité du statut de famille pour les Iphionidae.

Gonzalez *et al.* (2017a) étendent l'échantillonnage de cette étude et proposent une phylogénie en se basant sur les mêmes marqueurs moléculaires (18S, 28S, mtCOI, mt16S) et sur la morphologie. Enfin, Zhang *et al.* (2018) proposent encore une extension de la phylogénie (figure II.5) de Gonzalez *et al.*, (2017a), en incluant 15 nouvelles espèces, dont deux *Branchipolynoe* (Branchipolynoïnae), une *Branchinotogluma* (Branchinotogluminae), une *Levensteiniella* (Macellicephalinae), et un *Lepidonotopodium* (Lepidonotopodinae, une sous-famille qui n'avait pas encore été représentée dans les précédentes études). Cette dernière étude a utilisé une approche « shot-gun » de séquençage massif pour reconstruire le génome mitochondrial et les ARN ribosomiaux 18S et 28S de ces 15 espèces. Les soutiens des nœuds sont hétérogènes le long de l'arbre. Par exemple, dans les sous-familles Lepidonotinae et Polynoïnae qui sont les sous-familles avec le plus de représentants, certaines relations les plus récentes sont soutenues (*e.g.* *Bylgides elegans* et *B. sarsi* ; *Acholoe astericola* et *Polyeunoa scolopendrina*; *Halosydna* sp. et *H. brevisetosa*), alors que des relations de profondeur similaires ne sont pas résolues (*e.g.* : *Harmothoe* cf. *imbricata* et *Eunoe nodosa* ; *Malmgreniella mcintoshii* et *Neopolynoe paradoxa*, figure II.5). Dans ces mêmes sous-familles, les relations de profondeur moyenne (entre genres) ne sont globalement pas résolues. Par contre, les marqueurs utilisés dans cette étude permettent de résoudre les relations à l'échelle des sous-familles. En effet, le nœud rassemblant les espèces de Polynoïnae est soutenu (96% en bootstrap et soutien maximal en bayésien) et le nœud des Lepidonotopodinae à l'exception de *Hermenia verruculosa* est soutenu (91% en ML et soutien maximal en bayésien). Par ailleurs, les relations entre espèces dans les Macellicephalinae et dans les Iphioninae (deux sous-familles qui ont au moins quatre représentants) sont toutes entièrement résolues (figure II.5). Cette étude (Zhang *et al.*, 2018) et celles de Gonzalez *et al.* (2017a) confirment les conclusions de Norlinder *et al.* (2012)

pour les Iphioninae/Iphionidae. Bien que le clade contenant les Acoetidae et Iphioninae/Iphionidae soit soutenu (hormis en ML pour Gonzalez *et al.*, 2017a), ce résultat est une fois encore basé sur un jeu de données n'incluant qu'une seule espèce d'Acoetidae (*P. oerstedii*). Certaines relations importantes pour notre étude se dégagent tout de même de l'arbre de Zhang *et al.*, 2018. Tout d'abord, les Iphioninae/Iphionidae utilisés comprennent une espèce hydrothermale (*Thermiphione* sp.) et trois espèces non hydrothermales (toutes du genre *Iphione*). Ces quatre espèces forment un groupe monophylétique bien distinct des autres Polynoidae. Cette lignée est clairement différente de celle comprenant les huit autres espèces hydrothermales présentes dans l'étude de Zhang *et al.* (2018) (figure II.5). Dans la seconde lignée hydrothermale, l'ensemble des relations sont soutenues à la fois dans les analyses en maximum de vraisemblance et en bayésien. Les trois des quatre espèces de Branchipolynoinae décrites forment un clade soutenu, confirmant la monophylie de cette sous-famille. A l'inverse, les deux espèces de *Branchinotogluma* se positionnent de manière paraphylétique* à la base des Branchipolynoinae, suggérant que la sous-famille Branchinotogluminae n'est pas monophylétique. Enfin, les deux espèces hydrothermales placées parmi les Macellicephalinae ne groupent pas avec les Macellicephalinae non hydrothermaux, mais forment un groupe monophylétique avec *Lepidonotopodium okinawae*, ce qui suggère également que la sous-famille Macellicephalinae ne forme pas un ensemble naturel (c'est-à-dire un groupe monophylétique). Les résultats de cette étude suggèrent donc au moins deux événements indépendants de colonisation du milieu hydrothermal pour les espèces utilisées.

Cependant, ces deux études plus récentes (Gonzalez *et al.*, 2017a ; Zhang *et al.*, 2018) qui traitent de la phylogénie des Aphroditiformia, ne permettent pas de résoudre les questions importantes à l'échelle des Polynoidae. En effet, l'échantillonnage ne comporte qu'une seule espèce pour certaines sous-familles ce qui ne permet pas de tester leur validité. De plus, l'échantillonnage de ces études ne couvre pas non plus l'ensemble des sous-familles hydrothermales. Aucune étude phylogénétique plus particulièrement focalisée sur une famille parmi les Aphroditiformia n'a à ce jour été proposée. Cela ne permet donc pas de savoir si les marqueurs classiquement utilisés dans les phylogénies d'Aphroditiformes permettent, dans le cas d'un échantillonnage taxonomique ciblé et représentatif de la famille de résoudre les relations phylogénétiques à ce niveau.

II.1.4 But de l'étude

Le but de notre étude est de comprendre l'origine et l'évolution des espèces (et donc par extension des sous-familles) hydrothermales. Ceci nous amène à nous demander (i) si ces sous-familles (établies sur des critères morphologiques) sont bien des groupes naturels ou au contraire ne sont que des regroupements basés sur des convergences morphologiques et écologiques et (ii) si ces sous-familles forment un seul groupe monophylétique (indiquant un événement de colonisation unique du milieu hydrothermal) ou bien si plusieurs lignées évolutives ont colonisé ce milieu de manière indépendante. Afin de tester la monophylie des sous-familles à ce jour connues dans le milieu hydrothermal, l'échantillonnage taxonomique a grandement été élargi par rapport aux phylogénies moléculaires précédentes. Notre étude inclut les sous-familles hydrothermales Branchinotogluminae, Branchipolynoinae, Branchiplicatinae et Lepidonotopodinae (il ne manque que la sous-famille monospécifique Vampiropolynoinae) et les espèces hydrothermales des sous-familles Iphioninae et Macellicephalinae. Afin de comprendre l'origine des colonisations du milieu hydrothermal et de savoir combien de fois un tel événement s'est produit, l'échantillonnage taxonomique a également été élargi parmi les milieux non hydrothermaux, en incluant une grande proportion d'espèces placées dans les autres sous-familles selon des critères morphologiques.

Dans ce chapitre la méthode mise en œuvre est l'approche multi-marqueurs (une combinaison de trois marqueurs nucléaires et deux marqueurs mitochondriaux) obtenus par amplification par PCR suivie d'un séquençage Sanger. Cette approche permet de tirer profit des études précédentes, tout en produisant de nouvelles données pour des espèces complémentaires pour le même jeu de marqueurs. De plus, bien que les couvertures d'échantillonnage des précédentes études soient assez faibles, ces marqueurs permettent *a priori* de résoudre les relations à différentes profondeurs : certaines relations récentes (dans une sous-famille de Polynoidae) et certaines relations plus profondes (entre sous-familles de Polynoidae et entre familles d'Aphroditiformes). L'extension de l'échantillonnage taxonomique se fonde sur l'échantillonnage de milieux moins accessibles et donc méconnus. Cet échantillonnage inclut donc potentiellement des espèces nouvelles dont l'identification taxonomique peut s'avérer difficile, même aux rangs supra-spécifiques. L'identification morphologique des spécimens est donc complétée par une identification moléculaire afin de sélectionner un échantillonnage qui couvre le mieux possible à la fois la hiérarchie de la classification acceptée et la diversité des milieux où les Polynoidae sont présents.

Tableau II.2 Campagnes de récolte pour les Polynoidae utilisés dans cette étude, région géographique, type d'environnement et profondeur échantillonnée lors de la campagne. Les étagements des profondeurs renseignés sont les suivants : sub-littoral (0-200 m), bathyal (200-2000m) et abyssal (2000-6000m). Un astérisque dénote une campagne issue du programme TDSB (Topical Deep Sea Benthos).

| Campagne | Région | Environnements | Profondeur |
|------------------|---------------------|--|---------------------------------|
| ANT-XXIX-3 | Antarctique | Abyssal sédimentaire | Sub-littoral / Bathyal /Abyssal |
| ATOS | Atlantique | Sources hydrothermales | Abyssal |
| BigMAR | Atlantique | Sources hydrothermales | Bathyal / Abyssal |
| BIOPAPUA * | Pacifique Sud-Ouest | Monts sous-marin/ plateau continental | Sub-littoral / Bathyal |
| EPR1999 | Pacifique Est | Sources hydrothermales | Abyssal |
| EPR2001 | Pacifique Est | Sources hydrothermales | Abyssal |
| EPR2002 | Pacifique Est | Sources hydrothermales | Abyssal |
| EXBODI * | Pacifique Sud-Ouest | Monts sous-marin/ plateau continental | Sub-littoral / Bathyal |
| GoM1998 | Golfe du Mexique | Suintements froids | Bathyal |
| GoM2001 | Golfe du Mexique | Suintements froids | Bathyal |
| GoM2003 | Golfe du Mexique | Suintements froids | Bathyal / Abyssal |
| GoM2006 | Golfe du Mexique | Suintements froids | Bathyal / Abyssal |
| JdF98 | Pacifique NE | Sources hydrothermales | Abyssal |
| KANACONO * | Pacifique Sud-Ouest | Monts sous-marin/ plateau continental | Sub-littoral / Bathyal |
| KREVET_2013 | Antarctique | Substrat meuble | Sub-littoral / Bathyal |
| Lau2005 | Pacifique Sud-Ouest | Sources hydrothermales | Bathyal / Abyssal |
| Lau2006 | Pacifique Sud-Ouest | Sources hydrothermales | Bathyal / Abyssal |
| Lau2009 | Pacifique Sud-Ouest | Sources hydrothermales | Bathyal / Abyssal |
| Madang * | Pacifique Ouest | côtier/ plaine / plateau continental | Sub-littoral / Bathyal |
| MESCAL | EPR | Sources hydrothermales | Abyssal |
| MESCAL2 | EPR | Sources hydrothermales | Abyssal |
| POLARIS_2016 | Antarctique | Substrat rocheux | Sub-littoral |
| REVOLTA | Antarctique | Substrat meuble | Sub-littoral / Bathyal |
| TARASOC * | Pacifique Sud | Monts sous-marin/ plateau continental | Sub-littoral / Bathyal |
| WACS | Atlantique | Suintements froids | Bathyal / Abyssal |
| Walter's Shoal * | Indien | Monts sous-marin/ plateau continental | Sub-littoral / Bathyal |

II.2 Matériel et méthodes

II.2.1 Sélection des taxons pour l'étude

II.2.1.1 *Critères géographique et environnemental : diversité des milieux échantillonnés*

La majorité des spécimens utilisés dans cette étude provient de campagnes océanographiques qui couvrent des zones géographiques et des environnements variés (tableau II.2). Parmi elles : 16 campagnes ont été menées sur des environnements profonds chimiosynthétiques (suintements froids ou sources hydrothermales) ; 7 autres campagnes ont été menées à des profondeurs bathyales voire abyssales (au-delà de 2000 m de profondeur). Parmi celles-ci, plus de la moitié sont issues du programme Tropical Deep-Sea Benthos (TDSB), mené par le MNHN (Muséum National d'Histoire Naturelle) et l'IRD (Institut pour la Recherche et le Développement), et se déroulent principalement dans des habitats bathyaux variés de la zone tropicale Indo-Pacifique (Samadi *et al.*, 2010 ; Pante *et al.*, 2012). Ces campagnes ont pour but d'échantillonner la faune de milieux divers, difficiles d'accès et peu explorés pour en documenter la biodiversité. L'échantillonnage des milieux profonds a été complété par des espèces provenant d'environnements moins profonds (côtier à bathyal supérieur). Des échantillons de trois campagnes menées en Antarctique en milieux côtier à bathyal supérieur (entre 1 m et 300 m de fond) ont été inclus. D'autre part, certaines espèces ont été récoltées sur le littoral tempéré en Europe et aux Etats Unis. Les trois espèces d'*Arctonoe* (*A. fragilis*, *A. pulchra*, et *A. vittata*) ont été récoltées près de Friday Harbor Marine Lab (Etat de Washington, USA) et nous ont été fournies par Bruno Pernet (Washington State University, USA) qui les a identifiées morphologiquement. *Laetmonice hystrix*, *Acholoe astericola*, *Alentia gelatinosa*, *Harmothoe clavigera*, *Lepidonotus clava*, *Malmgrenia ljunghmani*, *Malmgrenia sp.*, *Pettibonesia furcosetosa* et *Subadyte pellucida* ont été récoltées sur les côtes du Finistère (Roscoff ou Brest). Les échantillons des campagnes TDSB et des campagnes Antarctiques KREVET 2013 et ANT-XXIX-3 et REVOLTA ont été récoltés soit par dragage (drague Warén), soit avec un chalut à perche. Les échantillons de la campagne POLARIS 2016 (Terre Adélie), ainsi que certains de Walter's Shoal ont été récoltés en plongée. L'ensemble des échantillons réunis ont été fixés juste après la récolte dans de l'éthanol à 85%.

Tableau II.3 Amorces utilisées pour les amplifications par PCR de l'étude des Polynoidae. Pour chaque amorce utilisée dans cette étude, sont renseignés : le marqueur ciblé, le nom de l'amorce, la séquence orientée (5'-3') de l'amorce, la référence de l'amorce, la taille du marqueur amplifié, les températures et les durées en secondes de chaque étape des cycles d'amplification par PCR.

| Marqueur | Amorce | Séquence (5'-3') | Sens | Référence | Taille amplifiée (pb) | Températures des étapes des cycles (°C) | Temps (s) |
|----------|-----------|----------------------------------|------|------------------------------|-----------------------|---|--------------|
| COI | LCO1490 | GGTCAACAAATCATAAAGATATTGG | F | Folmer <i>et al.</i> , 1994 | 658 | 94 ; 50 ; 72 | 30 ; 40 ; 60 |
| COI | HCO2198 | TAAACTTCAGGGTGACCAAAAAATCA | R | Folmer <i>et al.</i> , 1994 | | | |
| 16S | 16Sar | CGCCTGTTTAACAAAAACAT | F | Palumbi, 1996 | 486 | 94 ; 52 ; 72 | 30 ; 40 ; 60 |
| 16S | 16Sbr | CCGGTCTGAACTCAGATCACGT | R | Palumbi, 1996 | | | |
| 16S | 16SAnnF | GCGGTATCCTGACCGTRCWAAGGTA | F | Sjölin <i>et al.</i> , 2005 | 322 | 94 ; 52 ; 72 | 30 ; 40 ; 60 |
| 16S | 16SAnnR | TCCTAAGCCAACATCGAGGTGCCAA | R | Sjölin <i>et al.</i> , 2005 | | | |
| 18S | 18SF | GCCAGTAGCATATGCTTGCTC | F | Holland <i>et al.</i> , 1991 | 522 | 94 ; 50 ; 72 | 30 ; 40 ; 60 |
| 18S | 18SR | AGACTTGCCTCCAATGGATCC | R | Holland <i>et al.</i> , 1991 | | | |
| 28S | 28SD1F | ACCCSCTGAAYTTAAGCAT | F | McArthur et Koop, 1999 | 340 | 94 ; 52 ; 72 | 30 ; 40 ; 60 |
| 28S | 28SD1R | AACTCTCTCMTTCARAGTTC | R | Colgan <i>et al.</i> , 2003b | | | |
| 28S | 28SD9-10F | AAGGTAGCCAAATGCCTCATC | F | Hillis et Dixon, 1991 | 598 | 94 ; 52 ; 72 | 30 ; 40 ; 60 |
| 28S | 28SD9-10R | GTGAATTCTGCTTCATCAATGTAGGAAGAGCC | R | Hillis et Dixon, 1991 | | | |

II.2.1.2 Identification et sélection des spécimens récoltés

Stéphane Hourdez a divisé en morpho-espèces les spécimens récoltés lors des campagnes listées dans le tableau II.2 et attribués à la famille des Polynoïdae. A l'aide des clefs de détermination morphologique disponibles, il a ensuite identifié ces morpho-espèces au plus bas rang taxonomique possible. Toutes les morpho-espèces présentes sur des sites hydrothermaux sont incluses dans l'analyse phylogénétique. Pour les autres environnements, la sélection des spécimens visait à obtenir les plus grandes couvertures taxonomique et écologique possibles. L'objectif a donc été d'inclure le plus grand nombre de sous-familles et de genres, tout en évitant de biaiser la représentation des genres les plus diversifiés (ou les mieux échantillonnés). La sélection a donc été faite en limitant la redondance pour les morpho-espèces les plus représentées. Dans le cas de la sous-famille des Polynoïnae, la diversité spécifique est particulièrement importante et les phylogénies publiées à ce jour (Gonzalez *et al.*, 2017a ; Zhang *et al.*, 2018) suggèrent de la polyphylie* parmi les genres. Pour les espèces de Polynoïnae profondes, une attention particulière a été portée afin d'identifier plus précisément leur habitat et notamment les associations avec d'autres organismes (gorgone, éponges, etc....) ou le substrat (notamment bois coulés). Pour l'ensemble des morpho-espèces ainsi sélectionnées, jusqu'à trois spécimens ont été utilisés pour en extraire l'ADN afin de confirmer par un critère génétique la délimitation morphologique initiale.

II.2.2 Obtention et sélection des séquences moléculaire

II.2.2.1 Extraction de l'ADN et amplification des marqueurs

Les extractions des ADN totaux ont été effectuées en utilisant le protocole CTAB (CetylTrimethylAmmonium Bromide) avec un ajout de PVPP (PolyVinylPolyPyrrolidone) (Doyle et Dolye, 1987). Le CTAB est un détergent qui facilite la séparation des polysaccharides qui peuvent être retrouvés dans le mucus des annélides. L'ajout de PVPP permet d'éliminer les polyphénols, potentiellement inhibiteurs de PCR (Koonjul *et al.*, 1999).

Les amplifications PCR ont été effectuées dans un volume final de 25 µL et contiennent une concentration finale de tampon de réaction de 1X, 2 mM de MgCl₂, 0,2 µM d'amorces, 0,05 mM de dNTPs et 0,5 unités de polymérase d'ADN (UptiTherm, Interchim). Les cycles d'amplification consistent en une étape de dénaturation initiale de 3 minutes à 94°C, suivie par 35 cycles de dénaturation à 94 °C, hybridation et extension à 72°C puis terminent avec 5 minutes d'élongation finale à 72°C. Les températures d'hybridation et les cycles d'amplification pour les différents fragments amplifiés sont indiqués dans le tableau II.3. L'amplification du marqueur COI n'étant pas

toujours facile, un deuxième profile de cycle de PCR a été utilisé, qui comprend cinq cycles préliminaires avec une température d'hybridation des amorces plus basse (47°C).

Après amplification, les produits ont été visualisés sur un gel agarose à 1,5% par fluorescence sous éclairage UV. Le séquençage dans les deux directions des produits positifs par la méthode Sanger a été fait par la compagnie EUROFINS. Les données de séquençage ont été visualisées et les séquences des amorces ont été éliminées manuellement avec le logiciel Codon-Code Aligner v.7.0.1.

Plusieurs séquences de Polynoidae (sous-famille des Macellicephalinae et Polaruschakovinae) proviennent d'articles en cours de publication ont été obtenues par le biais d'une collaboration avec Paulo Bonifacio et Lenaïck Menot (Laboratoire Environnement Profond, Ifremer Brest), qui étudient la biodiversité des Polynoidae associés aux nodules polymétalliques dans le Pacifique Oriental.

II.2.2.2 Gènes sélectionnés

Les marqueurs COI et 16S permettent de résoudre les relations récentes (entre espèces d'un même genre ou entre des genres proches). Pour inférer les relations phylogénétiques plus profondes, il est généralement nécessaire d'utiliser plus de marqueurs afin d'améliorer le signal phylogénétique et de cibler des marqueurs plus conservés pour éviter la saturation du signal. De plus, concaténer plusieurs marqueurs indépendants, provenant à la fois des génomes mitochondriaux et nucléaires (ou éloignés dans le génome nucléaire) permet de plus fidèlement reconstruire la phylogénie des espèces et non seulement la phylogénie des gènes. Les marqueurs moléculaires choisis pour replacer les espèces hydrothermales de Polynoidae par rapport aux autres membres de la famille correspondent aux fragments pour lesquels des amorces universelles (qui peuvent s'hybrider sur un grand nombre de taxons) sont disponibles. De plus, nous avons ciblé les marqueurs déjà utilisés dans les études antérieures pour pouvoir bénéficier des données déjà acquises et disponibles sur les bases de données. En plus des fragments des gènes mitochondriaux COI et 16S, nous avons donc utilisé deux marqueurs nucléaires : un fragment du gène codant pour la sous-unité ribosomale 18S et deux fragments du gène codant pour la sous-unité ribosomale 28S (28SD1 et 28SD9-10).

II.2.2.3 Séquences provenant des bases de données et réduction du jeu de données

Afin de couvrir la plus grande diversité taxonomique possible, parmi les séquences de Polynoidae disponibles sur NCBI en août 2017, celles correspondant à des spécimens pour lesquels au moins deux des marqueurs sélectionnés étaient disponibles ont été téléchargées. Pour vérifier cette condition nous avons en priorité sélectionné des spécimens ayant un numéro de voucher*, qui

peuvent donc être considérés comme des hologénophores (tous les gènes séquencés proviennent du même spécimen ; Pleijel *et al.*, 2008), ou à défaut des spécimens identifiés par un numéro d'isolat.

Pour l'ensemble des spécimens sélectionnés dans l'échantillonnage des campagnes océanographiques, l'ensemble des gènes a été amplifié. Dans un premier temps, toutes les séquences acquises (téléchargées et produites) ont été vérifiées gène à gène afin de détecter les éventuelles contaminations. Pour ce faire, les alignements ont été effectués en utilisant l'algorithme MUSCLE (Edgar, 2004) avant d'être ajustés à la main, en particulier pour les jeux de données avec d'importantes insertions/délétions comme les gènes d'ARN ribosomiaux (16S, 18S, 28S). Puis, chaque gène a été analysé en construisant des arbres rapides grâce au logiciel MEGA v.5.05 (Tamura *et al.*, 2011) par la méthode de Neighbor-Joining sur des distances Kimura-2-paramètres (K2P). Lorsque des séquences sont identiques pour un marqueur alors qu'elles proviennent de spécimens attribués morphologiquement à des taxons distincts et pour lesquels des séquences différentes ont bien été obtenues pour les autres marqueurs, ces séquences sont considérées comme contaminées. Dans ce cas, les séquences pour ce marqueur ont été supprimées du jeu de données et/ou re-séquencées lorsque cela était possible.

Dans un second temps, comme il n'est pas nécessaire de conserver plusieurs spécimens appartenant à la même espèce pour reconstruire la phylogénie des Polynoidae, un seul spécimen par espèce, ou morpho-espèce, a été sélectionné parmi l'ensemble des données réunies (téléchargées et séquencées). Depuis 2003, le séquençage d'une partie du marqueur COI (ou COX1) comme «code barre» du vivant (Hebert *et al.*, 2003) s'est beaucoup développé (Ravara *et al.*, 2017). La séquence proposée comme «code barre» du vivant correspond aux 658 paires de bases (pb) situées à l'extrémité 5' du gène codant pour la Cytochrome Oxydase I (COI), amplifiée par le couple d'amorces issu de Folmer *et al.* (1994). Les spécimens ayant à la fois moins de 3% de divergence nucléotidique pour le marqueur COI (taxonomie moléculaire) et une identification morphologique identique sont considérés comme des spécimens de la même espèce. Pour certaines espèces, les séquences de COI sont particulièrement difficiles à amplifier et le jeu de données du gène 16S est le plus complet. De plus, bien que ce gène évolue moins rapidement que le COI (Chevaldonné *et al.*, 1998), il fournit un signal permettant de résoudre les relations phylogénétiques récentes (entre les espèces d'un même genre). Il a donc également été utilisé pour identifier moléculairement les spécimens appartenant à une même espèce. Les spécimens ayant à la fois moins de 1% de divergence nucléotidique pour le marqueur 16S et une identification morphologique identique sont considérés comme des spécimens de la même espèce. Les spécimens ayant ces deux marqueurs doivent avoir des distances inférieures aux seuils à la fois en COI et en 16S pour être considérés de la même espèce. Dans ce cas, le spécimen pour lequel le plus grand nombre de marqueurs séquencés a pu être obtenu a été

conservé dans la matrice finale. Les séquences associées à des vouchers ont été, autant que possible, retenues.

II.2.3 Analyses phylogénétiques

II.2.3.1 Echantillonnage taxonomique final et concaténation des marqueurs phylogénétiques :

Chaque gène a ensuite de nouveau été analysé en construisant des arbres rapides (NJ sur des distances K2P) afin de vérifier la congruence des topologies construites entre les différents marqueurs. En cas d'absence d'incongruence soutenue entre les marqueurs, l'analyse phylogénétique se fait en concaténant les alignements de tous les marqueurs phylogénétiques, afin d'obtenir une super-matrice. Pour l'analyse concaténée, nous avons gardé uniquement les spécimens vérifiant une des conditions suivantes :

- Au moins trois marqueurs ont été séquencés
- Si seulement deux marqueurs étaient disponibles, au moins un des deux était le COI ou le 16S qui ont servi à l'identification moléculaire

II.2.3.2 Reconstructions phylogénétiques

Les analyses phylogénétiques des séquences nucléiques ont été effectuées à la fois en utilisant les méthodes du Maximum de vraisemblance (ML) à l'aide du logiciel RAxML version 8.2.9 (Stamatakis *et al.*, 2014) et de l'Inférence Bayésienne (IB) avec MrBayes v3.2.6 (Ronquist *et al.*, 2012). Le modèle d'évolution GTR + Γ + I a été utilisé pour chaque partition et pour chaque position de codon (pour le COI) dans les deux analyses. Le modèle GTR (Generalised Time-Reversible) (Tavaré, 1986) est le modèle de substitution le plus neutre. Il implique que la fréquence de chacune des quatre bases A, T, C et G soit indépendante des autres et que les taux de substitution entre chaque paire de nucléotides soient différents. Ce modèle nécessite donc quatre paramètres de fréquences des bases et six taux de substitutions différents. Ce modèle est combiné à une distribution des taux de mutation des sites qui suit une loi gamma (Γ) (fixée à quatre catégories) et une proportion de sites invariables (I). Pour l'analyse ML, la robustesse des nœuds est estimée par la méthode du bootstrap avec 1 000 itérations. L'arbre est visualisé à l'aide du logiciel FigTree v1.4.3 (Rambaut et Drummond, 2012). L'analyse bayésienne a été effectuée sur le portail en ligne CIPRES (Miller *et al.*, 2010).

L'exploration de l'espace bayésien a été effectuée avec deux ensembles de chaînes de Markov Monte Carlo (MCMC) à partir de points de départ différents, sur $15 \cdot 10^6$ générations, avec une fréquence d'échantillonnage des arbres tous les 1 000 pas. La stabilité de chaque paramètre a été contrôlée à l'aide du logiciel Tracer 1.6 (Rambaut et Drummond, 2007), en vérifiant les valeurs d'« Estimate Sample Size » (ESS). L'ESS estime la quantité d'échantillons indépendants qui ont la valeur indiquée pour ce paramètre. Nous considérons la stabilité du paramètre quand ESS est supérieur à 200. La convergence des analyses a été vérifiée avec les valeurs de PSRF (Potential Scale Reduction Factor, proches de 1 si les analyses ont convergé). L'arbre consensus a été reconstruit en éliminant les premières 25% de générations échantillonnées comme un 'burn-in', et est là aussi visualisé à l'aide du logiciel FigTree v.1.4.3. Nous retiendrons comme significativement soutenus les nœuds présentant des valeurs d'au moins 80% de bootstrap et 0,95 de probabilité postérieure (pp).

Tableau II.4 Informations relatives à l'échantillonnage des genres et des espèces pour chaque sous-famille de Polynoïdæ dans notre étude. La première colonne indique le nombre de genres/espèces en commun entre NCBI et nos données de séquençage ; les deuxième et troisième colonnes indiquent le nombre de genres/espèces présents respectivement exclusivement (Excl.) dans NCBI et dans nos données de séquençage ; la quatrième colonne indique le nombre total de genres/espèces présents dans notre étude, par rapport au nombre total renseigné dans WoRMS. Les sous-familles sont ordonnées selon la profondeur à laquelle les espèces sont retrouvées. La dernière ligne comporte les informations relatives aux genres/espèces qui ne sont pas attribuées à une sous-famille en particulier. Un astérisque indique une sous-famille dans laquelle sont placées des espèces provenant des milieux hydrothermaux et des espèces provenant d'autres environnements ; Deux astérisques indiquent une sous-famille dans laquelle ne sont placées que des espèces hydrothermales.

| Etage | Sous-famille | Genres | | | | Espèces | | | |
|--------------------------------|-----------------------|-----------|-----------|------------------|---------|-----------|-----------|------------------|----------|
| | | En commun | Excl NCBI | Excl nos données | Total | En commun | Excl NCBI | Excl nos données | Total |
| Sub-littoral | Arctonoïnae | 1 | 2 | 0 | 3 / 8 | 2 | 2 | 1 | 5 / 18 |
| Sub-littoral | Gesiellinae | 0 | 1 | 0 | 1 / 1 | 0 | 1 | 0 | 1 / 1 |
| Sub-littoral | Lepidastheniinae | 0 | 1 | 0 | 1 / 4 | 0 | 1 | 0 | 1 / 51 |
| Sub-littoral | Lepidonotinae | 2 | 4 | 1 | 7 / 22 | 1 | 6 | 5 | 12 / 169 |
| Sub-littoral | Uncopolynoïnae | 0 | 0 | 0 | 0 / 1 | 0 | 0 | 0 | 0 / 1 |
| Bathyal | Admetellinae | 0 | 0 | 0 | 0 / 2 | 0 | 0 | 0 | 0 / 4 |
| Sub-littoral, Bathyal, Abyssal | Eulagiscinae | 0 | 1 | 2 | 3 / 2 | 0 | 1 | 4 | 5 / 6 |
| Sub-littoral, Bathyal, Abyssal | Iphioninae* | 2 | 0 | 1 | 3 / 4 | 0 | 2 | 7 | 9 / 13 |
| Sub-littoral, Bathyal, Abyssal | Macellicephalinae* | 1 | 9 | 1 | 11 / 16 | 1 | 14 | 4 | 19 / 35 |
| Sub-littoral, Bathyal, Abyssal | Polynoïnae | 5 | 11 | 8 | 24 / 48 | 4 | 27 | 18 | 49 / 438 |
| Bathyal, Abyssal | Branchinotogluminae* | 1 | 0 | 1 | 2 / 3 | 0 | 1 | 13 | 14 / 9 |
| Bathyal, Abyssal | Branchiplicatinae** | 0 | 0 | 1 | 1 / 1 | 0 | 0 | 1 | 1 / 1 |
| Bathyal, Abyssal | Branchipolynoïnae* | 0 | 0 | 1 | 1 / 1 | 0 | 0 | 6 | 6 / 3 |
| Bathyal, Abyssal | Lepidonotopodinae** | 0 | 0 | 2 | 2 / 2 | 0 | 0 | 6 | 6 / 9 |
| Bathyal, Abyssal | Macelloïdinae | 0 | 0 | 0 | 0 / 1 | 0 | 0 | 0 | 0 / 1 |
| Bathyal, Abyssal | Vampiropolynoïnae** | 0 | 0 | 0 | 0 / 1 | 0 | 0 | 0 | 0 / 1 |
| Abyssal | Bathyedithinae | 0 | 0 | 0 | 0 / 2 | 0 | 0 | 0 | 0 / 3 |
| Abyssal | Bathymacellinae | 0 | 0 | 0 | 0 / 1 | 0 | 0 | 0 | 0 / 1 |
| Abyssal | Polaruschakovinae | 0 | 1 | 0 | 1 / 2 | 0 | 2 | 0 | 2 / 3 |
| Abyssal, Hadal | Macellicephaloidinae | 0 | 0 | 0 | 0 / 9 | 0 | 0 | 0 | 0 / 9 |
| | <i>incertae sedis</i> | 1 | 1 | 3 | 5 / 56 | 0 | 3 | 5 | 8 / 121 |

II.3 Résultats - Discussion

II.3.1 Echantillonnage final

Les taux de réussite des amplifications par PCR n'ont pas été les mêmes pour les cinq fragments pour l'ensemble des taxons. En tout, l'ADN de 240 spécimens a été extrait, dont 27 pour lesquels l'extraction n'a finalement permis d'amplifier aucun gène. Onze spécimens appartiennent aux familles utilisées comme groupes externes (Acoetidae, Aphroditidae et Sigalionidae). A partir des 202 autres spécimens extraits, nous avons obtenu 147 séquences de 16S, 139 séquences de 18S, 98 séquences de COI, 130 séquences de 28SD1, et 107 séquences de 28SD9-10 (dont 98 spécimens ayant les 2 fragments de 28S). Pour 38 spécimens (19% des 202 spécimens de Polynoidae extraits), ni le fragment de COI ni le fragment de 16S n'ont pu être amplifiés, aussi ces spécimens n'ont pas été gardés pour les analyses. Seuls 49 spécimens (24%) ont pu être séquencés pour les 5 fragments ciblés. Les séquences de 982 Polynoidae ont été récupérées sur la base de données NCBI. Parmi ces derniers, 717 (73%) n'ont qu'une séquence de COI et 70 (7%) n'ont ni de séquence pour le fragment COI ni pour le fragment de 16S et n'ont donc pas été gardés pour les analyses. Après réduction du jeu de données, l'échantillonnage final comporte 87 nouveaux spécimens de Polynoidae provenant de notre étude et 92 spécimens provenant de NCBI ou des données fournies par Bonifacio et Menot. Parmi ces 179 spécimens, 5 sont attribués à des espèces avec les mentions *affinis* ou *confere* et 49 ne sont pas attribués à une espèce nominale (parmi eux, 13 ne sont attribués à aucun genre). Les 125 spécimens restant sont attribués à 96 espèces nominales et 59 genres nominaux différents (voir Annexe 1). Nos données de séquençage ont permis d'apporter 21 genres (dont 17 nommés) et 70 espèces (dont 42 avec un nom d'espèce) qui n'étaient pas présents dans la base de données de NCBI. Le jeu de données réuni dans notre étude couvre 13 des 20 sous-familles de Polynoidae (voir tableau II.4). Parmi les sept sous-familles manquantes se trouvent notamment les sous-familles abyssales Bathyedithinae, Bathymacellinae et Macelloidinae, la sous-famille abysso-hadale Macellicephaloidinae et la sous-famille monospécifique hydrothermale Vampiropolynoinae. Le tableau II.4 montre la répartition taxonomique parmi les sous-familles de Polynoidae couvertes par notre échantillonnage. Pour certaines sous-familles, le nombre d'espèces est plus grand que le nombre répertorié dans WoRMS, ce qui reflète la présence d'espèces potentiellement nouvelles dans nos données, dont certaines sont en cours de description (notamment dans les sous-familles Branchinotogluminae et Branchipolynoinae, *com. pers. Stéphane Hourdez* et dans la sous-famille Eulagiscinae, *com. pers. Paulo Bonifacio et Stéphane Hourdez*). Le tableau II.4 montre que la plupart des genres et des espèces provenant des milieux peu profonds sub-littoraux ont principalement été apportés par le jeu de données tiré de NCBI. Ceci est cohérent avec le fait que les milieux côtiers, plus

accessibles que les milieux profonds, aient en général été plus échantillonnés dans les autres études que la nôtre, pour laquelle nous avons échantillonné principalement des milieux profonds, difficiles d'accès.

Les sous-familles Polynoïnae et Macellicephalinae sont très représentées dans NCBI et sont les deux seules sous-familles avec des espèces profondes pour lesquelles l'apport de NCBI est plus important que celui de nos données de séquençage. Dans le cas des Macellicephalinae, ceci s'explique par l'apport de données moléculaires d'une étude focalisée sur les Macellicephalinae retrouvés sur les nodules profonds (Bonifacio et Menot, *in press*). Les données de NCBI nous permettent d'augmenter le nombre de sous-familles présentes dans notre phylogénie en apportant des séquences pour la sous-famille profonde Polaruschakovinae et les sous-familles sub-littorales Gesiellinae et Lepidastheniinae. Dans l'ensemble dans notre étude, les sous-familles profondes ont une meilleure couverture taxonomique que les sous-familles peu profondes, en grande partie car la très grande majorité des espèces décrites sont peu profondes et que la diversité profonde est certainement sous-estimée. Parmi les 13 sous-familles qui ne sont pas monogénériques, 8 sont représentées par au moins 2 genres dans notre phylogénie, permettant ainsi de tester leur monophylie. Parmi les sous-familles monogénériques, seule Branchipolynoïnae n'est pas monospécifique et est représentée par plusieurs espèces (ici cinq), permettant également de tester la monophylie de celle-ci.

Notre échantillonnage a permis d'obtenir une très bonne couverture dans les sous-familles des milieux chimiosynthétiques : 83% des genres et 80% des espèces décrites sont présents dans notre jeu de données. Plus de deux espèces par genre de Branchinotogluminae et Branchipolynoïnae sont présentes dans notre étude, permettant de tester la monophylie de ces deux sous-familles et des genres qui les composent. Pour la sous-famille Lepidonotopodinae, notre étude couvre les deux genres et six des neuf espèces décrites, permettant de vérifier la monophylie de la sous-famille et du genre *Lepidonotopodium*, qui n'est pas monospécifique. Dans le cas de la sous-famille Iphioninae, notre échantillonnage permet de couvrir trois genres sur quatre et environ la moitié des espèces placées dans cette sous-famille. Nous apportons de nouvelles espèces hydrothermales mais aussi de bois coulés et de milieu profond non chimiosynthétique. Cependant, pour pouvoir établir des hypothèses sur la colonisation du milieu hydrothermal, il est important d'échantillonner également les espèces provenant d'autres milieux.

Si le signal phylogénétique des marqueurs est suffisant, l'échantillonnage taxonomique réuni dans cette étude permet donc :

- de tester la monophylie de 9 sous-familles parmi les 13 présentes dans notre étude
- d'estimer le nombre de lignées ayant colonisé le milieu profond
- d'estimer le nombre de lignées ayant colonisé le milieu hydrothermal et de déterminer leur(s) groupe(s) frère(s)

II.3.2 Analyses phylogénétiques

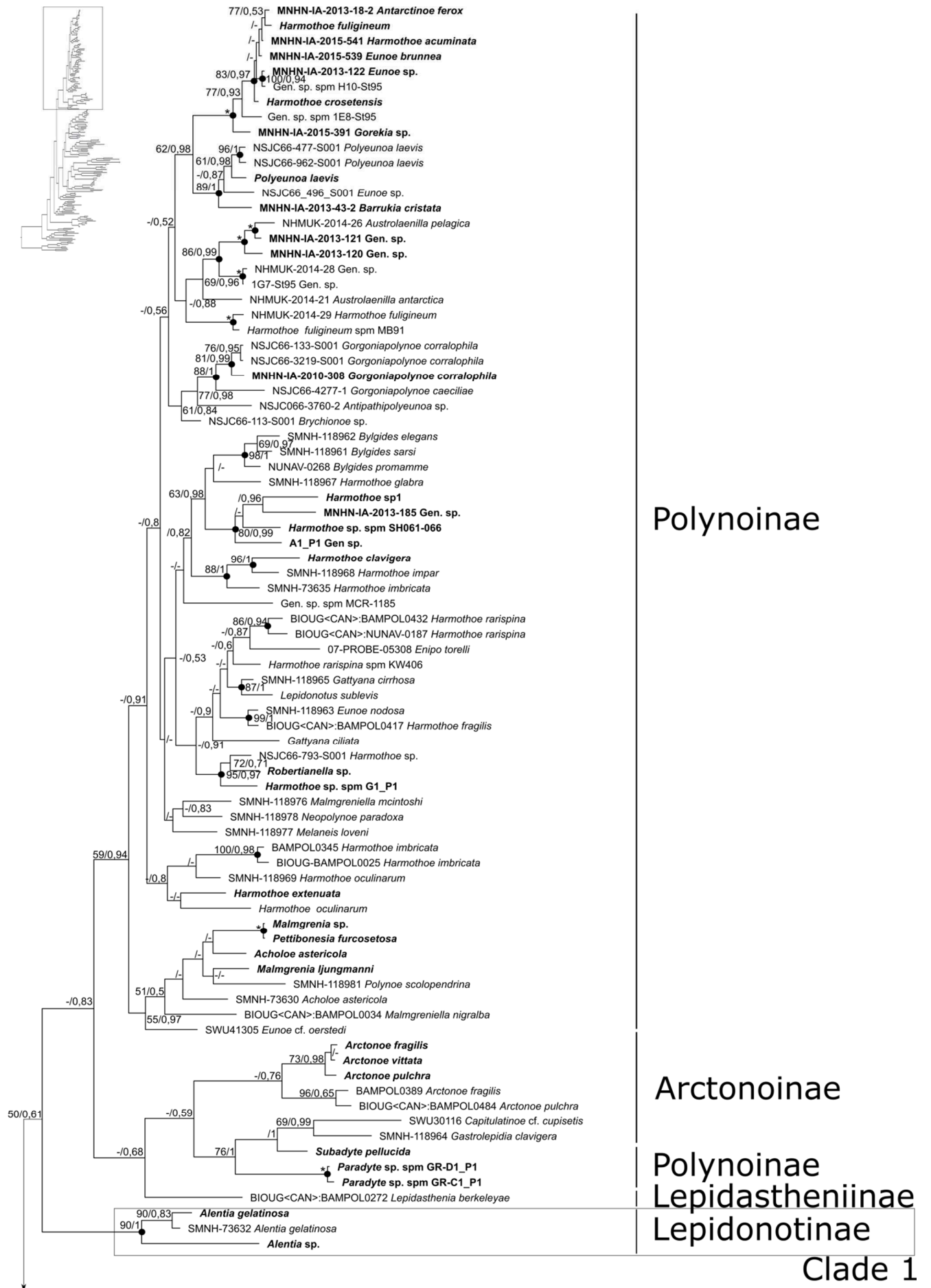
II.3.2.1 *Détails du jeu de données*

Le jeu de données concaténé des marqueurs 16S, COI, 18S, 28SD1 et 28SD9-10 a une longueur de 2824 pb dont 1583 positions sont conservées et 867 sont variables et informatives en parcimonie. L'alignement contient environ 33% de données manquantes. Les topologies obtenues en ML et en IB sont similaires (figure II.6) et les marqueurs utilisés permettent de résoudre les relations à différentes profondeurs phylogénétiques. Dans l'ensemble, la topologie de l'arbre est cohérente avec celles obtenues par Norlinder *et al.* (2012), Gonzalez *et al.* (2017a) et Zhang *et al.* (2018) sur des échantillonnages taxonomiques plus réduits.

II.3.2.2 *Relations phylogénétiques profondes et limites des sous-familles*

Parmi les Polynoidae de notre phylogénie, cinq grands clades sont soutenus dans les deux analyses (au moins 80% de bootstrap et 0,95 de probabilité postérieure (pp)) qui correspondent à différents rangs taxonomiques.

Le clade 1 (soutiens de 90% en ML et 1 pp en IB) est composé de trois spécimens attribués au genre *Alentia* et appartenant à deux morpho-espèces différentes. Ce genre est morphologiquement placé dans les Lepidonotinae, dont les autres représentants ne sont pas positionnés de façon soutenue dans cet arbre (respectivement 24% et 0,61 pp). Le genre type de cette sous-famille (*Lepidonotus*) n'étant pas monophylétique dans notre arbre, ce clade n'apporte pas d'information suffisante pour réviser la définition de la sous-famille Lepidonotinae. Le clade 2 (93%, 1 pp) regroupe les espèces des sous-familles Branchipolynoinae et Branchinotogluminae. Au sein de ce clade, les espèces du genre *Branchipolynoe* (seul genre attribué à la sous-famille Branchipolynoinae) forment un groupe monophylétique soutenu avec les deux spécimens attribués à *Branchinotogluma marianus* (87% et 1 pp). Les autres spécimens regroupés dans le clade 2 sont attribués aux genres *Branchinotogluma* et



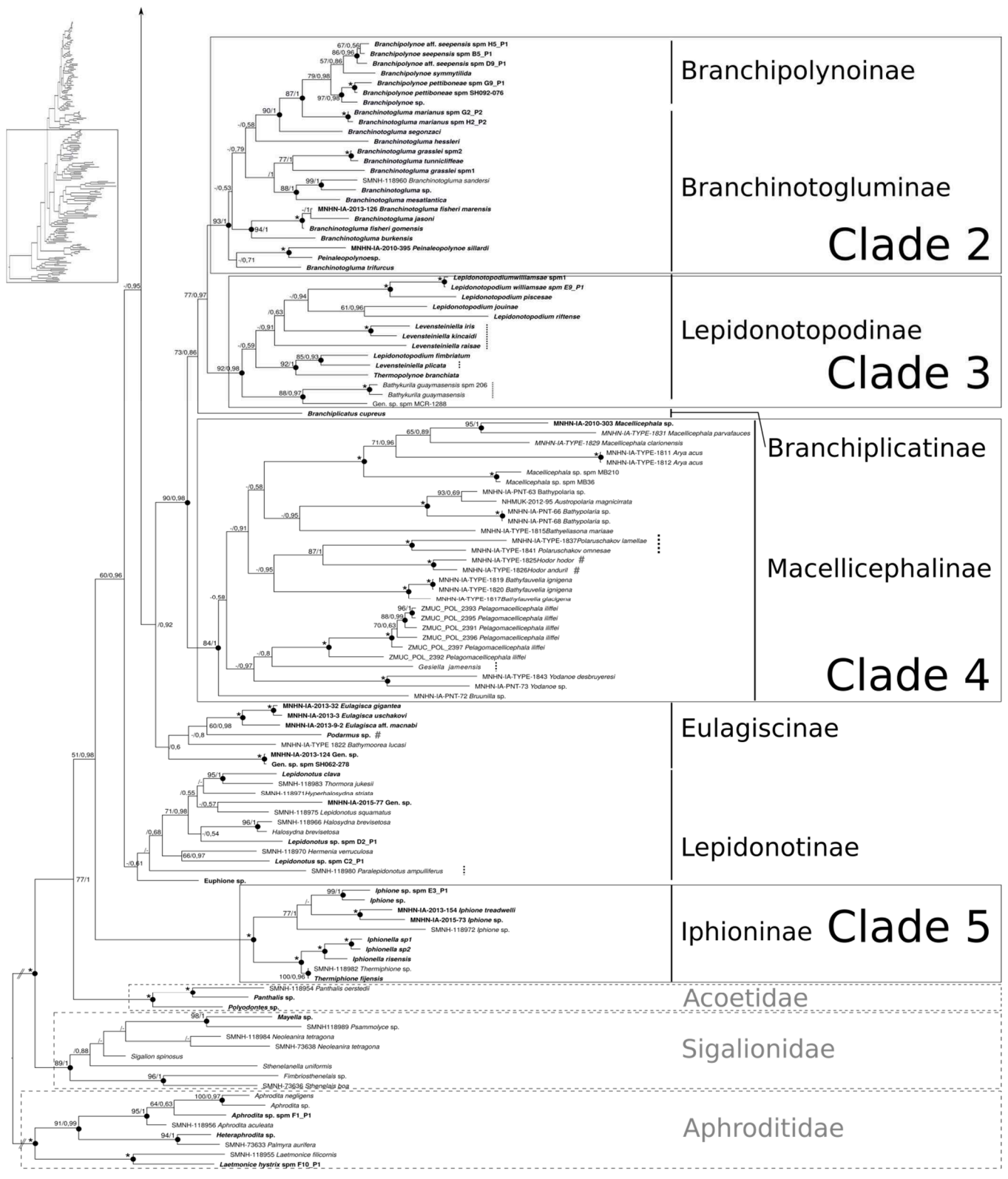


Figure II.6 Topologie bayésienne des relations phylogénétiques entre Polynoïdæ, obtenue avec la concaténation des marqueurs 16S, COI, 18S, 28SD1 et 28SD9-10. Les spécimens en gras proviennent de nos données de séquençage. Les nombres représentent les soutiens en ML / IB pour chaque nœud. Un tiret indique un soutien <50% en bootstraps ou <0,5 en pp. Un astérisque dénote des soutiens maximaux dans les deux analyses (100% en ML et 1 en IB). Une pastille indique un nœud soutenu (≥80% en ML et ≥0,95 en IB). Les sous-familles dans lesquelles sont morphologiquement placées les espèces sont renseignées. Les traits en pointillés verticaux indiquent que l'espèce appartient à une autre sous-famille (*Polaruschakov* spp. dans les Polaruschakovinae ; *Gesiella jameensis* dans les Gesiellinae ; *Levensteiniella* spp. et *Bathykurila guaymasensis* dans les Macellicephalinae ; *Paralepidonotus ampulliferus* dans les Polynoïdæ). Un dièse indique que le genre n'est placé dans aucune sous-famille (*Hodor* spp., *Podarmus* sp.). Les 5 clades numérotés représentent les grandes lignées soutenues à la fois en ML et en IB

Peinaleopolynoe (morphologiquement placés dans les Branchinotogluminae). Au sein du clade 2, les deux sous-familles (Branchipolynoïdæ et Branchinotogluminae) ne forment donc pas deux clades réciproquement monophylétiques. La paraphylie de ces sous-familles avait déjà été suggérée à partir d'un jeu de données incluant un nombre moins important d'espèces (Zhang *et al.*, 2018). Les auteurs de cette étude suggéraient que ces deux sous-familles sont synonymes, ce que nos résultats confirment. Afin de respecter le principe de priorité, il faudrait conserver le nom de Branchipolynoïdæ Pettibone, 1984. Le clade 3 (92%, 0,98 pp) contient les représentants des deux genres de notre jeu de données attribués à la sous-famille des Lepidonotopodinae (*Lepidonotopodium* et *Thermopolynoe*), les spécimens de 2 des 16 genres de Macellicephalinae de notre jeu de données (*Levensteiniella* et *Bathykurila*) et un spécimen récupéré dans NCBI qui n'a pu être attribué à aucun genre ni sous-famille. La sous-famille Lepidonotopodinae est donc monophylétique, à la condition de lui inclure le genre *Levensteiniella* et l'espèce *Bathykurila guaymasensis*. Nous proposons donc de redéfinir les Lepidonotopodinae en y incluant ce genre et cette espèce. Le clade 4 (84%, 1 pp) correspond à tous les autres genres de Macellicephalinae présents dans notre étude, ainsi que deux espèces du genre *Polaruschakov* et l'espèce *Gesiella jameensis*, morphologiquement placés respectivement dans les Polaruschakovinae et les Gesiellinae. Ce clade contient également deux espèces du genre *Hodor*, non attribué dans une sous-famille. Les espèces de la sous-famille Macellicephalinae à l'exclusion *Levensteiniella* et *Bathykurila*, forment avec les genres *Hodor*, *Polaruschakov* et l'espèce *Gesiella jameensis* un groupe monophylétique. Ce clade, qui inclut des spécimens attribués au genre type de la sous-famille (*Macellicephala* McIntosh, 1885), sera donc considéré comme définissant les Macellicephalinae. Enfin, le clade 5 (100% et 1 pp) correspond à tous les spécimens attribués à la sous-famille des Iphioninae.

En dehors de ces quatre sous-familles ainsi validées ou redéfinies (Branchipolynoïdæ, Macellicephalinae, Lepidonotopodinae, Iphioninae), l'arbre obtenu n'apporte aucune information au rang sous-familial. La vaste sous-famille des Polynoïdæ n'est notamment pas réunie dans un clade soutenu. Ceci s'explique potentiellement par le sous-échantillonnage particulièrement marqué dans cette sous-famille très diversifiée (438 espèces d'après WoRMS). Aussi, augmenter la couverture taxonomique dans ce groupe permettrait potentiellement d'améliorer le soutien de ce nœud.

II.3.2.3 Relations phylogénétiques récentes et limites des genres

Parmi les quatre sous-familles bien définies, les nœuds récents sont généralement bien soutenus, nous permettant d'évaluer la validité de la définition des genres représentés par plusieurs espèces. Dans les Iphioninae, le genre *Iphionella* (représenté par trois espèces différentes) est monophylétique. Le genre *Thermiphione* est représenté par deux spécimens, dont l'un n'est pas

identifié au niveau de l'espèce et provient de l'étude de Norlinder *et al.* (2012). La proximité génétique observée dans l'arbre, couplée au fait que ce dernier spécimen ait été récolté dans le bassin de Lau, tout comme *T. fijensis*, laisse penser que ces deux spécimens pourraient correspondre à l'espèce *T. fijensis*. Le genre *Iphione* forme un groupe monophylétique mais n'est soutenu que dans l'analyse bayésienne. Les spécimens attribués à ce genre n'ont, à l'exception près du spécimen MNHN-IA-2013-154 *Iphione treadwelli*, pas été assignés à des espèces connues. La distance génétique suggère cependant que les différents spécimens inclus dans l'analyse représentent bien des espèces différentes ce qui confirme donc la définition de ce genre.

Dans les Macellicephalinae, le genre type (*Macellicephalo*) est représenté par cinq espèces qui forment un groupe monophylétique soutenu à la condition de lui inclure l'espèce *Arya acus*. Le genre *Bathypolaria*, représenté par trois spécimens attribués à deux espèces, forme un groupe monophylétique soutenu (soutiens maximaux dans les deux analyses) avec l'espèce *Austropolaria magnicirrata*. Les genres *Bathyfauvelia*, *Hodor*, *Polaruschakov* et *Yodano* sont représentés chacun respectivement par deux espèces qui forment des clades soutenus, confirmant ainsi la monophylie de chacun de ces genres. Enfin, le genre *Pelagomacellicephalo* est représenté par six spécimens, tous attribués à l'espèce *P. iliffei*. L'ensemble de ces spécimens sont très divergents en 16S et COI (distances nucléotidiques de 0,17 en 16S et 0,19 en COI au maximum entre les spécimens ZMUC_POL_2392 et ZMUC_POL_2396) et sur cette base, semblent devoir être attribués à des espèces différentes. Si cette hypothèse est retenue, la monophylie du genre *Pelagomacellicephalo* est confirmée.

Dans les Lepidonotopodinae, les espèces du genre *Bathykurila* groupent de façon soutenue avec un spécimen non attribué à un genre. En revanche, *Lepidonotopodium fimbriatum*, espèce type du genre *Lepidonotopodium*, a pour espèce sœur *Levensteiniella plicata*. Ces deux espèces forment un clade soutenu avec l'espèce *Thermopolynoe branchiata*, seule espèce décrite du genre *Thermopolynoe*. *Levensteiniella kincaidi*, espèce type du genre *Levensteiniella* a pour espèce sœur *Levensteiniella iris* mais les relations avec les autres espèces de *Levensteiniella* ne sont pas résolues. Ce résultat suggère la synonymie de *Thermopolynoe* Miura 1994 et *Lepidonotopodium* Pettibone 1983 et la nécessité de redéfinir le contour des genres dans ce clade. Afin de respecter le principe de priorité, il faudrait conserver le nom de *Lepidonotopodium*.

Dans le clade 2, le genre *Branchipolynoe*, unique genre des Branchipolynoinae, est représenté par six espèces qui forment un clade soutenu. Les relations dans le genre *Branchinotogluma* (représenté par 13 espèces) ne sont pas toutes résolues, mais les espèces *B. marianus* et *B. segonzaci* se positionnent en paraphylie des *Branchipolynoe* avec de forts soutiens

(90% et 1 pp). L'espèce type de *Branchipolynoe*, *B. symmitilida* Pettibone 1984 est placée dans notre arbre phylogénétique parmi les autres espèces de ce genre. L'espèce type de *Branchinotogluma*, *B. hessleri* Pettibone 1985, n'est pas positionnée avec confiance dans l'arbre, mais ne se situe pas dans le clade avec les *Branchipolynoe*. En se basant sur ces données phylogénétiques, l'espèce *Branchinotogluma marianus* pourrait être attribuée au genre *Branchipolynoe*. Parmi les Branchinotogluminae, les deux espèces de *Peinaleopolynoe* présents dans notre étude forment un clade soutenu, confirmant la monophylie de ce genre.

Le clade 1 est formé uniquement de deux espèces du genre *Alentia*, monophylétique dans notre analyse. Ce genre est morphologiquement placé parmi les Lepidonotinae, dont le reste des espèces forme un clade peu soutenu (24% et 0,61 pp).

Le genre *Arctonoe*, représenté par cinq spécimens et seulement trois espèces nominales, forme un groupe peu soutenu (9 %, 0,76 pp). Dans ce genre, ni les spécimens attribués à *A. fragilis* ni ceux attribués à *A. pulchra* ne forment de clade monospécifique. Les spécimens *Arctonoe fragilis*, *Arctonoe vittata* et *Arctonoe pulchra* ont été identifiés par Bruno Pernet, tandis que les spécimens BIOUG<CAN>:BAMPOL0484 *Arctonoe pulchra* et BAMPOL0389 *Arctonoe fragilis* proviennent de la même étude (Carr *et al.*, 2011) et ont donc certainement été identifiés par la même personne, ne nous permettant pas de savoir si ces différences de placement dans l'arbre sont dues à des problèmes d'identifications ou des problèmes de délimitations entre ces espèces.

Parmi les genres attribués aux Polynoïdæ, seuls *Bylgides* et *Gorgoniapolynoe*, représentés chacun par au moins deux espèces, forment des groupes monophylétiques soutenus. Pour le genre *Harmothoe*, représenté par 22 spécimens attribués à 11 espèces nominales différentes et 4 espèces différentes non identifiées à l'espèce, les spécimens ne forment pas un clade distinct mais se répartissent dans ce groupe des Polynoïdæ. Quatre espèces sont représentées par plusieurs individus : *H. fuligineum* (3 spécimens), *H. imbricata* (3 spécimens), *H. ocularum* (2 spécimens) et *H. rarispina* (3 spécimens). Pour aucune de ces espèces nominales l'ensemble des spécimens ne forme de groupe monophylétique. Pour chacune des trois espèces nominales *H. fuligineum*, *H. imbricata* et *H. rarispina*, deux des trois individus séquencés groupent ensemble, mais le troisième est ailleurs dans l'arbre. De plus, les deux spécimens qui forment des clades proviennent soit de la même étude (*H. imbricata* Carr *et al.*, 2011 ; *H. rarispina* : Carr *et al.*, 2010) soit de deux études pour lesquelles les noms d'espèces ont été attribués par un même groupe d'identificateurs (*H. fuligineum* : Neal *et al.*, 2014 et Brasier *et al.*, 2016), tandis que le troisième spécimen provient d'une autre étude (*H. fuligineum* : nos données de séquençage, identification par S. Hourdez ; *H. imbricata* : Rousset *et al.*,

2007 ; *H. rarispina* Gonzalez *et al.*, 2017b). Là encore, il n'est pas possible de savoir si ces différences sont dues à des problèmes d'identification des spécimens ou à un défaut dans la délimitation des espèces. Ce résultat, ajouté à la dispersion dans l'arbre des spécimens attribués au genre *Harmothoe*, montre que les contours de ce genre sont mal définis et que les caractères morphologiques utilisés pour le définir résultent possiblement de convergences évolutives.

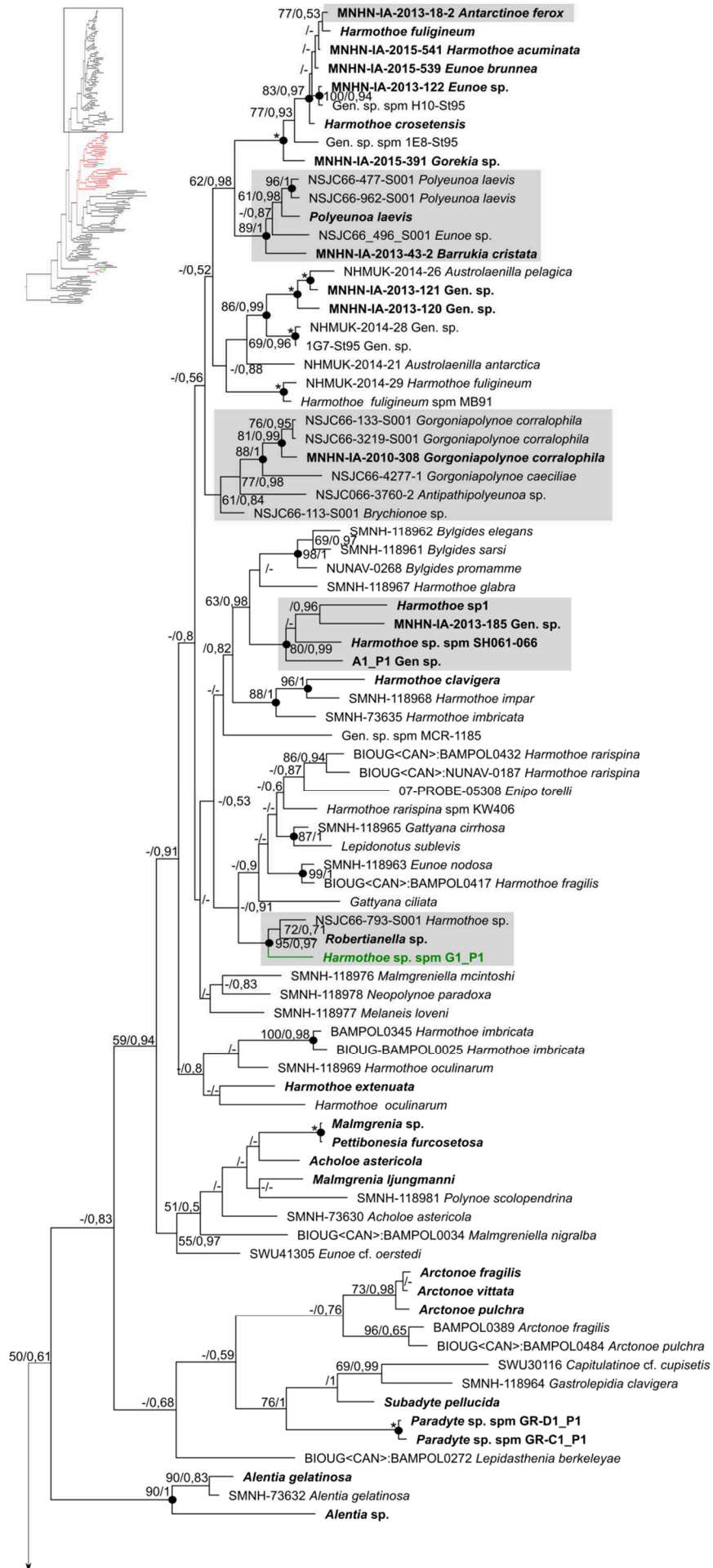
Parmi les genres attribués aux Polynoinae, un clade intéressant est retrouvé de manière soutenu, et ne comporte que des spécimens récoltés en Antarctique (soutiens maximaux dans les deux analyses). Ce clade comporte des spécimens attribués à différents genres tels qu'*Eunoe*, *Gorekia* et *Harmothoe*. Ce clade semble refléter une radiation géographique plus qu'une radiation écologique. La présence de trois autres spécimens attribués au genre *Eunoe* ailleurs dans l'arbre, couplé à la proximité génétique entre certains spécimens d'*Eunoe* et d'*Harmothoe* suggère que le genre *Eunoe* est polyphylétique. Les différences morphologiques des espèces antarctiques, qui ont conduit à les positionner dans des genres présents dans d'autres environnements pourraient résulter de convergence morphologiques au cours de cette radiation antarctique.

II.3.2.4 Colonisation du milieu profond

Les spécimens échantillonnés à des profondeurs d'au moins 200 mètres se répartissent à différents endroits de notre phylogénie (figure II.7). Ces spécimens sont soit isolés (cinq spécimens), soit se regroupent de manière soutenue dans les deux analyses (cinq clades) ou peu soutenue (deux groupes). Notre échantillonnage suggère donc qu'il y a eu jusqu'à 12 colonisations potentielles du milieu profond par les Polynoidae.

Lignée profonde principale

La lignée soutenue la plus diversifiée taxonomiquement ayant colonisé l'environnement profond (à partir de 200 m) est constituée des clades 2, 3 et 4 (soutiens de 90% et 0,98 en pp). Les Macellicephalinae (clade 4) sont une sous-famille constituée principalement d'espèces de l'environnement profond. Seules les espèces des genres *Pelagomacellicephala* et *Gesiella* habitent les grottes anchialines, un environnement peu profond qui peut être considéré comme analogue à l'environnement profond, car il est obscur, que peu de matière organique extérieure y parvient et que les conditions physico-chimiques y sont stables. Le spécimen *Branchipolynoe* sp. a été récolté entre 360 et 750 m et les spécimens *Peinaleopolynoe* sp. et *Branchipolynoe* aff. *seepensis* spm D9_P1 ont été récoltés vers 600 m de fond. Tous les autres ont été prélevés entre 1 200 et 4 000 m de profondeur. Les spécimens anchialins (qui vivent dans des points d'eau enclavés, reliés à l'océan par



Clade 1



Clade 2

Clade 3

Clade 4

Clade 5

↑ **Figure II.7 Topologie bayésienne des relations phylogénétiques entre Polynoïdae, obtenues avec la concaténation des marqueurs 16S, COI, 18S, 28SD1 et 28SD9-10.** Les spécimens en gras proviennent de nos données de séquençage. Les nombres représentent les soutiens en ML / IB pour chaque nœud. Un tiret indique un soutien <50% en bootstraps ou <0,5 en pp. Un astérisque dénote des soutiens maximaux dans les deux analyses (100% en ML et 1 en IB). Une pastille indique un nœud soutenu (≥80% en ML et ≥0,95 en IB). Les 5 clades numérotés représentent les grandes lignées soutenues à la fois en ML et en IB. Les couleurs sont relatives à l'environnement dans lequel sont retrouvées les espèces : en rouge les sources hydrothermales, en bleu les suintements froids, en vert les autres environnements enrichis en matière organique. Les spécimens récoltés à plus de 200m de fond sont encadré par un rectangle gris

voie souterraine) de *Pelagomacellicephala ifillei* forment un groupe monophylétique soutenu (soutiens maximaux dans les deux analyses) mais leur relation avec l'espèce anchialine *Gesiella jameensis* n'est pas soutenue, ne permettant pas d'émettre d'hypothèse sur le nombre de transitions entre les milieux anchialins et profonds dans cette sous-famille. Cette lignée couvre également les sous-familles Branchiplicatinae, Branchinotogluminae et Branchipolynoïnae, dont les espèces vivent toutes dans les milieux profonds.

Autres lignées profondes

Quatre autres lignées constituées uniquement d'espèces retrouvées à au moins de 200 m de profondeur sont retrouvées de manière soutenue dans notre arbre. Parmi elles, trois ne sont constituées que d'espèces morphologiquement attribuées à la sous-famille Polynoïnae.

La première (soutenue à 89% en ML et 1 de pp en IB) est composée de trois spécimens de *Polyeuna laevis* et des spécimens MNHN-IA-2013-122 *Eunoe* sp. et MNHN-IA-2013-43-2 *Barrukia cristata*. L'espèce *P. laevis* et certaines espèces du genre *Eunoe* sont des commensaux (Martin et Britayev, 1998). Les spécimens de *P. laevis* ont été récoltés sur des crinoïdes ou des coraux (scléactiniaires ou gorgones primnoïdes), ce qui confirme que le spectre d'hôtes colonisés est relativement large (comme suggéré par Alvaro *et al.*, 2014). La grande diversité génétique retrouvée sur le marqueur COI entre des spécimens de *P. laevis* reflète la présence d'espèces cryptiques, ce qui a également déjà été souligné dans des études antérieures (Alvaro *et al.*, 2014). Le spécimen d'*Eunoe* sp. a été récolté sur une holothurie, tandis que le spécimen de *Barrukia cristata* a, lui, été collecté sur un fond meuble. A l'exception de ce dernier spécimen, cette lignée profonde de Polynoïnae semble être spécialisée dans des associations avec des invertébrés (échinodermes et coraux).

La seconde lignée soutenue comprenant des espèces profondes placées dans les Polynoïnae est composée de quatre spécimens, dont deux attribués au genre *Harmothoe* (*Harmothoe sp1* et *Harmothoe* sp. spm SH061-066) et les deux autres non attribués à un genre particulier (MNHN-IA-

2013-185 Gen. sp. et A1_P1_Gen sp.). Ces spécimens proviennent de milieux profonds meubles ou rocheux, hormis MNHN-IA-2013-185 Gen. sp., qui a été prélevé sur une ophiure.

La troisième lignée de Polynoïnae des milieux profonds soutenue est constituée de trois spécimens non identifiés à l'espèce : deux du genre *Harmothoe* et un du genre *Robertianella*. Ce dernier a été récolté dans une éponge. Le spécimen NSJC66-793-S001 *Harmothoe* sp. provient de l'étude de Serpetti *et al.* (2016) sur les Polynoidae commensaux et est une espèce errante. Le spécimen *Harmothoe* sp. spm G1_P1 été collecté sur du bois coulé. Sa séquence pour le marqueur 16S était très proche (< 1% de différence nucléotidique) de celle d'un autre spécimen vivant dans une éponge siliceuse (non présent dans la phylogénie). Le COI de ce dernier spécimen n'a pas pu être amplifié, mais la similarité des séquences de 16S laisse penser que ces deux spécimens pourraient être de la même espèce. Il est donc possible que cette espèce soit spécialisée dans la colonisation de divers substrats organiques. Les associations biotiques telles que le commensalisme facultatif ou obligatoire sont courantes au sein des polychètes et particulièrement des Polynoidae (Britayev *et al.*, 2014). Des regroupements d'espèces partageant cette caractéristique écologique sont visibles dans notre arbre figure II.7, bien que les espèces soient taxonomiquement éloignées. Des auteurs ont récemment proposé une étude axée sur cet aspect écologique (Serpetti *et al.*, 2016). Pour chacune de ces trois lignées profondes de Polynoïnae, aucune relation avec un groupe frère n'est soutenue, ne permettant pas de discuter des milieux des espèces proches de celles-ci.

La dernière lignée soutenue constituée entièrement d'espèces profondes correspond à l'un des deux clades de la sous-famille Iphioninae (clade 5). Cette sous-famille est une petite lignée (4 genres pour 13 espèces décrites dans WorMS) bien couverte dans notre analyse phylogénétique (3 genres sont représentés par au moins 2 espèces). Dans cette sous-famille, une lignée est profonde (constituée des genres *Thermiphione* et *Iphionella*) et spécialisée des environnements enrichis en matière organique et une seconde lignée, plus généraliste, est constituée des espèces du genre *Iphione*, retrouvées sur différents milieux plutôt peu profonds (mais un spécimen, MNHN-IA-2013-154 *Iphione treadwelli*, a été récolté à près de 400 m).

Bien qu'ils ne soient soutenus dans aucune des analyses, deux autres regroupements d'espèces profondes sont visibles dans notre arbre et méritent d'être commentés. Le premier est constitué de spécimens placés dans la sous-famille Polynoïnae attribués aux genres *Gorgoniapolynoe*, *Antipathipolyeunoa* et *Brychionoe*. Pour le genre *Gorgoniapolynoe* qui forme un clade, notre échantillonnage comporte deux espèces nominales dont l'une est représentée par trois spécimens. Les spécimens NSJC066-3760-2 *Antipathipolyeunoa* sp. et NSJC66-113-S001 *Brychionoe*

sp. ont une position basale, bien que non soutenue, par rapport au genre *Gorgoniapolynoe*. Tous ces spécimens ont été récoltés sur des coraux, formant alors un autre regroupement d'espèces partageant le même type d'adaptation écologique. Le second regroupement de spécimens profonds non soutenu est constitué de sept spécimens qui correspondent, selon l'arbre, à six espèces différentes, dont deux non attribués à un genre, et quatre placées dans la sous-famille Eulagiscinae.

Enfin, cinq espèces profondes se placent de manière isolée dans l'arbre. Notamment, *Antarctinoe ferox* est récoltée à 283m de profondeur et se situe dans le clade discuté plus haut et composé uniquement de spécimens récoltés en Antarctique (composé des spécimens attribués à aux genres *Eunoe*, *Gorekia* et *Harmothoe*). Trois autres spécimens profonds se trouvent parmi des spécimens attribués à la sous-famille Lepidonotinae (*Lepidonotus* sp. spm C2_P1, *Euphione* sp. MNHN-IA-2015-77 Gen. sp.), mais sans soutien. Le dernier spécimen profond isolé fait partie de la sous-famille Iphioninae : MNHN-IA-2013-154 *Iphione treadwelli*.

II.3.2.5 Colonisation des environnements à base chimiosynthétique

Si la colonisation du milieu profond semble avoir eu lieu à plusieurs reprises au cours de l'évolution des Polynoidae, seules deux lignées comprennent des espèces associées aux suintements froids, aux accumulations de matière organique (carcasses de vertébrés, bois coulés) et aux sources hydrothermales.

La première lignée associée à ces trois environnements dans lesquels l'activité des bactéries chimiosynthétiques est possible, comprend les sous-familles Lepidonotopodinae (clade 3), Branchipolynoinae, Branchinotogluminae et Branchiplicatinae (ces trois dernières formant le clade 2). C'est la lignée la plus diversifiée taxonomiquement, et elle inclut des espèces provenant des trois types de milieux : les suintements froids (*Branchipolynoe seepensis* spm B5_P1, *Branchipolynoe* aff. *seepensis* spm D9_P1, *Branchipolynoe* sp., *Branchinotogluma fisheri gomensis* et *Peinaleopolynoe* sp.), les milieux enrichis en matière organique (MNHN-IA-2010-395 *Peinaleopolynoe sillardi* et les deux spécimens de *Bathykurila guaymasensis*) et les sources hydrothermales (tous les autres spécimens). Dans cette grande lignée, on distingue au moins six lignées différentes, correspondant aux milieux non hydrothermaux, dont quatre avec des espèces de suintements froids. Les espèces du genre *Branchipolynoe* sont commensales des bivalves Bathymodiolinae. *Branchipolynoe seepensis* Pettibone, 1986 a été décrit à partir de spécimens récoltés au niveau de suintements froids dans le Golfe du Mexique, sur l'escarpement de Floride (Pettibone, 1986). Plus tard, des spécimens ont été récoltés dans des moules des sources hydrothermales Atlantique et attribués à cette même espèce

(Segonzac, 1992 ; Desbruyères, 1994). La comparaison des distances génétiques entre des spécimens provenant de la Ride Médio-Atlantique et du Golfe du Mexique ont confirmé qu'il s'agit en réalité de deux espèces cryptiques (espèces génétiquement différentes mais morphologiquement identiques) (Chevaldonné *et al.*, 1998). En suivant cette hypothèse, nous avons nommé les spécimens morphologiquement similaires aux *B. seepensis*, mais prélevés ailleurs que sur les suintements froids de Floride, comme *Branchipolynoe* affinis *seepensis*. Les spécimens de notre étude *Branchipolynoe* aff. *seepensis* spm H5_P1 (collecté sur le site hydrothermal Atlantique Lucky Strike), *Branchipolynoe seepensis* spm B5_P1 (suintements froids de l'escarpement de Floride (localité type)) et *Branchipolynoe* aff. *seepensis* spm D9_P1 (sites de la pente de la Louisiane) forment une lignée, mais les distances génétiques entre ces trois spécimens suggèrent d'une part que les spécimens de suintements froids et de sources hydrothermales sont des espèces différentes, et d'autre part qu'il y a *a priori* au moins deux espèces différentes retrouvées sur les suintements froids du Golfe du Mexique. Bien que le clade ne soit pas soutenu, les spécimens de *Branchipolynoe* se regroupent, suggérant que leur similarité morphologique (qui justifie de les avoir placés dans un même genre), n'est pas due à une convergence évolutive résultant de l'adaptation au mode de vie commensal.

Dans cette lignée profonde, d'autres événements de colonisation des suintements froids sont visibles. Le spécimen *Branchipolynoe* sp., récolté dans l'océan Pacifique Ouest sur les suintements froids en Papouasie Nouvelle Guinée, forme un clade avec deux spécimens attribués à l'espèce *B. pettiboneae* récoltés sur les sources hydrothermales des Bassins de Manus et Lau. Là encore, les distances génétiques nous laissent penser qu'il s'agirait d'espèces différentes. *Branchinotogluma fisheri gomensis* (récolté au niveau d'un suintement froid dans le Golfe du Mexique) forme un clade avec les spécimens hydrothermaux *B. jasoni* (récolté dans le Bassin de Lau) et *B. fisheri marensis* (récolté sur la Dorsale Médio-Atlantique). Le dernier spécimen récolté sur des suintements froids est *Peinaleopolynoe* sp., frère de MNHN-IA-2010-395 *Peinaleopolinoe sillardi*, récolté sur un substrat organique. Enfin, les deux spécimens de *Bathykurila guaymasensis*, récoltés sur des carcasses de baleines (Glover *et al.*, 2005), forment un clade avec un spécimen non identifié provenant d'une source hydrothermale des Caïmans (Plouviez *et al.*, 2015). L'espèce *B. guaymasensis* est retrouvée à la fois sur les sources hydrothermales, les suintements froids et les milieux enrichis en matière organique comme les carcasses de baleine (Pettibone, 1989 ; Portail *et al.*, 2015). Le régime alimentaire basé sur le broutage de tapis bactérien de *Beggiatoa* a peut-être facilité la colonisation de divers milieux pour cette espèce, contrairement aux espèces qui vivent dans des habitats particuliers (comme les *Branchipolynoe* qui vivent dans les Bathymodiolinae ; Glover *et al.*, 2005). Seules des séquences de *Bathykurila guaymasensis* provenant des carcasses de baleines des bassins Santa Cruz et Santa Catalina sont disponibles sur les bases de données (Glover *et al.*, 2005), ne

permettant pas de vérifier moléculairement que les spécimens de ces trois environnements correspondent à une seule espèce et non pas à un complexe d'espèces cryptiques. Par ailleurs, le nombre de genres ou d'espèces de Polynoidae connus des environnements chimiosynthétiques non hydrothermaux est assez faible. Les Polynoidae identifiés au moins au niveau du genre retrouvés aujourd'hui sur des carcasses de mammifères sont *Bathyfauvelia* sp., *Bathykurila* cf. *guaymasensis* (Sumida *et al.*, 2016 ; Alfaro-Lucas *et al.*, 2017), *Bathykurila guaymasensis* (Glover *et al.*, 2005), *Peinaleopolynoe santacatalina*, *Harmothoe craigsmith* (Pettibone 1993) ou encore *Peinaleopolynoe sillardi*. D'autres Polynoidae ont été observés sur de tels substrats par différents auteurs, mais n'ont pas pu être assignés à un genre (Alfaro-Lucas *et al.*, 2017 ; Braby *et al.*, 2007). Les espèces de Polynoidae présentes sur les bois coulés en milieu profond sont des espèces d'*Austrolaenilla* (Amon, 2013), une espèce d'*Harmothoe* et au moins deux espèces d'*Iphionella*. Dans un article de revue sur les écosystèmes liés aux carcasses de baleines en milieu profond, Smith *et al.* (2015) suggèrent que, comme certaines espèces sont à la fois présentes dans ces milieux et dans l'environnement hydrothermal (comme avancé dans Glover *et al.* (2005)), cela permet d'envisager une transition écologique suivie d'une éventuelle spéciation d'un milieu vers l'autre, comme pour les moules Bathymodiolinae (Thubaut *et al.*, 2013). Les deux spécimens de *Bathykurila guaymasensis* forment un clade avec un spécimen non identifié récolté sur les sources hydrothermales. Du fait de la position dérivée de cette espèce, notre phylogénie ne permet pas de confirmer l'hypothèse d'une transition écologique depuis le milieu enrichi en matière organique (carcasse de vertébré) vers le milieu hydrothermal. Dans cette même lignée hydrothermale, le spécimen de l'espèce *Peinaleopolynoe sillardi* est le seul récolté sur une carcasse de mammifère dans le clade 2. Il est frère de *Peinaleopolynoe* sp., récolté sur les suintements froids, et le clade qu'ils forment est soutenu de manière maximale dans les deux analyses. Ce genre est l'un des seuls, avec *Bathykurila*, qui possède des espèces connues à la fois des milieux enrichis en matière organique, des suintements froids et des sources hydrothermales (Goffredi *et al.*, 2017). Cependant, notre échantillonnage n'incluant pas de spécimen de ce genre provenant des sources hydrothermales, nous ne pouvons pas vérifier que les espèces hydrothermales se positionnent bien avec les deux espèces de ce genre (*P. sillardi* et *P.* sp., respectivement des carcasses de vertébrés et des suintement froids) incluses dans notre étude. Les spécimens d'autres genres récoltés dans les suintements froids ont des positions terminales dans le clade 2 et groupent de manière soutenue avec des espèces de sources hydrothermales. Cela suggère que les suintements froids ont été colonisés plusieurs fois indépendamment au sein de cette grande lignée de Polynoidae.

Le représentant de la sous-famille monospécifique Branchiplicatinae (*Branchiplicatus cupreus*) est placé en groupe frère de l'ensemble des espèces des sous-familles Lepidonotopodinae,

Branchinotogluminae et Branchiplicatinae. Cependant le soutien est faible (73% et 0,86 pp) et ne permet pas de conclure que ces sous-familles forment un groupe monophylétique. Ces espèces, de milieux profonds permettant une production primaire chimiosynthétique locale, forment un clade soutenu (90% en ML et 0,98 pp) avec les espèces de la sous-famille des Macellicephalinae, sous-famille majoritairement profonde. Ainsi, cette grande lignée profonde (Macellicephalinae et espèces des environnements permettant une production primaire chimiosynthétique), très diversifiée taxonomiquement et écologiquement est donc constituée majoritairement d'espèces profondes, ce qui suggère une diversification dans le milieu profond relativement ancienne avec une spécialisation de certaines espèces pour les sources hydrothermales.

Le second clade qui comprend des espèces d'environnements permettant une production primaire chimiosynthétique se trouve dans la sous-famille des Iphioninae (clade 5). Ces espèces ont colonisé les sources hydrothermales (*Iphionella risensis* et *Thermiphione fijensis*) et les substrats organiques (bois coulés, *Iphionella* spp.). Les espèces du genre *Iphionella* sont retrouvées sur différents substrats comme des bois coulés (dans cette étude, deux espèces potentiellement nouvelles *Iphionella* sp 1 et *Iphionella* sp 2), les sources hydrothermales (*Iphionella risensis*) ou encore le milieu profond en bordure du talus continental (*Iphionella philippinensis* Pettibone, 1986, espèce non présente dans cette analyse). Au sein du genre *Iphionella*, notre arbre suggère une transition évolutive qui a suivi une transition écologique depuis le milieu hydrothermale vers les bois coulés. De plus, les espèces d'*Iphionella* forment un groupe monophylétique avec les deux espèces de *Thermiphione*, suggérant que l'ancêtre commun de cette lignée était présent sur les sources hydrothermales. La position basale des espèces hydrothermales par rapport à celles récoltées sur les bois coulés est à l'inverse de la théorie du « wooden step » (Distel *et al.*, 2000). Par ailleurs, les espèces d'environnements permettant une production primaire chimiosynthétique sont sœurs d'un clade composé uniquement d'espèces du genre *Iphione*, récoltées à des profondeurs généralement plus faibles. Les deux spécimens *Iphione* sp. spm E3_P1 et *Iphione* sp., qui forment un clade, ont été récoltées au niveau des récifs coralliens (milieu intertidal ou peu profond, < 40 m). Les spécimens MNHN-IA-2015-73 *Iphione* sp. et *Iphione treadwelli* ont été récoltés aux environs de 175 m et 400 m de profondeur, respectivement. Cette sous-famille a donc connu au moins deux transitions vers le milieu profond : une première au niveau de l'ancêtre commun aux *Thermiphione* et *Iphionella* et une seconde avant la spéciation ayant mené à *Iphione treadwelli*.

II.3.2.6 Adaptations à l'environnement hydrothermal pour le clade 2 3 4

Contrairement aux organismes qui dominent en biomasse l'environnement hydrothermal, les Polynoidae hydrothermaux n'ont pas développé de symbiose avec des bactéries chimiosynthétiques. Leur radiation adaptative dans ce milieu repose donc sur d'autres adaptations leur permettant de supporter les conditions locales et de tirer profit de la forte production primaire locale. Dans cette section sont présentés les éléments non encore publiés (Hourdez *com. pers.*) permettant de discuter quelles seraient les adaptations partagées par au sein des lignées de Polynoidae présentes dans les milieux profonds chimiosynthétiques.

Absence d'yeux

Toutes les espèces des clades 2-3-4 sont caractérisées par l'absence d'yeux. Cette caractéristique est également retrouvée chez d'autres espèces vivant dans les milieux profonds (Pettibone, 1976), ainsi que chez les Macellicephalinae des grottes anchialines peu profondes *Pelagomacellicephala* spp. et *Gesiella jameensis* (Gonzalez *et al.*, 2017c). L'absence d'yeux dans ces différentes lignées inféodées à des milieux en permanence obscurs, suggère que la fonction de ces organes est associée à un coût métabolique important. Selon cette hypothèse, la perte des yeux s'explique par la réallocation d'un budget métabolique dans d'autres fonctions, en réponse aux pressions de sélection propres à ces milieux.

Modification de l'appareil reproducteur

Plusieurs espèces de Polynoidae hydrothermales attribuées au clade 2 (sous-familles Branchinotogluminae et Branchipolynoïnae) et au clade 3 (sous famille Lepidonotopodinae) ont fait l'objet d'études anatomiques du système reproducteur : *Bathykurila guaymasensis*, *Branchipolynoe seepensis*, *Branchipolynoe* cf. *seepensis* et *Branchinotogluma* sp. (sous le nom *Opisthotrochopodus* sp.) (Gaudron *et al.*, 2017 ; Glover *et al.*, 2005 ; Jollivet *et al.*, 2000 ; Van Dover *et al.*, 1999).

L'appareil reproducteur de ces espèces est centralisé au niveau des segments terminaux, à partir du 11^{ème} segment et non réparti dans l'ensemble du corps comme pour la plupart des autres espèces de Polynoidae (Daly, 1972 ; Pernet, 2000). Une première observation d'un spécimen de *Macellicephala* sp. de la sous-famille Macellicephalinae (groupe frère des espèces retrouvées dans les environnements permettant la chimiosynthèse) a montré que ce spécimen possède également un système reproducteur centré au niveau des segments 10-12 (Hourdez *com. pers.*). Même si des observations d'autres espèces de Macellicephalinae sont nécessaires, cette donnée préliminaire

suggère que ce caractère pourrait être une synapomorphie du nœud soutenu qui réunit les clades 2, 3 et 4.

Chez les espèces hydrothermales examinées, la présence d'une spermathèque, et d'un sac ovigère chez les femelles ou de papilles néphridiennes pour transférer le sperme chez les mâles, suggèrent une fécondation interne. La présence de papilles néphridiennes chez les mâles des différentes espèces de *Lepidonotopodium*, *Thermopolynoe* et *Levensteiniella* semble également indiquer un système reproducteur centralisé et une fécondation interne pour ces genres du clade 3 (Hourdez, *com. pers.*). Classiquement, pour la plupart des espèces de Polynoidae, les gamètes (ovocytes chez les femelles et spermatozoïdes chez les mâles) sont libérés dans le liquide coelomique avant d'être émis dans l'environnement par les papilles néphridiennes segmentaires et la fécondation a lieu dans l'eau (Daly, 1972). Des sacs ovigères et des spermathèques ont également été documentées chez d'autres annélides des environnements hydrothermaux tels que les Alvinellidae ou les Siboglinidae (Desbruyères *et al.*, 1998 ; Jouin-Toulmond, 1997 ; Hilário, 2005 ; Zal *et al.*, 1994). Des modifications importantes du système reproducteur ont également été rencontrées chez des organismes hydrothermaux d'autres phylums, comme le caractère ovovipare rencontré chez les espèces hydrothermales d'un genre de nématodes, pour lequel les autres espèces non hydrothermales sont ovipares (Zeppilli *et al.*, *in prep*). De telles modifications anatomiques pourraient être une réponse adaptative aux environnements extrêmes instables (Tyler et Young 1999).

Par ailleurs, la période de reproduction n'est pas non plus la même entre les espèces hydrothermales et les espèces côtières de Polynoidae pour lesquelles des données sont disponibles (*e.g.* : parmi les Polynoïnae et les Arctonoïnae). Les espèces côtières étudiées jusqu'alors ont un comportement reproducteur saisonnier (régulé par la photopériode, la température, le cycle lunaire, etc ... Bentley and Pacey, 1992; Britayev, 1991 ; Daly, 1972 ; Watson *et al.*, 2000). Au contraire, les espèces profondes de sources hydrothermales, de suintements froids ou de carcasse de baleine *Branchinotogluma seepensis*, *Branchinotogluma. cf. seepensis*, *Bathykurila guaymasensis* et *Branchinotogluma* sp. se reproduisent de manière continue (Gaudron *et al.*, 2017 ; Glover *et al.*, 2005 ; Jollivet *et al.*, 2000 ; Van Dover *et al.* 1999). Enfin, les larves de Polynoidae ont généralement un développement planctotrophe (larves pouvant se nourrir dans la colonne d'eau) (Giangrande, 1997), tandis que les espèces des milieux permettant la chimiosynthèse citées précédemment (et placées dans les clades 2 et 3 de notre analyse), ont un développement lécithotrophe (larves pouvant se nourrir à partir de réserves vitellines). Ce type de développement peut permettre une dispersion des larves sur de longues distances, ce qui est utile pour des espèces vivant dans des milieux fragmentés (Gaudron *et al.*, 2017 ; Glover *et al.*, 2005 ; Van Dover *et al.* 1999).

II.3.2.6.1 Branchies

Toutes les espèces du clade 2 possèdent des branchies, un état de caractère très inhabituel chez les Polynoidae (Hourdez et Jouin-Toulmond, 1998 ; Hourdez *et al.*, 2001 ; Hourdez et Lallier, 2007). Leurs branchies sont ramifiées et émergent en deux touffes sur chaque parapode qui en porte (figure II.8 A et B). Cet état de caractère est une synapomorphie pour les genres et espèces de Branchinotogluminae et Branchipolynoïnae. Les espèces du clade 3 (Lepidonotopodinae) ne possèdent pas de branchie, à l'exception de *Thermopolynoe branchiata*. La structure anatomique des branchies de *T. branchiata* est très différente de la précédente : elles sont digitiformes et émergent comme des groupes sur les parapodes (figure II.8 C). Chez les Polynoidae, la présence de branchies n'est connue que pour les espèces du clade 2, *Thermopolynoe branchiata* et *Branchiplicatus cupreus*. La position de cette dernière espèce n'est pas soutenue dans l'arbre et la structure de ses branchies est encore différente des deux précédentes (elles forment des sacs longs et aplatis, figure II.8 D). Ces trois structures anatomiques très différentes suggèrent que ces branchies sont homoplasiques (également suggéré dans Gonzalez *et al.*, 2017a) et donc qu'elles résultent de convergences évolutives en réponse à l'hypoxie chronique rencontrée à proximité des sources hydrothermales (Hourdez, 2012). Cependant, toutes les espèces de Polynoidae qui vivent au niveau des sources hydrothermales ne possèdent pas de branchies, ce caractère n'est donc pas une adaptation indispensable à la colonisation de cet environnement.

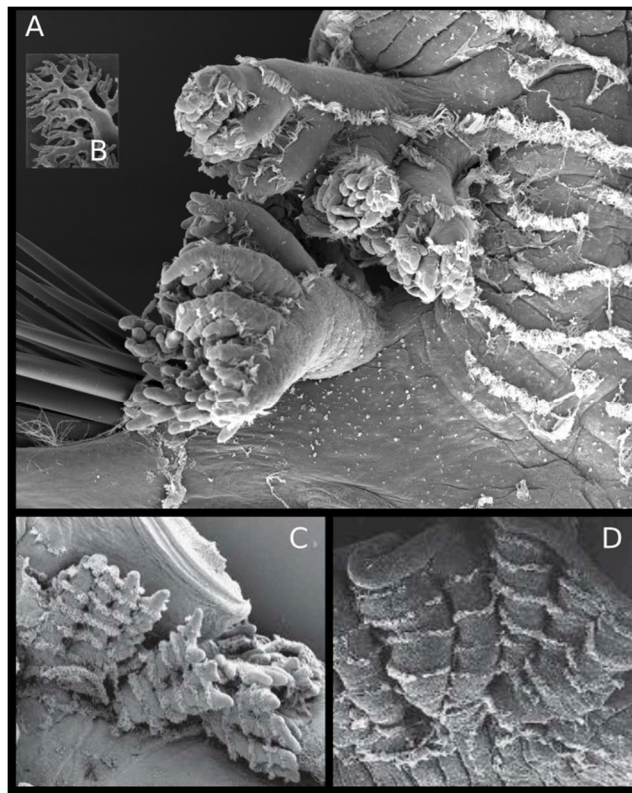


Figure II.8 Différents types de branchies rencontrées chez les Polynoidae des clades 2 et 3 et chez : (A) *Branchinotogluma sandersi* ; (B) *Branchipolynoe symmitilida* ; (C) *Thermopolynoe branchiata* et (D) *Branchiplicatus cupreus*.

II.3.2.7 Critères physiologiques : présence d'hémoglobine

Avant la découverte des espèces des environnements chimiosynthétiques, la famille des Polynoidae était l'une des rares familles d'annélides connues comme ne possédant pas d'hémoglobine circulante (Weber, 1978). Les espèces des Polynoidae connues alors possédaient des globines tissulaires dans leur système nerveux (neuroglobine) ou dans leurs muscles (myoglobine). Chez les espèces hydrothermales une globine circulante (hémoglobine) extracellulaire est trouvée en abondance dans le liquide cœlomique (Hourdez *et al.*, 1999). Elle a d'abord été mise en évidence et étudiée chez les Polynoidae commensaux Branchipolynoïnae *Branchipolynoe symmitilida* et *B. aff. seepensis* (Hourdez *et al.*, 1999). Les hémoglobines ont plus tard été séquencées également chez le Branchiplicatinae *Branchiplicatus cupreus*, les Branchinotogluminae *Branchinotogluma segonzaci* et *B. trifurcus* et le Lepidonotopodinae *Lepidonotopodium williamsae* (Projecto-Garcia *et al.*, 2017). L'hémoglobine peut avoir plusieurs fonctions. En particulier elle facilite la diffusion de l'oxygène du milieu extérieur vers l'intérieur et permet également de le stocker (Hourdez et Weber, 2005 ; Projecto-Garcia *et al.*, 2015). La simple présence d'hémoglobine a intrinsèquement une valeur adaptative dans les conditions d'hypoxie chronique du milieu hydrothermal, car elle permet (par sa capacité à le stocker), de tamponner les variations de concentration en oxygène dans l'environnement. Elle représente ainsi une réserve d'oxygène permettant de répondre aux besoins de l'organisme pendant environ une heure et demie chez *Branchipolynoe aff. seepensis* en cas d'anoxie environnementale complète (Hourdez et Lallier, 2007). De plus, grâce à sa forte affinité pour l'oxygène, elle facilite la diffusion vers la cavité cœlomique même avec de faibles concentrations environnementales (Hourdez *et al.*, 1999). La présence d'hémoglobine chez des organismes hydrothermaux appartenant à des lignées au sein desquelles les organismes habitant d'autres environnements en sont dépourvus a aussi été rapportée pour des copépodes (Fox, 1949 ; Sell, 2000 ; Hourdez *et al.*, 2000), ainsi que chez certaines patelles (*Rhynchopelta*, *Olgasolaris* ; Hourdez *com. pers.*).

Bien que toutes les espèces de Polynoidae d'environnements chimiosynthétiques possèdent de l'hémoglobine, il existe une certaine diversité de structure : les hémoglobines peuvent être mono- et tétra-domaines. La structure tétra-domaine est unique parmi les annélides (Hourdez *et al.*, 1999 ; Hourdez et Weber 2005) et résulte de la duplication en tandem d'un gène ancestral mono-domaine (Projecto-Garcia *et al.*, 2010). Bien qu'extracellulaires, l'analyse phylogénétique des séquences codant pour ces hémoglobines a montré qu'elles avaient une origine distincte des autres hémoglobines extracellulaires retrouvées chez les annélides (Projecto-Garcia *et al.*, 2010, 2017). Les données actuellement publiées (Branchipolynoïnae, Branchinotogluminae, Branchiplicatinae,

Lepidonotopodinae ; Projecto-Garcia *et al.*, 2017 ; Zhang *et al.*, 2017) suggèrent que l'hémoglobine était présente chez l'ancêtre commun à tous les Polynoidae d'environnements chimiosynthétiques. Une thèse est en cours dans l'équipe ABICE (UMR 7144) et nous savons que d'autres espèces possèdent également des hémoglobines (Hourdez et Le Layec *com. pers.*). Ainsi, les espèces du clade 3 (les Lepidonotopodinae *Lepidonotopodium pisceseae*, *Lepidonotopodium williamsae*, *Lepidonotopodium fimbriatum*, *Levensteiniella kincaidi*, *Levensteiniella raisae*, *Levensteiniella plicata* et *Thermopolynoe branchiata*) ne possèdent que des hémoglobines mono-domaines ; les Branchinotogluminae (clade 2 ; *Branchinotogluma segonzaci*, *B. hessleri*, et *B. jasoni*) possèdent aussi des hémoglobines tétra-domaines. Notre contexte phylogénétique suggère que la structure tétra-domaine est apparue dans l'ancêtre commun aux espèces du clade 2 (Branchinotogluminae et Branchipolynoidae). De plus, la présence d'hémoglobine retrouvée uniquement chez les organismes des milieux chimiosynthétiques suggère que c'est une adaptation clef nécessaire à la vie dans les environnements à production primaire chimiosynthétique, dans lesquels la concentration en oxygène peut parfois être très faible.

II.3.2.8 Le cas particulier des Iphioninae hydrothermaux

Les espèces d'Iphioninae retrouvées au niveau des milieux chimiosynthétiques sont caractérisées par l'absence d'yeux, tout comme de nombreuses espèces profondes (Pettibone 1976). Contrairement aux espèces des clades 2-3-4, les Iphioninae d'environnements permettant la production chimiosynthétique ne possèdent ni appareil reproducteur centralisé, ni branchie, ni hémoglobine. Dans cette sous-famille, les espèces décrites des sources hydrothermales sont généralement récoltées à la périphérie des écosystèmes hydrothermaux, dans des zones où il n'y a ni anomalie de température, ni présence de composés chimiques indicateurs de la présence de fluide. Cela laisse penser que les espèces retrouvées dans les milieux enrichis sont présentes de manière opportuniste, car la nourriture y est abondante, mais ne sont pas adaptées aux conditions extrêmes des sources hydrothermales.

II.4 Conclusions et perspectives

Nos données ajoutent des espèces dans tous les clades de la phylogénie des Polynoidae et montrent qu'une meilleure couverture de la diversité écologique permet de couvrir une importante diversité phylogénétique. Notre stratégie d'échantillonnage a permis de donner du soutien ou de mettre en évidence des clades regroupant des spécimens profonds (récoltés à plus de 200 m de fond) au sein des Polynoidae. Au total, jusqu'à 12 évènements de colonisation vers le milieu profond sont observables chez les Polynoidae, et seulement deux d'entre eux ont mené à des évènements de colonisation des environnements à base chimiosynthétiques. Il serait très intéressant d'obtenir des données de séquences pour *Vampiropolynoe embleyi*, qui est l'espèce hydrothermale pour laquelle la sous-famille Vampiropolynoinae a été créée (Marcus et Hourdez, 2002) et qui est la seule sous-famille hydrothermale non présente dans notre analyse phylogénétique. Nos données de séquence de l'ARN 16S suggèrent une claire parenté de cette espèce avec les autres espèces hydrothermales mais les autres marqueurs n'ont pas pu être séquencés. Cette espèce rare n'a été récoltée qu'à une seule occasion sur la Dorsale Juan de Fuca, sur des tapis bactériens épais formés après une éruption volcanique.

L'approche d'amplification des marqueurs par PCR suivi d'un séquençage Sanger a permis de résoudre un grand nombre de relations récentes, mais les branches profondes (relations entre les sous-familles) restent parfois peu soutenues. Il est possible que ces branches profondes correspondent à des évènements de radiation évolutive rapide et que l'on ne puisse donc pas résoudre ces nœuds avec confiance. Il est aussi possible que les marqueurs utilisés ne permettent pas de résoudre des divergences datant de cette période. Pour tester cette hypothèse, il est nécessaire d'augmenter le nombre de marqueurs phylogénétiques ou de définir de nouveaux marqueurs permettant d'obtenir de la résolution plus profonde que les marqueurs actuels. Zhang *et al.* (2018) ont proposé une phylogénie des Aphroditiformes en se basant sur les génomes mitochondriaux et les séquences de 18S et de 28S. La plupart de leurs nœuds sont soutenus, mais l'arbre n'est constitué que de 19 espèces. L'échantillonnage rassemblé dans notre étude couvre 13 des 20 sous-familles de Polynoidae et permettrait de constituer un jeu de données suffisant afin de tester l'intérêt du génome mitochondrial pour résoudre les relations phylogénétiques entre au sein des Polynoidae. Afin de tester cette approche pour les Polynoidae, l'approche par la méthode de shotgun mise en œuvre avec succès dans ce groupe (Zhang *et al.*, 2018) pourrait être adaptée. En effet, l'ordre des gènes mitochondriaux est variable dans la famille des Polynoidae, ce qui rend difficile l'approche par PCR longues (Zhang *et al.*, 2018). Cependant, étant donné que le mitogénome se comporte comme un seul locus (pas de recombinaison), il peut mener à des phylogénies des

espèces distordues, notamment si le taux d'évolution de ce marqueur varie entre les branches. Une autre alternative est de chercher de nouveaux marqueurs nucléaires, avec une approche de type transcriptomique. Dans l'équipe à Roscoff, une quinzaine de transcriptomes de Polynoidae ont été séquencés et assemblés, à partir d'espèces réparties dans quatre des cinq clades retrouvés dans cette étude (le clade 5 manque), et pourraient servir dans ce but.

Une meilleure résolution de cette phylogénie est nécessaire pour poser un cadre phylogénétique robuste permettant de comprendre l'histoire évolutive de certaines adaptations et la dynamique de diversification après la colonisation de ces environnements. Une phylogénie robuste est en effet nécessaire à l'utilisation de méthodes de reconstructions d'états ancestraux. Parmi ces adaptations, la seule qui soit partagée par l'ensemble des espèces chimiosynthétiques non Iphioninae est la présence d'hémoglobine, qui permet de transporter et stocker l'oxygène de l'environnement dans les tissus des organismes. Cette adaptation à l'environnement hypoxique semble être l'adaptation clef pré-requise à la colonisation des sources hydrothermales, puis *a posteriori* des autres environnements permettant la production chimiosynthétique (suintements froids, carcasses de vertébrés, bois coulés).

Chapitre III

Une approche NGS : le cas des crabes de la famille

Bythograeidae

III. UNE APPROCHE NGS : LE CAS DES CRABES DE LA FAMILLE BYTHOGRAEIDAE

III.1 Introduction

De nombreuses espèces de Décapodes sont présentes au niveau des sources hydrothermales. Ces espèces représentent jusqu'à 10% de la diversité des animaux identifiés (Desbruyères *et al.*, 2006). Une grande partie de ces espèces sont considérées comme opportunistes des milieux hydrothermaux (Martin et Haney, 2005). Parmi les espèces spécialisées des milieux hydrothermaux, on ne reconnaît actuellement que deux familles comme strictement associées à ces milieux. Ces deux familles appartiennent à l'infra-ordre des Brachyura : les Bythograeidae Williams, 1980 dans les sources profondes et les Xenograpsidae N. K. Ng, Davie, Schubart et Ng, 2007 présents sur des sources peu profondes. Les espèces incluses dans ces deux familles ont été découvertes dès les premières explorations des milieux hydrothermaux. Cependant, si la famille des Bythograeidae a été proposée dès la découverte de la première espèce, la famille Xenograpsidae n'a été décrite que 30 ans après la description du seul genre qu'elle inclut.

La famille de crabes Bythograeidae inclut actuellement 15 espèces (Leignel *et al.*, 2018 ; Guinot et Segonzac, 2018). La morphologie inhabituelle de ces crabes (notamment l'absence de pigmentation externe et des yeux très réduits) ainsi que les caractéristiques inhabituelles de l'environnement hydrothermal ont conduit à les classer dans une nouvelle famille de Brachyura. Les nombreuses études sur la physiologie et l'écologie des Bythograeidae renseignent en effet sur les traits biologiques spécifiques de ces prédateurs en relation avec les caractéristiques de leur milieu (Leignel *et al.*, 2018 pour une revue récente des données disponibles). Cependant, malgré cette cohérence morphologique et physiologique, la monophylie de cette famille n'est pas confirmée par une approche de phylogénie moléculaire.

Contrairement aux Bythograeidae, peu de données biologiques sont disponibles pour les trois espèces de la famille Xenograpsidae. Un comportement alimentaire très spécifique a été décrit pour l'espèce *Xenograpsus testudinatus*. En effet, deux fois par jour, au moment de l'étalement, les crabes *X. testudinatus* forment des agrégations à proximité des sources. Il a pu être montré que ces crabes se nourrissent des cadavres des organismes planctoniques qui meurent au contact de la plume toxique de la source hydrothermale et, à l'étalement, sédimentent à proximité des sources (Jeng *et al.*,

2004). Des données récentes confirment les adaptations comportementales de ces crabes à l'environnement hydrothermal (Dahms *et al.*, 2017). Contrairement aux Bythograeidae, la famille Xenograpsidae n'est proposée que tardivement. Cette proposition s'appuie sur des caractères morphologiques et une analyse phylogénétique qui suggère une divergence ancienne par rapport aux autres membres de la famille dans laquelle le genre était initialement placé.

Concernant les Bythograeidae, les phylogénies disponibles s'intéressent principalement aux relations entre les espèces qui le composent. Ainsi, en 2012, Mateos et collaborateurs proposent une phylogénie moléculaire incluant la quasi-totalité des espèces connues. Cette analyse utilise trois marqueurs nucléaires et trois marqueurs mitochondriaux. Dans cet article, les auteurs expliquent leur difficulté à enraciner la famille au sein des Brachyura. Les données disponibles pour d'autres Brachyura pour ces mêmes gènes ne leur permettent pas de reconstruire un arbre résolu pour les nœuds profonds. Dans ces jeux de données, seules les espèces du genre *Bythograea* forment un clade bien soutenu alors que les autres espèces de Bythograeidae ne sont pas positionnées de façon soutenue. Comme ils considèrent que la monophylie de la famille ne peut être mise en cause, ils proposent une analyse phylogénétique basée uniquement sur les espèces de Bythograeidae et enracinent leur arbre avec le clade des *Bythograea*. L'analyse concaténée des six gènes, enracinée sur les espèces du genre *Bythograea*, montre un groupe soutenu constitué des genres *Allograea*, *Austinograea* et *Gandalfus* alors que les positions des espèces *Segonzacia mesatlantica* et *Cyanograea predator* sont irrésolues.

Williams, en 1980, propose la famille Bythograeidae pour accommoder l'espèce *Bythograea thermydron*. L'espèce semble tellement divergente que Williams a également créé la superfamille des Bythograeoidea. Il propose des affinités éventuelles avec des familles telles que les Portunidae, les Potamidae, les Xanthidae ou encore les Goneplacidae. Depuis, de nombreuses autres hypothèses quant au groupe frère potentiel des Bythograeidae ont été proposées, tels que les Cancridae, Panopeidae, Trapeziidae ou les Xanthidae (Brösing *et al.*, 2007 ; Guinot *et al.*, 2002 ; Jamieson et Tudge, 2000 ; Leignel *et al.*, 2007 ; Von Sternberg *et al.*, 1999 ; Toullec *et al.*, 2002). Trouver le groupe frère des Bythograeidae au sein des Brachyura est ardu, car cet infra-ordre est très vaste : c'est le plus diversifié parmi les Décapodes avec environ 7000 espèces décrites (Guinot *et al.*, 2013). Les Brachyura sont historiquement divisés en quatre sections : les Dromiacea (6 familles), les Raninoidea (1 famille), les Cyclodorippoidea (3 familles) (ces trois sections formaient précédemment les Podotremata) et les Eubrachyura (89 familles) (Guinot, 1977 ; Ng *et al.*, 2008 ; Zhang *et al.*, 2011 : 165). Les Eubrachyura se subdivisent en deux sous-sections : les Thoracotremata (19 familles) et les Heterotremata (70 familles) et 3 familles dont l'emplacement est incertain. La morphologie des

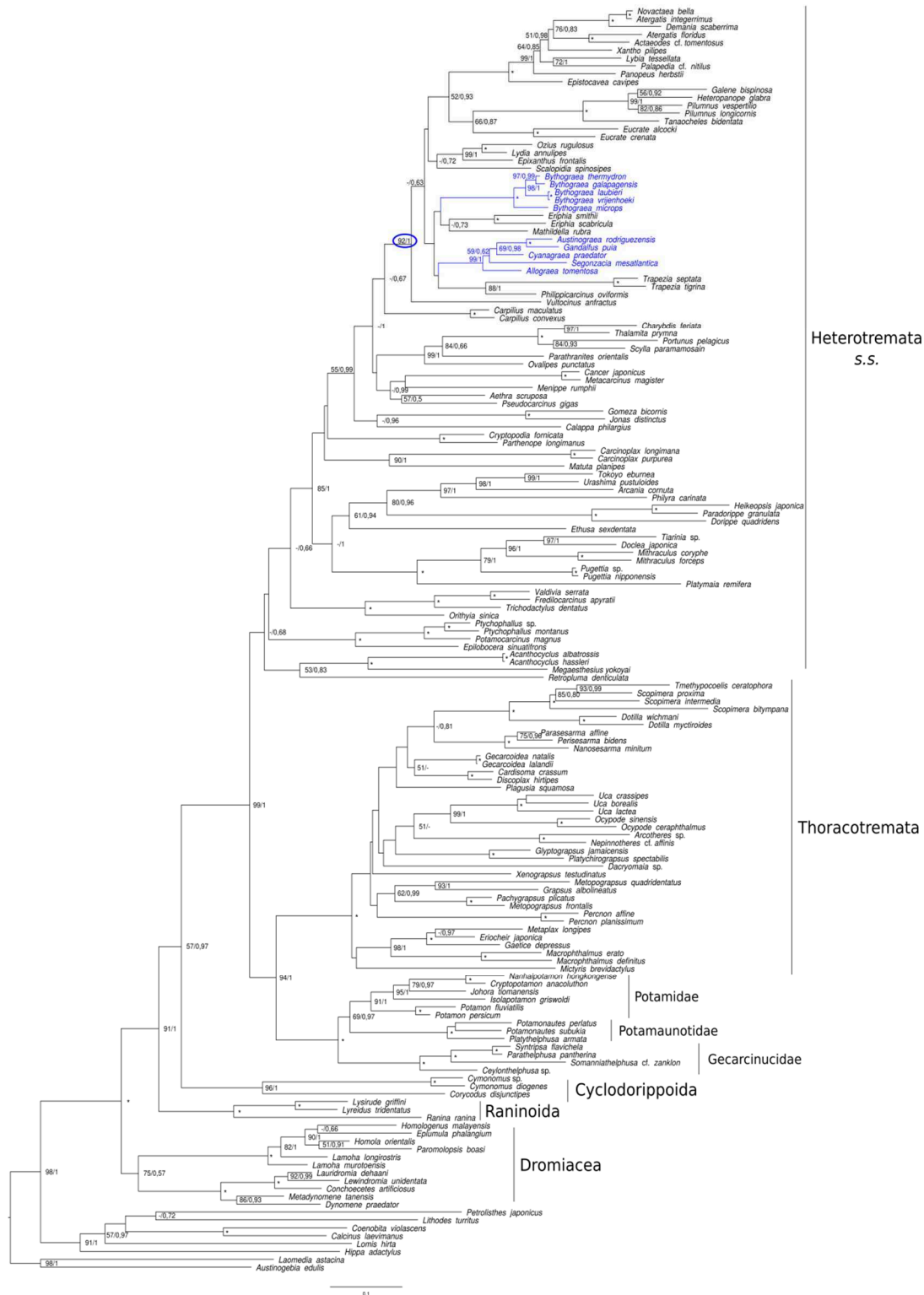


Figure III.1 Topologie RAxML (v8.2.9) des relations entre Brachyura, obtenues avec la concaténation de l'alignement des marqueurs NaK, H3, 12S et 16S. Le jeu de données est partitionné par gène et par position de codon pour les gènes NaK et H3. Le modèle GTR + Γ + I est appliqué pour chaque partition de l'analyse. Un rééchantillonnage est effectué avec 1 000 bootstraps. L'échantillonnage correspond à l'ensemble du jeu de données de Tsang *et al.* (2014) pour les marqueurs NaK, H3, 12S et 16S, auxquels ont été ajoutés les données de NaK, H3 et 16S des Bythograeidae provenant de l'étude de Mateos *et al.* (2012). Un astérisque indique un soutien maximal pour les deux analyses. Un tiret indique un support inférieur à 50% bootstraps ou 0,5 en probabilité postérieure. Les espèces de Bythograeidae sont colorées en bleu. Les sous-sections sont indiquées, sauf pour les familles Potamidae, Potamonotidae et Gecarcinidae, qui forment un groupe monophylétique à part.

Bythograeidae place ces crabes au sein des Heterotremata (Williams, 1980). Dans ce contexte, le problème d'enracinement rencontré par Mateos *et al.* (2012) peut s'expliquer de deux façons : (i) la divergence de cette lignée avec les autres Brachyura est trop ancienne pour être retracée avec les marqueurs choisis ; ou (ii) les espèces utilisées comme groupes externes potentiels ne sont pas les plus proches.

Une étude récente de Tsang *et al.* (2014) analyse les relations phylogénétiques au sein de l'infra-ordre des Brachyura. A ce jour, c'est l'étude qui combine à la fois le plus de marqueurs (six gènes nucléaires et deux mitochondriaux) et l'échantillonnage taxonomique le plus large au sein des Brachyura (140 espèces appartenant à 58 familles de Brachyura). Les nœuds qui correspondent aux familles sont assez bien résolus. Cependant, les relations entre les familles ne le sont pas ou seulement dans une seule des deux analyses proposées (maximum de vraisemblance/inférence Bayésienne). La famille des Bythograeidae n'est pas présente dans cette étude. Nous avons rencontré le Professeur Ling Ming Tsang, premier auteur de cette étude, qui explique que les espèces de Bythograeidae (pour lesquelles les données étaient disponibles pour ces gènes) ne se positionnaient pas de façon soutenue et suggéraient une possible polyphylie. Cette tendance est illustrée (figure III.1) par l'analyse réalisée sur les gènes de ce jeu de données pour lequel des données sont également disponibles pour les Bythograeidae du jeu de données de Mateos *et al.* (2012). La figure III.1 fournit le résultat obtenu sur le jeu de données concaténé pour les gènes NaK, H3, 12S et 16S. Cette analyse confirme le positionnement de l'ensemble des genres de Bythograeidae au sein des Heterotremata et confirme qu'ils se divisent en deux lignées dont les relations ne sont pas résolues. Ces deux lignées se placent dans l'arbre au sein d'un clade soutenu avec ces quatre gènes (92% bootstrap 1 pp. entouré en bleu figure III.1) qui contient les familles Eriphiidae, Euryplacidae, Galenidae, Mathildellidae, Oziidae, Panopeidae, Pilumnidae, Scalopidiidae, Tanaochelidae, Trapeziidae, Xanthidae, et Vultocinidae. Dans l'arbre de Tsang *et al.* (2014), produit avec huit gènes, ce clade était également retrouvé (à l'exception des Eryplacidae) de manière soutenue dans une seule des deux analyses (ML / IB). Dans l'étude de Tsang *et al.* (2014), les Goneplacidae sont également inclus dans ce clade. Pour résoudre la position des Bythograeidae, Ling Ming Tsang suggère de compléter l'analyse avec un nouveau jeu de marqueurs : les mitogénomes.

En 2013, deux études phylogénétiques incluant les Bythograeidae, basées sur les séquences des gènes codant pour les protéines mitochondriales ont été publiées : (i) Yang *et al.* (2013) séquent cinq nouveaux mitogénomes de décapodes hydrothermaux dont deux crabes Brachyura (Bythograeidae : *Austinograea alayseae* et *A. rodriguezensis*) et reconstruisent une phylogénie visant à dater la colonisation des sources hydrothermales par les décapodes ; (ii) Shen *et al.* (2013) fournissent 14 nouveaux mitogénomes de décapodes et s'intéressent aux relations phylogénétiques

entre les grands groupes de décapodes. A l'échelle des Brachyura, les échantillonnages taxonomiques de ces deux études sont très similaires : les deux études incluent sept crabes nageurs de la famille Portunidae, trois espèces de Varunidae, un crabe hydrothermal Xenograpsidae, un Potamidae et un Menippidae. L'étude de Yang *et al.* (2013) ajoute trois espèces de Bythograeidae à ce jeu de données, alors que celle de Shen *et al.* (2013) n'en comporte qu'une seule. Shen *et al.* (2013) ajoutent une espèce de Dromiidae. Ces deux études couvrent donc seulement 7 des 93 familles de Brachyura, la plupart d'entre elles n'étant représentées que par une seule espèce. Les données de ces deux études ne permettent donc pas de tester la monophylie des Bythograeidae ni d'identifier leur groupe-frère. L'étude de Yang *et al.* (2013) propose également une estimation de l'âge de la colonisation des milieux hydrothermaux, grâce à une analyse du registre fossile pour les Décapodes, de manière à calibrer l'horloge moléculaire. Bien que leur échantillonnage au sein des Brachyura soit parcellaire, ils datent le nœud à partir duquel divergent les familles Menippidae et Bythograeidae aux environs de 60 Ma, soit au début du Cénozoïque. Les études de Yang *et al.* (2013) et Shen *et al.* (2013) ont montré que les génomes mitochondriaux apportent une bonne résolution à l'échelle de l'infra-ordre.

A l'occasion d'une campagne récente (BICOSE2014, campagne IFREMER menée par Marie-Anne Cambon-Bonavita), Laure Corbari a rapporté au MNHN quelques spécimens de Bythograeidae de l'espèce *Segonzacia mesatlantica* dont la position relative aux autres Bythograeidae n'était pas résolue dans l'étude de Mateos *et al.* (2012). Les études basées sur les mitogénomes n'incluant pas cette espèce, il a été décidé de contribuer au projet avec le Professeur Ling Ming Tsang en séquençant le génome mitochondrial de l'espèce *Segonzacia mesatlantica*.

L'obtention des mitogénomes est classiquement réalisée par amplification par PCR de fragments chevauchants (cf. Shen *et al.*, 2013). Cette approche implique de dessiner des amorces spécifiques en utilisant les séquences disponibles pour des espèces proches. A condition d'avoir des données de référence, cette méthode est rapide et relativement peu coûteuse. Par ailleurs, cette approche ne fournit que le génome mitochondrial. Une seconde approche possible est l'obtention du mitogénome par une approche de séquençage de type shotgun. Dans cette approche, l'ADN total est extrait à partir des tissus puis fragmenté. Un faible effort de séquençage des banques construites à partir de ces extraits fragmentés permet d'obtenir les régions répétées du génome nucléaire (notamment les séquences ribosomales comme le 18S et le 28S) ainsi que l'ADN mitochondrial présent généralement en grand nombre de copies dans chaque cellule. Cette méthode permet donc d'obtenir, à faible coût et sans connaissance préalable, le mitogénome et certains gènes nucléaires classiquement utilisés en phylogénie des métazoaires.

L'objectif de cette étude était donc de tester cette seconde approche sur *Segonzacia mesatlantica* afin d'évaluer si elle permettrait, à moindre coût : (i) d'élargir les jeux de données mitogénomiques au sein des Brachyura, et (ii) de tester l'efficacité des génomes mitochondriaux pour résoudre des relations phylogénétiques à l'échelle des Brachyura. Ainsi, l'approche du séquençage en shotgun a été comparée à l'approche par PCR longues.

III.2 Matériel et méthodes

III.2.1 Echantillons biologiques et extraction d'ADN

Les deux spécimens utilisés ont été collectés lors de la campagne océanographique Ifremer BICOSE2014 menée sur la Ride Médio-Atlantique (MAR). Le spécimen MNHN-IU-2013-15617 a été récolté sur le site TAG (26°08' 14.07''N; 44°49'33.75''W lors de la plongée 573 Aspirateur 8) tandis que le spécimen MNHN-IU-2013-15603 a été récolté sur le site Snake Pit (23°22'6.32''N; 44°57' 11.66''W lors de la plongée 565 Aspirateur 3). A bord, Laure Corbari a préservé les spécimens dans l'alcool et les a identifiés. Les ADN des deux spécimens ont été extraits à l'aide du robot Epmotion 5075 (Eppendorf) en suivant les indications du fabricant (kit NucleoSpin 96 Tissu de Macherey-Nagel).

III.2.2 Séquençage du génome mitochondrial en PCR longues

L'approche des PCR longues est très similaire à celle d'une approche par PCR classique. L'enzyme permettant la réaction et le tampon de cette enzyme diffèrent de la PCR habituelle, et un allongement du temps d'élongation permet d'obtenir des amplicons pouvant faire jusqu'à une longueur de 10kb. Le génome mitochondrial ayant une taille de l'ordre de 15kb, seulement quelques fragments chevauchants permettent d'assembler un mitogénome complet. Une condition préalable pour obtenir des amplicons de grande taille est que l'ADN natif servant de matrice pour initier l'amplification ne soit pas trop fragmenté. L'approche par PCR repose sur la définition d'un couple d'amorces qui, en s'hybridant sur l'ADN dénaturé de part et d'autre du fragment ciblé, permettent de répliquer l'ADN à l'aide d'une polymérase. Les amorces universelles définies pour les quelques gènes mitochondriaux classiquement utilisés dans les approches de phylogénie ne sont généralement pas suffisantes pour amplifier le génome mitochondrial entier, car leur proximité sur le génome demanderait d'amplifier des fragments de très grande taille. Notamment, l'organisation des gènes doit être préalablement connue pour s'assurer de la longueur des fragments amplifiés avec les amorces disponibles et de leur degré de chevauchement. Pour cela, il est nécessaire de disposer de mitogénomes d'espèces les plus proches possible phylogénétiquement. A l'aide de ces références, de nouvelles amorces peuvent ainsi être dessinées. Cinq mitogénomes de quatre espèces de Bythograeidae sont disponibles dans la base de données NCBI et nous ont servi à dessiner des amorces et prédire la taille des fragments amplifiés par celles-ci. Ces cinq mitogénomes sont ceux de : *Austinograea alayseae* et *Austinograea rodriguezensis* (NC_020314 et NC_020312, Yang *et al.*, 2013), *Gandalfus puia* (NC_027414, Kim *et al.*, 2015), *Gandalfus yunohana* (NC_013713, Yang *et al.*, 2010) et *Austinograea alayseae* (KC851803, Kim *et al.*, 2014). Nous avons également utilisé

l'ensemble des séquences mitochondriales de *Segonzacia mesatlantica* disponibles dans NCBI à savoir les séquences partielles de COI, 16S et cytochrome oxydase b, issues de Mateos *et al.*, 2012 (JQ407473, JQ407447 et JQ407417, respectivement). Neuf amorces (tableau III.1) sont dessinées en plus des amorces universelles LCOI, HCOI (Folmer *et al.*, 1994) et 12S-R (Simon *et al.*, 1994) afin d'amplifier huit fragments se chevauchant. Les amorces peuvent s'hybrider même si leur séquence diffère de quelques nucléotides par rapport à la séquence d'ADN cible. De ce fait, les séquences correspondant aux zones d'hybridations d'amorces sont artéfactuellement identiques à celles des amorces. Les séquences correspondant à ces dernières ont donc été éliminées à la suite de l'assemblage* et une attention particulière a été portée pour que chaque zone d'hybridation d'une amorce soit également dans une zone de recouvrement entre fragments. Le dessin des amorces se fait avec l'outil *Add annotation* proposé par le logiciel Genious v.9.1.7 (Kearse *et al.*, 2012). Les amorces ont été dessinées dans les zones les plus conservées possible, en faisant l'hypothèse que la séquence de *Segonzacia mesatlantica* est suffisamment similaire aux séquences des autres espèces de Bythograeidae disponibles pour permettre aux amorces de s'hybrider. Chaque amorce est dessinée de manière à avoir une température d'hybridation (T_m) attendue autour de 60°C. L'absence d'homodimères (hybridation entre deux séquences d'une même amorce) et de hairpin (hybridation au sein d'une même amorce) ont été vérifiées sur le logiciel Genious. La possibilité d'hétérodimères (hybridation entre les séquences de deux amorces différentes (typiquement sens et anti-sens) n'a pas été vérifiée.

Les réactions de PCR longues ont été effectuées sur le spécimen MNHN-IU-2013-15617 en suivant le protocole de Hinsinger *et al.*, 2015 avec quelques modifications : les réactions de PCR se sont déroulées dans un volume total de 18 μ L comprenant du tampon de réaction pour la Taq LongAmp, 0,4 ng/ μ L de Sérum Albumine de Bovin (BSA), 3,5% de DMSO, 300 nM de chaque amorce, 300 μ M de dNTPs et 1 unité de Taq polymérase LongAmp (Biolabs). Après une étape initiale de dénaturation de 30 secondes à 94°C, l'ADN a été amplifié en faisant 37 cycles de 20 secondes de dénaturation à 94°C suivies de 30 secondes d'hybridation des amorces suivies de 15 minutes d'élongation à 65°C. L'amplification est finalisée par une dernière étape d'élongation de 15 minutes à 65°C. Deux températures d'hybridation, 50°C et 55°C, ont été testées. La PCR comportant l'amorce sens PM_COI-78 a été effectuée avec 2 amorces anti-sens mises en quantités équivalentes dans le mélange PCR : l'amorce PM_Asn-4618 et l'amorce PM_Asn-4631, car ces deux fragments sont différents de seulement 13 nucléotides. La présence d'un produit de la taille attendue a été vérifiée par révélation aux UV de la migration sur gel d'agarose à 0,75% des produits de PCR. L'ensemble des PCR ayant produit une amplification ont été séquencées en suivant la méthode

Tableau III.1 Nom, séquences (5'-3'), sens d'utilisation de l'amorce (F pour Forward = sens ; R pour Reverse = anti-sens) et provenance des amorces utilisées pour l'amplification du mitogénome de *Segonzacia mesatlantica*.

| Nom de l'amorce | Séquence (5'- 3') | Sens | Etude |
|-----------------|--------------------------------|------|----------------------------|
| PM_COX178 | AGTAGGAACATCTCTAAGTCTAATTATCCG | F | Cette étude |
| PM_COX1674 | AATAAGTGTTGGTAAAGAACGGGGTC | R | Cette étude |
| PM_Ala4502 | TGATATATGAGTTGCATTCATAAGGAGTTC | F | Cette étude |
| PM_Asn4618 | GTGACATACCTCTAGTTTGGTTTCTATC | R | Cette étude |
| PM_Asn4631 | TTTGTCAATTAACAGTGACATACCTCTAG | R | Cette étude |
| PM_Cytb9582 | GTAGATAACGCCACCTTAACTCGA | F | Cette étude |
| PM_cytb9720 | AGTAAGGGTGGGAATGGGAGTTTATC | R | Cette étude |
| PM_16s11425 | CAAGAGATAGAAACCGACCTGGC | F | Cette étude |
| PM_16s11909 | TGACCGTGCAAAGGTAGCATAATAATTAG | R | Cette étude |
| LCOI-1490 | GGTCAACAAATCATAAAGATATTG | F | Folmer, 1994 |
| HCOI-2198 | TAACTTCAGGGTGACCAAAAAATCA | R | Folmer, 1994 |
| 12S-R | AAACTAGGATTAGATACCC | R | Simon <i>et al.</i> , 1994 |

d'Hinsinger *et al.* (2015). Les reads* issus du séquençage ont été une première fois cartographiés de manière itérative (25 cycles) indépendamment sur les mitogénomes de référence de *Gandalfus yunohana* (NC_013713) et d'*Austinograea alayseae* (NC_020314) avec un maximum de 10% de nucléotides divergents autorisé par read avec le logiciel Geneious. Les séquences consensus ont été vérifiées à la main (vérification de l'alignement des reads, suppression des séquences correspondant aux amorces) avant de servir de nouvelle référence pour une seconde cartographie des reads de séquençage avec un maximum de 3% de nucléotides divergents tolérés par read (25 cycles itératifs). Les deux séquences consensus obtenues avec chacun des mitogénomes de référence ont ensuite été comparées pour vérifier la congruence des séquences. Les cadres de lecture des gènes codants ont été vérifiés à la main. L'annotation du mitogénome a été effectuée avec le logiciel Geneious et vérifiée avec le serveur web MITOS (Bernt *et al.*, 2013).

III.2.3 Préparation des banques pour le séquençage shotgun

La préparation de la banque pour le séquençage shotgun a été effectuée sur le spécimen MNHN-IU-2013-15603 en suivant le protocole *NEBNextVR Fast DNA Fragmentation & Library Prep Set* pour Ion Torrent™ (E6285L, New England Biolabs). Les fragments de taille de 200 à 500 pb ont été sélectionnés et la banque a été chargée sur une puce Ion 316V2 pour être séquencée (Hinsinger *et al.*, 2015) sur le séquenceur Ion Torrent personal machine (PGM) (Life Technologies, Carlsbad, CA) avec le kit de séquençage Ion PGMTMHi-QTM (séquençage au Service de Systématique Moléculaire (SSM) de l'UMS 2700 OMSI).

III.2.4 Cartographie des reads obtenus en shotgun

III.2.4.1 Assemblage du mitogénome de Segonzacia mesatlantica :

L'assemblage du génome mitochondrial grâce aux reads obtenus par la méthode du séquençage shotgun est effectué en cartographiant les reads sur un mitogénome de référence (NC_027414, *Gandalfus puia*) avec le logiciel Geneious v.9.1.7. 20% de nucléotides divergents sont tolérés entre le read et la référence.

Afin d'évaluer la proportion exacte de reads mitochondriaux parmi les reads séquencés, les reads obtenus par la méthode du séquençage shotgun ont également été cartographiés grâce au logiciel Geneious sur le génome mitochondrial nouvellement assemblé de *Segonzacia* avec les mêmes paramètres (maximum de 20% de divergence entre le read et la référence). Les reads de moins de 30 bases ne sont pas pris en compte dans cette analyse.

Les reads qui sont cartographiés sur la référence sont considérés comme des reads mitochondriaux lorsque :

- Les reads sont strictement identiques au mitogénome de référence ;
- Dans une région codante du mitogénome de référence, plusieurs reads identiques entre eux mais différents de la référence sont cartographiés, et que ce polymorphisme ne correspond pas à un codon stop ; dans ce cas la divergence nucléotidique est considérée comme du polymorphisme intra-spécifique).
- Les différences nucléotidiques occasionnant dans l'alignement avec la séquence de référence des indels (insertion/délétions) correspondent à des homopolymères (au moins trois nucléotides identiques à la suite) ; dans ce cas, la différence est considérée comme une erreur de séquençage.

Les reads cartographiés mais ne remplissant pas ces critères ont été interprétés comme des reads appartenant à des Numt (pour *Nuclear Mitochondrial DNA segment*, Lopez *et al.*, 1994), c'est-à-dire des copies de fragments d'ADN mitochondriaux transposées dans le génome nucléaire.

III.2.4.2 Recherche des gènes nucléaires 18S et 28S

Les gènes ribosomaux nucléaires 18S et 28S sont recherchés au sein des reads issus du séquençage en shotgun. L'ensemble des reads sont cartographiés sur plusieurs séquences de référence pour ces gènes, afin de vérifier la congruence des séquences obtenues. Pour le 18S, comme il n'existe aucune donnée pour les Bythograeidae, les séquences de références sont choisies au sein des Heterotremata, groupe dans lequel nous nous attendons à retrouver les Bythograeidae. Sans plus d'indication sur la proximité des différentes familles, les références sont prises dans des familles différentes. Les séquences de référence choisies pour la recherche du marqueur 18S sont celles de *Pseudocarcinus gigas* (HM638021) et *Meriola barrosi* (KY016806). Pour le 28S, une séquence est disponible dans NCBI pour un fragment de ce gène pour l'espèce *S. mesatlantica* (JQ407432) et est utilisée comme référence. Pour ce gène, la séquence disponible de *Carcinus maenas* (DQ079782) est plus longue (1838pb contre 637 pb pour *S. mesatlantica*) et est également utilisée comme référence. Dans l'ensemble de ces analyses, les reads sont cartographiés avec 3 itérations en tolérant jusqu'à 20% de différence entre les reads et la séquence référence. Pour chacun des deux gènes ribosomaux, les séquences obtenues pour un marqueur sont alignées entre elles afin de vérifier que les séquences retrouvées sont les mêmes. Elles sont ensuite tronquées afin de garder la même taille que leur séquence de référence et alignée avec l'ensemble des séquences du fragment correspondant présentes sur NCBI.

III.2.5 Analyses phylogénétiques

Les analyses phylogénétiques des séquences nucléiques ont été effectuées à la fois en utilisant les méthodes du Maximum de vraisemblance (ML) à l'aide des logiciels RAxML version 8.2.9 (Stamatakis *et al.*, 2014) et de l'Inférence Bayésienne (IB) avec MrBayes v3.2.6 (Ronquist *et al.*, 2012). Elles ont été effectuées en utilisant l'ensemble des mitogénomes de Brachyura disponibles dans la base de données NCBI (au 28 avril 2018), à partir de la concaténation des 13 protéines codantes et des 2 ARN ribosomiaux du génome mitochondrial. Les gènes ont été alignés un par un à l'aide de ClustalW (Sievers et Higgins, 2014), puis les cadres de lecture des gènes codant pour des protéines ont été vérifiés à la main. Lorsque plusieurs génomes mitochondriaux sont disponibles pour une même espèce, un seul d'entre eux est conservé pour l'analyse phylogénétique afin d'éviter la redondance au niveau spécifique. Cette redondance concerne *Eriocheir sinensis* (Varunidae), *Ranina ranina* (Ranidae), *Atergatis integerrimus* (Xanthidae), *Portunus pelagicus*, *P. sanguinolentus*, *Scylla paramamosain* et *S. serrata* (Portunidae), *Sinopotamon yangtsekiense* (Potamidae). Les deux spécimens d'*Austinograea alayseae* sont gardés car ils font partie de notre famille d'intérêt, les Bythograeidae, et que les séquences sont divergentes. Les analyses phylogénétiques finales (ML et IB) comportent 69 espèces de Brachyura réparties dans 22 familles.

Le logiciel RAxML n'accepte que le modèle d'évolution GTR, qui est un modèle neutre. Afin de comparer les résultats provenant des deux méthodes de reconstruction, nous utilisons le même modèle d'évolution dans les deux analyses. Nous avons donc utilisé le modèle d'évolution GTR avec la loi Γ afin de tolérer des variations du taux de substitution d'un nucléotide à un autre, en acceptant également une proportion de sites invariables I pour chaque marqueur et chaque position de codon dans le cas des 13 gènes codants (modèle GTR + Γ + I). La robustesse des nœuds a été estimée par la méthode du bootstrap avec 1 000 itérations. Dans la reconstruction par inférence bayésienne, deux analyses indépendantes ont été lancées, comportant chacune 8 chaînes de Markov, sur 10^9 générations avec une fréquence d'échantillonnage des arbres tous les 15 000 pas. Les arbres issus des analyses ML et IB ont été visualisés à l'aide du logiciel FigTree v1.4.3 (Rambaut et Drummond, 2012). Seuls les nœuds présentant à la fois 80% de bootstraps en ML et 0,95 de probabilité postérieure pour l'Inférence Bayésienne sont considérés comme soutenus.

Tableau III.2 Amorcés utilisés dans l'approche d'amplification du génome mitochondrial de *Segonzacia mesatlantica* par PCR longues. Nom de l'amorce sens, nom de l'amorce anti-sens, taille attendue du fragment amplifié, résultat de l'amplification pour les deux températures (50°C/55°C).

| Amorce sens | Amorce anti-sens | Taille attendue (kb) | Amp. des PCR (50°C/55°C) |
|-------------|----------------------------|----------------------|--------------------------|
| LCOI-1490 | PM_Asn4618 | 4,6 | Oui/oui |
| LCOI-1490 | PM_Asn4631 | 4,6 | Non/non |
| PM_coi78 | PM_Asn4618 & PM_Asn4631 | 4,6 | Oui/oui |
| PM_Ala4502 | PM_cytb9720 | 5,2 | Oui/oui |
| PM_Cytb9582 | PM_16S11909 | 2,4 | Oui/oui |
| PM_Cytb9582 | 12S-R | 3,4 | Oui/non |
| PM_16S11425 | HCOI-2198 | 5 | Non/non |
| PM_16S11425 | PM_coi674 | 5 | Oui/oui |

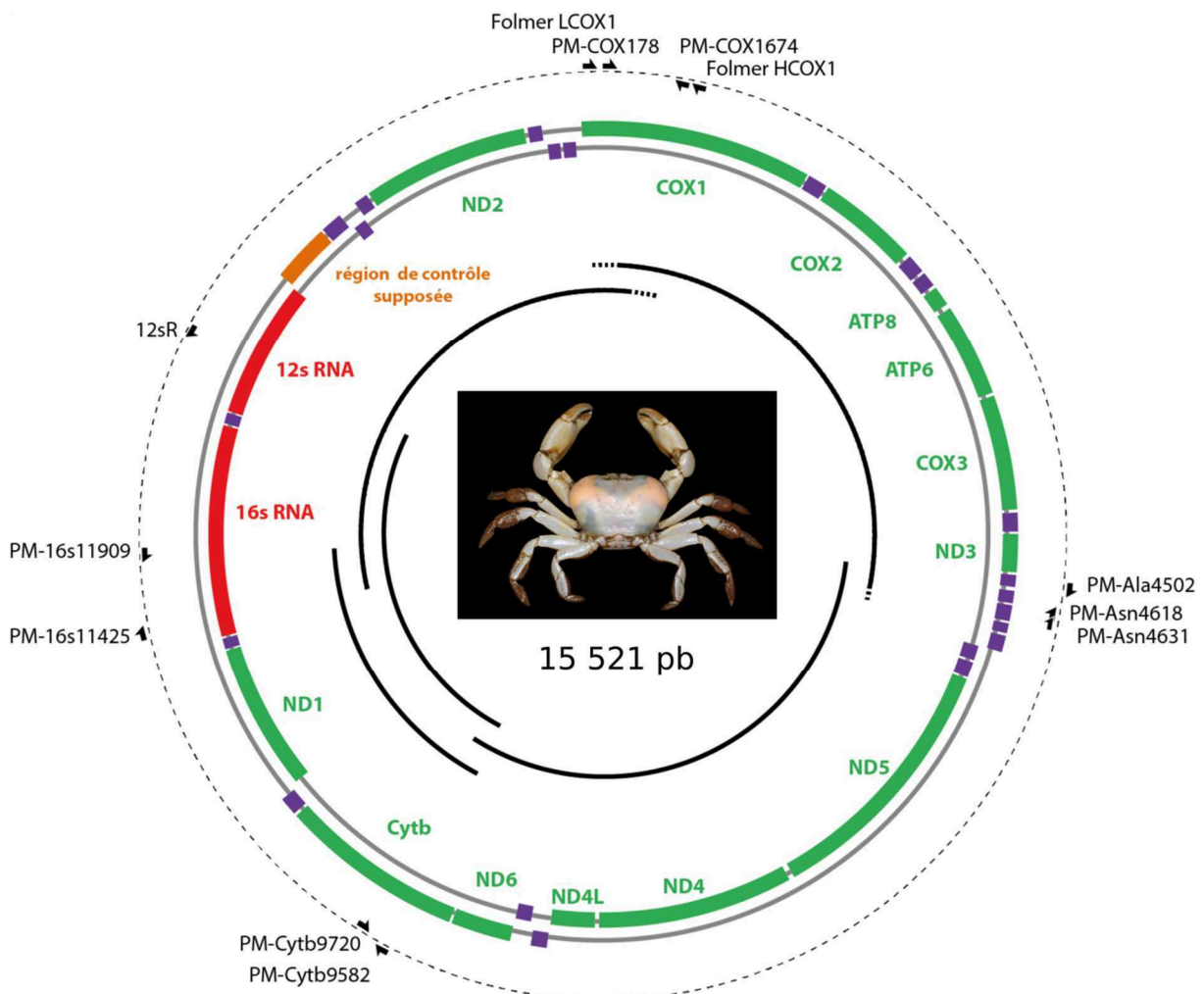


Figure III.2 Schéma du génome mitochondrial de *Segonzacia mesatlantica* (NC_035300). Les couleurs correspondent au type du gène codant pour des protéines (vert), ARN de transfert (violet), ARN ribosomaux (rouge) ; l'emplacement de la supposée région de contrôle est en orange. Le cercle extérieur correspond à l'emplacement des amorces pour les PCR longues. Les cercles relatifs aux gènes mitogénomiques correspondent au brin dit « lourd » (cercle externe) et au brin dit « léger » (cercle interne). Le cercle le plus interne schématise les fragments amplifiés par PCR longue, les pointillés indiquent que deux amorces ont été utilisées pour une amplification en PCR emboîtée. Le nom de chaque amorce indique le gène dans lequel elle se situe suivi de sa position (relative au mitogénome de référence).

III.3 Résultats - Discussion

L'approche de séquençage par la méthode des PCR longues a permis d'obtenir le génome mitochondrial de *Segonzacia mesatlantica*. Parmi les 18 PCR longues testées (9 couples d'amorces pour 2 conditions de températures d'hybridation), 11 ont produit un fragment de la taille attendue (tableau III.2). Le succès des PCR (*i.e.* nombre de PCR positives et intensité de la bande sur le gel d'agarose) est meilleur à une température d'hybridation de 50°C qu'à 55°C. Les 11 produits de PCR obtenus ont été séquencés pour produire environ 1,25 millions de reads. Une première cartographie de ces reads nous a permis d'établir que les banques séquencées sont très enrichies en ADN mitochondriaux (entre 14 500 et 19 000 reads sont cartographiés selon la référence utilisée), ce qui rend possible la reconstitution du génome mitochondrial en cartographiant les reads obtenus sur une référence qui n'a pas besoin d'être très proche phylogénétiquement (appartenant à un genre différent), du moment que l'organisation des gènes de la référence choisie est la même que dans le génome mitochondrial à assembler. Les séquences consensus finales obtenues après la cartographie des reads sur les séquences de *Gandalfus puia* et d'*Austinograea alayseae* sont très similaires : les seules différences sont des insertions artéfactuelles dues à des erreurs de séquençage dans des zones d'homopolymères. Le génome mitochondrial complet assemblé de *Segonzacia mesatlantica* a une longueur de 15 521 bp ; la séquence complète et annotée du génome mitochondrial est disponible dans la base de données NCBI sous les numéros d'accès KY541839 /NC_035300 (figure III.2) (Mandon *et al.*, 2017).

Le génome mitochondrial de *Segonzacia mesatlantica* contient 13 gènes qui codent pour des protéines, 2 gènes qui codent pour les sous-unités ribosomales 12S et 16S, 22 gènes d'ARN de transfert et une région de contrôle de 624 pb riche en AT (78,5%) (Annexe 2). Comme pour les autres Brachyura, parmi les 22 ARN de transfert, deux tRNA-Leu (anticodons UAA et UAG) et deux tRNA-Ser (anticodons UCU et UGA) sont retrouvés dans le mitogénome de *S. mesatlantica*. Les codons d'initiation sont : ATC, ATG, ATT ou GTG et les codons stop sont : TAA, TAG ou T- (qui s'étend en TAA lors de la polyadénylation).

Organisation 1



Organisation 2



Organisation 3



Organisation 4



Organisation 5



Organisation 6



Organisation 7



Figure III.3 Types d'organisation des gènes dans les génomes mitochondriaux de *Brachyura*. Les différentes organisations sont établies par rapport aux déplacements des gènes codant pour des protéines (en vert) et des ARNr (en rouge) et ne prennent pas en compte les modifications d'emplacement/direction des ARNt (en violet) et de la région de contrôle (en orange). Les noms des sept organisations sont arbitraires et ne reflètent pas la phylogénie.

L'organisation des gènes est un des éléments important à prendre en compte pour étendre l'échantillonnage taxonomique pour le mitogénome avec l'approche par PCR longue. La comparaison des 84 mitogénomes disponibles pour les Brachyura montre une grande variabilité de l'emplacement des séquences codant pour les ARNt, qui sont des gènes courts (~ 60-90 pb). Bien que les ARNt présentent des régions très conservées, il n'est pas judicieux de les utiliser comme zone d'amorçage pour une approche par PCR longues sur des taxons pour lesquels aucune donnée de référence proche n'est disponible, car leurs position change fréquemment (voir figure III.3). Cependant, leur changement de position n'a pas d'impact sur la réussite des PCR avec des amorces définies dans les autres gènes, car leur taille est particulièrement faible par rapport à celle des fragments amplifiés (plusieurs milliers de bases).

En ne considérant pas les changements relatifs à ces ARNt, il est possible de définir sept organisations des gènes mitochondriaux (voir figure III.3). Par ailleurs, si les amorces sont définies uniquement dans les régions codant pour les protéines, les sept organisations de gènes se réduisent à quatre. Si l'on considère uniquement 12 des 13 gènes protéiques (~ 9 000 pb), 55 génomes mitochondriaux de Brachyura (soit environ 66% du nombre de mitogénomes disponibles sur NCBI) ont la même organisation de gènes (correspondant aux organisations 1 à 4 sur la figure III.3). Les espèces qui ont un arrangement différent sont les espèces des familles Varunidae et Xenograpsidae, les Majidae du genre *Maja*, le Macrophthalmidae *Macrophthalmus japonicus* et le Gecarcinucidae *Somanniathelphusa boyangensis* (organisations 5 à 7, figure III.3). Parmi les gènes codant pour des protéines, six gènes (COX1, COX2, ATP8, ATP6, COX3, NAD3) couvrant 4 200 pb sont dans le même ordre pour l'ensemble des génomes mitochondriaux. L'organisation des gènes de *S. mesatlantica* est la même que celle des autres espèces de Bythograeidae. Cet ordre correspond à celui trouvé le plus fréquemment chez les Brachyura (correspond au groupe 1, figure III.3).

Dans le but de comparer l'approche de PCR longues avec l'approche Shotgun pour assembler le génome mitochondrial de *Segonzacia mesatlantica*, le génome de *Gandalfus puia* est utilisé comme référence pour la cartographie. Le séquençage shotgun a fourni 347 961 reads d'une longueur moyenne de 160 bases. Le nombre de reads est plus faible qu'attendu et leur longueur moyenne est deux fois plus courte que les 300 à 400 pb attendues, sans toutefois que nous ayons pu identifier la cause technique de ces différences. Lors de la cartographie, il faut noter que le caractère itératif de la méthode utilisée engendre des biais : en effet, un read qui va être cartographié sur la référence avec seulement une petite zone similaire à celle-ci va pouvoir engendrer la cartographie d'autres reads de plus en plus différents de la référence initiale, au fur et à mesure des étapes

itératives. Pour limiter de tels effets, ces reads artéfactuellement cartographiés (ainsi que les reads trop courts susceptibles d'introduire de tels biais) sont éliminés. Suite à cela, seulement 277 d'entre eux (soit moins de 0,1 %) ont pu être cartographiés sur le mitogénome de *Gandalfus puia* sans ambiguïté. Ces reads permettent de couvrir 13 402 bases du mitogénome de *G. puia* (soit une couverture de 86% de la longueur totale). Cependant la profondeur de séquençage est très faible et variable le long de la référence (entre 1 et 17X). De plus, on observe qu'une même région de la séquence de référence peut présenter des similarités avec plusieurs reads de séquences différentes. Dans le cas des marqueurs codant pour des protéines, la plupart de ces reads différents ont des cadres de lecture qui ne mènent pas à un codon stop, ce qui ne permet pas de discriminer les reads séquencés à partir du génome mitochondrial des reads provenant de Numt.

Afin d'identifier tous les reads de notre séquençage correspondant au génome mitochondrial et les différencier des Numt, nous les avons cartographiés sur le génome de *S. mesatlantica* assemblé grâce à l'approche des PCR longues. Parmi les 347 961 reads obtenus à partir du séquençage shotgun, avec les mêmes paramètres de cartographie que ceux utilisés avec le mitogénome de *Gandalfus puia*, 337 cartographient sur le mitogénome de *S. mesatlantica*. Parmi ces 337 reads, 35 sont plus petits que 30 bases et sont éliminés. Les 25 reads supplémentaires (par rapport à la cartographie sur *G. puia*) qui sont ainsi cartographiés se répartissent tout le long du génome mitochondrial, à raison de quelques reads par gène et cinq reads dans la région de contrôle. Ce résultat confirme que, dans les reads issus du séquençage, une très faible proportion de reads de taille supérieure à 30 pb correspond effectivement à l'ADN mitochondrial ciblé. En effet, seulement 302 reads plus longs que 30 bases ont été cartographiés, soit 0,09% des reads séquencés. Cette faible proportion de reads du séquençage shotgun cartographiés sur le mitogénome de *Segonzacia mesatlantica* indique que la quantité d'ADN mitochondriale séquencée est faible. Dans l'approche shotgun, l'hypothèse de travail est que le génome nucléaire, plus long que le génome mitochondrial, n'est présent qu'en un nombre faible de copies par cellule (*e.g.* deux pour un génome diploïde), alors qu'il y a un grand nombre de mitochondries (et donc de copies du mitogénome) par cellule. Une première cause d'écart à cette hypothèse est que les tissus utilisés pour extraire l'ADN pourraient être pauvre en mitochondries. L'extraction a cependant été effectuée à partir d'un périopode, et donc des muscles locomoteurs dans lesquels la densité en mitochondries est forte. Ce type de tissu a d'ailleurs été utilisé dans plusieurs études avec succès pour assembler le génome mitochondrial par une approche shotgun (Marquez *et al.*, 2014 ; Tan *et al.*, 2016a, b ; Yuhui *et al.*, 2017). Il est donc peu probable que la faible proportion de reads mitogénomiques par rapport aux reads nucléaires résulte du tissu utilisé, sauf si la méthode d'extraction a eu une efficacité différente sur les ADN génomiques

et mitochondriaux. Cette hypothèse est cependant peu probable à la vue du protocole très standard utilisé.

Un autre facteur pouvant affecter la proportion d'ADN mitochondrial séquencé est la taille de génome nucléaire. Cette situation a par exemple été mise en avant pour expliquer la faible couverture du mitogénome d'*Euphausia superba* dont le génome est estimé à près de 48 Gb (Leese *et al.*, 2012). Bien que grande relativement aux autres Brachyura, la taille du génome nucléaire de cette espèce est estimée à 4,76 Gb (Bonnivard *et al.*, 2009) et n'a pas une taille incompatible avec le succès de l'approche shotgun.

La dernière hypothèse est que le séquençage soit contaminé par de l'ADN d'organismes non ciblés (contamination bactérienne, par exemple). *Segonzacia mesatlantica* vit dans un environnement dans lequel les bactéries sont les producteurs primaires et sont trouvées libres ou en association avec les métazoaires (Van Dover, 2000). Cette espèce semble être à la fois nécrophage et prédatrice et se nourrit principalement de crevettes Alvinocarididae qui vivent en association avec des bactéries chimiosynthétiques (Desbruyères *et al.*, 2006 ; Colaço *et al.*, 2002). Pour tester cette hypothèse de contamination de nos extraits par de l'ADN bactérien, il faudrait cartographier l'ensemble des reads séquencés à partir des génomes de procaryotes disponibles en base de données sur nos reads de séquençage (*cf.* Miller *et al.*, 2008 ; Hun *et al.*, 2013).

Parmi les 302 reads cartographiant sur le mitogénome de *S. mesatlantica*, seulement 230 correspondent effectivement à des séquences mitochondriales. Les 72 autres reads (soit 23,8% des 302 reads) sont répartis sur tous les gènes du mitogénome assemblé, hormis les tRNA-Asp et tRNA-Arg et le gène codant pour la NADH 4L. L'ensemble de ces 72 reads sont donc interprétés comme des séquences nucléaires des gènes mitochondriaux (Numt). De telles séquences ont été identifiées chez d'autres Brachyura y compris chez les crabes Bythograeidae (Kim *et al.*, 2013). La présence de Numt dans la plupart des gènes mitochondriaux suggère que des copies de la quasi-totalité du génome mitochondrial seraient présentes dans le génome nucléaire. La taille des génomes a été évoquée comme un facteur pouvant conduire à un plus grand nombre de Numts, mais l'accumulation de données sur un plus grand nombre de taxons montre que bien d'autres facteurs interviennent (*e.g.* Richly et Leister 2004 ; Song *et al.*, 2014 ; Calabrese *et al.*, 2017). Ainsi, avec une plus grande profondeur de séquençage et des reads plus longs, il est probable que nous aurions fait plus facilement la différence entre les reads provenant de Numt et ceux provenant du génome mitochondrial. En effet, avec une profondeur de séquençage du génome mitochondrial adéquate, il est possible d'assembler un génome mitochondrial, même avec une référence éloignée (Machados *et al.*, 2015). La méthode du séquençage shotgun a d'ailleurs déjà permis d'assembler de nombreux

génomomes mitochondriaux de *Brachyura* (Tan *et al.*, 2016a, b ; Marquez *et al.*, 2014). Les difficultés rencontrées de l'approche shotgun reflètent donc plus probablement les problèmes techniques lors de notre réaction de séquençage qui a fourni une couverture et une profondeur insuffisante pour reconstruire le mitogénome.

L'autre intérêt de l'approche shotgun est d'obtenir, dans la même réaction de séquençage, des gènes nucléaires utilisables en phylogénie. Malgré le succès limité de notre approche shotgun, nous avons recherché les gènes ribosomiaux nucléaires 18S et 28S dans nos reads, ces gènes ribosomiaux étant présents en multiples copies dans le génome nucléaire. La recherche du fragment de 18S, réalisée avec *Meriola barrosi* et *Pseudocarcinus gigas* comme références, a fourni deux séquences (1 750 et 1 804 pb respectivement, identiques sur la partie partagée). Le long de ces séquences, la profondeur de séquençage varie entre 11X et 32X avec une moyenne de 20,6X. Aucune séquence de 18S de Bythograeidae n'est à ce jour disponible sur NCBI pour vérifier l'identité avec notre séquence.

Pour le 28S, nous disposons d'un fragment de 636 pb de 28S de *Segonzacia mesatlantica* mais avons également utilisé une séquence plus longue disponible pour *Carcinus maenas* (1 838pb). Ces deux références ont permis de retrouver le fragment de séquence recherché avec une profondeur variant entre 1 et 13X (moyenne de 7,1) pour la référence de 636 pb et entre 1X et 20X (moyenne de 9,1) pour la référence de 1 838 pb. La cartographie avec la référence *C. maenas* met en évidence des régions, absentes du fragment disponible pour *S. mesatlantica*, sur lesquelles de nombreux reads sont cartographiés. La séquence obtenue est identique à celle disponible sur NCBI.

La profondeur relativement faible le long des séquences assemblées pour ces deux gènes reflète probablement ici encore le succès limité du séquençage réalisé, mais confirme l'intérêt de ces approches pour obtenir ces séquences intéressantes pour la phylogénie sur des taxons pour lesquels ne sont disponibles que des références éloignées.

Un moyen de contourner les problèmes rencontrés avec l'approche shotgun que nous avons choisie, serait une approche d'enrichissement des banques en ADN mitochondrial avant le séquençage. Li *et al.* (2012) comparent les trois approches (PCR longue, enrichissement par capture et shotgun) et leurs résultats montrent que, bien qu'ayant une proportion d'ADN mitochondrial 2,5 fois moins importante qu'avec l'approche de PCR longues, leur enrichissement fournit une proportion de reads correspondant à de l'ADN mitochondrial 150 fois plus forte qu'avec un séquençage shotgun pour un même effort de séquençage. De plus, l'enrichissement par capture leur

permet facilement de discriminer les reads issus du mitogéome de ceux provenant de Numt grâce à la prévalence des reads mitochondriaux. Par ailleurs, une étude a montré que l'approche d'enrichissement de banque par capture peut permettre de cibler un grand nombre d'espèces pour lesquelles les données mitochondriales ne sont pas connues (Liu *et al.*, 2016). Les auteurs de cette étude utilisent les données mitochondriales d'insectes disponibles dans les banques publiques pour dessiner des sondes leur permettant de capturer les ADN mitochondriaux de branchiopodes, d'araignées mais également d'un chordé et, dans une moindre mesure, d'un échinoderme. D'une part, une telle approche d'enrichissement en ADN mitochondrial nous permettrait de nous affranchir à la fois du problème de profondeur de séquençage rencontré et du problème de discrimination des reads mitochondriaux par rapport aux reads provenant de Numt. D'autre part, cette approche nous permettrait, en dessinant des sondes à partir des données mitochondriales de *Brachyura* déjà disponibles dans les bases de données, d'élargir facilement l'échantillonnage à l'ensemble des *Brachyura*. Enfin, une telle approche enrichit les ADN mitochondriaux mais d'autres marqueurs nucléaires sont également séquencés (environ 55% de reads dans l'étude de Liu *et al.*, 2016). Nous pouvons supposer que parmi eux, de nombreux reads proviennent des ARN ribosomiaux.

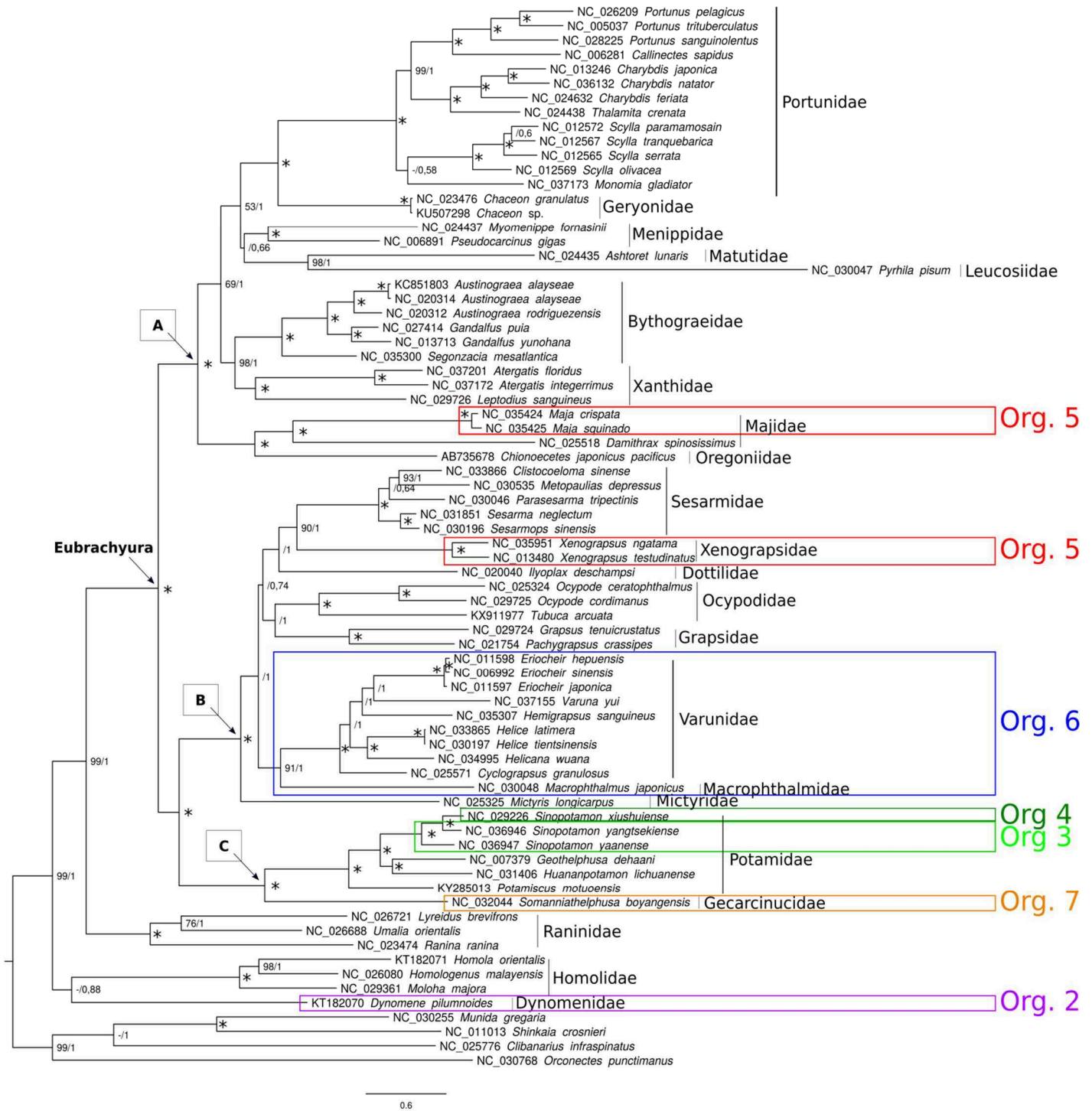


Figure III.4 Topologie bayésienne des relations phylogénétiques des Brachyura, basée sur l’alignement de la concaténation des séquences des 13 gènes mitogénomiques codant pour des protéines et des 2 rRNA pour 70 Brachyura et 4 Décapodes non-Brachyura. Les valeurs aux nœuds correspondent aux bootstraps /probabilités postérieures. Les astérisques indiquent un support maximal dans les deux analyses. L’organisation des gènes correspondant à la figure 2 est indiquée pour les espèces ayant une organisation différente de l’organisation 1. Trois clades soutenus (A, B et C) sont indiqués et sont cités dans le texte.

Afin d'évaluer la pertinence des mitogénomes pour répondre à la question de l'origine unique ou multiple des Bythograeidae, nous avons inclus la séquence du génome mitochondrial de *Segonzacia mesatlantica* dans le jeu de données mitogénomiques constitué à partir des bases de données publiques. L'échantillonnage taxonomique réuni est loin de couvrir la diversité des Brachyura mais apporte néanmoins des indications sur le pouvoir de résolution phylogénétique de ce marqueur. Dans nos analyses, la plupart des nœuds sont soutenus, que ce soit avec la méthode d'inférence Bayésienne (IB, $pp \geq 0,95$) ou avec la méthode du maximum de vraisemblance (ML, bootstraps $\geq 80\%$) (figure III.4). Les trois sections des Brachyura représentées dans notre échantillonnage (Dromiacea, Raninoida, et Eubrachyura) apparaissent comme monophylétiques. Les Eubrachyura sont retrouvés avec des supports maximaux et se divisent en trois grands clades, dont deux correspondent aux sous-sections Thoracotremata et Heterotremata (Guinot 1977 ; Ng *et al.*, 2008 ; Tsang *et al.*, 2014). Le premier clade (A, figure III.4) contient les familles Portunidae, Geryonidae, Matutidae, Menippidae, Bythograeidae, Xanthidae, Majidae et Oregoniidae qui font partie de la sous-section des Heterotremata (Zhang *et al.*, 2011 : 265). Toutes ces familles se situent dans le clade des Heterotremata pour l'étude de Tsang *et al.* (2014). Un second clade (B, figure III.4) contient les familles Varunidae, Macrophthalmidae, Sesarmidae, Xenograpsidae, Grapsidae, Dottilidae, Ocypodidae, Myctiridae, qui font partie de la sous-section des Thoracotremata. Le troisième clade (C, figure III.4) est constitué des espèces appartenant aux familles Potamidae et Gecarcinucidae et se positionne comme groupe frère du clade des Thoracotremata avec un soutien maximal alors qu'elles sont placées, dans la classification en cours, dans les Heterotremata. Cette incongruence entre la phylogénie et la classification a déjà été observée dans une étude de phylogénie qui utilise les mêmes marqueurs que notre étude avec un échantillonnage similaire (Yuhui *et al.*, 2017). La question du positionnement de ces deux familles ainsi que de la famille Potamonautidae (non représentée dans les données de mitogénomes), a déjà été soulevée par une étude morphologique qui proposait également de les placer au sein des Thoracotremata (Von Sternberg et Cumberlidge, 2001). L'étude de Tsang *et al.* (2014) identifie également un clade composé des trois familles Potamidae, Potamonautidae et Gecarcinucidae mais ne résout pas de façon soutenue sa position relativement aux autres Heterotremata. Ce résultat suggère qu'avec cet échantillonnage des lignées de Brachyura, l'approche de phylogénie mitogénomique permet de résoudre les relations à cette échelle de profondeur qui n'étaient pas résolues avec le jeu de marqueurs nucléaires utilisé par Tsang *et al.* (2014). Néanmoins, ce résultat pourrait être un artefact résultant de l'échantillonnage taxonomique très incomplet pour les mitogénomes par rapport à celui utilisé par ces auteurs.

Tableau III.3 Diversité familiale des Brachyura (Ng *et al.*, 2008 ; Zhang *et al.*, 2011 : 165). Le tableau indique le nombre d'espèces des familles échantillonnées et précise le statut monophylétique de chacune de ces familles, dans l'étude de Tsang *et al.* (2014) et dans la présente étude. La monophylie d'une famille multi-générique/mono-générique est testée si au moins deux genres/deux espèces de cette famille sont échantillonnés. « ? » indique que la monophylie de la famille ne peut être testée par l'échantillonnage de l'étude. * Monophylétique si l'espèce *Panopeus herbstii* (Panopeidae) est incluse dans les Xanthidae.

| Section | Sous-section | Super famille | Famille | Nombre espèces Tsang <i>et al.</i> (2014) | Monophylie famille dans Tsang <i>et al.</i> (2014) | Nombre espèces (présente étude) | Monophy -lie famille (présente étude) | |
|----------------|-----------------|-----------------|--------------------|---|--|---------------------------------|---------------------------------------|--|
| Eubrachyura | Heterotremata | Aethroidea | Aethridae | 1 | ? | | | |
| | | Bellioidea | Belliidae | 2 | ? | | | |
| | | Bythograeoidea | Bythograeidae | | | 5 | oui | |
| | | Calappoidea | Calappidae | 1 | ? | | | |
| | | | Matutidae | 1 | ? | 1 | ? | |
| | | Cancroidea | Atelecyclidae | | | | | |
| | | | Cancridae | 2 | oui | | | |
| | | Carpilioidea | Carpiliidae | 2 | oui | | | |
| | | Cheiragonoidea | Cheiragonidae | | | | | |
| | | Corystoidea | Corystidae | 2 | oui | | | |
| | | | Dacryopilumnidae | | | | | |
| | | Dairoidea | Dairidae | | | | | |
| | | | Dorippoidea | Dorippidae | 3 | oui | | |
| | | | Ethusidae | 1 | ? | | | |
| | | | Eriphioidea | Dairodidae | | | | |
| | | Eriphiidae | | 2 | ? | | | |
| | | Hypothalassidae | | | | | | |
| | | Menippidae | | 2 | non | 2 | oui | |
| | | Oziidae | | 3 | oui | | | |
| | | | Platyxanthidae | | | | | |
| | | Gecarcinucoidea | Gecarcinucidae | 4 | oui | 1 | ? | |
| | | Goneplacoidea | Acidopsidae | | | | | |
| | | | Chasmocarcinidae | 1 | ? | | | |
| | | | Conleyidae | | | | | |
| | | | Euryplacidae | 2 | ? | | | |
| | | | Goneplacidae | 2 | ? | | | |
| | | | Litocheiridae | | | | | |
| | | | Mathildellidae | 1 | ? | | | |
| | | | Neommatocarcinidae | | | | | |
| | | | Progeryonidae | | | | | |
| | | Hexapodoidea | Scalopidiidae | 1 | ? | | | |
| | | | Vultocinidae | 1 | ? | | | |
| | | Hexapodoidea | Hexapodidae | | | | | |
| | | Leucosioidea | Iphiculidae | | | | | |
| | | | Leucosiidae | 4 | oui | 1 | ? | |
| | | Majoidea | Epialtidae | 3 | non | | | |
| | | | Hymenosomatidae | | | | | |
| | | | Inachidae | 1 | ? | | | |
| | | | Inachoididae | | | | | |
| | | | Majidae | 3 | non | 3 | oui | |
| | | | Oregoniidae | | | 1 | ? | |
| | | Orithyoidea | Orithyiidae | 1 | ? | | | |
| Palicoidea | Crossotonotidae | | | | | | | |
| Palicoidea | Palicidae | | | | | | | |
| Parthenopoidea | Parthenopidae | 2 | oui | | | | | |
| Pilumnoidea | Galenidae | 1 | ? | | | | | |
| | Pilumnidae | 3 | non soutenu | | | | | |
| | Tanaochelidae | 1 | ? | | | | | |
| Portunoidea | Carcinidae | | | | | | | |
| | Catopridae | | | | | | | |
| | Geryonidae | | | 2 | ? | | | |
| | Macropipidae | | | | | | | |
| | Primelidae | | | | | | | |

| | | | | | | |
|------------------|---------------------|--------------------|-----------------------|-----------------------|-----|-----|
| | | Polybiidae | 2 | non soutenu | | |
| | | Portunidae | 4 | oui | 13 | oui |
| | | Thiidae | | | | |
| | Potamoidea | Potamidae | 6 | oui | 7 | oui |
| | Potamoidea | Potamonautidae | 3 | oui | | |
| | Pseudothelphusoidea | Pseudothelphusidae | 4 | oui | | |
| | Pseudozoidea | Caecopilumnidae | | | | |
| | | Pilumnoididae | | | | |
| | | Planopilumnidae | | | | |
| | | Pseudoziidae | | | | |
| | Retroplumoidea | Retroplumidae | 1 | ? | | |
| | Trapezioidea | Domeciidae | | | | |
| | | Tetraliidae | | | | |
| | Trapezioidea | Trapeziidae | 3 | oui | | |
| | Trichodactyloidea | Trichodactylidae | 3 | oui | | |
| | Xanthoidea | Panopeidae | 1 | ? | | |
| | | Pseudorhombilidae | | | | |
| | | Xanthidae | 9 | non* | 3 | oui |
| Thoracotremata | Cryptochiroidea | Cryptochiridae | 1 | ? | | |
| | Grapsoidea | Gecarcinidae | 4 | oui | | |
| | | Glyptograpsidae | 2 | oui | | |
| | | Grapsidae | 4 | oui | 2 | oui |
| | | Percnidae | 2 | oui | | |
| | | Plagusiidae | 1 | ? | | |
| | | Sesarmidae | 3 | oui | 5 | oui |
| | | Varunidae | 3 | oui | 9 | oui |
| | Xenograpsidae | 1 | ? | 2 | oui | |
| | Ocypodoidea | Camptandriidae | | | | |
| | | Dotillidae | 6 | oui | 1 | ? |
| | | Heloeiidae | | | | |
| | | Macrophthalmidae | 2 | non | 1 | ? |
| | | Mictyridae | 1 | ? | 1 | ? |
| | | Ocypodidae | 5 | oui | 3 | oui |
| | Pinnotheroidea | Ucididae | | | | |
| | | Xenophthalmidae | | | | |
| | | Pinnotheroidea | Aphanodactylidae | | | |
| | | Pinnotheroidea | Pinnotheridae | 2 | oui | |
| | | NA | NA | Brankocleistostomidae | | |
| | NA | NA | Garthopilumnidae | | | |
| | NA | NA | Lazarocleistostomidae | | | |
| Cyclodorip-poida | Cyclodorippoidea | Cyclodorippidae | 1 | ? | | |
| | | Cymonomidae | 2 | ? | | |
| | | Phyllotymolinidae | | | | |
| Dromiacea | Dromioidea | Dromiidae | 3 | oui | | |
| | | Dynomenidae | 2 | oui | 1 | ? |
| | Homolodromioidea | Homolodromiidae | | | | |
| | Homoloidea | Homolidae | 5 | non | 3 | oui |
| | | Latreilliidae | 1 | ? | | |
| | | Poupiniidae | | | | |
| Raninoidea | Raninoidea | Raninidae | 3 | oui | 3 | oui |

Notre échantillonnage permet d'observer la résolution apportée par les génomes mitochondriaux à une échelle plus récente : en particulier au sein des familles Varunidae, Potamidae, Portunidae ou Bythograeidae, car elles ont respectivement 9, 6, 13 et 6 espèces pour les représenter. La hiérarchisation des espèces dans ces familles est cohérente avec la taxonomie morphologique : pour toutes ces familles, les genres sont monophylétiques et les soutiens sont maximaux à la fois en vraisemblance et en bayésien. Dans les Varunidae, les relations entre les genres ne sont pas soutenues, mais elles le sont dans les familles Bythograeidae, Potamidae et Portunidae. Ce résultat indique que les génomes mitochondriaux permettent à la fois de résoudre les relations phylogénétiques à une échelle récente (intra-genre) et dans la plupart des cas de résoudre les relations phylogénétiques entre les genres. A une échelle intermédiaire : les relations entre les familles sont majoritairement soutenues dans les clades A et C mais pas dans le clade B (figure III.4). Cependant, notre échantillonnage en nombre de familles est trop faible pour pouvoir conclure sur la résolution des relations à cette profondeur phylogénétique.

De plus, dans notre étude, les 13 lignées correspondant aux familles pour lesquelles la monophylie est testable avec notre échantillonnage taxonomique (voir tableau III.4) sont monophylétiques. Parmi ces lignées, 11 sont également présentes dans l'étude de Tsang *et al.* (2014) avec un échantillonnage leur permettant également de tester la monophylie au niveau de la famille (tableau III.4). Nos résultats sont similaires pour les familles Grapsidae, Ocypodidae, Portunidae, Potamidae, Raninidae, Sesarnidae et Varunidae pour lesquelles les monophylies sont retrouvées dans les deux études. Cependant, nos résultats obtenus avec les données mitochondriales sont en contradiction avec ceux de Tsang *et al.* (2014) pour les familles Homolidae, Majidae, Menippidae et Xanthidae, pour lesquelles ils ne retrouvent pas de monophylie. Ces contradictions peuvent s'expliquer soit par le fait que nous n'utilisons pas les mêmes marqueurs et que les histoires évolutives de ces marqueurs diffèrent, soit ces différences reflètent un artéfact dans l'échantillonnage taxonomique entre les deux études. Dans le cas de la famille Homolidae, les cinq espèces d'Homolidae sont en paraphylie avec l'espèce *Eplumula phalangium* (Latreillidae) dans l'étude de Tsang *et al.* (2014), tandis que nous n'avons pas cette dernière espèce (ni de représentant de cette famille) dans notre jeu de données.

Les auteurs trouvent que les Majidae sont en paraphylie avec la famille Epialtidae, qui est également absente de notre échantillonnage. Dans le cas de la famille Menippidae, les auteurs de l'étude ont l'espèce *Pseudocarcinus gigas* (présente dans notre échantillonnage) et *Menippe rumphii* (absente de notre échantillonnage), qui ne se regroupent pas. Notre jeu de données contient l'espèce *Myomenippe fornasini*, qui forme un clade avec *P. gigas*. Enfin, les auteurs de l'étude

trouvent l'espèce *Panopeus herbstii* (Panopeidae) au milieu des neuf espèces de Xanthidae de leur jeu de données, tandis que nous n'avons pas cette espèce dans notre échantillonnage. Ces familles telles qu'elles sont actuellement définies sont donc certainement polyphylétiques, et certaines espèces devraient être transférées dans d'autres familles, ce que l'échantillonnage de Tsang *et al.* (2014) montre, contrairement à nos données. Ces résultats appuient l'hypothèse que les contradictions retrouvées entre l'arbre de Tsang *et al.* (2014) et notre arbre sont occasionnées par un échantillonnage moins large dans notre étude. De plus, dans cet arbre, le positionnement des Bythograeidae au sein des Heterotremata *s.s.* est cohérent avec les résultats obtenus en intégrant les Bythograeidae dans le jeu de données de Tsang *et al.* (2014) (figure III.1).

Par ailleurs, notre analyse phylogénétique contient trois familles qui sont absentes dans l'étude de Tsang *et al.* (2014) : les Bythograeidae, les Geryonidae et les Oregoniidae. La monophylie de la famille des crabes hydrothermaux profonds Bythograeidae est retrouvée avec des soutiens maximaux, en groupe frère des Xanthidae. Dans la phylogénie basée sur des génomes mitochondriaux de Yang *et al.* (2013), l'espèce retrouvée en groupe frère des Bythograeidae est *Pseudocarcinus gigas* (Menippidae). Cependant leur échantillonnage des Heterotremata se limite aux familles Bythograeidae, Menippidae, Portunidae, Potamidae, Varunidae et Xenograpsidae et n'inclut pas les espèces à position intermédiaire entre les Bythograeidae et *P. gigas*, présentes dans notre jeu de données. Notre analyse phylogénétique divise les Bythograeidae dans trois lignées soutenues sans ambiguïté : la première contient les espèces *Austinograea alayseae* et *A. rodriguezensis*, la deuxième comprend les espèces *Gandalfus yunohana* et *G. puia* et la troisième n'est constituée que de *S. mesatlantica*, sœur de toutes les autres espèces de Bythograeidae dont le génome mitochondrial est disponible dans les bases de données. Cette topologie est cohérente avec celle retrouvée dans l'arbre figure III.1, et n'est pas en contradiction avec celle obtenue par Mateos *et al.* (2012), en utilisant des marqueurs mitochondriaux et nucléaires. L'arbre que les auteurs présentent dans l'article est obtenu par l'analyse concaténée de l'ensemble des marqueurs ; cependant, n'ayant pas retrouvé la monophylie de la famille avec les différents gènes dont ils disposent, ils enracinent l'arbre obtenu sur les données concaténées avec le clade des *Bythograea*. Leur analyse ne permet donc, comme dans notre étude, que de discuter des relations entre les Bythograeidae n'appartenant pas au genre *Bythograea*. Notre analyse est effectuée avec un plus grand nombre de marqueurs mitochondriaux et fournit des nœuds soutenus. Elle confirme la proximité des espèces de Bythograeidae non *Bythograea* entre elles et leur position au sein des Heterotremata, à proximité

des Xanthidae. La question de la position du genre *Bythograea*, et donc de la monophylie des Bythograeidae reste pour le moment irrésolue.

Cette analyse phylogénétique apporte des éléments supplémentaires aux questions soulevées dans la section précédente, sur la difficulté à étendre l'échantillonnage des mitogénomes par PCR longues. D'après les analyses morphologiques, les espèces de Bythograeidae se situent dans les Heterotremata *sensu stricto*, ce qui est corroboré par l'analyse phylogénétique effectuée avec les marqueurs H3, NaK, 16S et 12S (figure III.1). Ainsi, les espèces de Bythograeidae se situent dans le clade B (figure III.3) dans l'analyse phylogénétique basée sur les génomes mitochondriaux. Parmi les neuf familles du clade A, seul le genre *Maja* de la famille des Majidae a une organisation des gènes codants différente de l'organisation 1 (figures III.3 et III.4). Cependant, seules 11 des 70 familles d'Heterotremata (Ng *et al.*, 2008) sont représentées ici, il n'est donc pas exclu que d'autres lignées non présentes dans notre étude possèdent une organisation des gènes différente. Par ailleurs, les positions des ARNt diffèrent au sein de ce groupe pour les espèces *Damithrax Spinosissimus* (Majidae) et les deux espèces du genre *Atergatis* (Xanthidae). Ainsi, dans les familles proches des Bythograeidae, l'organisation des gènes codant pour des protéines semble conservée, ce qui faciliterait l'utilisation de l'approche de PCR longues pour élargir l'échantillonnage taxonomique à l'ensemble de ce clade en utilisant cette approche.

III.4 Conclusion et perspectives

Les résultats de phylogénie avec l'ensemble des mitogénomes de *Brachyura* disponibles indiquent que ce type de données est particulièrement intéressant en phylogénie. En effet, les génomes mitochondriaux sont composés de gènes ayant des taux d'évolution variés qui permettent de résoudre les relations à différentes profondeurs phylogénétiques.

Afin de séquencer le génome mitochondrial du crabe hydrothermal profond d'Atlantique *Segonzacia mesatlantica*, deux approches ont été testées dans notre étude : une approche shotgun et une approche par PCR longues. L'approche shotgun n'a pas permis d'assembler directement le mitogénome en raison d'un problème technique lors de la réaction de séquençage, rendant la profondeur de séquençage insuffisante. De plus, nous avons rencontré une autre difficulté due à la présence de Numt représentant près de 30% des reads 'mitochondriaux'. L'approche par PCR longues, elle, requiert une préparation supplémentaire mais a permis d'assembler le génome mitochondrial complet, y compris la région de contrôle.

L'étude de Tsang *et al.* (2014) possède l'échantillonnage taxonomique le plus large au sein des *Brachyura*. Replacer les *Bythograeidae* au sein des espèces présentes dans leur étude permettrait de tester l'hypothèse de la monophylie de la famille et donc de comprendre l'origine évolutive de la colonisation des sources hydrothermales profondes chez les crabes *Brachyura*. Pour ce faire, l'approche shotgun reste valable car son efficacité a déjà été démontrée pour assembler des mitogénomes de *Brachyura* et demande peu de préparation. L'approche par PCR longues est également envisageable, d'autant plus que les *Bythograeidae* se situent dans le clade des *Heterotremata sensu stricto*, dans lequel l'organisation des gènes codant pour des protéines semble conservée. Cette dernière approche requiert aussi un effort de séquençage moins important pour une couverture équivalente à celle du shotgun, permettant d'envisager un échantillonnage sur un plus grand nombre d'espèces.

Chapitre IV

Une approche NGS : le cas
des crevettes de la famille

Alvinocarididae

IV. UNE APPROCHE NGS : LE CAS DES CREVETTES DE LA FAMILLE ALVINOCARIDIDAE

IV.1 Introduction

Les travaux récents attribuent 33 espèces à la famille Alvinocarididae, dont 29 présentes exclusivement à proximité des sources hydrothermales, 2 exclusivement dans les environnements de suintements froids et 2 partagées entre ces 2 environnements (Nye *et al.*, 2012 ; Teixeira *et al.*, 2013 ; Vereshchaka *et al.*, 2015 ; Komai *et al.*, 2016 ; Wang et Sha, 2016, 2017). La famille est inconnue dans tout autre milieu et reste donc considérée comme un des taxons de haut rang spécialisés, ou « endémiques », des sources hydrothermales (McArthur et Tunnicliffe, 1998 ; Wolf, 2005 ; Samadi, 2015). Pour comprendre comment la colonisation vers ces milieux a eu lieu au sein des crevettes Caridea, il faut pouvoir reconstituer les habitats des ancêtres hypothétiques des lignées de Caridea. Cela ne peut être effectué que dans un contexte phylogénétique robuste, c'est-à-dire avec un échantillonnage représentatif du groupe ciblé, couplé à une bonne résolution de l'ensemble des nœuds. L'objectif général de ce travail était donc de contribuer à positionner la famille endémique des Alvinocarididae parmi les familles de Caridea, afin de comprendre l'origine évolutive de la spécialisation de cette lignée aux sources hydrothermales. Dans cette introduction seront abordés, dans un premier temps, l'état de la classification de la famille des Alvinocarididae et de sa position au sein des Caridea et, dans un second temps, la question des outils et méthodes à développer pour parvenir à résoudre sa position phylogénétique au sein des Caridea.

IV.1.1 Systématique des Alvinocarididae

L'espèce type de la famille, *Alvinocaris lusca* Williams et Chace, 1982, a été découverte dès les premières explorations des sites hydrothermaux sur la Ride des Galápagos. Au moment de sa description, l'espèce est placée dans la famille des Bresiliidae. Cette famille, créée lors de la description de l'espèce *Bresilia atlantica* Calman, 1896 à partir d'un unique spécimen récolté au Sud-Ouest de l'Irlande à 1 370 m de profondeur, inclut des espèces rares, d'environnement profonds, difficiles à explorer (Forest, 1977 ; Forest et Cals, 1977 ; Chace et Brown, 1978 ; Wilson et Gore, 1979).

Peu après la découverte d'*A. lusca*, en s'appuyant sur une étude phylogénétique basée sur la morphologie, Christoffersen propose de classer cette espèce dans une nouvelle famille, les

Alvinocarididae Christoffersen, 1986. Dans les années qui suivirent, d'autres espèces de crevettes sont découvertes, principalement au niveau des sources hydrothermales mais également sur des sites de suintements froids. Dans un premier temps, les auteurs de ces nouvelles espèces ne discutent généralement pas de la nouvelle famille proposée par Christoffersen et continuent de les placer dans la famille des Bresiliidae (Williams et Rona, 1986 ; Williams, 1988 ; Martin et Hessler, 1990 ; Martin et Christiansen, 1995 ; Williams et Dobbs, 1995). Cette position reflète celle que Chace formalise en 1992 dans sa révision de la classification des Caridea. Cet auteur considère en effet que les descriptions de familles monotypiques, généralement érigées pour classer des espèces de crevettes rares et extrêmement spécialisées, reposent sur un nombre insuffisant d'espèces et de spécimens pour être définies de manière satisfaisante. Parmi ces taxons, l'espèce *Discias serrifer* Rathburn, 1902 a été décrite à partir de trois spécimens et positionnée en raison d'une morphologie singulière dans un nouveau genre et une nouvelle famille : les Disciidae. La découverte d'une nouvelle espèce, *Pseudocheles enigma* Chace et Brown, 1978, également décrite à partir de trois spécimens et positionnée dans un nouveau genre, amène les auteurs à discuter des contours des Disciidae et Bresiliidae. Les auteurs montrent que les caractères de cette nouvelle espèce la rapprochent à la fois des Disciidae et des Bresiliidae. Ils proposent donc de synonymiser ces deux familles afin d'éviter l'inflation taxonomique. De même, la découverte dans des grottes anchialines d'une nouvelle espèce, conduit Hart et Manning (1986) à décrire une nouvelle famille monotypique Agostocarididae, proche des Bresiliidae. Ici encore Chace (1992) suggère que les données sont trop réduites pour faire figurer cette nouvelle famille dans sa classification des Caridea.

Les explorations des milieux hydrothermaux et des suintements froids ont permis d'accroître le nombre d'espèces d'Alvinocarididae. Ainsi en 1997, lorsque Vereshchaka décrit la nouvelle espèce *Mirocaris keldyshi* Vereshchaka, 1997, il reconsidère la classification au rang familial de ces espèces. Il place ainsi les genres *Alvinocaris*, *Chorocaris* (Martin et Hessler, 1990), *Opapele* Williams et Dobbs, 1995 et *Rimicaris* Williams et Rona, 1986 dans la famille des Alvinocarididae dont il redéfinit les contours. Dans cet article, Vereshchaka propose également de classer *M. keldyshi* dans une nouvelle famille : les Mirocarididae. En 2003, Komai et Segonzac, synonymisent cette nouvelle famille avec les Alvinocarididae et précisent de nouveau les contours de la famille. Il faut souligner que, contrairement à l'étude de Christoffersen, ces révisions ne sont pas fondées sur une approche phylogénétique, que ce soit avec des caractères morphologiques ou moléculaires.

Depuis cette révision, cinq genres ont été ajoutés à la famille : *Nautilocaris* Komai et Segonzac, 2004, *Shinkaicaris* (Kikuchi et Hashimoto, 2000), *Alvinocaridinides* Komai et Chan, 2010, *Keldyshicaris* Vereshchaka, Kulagin et Lunina, 2015 et *Manuscaris* Komai et Tsuchida, 2015. Parmi ces neuf genres, quatre sont monospécifiques et trois ne comprennent que deux espèces. Ce découpage

repose sur des caractères morphologiques et a récemment été rediscuté dans une révision taxonomique par Vereshchaka *et al.* (2015). Basé sur une analyse de phylogénie moléculaire (voir encadré 2 pour plus de détails), les auteurs de cet article s'intéressent aux relations entre les espèces au sein de la famille mais ne testent pas la monophylie de la famille ni *a fortiori* son positionnement dans la superfamille des Bresilioidea ou au sein des autres familles de Caridea.

IV.1.2 Position des Alvinocarididae au sein des Caridea

IV.1.2.1 *Diversité des Caridea*

La famille Alvinocarididae s'inscrit dans l'infra-ordre des crevettes Caridea. Les Caridea sont caractérisées par les caractères morphologiques suivants : (i) les femelles portent leurs œufs jusqu'à l'éclosion ; (ii) absence de pince sur le 3^{ème} péréiopode ; (iii) la pleurite du second segment (ou métamère) chevauche celles du 1^{er} et du 3^{ème} (Chan, 1998), même si la prise en compte d'autres caractères morphologiques a été proposée (De Grave et Goulding, 2011). L'infra-ordre des Caridea est un groupe très diversifié, de par le nombre d'espèces présentes et de par la remarquable diversité d'adaptations que l'on peut mettre en relation avec l'hétérogénéité des niches écologiques occupées (Bracken *et al.*, 2009a ; De Grave et Moosa, 2004). Cet infra-ordre comporte environ 3 500 espèces valides réparties en près de 390 genres (De Grave et Fransen, 2011). C'est le second groupe de décapodes le plus diversifié après les crabes Brachyours (~7 000 espèces d'après Guinot *et al.*, 2013). Les espèces de Caridea sont connues sous toutes les latitudes (des milieux polaires aux milieux tropicaux, *e.g.* Clarke, 1993 ; Ditter *et al.*, 2015) et dans tous les types de milieux aquatiques (marins, eaux douces ou saumâtres *e.g.* Von Rintelen *et al.*, 2012), depuis la surface jusqu'aux grandes profondeurs (*e.g.* Chan *et al.*, 2010), dans les environnements chimiosynthétiques des suintements froids et des sources hydrothermales (*e.g.* Vereshchaka *et al.*, 2015), dans des grottes anchialines (*e.g.* Alvarez *et al.*, 2005), dans les eaux douces lacustres (*e.g.* De Grave *et al.*, 2008) ou les torrents de montagne (*e.g.* Richardson et Cook, 2006). Certaines espèces du genre *Merguia* sont même semi-terrestres (Baeza 2010). Les espèces de la famille des Alvinocarididae ne sont pas les seules espèces de Caridea connues des milieux chimiosynthétiques. En effet, 22 espèces de 8 familles différentes (Crangonidae, Glyphocrangonidae, Hippolytidae, Nematocarcinidae, Ophlophoridae, Palaemonidae, Pandalidae et Stylodactylidae) sont également identifiées dans ces milieux (Martin et Haney, 2005). Cependant, ces lignées ne sont pas inféodées aux environnements chimiosynthétiques car les autres espèces de ces huit familles sont présentes dans d'autres environnements (De Grave et Fransen, 2011).

Encadré 2 : Relations au sein de la famille des Alvinocarididae

Les données moléculaires analysées dans la révision de Vereshchaka *et al.*, 2015 ont été complétées avec l'ensemble des séquences disponibles dans les bases de données publiques ainsi que par les données disponibles pour des spécimens conservés au MNHN. L'analyse phylogénétique de ces données supplémentaires permet de mettre à jour les résultats de cet article et de fournir un arbre mieux résolu en analyse bayésienne (figure IV.1).

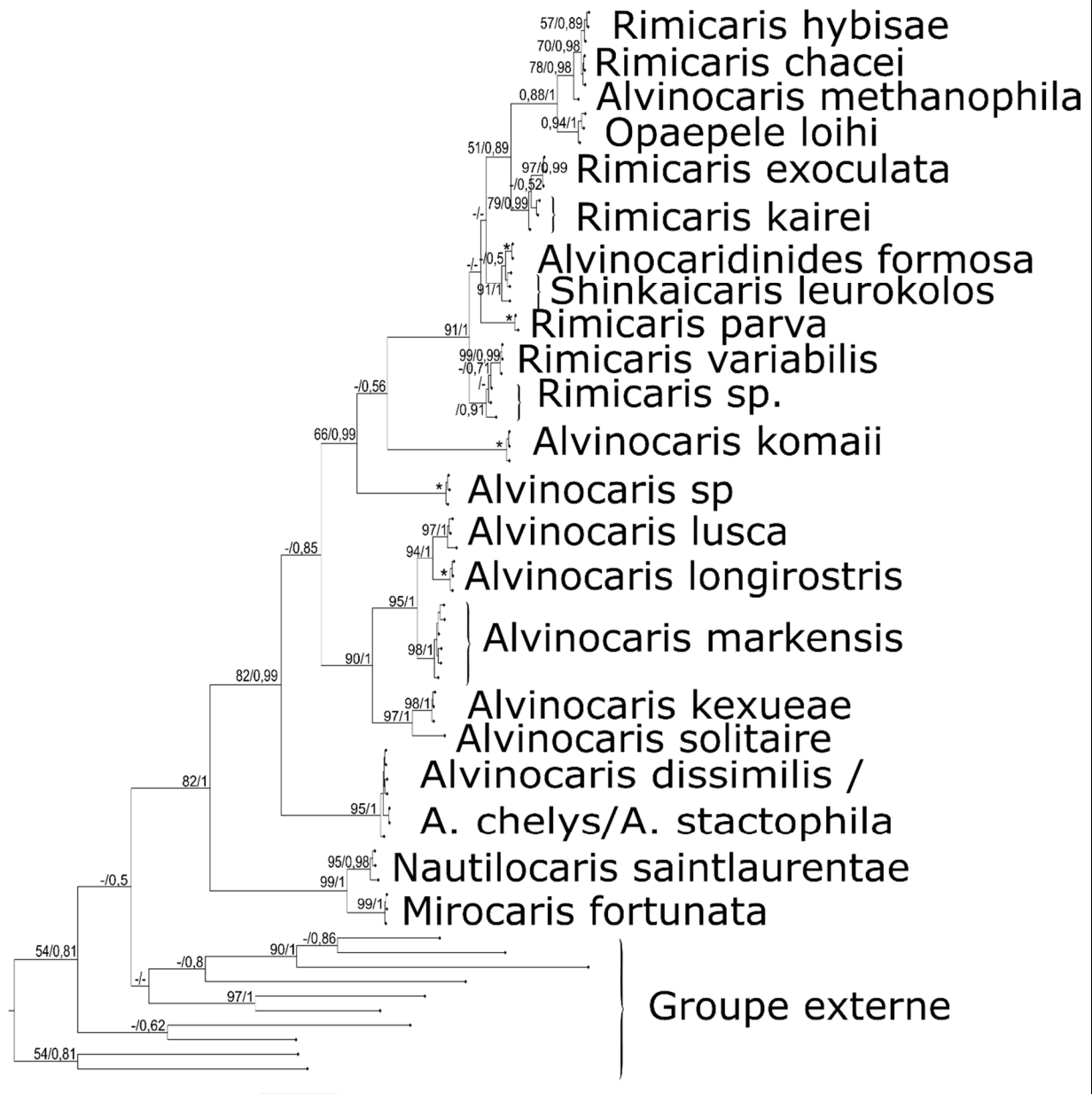


Figure IV.1 Topologie de l'inférence bayésienne (IB) des relations phylogénétiques entre espèces d'Alvinocarididae déterminées par l'alignement du marqueur COI. Seules trois séquences par clade COI soutenu ont été conservées pour l'analyse finale. Les analyses ML et IB ont été faites avec RAXML v.8.2.9 et MrBayes v.3.2.6 respectivement, en utilisant le modèle GTR+Γ+I pour chaque position de codon. Un rééchantillonnage est effectué avec 1 000 bootstraps dans le cas de l'analyse ML. L'analyse bayésienne a été effectuée avec 2 10⁶ générations. Un astérisque indique un soutien maximal pour les deux analyses. Un tiret indique un soutien inférieur à 50% bootstraps ou 0,5 en probabilité postérieure.

Cette reconstruction phylogénétique confirme que le genre *Mirocaris* est positionné en groupe frère de tous les autres Alvinocarididae, dans un clade qui inclut le genre *Nautilocaris*. Elle permet aussi de souligner que les incongruences entre les résultats moléculaires et morphologiques ne sont pas traitées de la même façon pour tous les taxons. Par exemple : Vereshchaka *et al.*, 2015 proposent de synonymiser les genres *Rimicaris* et *Chorocaris* qui forment un clade bien défini, mais conservent *Opaepele loihi* dans un genre distinct, bien que cette espèce se positionne aussi dans ce clade. Dans le cas de *Rimicaris* et *Chorocaris*, les auteurs justifient la synonymie à la fois par l'analyse morphologique et par la proximité génétique de ces différentes espèces, également mise en évidence avec d'autres marqueurs (Teixeira *et al.*, 2013).

Notre analyse permet également de discuter l'hypothèse proposée par Shank *et al.* (1998), selon laquelle les zones de suintements froids auraient servi d'environnement tremplin dans la colonisation du milieu hydrothermal. Cette reconstruction phylogénétique sur les données de COI contredit cette hypothèse car les espèces des suintements froids (*Alvinocaris longirostris*, *A. markensis*, *A. methanophila* et *A. stactophila*) ne forment pas un groupe monophylétique au sein des Alvinocarididae (figure IV.1). Cet arbre reste cependant incomplet car toutes les espèces connues ne sont pas représentées, notamment les espèces des genres *Keldyshicaris* et *Manuscaris*. De plus, l'arbre n'est reconstruit qu'à partir d'un seul marqueur et pourrait ne pas refléter la phylogénie des espèces. Reconstruire une phylogénie à partir d'un second marqueur indépendant du premier permettrait de comparer les deux topologies obtenues et d'avoir une meilleure confiance dans l'histoire évolutive de ces lignées, si les topologies reconstruites sont identiques. Par ailleurs, l'ensemble des nœuds de l'arbre obtenu (figure IV.1) n'ont pas de soutiens maximaux. Obtenir une phylogénie à partir d'une concaténation de marqueurs permettrait également d'augmenter la résolution des relations qui ne sont pas encore bien définies.

Depuis De Grave *et al.* (2009), plusieurs modifications de la classification des Caridea ont été proposées et conduisent à comptabiliser 40 familles ayant des représentants actuels au sein de cet infra-ordre (Aznar-Cormano *et al.*, 2015 ; Bracken *et al.*, 2010 ; Chan *et al.*, 2010 ; De Grave *et al.*, 2010, 2014 ; Mejia-Ortiz *et al.*, 2017 ; Short *et al.*, 2013 ; Burukovsky, 2012). Parmi ces familles, plusieurs sont très diversifiées à la fois en nombre d'espèces (*e.g.* 981 espèces de Palaemonidae, 663 d'Alpheidae) et en type d'environnement. Par exemple, les Atyidae sont des espèces amphidromes (dont les différents stades de vies se déroulent à la fois dans des eaux douces et marines) ou encore les Palaemonidae, qui sont composés d'espèces d'eaux douces mais également d'espèces marines (De Grave et Fransen, 2011). Seules cinq familles de Caridea sont décrites comme spécialisées et associées à des environnements très restreints. Les Alvinocarididae ne sont retrouvés que dans les environnements chimiosynthétiques. Trois familles, peu diversifiées, ne regroupent chacune que des espèces spécialisées des milieux anchialins : les familles Agostocarididae (un genre, trois espèces, placée dans les Bresilioidea) et Barbouriidae (trois genres et huit espèces, placée dans les Alpheoidea) (De Grave et Fransen, 2011), auxquelles s'ajoute la famille Anchialocarididae,

récemment décrite et composée d'une seule espèce (Mejia-Ortiz *et al.*, 2017). Le positionnement de cette dernière famille monospécifique n'est pas résolu par les auteurs qui la rapprochent des Agostocarididae mais souligne également une proximité avec les Atyidae. Enfin la famille Lipkiidae (placée dans les Nematocarcinoidea) avec deux genres monospécifiques est proposée pour classer deux espèces de très grande taille, retrouvées au niveau des milieux de substrats durs profonds subantarctiques (d'Acoz et De Grave, 2018). Toutes ces familles sont proposées pour classer de rares espèces, à la fois récoltées dans des milieux difficiles d'accès (ou peu explorés) et dont la morphologie est très divergente de celle connue chez les autres Caridea.

IV.1.2.2 Le contexte phylogénétique

La monophylie de l'infra-ordre est bien établie (*e.g.* Aznar-Cormano *et al.*, 2015 ; Bracken *et al.*, 2009b, 2010 ; Shen *et al.*, 2013). Une étude phylogénétique récente a tenté de résoudre les relations phylogénétiques à l'échelle des Caridea en combinant cinq marqueurs à un échantillonnage taxonomique comprenant 94 espèces de Caridea (Aznar-Cormano *et al.*, 2015). Cette étude visait à compléter les études précédentes qui incluaient soit un grand nombre de taxons (Bracken *et al.*, 2009b) soit un grand nombre de marqueurs (Li *et al.*, 2011). C'est à ce jour l'étude la plus complète à la fois du point de vue de la couverture taxonomique des familles de Caridea et du point de vue du nombre de marqueurs utilisés. Leur échantillonnage taxonomique permet de tester la monophylie de 13 familles, et celle-ci n'est pas retrouvée pour les familles Lysmatidae, Hippolytidae (*sensu stricto*) et Palaemonidae. Dans cette étude cependant, les relations interfamiliales ne sont pas résolues. Néanmoins, l'analyse bayésienne a différencié deux clades majeurs : le clade I composé des 4 familles tandis que le clade II est beaucoup plus diversifié (21 familles, dont les Alvinocarididae). Dans cette étude, la famille des crevettes hydrothermales Alvinocarididae est représentée par huit espèces appartenant à cinq genres différents. Grâce à l'échantillonnage large des autres familles de Caridea, les auteurs confirment la monophylie de la famille des Alvinocarididae par une approche moléculaire. La monophylie de cette lignée confirme que la colonisation du milieu hydrothermal correspond à un événement unique, suivi d'une diversification spécifique. L'absence de résolution entre les lignées de rang familial ne permet cependant pas de discuter de l'origine de cette colonisation.

Une étude phylogénétique plus récente a été proposée, qui s'intéresse à la famille des Pasiphaeidae (Liao *et al.*, 2017). Cette étude inclut 33 espèces de Pasiphaeidae, et entre 1 et 3 espèces parmi 16 autres familles de Caridea, dont les Alvinocarididae, pour constituer le groupe externe de leur analyse. L'arbre phylogénétique est obtenu par l'analyse combinée de six marqueurs

phylogénétiques (12S, 16S, H3, NaK, l'énolase et l'ATP β). Cette étude propose d'ajouter le marqueur ATP β à l'étude des relations des familles de Caridea. Cependant, la pertinence de ce marqueur ne peut être correctement évaluée en dehors des Pasiphaeidae, dans la mesure où la grande majorité des données concernant le groupe externe est tirée des bases de données publiques. Les données pour ce gène ne sont disponibles que pour trois espèces du groupe externe, correspondant à trois familles distinctes (dont une espèce d'Alvinocarididae). L'analyse de l'alignement des séquences d'ATP β montre que le gène est très variable avec une divergence maximale observée de 0,18 entre deux espèces de Pasiphaeidae (*Psathyrocaris hawaiiensis* et *Leptochela japonica*). Les auteurs utilisent également le marqueur NaK. Ce marqueur a été exclu de l'étude d'Aznar-Cormano *et al.* (2015) en raison de l'importante saturation à l'échelle des Caridea, corrélée à une difficulté d'amplification par PCR pour tous les taxons. Les auteurs de Liao *et al.* (2017) n'utilisent pas les marqueurs 18S et 28S qui sont pourtant amplifiables aisément à l'échelle des Caridea (Aznar-Cormano *et al.*, 2015). La topologie de l'arbre obtenu par Liao *et al.* (2017) résulte donc probablement de données manquantes distribuées de façon hétérogène entre les groupes interne et externe, et d'un choix de marqueurs saturés à l'échelle des Caridea. Cette analyse suggère que le groupe de Pasiphaeidae pourrait être groupe frère des Alvinocarididae. Il faudrait confirmer ce résultat en étendant l'échantillonnage du gène ATP β à l'ensemble du groupe externe et en ajoutant les gènes inclus dans l'étude d'Aznar-Cormano *et al.* (2015) pour l'ensemble des espèces. La topologie obtenue pour les espèces échantillonnées dans le groupe externe de cette étude diffère de celle d'Aznar-Cormano *et al.* (2015). Notamment, les clades I et II d'Aznar-Cormano *et al.* (2015) ne sont pas retrouvés.

Si ces deux études (Aznar-Cormano *et al.*, 2015 ; Liao *et al.*, 2017) ont apporté des informations concernant les relations internes de plusieurs familles de Caridea, la question du groupe frère des Alvinocarididae reste encore non résolue.

IV.1.2.3 La transcriptomique pour sous-échantillonner le génome de manière ciblée

L'ensemble des marqueurs testés à ce jour à l'échelle des Caridea par amplification PCR et séquençage Sanger ne fournissent pas un signal phylogénétique suffisamment profond pour résoudre les nœuds correspondant aux relations interfamiliales. L'objectif du travail de cette thèse a donc été de développer de nouveaux marqueurs par des approches utilisant les nouvelles technologies de séquençage et de tester leur résolution phylogénétique à cette échelle taxonomique.

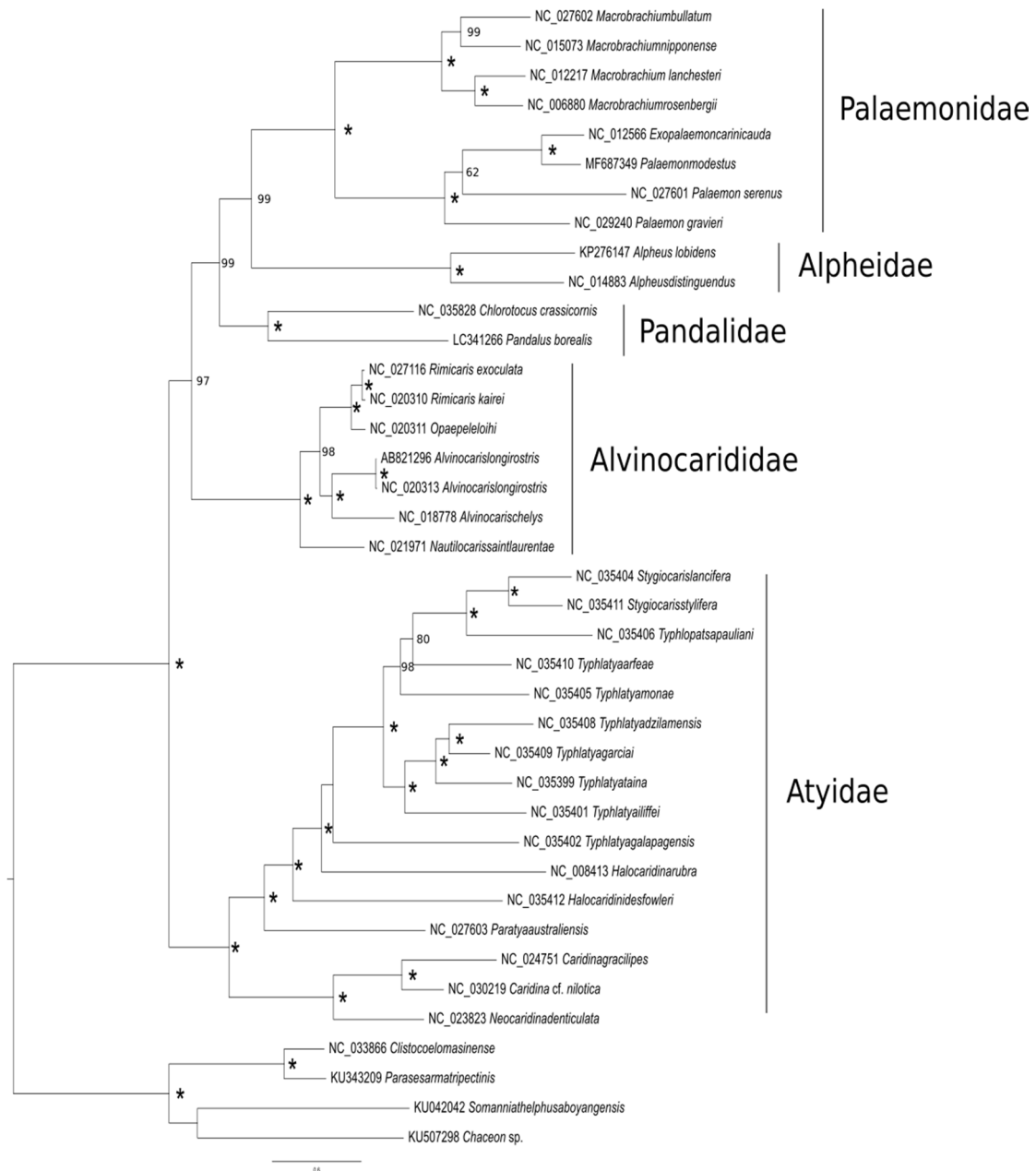


Figure IV.2 Topologie RAxML des relations obtenues en maximum de vraisemblance à partir de la concaténation des séquences mitochondriales des 13 gènes codant pour des protéines et des ARN ribosomiaux 12S et 16S. Le modèle GTR +I +A a été appliqué pour chaque partition (par gène et par position de codon pour les gènes codants) et un rééchantillonnage de 1 000 bootstraps a été utilisé. Un astérisque dénote d'un soutien maximal (100%). Les soutiens inférieurs à 50 bootstraps ne sont pas affichés. Les familles dans lesquelles sont placées les espèces sont indiquées. La famille des Atyidae a été sous-échantillonnée : un seul spécimen d'*Halocaridina rubra* et 7 des 13 espèces du genre *Typhlatya* ont été inclus dans l'analyse.

L'approche choisie par l'équipe avant le démarrage de cette thèse était de développer un jeu de marqueurs conservés, définis à partir de données de séquençage de transcriptomes d'un échantillonnage de familles de Caridea. Le choix de cette approche a été fait après évaluation de plusieurs stratégies possibles.

(1) Approche mitogénomique. Le séquençage de mitogénomomes complets est réalisable à grande échelle taxonomique et peu coûteuse (Hisinger *et al.*, 2015). L'analyse phylogénétique (figure IV.2) montre que le génome mitochondrial semble avoir une résolution suffisante pour inférer les relations interfamiliales des Caridea, et a déjà été utilisée pour discuter de la colonisation des sources hydrothermales par les Caridea (Yang *et al.*, 2013). Les deux principaux écueils de l'approche mitogénomique sont que (i) on ne retrace l'histoire que d'un seul locus et (ii) les variations dans les taux d'évolution ou les réarrangements de l'ordre des gènes mitochondriaux peuvent induire des distorsions importantes de cet arbre.

(2) Approche génomique « shotgun ». L'ADN est fragmenté physiquement puis l'ensemble est séquencé de façon aléatoire. Avec un faible effort de séquençage (et donc un coût moindre) cette méthode permet de séquencer des génomes complets s'ils sont de petite taille (*e.g.* génome bactérien) ou de séquencer les régions présentes en multiples copies dans l'extrait d'ADN total, comme notamment les séquences ribosomales (présentes en grand nombre de copies dans le génome nucléaire) et le génome mitochondrial (présent en copies multiples dans les cellules). Dans le cas des Caridea, les données disponibles montrent que la taille de leur génome est très variable et peut atteindre environ 40 Gbases (Bonnivard *et al.*, 2009 ; Rees *et al.*, 2008). En raison de cette variabilité de la taille du génome nucléaire, cette méthode n'a pas été retenue, car il a été estimé que fournir un effort de séquençage suffisant pour s'assurer de couvrir les mêmes régions d'une espèce à l'autre était trop coûteux.

(3) Approche génomique réduite par la méthode « Restriction site Associated DNA » (RAD-Seq). Cette approche vise à réduire le séquençage du génome à un jeu de fragments d'ADN homologues* entre les spécimens comparés. Pour cela, l'ADN génomique est digéré par une enzyme de restriction et les fragments générés sont séquencés. Les sites de restrictions peuvent se situer à la fois dans des régions codantes et non codantes du génome. Cette approche permet ainsi potentiellement d'analyser des régions du génome avec différents niveaux de polymorphisme. Avec cette approche, le niveau de conservation des positions des sites de restriction dans le génome détermine l'échelle possible de comparaison taxonomique des marqueurs. Les études disponibles montrent que cette approche ne permet généralement pas de résoudre des relations très distantes, car le nombre de locus partagés diminue rapidement avec la distance phylogénétique (Wagner *et al.*, 2013 ; Herrera *et*

al., 2015 ; Pante *et al.*, 2015). Ainsi, l'échelle de résolution phylogénétique visée par notre étude ne correspond pas à celle apportée par cette approche.

(4) Approche génomique réduite par transcriptomique. Cette approche cible les parties codantes du génome. Elle est intéressante à grande échelle phylogénétique dans la mesure où les organismes partagent des gènes très conservés, qui sont liés au fonctionnement général des cellules (appelés des gènes de « ménage »). Plusieurs études de phylogénomique ont montré la pertinence de cette approche pour résoudre les relations phylogénétiques à grande échelle, notamment celles réalisées par O'Hara *et al.* (2014) sur les ophiures ou Borner *et al.* (2014) sur les Ecdysozoa. Les transcriptomes peuvent être utilisés en phylogénomique de différentes façons : soit en produisant des transcriptomes pour l'ensemble des taxons ciblés, soit en définissant, à partir des transcriptomes d'un petit nombre de taxons, un sous-ensemble de marqueurs séquencés ensuite par différentes méthodes sur un plus large échantillonnage. C'est cette dernière approche (recherche de nouveaux marqueurs à partir de transcriptomes), plus adaptée au large échantillonnage taxonomique que notre objectif requiert, qui est sélectionnée dans notre étude.

IV.1.3 Objectif de notre étude

La principale difficulté des approches de transcriptomiques est que l'ARN total ne peut être extrait qu'à partir de tissus frais ou conservés avec une méthode adaptée, impliquant notamment de pouvoir immédiatement congeler ou conserver les échantillons dans une solution adéquate (RNA-later). Il est donc difficile de réunir un échantillonnage taxonomiquement large incluant notamment des taxons de milieux difficiles d'accès. En effet, le matériel disponible dans les collections des muséums est conservé en alcool, ce qui ne permet pas la conservation des ARN. La stratégie mise en œuvre a donc été de cibler des espèces représentant la diversité des lignées mises en évidence par l'étude d'Aznar-Cormano *et al.* (2015) afin de les utiliser pour définir des marqueurs pouvant être séquencés par d'autres méthodes sur un échantillonnage plus large.

Au début de ce projet, très peu de données de transcriptomes étaient disponibles dans les bases de données publiques, et celles disponibles appartenaient uniquement à la famille des Palaemonidae. Depuis, de nombreuses données ont été publiées et sont venues appuyer le choix de cette méthode.

Parmi l'ensemble des études phylogénétiques basées sur des approches génomiques (études phylogénomiques), l'une d'elles porte sur la phylogénie de la classe des Ophiuroidea et résout les relations de parenté entre des espèces d'ophiures ayant divergé il y a environ 240 Ma (O'Hara *et al.*, 2014). D'après la reconstruction phylogénétique de Décapodes calibrée à partir de données fossiles de Bracken *et al.* (2010), la divergence des Caridea est estimée à environ 260 Ma. Plutôt que de définir un nouveau jeu de marqueurs liés à l'échantillonnage taxonomique très restreint sur lequel

nous avons pu obtenir des transcriptomes, nous avons donc tenté de retrouver les marqueurs de l'étude sur les Ophiuroidea (O'Hara *et al.*, 2014). En effet, si ces marqueurs sont effectivement présents dans les transcriptomes de Caridea, ils devraient permettre de résoudre les relations au sein de l'infra-ordre. D'autre part, si cette hypothèse est confirmée, cet ensemble de marqueurs pourrait constituer un jeu standardisé, utilisable pour résoudre des relations phylogénétiques profondes chez les métazoaires.

IV.2 Matériel et méthodes

IV.2.1 Echantillonnage

Douze spécimens, répartis parmi les grandes lignées de Caridea définies par la phylogénie d'Aznar-Cormano *et al.* (2015), ont été échantillonnés pour produire des transcriptomes. L'obtention d'un transcriptome implique de fixer rapidement les spécimens dans une solution de RNA-later, ou de les congeler rapidement (en les plongeant dans l'azote liquide ou dans la carboglace) pour préserver les ARN. Pour la fixation au RNA-later les spécimens sont conservés 24h à 4°C dans cette solution puis transférés au congélateur à -20°C. Les spécimens ont été identifiés à l'espèce ou au genre en se basant sur leur morphologie. Sept spécimens ont été échantillonnés lors de campagnes océanographiques : *Rimicaris exoculata* (BICOSE, 2014) ; *Periclimenes brevicarpalis* et *Alpheus* sp. (KAVIENG, 2014) ; *Oplophorus gracilirostris* (MADEEP, 2014), *Heterocarpus* aff. *dorsalis* (SALOMON 2, 2004) ; *Heterocarpus parvispina* (MADEEP, 2014) ; *Glyphocrangon* sp. (SALOMON 2, 2004) ; quatre spécimens ont été échantillonnés à la station biologique de Roscoff (CNRS-UPMC) en 2014 : *Athanas nitescens*, *Caridion steveni*, *Crangon crangon* et *Hippolyte varians* ; et l'espèce *Atyopsis moluccensis* est issue du commerce aquariophile (deux spécimens).

L'échantillonnage taxonomique au sein des Caridea est dans un second temps élargi en recherchant des données complémentaires dans la littérature et dans les bases de données génomiques. Pour ce faire, toutes les études (correspondant aux *projets* sur EMBL) portant sur des Caridea sont recherchées sur la base de données EMBL. Une sélection au sein des données disponibles est effectuée en s'appuyant sur les critères suivants :

- Les données concernent une espèce différente de celles de notre échantillonnage taxonomique ;
- Si plusieurs jeux de données sont disponibles pour une même espèce, les données déjà assemblées ont été utilisées ;
- Dans le cas de données intéressantes du point de vue taxonomique mais non assemblées (reads* de séquençage bruts), seules les données obtenues par la méthode de séquençage par *paired-end* ont été retenues.

IV.2.2 Méthodes

IV.2.2.1 *Extraction et quantification des ARN*

Les ARN totaux des spécimens échantillonnés sont extraits à partir du muscle de l'abdomen. Ce tissu est choisi car sa fonction est bien identifiée et similaire quels que soient les organismes et les environnements considérés. La population d'ARN transcrits chez les différents organismes échantillonnés devrait donc être similaire. Pour les plus petits spécimens identifiés comme *Alpheus* sp., *Athanas nitescens*, *Caridion steveni*, *Hippolyte varians* et *Periclimenes brevicarpalis*, l'ensemble de l'abdomen est utilisé afin d'extraire suffisamment d'ARN. Pour les espèces de plus grande taille, une dissection rapide (afin de limiter la dégradation des ARN en ne laissant pas trop longtemps l'échantillon à température ambiante), permet d'éliminer le tube digestif de la masse des muscles abdominaux. Cette dissection permet notamment de minimiser la contamination par des bactéries ou des fragments de nourriture présents dans le tractus digestif. La dissection est réalisée pour les spécimens identifiés comme *Atyopsis moluccensis*, *Crangon crangon*, *Glyphocrangon* sp., *Heterocarpus* aff. *dorsalis*, *Heterocarpus parvispina*, *Oplophorus gracilirostris* et *Rimicaris exoculata*. Les extractions au Tri-Reagent sont réalisées selon le protocole du fabricant (Sigma Aldrich). Après un broyage dans ce réactif, l'ajout de chloroforme permet de séparer deux phases. Suite à cela, les protéines sont précipitées et se retrouvent à l'interface entre les deux phases, tandis que les ARN se trouvent dans la phase aqueuse. Après récupération de cette phase, les ARN sont précipités par addition d'éthanol pur et centrifugation à 4°C. Après remise en suspension dans de l'eau, la qualité et la concentration des extraits ARN sont déterminées par spectrophotométrie, à l'aide d'un NanoDrop. Le rapport d'absorbance A_{260}/A_{230} sert d'indice de contamination par des polysaccharides ou polyphénols (potentiels perturbateurs des réactions de préparation de banques). Les échantillons ayant une valeur de ce ratio supérieure à 1,5 sont conservés. Le ratio A_{260}/A_{280} est, lui, indicateur de contamination protéique et une valeur minimale de 1,8 sert de critère de qualité. Cette étape a été réalisée avec l'aide de Dario Zuccon (ingénieur dans l'équipe de Sarah Samadi).

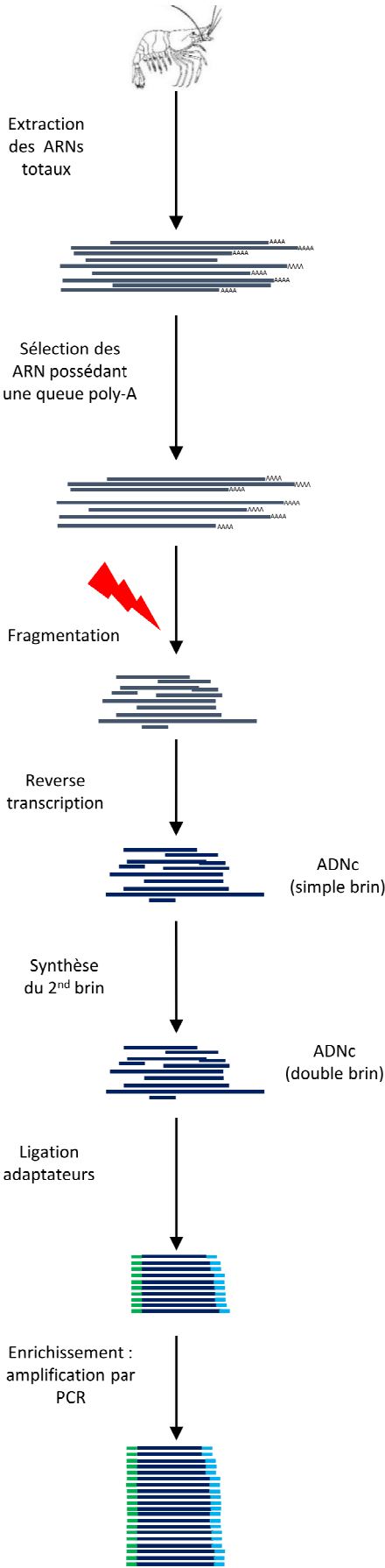


Figure IV.3 Extraction des ARN et préparation des banques pour le séquençage haut-débit Illumina.

IV.2.2.2 Construction des banques pour le séquençage haut-débit Illumina et séquençage

La production des banques est confiée au centre d'innovation Génome Québec et Université McGill. Les banques sont produites à l'aide du kit Illumina TruSeq® RNA ; les fragments d'ADN sont séquencés sur une longueur de 100 pb par les deux extrémités (méthode *paired-end*) sur l'instrument Illumina HiSeq 2000 PE100.

La figure IV.3, résume les différentes étapes de ce protocole. Les ARN messagers (possédant une queue poly-A) sont sélectionnés sur billes magnétiques liées à des séquences poly-T. Une réaction enzymatique en deux étapes permet de rétrotranscrire les ARN sous forme d'ADN, utilisables pour le séquençage. Tout d'abord une retro-transcription des ARN en ADN complémentaire (ADNc) par une transcriptase inverse suivie de la synthèse du brin complémentaire à l'aide d'une polymérase. Les molécules d'ADN ainsi obtenues sont fragmentées par la chaleur en présence de cations divalents. Des adaptateurs, nécessaires pour le séquençage, sont liés aux fragments d'ADN. Ces adaptateurs permettent de fixer les fragments à la puce de séquençage et d'initier la réaction de séquençage. La présence d'une séquence index dans l'adaptateur permet de plus, de séquencer plusieurs banques en parallèle sur une même ligne de puce. Une étape d'enrichissement par amplification PCR de la banque est finalement effectuée. Le séquençage des fragments d'ADN peut être réalisé à partir d'une seule extrémité (*single-end*) ou des deux extrémités (*paired-end*). Dans le cas du séquençage *single-end*, chaque fragment d'ADNc est séquencé à partir d'une seule extrémité, on obtient donc une séquence (appelée *read*) par fragment d'ADNc séquencé. Dans le cas du *paired-end*, chaque fragment d'ADNc est séquencé à partir de ses deux extrémités, on obtient donc deux reads par fragment séquencé (100 bases pour chaque read dans notre étude) ; la taille de l'insert non séquencé entre les deux reads dépend de la longueur de séquençage et de la fraction de taille des fragments d'ADNc obtenus lors de la préparation des banques. Dans le cas d'organismes modèles, l'assemblage des reads peut être effectué par cartographie (ou *mapping*) sur des données de références (*i.e.* génome complet assemblé ou transcriptome de la même espèce ou d'une espèce proche). Dans le cas des organismes non modèles, les génomes ou transcriptome de référence sont souvent absents des bases de données, et il est nécessaire d'en obtenir un par assemblage dit *de novo*.

Deux éléments sont primordiaux pour assurer la qualité de l'assemblage *de novo* : la profondeur* (nombre de fois où une même base est séquencée) et la couverture* (nombre de nucléotides séquencés au moins une fois) du séquençage. Un effort de séquençage important

Sélection des reads

Assemblage de novo

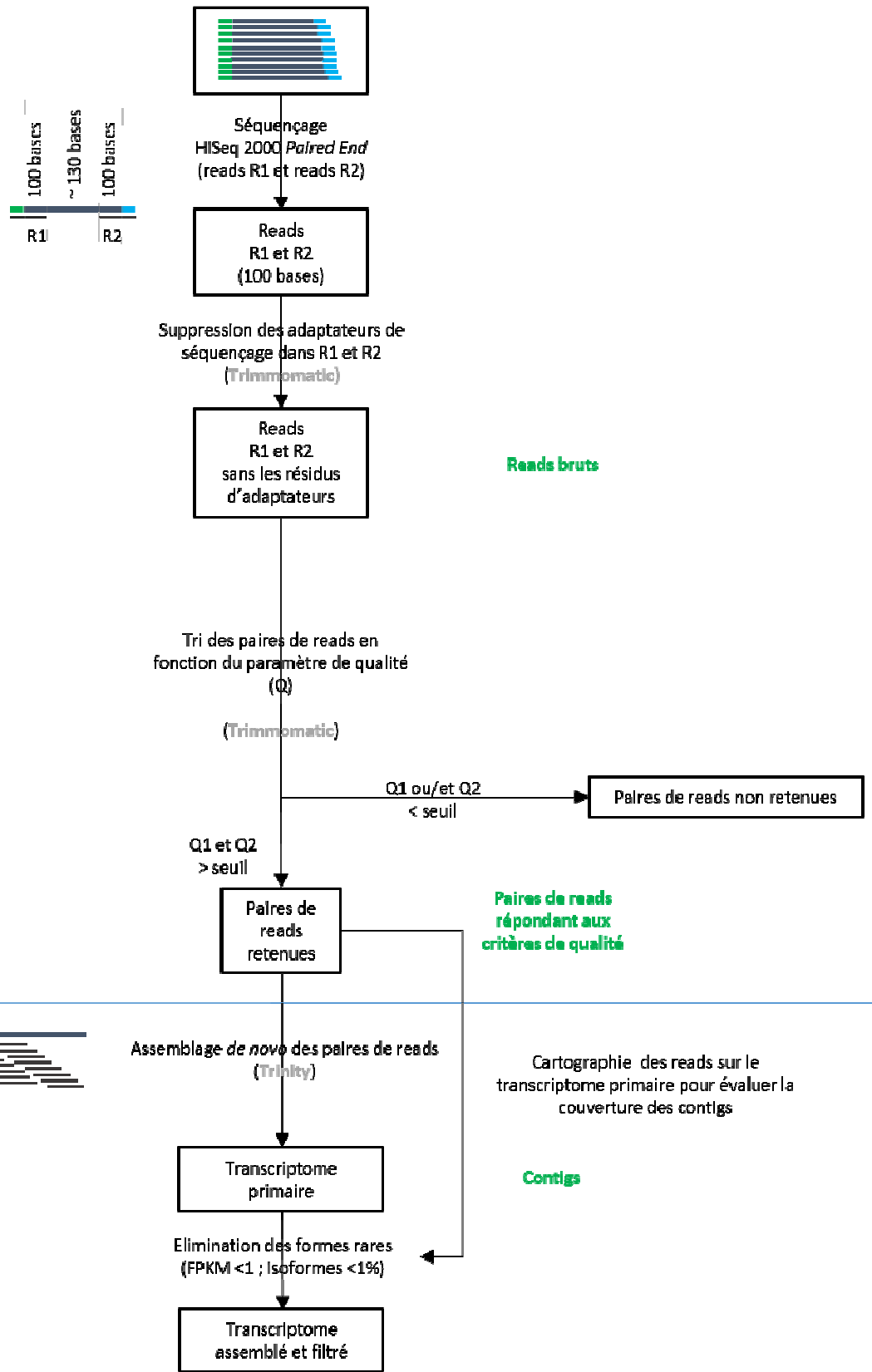


Figure IV.4 Représentation schématique des étapes de la sélection des reads en fonction de la qualité, de l'assemblage des reads en contigs et de la filtration des transcriptomes primaires.

permet à la fois d'augmenter la profondeur et la couverture de séquençage, donc d'augmenter la probabilité de séquencer des marqueurs homologues pour les différentes espèces, en particulier pour les gènes peu exprimés (et donc peu représentés dans les banques). Afin d'obtenir la couverture la plus complète possible avec la plus grande profondeur possible, chaque banque est séquencée en *paired-end* sur deux lignes de séquençage et l'ensemble des banques est séquencé sur quatre lignes différentes (répliques expérimentales). Le séquençage en Illumina HiSeq 2000 fournit environ 187 millions de reads appariés par ligne de séquençage. Ainsi, avec un spécimen séquencé pour chacune des 11 espèces échantillonnées, nous devrions obtenir environ 187 millions de reads appariés x 4 lignes / 11 banques, soit : 2x70 millions de reads *paired end* par banque.

IV.2.2.3 Sélection des reads de séquençage

Les différentes étapes depuis le séquençage de la banque d'ADNc jusqu'à l'obtention d'un transcriptome assemblé et filtré sont schématisées sur la figure IV.4. Le séquençage des banques produit un nombre très important de reads de 100 bases, fournis sous le format fastq. Dans ce format, à chaque read sont associés : des scores de qualité (Phred quality scores) pour chaque nucléotide séquencé, et des informations relatives à la position sur la puce du fragment d'ADNc, à partir duquel le read a été produit. Cette dernière information permet d'associer les deux reads correspondant à un même fragment d'ADNc.

Le prestataire de séquençage filtre les données brutes en retirant de l'extrémité 5' du read la séquence correspondant à l'adaptateur. Ce premier filtrage est cependant insuffisant. En effet, d'une part, dans le cas d'un fragment d'ADNc d'une longueur inférieure à la longueur de séquençage (dans notre cas : 100 bases), la séquence obtenue peut inclure tout ou partie du second adaptateur. D'autre part, lors de la confection des banques de séquençage, des accolements entre deux adaptateurs peuvent survenir et ne pas être détectés lors de ce premier filtrage. De plus, la qualité de chaque base décroît au cours du séquençage d'un fragment d'ADNc, rendant les extrémités 3' des reads plus sujettes aux erreurs (Fuller *et al.*, 2009). Ainsi, pour compléter cette filtration, le logiciel Trimmomatic-0.35 (Bolger *et al.*, 2014, voir encadré 3) permet notamment de supprimer les séquences d'adaptateurs (ILLUMINACLIP), et d'éliminer les portions de reads (SLIDINGWINDOW) et les extrémités (LEADING et TRAILING) de faible qualité. Enfin, seules les paires de reads ayant chacun une longueur supérieure à 50 nucléotides sont conservées (MINLEN).

Encadré 3 : L'outil Trimmomatic

Afin de supprimer les adaptateurs, l'étape ILLUMINACLIP de l'outil Trimmomatic procède par alignement des reads et des séquences des adaptateurs. Dans un premier temps, si la séquence de l'adaptateur s'aligne sur le read (avec une tolérance de deux nucléotides de différence au sein de l'alignement), l'adaptateur est supprimé. Dans un second temps, afin de détecter la présence de fragments d'adaptateur dans les reads (recherche dite en mode simple) ou/et d'accolements d'adaptateurs (recherche en mode palindromique), des scores d'alignement sont calculés. Les valeurs seuil des scores d'alignement utilisées dans notre étude sont respectivement de 15 pour le mode simple et de 40 pour le mode palindromique.

L'étape SLIDINGWINDOW de l'outil Trimmomatic permet, via une fenêtre coulissante à partir de l'extrémité 5', de couper les régions 3' des reads ayant des qualités faibles de séquençage. Dans notre étude, nous avons utilisé une fenêtre coulissante de quatre nucléotides sur laquelle une qualité moyenne est calculée à partir des scores de qualité de chaque nucléotide. Tant que la qualité moyenne est supérieure ou égale à 20, la fenêtre se déplace d'un nucléotide vers l'extrémité 3'. Lorsque la qualité moyenne de la fenêtre devient inférieure à 20, alors l'extrémité 3' du read est coupée ainsi que le dernier nucléotide (en 3') de la fenêtre coulissante.

Les étapes LEADING et TRAILING de Trimmomatic sont utilisées pour tronquer et éliminer les nucléotides aux extrémités du read tant que le score de qualité des nucléotides se trouvant aux extrémités est inférieur à 25.

Enfin, seules les paires de read pour lesquelles chaque read a une longueur supérieure ou égale à 50 nucléotides sont gardées pour l'assemblage (étape MINLEN).

IV.2.2.4 Assemblage de novo

Aucun assemblage de génome complet de décapode, *a fortiori* de Caridea, ni aucun jeu de données de séquençage avec une profondeur et une couverture suffisante pour produire un premier assemblage de génome, ne sont disponibles. Nos données sont donc assemblées *de novo*, c'est-à-dire sans référence génomique.

Dans le cas de l'assemblage *de novo*, une préparation initiale est nécessaire avant l'assemblage des reads en contigs* : chaque read est découpé en mots d'une longueur nucléotidique k (ces mots sont appelés k-mer*). Cette préparation consiste à normaliser le jeu de données afin de réduire la représentation des reads redondants et à éliminer les reads aberrants. Cette normalisation permet de diminuer le temps de calcul global de l'assemblage. La normalisation du jeu de données est réalisée par la méthode de couverture de k-mers à l'aide de l'outil Jellyfish version 1.1.11 (Marçais et Kingsford, 2011). Cet outil permet de découper l'ensemble des reads en mots de

longueur k , puis d'établir un catalogue de l'ensemble des mots uniques et enfin de calculer leur fréquence dans la population de reads. Cette méthode permet de détecter des reads correspondant à des transcrits* fortement exprimés (reads couverts par des k -mers de forte fréquence) et des reads aberrants (reads couverts par des k -mers de fréquences hétérogènes).

Deux méthodes d'assemblage *de novo* sont possibles. La première méthode utilise une unique valeur de k durant le processus d'assemblage (SK pour Simple K-mer) tandis que la seconde méthode utilise plusieurs valeurs de k durant le processus d'assemblage (MK pour Multiple k-mer). D'après Robertson *et al.* (2010), de faibles valeurs de k -mer permettent d'optimiser l'assemblage pour les gènes faiblement exprimés tandis que des valeurs importantes de k -mer permettent d'optimiser l'assemblage pour les gènes fortement exprimés. Zhao *et al.* (2011) ont analysé la qualité des assemblages produits par des logiciels SK et des logiciels MK, en observant plusieurs paramètres : (i) le nombre de reads que l'on peut *a posteriori* cartographier (remapping) sur le transcriptome nouvellement assemblé et (ii) le nombre de contigs assemblés *de novo* s'alignant sur le génome de référence des espèces sélectionnées (vrais positifs). Dans leur étude, Trinity (suite logicielle SK) est l'assembleur qui comptabilise le plus de reads qui sont cartographiés *a posteriori* sur les contigs assemblés, logiciels SK et MK confondus (paramètre (i)). Par ailleurs, alors que le nombre de gènes assemblés avec les logiciels SK est généralement plus faible que celui obtenu avec les logiciels MK, Trinity est le seul logiciel SK à avoir des valeurs similaires aux logiciels MK (paramètre (ii)). Enfin, les logiciels utilisant la méthode MK produisent plus de faux positifs et de contigs correspondant à des transcrits artificiellement fusionnés (chimères d'assemblage) que les méthodes SK (paramètre (ii)). Ces résultats sont cohérents avec l'étude de Clarke *et al.* (2013) qui recommande également l'utilisation de Trinity. Nous avons donc utilisé la suite logicielle Trinity version 2.2.0 (Hass *et al.*, 2013) pour assembler les données produites dans cette étude. Dans le logiciel Trinity la longueur du k -mer est fixée à 25 et ne peut être modifiée.

Trinity est une suite logicielle qui combine trois logiciels indépendants : *Inchworm*, *Chrysalis* et *Butterfly* (Grabherr *et al.*, 2011). Trinity est basé sur la construction de graphes de De Bruijn, qui représentent la complexité des gènes avec leurs épissages alternatifs et permettent de différencier les transcrits provenant de gènes paralogues. *Inchworm* (figure IV.5 (a)) sélectionne le k -mer le plus fréquent et l'assemble de manière itérative avec le k -mer le plus fréquent ayant $k-1$ nucléotides chevauchants. *Inchworm* ne permet pas de comprendre toute la complexité du transcriptome : dans le cas d'un gène avec des épissages alternatifs, seule la forme la plus abondante est assemblée sur

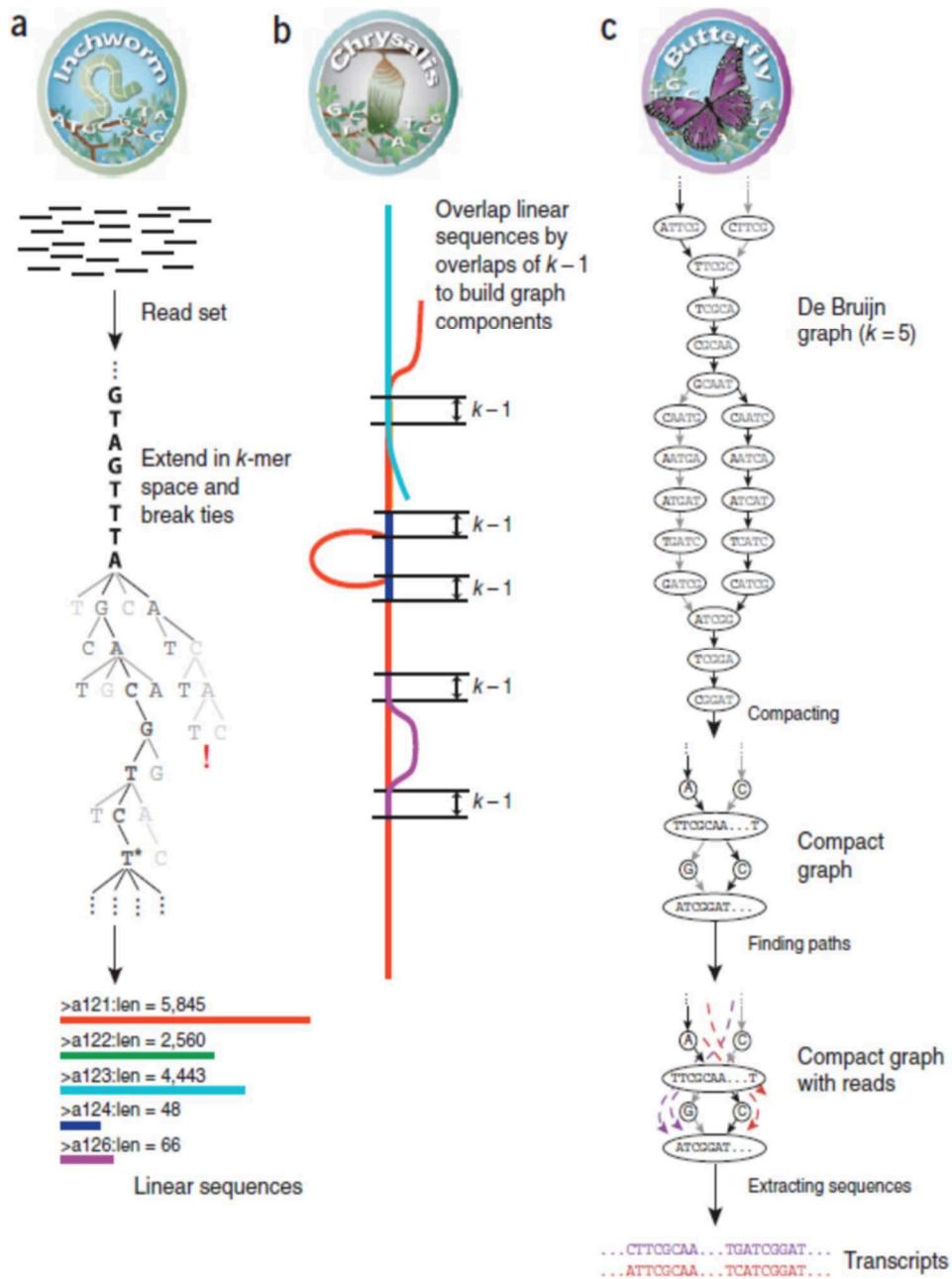


Figure IV.5 Illustration des trois étapes de la suite logicielle Trinity : Inchworm (a), Chrysalis (b) et Butterfly (c). Illustration tirée de Grabherr *et al.* (2011).

toute sa longueur lors de cette étape ; les autres formes ne sont, à cette étape, que partiellement assemblées. Dans un second temps, *Chrysalis* (figure IV.5 (b)) regroupe les contigs d'*Inchworm* ayant k-1 nucléotides chevauchants et chaque regroupement s'appelle un composant. Un graphe de De Bruijn est construit pour chaque composant. Les chemins retrouvés dans ces graphes sont constitués de nœuds correspondant aux séquences des k-mer et de branches reliant les k-mer qui ont k-1 nucléotides en communs. Lors de la dernière étape, *Butterfly* (figure IV.5 (c)) trace les chemins que prennent les reads et les paires de reads dans les graphes de De Bruijn obtenus lors de l'étape précédente. Comme les reads sont plus longs que k, cette étape permet d'éliminer d'éventuels faux contigs (chimères d'assemblage). *Butterfly* permet de reconstruire le transcrit sur toute sa longueur en détectant les épissages alternatifs (aussi appelés isoformes*) d'un même gène ainsi que les gènes paralogues.

A l'issu de Trinity, nous obtenons un transcriptome primaire (voir figure IV.4) constitué d'un ensemble de contigs. Sur ce transcriptome primaire, les paires de reads retenues à l'issue du traitement par le logiciel Trimmomatic sont cartographiées à l'aide du logiciel Bowtie v2.2.6 (Langmead et Salzberg, 2012). Le logiciel RSEM (pour *RNA-Seq by Expectation Maximization*) v1.2.31 (Li et Dewey, 2011) permet de calculer les abondances relatives de chaque contig et donc la profondeur de séquençage des contigs, relativement à l'effort de séquençage. Cette étape permet d'éliminer les contigs couverts avec une faible profondeur correspondant potentiellement à des contigs chimériques sans réalité biologique. Bien que Li et Dewey (2011) conseillent l'utilisation du TPM (Transcript Per Million) dans leur article, nous utilisons le calcul du FPKM* (Fragments Per Kilobase Of Exon Per Million Fragments Mapped), qui est plus communément utilisé. Pour cela le nombre de reads cartographiés sur un contig est normalisé par la longueur de ce contig ainsi que par le nombre total de reads cartographiés sur l'ensemble des contigs de l'assemblage *de novo*. Les contigs ayant des valeurs de FPKM <1 et les isoformes rares, représentant moins de 1% de l'effort de séquençage sont éliminées.

Les outils Jellyfish, Trinity, Bowtie et RSEM sont implémentés sur l'environnement virtuel Galaxy version r20140717 mis en place par la plateforme ABiMS de Roscoff.

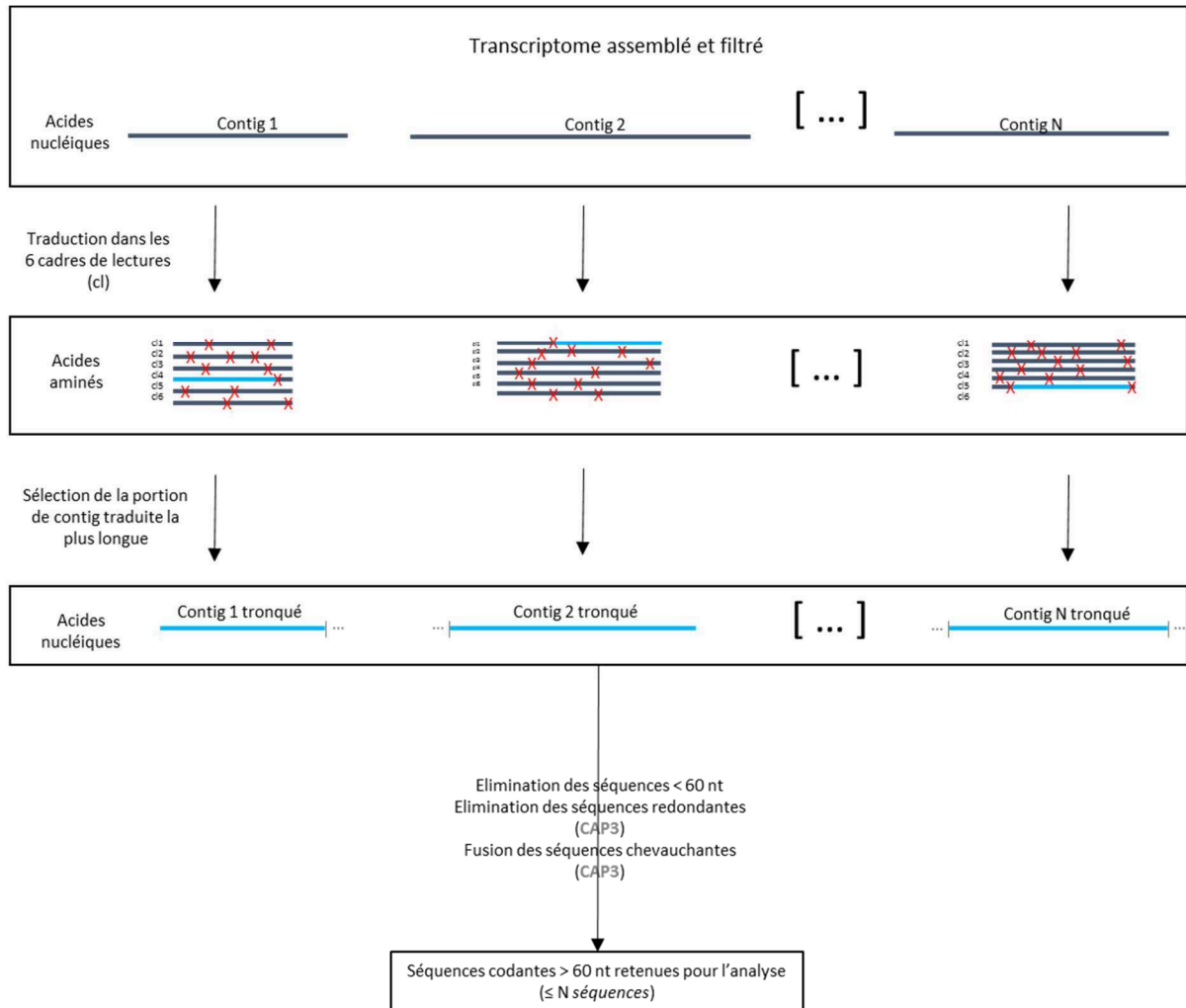


Figure IV.6 Représentation schématique des étapes du traitement post assemblage. Les contigs sont traduits selon les six cadres de lecture ouverts. La séquence nucléotidique correspondant au plus long fragment traduit est sélectionnée. Seules les séquences codantes de plus de 60 nucléotides sont conservées pour la suite des analyses. Les croix rouges sur le schéma correspondent à l'emplacement de codons stop.

IV.2.2.5 *Traitement post-assemblage*

La constitution des banques par fragmentation des ADN complémentaires induit une moins bonne représentation des extrémités des ARNm dans les banques. Ils sont en effet statistiquement de plus petite longueur, donc plus facilement éliminés lors des étapes de purification. Ainsi, les potentielles erreurs de séquençage ont plus de poids dans les extrémités des contigs assemblés. Par ailleurs, les extrémités des ARNm sont des régions transcrites mais non traduites (ou UTR pour Untranslated Transcribed Region*) qui ne sont pas des régions d'intérêt pour notre étude. Pour éliminer les biais de séquençage et éliminer les parties non codantes de l'assemblage, les séquences sont traduites selon les six cadres de lecture ouverts (ORF, Open Reading Frame) possibles (figure IV.6). Le code génétique utilisé est le code nucléaire. Pour chacun des six ORF possibles, la traduction peut produire un ou plusieurs fragments : (i) entre une des extrémités du contig et un codon stop, (ii) entre deux codons stop, ou (iii) sur la longueur totale du contig, en l'absence de codon stop. Pour chaque contig, le plus long fragment traduit selon un des six cadres de lecture ouverts est sélectionné. Malgré la sélection des ARNm par la présence de la queue poly-A, certains ARN non traduits (*e.g.* ARN ribosomiques, ARN de transferts, ARNm mitochondriaux, etc...) peuvent avoir été intégrés dans les banques en raison de leur forte abondance dans les cellules. On peut émettre l'hypothèse que pour ces ARN, aucun des six cadres de lecture ne permettra de trouver une séquence d'acides aminés de grande longueur. Aussi, les contigs pour lesquels les fragments traduits les plus longs font moins de 60 nucléotides sont supprimées. La population de contigs tronqués et filtrés par la taille est ensuite traitée avec l'outil CAP3 (Huang et Madan, 1999). Cet outil est utilisé à cette étape pour détecter la présence de contigs identiques sur toute ou partie de leur longueur. Cet outil permet ainsi d'éliminer la redondance et de fusionner les contigs chevauchants. Les seuils de similarité et de chevauchement retenus sont les valeurs par défaut, soit respectivement de 100% de similarité pour la redondance et 90% de similarité sur 40 nucléotides pour le chevauchement.

IV.2.2.6 *Caractérisation des assemblages*

Les principales métriques utilisées dans les études transcriptomiques afin de caractériser un assemblage (Moreton *et al.*, 2016) sont : le nombre de contigs assemblés, la moyenne et la médiane des longueurs des contigs ou encore le N50 (voir encadré 4). Ces métriques permettent de caractériser la distribution du nombre et de la taille des contigs et ainsi de comparer des assemblages entre eux. C'est un premier élément d'évaluation de la qualité des assemblages. En effet, le nombre de contigs seul ne donne pas une évaluation du nombre de transcrits assemblés car si les données sont insuffisantes, un même transcrit peut être fragmenté dans l'assemblage. La longueur moyenne

de l'assemblage permet de comparer relativement au nombre de contigs, l'importance de cette fragmentation. Ces métriques sont fournies pour l'ensemble des données issues du traitement post-assemblage utilisé dans notre étude.

Encadré 4 : Le N50

Le N50 a été utilisé initialement pour caractériser les assemblages du génome humain (Lander *et al.*, 2001). Dans le cas d'un génome, la discontinuité des séquences d'ADN reflète l'organisation de l'ADN total en chromosomes, l'assemblage complet devrait donc être composé d'un petit nombre de contigs de grande taille. Dans ce cas : plus le N50 est grand, plus les contigs assemblés sont longs : donc plus le N50 est grand, plus l'assemblage est complet. Cependant, dans le cas d'un transcriptome, les contigs correspondent aux transcrits. Le nombre et la taille des transcrits peut varier d'un tissu à l'autre et d'une espèce à l'autre, ce qui rend cette métrique peu adaptée pour évaluer la qualité d'un assemblage de transcriptome. Le N50 est notamment utilisé pour comparer des assemblages transcriptomiques obtenus par des logiciels différents à partir d'une même population de reads, ou pour comparer des assemblages obtenus avec différents sous-échantillonnages dans une population de reads (Clarke *et al.*, 2013). Le N50 n'est donc pas approprié pour comparer des assemblages de différentes espèces ou effectués dans des conditions différentes. Cependant, cette métrique est fréquemment renseignée dans les études transcriptomiques, mais sans être particulièrement commentée. Dans notre étude, nous nous servons uniquement des valeurs de N50 calculées pour chaque espèce pour comparer les résultats des différentes étapes de l'assemblage afin de rendre compte des impacts des traitements effectués.

Calcul du N50

Le N50 est la taille du contig tel que 50% des bases de l'assemblage sont comprises dans des contigs de taille supérieure ou égale à cette taille.

Soit l'assemblage comprenant 12 contigs rangés selon leur longueur nucléotidique. Les longueurs nucléotidiques de ces contigs sont :

12 ; 16 ; 26 ; 30 ; 30 ; 30 ; 39 ; 42 ; 48 ; 59 ; 60 ; 62

Somme = 454 nucléotides assemblés.

La moitié de l'assemblage est donc de 227 nucléotides.

$62 + 60 + 59 = 181.$

$62 + 60 + 59 + 48 = 229.$

Le N50 de cet assemblage est de 48, car c'est la longueur du plus petit contig pour laquelle la somme des longueurs des contigs de longueurs supérieures ou égales à 48 correspond à au moins 50% des nucléotides assemblés.

IV.2.3 Recherche et sélection de marqueurs phylogénétiques

IV.2.3.1 *Recherche de marqueurs pour la phylogénie : le principe des blasts réciproques*

Notre recherche s'appuie sur l'étude de O'Hara *et al.* (2014), visant à résoudre les relations phylogénétiques profondes au sein de la classe des Ophiuroidea en utilisant des marqueurs issus de transcriptomes. Dans cette étude les 425 marqueurs ont été sélectionnés en utilisant le protéome de l'oursin *Strongylocentrotus purpuratus* pour rechercher dans les transcriptomes de 52 espèces d'ophiures, d'une espèce de crinoïde, d'une espèce d'holothurie et de deux espèces d'astéries, des marqueurs orthologues*, afin de résoudre les relations phylogénétiques profondes au sein de la classe des Ophiuroidea. Ces protéines sont conservées à une échelle plus profonde que celle de la classe des Ophiuroidea car les auteurs de l'étude ont pu identifier et annoter ces protéines à partir du protéome de *Danio rerio* (espèce éloignée du groupe d'intérêt par environ 630 Ma, d'après dos Reis *et al.*, 2015). De plus, les auteurs de cette étude ont fait un important travail afin de ne pas sélectionner des gènes qui sont présents en multi-copies dans leur échantillonnage. Nous avons donc choisi d'utiliser ce même lot de marqueurs pour rechercher, par la méthode des blasts réciproques (ou RBBH pour *Reciprocal Best Blast Hit*) (i) des marqueurs au sein des transcriptomes de Caridea réunis pour cette étude ainsi que (ii) dans les décapodes choisis pour former le groupe externe. Enfin, l'approche phylogénomique est nouvelle au sein de l'équipe au muséum. Aussi, il paraît plus simple de s'appuyer sur un ensemble de marqueurs prédéfini plutôt que de faire à nouveau tout le travail de recherche de marqueurs. Nous utilisons donc les séquences protéiques de l'oursin de ces 425 marqueurs pour rechercher des marqueurs dans les transcriptomes réunis pour notre étude.

La méthode des blasts réciproques (figure IV.7) peut être utilisée pour identifier des séquences comme potentiellement orthologues. Dans cette méthode, couramment utilisée (Tatusov *et al.*, 1997 ; Rivera *et al.*, 1998 ; Ward et Moreno-Hagelsieb, 2014), deux séquences sont considérées orthologues si chacune est le meilleur résultat de blast de l'autre. C'est la méthode que nous utilisons pour identifier des marqueurs orthologues au sein des transcriptomes de notre étude, en prenant comme référence les 425 protéines de l'oursin sélectionnées dans l'étude d'O'Hara *et al.* (2014).

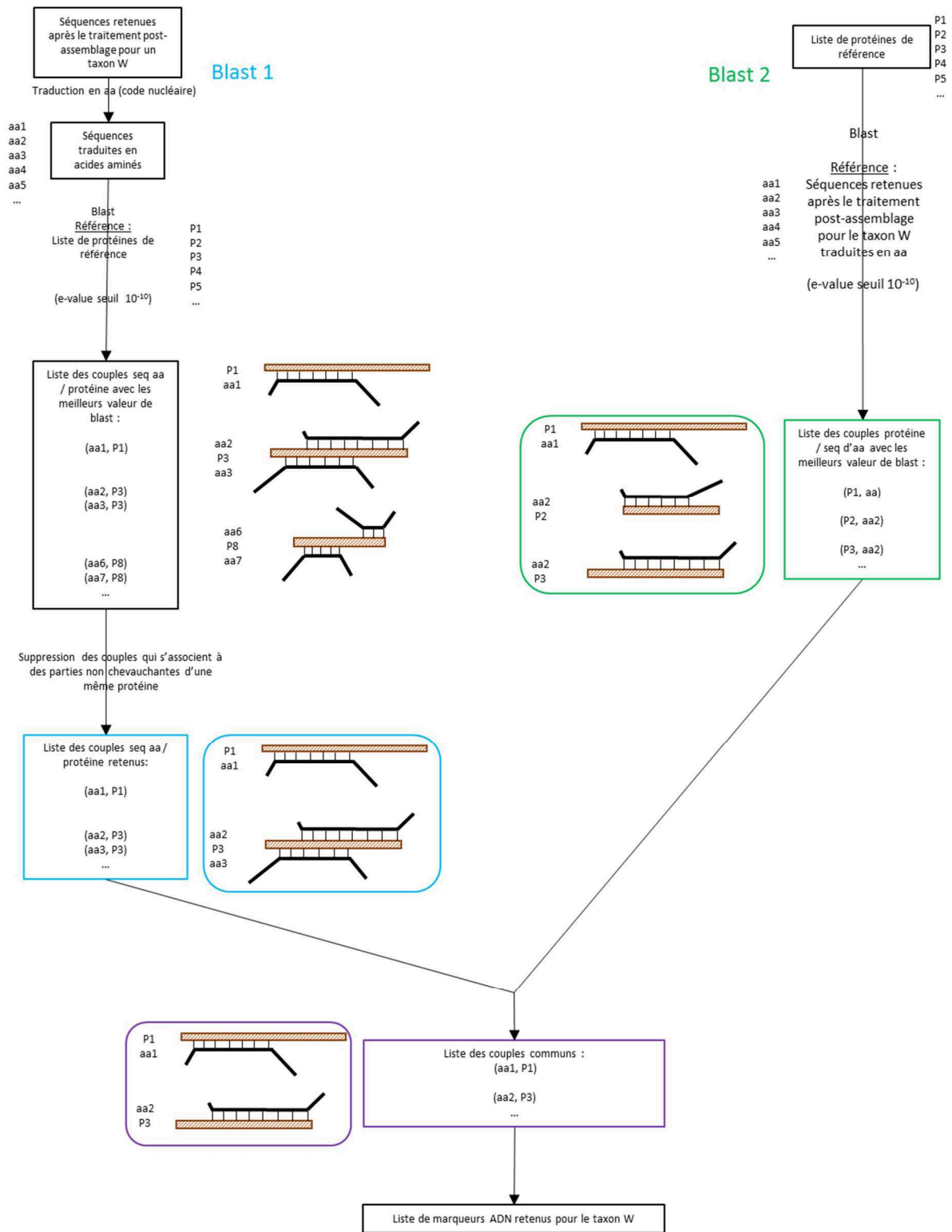
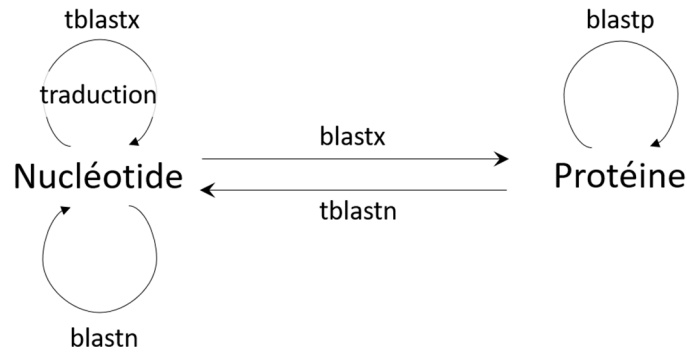


Figure IV.7 Représentation schématique du principe des blast réciproques. Soit un ensemble de séquences A et un ensemble de séquences B. Lors du premier blast, chaque séquence de A (la requête) est recherchée au sein de B (la référence) avec une valeur seuil d'e-value déterminée. Seul le meilleur résultat de blast pour chaque séquence de A (e-value la plus proche de 0) est conservé. Lors du second blast, chaque séquence de B (la requête) est recherchée au sein de A (la référence) avec la même valeur seuil d'e-value. Pareillement, seul le meilleur résultat de blast pour chaque séquence de B est conservé. Si les meilleurs résultats de blast sont symétriques, le couple de séquences concernées est considéré comme orthologue.

Encadré 5 : Les différents types de blast

Présentation des différents types de blast en fonction de la nature des données utilisées (ADN / acide aminé) pour la requête (du côté initial de la flèche) et pour la référence (du côté de la pointe de la flèche). *Par exemple, faire un blast d'une séquence ADN contre une base de données protéique est un blastx.*



Dans le cas de notre étude, la méthode du blast réciproque est appliquée à chaque transcriptome de Caridea et de Décapode non Caridea de la manière décrite ci-après.

Lors du premier blast, chaque séquence codante du Décapode considéré issue du traitement post-assemblage (la requête) est comparée par blast aux 425 protéines de l'oursin (la référence) (*blastx*, avec une valeur d'e-value seuil de 10^{-10} ; voir encadré 5 pour une explication des différents types de blast). Seul le meilleur résultat de blast pour chaque séquence du Décapode est conservé, résultant en une liste de couples : séquence d'acides aminés du Décapode / protéine de l'oursin. Dans ce premier blast, une même protéine d'oursin peut être associée à différentes séquences du Décapode considéré. Dans ce cas, deux situations peuvent être rencontrées : (1) Les couples s'associent sur des parties non chevauchantes d'une même protéine. Dans ce cas, il n'est pas possible de déterminer si les séquences du Décapode associées à une même protéine correspondent à deux transcrits distincts ou à un même transcrit qui n'a pu être assemblé entièrement. Ces résultats de blast ne sont donc pas retenus ; ou (2) Les couples s'associent sur des parties chevauchantes d'une même protéine. Dans ce cas la séquence ayant le meilleur résultat de blast (c'est-à-dire la valeur d'e-value la plus faible) est conservée, l'autre étant considérée comme un potentiel paralogue*.

Lors du second blast, chaque protéine d'oursin (la requête) est comparée par blast (*tblastn*, avec la même valeur seuil d'e-value de 10^{-10}) aux séquences codantes du Décapode considéré (la référence). De même, seul le meilleur résultat de blast pour chaque protéine de l'oursin est conservé, résultant en une liste de couples : protéine d'oursin / séquence d'acides aminés du Décapode.

Enfin, les deux listes de couples résultant des deux blasts sont comparées et seuls les couples retrouvés communs aux deux listes sont conservés.

Pour chacune des 425 protéines d'oursin, les séquences retenues par la méthode RBBH pour les espèces de Décapodes utilisées dans cette étude sont blastées entre elles (*tblastx*, e-value seuil de 10^{-10}). Les extrémités qui ne sont similaires à aucune autre séquence sont tronquées. Les séquences sont ensuite alignées entre elles de manière automatique à l'aide de l'outil MAFFT v7.299b (Kato et al., 2002).

IV.2.3.2 *Choix et sélection de marqueurs pour la phylogénie*

Le but de notre étude est de retrouver le plus grand nombre de marqueurs possible, pour le plus d'espèces possible. Afin de maximiser les chances de trouver des marqueurs transcriptomiques partagés par l'ensemble des lignées (correspondant aux familles) de Caridea, nous ciblons :

1/ des transcriptomes extraits dans des conditions similaires et exprimant *a priori* la même population d'ARNm ;

2/ des marqueurs retrouvés dans le plus grand nombre de familles possible ;

La production de transcriptomes par l'équipe au muséum a été ciblée pour remplir ces deux critères et sont donc tous utilisés pour sélectionner les marqueurs de notre étude. De plus, les transcriptomes issus de la littérature et complétant la couverture taxonomique au rang des familles sont également inclus dans la recherche de marqueurs. Dès lors que, pour au moins une espèce de cette sélection de transcriptomes, un marqueur testé résulte en plusieurs couples non chevauchants d'une même protéine, ce marqueur n'est plus considéré pour l'ensemble des espèces de la sélection.

La qualité du signal phylogénétique, mais aussi les méthodes utilisées pour séquencer les marqueurs choisis à partir d'extraits d'ADN sur un plus grand panel d'espèces sont dépendantes de la divergence nucléotidique. Une divergence trop forte implique en effet des difficultés d'amplification par PCR et de capture par hybridation. Les p-distances nucléotidiques (correspondant au nombre de nucléotides différents entre les séquences comparées) sont donc calculées pour l'ensemble des marqueurs partagés par l'ensemble des familles considérées dans cette étude. L'objectif de l'étude étant de résoudre les nœuds profonds et de définir des marqueurs utilisables de façon standardisée sur d'autres taxons, le signal phylogénétique des 20 marqueurs les plus conservés est étudié et comparé avec les 20 marqueurs les moins conservés.

IV.2.4 Orthologie des marqueurs sélectionnés

IV.2.4.1 Rappel des concepts d'homologie/orthologie/paralogie/homoplasie

Encadré 6 : Définitions

Homologie : relation entre deux caractères similaires ayant divergé à partir d'un même ancêtre commun.

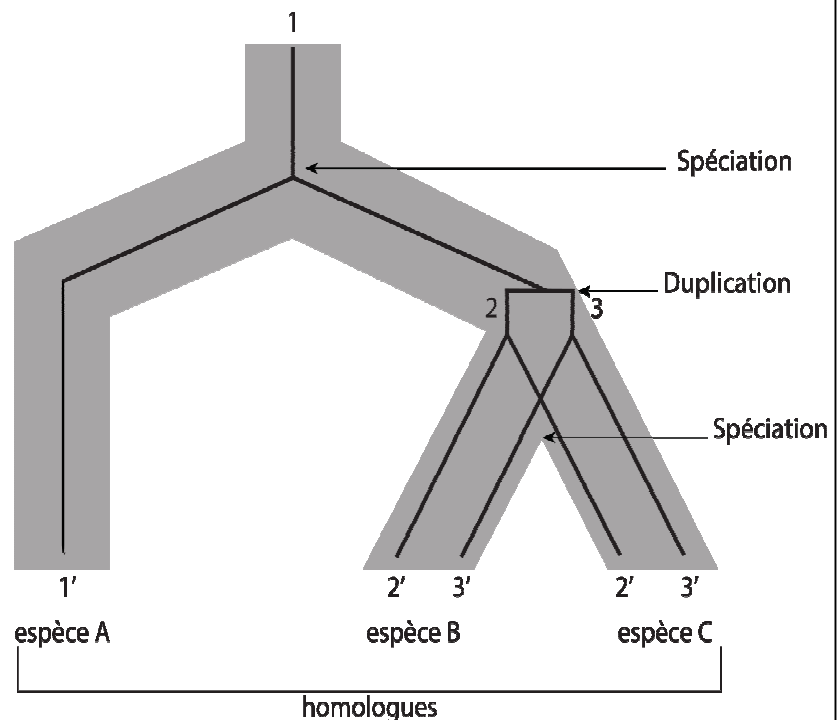
Orthologie : la divergence entre les caractères homologues résulte uniquement de la spéciation.

Paralogie : la divergence entre les caractères homologues résulte d'un événement de duplication chez un ancêtre commun.

Homoplasie : Relation entre deux caractères dont la similarité résulte d'une convergence évolutive à partir de deux caractères chez deux ancêtres distincts.

Figure IV.8 Schéma explicatif des relations d'orthologie et de paralogie au sein d'une famille de gènes.

Les relations entre les espèces A, B et C sont explicitées par les rectangles grisés. Les relations de parentés entre les copies du gène issues de l'état ancestral 1 sont explicitées par les traits noirs. L'espèce A n'a qu'une copie du gène considéré (copie 1'). Suite à un événement de duplication du gène antérieur à la spéciation conduisant aux espèces B et C, ces deux dernières espèces ont deux copies de ce gène. Entre les espèces B et C, les copies du gène notées avec le même chiffre sont orthologues, tandis que des copies avec des chiffres différents sont dites paralogues. L'ensemble des copies sont homologues. La copie 1' de l'espèce A sert de groupe externe pour les deux couples d'orthologues présentes dans les espèces B et C.



IV.2.4.2 Vérification de l'homologie et de l'orthologie des séquences des marqueurs sélectionnés

Les marqueurs sélectionnés avec la méthode des RBBH sont des marqueurs ayant une séquence similaires entre les taxons étudiés. Suite à cette étape, il est nécessaire de vérifier que, pour chacun des 40 marqueurs, cette similarité peut raisonnablement être considérée comme traduisant l'orthologie entre les taxons (voir encadré 6) et dans le cas de séquences douteuses, de chercher une séquence plus probablement orthologue que celle trouvée à l'aide du RBBH.

Pour chaque marqueur, trois étapes successives, basées respectivement sur des critères fonctionnel, topologique et de distance détaillés ci-après, permettent de détecter les séquences non orthologues et de les supprimer du jeu de données. A l'issue de ces trois étapes, pour remplacer les séquences ainsi éliminées, une séquence plus probablement orthologue est cherchée à nouveau dans les transcriptomes des espèces concernées selon la procédure détaillée ci-après. L'hypothèse d'orthologie n'est pas rejetée pour les séquences qui ont répondu à tous les critères. A l'issue de cette procédure, les séquences sont réalignées à la main avec le logiciel MEGA v.5.0.5 (Tamura *et al.*, 2011) pour chacun des marqueurs.

Critère fonctionnel du gène

Les séquences recherchées sont des copies fonctionnelles d'un gène et ne doivent donc pas contenir de codon stop. La présence d'un codon stop peut être expliquée de deux manières : soit la séquence considérée ne correspond pas à un transcrit mais est un artéfact de séquençage ou une chimère créée lors du traitement post-assemblage, soit cette séquence est un pseudogène, c'est-à-dire correspond à une copie paralogue non fonctionnelle ou qui ne code pas pour la même protéine. Pour évaluer ce critère, chaque séquence est traduite en acide aminés. Les séquences non fonctionnelles sont retirées de l'alignement.

Critères topologiques

Dans un second temps, un arbre basé sur les nombres de différences nucléotidiques a été reconstruit avec la méthode de phylogénie de Neighbor-Joining pour les séquences qui ont répondu au critère fonctionnel. Pour chaque marqueur, les séquences de Caridea groupant avec les espèces du groupe externe sont considérées comme douteuses. De même, dans les cas où plusieurs espèces d'une même famille sont incluses dans le jeu de données, si une espèce se positionne en dehors de

cette famille, la séquence est considérée comme potentiellement paralogue. Les séquences n'ayant pas rempli ces critères topologiques sont douteuses et également retirées de l'alignement.

Critère des distances nucléotidiques

Etant donnée que la distance nucléotidique entre deux séquences orthologues est attendue à être plus faible que la distance entre deux séquences non orthologue (homoplasique* ou paralogue*), la présence d'une séquence non orthologue dans un ensemble de séquences orthologues, doit pouvoir se distinguer en analysant la distribution des distances nucléotidiques par paire de séquences. Pour chacun des marqueurs, une valeur seuil de distance nucléotidique doit donc être définie au-delà de laquelle deux séquences ne sont pas considérées comme orthologues. Cependant, la valeur de ce seuil dépend de l'échantillonnage taxonomique disponible pour chaque marqueur. Ainsi, la détermination et l'application de ce critère sont faites au cas par cas et sont détaillées dans la partie Résultats et discussion. Aussi, si deux séquences d'espèces de genres différents sont strictement identiques (distance nucléotidique = 0), nous avons considéré qu'il s'agissait d'une contamination. Dans ce cas, les deux séquences sont considérées comme douteuses, car il ne nous est parfois difficile de déterminer la direction de la contamination.

Recherche de la copie orthologue dans les cas des espèces n'ayant pas répondu aux trois critères

Lorsque des séquences présentent une orthologie douteuse (car ne répondant pas à un l'un des trois critères décrits précédemment), une nouvelle copie est recherchée dans le transcriptome de chacune des espèces concernées. Pour affiner la recherche, et au lieu de partir de la protéine correspondante de l'oursin (O'hara *et al.*, 2014), nous utilisons les séquences nucléotidiques de *Caridea* retenues comme potentiellement orthologues. Pour ce faire, la même méthode des blast réciproques (cette fois-ci seulement en *blastn*) décrite précédemment est utilisée, avec comme valeur seuil e-value < 10⁻¹⁰.

A l'issue de chaque blast réciproque, plusieurs résultats sont possibles :

- (i) Aucune séquence n'est retrouvée dans le transcriptome de l'espèce considérée. L'espèce est alors manquante dans le jeu de données pour ce marqueur.
- (ii) La séquence retrouvée est différente de celle rejetée lors des trois étapes de vérification. La séquence est incluse dans l'alignement du marqueur considéré et est évaluée selon les trois étapes de vérification. Si cette séquence ne répond pas aux trois critères, le second résultat des blast réciproques est alors considéré et est à son tour évalué selon les trois étapes de vérification.

- (iii) La séquence retrouvée est la même que celle qui n'a pas répondu aux critères de vérification. Dans ce cas, le second résultat des blasts réciproques est pris en compte, et peut donc correspondre soit au cas (i) si la p value de ce résultat est supérieur au seuil retenu, soit au cas (ii).

Dans le cas où les séquences d'espèces de genres différents étaient strictement identiques (contamination), plusieurs résultats de blast sont envisageables :

- Le meilleur résultat du blast réciproque donne pour l'une des deux espèces la même séquence que celle d'origine et une séquence différente pour l'autre. On considère que la première espèce a contaminé la seconde. Dans ce cas, la première espèce garde la même séquence que celle d'origine et la seconde espèce la nouvelle séquence, qui est vérifiée selon les trois étapes de vérification.
- Le meilleur résultat de blast réciproque donne la même séquence qu'à l'origine pour les deux espèces. Dans ce cas, les deux séquences sont supprimées et les deux espèces sont manquantes dans le jeu de données pour ce marqueur.
- Le meilleur résultat des blast réciproques fournit deux séquences différentes des séquences d'origines. Ces deux nouvelles séquences sont conservées et sont vérifiées selon les trois étapes de vérification.

Par ailleurs, une recherche de séquence par blast réciproque (*blastn*) est réalisée de la même manière, pour les espèces pour lesquelles aucune séquence n'avait été trouvée en prenant comme référence les protéines de l'oursin *Strongylocentrotus purpuratus*.

Ce processus de vérification en trois étapes peut être considéré comme une boucle de vérification. En effet, il est itératif et ne s'arrête que lorsqu'aucune nouvelle séquence n'est proposée à l'issue de la recherche par blast réciproque d'une séquence plus probablement orthologue.

IV.2.4.3 Comparaison des marqueurs avant et après vérification

Afin d'évaluer l'importance des modifications engendrées par le remplacement des séquences n'ayant pas respecté les trois critères, plusieurs représentations sont proposées. D'une part, la distribution des distances nucléotidiques pour chacun des marqueurs est représentée, en différenciant par la couleur les distances calculées entre deux spécimens d'une même famille de celles calculées pour deux spécimens attribués à des familles différentes. Comme une séquence correspondant à une copie non orthologue aux autres séquences pour ce marqueur entraîne des distances nucléotidiques importantes avec cette séquence, les distributions des distances devraient

être plus larges avant que les copies paralogues n'aient été supprimées. De plus, au moins pour les marqueurs les plus conservés, nous nous attendons à ce que les distances au sein d'un même marqueur se hiérarchisent d'une manière cohérente avec la taxonomie, c'est-à-dire que globalement les distances nucléotidiques intrafamiliales doivent être inférieures aux distances interfamiliales. De plus, cette hiérarchie devrait se retrouver plus clairement dans les données qui ont été vérifiées.

D'autre part, les moyennes, écart-types et ratios variance/moyenne (RVM, un indice statistique de dispersion), sont calculés pour chaque marqueur, avant et après vérification de l'orthologie et (eventuellement) remplacement des séquences. Ces mesures statistiques sont illustrées sous forme de diagrammes en boîte. Pour cette représentation, nous avons différencié à la fois le jeu de marqueurs (les 20 marqueurs plus conservés et les 20 marqueurs plus divergents), et le type de distance (intra- ou interfamiliale) au sein de chacun de ces deux jeux de données.

IV.2.5 Analyses phylogénétiques

IV.2.5.1 *Choix du groupe externe*

Il est recommandé d'utiliser au moins deux groupes externes pour ancrer une phylogénie, afin de tester la monophylie du groupe interne (Barriel et Tassy, 1998). De plus, il est conseillé d'utiliser un premier groupe externe relativement proche du groupe interne et un second plus éloigné phylogénétiquement. Ainsi, nous utiliserons un groupe externe composé de trois espèces de l'ordre des Décapodes. Selon Kenny *et al.* (2014) et Shen *et al.* (2013), les Caridea sont le groupe frère de tous les autres Décapodes Pleocyemata. Ainsi, le groupe externe peut être choisi au sein de n'importe quel autre infra-ordre des Pleocyemata. Nous avons donc sélectionné parmi des données disponibles déjà assemblées : un crabe brachyoure (*Hyas araneus*) et un anomoure (*Petrolisthes cinctipes*). Nous avons également ajouté le transcriptome d'un Dendrobranchiata (groupe frère des Pleocyemata), *Penaeus vannamei*, à l'analyse.

IV.2.5.2 *Analyses phylogénétiques*

Les marqueurs avec moins de la moitié de l'échantillonnage taxonomique complet ne sont pas gardés pour notre étude, car ils comportent trop de données manquantes. Trois jeux de données ont été utilisés pour les analyses phylogénétiques. Un premier jeu constitué de la concaténation des marqueurs les plus conservés, un second constitué de la concaténation des marqueurs les plus

divergents, et un troisième constitué de la concaténation de l'ensemble de ces marqueurs. L'homoplasie étant potentiellement importante (principalement pour les marqueurs les plus divergents), ces trois jeux de données ont été analysés d'une part avec un groupe externe composé de Décapodes non Caridea et d'autre part avec les taxons appartenant à la famille des Atyidae (soit un total de six jeux de données). En effet, d'après Aznar-Cormano *et al.* (2015), les Atyidae appartiennent au clade I, frère du clade qui inclut les Alvinocarididae, dont l'origine évolutive est la motivation de cette étude.

Les analyses phylogénétiques des séquences nucléotidiques ont été effectuées à la fois en utilisant les méthodes du Maximum de vraisemblance (ML) et de l'Inférence Bayésienne (IB).

Maximum de vraisemblance

Le logiciel RAxML version 8.2.9 (Stamatakis *et al.*, 2014) a été utilisé pour l'analyse de Maximum de vraisemblance. Le modèle d'évolution GTR + Γ + I a été utilisé pour chaque partition, pour chaque position de codon. Le modèle GTR (Generalised Time-Reversible) (Tavaré 1986) est le modèle de substitution le plus neutre qui existe. Il implique que la fréquence de chacune des quatre bases A, T, C et G soit indépendante des autres et que les taux de substitution entre chaque paire de nucléotides soient différents. Ce modèle nécessite donc quatre paramètres de fréquences des bases et six taux de substitutions différents. Ce modèle est combiné à une distribution des taux de mutation des sites qui suit une loi gamma (Γ) (fixée à 4 catégories) et une proportion de sites invariables (I). La robustesse des nœuds est estimée par la méthode du bootstrap avec 1 000 itérations. L'arbre est visualisé à l'aide du logiciel FigTree v1.4.3 (Rambaut et Drummond, 2012). Un nœud est considéré comme étant soutenu pour des valeurs de bootstrap ≥ 80 %.

Inférence bayésienne

L'inférence bayésienne a été faite avec MrBayes v3.2.6 (Ronquist *et al.*, 2012). Les analyses ont été effectuées sur le portail en ligne CIPRES (Miller *et al.*, 2010). Elles ont été réalisées sous le modèle d'évolution GTR + Γ + I. Deux analyses sont systématiquement lancées, comportant chacune 8 chaînes de Markov, sur $150 \cdot 10^6$ générations (pour les jeux de données basés sur les marqueurs conservés d'une part et les marqueurs divergents d'autre part) et $175 \cdot 10^6$ générations (pour les jeux de données constitués de l'ensemble des marqueurs), avec une fréquence d'échantillonnage des arbres tous les 5 000 pas. La stabilité de chaque paramètre a été contrôlée à l'aide du logiciel Tracer v.1.6 (Drummond et Rambaut, 2007), en vérifiant les « Estimate Sample Size » (ESS). L'ESS

estime la quantité d'échantillons indépendants qui ont la valeur indiquée pour ce paramètre. Nous considérons que le paramètre est stable quand les ESS sont supérieures à 200. La convergence des analyses a été vérifiée avec les valeurs de PSRF (Potential Scale Reduction Factor, égales à 1 si les analyses ont convergé). L'arbre consensus a été reconstruit en éliminant les premières 25% générations échantillonnées comme un 'burn-in'. Un nœud est considéré comme soutenu quand il présente des valeurs de probabilité postérieure $\geq 0,95$.

Tableau IV.1 Liste des espèces dont l'ARN a été extrait au MNHN. Superfamille et famille des espèces ; numéro MNHN du spécimen extrait ; informations sur la mission océanographique, la station et l'année de récolte du spécimen (les informations en italique correspondent aux échantillonnages ne provenant pas de campagnes océanographiques) ; méthode de préservation des tissus (RNA-later / azote liquide) le cas échéant ('frais' indique que l'ARN a été extrait immédiatement après prélèvement du spécimen) ; tissu prélevé pour l'extraction d'ARN en précisant lorsque le tube digestif (TD) a pu être enlevé ; séquençage ou non des ARN extraits du spécimen.

| Superfamille | Famille | Espèce | numéro MNHN | Mission (station), année de récolte | Préserv. des tissus | Tissu prélevé pour extraction ARN | Séquéncé |
|---------------|-----------------|--|--------------------|---|---------------------|-----------------------------------|----------|
| Alpheoidea | Alpheidae | <i>Alpheus</i> sp. | MNHN-IU-2014-2799 | Kavieng2014 (KM21), 2014 | RNA-later | abdomen | non |
| Alpheoidea | Alpheidae | <i>Athanas nitescens</i> | MNHN-IU-2014-10152 | <i>Hors camp. (Roscoff, côtier), 2014</i> | frais | abdomen | oui |
| Alpheoidea | Hippolytidae | <i>Caridion steveni</i> | MNHN-IU-2014-10156 | <i>Hors camp. (Roscoff, côtier), 2014</i> | frais | abdomen | oui |
| Alpheoidea | Hippolytidae | <i>Hippolyte varians</i> | MNHN-IU-2014-10158 | <i>Hors camp. (Roscoff, côtier), 2014</i> | frais | abdomen | non |
| Atyoidea | Atyidae | <i>Atyopsis moluccensis</i> | MNHN-IU-2014-10151 | <i>Hors camp. (commerce aquariophile), 2014</i> | frais | abdomen, pas TD | oui |
| Bresilioidea | Alvonocarididae | <i>Rimicaris exoculata</i> | MNHN-IU-2014-10153 | BICOSE2014 (PL568), 2014 | Azote liquide | abdomen, pas TD | oui |
| Crangonoidea | Crangonidae | <i>Crangon crangon</i> | MNHN-IU-2014-10155 | <i>Hors camp. (Roscoff, côtier), 2014</i> | frais | abdomen, pas TD | oui |
| Crangonoidea | Crangonidae | <i>Glyphocrangon</i> sp. | MNHN-IU-2014-10159 | Salomon2 (CP2186), 2004 | Azote liquide | abdomen, pas TD | non |
| Oplophoroidea | Oplophoridae | <i>Oplophorus gracilirostris</i> | MNHN-IU-2014-10154 | MADEEP (CP4302), 2014 | RNA-later | muscle abdomen | oui |
| Palaemonoidea | Palaemonidae | <i>Periclimenes brevicarpalis</i> | MNHN-IU-2014-2819 | Kavieng2014 (KM70), 2014 | RNA-later | abdomen | oui |
| Pandaloidea | Pandalidae | <i>Heterocarpus</i> aff. <i>dorsalis</i> | MNHN-IU-2014-10157 | Salomon2 (CP2186), 2004 | Azote liquide | muscle abdomen | oui |
| Pandaloidea | Pandalidae | <i>Heterocarpus parvispina</i> | MNHN-IU-2014-10160 | MADEEP (DW4314), 2014 | RNA-later | muscle abdomen | non |

IV.3 Résultats - Discussion

IV.3.1 Echantillonnage

IV.3.1.1 *Transcriptomes acquis dans l'équipe du MNHN*

Sur les 12 échantillons d'ARN extraits dans l'équipe, 8 avaient le niveau de qualité requis pour être séquencés par le centre d'innovation Génome Québec et Université McGill : *Atyopsis moluccensis*, *Athanas nitescens*, *Caridion steveni*, *Crangon crangon*, *Heterocarpus* aff. *dorsalis*, *Oplophorus gracilirostris*, *Pandalus latirostris* et *Rimicaris exoculata*. Le séquençage a fourni en moyenne 100 millions de reads en *paired-end* par transcriptome, ce qui correspond à l'attendu dans nos conditions de séquençage. Les huit espèces dont les transcriptomes ont été séquencés appartiennent à huit familles différentes (voir tableau IV.1) : Alvinocarididae, Atyidae, Alpheididae, Crangonidae, Hippolytidae, Oplophoridae, Palaemonidae et Pandalidae. Sept d'entre elles se situent dans le clade II de l'étude d'Aznar-Cormano *et al.* (2015) qui est le clade le plus diversifié, avec 21 familles de Caridea. L'espèce *Atyopsis moluccensis* (Atyidae) est l'unique représentant du clade I, qui contient quatre familles de Caridea et qui est groupe frère de toutes les autres familles de Caridea présentes dans leur étude.

IV.3.1.2 *Transcriptomes récupérés à partir des bases de données publiques*

Dans la base de données EMBL, 31 études de transcriptomique ciblant 20 espèces de Caridea séquencées en *paired-end* sont enregistrées. Les familles Atyidae et Palaemonidae couvrent la majeure partie des données disponibles avec 5 et 23 études menées sur au moins un membre de ces familles, respectivement. Une des cinq études sur les Atyidae est menée sur trois espèces d'Atyidae différentes et à la fois sur deux espèces d'Alpheididae. Les trois études ne concernant ni des Atyidae ni des Palaemonidae concernent un Alvinocarididae, un Hippolytidae et un Pandalidae.

Les transcriptomes sélectionnés à partir de la littérature proviennent d'études différentes. Aussi, le nombre de spécimens utilisés, les conditions expérimentales et les tissus à partir desquels ont été faites les extractions d'ARN sont très diverses (voir tableau IV.2 pour plus de détails sur les jeux de données sélectionnés). Les transcriptomes de huit espèces sont disponibles sous forme de données déjà assemblés : *Antecaridina lauensis*, *Caridina rubella*, *Chorismus antarcticus*,

Tableau IV.2 Liste des transcriptomes utilisés provenant de la littérature. Pour chaque espèce sélectionnée pour notre étude, sont renseignés : la famille dans laquelle l'espèce est classée, le type de données qui est utilisé dans cette étude (contigs déjà assemblés ou données de séquençage brute), le numéro de l'étude sous lequel les données peuvent être retrouvées dans la base de données EMBL-ENA, les numéros de séquençage à partir desquels les données ont été réassemblées, la référence de l'article scientifique lorsque le transcriptome provient d'une étude publiée, le nombre d'individus (N) à partir desquels tout ou partie des ARN ont été extraits et les tissus à partir desquels les ARN ont été extraits

| Espèce | Type de donnée utilisée | Etude dans EMBL-ENA | Réf des runs dans EMBL-ENA | Référence | N | Tissu |
|--|-------------------------|---------------------|---------------------------------------|------------------------------------|-----------|---|
| <i>Antecaridina lauensis</i> | contigs assemblés | PRJNA244748 | | Havird et Santos, 2014 | 1 | Individu entier |
| <i>Caridina multidentata</i> | Reads | PRJDB4543 | DRR054560 | | | Mélange de tissus musculaires et neuronaux |
| <i>Caridina rubella</i> | contigs assemblés | PRJNA244748 | | Havird et Santos, 2014 | 1 | Individu entier |
| <i>Chorismus antarcticus</i> | contigs assemblés | PRJNA359743 | | Toullec <i>et al.</i> , 2017 | 20 | 2 thorax, 2 abdomens et 20 pédoncules oculaires |
| <i>Halocaridina rubra</i> | contigs assemblés | PRJNA304256 | | Havird et Santos, 2016 | 1 | Individu entier |
| <i>Halocaridinides trigonophthalma</i> | contigs assemblés | PRJNA244748 | | Havird et Santos, 2014 | 1 | Individu entier |
| <i>Macrobrachium nipponense</i> | contigs assemblés | GSE74684 | | Xu <i>et al.</i> , 2016 | 36 | Hépatopancréas |
| <i>Macrobrachium olfersii</i> | Reads | PRJNA298624 | SRR2637704 | Jaramillo <i>et al.</i> , 2016 | plusieurs | Individu entier |
| <i>Macrobrachium rosenbergii</i> | Reads | PRJNA72407 | SRR345608, SRR345609, SRR345610 | | | Branchies, hépatopancréas, muscles |
| <i>Metabetaeus lohena</i> | contigs assemblés | PRJNA244748 | | Havird et Santos, 2014 | 1 | Individu entier |
| <i>Metabetaeus minutus</i> | contigs assemblés | PRJNA244748 | | Havird et Santos, 2014 | 1 | Individu entier |
| <i>Neocaridina denticulata</i> | Reads | PRJNA240382 | SRR1185328 | | | Individu entier |
| <i>Palaemon carinicauda</i> | Reads | PRJNA233983 | SRR1105776 | | | Céphalothorax |
| <i>Palaemon serratus</i> | Reads | PRJNA343473 | SRR4341163 | Perina <i>et al.</i> , 2017 | 1 | Muscle abdominal |
| <i>Palaemon argentinus</i> | Reads | PRJNA309860 | SRR3124666 | | | Individu entier |
| <i>Pandalus latirostris</i> | Reads | PRJDA66833 | DRR001119, DRR001121 | Kawahara-Miki <i>et al.</i> , 2011 | 10 | Ovaires |
| <i>Paratya australiensis</i> | Reads | PRJNA326883 | SRR3720910 | Bain <i>et al.</i> , 2016 | 1 | Individu entier |
| <i>Rimicaris sp.</i> | Reads | PRJNA345278 | SRR5648366 | Zhang <i>et al.</i> , 2017 | 1 | Branchies |

Halocaridinides trigonophthalma, *Metabetaeus lohena*, *Metabetaeus minutus*, *Halocaridina rubra* et *Macrobrachium nipponense*. Le transcriptome de *Macrobrachium nipponense* a été récupéré directement grâce à l'auteur de l'étude qui nous a envoyé les données assemblées. Les données (assemblées ou non) disponibles en base de données après mai 2017 n'ont pas été prises en compte pour cette étude dans la mesure où elles n'apportaient pas de couverture taxonomique supplémentaire. C'est le cas pour les espèces de Palaemonidae *Palaemon suttkusi* et *Palaemon sinensis*.

L'analyse finale comporte donc les transcriptomes de 18 espèces de Caridea disponibles dans la littérature en plus des 8 espèces produites par l'équipe au MNHN. Ces 26 espèces couvrent en tout 8 familles de Caridea. Les données provenant de la littérature n'ajoutent pas de famille supplémentaire à celles de notre jeu de données initial, mais complètent l'échantillonnage au sein de six des huit familles déjà représentées. Notre analyse comporte donc plusieurs transcriptomes pour les familles Alpheidae (N=3 espèces), Alvinocarididae (N=2 espèces), Atyidae (N=8 espèces), Hippolytidae (N=2 espèces), Palaemonidae (N= 7 espèces) et Pandalidae (N=2 espèces). Seules les familles Crangonidae et Oplophoridae n'ont chacune qu'un seul représentant : *Crangon crangon* et *Oplophorus gracilirostris* respectivement. L'utilisation de plusieurs espèces au sein des familles est particulièrement intéressante car elle permet de comparer des distances génétiques inter- et intra-familiales pour les marqueurs génétiques sélectionnés et de potentiellement mettre en évidence des problèmes de paralogie. Ainsi les familles Atyidae et Palaemonidae, pour lesquelles nous avons huit et sept espèces respectivement, sont de bonnes candidates pour évaluer le signal des marqueurs sélectionnés.

IV.3.2 Assemblages des transcriptomes

IV.3.2.1 *Caractérisation des assemblages*

Métriques usuelles

Dans le cas des spécimens séquencés dans l'équipe au MNHN, de 88 à 92% des reads avaient des qualités suffisantes pour être conservés à l'issue du tri opéré par Trimmomatic (voir Annexe 3). L'assemblage primaire des transcriptomes avec Trinity a permis d'assembler entre 110 312 contigs (pour l'espèce *Crangon crangon*) et 265 779 contigs (*Periclimenes brevicarpalis*). Après élimination des formes rares en se basant sur les valeurs de FPKM et suppression des isoformes représentant moins de 1% de l'effort d'échantillonnage, une importante partie des contigs a été supprimée et entre 24 134 (*Atyopsis moluccensis*) et 88 910 (*Caridion steveni*) contigs sont conservés. Suite à l'étape de traitement post-assemblage permettant d'extraire les parties codantes, entre 19 830 et 77 406 séquences codantes sont conservées (pour les espèces d'*Atyopsis moluccensis* et de *Caridion steveni* respectivement).

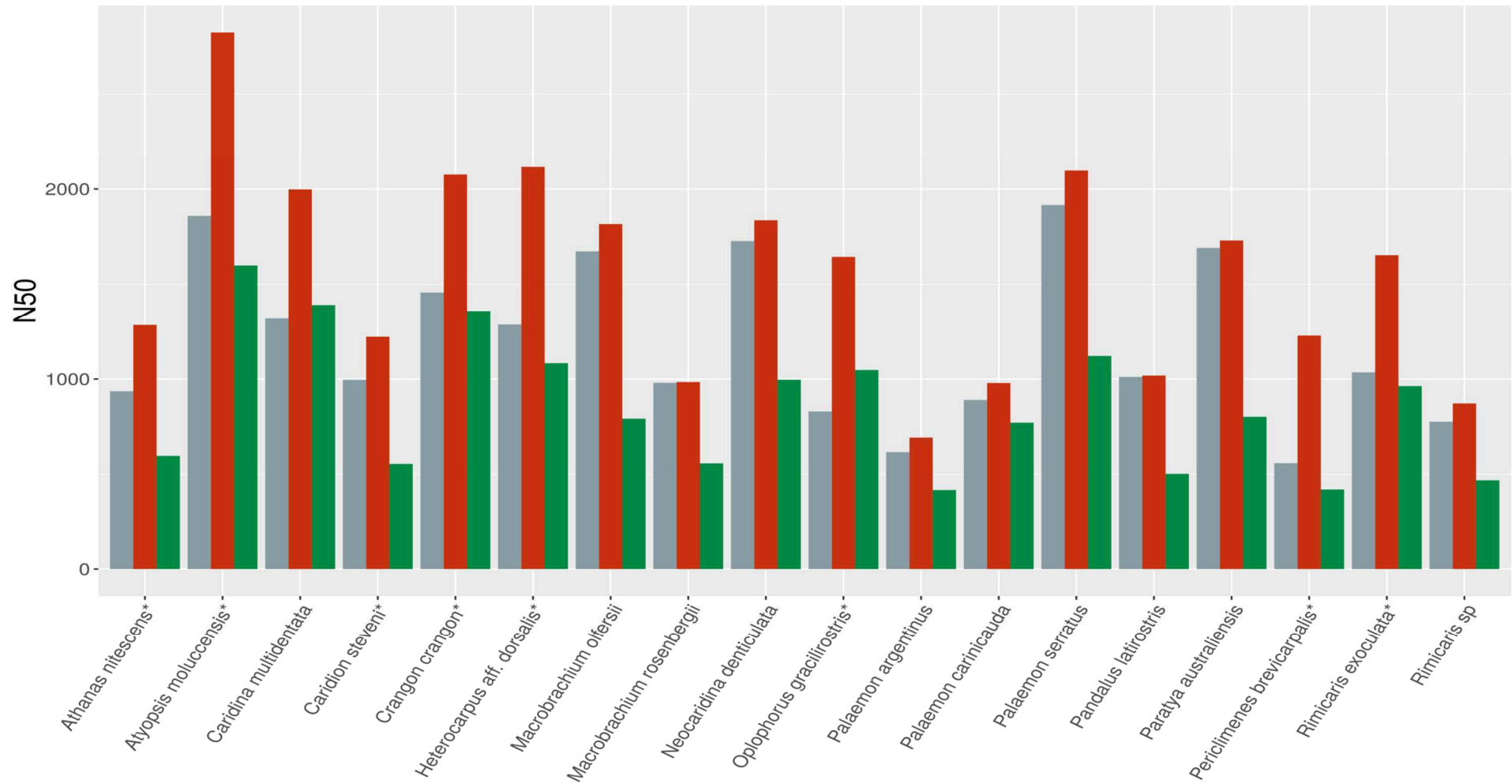


Figure IV.9 Evolution des N50 pour chacune des trois étapes de l'assemblage, pour chaque espèce ayant été assemblée durant cette étude : à l'issue de l'assemblage primaire (gris), après la filtration des contigs peu représentés (rouge) et à la suite du traitement post-assemblage (vert). Les espèces séquencées dans l'équipe sont indiquées par un astérisque.

En considérant l'ensemble des espèces à la fois séquencées dans l'équipe et récupérées dans les bases de données, entre 70 et 100% des reads de séquençage ont été utilisés pour les différents assemblages (respectivement dans le cas de *Pandalus latirostris* et *Halocaridina rubra* -aucun tri des données brutes n'avait été fait avant assemblage pour cette espèce). Après l'étape de traitement post-assemblage, les transcriptomes utilisés pour cette étude comportent entre 15 514 contigs pour l'espèce *Caridina rubella* et 239 512 contigs pour l'espèce *Chorismus antarcticus*.

N50

Après filtration des contigs peu couverts et des isoformes rares, les valeurs de longueurs médiane et moyenne (et donc les N50 des assemblages) augmentent pour tous les transcriptomes (voir tableau Annexe 4 et figure IV.9). Cela indique que ce sont généralement des contigs courts qui ont des valeurs de FPKM ou des abondances d'isoformes faibles, qui sont supprimés lors de cette étape de filtration. Cette augmentation n'est pas uniforme, elle peut être très faible comme dans le cas de *M. rosenbergii* (augmentation du N50 de moins de 1% après la filtration) ou forte comme dans le cas de *Periclimenes brevicarpalis* (augmentation du N50 de 54,5%). Au sein des 8 transcriptomes produits par l'équipe, *Caridion steveni* est l'espèce pour laquelle l'augmentation est la plus faible (augmentation du N50 de 18,5%) et *Periclimenes brevicarpalis* l'espèce avec la plus grande augmentation du N50. Les huit espèces séquencées dans l'équipe ont des valeurs de N50 similaires aux espèces provenant des bases de données. Après le traitement post assemblage, les valeurs de N50 diminuent car les contigs sont tronqués afin de ne garder que la partie codante pour des protéines nucléaires et éliminer les extrémités 5' et 3' UTR. La plus petite diminution du transcriptome est observée chez *Palaemon carinicauda* (le N50 diminue de 21,2 % entre l'assemblage filtré et le jeu de données post-traitement) et la diminution la plus importante est observée pour *Antecaridina lauensis* (le N50 diminue de 70,7%, donnée non présentée car transcriptome non assemblé dans le cadre de cette étude). Les 8 transcriptomes produits dans l'équipe ont des valeurs de diminution des N50 similaires à celles des transcriptomes issus de la littérature, la diminution des valeurs allant de 34,5% à 66,1% pour les espèces *Crangon crangon* et *Periclimenes brevicarpalis* respectivement.

IV.3.3 Recherche et caractéristiques des 425 marqueurs recherchés

IV.3.3.1 Marqueurs retrouvés dans les transcriptomes

Les résultats des blast réciproques effectués pour les espèces séquencées dans l'équipe au muséum contre les 425 protéines du jeu de données de l'oursin fournissent entre 351 et 371

marqueurs (pour *Oplophorus gracilirostris* et *Crangon crangon* respectivement), soit plus de 83% des marqueurs recherchés. Pour les 18 espèces issues de la littérature, entre 266 (pour *Caridina rubella*) et 381 (pour *Neocaridina denticulata*) marqueurs sont retrouvés. Les transcriptomes des espèces *Caridina rubella* et *Metabetaeus minutus* sont les seuls pour lesquels la proportion de protéines identifiées est inférieure à 70%. Les marqueurs mis en évidence dans l'étude des Ophiuroidea sont donc suffisamment conservés pour être retrouvés au sein d'un groupe aussi éloigné phylogénétiquement que les Caridea (divergence estimée entre les Caridea et les Ophiuroidea à ~630 Ma d'après dos Reis *et al.* (2015) ou ~760 Ma d'après Gaya *et al.* (2015)).

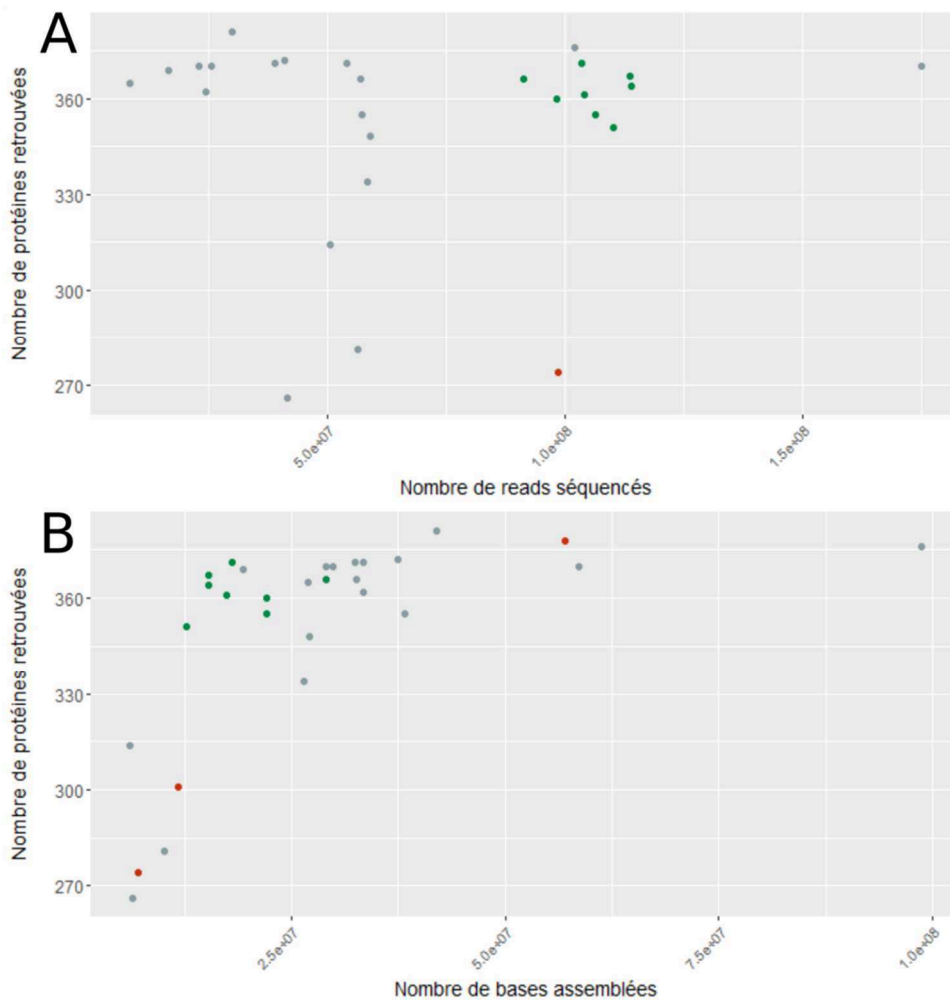


Figure IV.10 Graphiques représentant le nombre de gènes retrouvés en fonction (A) du nombre de reads séquencés et (B) du nombre total de bases effectivement assemblées. Les couleurs différencient les transcriptomes séquencés dans l'équipe (vert), les transcriptomes de Caridea issus de la littérature (gris) et les transcriptomes des espèces non Caridea (rouge). Le spécimen de *Petrolisthes cinctipes* n'est pas représenté, car nous n'avons pas l'information du nombre de reads de séquençage mais 301 marqueurs ont été retrouvés. De même, le spécimen *Penaeus vannamei* n'est pas représenté dans le graphique (A), car il a été séquencé en *single-end* tandis que tous les autres proviennent de séquençage en *paired-end*.

Plusieurs hypothèses pourraient expliquer la variation du nombre de marqueurs retrouvés selon les transcriptomes : relation avec le nombre de reads séquencés, couverture du transcriptome assemblé, type de tissu utilisé pour l'extraction ou encore état physiologique du spécimen et dégradation de l'ARN.

La figure IV.10 (A et B) représente le nombre de marqueurs provenant de l'oursin retrouvés dans les transcriptomes des Caridea, en fonction du nombre de reads séquencés et du nombre de bases assemblées. Dans l'ensemble, l'effort de séquençage (figure IV.10 A) ne permet pas d'expliquer les différences de nombre de gènes retrouvés. Les espèces *Caridina rubella*, *Hyas araneus*, *Metabetaeus minutus* et *Palaemon argentinus* se différencient des autres espèces par le nombre de marqueurs retrouvés nettement plus faible que pour le reste des transcriptomes (266, 274, 281 et 314 respectivement, tandis que les autres ont au moins 334 marqueurs retrouvés). Ces quatre espèces ont entre $4,2 \cdot 10^7$ et $9,9 \cdot 10^7$ reads séquencés, des valeurs semblables à d'autres espèces pour lesquelles plus de gènes ont été retrouvés. 370 marqueurs ont été retrouvés dans le transcriptome de *Macrobrachium nipponense*, l'espèce qui a le plus important nombre de reads séquencés (environ 175 millions). Le transcriptome dans lequel le plus de marqueurs ont été retrouvés est celui de *Neocaridina denticulata* (381), pour lequel environ 30 millions de reads ont été séquencés. Les transcriptomes produits dans l'équipe font partie des transcriptomes ayant le plus grand nombre de reads séquencés. Ils se regroupent et ont des efforts de séquençage similaires aux transcriptomes de *Chorismus antarcticus* et d'*Hyas araneus* (soit : entre $98 \cdot 10^6$ et $102 \cdot 10^6$ reads séquencés). Ils possèdent pourtant le même ordre de grandeur de marqueurs retrouvés que les autres transcriptomes.

La figure IV.10 B présente le nombre de marqueurs tirés de l'oursin retrouvés en fonction du nombre de bases assemblées, qui est un proxy de la couverture de l'assemblage. Dans l'ensemble, en dessous de $12 \cdot 10^6$ bases assemblées le nombre de marqueurs retrouvés ne dépasse pas 301, mais une augmentation est visible avant d'atteindre un plateau entre 360 et 370 marqueurs retrouvés pour toutes les autres espèces (figure IV.10 B). Les espèces ayant la couverture la plus faible correspondent aux espèces dans lesquelles le moins de marqueurs sont retrouvés : *Caridina rubella*, *Hyas araneus*, *Metabetaeus minutus*, *Petrolisthes cinctipes* et *Palaemon argentinus*. Cependant, le transcriptome de *Chorismus antarcticus*, qui a la plus grande couverture (avec environ $99 \cdot 10^6$ bases assemblées), ne correspond pas au transcriptome pour lequel le plus grand nombre de marqueurs est identifié. Le nombre maximal de marqueurs identifiés correspond au transcriptome de *N. denticulata* (381) avec une couverture de $42 \cdot 10^6$ bases. Par ailleurs, les transcriptomes produits dans l'équipe se retrouvent également ensemble et ont des couvertures d'assemblage similaires aux

transcriptomes de *Caridina multidentata*, *Pandalus latirostris*, *Halocaridina rubra*, *Antecaridina lauensis*, *Macrobrachium nipponense* et *Macrobrachium olfersii*. Bien que les transcriptomes produits dans l'équipe fassent partie des transcriptomes possédant les plus grandes profondeurs de séquençage (figure IV.10 A), ils font partie des transcriptomes ayant des couvertures d'assemblage plutôt faibles (moins de $30 \cdot 10^6$ bases assemblées, figure IV.10 B).

Les ARN des transcriptomes provenant de la littérature ont été extraits à partir de différents types de tissus, tandis que les transcriptomes produits au sein de l'équipe étaient extraits sur un même tissu avec le même protocole et avec le même effort de séquençage, ce qui pourrait expliquer le regroupement des transcriptomes produits dans l'équipe (figures IV.10 A et B). Les ARN des transcriptomes provenant de la littérature ont été extraits soit à partir d'individus entiers (ou de mélanges de tissus), soit à partir de tissus spécifiques (et proviennent donc de types de tissus très différents), sauf pour les dix études ayant utilisé des spécimens entiers (voir tableau IV.2). Le nombre de marqueurs identifiés dans les transcriptomes provenant de spécimens entiers se situe entre 266 et 381 (340,8 en moyenne). Le nombre de marqueurs retrouvés dans les transcriptomes provenant de tissus abdominaux (les 8 produits dans l'équipe et *Palaemon serratus*) se situent entre 351 et 371 (362,9 en moyenne). Nous nous attendons à identifier plus de marqueurs dans les séquençages produits à partir de spécimens entiers, car la population d'ARN extraits est plus complète. Cependant, les nombres moyens de marqueurs identifiés sont très similaires entre les transcriptomes produits à partir de spécimens entiers et de tissus musculaires, bien que le minimum et le maximum soient plus élevés dans le cas des séquençages d'ARN extraits de spécimens entiers. Le type de tissu utilisé lors de l'extraction des ARN ne semble donc pas influencer non plus le nombre de marqueurs retrouvés. Cependant, il faudrait plus de répliques (plus de transcriptomes assemblés) pour chacun des types de tissus testés afin de tester plus robustement l'effet de l'impact du tissu utilisé sur le nombre de marqueurs retrouvés.

En plus des différences de tissus choisis pour l'extraction et des conditions physiologiques des organismes, de multiples autres paramètres peuvent expliquer des différences entre les populations d'ARN séquencées. Par exemple : la rapidité de la mise en conservation du spécimen avant son extraction (les ARN ont une demi vie courte et se dégradent rapidement), ou encore des biais expérimentaux durant l'étape d'extraction des ARN (biais dépendant du kit utilisé ou encore de l'expérimentateur). La similarité (et donc le regroupement, voir figure IV.10 B) des transcriptomes produits dans l'équipe illustre potentiellement l'importance de ces effets. En effet, ils ont été extraits par le même expérimentateur en utilisant la même méthode, et la confection des banques a été faite par le même organisme extérieur en suivant un seul protocole.

IV.3.3.2 Distribution des distances nucléotidiques par paires d'espèces et sélection des marqueurs

La figure IV.11 représente le nombre de marqueurs retrouvés à partir des protéines du transcriptome de *Strongylocentrotus purpuratus* parmi les 26 transcriptomes de Caridea utilisés dans notre étude. Les données transcriptomiques récupérées dans la littérature n'apportent pas d'information taxonomique supplémentaire (pas de nouvelles familles) par rapport à celles présentes dans l'échantillonnage effectué au sein de l'équipe au muséum. Ainsi, de façon à ne pas déséquilibrer l'échantillonnage taxonomique et ainsi à ne pas donner une importance différente aux familles selon le nombre d'espèces séquencées, seuls les marqueurs présents dans les huit transcriptomes séquencés dans l'équipe sont utilisés pour sélectionner les marqueurs moléculaires utilisés dans la suite de notre étude. Les huit transcriptomes produits par l'équipe au muséum partagent 265 marqueurs (figure IV.11). Pour chacun de ces 265 marqueurs, les distances nucléotidiques ont été calculées par paire d'espèces pour l'ensemble des 29 transcriptomes utilisés : 18 transcriptomes de la littérature, les 8 transcriptomes produits par l'équipe ainsi que 3 transcriptomes d'espèces non Caridea (groupe externe). Les distances nucléotidiques moyennes par marqueur entre les 8 transcriptomes produits dans l'équipe (ligne noire sur la figure IV.12) se répartissent entre 0,06 pour le marqueur correspondant à *tpm3* (une isoforme de la tropomyosine) et 0,45 pour le marqueur correspondant à *her6* (un facteur de transcription). La distribution des distances par paire d'espèces pour chacun des 265 marqueurs rangés par distance nucléotidique moyenne croissante (figure IV.12) permet de dégager plusieurs informations. Notamment, il est possible de distinguer trois zones principales dans les distributions de distances nucléotidiques: les deux premiers marqueurs avec des distances moyennes faibles de 0,06 et 0,10 (donc très conservés), puis une zone dans laquelle les distances augmentent plus modérément, avant une nouvelle augmentation rapide des distances pour les 7 derniers marqueurs. Dans l'ensemble, pour chacun des marqueurs, les distances entre paire d'espèces est cohérente avec la hiérarchie linnéenne : les distances nucléotidiques entre espèces appartenant à une même famille (en rouge dans la figure IV.12) sont généralement inférieures aux distances nucléotidiques entre espèces appartenant à des familles différentes (en gris dans la figure IV.12). La prise en compte d'une séquence qui n'est pas la copie orthologue du marqueur considéré peut expliquer les exceptions à cette tendance, car toutes les distances qui lui sont appariées sont relativement grandes par rapport aux autres distances calculées pour ce marqueur. Si des contaminations aient eu lieu, cela va avoir pour conséquence de réduire à 0 des distances entre espèces de familles différentes. De plus, dans le cas de familles polyphylétiques,

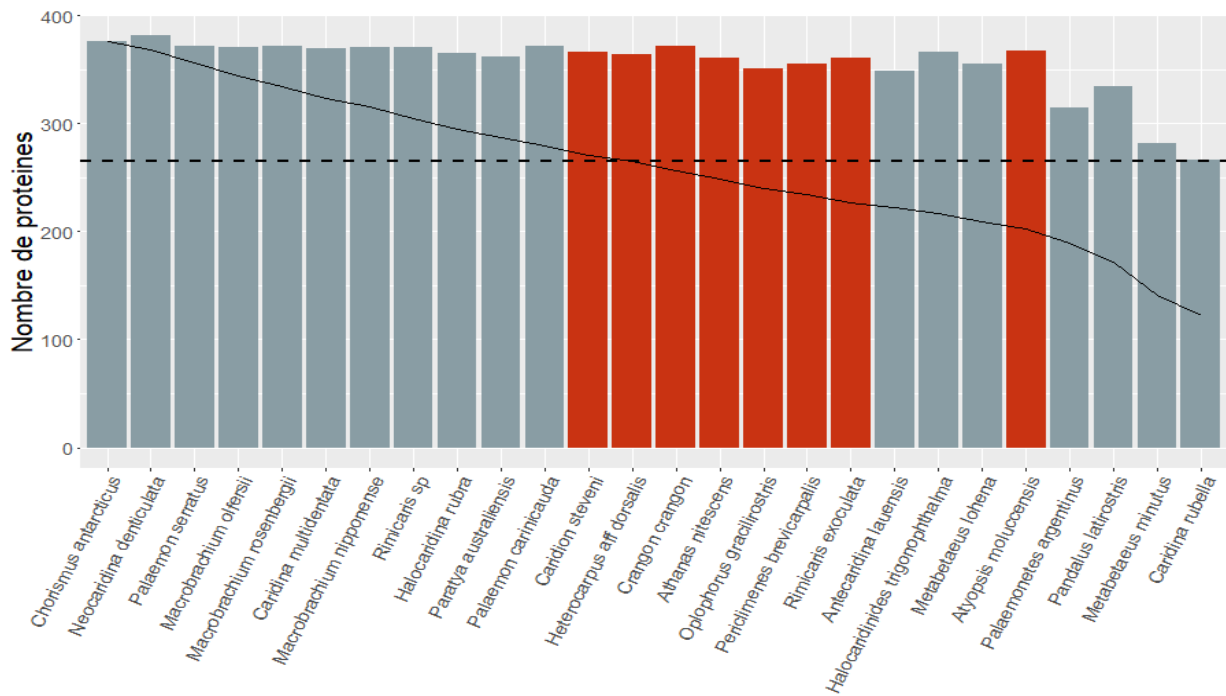


Figure IV.11 Nombre de marqueurs identifiés dans les 26 transcriptomes de Caridea retenus. Les couleurs correspondent à l'origine du transcriptome : la littérature (gris), produits par l'équipe (rouge). Les transcriptomes sont rangés tel que le nombre de marqueurs partagés reste maximal en ajoutant les espèces. La ligne en pointillée indique le nombre de marqueurs partagés entre les transcriptomes produits dans l'équipe (265). La ligne pleine indique le nombre de marqueurs en commun lorsqu'on ajoute un transcriptome. Ainsi, 124 marqueurs sont communs lorsqu'on considère toutes les espèces de Caridea utilisées dans notre étude.

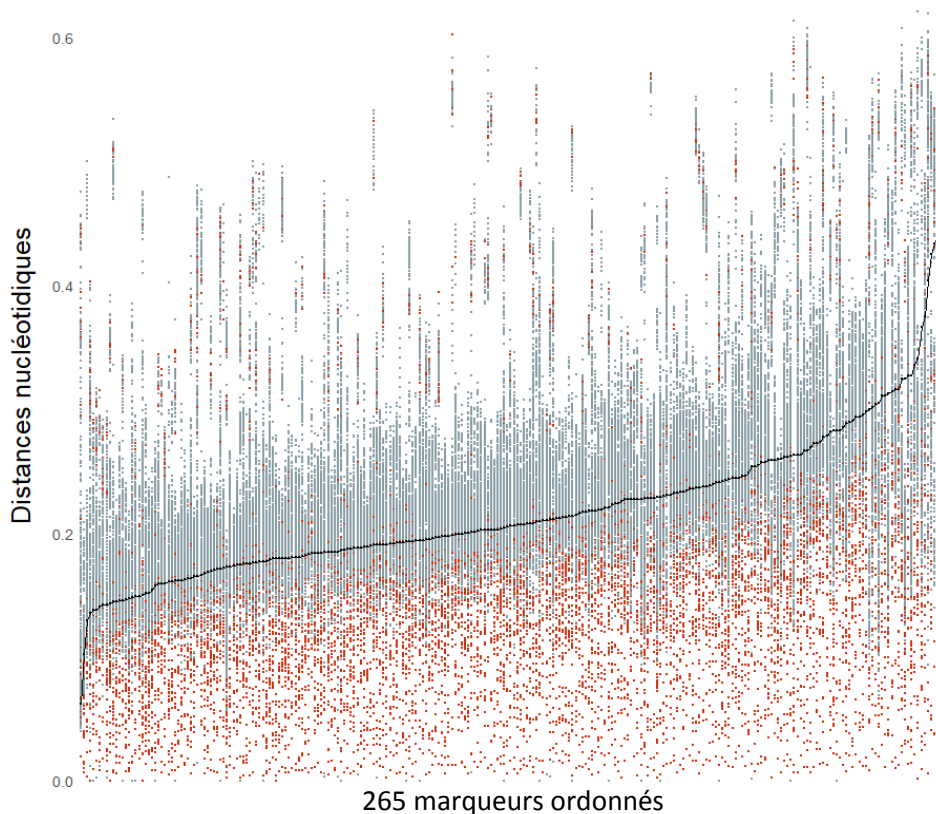


Figure IV.12 Distribution des distances nucléotidiques entre chaque paire d'espèces pour les 29 espèces sélectionnées et pour les 265 marqueurs partagés entre les 8 transcriptomes de l'équipe. Les marqueurs sont rangés par ordre croissant de leur distance nucléotidique moyenne parmi les 8 transcriptomes séquencés dans l'équipe. Les distances représentées en rouge sont calculées entre 2 espèces de la même famille ; les distances représentées en gris sont calculées entre 2 espèces de famille différentes ; le trait noir représente la moyenne calculée pour les huit transcriptomes, pour chaque marqueur.

des distances intrafamiliales importantes (comparable à des distances interfamiliales) peuvent être observées (par exemple les Palaemonidae, Aznar-Cormano *et al.*, 2015).

Le but de cette étude est de définir un ensemble de marqueurs suffisamment conservés pour résoudre les relations phylogénétiques profondes des Caridea. Les marqueurs ayant moins de 15% de divergence moyenne sont au nombre de 20. Ces 20 marqueurs les plus conservés semblent les plus favorables pour une analyse phylogénétique profonde. Afin de tester cette hypothèse, nous avons également sélectionné les 20 gènes les plus divergents pour une analyse phylogénétique. Afin de vérifier si les séquences des marqueurs retrouvés par la méthode des blast réciproques sont réellement toutes des copies orthologues, nous avons travaillé sur ces 40 marqueurs. Les 20 marqueurs les plus conservés correspondent principalement à des protéines ribosomales et des ubiquitines (annotations faites pour les Ophiuroidea ; O'Hara *et al.*, 2015) et ont une valeur de distance moyenne de 0,137 (avec en moyenne des écart-types de 0,023). La distance moyenne des 20 marqueurs les plus variables est 0,353 avec en moyenne, des écart-types bien plus importants que pour le jeu de marqueurs conservés (0,115).

IV.3.4 Evaluation de l'intérêt des marqueurs sélectionnés pour des études phylogénétiques

IV.3.4.1 Détection et élimination des séquences paralogues ou ne correspondant pas au locus recherché

Troisième étape de vérification, basée sur le critère des distances nucléotidiques

La dernière étape d'élimination des séquences paralogues dépend de l'échantillonnage taxonomique. En effet, elle est basée sur des valeurs seuil de divergence moyenne entre paires d'espèces, et prend donc en compte la redondance taxonomique au sein des familles échantillonnées. Les distances interfamiliales peuvent être importantes par rapport aux distances intrafamiliales. De ce fait, si l'échantillonnage taxonomique engendre un déséquilibre en faveur des distances interfamiliales par rapport aux intrafamiliales (grande diversité familiale, peu de redondance taxonomique au sein des familles), la valeur moyenne des distances calculée sera grande et peut masquer des distances engendrées par une séquence paralogue. A l'inverse pour un échantillonnage taxonomique déséquilibré en faveur des distances intrafamiliales (importante redondance taxonomique au sein des familles mais peu de familles différentes), la valeur moyenne des distances calculée sera faible. Dans ce cas, les distances calculées à partir d'une espèce un peu éloignée phylogénétiquement des autres seront systématiquement plus importantes que la valeur seuil estimée et sa séquence sera considérée comme paralogue (faux positif). Ainsi, afin que le déséquilibre dans l'échantillonnage taxonomique (certaines familles sont beaucoup plus représentées que d'autres ; typiquement les Atyidae avec 8 espèces et les Palaemonidae avec 7 espèces) n'engendre pas de tels biais, chaque paire d'espèces est classée parmi quatre catégories et une valeur seuil est calculée pour chaque catégorie. Les espèces du groupe externe ne sont pas prises en compte dans ces catégories. Ces catégories sont :

- (i) paire interfamiliale d'espèces ;
- (ii) paire intrafamiliale d'espèces pour deux espèces qui ne sont ni Atyidae ni Palaemonidae ;
- (iii) paire intrafamiliale d'espèces entre deux espèces d'Atyidae ;
- (iv) paire intrafamiliale d'espèces entre deux espèces de Palaemonidae.

Une valeur seuil est calculée pour chacune de ces catégories, à partir des distances entre les paires d'espèces qu'elle comporte, à l'exception de la catégorie (ii), pour laquelle les distances

intrafamiliales concernant des Atyidae et des Palaemonidae sont également prises en compte dans le calcul, afin de toujours calculer des moyennes à partir de plus de trois valeurs. De même, pour les catégories (iii) et (iv), il a préalablement été vérifié que les familles Atyidae et Palaemonidae sont représentées par plus de deux espèces. La valeur seuil correspond à la somme de la moyenne arithmétique et de l'écart-type des distances appariées considérées. Une séquence est considérée comme paralogue si plus de la moitié des distances qui lui sont relatives dépassent les seuils calculés.

Cependant, en gardant ce calcul pour déterminer les valeurs seuils lors des boucles de vérification suivantes (une boucle correspondant à l'ensemble formé par les trois étapes de vérification décrites en matériel et méthodes), plusieurs séquences ne répondent pas au critère de distance : 16 séquences lors de la 2^{ème} boucle, 9 lors de la 3^{ème} boucle et 5 lors de la 4^{ème} boucle. Une seule séquence parmi ces 30 a été remplacée à la suite de la recherche d'une séquence plus probablement orthologue, aucun résultat de blast n'a été trouvé pour les autres. Le fait de ne pas obtenir de résultat de blast signifie soit (i) que les séquences qui n'ont pas répondu au critère de distance correspondaient à la copie orthologue ; dans ce cas, le seuil fixé est trop stringent et exclut des séquences qui sont biologiquement plus divergentes que la gamme tolérée par le seuil ; (ii) soit que la copie orthologue de ce gène n'existe pas dans le transcriptome assemblé. Dans ce dernier cas, cela peut être dû à un assemblage incomplet ne permettant pas de retrouver la copie orthologue, ou au fait que cette copie soit éteinte dans la lignée concernée. La valeur seuil calculée pour l'étape de vérification basée sur les distances nucléotidiques est *a priori* trop stricte et supprime des séquences qui sont simplement différentes biologiquement. Pour cette raison, à partir de la deuxième boucle de vérification le calcul des valeurs seuil pour le critère basé sur les distances nucléotidiques diffère. La valeur seuil est alors calculée comme la somme de la moyenne et de l'écart-type des paires de distances entre les séquences des espèces du groupe externe et toutes les autres séquences. Ce seuil n'est pas appliqué dès la première boucle de vérification, car le trop grand nombre de séquences paralogues (donc avec des distances plus proches du groupe externe que du groupe interne) aurait artificiellement augmenté la valeur seuil.

Cas des contaminations entre les banques

Des contaminations (distances nucléotidiques nulles) entre espèces de genres différents ont été détectées dans 5 des 20 marqueurs conservés. Ces contaminations ont été retrouvées entre les couples d'espèces :

- *Caridina rubella* et *Metabetaeus lohena*, provenant de l'étude de Havird et Santos (2014) (contamination retrouvée pour deux marqueurs) ;
- *Atyopsis moluccensis* et *Rimicaris exoculata* provenant de notre étude (contamination retrouvée pour deux marqueurs) ;
- *Caridion steveni* et *Periclimenes brevicarpalis* provenant de notre étude (contamination retrouvée pour un marqueur)

Chacun de ces trois couples est composé d'espèces ayant été séquencées sur la même ligne de séquençage. A l'échelle des reads, il n'est pas rare de rencontrer de la contamination, pour des séquençages effectués sur les mêmes équipements (Ballenghien *et al.*, 2017). Cependant, retrouver des contaminations à l'échelle des contigs signifie que le nombre de reads issus de contamination est potentiellement suffisamment important pour permettre de reconstituer des transcrits *de novo* (entiers ou partiels) provenant d'une autre banque de séquençage.

Détection et élimination des séquences paralogues

Après élimination des paralogues, deux marqueurs parmi les marqueurs divergents ne sont pas conservés : le marqueur codant pour la protéine *elk* (un facteur de transcription) pour lequel des séquences ont été trouvées pour 14 espèces et le marqueur codant pour une protéine de la famille des cyclin-dependent kinases (*cdk9*) pour lequel des séquences ont été trouvées seulement pour 10 espèces. De plus, un marqueur parmi le groupe de marqueurs les plus conservés n'est pas non plus conservé pour l'analyse car il est très similaire à un autre marqueur de ce groupe de marqueurs (les 80 premiers acides aminés sont identiques) et ces deux marqueurs correspondent peut être simplement à des transcrits issus d'épissages alternatifs d'un unique ARNm. Par ailleurs, il faut noter que la séquence de la protéine de référence que nous avons utilisée pour chercher l'un de ces deux marqueurs a été supprimée de la base de données NCBI. Ainsi, parmi les 20 marqueurs conservés et les 20 marqueurs divergents, respectivement 19 et 18 marqueurs ont été utilisés pour la suite des analyses.

Parmi les marqueurs les plus divergents, un plus grand nombre de séquences n'a pas passé au moins une des trois étapes de vérification (en moyenne 4,9 séquences/marqueur) par rapport aux plus conservés (en moyenne 2,8/marqueur). Parmi ces séquences ne répondant pas au moins à un des trois critères, la procédure mise en œuvre a permis de trouver une séquence plus probablement orthologue dans environ 74% des cas pour les gènes conservés, contre seulement environ 48% pour

les gènes divergents. Dans le reste des cas, aucune séquence alternative n'a été trouvée et la donnée est considérée comme manquante.

Sur les 29 transcriptomes sélectionnés, les marqueurs conservés comportent entre 0 et 6 séquences manquantes (en moyenne 0,9 par marqueur). Après recherche des séquences plus probablement orthologues, la recherche de marqueurs par blast réciproques a permis d'obtenir des séquences pour 35% d'entre elles, les 65% restant n'ont pas été retrouvées et sont traitées comme des données manquantes. Dans le cas des marqueurs divergents, les marqueurs comportaient entre 0 et 4 séquences manquantes (en moyenne 2 par marqueur), pour lesquelles une séquence de remplacement a été obtenue par blast réciproques dans seulement 11% des cas.

La matrice finale pour les 19 gènes conservés n'a en tout que 4,7% de données manquantes et celle pour les 18 marqueurs les plus divergents contient 14,9% de données manquantes.

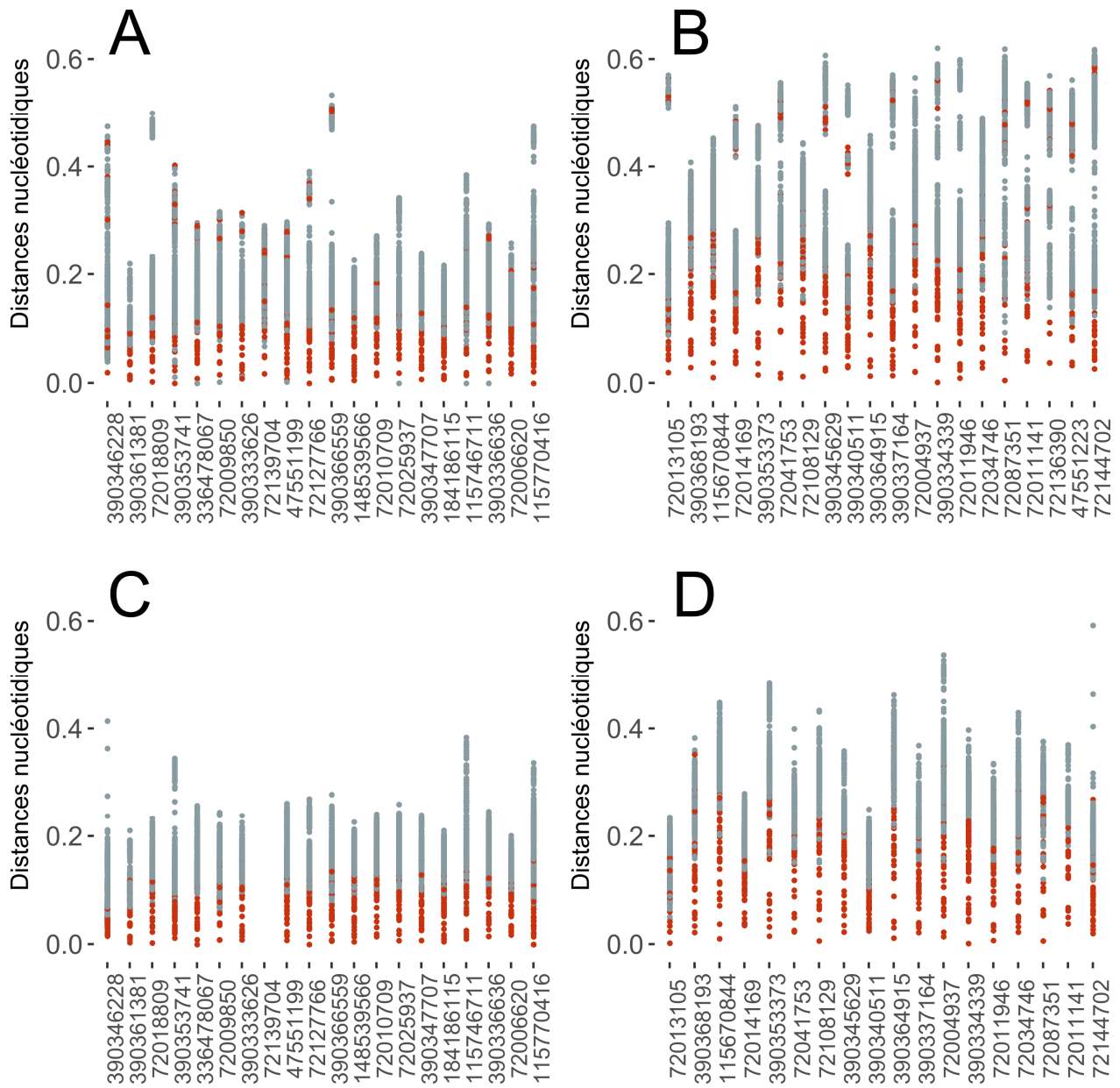


Figure IV.13 Distances nucléotidiques entre paires d'espèces pour les 40 marqueurs sélectionnés pour les analyses. Les graphiques présentent les distances calculées sur les données avant (en haut) et après vérification de l'orthologie (en bas), pour les marqueurs conservés (à gauche) et les marqueurs divergents (à droite). Les points gris correspondent aux distances interfamiliales et les points rouges aux distances intrafamiliales.

IV.3.4.2 Effet des vérifications d'orthologie sur les distances nucléotidiques

Les distances nucléotidiques entre paires d'espèces ont été calculées pour les 37 marqueurs retenus (19 conservés et 18 divergents) sur les séquences brutes et après vérification de l'orthologie. La figure IV.13 présente, pour chacun des marqueurs sélectionnés, les distances par paires d'espèces avant et après vérification de l'orthologie des séquences. Les figure IV.13 A et B (séquences brutes, non nettoyées) correspondent à un zoom des zones les plus conservées et les plus divergentes, respectivement, de la figure IV.12. Parmi les gènes conservés avant vérification, une discontinuité dans la distribution des valeurs de distances appariées est visible pour 4 marqueurs sur 20 (figure IV.13 A). Ces discontinuités correspondent à un ensemble de distances appariées bien plus importantes que le reste des distances calculées pour ce marqueur. Pour les gènes divergents, 13 locus sur 20 présentent au moins une discontinuité, parfois deux (figure IV.13 B). Ces discontinuités sont éliminées à la suite de la vérification de l'orthologie (figures IV.13 C et D), montrant que la boucle de vérification a bien éliminé les séquences non homologues. Les 3 distances supérieures à 0,4 dans le cas du dernier marqueur (72144702) impliquent systématiquement l'espèce *Caridina rubella*, car la séquence pour cette espèce ne couvre que la moitié de l'alignement, et ne recouvre les séquences des espèces *Metabetaeus minutus*, *M. lohena* et *Hyas araneus* que de quelques nucléotides.

La figure IV.14 présente les dispersions des moyennes et des ratios variance-moyennes (RVM) des distances par paires d'espèces calculés sur chaque marqueur. Après vérification de l'orthologie, nous avons bien une diminution des valeurs moyennes de distance : pour les marqueurs conservés, la médiane passe de 0,13 à 0,092 pour les distances intrafamiliales, et de 0,17 à 0,15 pour les distances interfamiliales ; pour les marqueurs divergents, la médiane passe de 0,22 à 0,16 pour les distances intrafamiliales et de 0,33 à 0,25 pour les distances interfamiliales (figure IV.14)

Les discontinuités visibles avant élimination des paralogues (figure IV.13 A et B) expliquent les importantes valeurs de RVM retrouvées aussi bien dans le cas des marqueurs conservés que divergents. De même, le resserrement des distances visible (figure IV.13 C et D) se retrouve dans les données moyennées par marqueur (figure IV.14 B) : dans le cas des marqueurs conservés, les valeurs médianes de RVM passent de 0,034 à 0,016 pour les distances intrafamiliales et de 0,01 à $2,7 \cdot 10^{-3}$ dans le cas des distances interfamiliales.

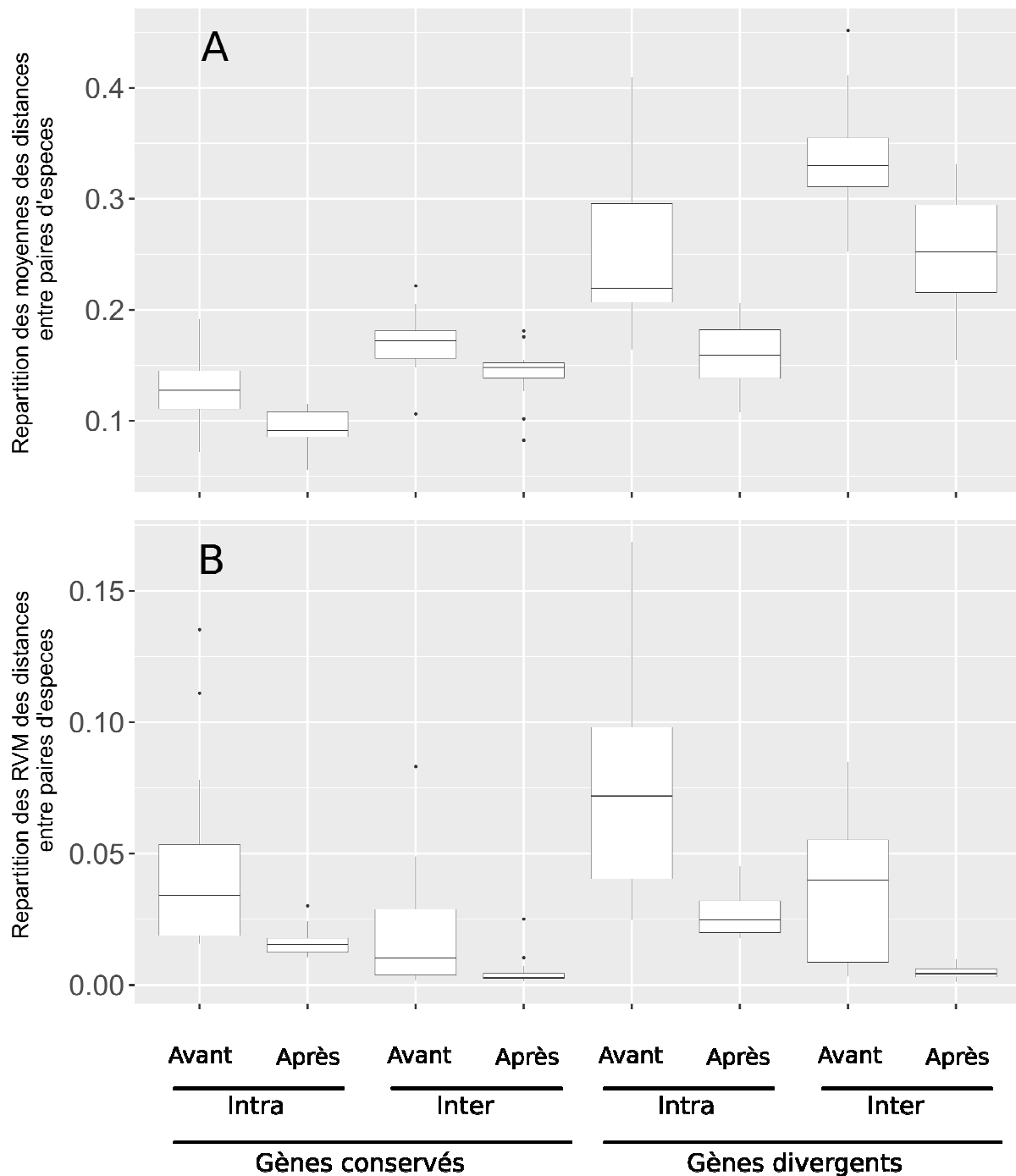


Figure IV.14 Diagrammes en boîtes représentant la répartition des moyennes des distances entre paires d'espèces (A) et des Ratios Variance Moyenne (B) pour les marqueurs sélectionnés, avant et après vérification de l'orthologie des séquences. Ces calculs ont été faits pour des distances appariées entre espèces d'une même famille (intra-famille : « Intra ») et entre espèces n'appartenant pas à la même famille (inter-famille : « Inter »). Les diagrammes en boîtes représentent le 1er quartile, la médiane et le 3ème quartile des données considérées ; les ymin et ymax représentés par les moustaches correspondent à la plus petite/grande observation supérieure/inférieure ou égale au 1er /3ème quartile plus/moins 1,5 fois l'écart interquartile.

En accord avec la hiérarchie linnéenne, les moyennes des distances interfamiliales sont donc plus grandes que les distances intrafamiliales (figure IV.14 A) : la médiane est de 0,092 pour les distances intrafamiliales contre 0,15 pour les distances interfamiliales dans le cas des données qui ont été vérifiées et dans le cas des marqueurs conservés ; la médiane est de 0,16 pour les distances intrafamiliales contre 0,25 pour les distances interfamiliales dans le cas des marqueurs divergents après vérification de l'orthologie. D'autre part, le graphique montre que les médianes des RVM des distances interfamiliales sont plus basses que celles des distances intrafamiliales (à la fois pour les données avant et après vérification de l'orthologie). Ainsi, dans le cas des marqueurs conservés et après vérification, la médiane des RVM est de $1,6 \cdot 10^{-2}$ pour les distances intrafamiliales et de $2,7 \cdot 10^{-3}$ pour les distances interfamiliales. Un échantillonnage hétérogène au sein des familles est peut être responsable de cette différence. En effet, notre jeu de données comporte plusieurs espèces d'un même genre pour certaines familles et plusieurs genres pour d'autres.

Un résultat remarquable est que dans le cas des marqueurs divergents, l'écart interquartile des moyennes des distances interfamiliales est plus important après qu'avant vérification de l'orthologie (figure IV.14 A). En effet avant vérification, les moyennes des distances pour chacun des marqueurs sont grandes (médiane à 0,33) avec des ordres de grandeur similaires les unes aux autres (écart interquartile de 0,054) et les RVM correspondants sont importants : la médiane est à 0,04 avec un écart interquartile de 0,048, signifiant qu'au sein de chaque marqueur les distances sont très dispersées autour de la moyenne (figure IV.14 B). Après la vérification de l'orthologie, la médiane de la dispersion des moyennes est plus faible (0,25) mais l'écart interquartile est plus important (0,084) (figure IV.14 A), ce qui signifie qu'après vérification, les valeurs des divergences moyennes entre les marqueurs sont plus différentes. Cependant, non seulement les RVM correspondants sont très faibles (médiane de $4,4 \cdot 10^{-3}$), mais l'écart interquartile l'est également ($3,2 \cdot 10^{-3}$) traduisant le fait que les dispersions des distances sont très proches de la valeur moyenne pour chaque marqueur. Ainsi dans ce groupe, nous trouvons donc après vérification, une plus grande dispersion des moyennes, même si pour chaque marqueur, la dispersion des valeurs est très faible.

N = 29

N = 26

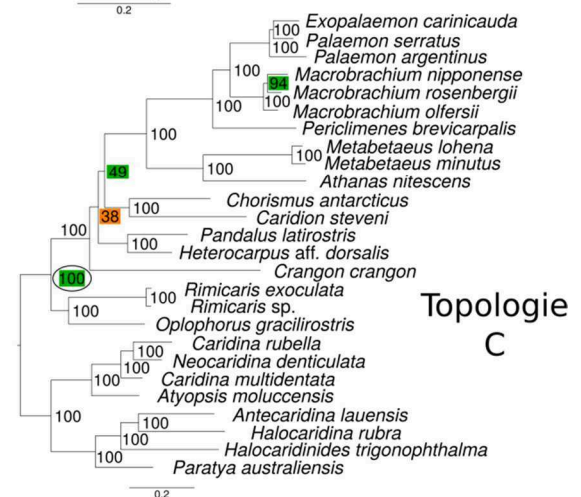
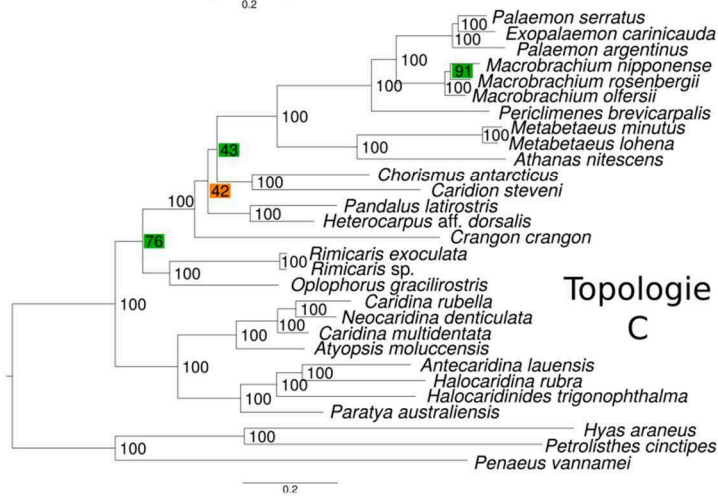
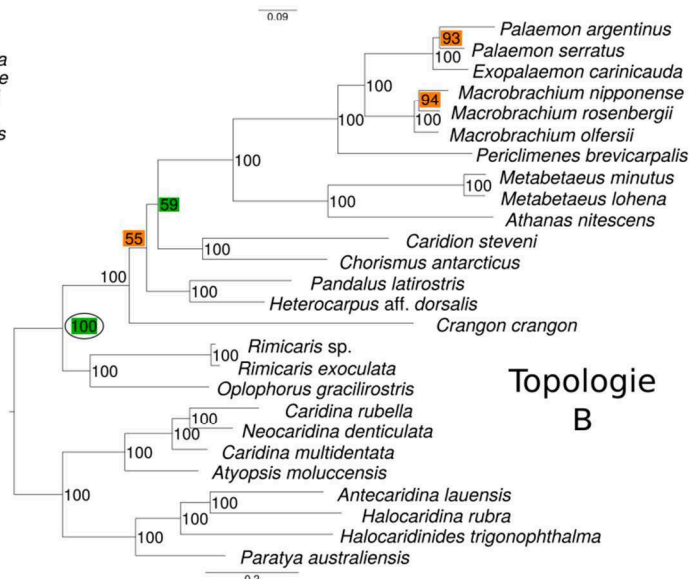
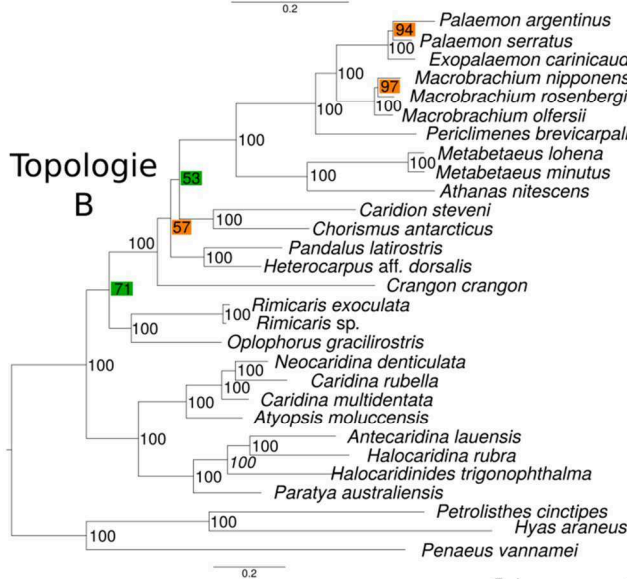
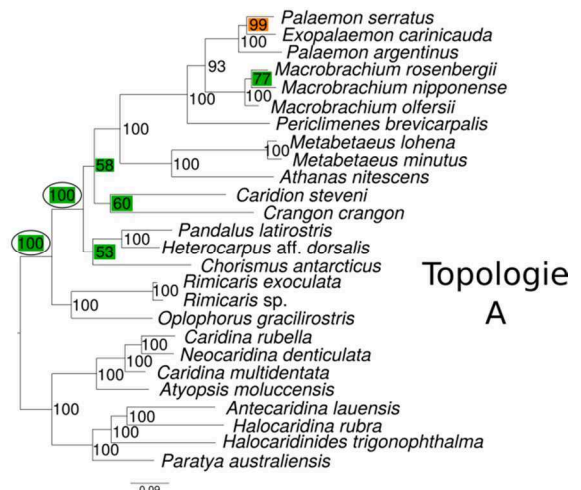
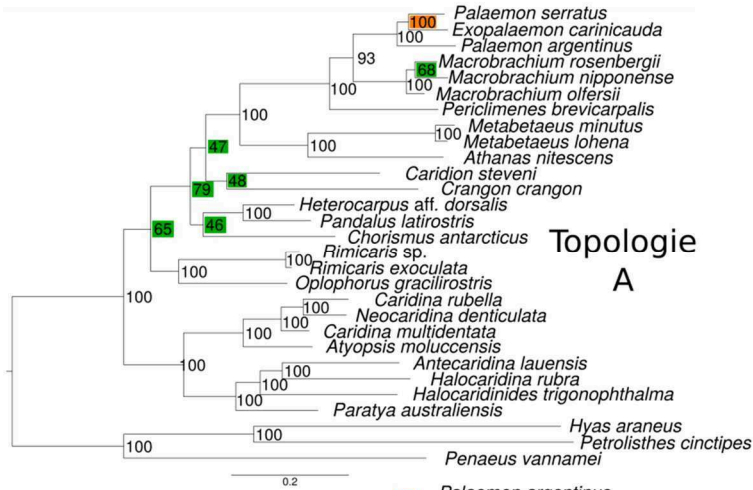


Figure IV.15 Topologies ML des analyses phylogénétiques pour les 19 marqueurs conservés (en haut), 18 divergents (milieu) et de la concaténation des 37 marqueurs (bas) pour les analyses effectuées avec 29 taxons (à gauche) et 26 taxons (à droite). Les topologies A et B correspondent aux topologies avec les mêmes noms sur la figure IV.16. Les couleurs indiquent une différence dans le soutien du nœud entre l'analyse à 29 et 26 taxons, pour un même jeu de données : en vert, le soutien est plus élevé dans l'analyse à 26 taxons, en orange, le soutien est plus faible dans l'analyse à 26 taxons. Les nœuds entourés correspondent aux relations soutenues dans l'analyse à 26 taxons mais pas soutenues dans l'analyse à 29 taxons. Pour faciliter la lecture, les noms des clades I, II.a et II.b utilisés dans le texte ne sont indiqués que sur une des topologies.

N = 29

N = 26

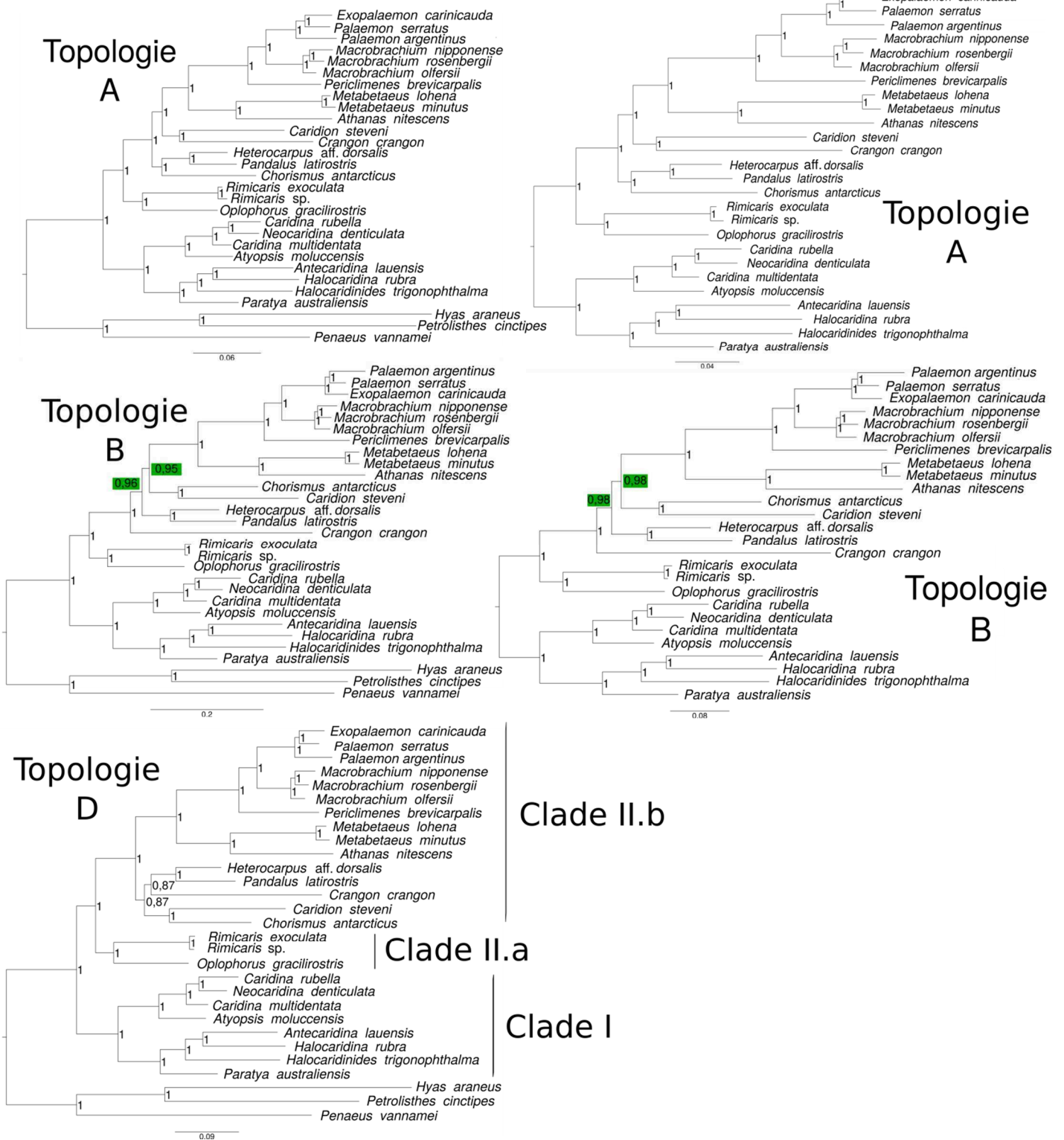


Figure IV.16 Topologies IB des analyses phylogénétiques pour les 19 marqueurs conservés (en haut), 18 marqueurs (milieu) et de la concaténation des 37 marqueurs (bas) pour les analyses effectuées avec 29 taxons (à gauche) et 26 taxons (à droite). Les topologies A et B correspondent aux topologies avec les mêmes noms sur la figure IV.15. Les couleurs indiquent une différence dans le soutien du nœud entre l'analyse à 29 et 26 taxons, pour un même jeu de données : en vert, le soutien est plus élevé dans l'analyse à 26 taxons, en orange, le soutien est plus faible dans l'analyse à 26 taxons. Les nœuds entourés correspondent aux relations soutenues dans l'analyse à 26 taxons mais pas soutenues dans l'analyse à 29 taxons.

IV.3.5 Analyses phylogénétiques

En partant des trois jeux de données assemblés (les 19 marqueurs conservés, les 18 divergents, et l'ensemble des 37 marqueurs) avec 29 taxons incluant les groupes externes non Caridea ou 26 taxons en utilisant les espèces d'Atyidae en groupe externe des autres Caridea, nous avons produit des phylogénies en IB et en ML. Nous n'avons pas pu reconstruire l'arbre bayésien à partir de la concaténation des 37 marqueurs pour les 26 taxons, car les paramètres estimés lors de l'analyse bayésienne ne se sont jamais stabilisés (ESS < 200).

Dans cinq des six analyses effectuées avec 29 taxons, trois grands clades sont retrouvés de manière soutenue pour chacune des analyses (figures IV.15 et IV.16). Le clade I correspond à la famille des Atyidae et sa position comme groupe frère de l'ensemble des autres Caridea présents dans notre étude est soutenue dans toutes les analyses. Ce résultat est en accord avec Aznar-Cormano *et al.* (2015) et nous permet également de pouvoir faire les analyses pour chacun des trois jeux de données en utilisant ce clade pour enraciner le reste des Caridea. Le clade II correspond à l'ensemble des autres espèces, au sein duquel nous pouvons distinguer le clade II.a qui comprend les deux Alvinocarididae et l'espèce d'Oplophoridae, et le clade II.b comprenant les espèces des familles Alpheidae, Crangonidae, Hippolytidae, Palaemonidae et Pandalidae. Les relations phylogénétiques au sein des clades I et II.a sont les mêmes dans l'ensemble des analyses. Les Atyidae (clade I) sont monophylétiques et deux clades se distinguent. La monophylie du genre *Caridina* n'est pas retrouvée avec nos données transcriptomiques, confirmant une étude publiée avec des marqueurs phylogénétiques classiques (Von Rintelen *et al.*, 2012). Le clade II.a contient les deux espèces d'Alvinocarididae et *Oplophorus gracilirostris*, une espèce de la famille Oplophoridae. Le nœud qui relie le clade II.a au clade II.b n'est pas soutenu dans les trois analyses ML pour lesquelles le groupe externe est composé d'espèces non-Caridea, mais il l'est lorsque l'enracinement est fait sur les Atyidae. Ceci montre l'importance du choix des groupes externes dans les reconstructions phylogénétiques. Dans le cas où le groupe externe est trop éloigné, plusieurs événements de substitutions nucléotidiques peuvent avoir eu lieu sur un même nucléotide entre les espèces comparées, et finalement aboutir à une convergence nucléotidique homoplasique entre les deux séquences. De ce fait, certaines espèces peuvent se rapprocher artificiellement du groupe externe. Cela se manifeste par des supports plus faibles aux nœuds. Bien que ce phénomène soit principalement attendu pour les marqueurs les plus divergents, il est également rencontré dans le jeu de données conservé. Cela signifie qu'il faudrait trouver des marqueurs encore plus conservés. Il est aussi possible que la duplication de locus (ou la perte secondaire d'un locus initialement

dupliqué) dans l'une de ces lignées génère l'inclusion d'un paralogue dans l'alignement, perturbant ainsi la phylogénie.

En ce qui concerne les espèces du clade II.b, il y a des variations de topologie selon les analyses et les jeux de données. Ces différences concernent d'une part les relations interfamiliales entre les espèces *Chorismus antarcticus*, *Crangon crangon*, *Caridion steveni* et les deux espèces de Pandalidae et d'autre part les relations au sein des Palaemonidae entre *Palaemon serratus*, *P. carinicauda* et *Palaemon argentinus*. Seules les deux analyses bayésiennes réalisées avec les 19 marqueurs conservés ont des soutiens maximaux à tous les nœuds.

Sur les 12 phylogénies effectuées, 4 topologies différentes sont retrouvées, mais les nœuds où l'on observe les différences n'ont pas de soutien. Les analyses réalisées avec les 19 marqueurs conservés mènent à une topologie (A) dans laquelle les Hippolytidae sont polyphylétiques car *Crangon crangon* (Crangonidae) forme un groupe avec *Caridion steveni* (soutenu en IB mais pas en ML) et *Chorismus antarcticus* forme un groupe avec les Pandalidae (soutenu en IB mais pas en ML). Les analyses réalisées avec les 18 marqueurs les plus divergents fournissent une seconde topologie (B) dans laquelle les Hippolytidae sont monophylétiques (supports maximaux en IB et ML) et le Crangonidae *Crangon crangon* se place en groupe frère de toutes les autres espèces du clade II.b. Une troisième topologie (C) est retrouvée pour l'analyse de maximum de vraisemblance (ML) effectuée en concaténant l'ensemble des 37 marqueurs. Elle est similaire à la seconde topologie (B) hormis au sein des Palaemonidae pour lesquels : en topologie (C) *Palaemon serratus* est plus proche de *Palaemon carinicauda* que *Palaemon argentinus*, tandis qu'en topologie (B) *P. serratus* est plus proche de *P. argentinus* que de *P. carinicauda*. Enfin, la dernière topologie (D) est retrouvée pour l'analyse bayésienne effectuée en concaténant l'ensemble des 37 marqueurs (avec 29 taxons) : les Hippolytidae (monophylétiques dans cette analyse), les Pandalidae (monophylétiques dans cette analyse) et le Crangonidae forment un groupe monophylétique. Cependant les soutiens de ce clade sont faibles : 0,87 (enracinement sur les espèces non Caridea).

Les analyses ML de cette étude montrent que choisir un groupe externe le plus proche possible du groupe d'intérêt permet de mieux résoudre les relations profondes qu'en utilisant un jeu de marqueurs plus conservés avec un enracinement sur des espèces éloignées. Par contre, les nœuds récents sont mieux résolus en utilisant les 18 marqueurs les plus divergents que les autres jeux de données (relations entre les couples d'espèces *Palaemon argentinus* / *Palaemon serratus* et *Macrobrachium nipponense* / *M. rosenbergii*).

L'analyse bayésienne fournit un arbre ayant des soutiens maximaux pour le jeu de données avec les marqueurs conservés (figure IV.16), mais la topologie est différente et les soutiens

diminuent (deux nœuds à 0,87) lorsqu'on leur ajoute les 18 marqueurs divergents. Cela est peut-être dû au bruit apporté par les marqueurs divergents (convergences nucléotidiques artéfactuelles). Il est nécessaire d'augmenter le nombre de taxons pour pouvoir tester cette hypothèse.

Dans notre étude, en plus des Alvinocarididae, la monophylie est retrouvée pour quatre des cinq familles représentées par au moins deux espèces (Alpheidae, Atyidae, Palaemonidae et Pandalidae). Pour les Hippolytidae, la monophylie n'est pas retrouvée avec le jeu de données des 19 gènes conservés (en IB : support maximal pour le clade *Caridion steveni* et *Crangon crangon*, un Crangonidae). De Grave et collaborateurs (2014) séparent les Hippolytidae en 5 familles distinctes. Dans cette étude, l'espèce *Chorismus antarcticus* fait partie des espèces incluses dans la phylogénie moléculaire, ce qui n'est pas le cas du genre *Caridion*. Nos résultats suggèrent que ce genre n'appartient pas à la famille des Hippolytidae s.s.. Par ailleurs, notre étude permet de mettre en évidence quelques relations inter-familiales qui sont retrouvées dans toutes les analyses, telles que Palaemonidae/Alpheidae d'une part et Oplophoridae/Alvinocarididae d'autre part.

IV.4 Conclusions et perspectives

Les marqueurs développés dans cette étude ont permis de résoudre une partie des relations interfamiliales au sein des huit familles de Caridea échantillonnées. Bien que l'échantillonnage taxonomique soit plus limité que dans l'étude d'Aznar-Cormano *et al.* (2015), la résolution obtenue sur certains nœuds est encourageante, quelle que soit l'analyse (voir arbre consensuel, figure IV.17).

De nombreuses corrections ont cependant dû être apportées pour vérifier l'orthologie des marqueurs, notamment les plus divergents, et il est possible que les marqueurs utilisés aient des divergences telles qu'ils apportent plus de bruit (saturation du signal) que d'information phylogénétique. Afin de fixer les relations profondes et savoir si les différences de topologie s'expliquent par une saturation du signal phylogénétique, il est nécessaire d'augmenter à la fois le nombre de taxons et de lignées échantillonnées, et le nombre de marqueurs utilisés.

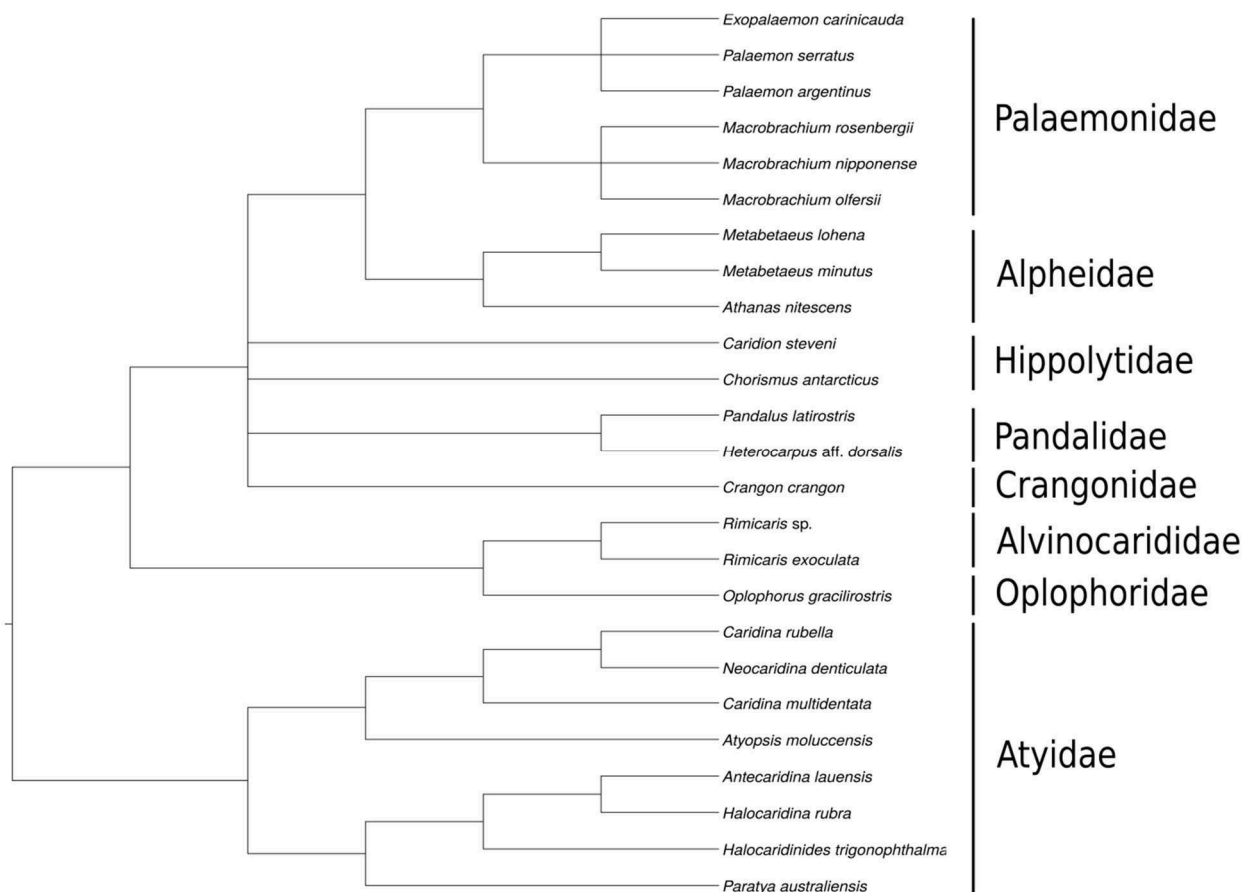


Figure IV.17 Topologie récapitulative des relations au sein des 26 espèces de Caridea présentes dans notre échantillonnage. Les polytomies représentent les nœuds qui ne sont pas entièrement soutenus dans l'ensemble des analyses effectuées avec 26 taxons.

Encadré 7 : BUSCO

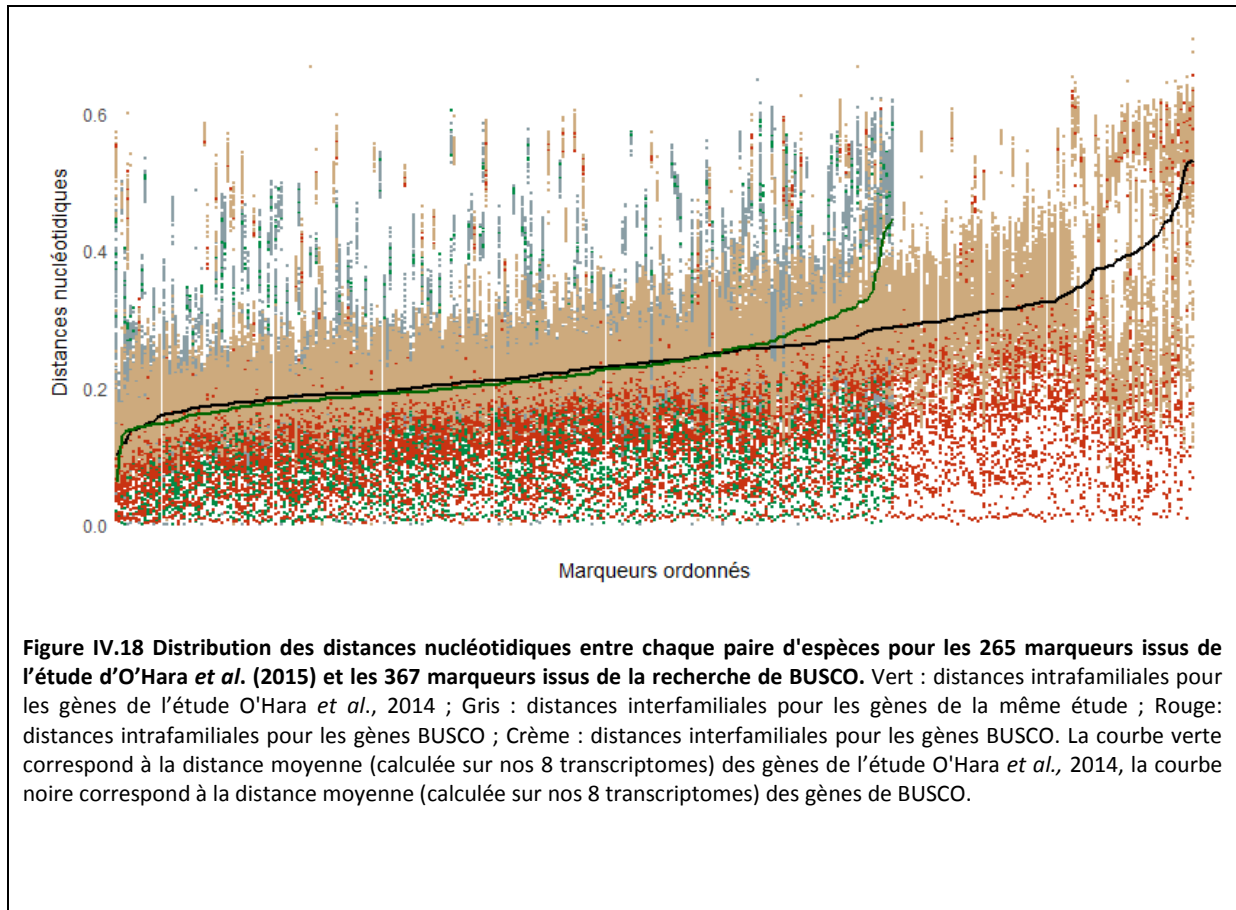
Le transcriptome est complet si tous les ARNm de l'organisme ont été assemblés. Dans le cas d'un assemblage produit à l'aide d'un génome de référence, la complétude de l'assemblage produit est estimée par la taille de l'assemblage relativement à la taille des régions codant pour des protéines dans le génome de référence. Dans le cas d'un assemblage *de novo*, la complétude ne peut être évaluée que de manière indirecte, par comparaison avec des catalogues de protéines identifiées pour des organismes proches. Dans cette optique, Simão *et al.* (2015) ont développé l'outil BUSCO (Benchmarking Universal Single-Copy Orthologs), qui fournit des catalogues de marqueurs orthologues (*i.e.* marqueurs présents dans des espèces actuelles par héritage de leur dernier ancêtre commun), sélectionnés à partir de la base de données OrthoDB v9.1 (Zdobnov *et al.*, 2017) pour plusieurs grands clades à différentes profondeurs phylogénétiques. Les marqueurs retenus dans ce catalogue sont présents en simple copie pour au minimum 90% des espèces du corpus considéré. Pour évaluer la qualité de nos assemblages, nous cherchons pour chacun d'eux les 1 066 protéines du catalogue fourni par BUSCO v3.0.1 pour le taxon Arthropoda.

Pour chacune des protéines de ces catalogues, nous classons ces protéines comme :

- Complètes, c'est-à-dire que la protéine a été retrouvée dans l'assemblage avec une longueur supérieure ou égale à 95% de la taille de la protéine dans le catalogue ; elles sont alors soit en simple copie (sc), soit dupliquées (d) si plusieurs contigs correspondent à une même protéine.
- Fragmentées (F), si dans nos assemblages elles sont plus courtes que 95% de la taille de la protéine dans le catalogue.
- Manquantes (M) quand elles sont absentes de l'assemblage.

Ainsi, plus le nombre de protéines retrouvées est grand plus le transcriptome assemblé est considéré comme complet. Les protéines retrouvées dupliquées peuvent résulter de l'hétérozygotie (*i.e.* assemblage contenant plusieurs haplotypes, Simão *et al.*, 2015) ou d'une inadéquation du catalogue vis-à-vis des Caridea pour une partie de ces protéines (*i.e.* gènes qui ne seraient pas en simple copie chez les Caridea). Les résultats de BUSCO permettent d'évaluer la complétude de chaque transcriptome assemblé relativement au catalogue disponible mais aussi d'évaluer la complétude relative entre les transcriptomes que nous avons assemblés. Les gènes du corpus BUSCO sont recherchés dans les séquences codantes obtenues après le traitement post-assemblage pour chacune des espèces utilisées lors de cette étude.

Nous avons recherché les gènes du corpus Arthropode dans les transcriptomes utilisés dans notre étude, et la plupart d'entre eux sont retrouvés, dont 367 partagés entre les 8 transcriptomes produits dans notre équipe. Les distances moyennes obtenues (calculées sur les huit transcriptomes produits dans notre équipe) sont similaires à celles obtenues en cherchant les marqueurs de l'oursin (figure IV.18), et les 13 premiers marqueurs ont moins de 15% de divergence nucléotidique en moyenne (contre 20 pour les marqueurs identifiés à partir des protéines de l'oursin).



Afin d'augmenter le nombre de marqueurs conservés, il serait premièrement possible d'utiliser plus de marqueurs que les 20 les plus conservés parmi les 265 retrouvés dans cette étude. Secondement, il serait possible de rechercher par la méthode des blast réciproques les gènes partagés les plus conservés parmi les transcriptomes de Caridea disponibles, ou encore d'utiliser des jeux de marqueurs prédéfinis et très conservés (par exemple jeu de données BUSCO). Les gènes des corpus BUSCO (voir encadré 7) sont utilisés classiquement pour évaluer la complétude d'un assemblage (Waterhouse *et al.*, 2017) mais peuvent également servir pour résoudre des phylogénies profondes (Ioannidis *et al.*, 2017 ; Shen *et al.*, 2016).

Par ailleurs, le but ultime est de construire une phylogénie avec un échantillonnage taxonomique comparable à celui de l'étude d'Aznar-Cormano *et al.* (2015), afin de placer la famille des Alvinocarididae parmi les 28 familles de leur étude et non plus seulement par rapport aux sept autres familles proposées dans la présente étude. Pour élargir l'échantillonnage taxonomique, une approche d'amplification par PCR sur ADN génomique (compatible avec la conservation en alcool) a été tentée, cependant aucun résultat concluant n'a été obtenu. Ceci pourrait s'expliquer par la

présence d'introns. En l'absence de génome de référence nous n'avons pas d'information sur les limites introns/exons ni sur la longueur des introns. Une seconde approche pour augmenter l'échantillonnage taxonomique est l'approche de capture d'exon (Hugall *et al.*, 2015). En effet, en l'absence d'information sur les limites introns-exons, il est possible de récupérer les séquences en procédant par capture par hybridation à l'aide de sondes de capture spécifiques sur de l'ADN génomique fragmenté. Les séquences issues des transcriptomes permettent de dessiner plusieurs sondes chevauchantes pour chaque marqueur. Ainsi, même si les sondes correspondant aux limites intron/exons ne pourront pas s'hybrider sur l'ADN, les sondes voisines s'hybrideront en amont et en aval de l'intron, permettant de capturer la séquence de l'exon. L'échantillonnage transcriptomique de notre étude nous permettrait de confectionner des sondes pour les 37 marqueurs, en couvrant pratiquement la totalité des taxons, avec *a priori* une divergence maximale de 20% entre la séquence de la sonde et la séquence ciblée. Par ailleurs, apporter des transcriptomes supplémentaires de lignées non présentes dans notre étude permettrait d'être à la fois plus confiant dans la couverture taxonomique ciblée par les sondes d'hybridation, et également de dessiner plus de sondes, qui seraient alors moins divergentes des séquences ciblées.

Chapitre V

Conclusions et
perspectives générales
de la thèse

V. CONCLUSIONS ET PERSPECTIVES GENERALES DE LA THESE

Pour obtenir une phylogénie qui fournisse des informations robustes sur l'évolution des taxons, nous cherchons idéalement à optimiser à la fois l'échantillonnage taxonomique et le nombre de marqueurs moléculaires permettant la résolution des relations phylogénétiques à la profondeur taxonomique souhaitée. Cependant, le nombre, le choix des marqueurs et la couverture taxonomique optimale ne sont pas standardisés et doivent évidemment être adaptés à chaque problématique et pour chaque groupe taxonomique. De plus, l'optimisation conjointe de ces deux paramètres reste la plupart du temps très compliquée à mettre en œuvre. De ce constat émerge la nécessité de comprendre les impacts respectifs de la couverture taxonomique et des marqueurs considérés pour dans la résolution d'une phylogénie. Cette nécessité est d'autant plus vraie que (i) l'obtention d'un large échantillonnage et (ii) l'optimisation de marqueurs sont grandement mises à l'épreuve lorsque l'on veut établir une phylogénie sur des organismes non modèles, dont certains représentants vivent dans des milieux difficiles d'accès.

Classiquement utilisée dans les reconstructions phylogénétiques, l'approche d'amplification par PCR de plusieurs gènes suivie d'un séquençage Sanger permet d'obtenir plusieurs marqueurs potentiellement indépendants, et peut typiquement être appliquée sur un large échantillonnage taxonomique. Cependant, si ces marqueurs peuvent suffire pour résoudre des phylogénies récentes, ils ne peuvent bien souvent pas résoudre des phylogénies profondes. Pour ce faire, il est alors nécessaire d'augmenter le signal phylogénétique, notamment en optimisant les marqueurs moléculaires utilisés. Depuis l'avènement des nouvelles technologies de séquençage, la maximisation du nombre de marqueurs moléculaires a été grandement facilitée et offre la possibilité de définir de nouveaux marqueurs fournissant un signal phylogénétique adapté à chaque problématique. Durant cette thèse, 3 approches ont été abordées (amplification de plusieurs gènes par PCR, utilisation de génomes mitochondriaux et analyse de transcriptomes), pour amener à des avancées dans plusieurs groupes de taxons hydrothermaux : les vers de la famille Polynoidae, les crabes de la famille Bythograeidae et les crevettes de la famille Alvinocarididae. Les résultats obtenus dans ces différents phylums illustrent bien les capacités et limites des approches pour résoudre ces problèmes couramment rencontrés en phylogénie.

La première partie de ce projet de thèse a étudié les relations phylogénétiques d'une famille très diverse d'annélides, les Polynoidae, et a notamment permis d'aborder la problématique de

l'optimisation du nombre de taxons dans les études de phylogénie. Sans information préalable sur le contexte phylogénétique, nous avons cherché à établir un échantillonnage incluant le plus grand nombre de sous-familles, ainsi que des espèces provenant des contextes écologiques variés. Les données disponibles sur les bases de données étaient en effet centrées sur les milieux les plus accessibles ou d'intérêt particulier (biais classique des bases de données). La reconstruction phylogénétique que nous proposons, basée sur cinq fragments de gènes amplifiés par PCR, nous a permis de produire une phylogénie avec un échantillonnage centré exclusivement sur cette famille, alors que les études antérieures étaient centrées sur des familles proches ou sur les relations interfamiliales. Cette phylogénie nous permet de montrer avec confiance que les Polynoidae ont colonisé les sources hydrothermales au moins à deux reprises au cours de leur évolution. De plus, et à l'inverse de l'hypothèse d'une transition évolutive depuis les substrats organiques vers les suintements froids puis les sources hydrothermales, notre phylogénie suggère que les sources hydrothermales ont été colonisées en premier parmi ces environnements à production primaire chimiosynthétique. En effet, les espèces retrouvées au niveau des sources hydrothermales sont entremêlées avec les espèces provenant d'environnements non hydrothermaux à production chimiosynthétique et ces dernières ont des positions dérivées dans l'arbre phylogénétique. Cependant cette approche multigène ne nous a pas permis de résoudre toutes les relations phylogénétiques profondes, en particulier entre les sous-familles. Cela souligne l'importance d'augmenter le nombre de marqueurs et de cibler ceux qui permettront de résoudre les relations à cette profondeur. Une telle limite de résolution a également été rencontrée chez les Caridea et les Brachyura (Aznar-Cormano *et al.*, 2015 ; Tsang *et al.*, 2014).

Dans le but de tester les effets de la maximisation et de l'optimisation des marqueurs sur l'amélioration du signal phylogénétique, nous avons étudié le cas des crevettes de l'infra-ordre Caridea. Grâce à une approche transcriptomique, de nouveaux marqueurs ont été déterminés, et nous ont effectivement permis de préciser les relations entre huit familles de Caridea en obtenant des relations phylogénétiques profondes soutenues. Malgré le potentiel de cette approche (chapitre III), l'étude des relations phylogénétiques basée sur une approche d'assemblage de transcriptomes est difficilement généralisable à un grand nombre de taxons, notamment en raison de verrous techniques et économiques.

La troisième approche abordée dans cette thèse (chapitre II), évalue la capacité des génomes mitochondriaux complets à résoudre des relations phylogénétiques profondes à l'échelle de l'infra-ordre des crabes Brachyura et replacer la famille hydrothermale Bythograeidae dans ce contexte.

Cette approche se base sur le séquençage de génomes mitochondriaux complets (permettant d'obtenir un grand nombre de marqueurs avec des taux d'évolution différents) et peut facilement s'étendre sur une couverture taxonomique large. Cette approche est celle qui permet le plus facilement des trois, d'allier à la fois un large échantillonnage taxonomique et un jeu de marqueurs optimisé, permettant de résoudre les relations aux profondeurs visées. La diversité des taux d'évolution des gènes de ce marqueur permet de résoudre à la fois les relations dans les familles, entre les familles et les relations profondes entre les sous-sections de cet infra-ordre. Cependant, les marqueurs mitochondriaux ne correspondent qu'à un seul locus et partagent une même histoire évolutive, ce qui peut apporter un biais important lors de la reconstruction phylogénétique. Afin d'utiliser à la fois des séquences mitochondriales et des séquences nucléaires, il serait, dans le cas des Brachyura, intéressant de séquencer les génomes mitochondriaux des espèces (si possible sur les mêmes spécimens) de l'étude de Tsang *et al.* (2014). Ceci permettrait certainement de replacer la famille Bythograeidae parmi les Brachyura de manière plus précise et satisfaisante. De plus, ceci pourrait limiter le biais potentiel de la dépendance des marqueurs mitochondriaux, tout en bénéficiant d'un échantillonnage taxonomique déjà défini et homogène au sein des Brachyura. Il est également nécessaire d'ajouter à cette analyse les espèces des genres *Bythograea* et *Cyanograea*, pour le moment jamais intégrés dans les études phylogénétiques publiées et couvrant un échantillonnage taxonomique large, afin de tester la monophylie de la famille Bythograeidae.

Des études récentes proposent de combiner à la fois le nombre de marqueurs séquencés par des approches de NGS et un échantillonnage taxonomique important. En effet, il est possible d'utiliser, dans un premier temps, une analyse transcriptomique (comme il a été fait dans le cas des Caridea), afin de définir des nouveaux marqueurs, qui seront dans un second temps recherchés dans un échantillonnage taxonomique plus large. Plusieurs études ont résolu des phylogénies en combinant l'approche transcriptomique (pour définir des marqueurs) et une approche d'enrichissement par capture (pour élargir l'échantillonnage taxonomique). Parmi elles, Bank *et al.* (2017) ont séquencé 913 gènes en simple copie pour 25 espèces, grâce à un enrichissement par capture d'exons et ont résolu les relations phylogénétiques dans une famille d'hyménoptères (divergence d'environ 200 Ma ; Rainford *et al.*, 2014). De la même façon, les études d'O'Hara *et al.* (2014) et Hugall *et al.* (2015) ont permis de séquencer, pour 380 espèces, 425 marqueurs issus de l'étude de 52 transcriptomes d'Ophiures, permettant de résoudre des relations phylogénétiques profondes entre les différents ordres de la classe des Ophiuroidea (divergence d'environ 240 Ma, O'Hara *et al.*, 2014). Cette approche de capture de marqueurs ciblés ne nécessite pas de

conservation des échantillons particulière et semble bien adaptée au matériel issu des collections de muséums d'histoire naturelle (Wood *et al.*, 2018), sauf pour les échantillons qui ont été conservés dans le formol. Au cours de cette thèse, de nouveaux marqueurs (à la fois des marqueurs conservés et divergents au sein des Caridea) ont été définis. Une étude préliminaire suggère que, dans le cas des marqueurs les plus conservés, une seule séquence consensus de sondes de capture suffirait pour cibler les marqueurs parmi la diversité des espèces du clade I d'Aznar-Cormano *et al.* (2015), le plus diversifié des Caridea.

En séquençant les transcriptomes d'espèces réparties parmi les différentes lignées mises en évidence dans notre étude sur la phylogénie des vers Polynoidae, la définition de nouveaux marqueurs pourrait être effectuée par une approche similaire à celle utilisée chez les Caridea. Ceci ouvrirait de nouvelles perspectives sur la compréhension des relations phylogénétiques et sur l'histoire évolutive de ces espèces hydrothermales. Cette approche a d'ailleurs déjà débuté au sein de l'équipe ABICE, avec une quinzaine de transcriptomes séquencés et assemblés. Une recherche de similarité des séquences par blast réciproques parmi les transcriptomes deux à deux permettra d'identifier des marqueurs partagés à l'échelle de cette famille d'annélides. L'échantillonnage taxonomique couvrant la diversité du groupe d'intérêt de manière homogène mis en place lors de cette étude pourra servir afin de capturer les marqueurs identifiés à l'aide de sondes de capture d'exons.

Obtenir des phylogénies résolues pour un grand nombre d'espèces couvrant une diversité taxonomique représentative de chacun de ces trois taxons, permettra de mieux comprendre les événements de colonisation des sources hydrothermales. Par exemple, dans le cas des vers Polynoidae, nous avons mentionné la possibilité d'une radiation rapide une fois l'environnement hydrothermal colonisé. Cette hypothèse est appuyée par des branches internes relativement courtes dans la lignée comprenant les clades 2 et 3, correspondant à des espèces qui habitent les environnements permettant une production chimiosynthétique locale. Davis *et al.* (2018) ont montré chez les Caridea que la transition vers l'environnement d'eau douce était associée à une diversification nette plus importante que dans le milieu marin. Cette diversification nette est la résultante d'un taux de spéciation plus important dans ces milieux, et, bien que dans une moindre mesure, d'un taux d'extinction également plus fort en eau douce qu'en milieu marin. De par notre résultat suggérant un même scénario chez les Polynoidae lors de leur colonisation des sources

hydrothermales, il serait intéressant d'appliquer l'approche de Davis *et al.* (2018) pour tester cette hypothèse quant à leur dynamique de diversification dans les environnements hydrothermaux.

D'autre part, les relations phylogénétiques restent encore non résolues pour d'autres taxons originaux provenant des sources hydrothermales. La plupart des 14 familles qualifiées d'endémiques des sources hydrothermales identifiées par Wolf, (2005) ont aujourd'hui fait l'objet d'études phylogénétiques et doivent inclure des espèces récoltées sur d'autres environnements pour former des groupes monophylétiques. Cependant, la monophylie de la famille des Alvinellidae, annélides polychètes exclusivement retrouvés à proximité des sources hydrothermales du Pacifique, n'a pas encore été testée. Une équipe américaine (G. Rouse, Scipps) travaille actuellement à partir de données transcriptomiques, pour fournir une phylogénie des Terebellomorpha, sous-ordre dans lequel se situe cette famille.

Par ailleurs, une étude phylogénétique menée sur les cirripèdes du super-ordre Thoracica suggère que ce groupe a connu au moins deux évènements de colonisation du milieu hydrothermal (Herrera *et al.*, 2015). Cependant, parmi les deux clades hydrothermaux de cirripèdes identifiés, le plus diversifié taxonomiquement comprend des espèces morphologiquement placées dans deux ordres différents. Ce résultat suggère que les caractères morphologiques utilisés pour l'identification des ordres (*e.g.* le nombre de plaques du capitulum) résultent probablement de convergences évolutives (ce qui avait déjà été discuté auparavant ; Pérez-Losada *et al.*, 2014), et appuie la nécessité de construire une phylogénie à partir de marqueurs moléculaires. L'étude d'Herrera *et al.* (2015) contient des représentants pour chacune des familles hydrothermales de cirripèdes. Toutefois, les genres *Imbricaverruca* (Neoverrucidae), *Archeolepas* et *Neolepas* (Eolepadidae) ne sont pas présents dans leur étude. Par ailleurs, la famille Chionelasmatidae comporte deux genres : *Eochionelasmus* retrouvé au niveau des sources hydrothermales et *Chionelasmus*, non hydrothermal et dont les espèces sont récoltées sur des coraux profonds (Buckeridge, 1998). De plus, la super famille Chionelasmatoidea dans laquelle sont placés les Chionelasmatidae comprend également la famille monogénérique non hydrothermale Waikalasmatidae. Les genres *Waikalasma* et *Chionelasmus* ne sont cependant pas présents dans l'étude de Herrera *et al.* (2015). Une étude récente (Chan *et al.*, 2017) suggère un lien de parenté entre ces deux genres, mais l'échantillonnage taxonomique de cette dernière étude n'inclut pas les taxons hydrothermaux cités précédemment. Ainsi, il serait intéressant de reconstruire une phylogénie qui contienne, au minimum, l'ensemble des genres utilisés dans ces deux études, afin de tester (i) si les genres hydrothermaux non représentés

dans l'étude d'Herrera *et al.* (2015) se regroupent parmi l'une des deux lignées hydrothermales et (ii) si les regroupements taxonomiques familiaux et super-familiaux effectués sur la base des caractères morphologiques correspondent à des regroupements naturels (monophylétiques), en particulier dans le cas de la super famille Chionelasmatoidea. En effet, l'approche menée par Herrera *et al.* (2015) (approche multigène suivie d'un séquençage Sanger) ne leur permet pas de résoudre les nœuds profonds. Ainsi, les relations entre les deux clades hydrothermaux qu'ils ont mis en évidence restent inconnues. Les temps de divergences inférés durant leur étude sont longs : les auteurs ont estimé que le clade comprenant les deux lignées hydrothermales date d'environ 350 Ma, et il serait judicieux d'utiliser des marqueurs permettant de résoudre des relations anciennes. Aussi, étant donné que peu d'informations sont disponibles sur les tailles de génomes de cirripèdes, l'approche de séquençage de transcriptomes (telle que celle que nous avons utilisée pour les Caridea), parmi un sous ensemble d'espèces de différentes lignées du super-ordre de Thoracica, réparties dans l'arbre pourrait être appliquée dans le cas des cirripèdes afin de définir de nouveaux marqueurs phylogénétiques. Dans un second temps, ces marqueurs pourront être cherchés chez d'autres espèces de ce super-ordre, afin d'élargir l'échantillonnage.

Pour conclure, à travers l'étude de trois taxons hydrothermaux, ce travail de thèse a permis de souligner les capacités et les limitations de différentes approches moléculaires (amplification de plusieurs gènes par PCR, utilisation de génomes mitochondriaux et analyse de transcriptomes) en phylogénie. Dans le cas particulier de notre étude, des hypothèses ont été émises quant aux histoires évolutives de taxons hydrothermaux, en les replaçant dans un contexte global de colonisation des milieux extrêmes. De manière générale, nous avons également soulevé certains points de réflexion qui établissent les bases de futurs développements et applications de telles approches, qui permettront de nombreuses avancées en phylogénie moléculaire dans un avenir proche.

Bibliographie

VI. BIBLIOGRAPHIE

- Alfaro-Lucas, J. M., Shimabukuro, M., Ferreira, G. D., Kitazato, H., Fujiwara, Y., & Sumida, P. Y. G. (2017). Bone-eating *Osedax* worms (Annelida: Siboglinidae) regulate biodiversity of deep-sea whale-fall communities. *Deep Sea Research Part II: Topical Studies in Oceanography*, (146), pp. 4-12. <https://doi.org/10.1016/j.dsr2.2017.04.011>
- Alvarez, F., Iliffe, T. M., & Villalobos, J. L. (2005). New species of the genus *Typhlatya* (Decapoda: Atyidae) from anchialine caves in Mexico, the Bahamas, and Honduras. *Journal of Crustacean Biology*, 25(1), pp. 81-94.
- Alvaro, M. C., Barnich, R., Modica, M. V., Barco, A., Oliverio, M., & Schiaparelli, S. (2014, August). Phylogenetic relationships and phylogeography in Antarctic symbiotic polyxenous polychaetes. In *XXXIII SCAR Biennial Meetings and Open Science Conference, Auckland*.
- Amon, D. J. (2013). *Bone-eating worms and wood-eating bivalves: characterising the ecology of deep-sea organic falls from multiple ocean basins* (PhD Thesis). University of Southampton.
- Ayhong, S. T., Lowy, J. K., Alonso, M., Bamber, R. N., Boxshall, G. A., Castro, P., ... Svavarsson, J. (2011). Crustacea. In *Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness*. pp. 165-191. Magnolia press.
- Aznar-Cormano, L., Brisset, J., Chan, T.-Y., Corbari, L., Puillandre, N., Utge, J., ... Samadi, S. (2015). An improved taxonomic sampling is a necessary but not sufficient condition for resolving inter-families relationships in Caridean decapods. *Genetica*, 143(2), pp. 195-205. <https://doi.org/10.1007/s10709-014-9807-0>
- Baeza, J. A. (2010). Observations on the sexual system and the natural history of the semi-terrestrial shrimp *Merguia rhizophorae* (Rathbun, 1900): Protandry in a semi-terrestrial shrimp. *Invertebrate Biology*, 129(3), pp. 266-276. <https://doi.org/10.1111/j.1744-7410.2010.00200.x>
- Bain, P. A., Gregg, A. L., & Kumar, A. (2016). *De novo* assembly and analysis of changes in the protein-coding transcriptome of the freshwater shrimp *Paratya australiensis* (Decapoda: Atyidae) in response to acid sulfate drainage water. *BMC Genomics*, 17(1). <https://doi.org/10.1186/s12864-016-3208-y>
- Ballenghien, M., Faivre, N., & Galtier, N. (2017). Patterns of cross-contamination in a multispecies population genomic project: detection, quantification, impact, and solutions. *BMC Biology*, 15(1). <https://doi.org/10.1186/s12915-017-0366-6>
- Bank, S., Sann, M., Mayer, C., Meusemann, K., Donath, A., Podsiadlowski, L., ... Niehuis, O. (2017). Transcriptome and target DNA enrichment sequence data provide new insights into the phylogeny of vespidae wasps (Hymenoptera: Aculeata: Vespidae). *Molecular Phylogenetics and Evolution*, (116), pp. 213-226. <https://doi.org/10.1016/j.ympev.2017.08.020>
- Barriel, V., & Tassy, P. (1998). Rooting with multiple outgroups: consensus versus parsimony. *Cladistics*, 14(2), pp. 193-200.
- Bentley, M. G., & Pacey, A. A. (1992). Physiological and environmental control of reproduction in polychaetes. *Oceanography and Marine Biology: An Annual Review*.
- Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritsch, G., ... Stadler, P. F. (2013). MITOS: Improved *de novo* metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, 69(2), pp. 313-319. <https://doi.org/10.1016/j.ympev.2012.08.023>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), pp. 2114-2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bonnivard, E., Catrice, O., Ravaux, J., Brown, S. C., & Higuier, D. (2009). Survey of genome size in 28 hydrothermal vent species covering 10 families. *Genome*, 52(6), pp. 524-536. <https://doi.org/10.1139/G09-027>
- Borner, J., Rehm, P., Schill, R. O., Ebersberger, I., & Burmester, T. (2014). A transcriptome approach to ecdysozoan phylogeny. *Molecular Phylogenetics and Evolution*, (80), pp. 79-87. <https://doi.org/10.1016/j.ympev.2014.08.001>

- Braby, C. E., Rouse, G. W., Johnson, S. B., Jones, W. J., & Vrijenhoek, R. C. (2007). Bathymetric and temporal variation among *Osedax* boneworms and associated megafauna on whale-falls in Monterey Bay, California. *Deep Sea Research Part I: Oceanographic Research Papers*, 54(10), pp. 1773–1791. <https://doi.org/10.1016/j.dsr.2007.05.014>
- Bracken, H. D., Toon, A., Felder, D. L., Martin, J. W., Finley, M., Rasmussen, J., ... & Crandall, K. A. (2009a). The decapod tree of life: compiling the data and moving toward a consensus of decapod evolution. *Arthropod Systematics & Phylogeny*, 67(1), pp. 99–116.
- Bracken, H. D., De Grave, S., & Felder, D. L. (2009b). Phylogeny of the infraorder Caridea based on mitochondrial and nuclear genes (Crustacea: Decapoda). *Decapod crustacean phylogenetics*, 28, pp. 1–305.
- Bracken, H. D., De Grave, S., Toon, A., Felder, D. L., & Crandall, K. A. (2010). Phylogenetic position, systematic status, and divergence time of the Procarididea (Crustacea: Decapoda). *Zoologica Scripta*, 39(2), pp. 198–212. <https://doi.org/10.1111/j.1463-6409.2009.00410.x>
- Brasier, M. J., Wiklund, H., Neal, L., Jeffreys, R., Linse, K., Ruhl, H., & Glover, A. G. (2016). DNA barcoding uncovers cryptic diversity in 50% of deep-sea Antarctic polychaetes. *Royal Society Open Science*, 3(11). <https://doi.org/10.1098/rsos.160432>
- Britayev, T. A. (1991). Life cycle of the symbiotic scaleworm *Arctonoe vittata* (Polychaeta: Polynoidae). In *Systematics, Biology and Morphology of World Polychaeta. Proceedings of the Second International Polychaeta Conference. Ophelia Supplement* (Vol. 5, pp. 305–312).
- Britayev, T. A., Gil, J., Altuna, A., Calvo, M., & Martin, D. (2014). New symbiotic associations involving polynoids (Polychaeta, Polynoidae) from Atlantic waters, with redescription of *Parahololepidella greeffi* (Augener, 1918) and *Gorgoniapolynoe caeciliae* (Fauvel, 1913). *Memoirs of the National Museum Victoria* 71, pp. 27–43
- Brösing, A., Richter, S., & Scholtz, G. (2007). Phylogenetic analysis of the Brachyura (Crustacea, Decapoda) based on characters of the foregut with establishment of a new taxon. *Journal of Zoological Systematics and Evolutionary Research*, 45(1), pp.20–32. <https://doi.org/10.1111/j.1439-0469.2006.00367.x>
- Buckeridge, J. S. (1998). A new coral inhabiting barnacle of the genus *Chionelasmus* (Cirripedia, Balanomorph) from New Caledonia, South-West Pacific. *ZOOSYSTEMA-PARIS*-, 20, pp. 167–176.
- Calabrese, F. M., Balacco, D. L., Preste, R., Diroma, M. A., Forino, R., Ventura, M., & Attimonelli, M. (2017). NumtS colonization in mammalian genomes. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-16750-2>
- Carr, C. M., Hardy, S. M., Brown, T. M., Macdonald, T. A., & Hebert, P. D. N. (2011). A Tri-Oceanic Perspective: DNA Barcoding Reveals Geographic Structure and Cryptic Diversity in Canadian Polychaetes. *PLoS ONE*, 6(7), e22232. <https://doi.org/10.1371/journal.pone.0022232>
- Cavanaugh, C. M., Gardiner, S. L., Jones, M. L., Jannasch, H. W., & Waterbury, J. B. (1981). Prokaryotic cells in the hydrothermal vent tube worm *Riftia pachyptila* Jones: possible chemoautotrophic symbionts. *Science*, pp. 340–342.
- Colgan, D. J., Ponder, W. F., Beacham, E., & Macaranas, J. M. (2003). Molecular phylogenetic studies of Gastropoda based on six gene segments representing coding or non-coding and mitochondrial or nuclear DNA. *Molluscan Research*, 23, pp. 123–148.
- Chace Jr, F. A., & Brown, D. E. (1978). A new polychelate shrimp from the Great Barrier Reef of Australia and its bearing on the family Bresiliidae (Crustacea: Decapoda: Caridea). *Proceedings of the Biological Society of Washington*, 91(3), pp. 756–766.
- Chace, F. A. (1992). On the classification of the Caridea (Decapoda). *Crustaceana*, 63(1), pp. 70–80.
- Chan, B. K. K., Corbari, L., Rodriguez Moreno, P. A., & Tsang, L. M. (2017). Molecular phylogeny of the lower acorn barnacle families (Bathylasmatidae, Chionelasmatidae, Pachylasmatidae and Waikalasmatidae) (Cirripedia: Balanomorph) with evidence for revisions in family classification. *Zoological Journal of the Linnean Society*, 180(3), pp. 542–555.
- Chan, T.-Y., Lei, H. C., Li, C. P., & Chu, K. H. (2010). Phylogenetic analysis using rDNA reveals polyphyly of Oplophoridae (Decapoda: Caridea). *Invertebrate Systematics*, 24(2), p. 172. <https://doi.org/10.1071/IS09049>

- Chevaldonné, P., Jollivet, D., Feldman, R. A., Desbruyères, D., Lutz, R. A., & Vrijenhoek, R. C. (1998). Commensal scale-worms of the genus *Branchipolynoe* (Polychaeta: Polynoidae) at deep-sea hydrothermal vents and cold seeps. *Cahier de Biologie Marine*, 39(3/4), pp. 347-350.
- Christoffersen, M. L. (1986). Phylogenetic relationships between Oplophoridae, Atyidae, Pasiphaeidae, Alvinocarididae fam. n., Bresiliidae, Psalidopodidae and Disciadidae (Crustacea Caridea Atyoidea). *Boletim de Zoologia*, 10(10), pp. 273-281.
- Clarke, A. (1993). Egg size and egg composition in polar shrimps (Caridea; Decapoda). *Journal of Experimental Marine Biology and Ecology*, 168(2), pp. 189-203.
- Clarke, K., Yang, Y., Marsh, R., Xie, L., & K., Z. K. (2013). Comparative analysis of *de novo* transcriptome assembly. *Science China Life Sciences*, 56(2), pp.156-162. <https://doi.org/10.1007/s11427-013-4444-x>
- Colaço, A., Dehairs, F., & Desbruyères, D. (2002). Nutritional relations of deep-sea hydrothermal fields at the Mid-Atlantic Ridge: a stable isotope approach. *Deep Sea Research Part I: Oceanographic Research Papers*, 49(2), pp. 395-412.
- Corliss, J. B., Dymond, J., Gordon, L. I., Edmond, J. M., von Herzen, R. P., Ballard, R. D., ... & van Andel, T. H. (1979). Submarine thermal springs on the galapagos rift. *Science*, 203(4385), pp. 1073-1083.
- von Cosel, R., & Olu, K. (1998). Gigantism in Mytilidae. A new *Bathymodiolus* from cold seep areas on the Barbados accretionary prism. *Comptes Rendus de l'Académie des Sciences-Series III-Sciences de la Vie*, 321(8), pp. 655-663.
- D'Acoz, C. D., & De Grave, S. (2018). A new genus and species of large-bodied caridean shrimp from the Crozet Islands, Southern Ocean (Crustacea, Decapoda, Lipkiidae) with a checklist of Antarctic and sub-Antarctic shrimps. *Zootaxa*, 4392(2), p. 201. <https://doi.org/10.11646/zootaxa.4392.2.1>
- Dahms, H.-U., Tseng, L.-C., & Hwang, J.-S. (2017). Are vent crab behavioral preferences adaptations for habitat choice? *PloS one*, 12(9), e0182649.
- Daly, J. M. (1972). The maturation and breeding biology of *Harmothoe imbricata* (Polychaeta: Polynoidae). *Marine Biology*, 12(1), pp. 53-66.
- De Grave, S., & Moosa, M. K. (2004). A new species of the enigmatic shrimp genus *Pseudosquilla* (Decapoda: Bresiliidae) from Sulawesi (Indonesia), with the designation of a new family Pseudosquillidae. *Crustacean Research*, 33, pp.1-9.
- De Grave, S., Cai, Y., & Anker, A. (2008). Global diversity of shrimps (Crustacea: Decapoda: Caridea) in freshwater. *Hydrobiologia*, 595(1), pp. 287-293. <https://doi.org/10.1007/s10750-007-9024-2>
- De Grave, N., Pentcheff, D., & Ahyong, S. T. (2009). A classification of living and fossil genera of decapod crustaceans. *Raffles Bulletin of Zoology*, pp. 1-109.
- De Grave, S., Chan, T.-Y. Y., & Chu, K. H.. (2010). On the Systematic Position of *Galatheacaris abyssalis* (Decapoda: Galatheacaridoidea). *Journal of Crustacean Biology*, 30(3), pp. 521-527. <https://doi.org/10.1651/10-3278.1>
- De Grave, S., & Franssen, C. H. J. M. (2011). Carideorum catalogus: the recent species of the dendrobranchiate, stenopodidean, procarididean and caridean shrimps (Crustacea: Decapoda). *Zoologische Mededelingen*, (9).
- De Grave, S., & Goulding, L. Y. D. (2011). Comparative morphology of the pereopod 1 carpo-propodal (P1-CP) antennal flagellar grooming brush in caridean shrimps (Crustacea, Decapoda). *Zoologischer Anzeiger - A Journal of Comparative Zoology*, 250(4), pp. 280-301. <https://doi.org/10.1016/j.jcz.2011.08.003>
- De Grave, S., Li, C. P., Tsang, L. M., Chu, K. H., & Chan, T.-Y. (2014). Unweaving hippolytoid systematics (Crustacea, Decapoda, Hippolytidae): resurrection of several families. *Zoologica Scripta*, 43(5), pp. 496-507. <https://doi.org/10.1111/zsc.12067>
- Desbruyères, D., Alayse-Danet, A. M., & Ohta, S. (1994). Deep-sea hydrothermal communities in Southwestern Pacific back-arc basins (the North Fiji and Lau Basins): Composition, microdistribution and food web. *Marine Geology*, 116(1-2), pp. 227-242.
- Desbruyères, D., Segonzac, M., & Bright, M. (2006). *Handbook of Deep-Sea Hydrothermal Vent Fauna* (Second compl. rev. ed). Linz: Art & Publishing.

- Ditter, R. E., Goebel, A. M., & Erdman, R. B. (2015). First Record of the anchialine shrimp *Parhippolyte sterreri* (Decapoda, Barbouriidae) from San Salvador Island, Bahamas, with observations on *Barbouria cubensis*. *Marine Biodiversity Records*, 8.
- Dos Reis, M., Thawornwattana, Y., Angelis, K., Telford, M. J., Donoghue, P. C. J., & Yang, Z. (2015). Uncertainty in the Timing of Origin of Animals and the Limits of Precision in Molecular Timescales. *Current Biology*, 25(22), pp. 2939–2950. <https://doi.org/10.1016/j.cub.2015.09.066>
- Doyle, J. J., & Doyle, J. L. (1987). A rapid procedure for DNA purification from small quantities of fresh leaf tissue. *Phytochemical Bulletin*, 19, pp. 11–15.
- Dubilier, N., Bergin, C., & Lott, C. (2008). Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nature Reviews Microbiology*, 6(10), pp. 725–740. <https://doi.org/10.1038/nrmicro1992>
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), pp. 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Fauchald, K. (1977). The polychaete worms. Definitions and keys to the orders, families and genera. *Natural History Museum of Los Angeles County, Science Series*, 28, pp. 1–190.
- Felbeck, H. (1981). Chemoautotrophic potential of the hydrothermal vent tube worm, *Riftia pachyptila* Jones (Vestimentifera). *Science*, 213(4505), pp. 336–338.
- Folmer, O., Blak, M., Hoej, W., Vrijenhoek, R.. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3(5), 294–299.
- Forbes, E. (1844). Report on the Mollusca and Radiata of the Aegean Sea, and on their distribution, considered as bearing on geology, *Report of the British Association for the Avancement of Science for 1843*. pp. 129–193.
- Forest, J. (1977). Un groupement injustifié: la superfamille des Bresilioidea. Remarques critiques sur les statuts des familles réunies sous ce nom (Crustacea Decapoda Caridea). *Bulletin de Museum National d'Histoire Naturelle*(3), (475), pp. 869–888.
- Forest, J., & Cals, P. (1977). Une deuxième espèce du genre *Bresilia* Caïman, *B. corsicana* sp. nov. Comparaison avec *B. atlantica* Caïman (Crustacea Decapoda Bresiliidae). *Bulletin de Museum National d'Histoire Naturelle*(3), pp. 549–565
- Fox, H. M. (1949). Haemoglobin in crustacea. *Nature*, 164(4158), p. 59.
- Fuller, C. W., Middendorf, L. R., Benner, S. A., Church, G. M., Harris, T., Huang, X., ... Vezenov, D. V. (2009). The challenges of sequencing by synthesis. *Nature Biotechnology*, 27(11), pp. 1013–1023. <https://doi.org/10.1038/nbt.1585>
- Gaudron, S. M., Lefebvre, S., Jorge, A. N., Gaill, F., & Pradillon, F. (2012). Spatial and temporal variations in food web structure from newly-opened habitat at hydrothermal vents. *Marine environmental research*, (77), pp. 129–140.
- Gaudron, S. M., Hourdez, S., & Olu, K. (2017). Aspects on gametogenesis, fertilization and embryogenesis of two deep-sea polychaetes from Eastern Atlantic cold seeps. *Deep Sea Research Part I: Oceanographic Research Papers*, (129), pp. 59–68. <https://doi.org/10.1016/j.dsr.2017.10.003>
- Gaya, E., Fernández-Brime, S., Vargas, R., Lachlan, R. F., Gueidan, C., Ramírez-Mejía, M., & Lutzoni, F. (2015). The adaptive radiation of lichen-forming Teloschistaceae is associated with sunscreens pigments and a bark-to-rock substrate shift. *Proceedings of the National Academy of Sciences*, 112(37), pp. 11600–11605. <https://doi.org/10.1073/pnas.1507072112>
- Genomic Resources Development Consortium, Havird, J. C., & Santos, S. R. (2014). Genomic Resources Notes accepted 1 June 2014 - 31 July 2014. *Molecular Ecology Resources*, 14(6), 1322–1322. <https://doi.org/10.1111/1755-0998.12315>
- Glover, A. G., Goetze, E., Dahlgren, T. G., & Smith, C. R. (2005). Morphology, reproductive biology and genetic structure of the whale-fall and hydrothermal vent specialist, *Bathypurila guaymasensis* Pettibone, 1989 (Annelida: Polynoidea). *Marine Ecology*, (26), pp. 223–234. <https://doi.org/10.1111/j.1439-0485.2005.00060.x>

- Goffredi, S. K., Orphan, V. J., Rouse, G. W., Jahnke, L., Embaye, T., Turk, K., ... Vrijenhoek, R. C. (2005). Evolutionary innovation: a bone-eating marine symbiosis. *Environmental Microbiology*, 7(9), 1369-1378. <https://doi.org/10.1111/j.1462-2920.2005.00824.x>
- Goffredi, S. K., Johnson, S., Tunnicliffe, V., Caress, D., Clague, D., Escobar, E., ... Vrijenhoek, R. (2017). Hydrothermal vent fields discovered in the southern Gulf of California clarify role of habitat in augmenting regional diversity. *Proceedings of the Royal Society B: Biological Sciences*, 284(1859), 20170817. <https://doi.org/10.1098/rspb.2017.0817>
- Gonzalez, B. C., Martínez, A., Borda, E., Iliffe, T. M., Eibye-Jacobsen, D., & Worsaae, K. (2017a). Phylogeny and systematics of Aphroditiformia. *Cladistics*, 34(3), pp. 225-259.
- Gonzalez, B. C., Petersen, H. C. B., Di Domenico, M., Martínez, A., Armenteros, M., García-Machado, E., ... & Worsaae, K. (2017b). Phylogeny and biogeography of the scaleless scale worm Pisone (Sigalionidae, Annelida). *Ecology and evolution*, 7(9), pp. 2894-2915.
- Gonzalez, B. C., Worsaae, K., Fontaneto, D., & Martínez, A. (2017c). Anophthalmia and elongation of body appendages in cave scale worms (Annelida: Aphroditiformia). *Zoologica Scripta*, 47(1), pp. 106-121.
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), pp. 644-652. <https://doi.org/10.1038/nbt.1883>
- Guggolz, T., Lins, L., Meißner, K., & Brandt, A. (2018). Biodiversity and distribution of polynoid and spionid polychaetes (Annelida) in the Vema Fracture Zone, tropical North Atlantic. *Deep Sea Research Part II: Topical Studies in Oceanography*, (148), pp. 54-63. <https://doi.org/10.1016/j.dsr2.2017.07.013>
- Guinot, D., Hurtado, L. A., & Vrijenhoek, R. (2002). New genus and species of brachyuran crab from the southern East Pacific Rise (Crustacea Decapoda Brachyura Bythograeidae). *Comptes Rendu des Biologie*, 325(11), pp. 1143-1152.
- Guinot, D., Tavares, M., & Castro, P. (2013). Significance of the sexual openings and supplementary structures on the phylogeny of brachyuran crabs (Crustacea, Decapoda, Brachyura), with new nomina for higher-ranked podotreme taxa. *Zootaxa*, 3665(1), 1. <https://doi.org/10.11646/zootaxa.3665.1.1>
- Guinot, D., & Segonzac, M. (2018). A review of the brachyuran deep-sea vent community of the western Pacific, with two new species of *Austinograea* Hessler & Martin, 1989 (Crustacea, Decapoda, Brachyura, Bythograeidae) from the Lau and North Fiji Back-Arc Basins. *Zoosystema*, 40(1), pp. 1-36. <https://doi.org/10.5252/zoosystema2018v40a5>
- Hart, C. W., & Manning, R. B. (1986). Two New Shrimps (Procarididae and Agostocarididae, New Family) from Marine Caves of the Western North Atlantic. *Journal of Crustacean Biology*, 6(3), p. 408. <https://doi.org/10.2307/1548181>
- Havird, J. C., & Santos, S. R. (2016). Developmental Transcriptomics of the Hawaiian Anchialine Shrimp *Halocaridina rubra* Holthuis, 1963 (Crustacea: Atyidae). *Integrative and Comparative Biology*, 56(6), pp. 1170-1182. <https://doi.org/10.1093/icb/icw003>
- Hebert, P. D., Cywinska, A., & Ball, S. L. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, 270(1512), pp. 313-321.
- Herrera, S., Watanabe, H., & Shank, T. M. (2015). Evolutionary and biogeographical patterns of barnacles from deep-sea hydrothermal vents. *Molecular Ecology*, 24(3), pp. 673-689. <https://doi.org/10.1111/mec.13054>
- Hillis, D. M., & Dixon, M. T. (1991). Ribosomal DNA: molecular evolution and phylogenetic inference. *The Quarterly review of biology*, 66(4), pp. 411-453.
- Hinsinger, D. D., Debruyne, R., Thomas, M., Denys, G. P. J., Mennesson, M., Utage, J., & Dettai, A. (2015). Fishing for barcodes in the Torrent: from COI to complete mitogenomes on NGS platforms. *DNA Barcodes*, 3(1). <https://doi.org/10.1515/dna-2015-0019>
- Holland, P. W. H., Hacker, A., & Williams, N. A. (1991). A molecular analysis of the phylogenetic affinities of *Saccoglossus cambrensis* Brambell & Cole (Hemichordata). *Philosophical Transactions of the Royal Society B*, 332(1264), pp. 185-189.
- Hourdez, S., & Jouin-Toulmond, C. (1998). Functional anatomy of the respiratory system of *Branchiopolynoe* species (Polychaeta, Polynoidae), commensal with *Bathymodiolus* species (Bivalvia, Mytilidae) from deep-sea hydrothermal vents. *Zoomorphology*, 118(4), pp. 225-233.

- Hourdez, S., & Lallier, F. H. (2007). Adaptations to hypoxia in hydrothermal-vent and cold-seep invertebrates. *Reviews in Environmental Science and Bio/Technology*, 6(1-3), pp. 143–159. <https://doi.org/10.1007/s11157-006-9110-3>
- Hourdez, S., Lallier, F. H., Martin-Jézéquel, V., Weber, R. E., & Toulmond, A. (1999). Characterization and functional properties of the extracellular coelomic hemoglobins from the deep-sea, hydrothermal vent scaleworm *Branchipolynoe symmytilida*. *Proteins: Structure, Function, and Bioinformatics*, 34(4), pp. 435–442.
- Hourdez, S., Lamontagne, J., Peterson, P., Weber, R., & Fisher, C. (2000). Hemoglobin from a deep-sea hydrothermal-vent copepod. *The Biological Bulletin*, 199(2), pp. 95–99. <https://doi.org/10.2307/1542868>
- Hourdez, S., & Weber, R. (2005). Molecular and functional adaptations in deep-sea hemoglobins. *Journal of Inorganic Biochemistry*, 99(1), pp. 130–141. <https://doi.org/10.1016/j.jinorgbio.2004.09.017>
- Hourdez, S., Frederick, L. A., Scherneck, A., & Fisher, C. R. (2001). Functional respiratory anatomy of a deep-sea orbiiniid polychaete from the Brine Pool NR-1 in the Gulf of Mexico. *Invertebrate Biology*, 120(1), pp. 29–40.
- Hourdez, H. (2012). Hypoxic Environments. In E. Bell (Ed.). *Life at extremes: Environments, organisms, and strategies for survival* (pp. 438–453). Cabi.
- Huang, X., & Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome research*, 9(9), pp. 868–877.
- Hugall, A. F., O'hara, T. D., Hunjan, S., Nilsen, R., & Moussalli, A. (2015). An exon-capture system for the entire class Ophiuroidea. *Molecular biology and evolution*, 33(1), pp. 281–294.
- Hutchison, C. A. (2007). DNA sequencing: bench to bedside and beyond. *Nucleic Acids Research*, 35(18), pp. 6227–6237. <https://doi.org/10.1093/nar/gkm688>
- Imajima, M. (2005). Deep-sea benthic polychaetous annelids from around Nansei Islands. *National Science Museum Monographs*, (29), pp. 37–99.
- Ioannidis, P., Simao, F. A., Waterhouse, R. M., Manni, M., Seppey, M., Robertson, H. M., ... Zdobnov, E. M. (2017). Genomic features of the damselfly *Calopteryx splendens* representing a sister clade to most insect orders. *Genome Biology and Evolution*, evx006. <https://doi.org/10.1093/gbe/evx006>
- Jaramillo, M. L., Guzman, F., Paese, C. L. B., Margis, R., Nazari, E. M., Ammar, D., & Müller, Y. M. R. (2016). Exploring developmental gene toolkit and associated pathways in a potential new model crustacean using transcriptomic analysis. *Development Genes and Evolution*, 226(5), pp. 325–337. <https://doi.org/10.1007/s00427-016-0551-6>
- Jeng, M. S., Ng, N. K., & Ng, P. K. L. (2004). Feeding behaviour: hydrothermal vent crabs feast on sea 'snow'. *Nature*, 432(7020), p. 969.
- Jollivet, D., Empis, A., Baker, M. C., Hourdez, S., Comtet, T., Jouin-Toulmond, C., ... & Tyler, P. A. (2000). Reproductive biology, sexual dimorphism, and population structure of the deep sea hydrothermal vent scale-worm, *Branchipolynoe seepensis* (Polychaeta: Polynoidae). *Journal of the Marine Biological Association of the United Kingdom*, 80(1), pp. 55–68.
- Jouin-Toulmond, C., Zal, F., & Hourdez, S. (1997). Genital apparatus and ultrastructure of the spermatozoa in *Alvinella pompejana* (Annelida: Polychaeta). *Cahiers de Biologie Marine*, (2).
- Jones, M. L. (1981). *Riftia pachyptila*, new genus, new species, the vestimentiferan from the Galapagos Rift geothermal vents (Pogonophora). *Proceedings of the National Academy of Sciences*, (93), pp. 1295–1313.
- Jones, M. L. (1985). On the Vestimentifera, new phylum: six new species, and other taxa, from hydrothermal vents and elsewhere. *Biol Soc Wash Bull*, (6), pp. 117–158.
- Jones, M. L., & Gardiner, S. L. (1988). Evidence for a transient digestive tract in Vestimentifera. *Proc. Biol. Soc. Wash*, (101), pp. 423–433.
- Jumars, P. A., Dorgan, K. M., & Lindsay, S. M. (2015). Diet of worms emended: an update of polychaete feeding guilds. *Annual Review of Marine Science*, 7(1), pp. 497–520. <https://doi.org/10.1146/annurev-marine-010814-020007>
- Kawahara-Miki, R., Wada, K., Azuma, N., & Chiba, S. (2011). Expression Profiling without Genome Sequence Information in a Non-Model Species, Pandalid Shrimp (*Pandalus latirostris*), by Next-Generation Sequencing. *PLoS ONE*, 6(10), e26043. <https://doi.org/10.1371/journal.pone.0026043>

- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), pp. 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Kenk, V. C., Wilson, B., R. (1985). A new mussel (Bivalvia, Mytilidae) from hydrothermal vents in the Galapagos Rift zone. *Malacologia*, (26), pp. 253-271.
- Kenny, N., Sin, Y., Shen, X., Zhe, Q., Wang, W., Chan, T., ... Hui, J. (2014). Genomic Sequence and Experimental Tractability of a New Decapod Shrimp Model, *Neocaridina denticulata*. *Marine Drugs*, 12(3), pp. 1419–1437. <https://doi.org/10.3390/md12031419>
- Kim, S.-J., Lee, K. Y., & Ju, S.-J. (2013). Nuclear mitochondrial pseudogenes in *Austinograea alayseae* hydrothermal vent crabs (Crustacea: Bythograeidae): effects on DNA barcoding. *Molecular Ecology Resources*, 13(5), pp. 781–787. <https://doi.org/10.1111/1755-0998.12119>
- Kim, S.-J., Kim, H. S., & Ju, S.-J. (2014). Mitochondrial genome of the hydrothermal vent crab *Austinograea alayseae* (Crustacea: Bythograeidae): genetic differences between individuals from Tofua Arc and Manus Basin. *Mitochondrial DNA*, 25(4), 251–252. <https://doi.org/10.3109/19401736.2013.800489>
- Kim, S.-J., Moon, J.-W., & Ju, S.-J. (2015). Complete mitochondrial genome of the blind vent crab *Gandalfus puia* (Crustacea: Bythograeidae) from the Tonga Arc. *Mitochondrial DNA*, 1–2. <https://doi.org/10.3109/19401736.2015.1046162>
- Kojima, S., Hashimoto, T., Hasegawa, M., Murata, S., Ohta, S., Seki, H., & Okada, N. (1993). Close phylogenetic relationship between vestimentifera (tube worms) and annelida revealed by the amino acid sequence of elongation factor-1a. *Journal of Molecular Evolution*, 37(1), pp. 66–70.
- Komai, T., Menot, L., & Segonzac, M. (2016). New records of caridean shrimp (Crustacea: Decapoda) from hydrothermally influenced fields off Futuna Island, Southwest Pacific, with description of a new species assigned to the genus *Alvinocaridinides* Komai & Chan, 2010 (Alvinocarididae). *Zootaxa*, 4098(2), 298. <https://doi.org/10.11646/zootaxa.4098.2.5>
- Komai, T., & Segonzac, M. (2003). Review of the hydrothermal vent shrimp genus *Mirocaris*, redescription of *M. fortunata* and reassessment of the taxonomic status of the family Alvinocarididae (Crustacea: Decapoda: Caridea). *Cahiers de Biologie Marine*, 44(2), pp. 199–215.
- Koonjul, P. K., Brandt, W. F., Lindsey, G. G., & Farrant, J. M. (1999). Inclusion of polyvinylpyrrolidone in the polymerase chain reaction reverses the inhibitory effects of polyphenolic contamination of RNA. *Nucleic acids research*, 27(3), pp. 915–916.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), pp. 357–359. <https://doi.org/10.1038/nmeth.1923>
- Leignel, V., Cibois, M., Moreau, B., & Chénais, B. (2007). Identification of new subgroup of HSP70 in Bythograeidae (hydrothermal crabs) and Xanthidae. *Gene*, 396(1), pp. 84–92. <https://doi.org/10.1016/j.gene.2007.02.025>
- Leignel, V., Hurtado, L. A., & Segonzac, M. (2018). Ecology, adaptation and acclimatisation mechanisms of Bythograeidae Williams, 1980, a unique endemic hydrothermal vent crabs family: current state of knowledge. *Marine and Freshwater Research*, 69(1), p. 1. <https://doi.org/10.1071/MF17007>
- Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics*, 12(1), p. 323.
- Li, C. P., De Grave, S., Chan, T.-Y., Lei, H. C., & Chu, K. H. (2011). Molecular systematics of caridean shrimps based on five nuclear genes: Implications for superfamily classification. *Zoologischer Anzeiger - A Journal of Comparative Zoology*, 250(4), pp. 270–279. <https://doi.org/10.1016/j.jcz.2011.04.003>
- Li, M., Schroeder, R., Ko, A., & Stoneking, M. (2012). Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTs. *Nucleic Acids Research*, 40(18), p. 137. <https://doi.org/10.1093/nar/gks499>
- Li, Y., Kocot, K. M., Whelan, N. V., Santos, S. R., Waits, D. S., Thornhill, D. J., & Halanych, K. M. (2017). Phylogenomics of tubeworms (Siboglinidae, Annelida) and comparative performance of different reconstruction methods. *Zoologica Scripta*, 46(2), pp. 200–213. <https://doi.org/10.1111/zsc.12201>

- Liao, Y., De Grave, S., Ho, T. W., Ip, B. H. Y., Tsang, L. M., Chan, T.-Y., & Chu, K. H. (2017). Molecular phylogeny of Pasiphaeidae (Crustacea, Decapoda, Caridea) reveals systematic incongruence of the current classification. *Molecular Phylogenetics and Evolution*, *115*, pp. 171–180. <https://doi.org/10.1016/j.ympev.2017.07.021>
- Lister, C. R. B. (1972). On the thermal balance of a mid-ocean ridge. *Geophysical Journal International*, *26*(5), pp. 515–535.
- Liu, S., Wang, X., Xie, L., Tan, M., Li, Z., Su, X., ... Zhou, X. (2016). Mitochondrial capture enriches mito-DNA 100 fold, enabling PCR-free mitogenomics biodiversity analysis. *Molecular Ecology Resources*, *16*(2), pp. 470–479. <https://doi.org/10.1111/1755-0998.12472>
- Lonsdale, P. (1977). Clustering of suspension-feeding macrobenthos near abyssal hydrothermal vents at oceanic spreading centers. *Deep-Sea Research*, *24*(9), pp. 857–863.
- Lopez, J. V., Yuhki, N., Masuda, R., Modi, W., & O'Brien, S. J. (1994). Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *Journal of Molecular Evolution*, *39*(2), pp. 174–190.
- Lorion, J., Buge, B., Cruaud, C., & Samadi, S. (2010). New insights into diversity and evolution of deep-sea Mytilidae (Mollusca: Bivalvia). *Molecular Phylogenetics and Evolution*, *57*(1), pp. 71–83. <https://doi.org/10.1016/j.ympev.2010.05.027>
- Katoh, K., Misawa, K., Kuma, K. I., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research*, *30*(14), pp. 3059–3066.
- Mandon, P., Aznar-Cormano, L., Hourdez, S., & Samadi, S. (2017). Assembly of the mitochondrial genome of the hydrothermal vent crab *Segonzacia mesatlantica* and detection of potential nuclear pseudogenes. *Mitochondrial DNA Part B*, *2*(1), pp. 291–293. <https://doi.org/10.1080/23802359.2017.1318674>
- Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, *27*(6), pp. 764–770. <https://doi.org/10.1093/bioinformatics/btr011>
- Marcus, J., & Hourdez, S. (2002). A new species of scale-worm (Polychaeta: Polynoidae) from Axial Volcano, Juan de Fuca Ridge, northeast Pacific. *Proceedings of the Biological Society of Washington*, *115*(2), pp. 341–349.
- Márquez, E. J., Hurtado-Alarcón, J. C., Isaza, J. P., Alzate, J. F., & Campos, N. H. (2014). Mitochondrial genome of the Caribbean king crab *Damithrax spinosissimus* (Lamarck, 1818) (Decapoda: Majidae). *Mitochondrial DNA*, pp. 1–2. <https://doi.org/10.3109/19401736.2014.961140>
- Martin, D., & Britayev, T. (1998). Symbiotic Polychaetes: Review of known species. *Oceanography and Marine Biology: An Annual Review* *36*(1), pp. 217–340. <https://doi.org/10.1016/j.catena.2010.08.006>
- Martin, J. W., & Haney, T. A. (2005). Decapod crustaceans from hydrothermal vents and cold seeps: a review through 2005. *Zoological Journal of the Linnean Society*, *145*(4), pp. 445–522.
- Martin, J. W., Hessler, R., R. (1990). *Chorocaris vandoverae*, a new genus and species of hydrothermal vent shrimp (Crustacea, Decapoda, Bresiliidae). *Contributions in Science*, (417), pp. 1-11.
- Martin, J. W., & Christiansen, J. C. (1995). A new species of the shrimp genus *Chorocaris* Martin & Hessler, 1990 (Crustacea: Decapoda: Bresiliidae) from hydrothermal vent fields along the Mid-Atlantic Ridge. *Proceedings – Biological Society of Washington*, (108), pp. 220-220.
- Mateos, M., Hurtado, L. A., Santamaria, C. A., Leignel, V., & Guinot, D. (2012). Molecular Systematics of the Deep-Sea Hydrothermal Vent Endemic Brachyuran Family Bythograeidae: A Comparison of Three Bayesian Species Tree Methods. *PLoS ONE*, *7*(3), e32066. <https://doi.org/10.1371/journal.pone.0032066>
- McArthur, A. G., & Koop, B. F. (1999). Partial 28S rDNA sequences and the antiquity of hydrothermal vent endemic gastropods. *Molecular Phylogenetics and Evolution*, *13*(2), pp. 255–274.
- McArthur, A. G., & Tunnicliffe, V. (1998). Relics and antiquity revisited in the modern vent fauna. *Geological Society, London, Special Publications*, *148*(1), pp. 271–291. <https://doi.org/10.1144/GSL.SP.1998.148.01.15>
- Mejía-ortíz, L. M., Yañez, G., & López-Mejía, M. (2017). Anchialocarididae, a new family of anchialine decapods and a new species of the genus *Agostocaris* from Cozumel Island, Mexico. *Crustaceana*, *90*(4), pp. 381-398.
- Metzker, M. L. (2005). Emerging technologies in DNA sequencing. *Genome research*, *15*(12), pp. 1767–1776.

- Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *Gateway Computing Environments Workshop (GCE)*, pp. 1–8.
- Miller, W., Drautz, D. I., Janecka, J. E., Lesk, A. M., Ratan, A., Tomsho, L. P., ... Schuster, S. C. (2008). The mitochondrial genome sequence of the Tasmanian tiger (*Thylacinus cynocephalus*). *Genome Research*, *19*(2), pp. 213–220. <https://doi.org/10.1101/gr.082628.108>
- Miura, T. (1991). Two new branchiate scale worms (Polynoidae: Polychaeta) from the hydrothermal vent of the Okinawa Trough and the volcanic seamount off Chichijima Island. *Proceedings of the Biological Society of Washington*, (104), pp. 166-174.
- Miura, T. (1994). Two new scale-worms (Polynoidae: polychaeta) from the Lau back-arc and North Fiji basins, South Pacific Ocean. *Proceedings of the Biological Society of Washington*, (107), pp. 532-543.
- Miyazaki, J.-I., Martins, L. de O., Fujita, Y., Matsumoto, H., & Fujiwara, Y. (2010). Evolutionary Process of Deep-Sea *Bathymodiolus* Mussels. *PLoS ONE*, *5*(4), e10363. <https://doi.org/10.1371/journal.pone.0010363>
- Moalic, Y., Desbruyères, D., Duarte, C. M., Rozenfeld, A. F., Bachraty, C., & Arnaud-Haond, S. (2011). Biogeography Revisited with Network Theory: Retracing the History of Hydrothermal Vent Communities. *Systematic Biology*, *61*(1), p. 127. <https://doi.org/10.1093/sysbio/syr088>
- Moreton, J., Izquierdo, A., & Emes, R. D. (2016). Assembly, Assessment, and Availability of *De novo* Generated Eukaryotic Transcriptomes. *Frontiers in Genetics*, (6). <https://doi.org/10.3389/fgene.2015.00361>
- Muir, A. I. (1982). Generic Characters in the Polynoinae (Annelida, Polychaeta), with Notes on the Higher Classification of Scale-worms (Aphroditacea). *British Museum (Natural History)*.
- Neal, L., Wiklund, H., Muir, A. I., Linse, K., & Glover, A. G. (2014). The identity of juvenile Polynoidae (Annelida) in the Southern Ocean revealed by DNA taxonomy, with notes on the status of *Herdmanella gracilis* Ehlers sensu Augener. *Memoirs of Museum Victoria*, (71), pp. 203–216. <https://doi.org/10.24199/j.mmv.2014.71.16>
- Nelson, D. C., Hagen, K. D., & Edwards, D. B. (1995). The gill symbiont of the hydrothermal vent mussel *Bathymodiolus thermophilus* is a psychrophilic, chemoautotrophic, sulfur bacterium. *Marine Biology*, *121*(3), pp. 487–495.
- Newman, W. A. (1985). The abyssal hydrothermal vent invertebrate fauna. A glimpse of antiquity? *Bulletin of the Biological Society of Washington*, (6), pp. 231-242.
- Ng, P. K., Guinot, D., & Davie, P. J. (2008). Systema Brachyurorum: Part I. An annotated checklist of extant brachyuran crabs of the world. *The Raffles Bulletin of Zoology*, *17*(1), pp. 1-286.
- Norlinder, E., Nygren, A., Wiklund, H., & Pleijel, F. (2012). Phylogeny of scale-worms (Aphroditiformia, Annelida), assessed from 18SrRNA, 28SrRNA, 16SrRNA, mitochondrial cytochrome c oxidase subunit I (COI), and morphology. *Molecular Phylogenetics and Evolution*, *65*(2), pp. 490-500. <https://doi.org/10.1016/j.ympev.2012.07.002>
- Nørrevang, A. (1970). The position of Pogonophora in the phylogenetic system. *Journal of Zoological Systematics and Evolutionary Research*, *8*(1), pp. 161–172.
- Nye, V., Copley, J., & Plouviez, S. (2012). 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*, *92*(5), pp. 1057-1072. <https://doi.org/10.1017/S0025315411002001>
- O'Hara, T. D., Hugall, A. F., Thuy, B., & Moussalli, A. (2014). Phylogenomic Resolution of the Class Ophiuroidea Unlocks a Global Microfossil Record. *Current Biology*, *24*(16), pp. 1874-1879. <https://doi.org/10.1016/j.cub.2014.06.060>
- Palumbi, S. R. (1996). Nucleic acids II: the polymerase chain reaction. *Molecular systematics*, pp. 205-247.
- Pante, E., Corbari, L., Thubaut, J., Chan, T. Y., Mana, R., Boisselier, M. C., ... & Samadi, S. (2012). Exploration of the deep-sea fauna of Papua New Guinea. *Oceanography*, *25*(3), pp. 214-225.
- Pante, E., Abdelkrim, J., Viricel, A., Gey, D., France, S. C., Boisselier, M.-C., & Samadi, S. (2015). Use of RAD sequencing for delimiting species. *Heredity*, *114*(5), p. 450.
- Perina, A., González-Tizón, A. M., Meilán, I. F., & Martínez-Lage, A. (2017). De novo transcriptome assembly of shrimp *Palaemon serratus*. *Genomics Data*, (11), pp. 89-91. <https://doi.org/10.1016/j.gdata.2016.12.009>

- Pernet, B. (2000). Reproduction and development of three symbiotic scale worms (Polychaeta: Polynoidae). *Invertebrate Biology*, 119(1), pp. 45–57.
- Pettibone, M. H. (1976). Revision of the Genus *Macellicephalo* McIntosh and the Subfamily Macellicephalinae Hartmann-SchrAder (Polychaeta: Polynoidae).
- Pettibone, M. H. (1983). A new scale worm (Polychaeta: Polynoidae) from the hydrothermal rift-area off western Mexico at 21 N. *Proceedings of the Biological Society of Washington*, 96(3), pp. 392-399.
- Pettibone, M. H. (1984). A new scale-worm commensal with deep-sea mussels on the Galapagos hydrothermal vent (Polychaeta: Polynoidae). *Proceedings of the Biological Society of Washington*, (97), pp. 226-239.
- Pettibone, M. H. (1985a). An additional new scale worm (Polychaeta: Polynoidae) from the hydrothermal rift area off western Mexico at 21 N.
- Pettibone, M. H. (1985b). Additional branchiate scale-worms (Polychaeta: Polynoidae) from Galapagos hydrothermal vent and rift-area off Western Mexico at 21 N.
- Pettibone, M. H. (1986). Review of the Iphioninae (Polychaeta: Polynoidae) and Revision of *Iphione Cimex* Quatrefages, *Gattyana Deludens* Fauvel and *Harmothoe Iphionelloides* Johnson (Harmothoinae).
- Pettibone, M. H. (1989). Polynoidae and Sigalionidae (Polychaeta) from the Guaymas Basin, with descriptions of two new species, and additional records from hydrothermal vents of the Galapagos Rift, 21 N, and seep-sites in the Gulf of Mexico (Florida and Louisiana).
- Pettibone, M. H. (1993). Polynoid polychaetes associated with a whale skeleton in the bathyal Santa Catalina Basin. *Proceedings of the Biological Society of Washington*, 106(4), pp. 678-688.
- Pleijel, F., Dahlgren, T. G., & Rouse, G. W. (2009). Progress in systematics: from Siboglinidae to Pogonophora and Vestimentifera and back to Siboglinidae. *Comptes Rendus Biologies*, 332(2-3), pp. 140-148. <https://doi.org/10.1016/j.crv.2008.10.007>
- Pleijel, F., Jondelius, U., Norlinder, E., Nygren, A., Oxelman, B., Schander, C., ... Thollesson, M. (2008). Phylogenies without roots? A plea for the use of vouchers in molecular phylogenetic studies. *Molecular Phylogenetics and Evolution*, 48(1), pp. 369-371. <https://doi.org/10.1016/j.ympev.2008.03.024>
- Plouviez, S., Jacobson, A., Wu, M., & Van Dover, C. L. (2015). Characterization of vent fauna at the Mid-Cayman Spreading Center. *Deep Sea Research Part I: Oceanographic Research Papers*, (97), pp. 124-133. <https://doi.org/10.1016/j.dsr.2014.11.011>
- Portail, M., Olu, K., Escobar-Briones, E., Caprais, J. C., Menot, L., Waeles, M., ... Sarrazin, J. (2015). Comparative study of vent and seep macrofaunal communities in the Guaymas Basin. *Biogeosciences*, 12(18), pp. 5455-5479. <https://doi.org/10.5194/bg-12-5455-2015>
- Projecto-Garcia, J., Zorn, N., Jollivet, D., Schaeffer, S. W., Lallier, F. H., & Hourdez, S. (2010). Origin and Evolution of the Unique Tetra-Domain Hemoglobin from the Hydrothermal Vent Scale Worm Branchipolynoe. *Molecular Biology and Evolution*, 27(1), pp. 143–152. <https://doi.org/10.1093/molbev/msp218>
- Projecto-Garcia, J., Jollivet, D., Mary, J., Lallier, F. H., Schaeffer, S. W., & Hourdez, S. (2015). Selective forces acting during multi-domain protein evolution: the case of multi-domain globins. *SpringerPlus*, 4(1), p. 354.
- Projecto-Garcia, J., Le Port, A.-S., Govindji, T., Jollivet, D., Schaeffer, S. W., & Hourdez, S. (2017). Evolution of Single-Domain Globins in Hydrothermal Vent Scale-Worms. *Journal of Molecular Evolution*, 85(5-6), pp. 172-187. <https://doi.org/10.1007/s00239-017-9815-7>
- Rainford, J. L., Hofreiter, M., Nicholson, D. B., & Mayhew, P. J. (2014). Phylogenetic Distribution of Extant Richness Suggests Metamorphosis Is a Key Innovation Driving Diversification in Insects. *PLoS ONE*, 9(10), e109085. <https://doi.org/10.1371/journal.pone.0109085>
- Rambaut, A., & Drummond, A. J., (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7(1), p. 214. <https://doi.org/10.1186/1471-2148-7-214>

- Ravara, A., Ramos, D., Teixeira, M. A. L., Costa, F. O., & Cunha, M. R. (2017). Taxonomy, distribution and ecology of the order Phyllodocida (Annelida, Polychaeta) in deep-sea habitats around the Iberian margin. *Deep Sea Research Part II: Topical Studies in Oceanography*, (137), pp. 207–231. <https://doi.org/10.1016/j.dsr2.2016.08.008>
- Read, G.; Fauchald, K. (Ed.) (2018). World Polychaeta database. Polynoidae Kinberg, 1856. Accessed through: World Register of Marine Species at: <http://www.marinespecies.org/aphia.php?p=taxdetails&id=939> on 2018-07-17
- Rees, D. J., Belzile, C., Glémet, H., & Dufresne, F. (2008). Large genomes among caridean shrimp. *Genome*, 51(2), pp. 159–163. <https://doi.org/10.1139/G07-108>
- Richardson, A. J., & Cook, R. A. (2006). Habitat use by caridean shrimps in lowland rivers. *Marine and Freshwater Research*, 57(7), p. 695. <https://doi.org/10.1071/MF05160>
- Richly, E., & Leister, D. (2004). NUMTs in sequenced eukaryotic genomes. *Molecular biology and evolution*, 21(6), pp. 1081–1084.
- Rivera, M. C., Jain, R., Moore, J. E., & Lake, J. A. (1998). Genomic evidence for two functionally distinct gene classes. *Proceedings of the National Academy of Sciences*, 95(11), pp. 6239–6244.
- Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S. D., ... Birol, I. (2010). *De novo* assembly and analysis of RNA-seq data. *Nature Methods*, 7(11), pp. 909–912. <https://doi.org/10.1038/nmeth.1517>
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, 61(3), pp. 539–542. <https://doi.org/10.1093/sysbio/sys029>
- Rouse, G. W., & Fauchald, K. (1995). The articulation of annelids. *Zoologica Scripta*, 24(4), pp. 269–301.
- Rouse, G. W., & Fauchald, K. (1997). Cladistics and polychaetes. *Zoologica Scripta*, 26(2), pp. 139–204.
- Rousset, V., Pleijel, F., Rouse, G. W., Erséus, C., & Siddall, M. E. (2007). A molecular phylogeny of annelids. *Cladistics*, 23(1), pp. 41–63. <https://doi.org/10.1111/j.1096-0031.2006.00128.x>
- Samadi, S., Corbari, L., Lorion, J., Hourdez, S., & Haga, T. (2010). Biodiversity of deep-sea organisms associated with sunken-wood or other organic remains sampled in the tropical Indo-Pacific. *Cahiers de Biologie Marine*, 2010, vol. 51, no 4, p. 459.
- Samadi, S. (2015). Evolutionary origins of hydrothermal vents metazoans. *BIO Web of Conferences*, 4, 00007. <https://doi.org/10.1051/bioconf/20150400007>
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences*, 74(12), pp. 5463–5467.
- Sarvala, J. (1971). Ecology of *Harmothoe sarsi* (Malmgren)(Polychaeta, Polynoidae) in the northern Baltic area. In *Annales Zoologici Fennici*. JSTOR. pp. 231–309
- Sell, A. F. (2000). Life in the extreme environment at a hydrothermal vent: haemoglobin in a deep-sea copepod. *Proceedings of the Royal Society B: Biological Sciences*, 267(1459), pp. 2323–2326. <https://doi.org/10.1098/rspb.2000.1286>
- Serpetti, N., Taylor, M. L., Brennan, D., Green, D. H., Rogers, A. D., Paterson, G. L. J., & Narayanaswamy, B. E. (2017). Ecological adaptations and commensal evolution of the Polynoidae (Polychaeta) in the Southwest Indian Ocean Ridge: A phylogenetic approach. *Deep Sea Research Part II: Topical Studies in Oceanography*, (137), pp. 273–281. <https://doi.org/10.1016/j.dsr2.2016.06.004>
- Shank, T. M., Lutz, R. A., & Vrijenhoek, R. C. (1998). Molecular systematics of shrimp (Decapoda: Bresiliidae) from deep-sea hydrothermal vents, I: Enigmatic "small orange" shrimp from the Mid-Atlantic ridge are juvenile *Rimicaris exoculata*. *Molecular marine biology and biotechnology*, 7(2), pp. 88–96.
- Shen, H., Braband, A., & Scholtz, G. (2013). Mitogenomic analysis of decapod crustacean phylogeny corroborates traditional views on their relationships. *Molecular Phylogenetics and Evolution*, 66(3), pp. 776–789. <https://doi.org/10.1016/j.ympev.2012.11.002>
- Shen, X.-X., Zhou, X., Kominek, J., Kurtzman, C. P., Hittinger, C. T., & Rokas, A. (2016). Reconstructing the backbone of the Saccharomycotina yeast phylogeny using genome-scale data. *G3: Genes, Genomes, Genetics*, pp. 3–116.

- Shields, M. A., Glover, A. G., & Wiklund, H. (2013). Polynoid polychaetes of the Mid-Atlantic Ridge and a new holothurian association. *Marine Biology Research*, 9(5-6), pp. 547-553. <https://doi.org/10.1080/17451000.2012.749992>
- Short, J. W., Humphrey, C. L., & Page, T. J. (2013). Systematic revision and reappraisal of the Kakaducarididae Bruce (Crustacea: Decapoda: Caridea) with the description of three new species of *Leptopalaemon* Bruce & Short. *Invertebrate Systematics*, 27(1), p. 87. <https://doi.org/10.1071/IS12016>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), pp. 3210-3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Sjölin, E., Erséus, C., & Källersjö, M. (2005). Phylogeny of Tubificidae (Annelida, Clitellata) based on mitochondrial and nuclear sequence data. *Molecular phylogenetics and evolution*, 35(2), pp. 431-441.
- Smith, C. R., Glover, A. G., Treude, T., Higgs, N. D., & Amon, D. J. (2015). Whale-Fall Ecosystems: Recent Insights into Ecology, Paleoecology, and Evolution. *Annual Review of Marine Science*, 7(1), pp. 571-596. <https://doi.org/10.1146/annurev-marine-010213-135144>
- Song, H., Moulton, M. J., & Whiting, M. F. (2014). Rampant Nuclear Insertion of mtDNA across Diverse Lineages within Orthoptera (Insecta). *PLoS ONE*, 9(10), e110508. <https://doi.org/10.1371/journal.pone.0110508>
- Southward, A. J., Southward, E. C., Dando, P. R., Rau, G. H., Felbeck, H., & Flügel, H. (1981). Bacterial symbionts and low $^{13}\text{C}/^{12}\text{C}$ ratios in tissues of Pogonophora indicate unusual nutrition and metabolism. *Nature*, 293(5834), p. 616.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), pp. 1312-1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Sui, J., & Li, X. (2017). A new species and new record of deep-sea scale-worms (Polynoidae: Polychaeta) from the Okinawa Trough and the South China Sea. *Zootaxa*, 4238(4), p. 562. <https://doi.org/10.11646/zootaxa.4238.4.4>
- Sumida, P. Y. G., Alfaro-Lucas, J. M., Shimabukuro, M., Kitazato, H., Perez, J. A. A., Soares-Gomes, A., ... Fujiwara, Y. (2016). Deep-sea whale fall fauna from the Atlantic resembles that of the Pacific Ocean. *Scientific Reports*, 6(1). <https://doi.org/10.1038/srep22139>
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, 28(10), pp. 2731-2739. <https://doi.org/10.1093/molbev/msr121>
- Tan, M. H., Gan, H. M., Lee, Y. P., & Austin, C. M. (2016a). The complete mitogenome of the moon crab *Ashtoret lunaris* (Forsk., 1775), (Crustacea; Decapoda; Matutidae). *Mitochondrial DNA*, 27(2), pp. 1313-1314. <https://doi.org/10.3109/19401736.2014.945572>
- Tan, M. H., Gan, H. M., Lee, Y. P., & Austin, C. M. (2016b). The complete mitogenome of the swimming crab *Thalamita crenata* (Rüppell, 1830) (Crustacea; Decapoda; Portunidae). *Mitochondrial DNA*, 27(2), pp. 1275-1276. <https://doi.org/10.3109/19401736.2014.945553>
- Tatusov, R. L., Koonin, E. V., & Lipman, D. J. (1997). A genomic perspective on protein families. *Science*, 278(5338), pp. 631-637.
- Teixeira, S., Olu, K., Decker, C., Cunha, R. L., Fuchs, S., Hourdez, S., ... Arnaud-Haond, S. (2013). High connectivity across the fragmented chemosynthetic ecosystems of the deep Atlantic Equatorial Belt: efficient dispersal mechanisms or questionable endemism? *Molecular Ecology*, 22(18), pp. 4663-4680. <https://doi.org/10.1111/mec.12419>
- Thubaut, J., Puillandre, N., Faure, B., Cruaud, C., & Samadi, S. (2013). The contrasted evolutionary fates of deep-sea chemosynthetic mussels (Bivalvia, Bathymodiolinae). *Ecology and Evolution*, 3(14), pp. 4748-4766. <https://doi.org/10.1002/ece3.749>
- Toullec, J.-Y., Corre, E., Mandon, P., Gonzalez-Aravena, M., Ollivaux, C., & Lee, C.-Y. (2017). Characterization of the neuropeptidome of a Southern Ocean decapod, the Antarctic shrimp *Chorismus antarcticus*: Focusing on a new decapod ITP-like peptide belonging to the CHH peptide family. *General and Comparative Endocrinology*, (252), pp. 60-78. <https://doi.org/10.1016/j.ygcen.2017.07.015>

- Toullec, J.-Y., Vinh, J., Le Caer, J.-P., Shillito, B., & Soyez, D. (2002). Structure and phylogeny of the crustacean hyperglycemic hormone and its precursor from a hydrothermal vent crustacean: the crab *Bythograea thermydron*. *Peptides*, 23(1), pp. 31–42.
- Tsang, L. M., Schubart, C. D., Ah Yong, S. T., Lai, J. C. Y., Au, E. Y. C., Chan, T.-Y., ... Chu, K. H. (2014). Evolutionary History of True Crabs (Crustacea: Decapoda: Brachyura) and the Origin of Freshwater Crabs. *Molecular Biology and Evolution*, 31(5), pp. 1173–1187. <https://doi.org/10.1093/molbev/msu068>
- Tunnicliffe, V. (1992). The Nature and Origin of the Modern Hydrothermal Vent Fauna. *PALAIOS*, 7(4), p. 338. <https://doi.org/10.2307/3514820>
- Tunnicliffe, V., McArthur, A. G., & McHugh, D. (1998). A biogeographical perspective of the deep-sea hydrothermal vent fauna. In *Advances in marine biology*. (34), pp. 353–442. Academic Press.
- Tyler, P. A., & Young, C. M. (1999). Reproduction and dispersal at vents and cold seeps. *Journal of the Marine Biological Association of the United Kingdom*, 79(2), pp. 193–208.
- Tyler, P. A., German, C. R., Ramirez-Llodra, E., & Van Dover, C. L. (2002). Understanding the biogeography of chemosynthetic ecosystems. *Oceanologica Acta*, 25(5), pp. 227–241.
- Uschakov, P. V. (1977). Phylogenetic relationship in the family Polynoidae (Polychaeta). *Essays on the polychaetous annelids in memory of Dr Olga Hartman*. Allan Hancock Foundation, Los Angeles, pp. 29–38.
- Van Dover, C. L., Trask, J., Gross, J., & Knowlton, A. (1999). Reproductive biology of free-living and commensal polynoid polychaetes at the Lucky Strike hydrothermal vent field (Mid-Atlantic Ridge). *Marine Ecology Progress Series*, (181), pp. 201–214.
- Van Dover, C. (2000). The ecology of deep-sea hydrothermal vents. Princeton University Press.
- Van Dover, C. (2002). Trophic relationships among invertebrates at the Kairei hydrothermal vent field (Central Indian Ridge). *Marine Biology*, 141(4), pp. 761–772.
- Vereshchaka, A. L. (1997). A new family for a deep-sea caridean shrimp from North Atlantic hydrothermal vents. *Journal of the Marine Biological Association of the United Kingdom*, 77(2), pp. 425–438.
- Vereshchaka, A. L., Kulagin, D. N., & Lunina, A. A. (2015). Phylogeny and New Classification of Hydrothermal Vent and Seep Shrimps of the Family Alvinocarididae (Decapoda). *PLOS ONE*, 10(7), e0129975. <https://doi.org/10.1371/journal.pone.0129975>
- Voight, J. R. (2000). A deep-sea octopus (*Graneledone* cf. *boreopacifica*) as a shell-crushing hydrothermal vent predator. *Journal of Zoology*, 252(3), pp. 335–341.
- Von Rintelen, K., Page, T. J., Cai, Y., Roe, K., Stelbrink, B., Kuhajda, B. R., ... von Rintelen, T. (2012). Drawn to the dark side: A molecular phylogeny of freshwater shrimps (Crustacea: Decapoda: Caridea: Atyidae) reveals frequent cave invasions and challenges current taxonomic hypotheses. *Molecular Phylogenetics and Evolution*, 63(1), pp. 82–96. <https://doi.org/10.1016/j.ympev.2011.12.015>
- Von Sternberg, R., & Cumberlidge, N. (2001). On the heterotreme-thoracotreme distinction in the Eubrachyura de Saint Laurent, 1980 (Decapoda, Brachyura). *Crustaceana*, 74(4), pp. 321–338.
- Von Sternberg, R., Cumberlidge, N., & Rodriguez, G. (1999). On the marine sister groups of the freshwater crabs (Crustacea: Decapoda: Brachyura). *Journal of Zoological Systematics and Evolutionary Research*, 37(1), pp. 19–38.
- Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., ... Seehausen, O. (2013). Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, 22(3), pp. 787–798. <https://doi.org/10.1111/mec.12023>
- Wang, Y., & Sha, Z. (2016). A new species of the genus *Manuscaris* Komai & Tsuchida, 2015 (Decapoda, Caridea, Alvinocarididae) from the Manus Basin hydrothermal vents. *Crustaceana*, 89(13), pp. 1541–1550.
- Wang, Y.-R., & Sha, Z.-L. (2017). A new species of the genus *Alvinocaris* Williams and Chace, 1982 (Crustacea: Decapoda: Caridea: Alvinocarididae) from the Manus Basin hydrothermal vents, Southwest Pacific. *Zootaxa*, 4226(1), p. 126. <https://doi.org/10.11646/zootaxa.4226.1.7>

- Ward, N., & Moreno-Hagelsieb, G. (2014). Quickly Finding Orthologs as Reciprocal Best Hits with BLAT, LAST, and UBLAST: How Much Do We Miss? *PLoS ONE*, *9*(7), e101850. <https://doi.org/10.1371/journal.pone.0101850>
- Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., ... Zdobnov, E. M. (2017). BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Molecular Biology and Evolution*, *35*(3), pp. 543-548. <https://doi.org/10.1093/molbev/msx319>
- Watson, G. J., Langford, F. M., Gaudron, S. M., & Bentley, M. G. (2000). Factors influencing spawning and pairing in the scale worm *Harmothoe imbricata* (Annelida: Polychaeta). *The Biological Bulletin*, *199*(1), pp. 50–58.
- Webb, M. (1964). Additional notes on *sclerolinum brattstromi* (pogonophora) and the establishment of a new family, sclerolinidae. *Sarsia*, *16*(1), pp. 47-58. <https://doi.org/10.1080/00364827.1964.10409533>
- Webb, M. (1969). *Lamellibrachia barhami*, gen. nov., sp. nov., (Pogonophora), from the Northeast Pacific. *Bulletin of Marine Science*, *19*(1), pp. 18–47.
- Weber, R. E. (1978). Respiratory pigments. In *Physiology of annelids*. pp. 393-446. Academic Press.
- Williams, A. B. (1980). A new crab family from the vicinity of submarine thermal vents on the Galapagos Rift (Crustacea: Decapoda: Brachyura). *Proceedings of the Biological society of Washington*, *93*(2), pp. 443–472.
- Williams, A. B. (1988). New marine decapod crustaceans from waters influenced by hydrothermal discharge, brine and hydrocarbon seepage.
- Williams, A. B., & Dobbs, F. C. (1995). A new genus and species of caridean shrimp (Crustacea: Decapoda: Bresiliidae) from hydrothermal vents on Loihi Seamount, Hawaii. *Proceedings of the Biological society of Washington*, (108), pp. 228–228.
- Williams, A. B., & Rona, P. A. (1986). Two New Caridean Shrimps (Bresiliidae) from a Hydrothermal Field on the Mid-Atlantic Ridge. *Journal of Crustacean Biology*, *6*(3), p. 446. <https://doi.org/10.2307/1548184>
- Wilson, K. A., & Gore, R. H. (1979). Studies on decapod Crustacea from the Indian River region of Florida. XVI. The second known specimen and first continental record for *Discias serratirostris* Lebour, 1949 (Caridea, Bresiliidae). *Crustaceana*, *37*(2), pp. 311–315.
- Wolff, T. (2005). Composition and endemism of the deep-sea hydrothermal vent fauna. *Cahiers de Biologie marine*, *46*(2), pp. 97-104.
- Wood, H. M., González, V. L., Lloyd, M., Coddington, J., & Scharff, N. (2018). Next-generation museum genomics: Phylogenetic relationships among palpimanoid spiders using sequence capture techniques (Araneae: Palpimanoidea). *Molecular Phylogenetics and Evolution*. <https://doi.org/10.1016/j.ympev.2018.06.038>
- Xu, Z., Li, T., Li, E., Chen, K., Ding, Z., Qin, J. G., ... Ye, J. (2016). Comparative transcriptome analysis reveals molecular strategies of oriental river prawn *Macrobrachium nipponense* in response to acute and chronic nitrite stress. *Fish & Shellfish Immunology*, (48), pp. 254-265. <https://doi.org/10.1016/j.fsi.2015.12.005>
- Yang, J. S., Nagasawa, H., Fujiwara, Y., Tsuchida, S., & Yang, W. J. (2010). The complete mitogenome of the hydrothermal vent crab *Gandalfus yunohana* (Crustacea: Decapoda: Brachyura): a link between the Bythograeoidea and Xanthoidea. *Zoologica Scripta*, *39*(6), pp. 621-630.
- Yang, J.-S., Lu, B., Chen, D.-F., Yu, Y.-Q., Yang, F., Nagasawa, H., ... Yang, W.-J. (2013). When Did Decapods Invade Hydrothermal Vents? Clues from the Western Pacific and Indian Oceans. *Molecular Biology and Evolution*, *30*(2), pp. 305-309. <https://doi.org/10.1093/molbev/mss224>
- Yuhui, X., Lijun, Z., Yue, H., Xiaoqi, W., Chen, Z., Huilun, Z., ... Hongying, S. (2017). Complete mitochondrial genomes from two species of Chinese freshwater crabs of the genus *Sinopotamon* recovered using next-generation sequencing reveal a novel gene order (Brachyura, Potamidae). *ZooKeys*, *705*, pp. 41-60. <https://doi.org/10.3897/zookeys.705.11852>
- Zal, F., Desbruyeres, D., & Jouin-Toulmond, C. (1994). Sexual dimorphism in *Paralvinella grasslei*, a polychaete annelid from deep-sea hydrothermal vents. *Comptes rendus de l'Académie des sciences. Série 3, Sciences de la vie*, *317*(1), pp. 42-48.

Bibliographie

- Zdobnov, E. M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R. M., Simão, F. A., Ioannidis, P., ... Kriventseva, E. V. (2017). OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Research*, *45*(1), pp. 744-749. <https://doi.org/10.1093/nar/gkw1119>
- Zhang, J., Sun, Q., Luan, Z., Lian, C., & Sun, L. (2017). Comparative transcriptome analysis of *Rimicaris* sp. reveals novel molecular features associated with survival in deep-sea hydrothermal vent. *Scientific Reports*, *7*(1). <https://doi.org/10.1038/s41598-017-02073-9>
- Zhang, Y., Sun, J., Rouse, G. W., Wiklund, H., Pleijel, F., Watanabe, H. K., ... Qiu, J.-W. (2018). Phylogeny, evolution and mitochondrial gene order rearrangement in scale worms (Aphroditiformia, Annelida). *Molecular Phylogenetics and Evolution*, (125), pp. 220-231. <https://doi.org/10.1016/j.ympev.2018.04.002>
- Zhao, Q.-Y., Wang, Y., Kong, Y.-M., Luo, D., Li, X., & Hao, P. (2011). Optimizing *de novo* transcriptome assembly from short-read RNA-Seq data: a comparative study. In *BMC bioinformatics* (12). p. 2. BioMed Central.

Annexes

VII. ANNEXES

Annexe 1 Echantillonnage représenté dans la phylogénie des Polynoidae. La colonne taxon indique la famille (en bleu foncé et majuscule), la sous famille (en bleu clair et souligné) et le nom d'espèce pour chaque spécimen présent dans notre analyse phylogénétique. La deuxième colonne indique le nom correspondant dans l'arbre. La troisième colonne donne l'indication de voucher (le cas échéant). Les colonnes suivantes fournissent les numéros d'accessions NCBI des séquences téléchargées. Nos données de séquençage sont en gras, et n'ont pas encore de numéro d'accession NCBI (remplacé par « XXXX »).

| Taxon | Nom dans l'arbre | Voucher | 16S | 18S | COI | 28SD 1 | 28SD 9-10 |
|---|--|-----------------------|--------------|--------------|--------------|--------------|--------------|
| ACOETIDAE | | | | | | | |
| <i>Panthalis oerstedii</i> | SMNH-118954 <i>Panthalis oerstedii</i> | SMNH118954 | JN852 881 | | | JN8528 45 | |
| <i>Panthalis</i> sp. | <i>Panthalis</i> sp. | | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Polyodontes</i> sp. | <i>Polyodontes</i> sp. | | XXXX | XXXX | XXXX | XXXX | XXXX |
| | | | | | | | |
| APHRODITIDAE | | | | | | | |
| <i>Aphrodita aculeata</i> | SMNH-118956 <i>Aphrodita aculeata</i> | SMNH118956 | JN852 882 | | | JN8528 46 | |
| <i>Aphrodita negligens</i> | <i>Aphrodita negligens</i> | | | AY894 294 | AY894 309 | | |
| <i>Aphrodita</i> sp. | <i>Aphrodita</i> sp. | | | AY894 295 | AY894 310 | | DQ790 024 |
| <i>Aphrodita</i> sp. | <i>Aphrodita</i> sp. spm F1_P1 | | XXXX | XXXX | | XXXX | XXXX |
| <i>Heteraphrodita</i> sp. | <i>Heteraphrodita</i> sp. | | XXXX | XXXX | | XXXX | XXXX |
| <i>Laetmonice filicornis</i> | SMNH-118955 <i>Laetmonice filicornis</i> | SMNH118955 | JN852 883 | JN852 816 | JN852 919 | JN8528 47 | |
| <i>Laetmonice hystrix</i> | <i>Laetmonice hystrix</i> spm F10_P1 | | XXXX | XXXX | | XXXX | |
| <i>Palmyra aurifera</i> | SMNH-73633 <i>Palmyra aurifera</i> | SMNH73633 | | AY839 571 | AY839 583 | | |
| | | | | | | | |
| POLYNOIDAE | | | | | | | |
| <u>Arctonoinae</u> | | | | | | | |
| <i>Arctonoe fragilis</i> | <i>Arctonoe fragilis</i> | | XXXX | | | XXXX | |
| <i>Arctonoe fragilis</i> | BAMPOL0389 <i>Arctonoe fragilis</i> | BAMPOL0389 | | | HM47 3300 | | HM473 199 |
| <i>Arctonoe pulchra</i> | BIOUG<CAN>:BAMPOL0484 <i>Arctonoe pulchra</i> | BIOUG<CAN>:BAMPOL0484 | | | HM47 3309 | | HM473 200 |
| <i>Arctonoe pulchra</i> | <i>Arctonoe pulchra</i> | | XXXX | | | XXXX | |
| <i>Capitulatinoe</i> cf. <i>cupisetis</i> | SWU30116 <i>Capitulatinoe</i> cf. <i>cupisetis</i> | SWU30116 | KF919 303 | KF919 301 | | KF919 302 | |
| <i>Arctonoe vittata</i> | <i>Arctonoe vittata</i> | | XXXX | XXXX | | XXXX | |
| <i>Gastrolepidia clavigera</i> | SMNH-118964 <i>Gastrolepidia clavigera</i> | SMNH118964 | JN852 893 | JN852 825 | JN852 927 | JN8528 55 | |
| <u>Branchinotogluminae</u> | | | | | | | |
| <i>Branchinotogluma burkensis</i> | <i>Branchinotogluma burkensis</i> | SH092-378 | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Branchinotogluma fisheri gomensis</i> | <i>Branchinotogluma fisheri gomensis</i> | | XXXX | | | XXXX | |

| | | | | | | | |
|---|--|-------------------|--------------|--------------|--------------|--------------|------|
| <i>Branchinotogluma fisheri marensis</i> | MNHN-IA-2013-126 <i>Branchinotogluma fisheri marensis</i> | MNHN-IA-2013-126 | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Branchinotogluma grasslei</i> | <i>Branchinotogluma grasslei</i> spm1 | | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Branchinotogluma grasslei</i> | <i>Branchinotogluma grasslei</i> spm2 | | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Branchinotogluma hesleri</i> | <i>Branchinotogluma hesleri</i> | | XXXX | | XXXX | XXXX | XXXX |
| <i>Branchinotogluma jasoni</i> | <i>Branchinotogluma jasoni</i> | | XXXX | XXXX | | XXXX | XXXX |
| <i>Branchinotogluma marianus</i> | <i>Branchinotogluma marianus</i> spm G2_P2 | SH092-136 | XXXX | XXXX | XXXX | XXXX | |
| <i>Branchinotogluma marianus</i> | <i>Branchinotogluma marianus</i> spm H2_P2 | SH092-136 | XXXX | XXXX | XXXX | XXXX | |
| <i>Branchinotogluma mesatlantica</i> | <i>Branchinotogluma mesatlantica</i> | | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Branchinotogluma sandersi</i> | SMNH-118960 <i>Branchinotogluma sandersi</i> | SMNH118960 | JN852 889 | JN852 821 | JN852 923 | JN8528 51 | |
| <i>Branchinotogluma segonzaci</i> | <i>Branchinotogluma segonzaci</i> | SH092-937 | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Branchinotogluma</i> sp. | <i>Branchinotogluma</i> sp. | | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Branchinotogluma trifurcus</i> | <i>Branchinotogluma trifurcus</i> | SH092-136 | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Opisthotrochopodus tunnicliffeae</i> | <i>Opisthotrochopodus tunnicliffeae</i> | | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Peinaleopolynoe sillardi</i> | MNHN-IA-2010-395 <i>Peinaleopolynoe sillardi</i> | MNHN-IA-2010-395 | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Peinaleopolynoe</i> sp. | <i>Peinaleopolynoe</i> sp. | SH111-322 | XXXX | XXXX | | XXXX | XXXX |
| <u>Branchiplicatinae</u> | | | | | | | |
| <i>Branchiplicatus cupreus</i> | <i>Branchiplicatus cupreus</i> | | XXXX | XXXX | XXXX | | XXXX |
| <u>Branchipolynoinae</u> | | | | | | | |
| <i>Branchipolynoe</i> aff. <i>seepensis</i> | <i>Branchipolynoe</i> aff. <i>seepensis</i> spm D9_P1 | | XXXX | XXXX | XXXX | XXXX | |
| <i>Branchipolynoe</i> aff. <i>seepensis</i> | <i>Branchipolynoe</i> aff. <i>seepensis</i> spm H5_P1 | | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Branchipolynoe pettiboneae</i> | <i>Branchipolynoe pettiboneae</i> spm G9_P1 | | XXXX | | XXXX | XXXX | XXXX |
| <i>Branchipolynoe pettiboneae</i> | <i>Branchipolynoe pettiboneae</i> spm SH092-076 | SH092-076 | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Branchipolynoe seepensis</i> | <i>Branchipolynoe seepensis</i> spm B5_P1 | | | XXXX | XXXX | XXXX | |
| <i>Branchipolynoe</i> sp. | <i>Branchipolynoe</i> sp. | | XXXX | | XXXX | XXXX | |
| <i>Branchipolynoe symmytilida</i> | <i>Branchipolynoe symmytilida</i> | | XXXX | XXXX | | XXXX | |
| <u>Eulagiscinae</u> | | | | | | | |
| <i>Bathymoorea lucasi</i> | MNHN-IA-TYPE 1822 <i>Bathymoorea lucasi</i> | MNHN-IA-TYPE 1822 | MH23 3165 | MH23 3223 | MH23 3266 | | |
| <i>Eulagisca</i> aff. <i>macnabi</i> | MNHN-IA-2013-9-2 <i>Eulagisca</i> aff. <i>macnabi</i> | MNHN-IA-2013-9-2 | XXXX | XXXX | XXXX | | XXXX |
| <i>Eulagisca gigantea</i> | MNHN-IA-2013-32 <i>Eulagisca gigantea</i> | MNHN-IA-2013-32 | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Eulagisca uschakovi</i> | MNHN-IA-2013-3 <i>Eulagisca uschakovi</i> | MNHN-IA-2013-3 | XXXX | XXXX | XXXX | XXXX | XXXX |

| | | | | | | | |
|------------------------------------|--|-----------------------|--------------|--------------|--------------|---------------|--------------|
| <i>Eulagiscinae</i> sp. | MNHN IA-2013-124 Gen. sp. | MNHN-IA-2013-124 | XXXX | XXXX | XXXX | | XXXX |
| <u>Gesiellinae</u> | | | | | | | |
| <i>Gesiella jameensis</i> | <i>Gesiella jameensis</i> | | KY454 412 | KY454 403 | KY454 429 | KW502 _285 | |
| <u>Iphioninae</u> | | | | | | | |
| <i>Iphione</i> sp. | MNHN-IA-2015-73 <i>Iphione</i> sp. | MNHN-IA-2015-73 | XXXX | XXXX | | | XXXX |
| <i>Iphione</i> sp. | SMNH-118972 <i>Iphione</i> sp. | SMNH118972 | JN852 886 | JN852 819 | JN852 921 | | |
| <i>Iphione</i> sp. | <i>Iphione</i> sp. spm E3_P1 | | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Iphione</i> sp. | <i>Iphione</i> sp. | | XXXX | | | XXXX | XXXX |
| <i>Iphione treadwelli</i> | MNHN-IA-2013-154 <i>Iphione treadwelli</i> | MNHN-IA-2013-154 | XXXX | XXXX | | | XXXX |
| <i>Iphionella risensis</i> | <i>Iphionella risensis</i> | | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Iphionella</i> sp. | <i>Iphionella</i> sp1 | | XXXX | | XXXX | XXXX | XXXX |
| <i>Iphionella</i> sp. | <i>Iphionella</i> sp2 | | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Thermiphione fijensis</i> | <i>Thermiphione fijensis</i> | SH092-748 | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Thermiphione</i> sp. | SMNH-118982 <i>Thermiphione</i> sp. | SMNH118982 | JN852 887 | JN852 820 | JN852 922 | JN8528 49 | |
| <u>Lepidastheniinae</u> | | | | | | | |
| <i>Lepidasthenia berkeleyae</i> | BIOUG<CAN>:BAMPOL0272 <i>Lepidasthenia berkeleyae</i> | BIOUG<CAN>:BAMPOL0272 | | | HM47 3443 | | HM473 230 |
| <u>Lepidonotinae</u> | | | | | | | |
| <i>Alentia gelatinosa</i> | <i>Alentia gelatinosa</i> | | XXXX | XXXX | | XXXX | XXXX |
| <i>Alentia gelatinosa</i> | SMNH-73632 <i>Alentia gelatinosa</i> | SMNH73632 | | AY839 566 | AY839 577 | | |
| <i>Alentia</i> sp. | <i>Alentia</i> sp. | | XXXX | XXXX | | XXXX | XXXX |
| <i>Euphione</i> sp. | <i>Euphione</i> sp. | | | XXXX | XXXX | XXXX | XXXX |
| <i>Halosydna brevisetosa</i> | <i>Halosydna brevisetosa</i> | | | AY894 298 | AY894 313 | | |
| <i>Halosydna brevisetosa</i> | SMNH-118966 <i>Halosydna brevisetosa</i> | SMNH118966 | JN852 895 | JN852 827 | | JN8528 57 | |
| <i>Hermenia verruculosa</i> | SMNH-118970 <i>Hermenia verruculosa</i> | SMNH118970 | JN852 899 | JN852 830 | JN852 931 | JN8528 61 | |
| <i>Hyperhalosydna striata</i> | SMNH-118971 <i>Hyperhalosydna striata</i> | SMNH118971 | JN852 900 | JN852 831 | JN852 932 | JN8528 62 | |
| <i>Lepidonotus clava</i> | <i>Lepidonotus clava</i> | | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Lepidonotus</i> sp. | <i>Lepidonotus</i> sp. spm C2_P1 | | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Lepidonotus</i> sp. | <i>Lepidonotus</i> sp. spm D2_P1 | | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Lepidonotus squamatus</i> | SMNH-118975 <i>Lepidonotus squamatus</i> | SMNH118975 | JN852 903 | | | JN8528 65 | |
| <i>Lepidonotus sublevis</i> | <i>Lepidonotus sublevis</i> | | | | AY894 317 | DQ790 039 | DQ790 039 |
| <i>Thormora jukesii</i> | SMNH-118983 <i>Thormora jukesii</i> | SMNH118983 | JN852 910 | JN852 840 | JN852 941 | JN8528 71 | |
| <u>Lepidonotopodinae</u> | | | | | | | |
| <i>Lepidonotopodium fimbriatum</i> | <i>Lepidonotopodium fimbriatum</i> | | XXXX | | XXXX | XXXX | XXXX |
| <i>Lepidonotopodium jouinae</i> | <i>Lepidonotopodium jouinae</i> | | XXXX | | XXXX | | |
| <i>Lepidonotopodium piscesae</i> | <i>Lepidonotopodium piscesae</i> | | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Lepidonotopodium riftense</i> | <i>Lepidonotopodium riftense</i> | | XXXX | XXXX | XXXX | XXXX | |

| | | | | | | | |
|-------------------------------------|--|-------------------------|-------------|-------------|-------------|-------------|-------------|
| <i>Lepidonotopodium williamsae</i> | <i>Lepidonotopodium williamsae</i> spm1 | | XXXX | XXXX | | XXXX | XXXX |
| <i>Lepidonotopodium williamsae</i> | <i>Lepidonotopodium williamsae</i> spm E9_P1 | | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Thermopolynoe branchiata</i> | <i>Thermopolynoe branchiata</i> | SH092-064 | | XXXX | XXXX | XXXX | XXXX |
| Macellicephalinae | | | | | | | |
| <i>Arya acus</i> | MNHN-IA-TYPE-1811 <i>Arya acus</i> | MNHN-IA-TYPE 1811 | MH23 3179 | MH23 3231 | MH23 3277 | | |
| <i>Arya acus</i> | MNHN-IA-TYPE-1812 <i>Arya acus</i> | MNHN-IA-TYPE 1812 | MH23 3182 | MH23 3232 | MH23 3279 | | |
| <i>Austropolaria magnicirrata</i> | NHMUK-2012-95 <i>Austropolaria magnicirrata</i> | NHMUK:2012.95 | JX863 896 | JX863 895 | | | |
| <i>Bathyeliasona mariaae</i> | MNHN-IA-TYPE-1815 <i>Bathyeliasona mariaae</i> | MNHN-IA-TYPE 1815 | MH23 3149 | MH23 3204 | MH23 3249 | | |
| <i>Bathyfauvelia glacigena</i> | MNHN-IA-TYPE-1817 <i>Bathyfauvelia glacigena</i> | MNHN-IA-TYPE 1817 | MH23 3160 | MH23 3218 | MH23 3274 | | |
| <i>Bathyfauvelia ignigena</i> | MNHN-IA-TYPE-1820 <i>Bathyfauvelia ignigena</i> | MNHN-IA-TYPE 1820 | MH23 3161 | | MH23 3252 | | |
| <i>Bathyfauvelia ignigena</i> | MNHN-IA-TYPE-1819 <i>Bathyfauvelia ignigena</i> | MNHN-IA-TYPE 1819 | MH23 3200 | MH23 3246 | MH23 3262 | | |
| <i>Bathykurila guaymasensis</i> | <i>Bathykurila guaymasensis</i> spm 206 | | | DQ07 4765 | DQ07 4769 | | |
| <i>Bathykurila guaymasensis</i> | <i>Bathykurila guaymasensis</i> | | | DQ07 4765 | DQ07 4780 | | |
| <i>Bathypolaria</i> sp. | MNHN-IA-PNT-63 <i>Bathypolaria</i> sp. | MNHN-IA-PNT 63 | MH23 3151 | MH23 3206 | MH23 3281 | | |
| <i>Bathypolaria</i> sp. | MNHN-IA-PNT-66 <i>Bathypolaria</i> sp. | MNHN-IA-PNT 66 | MH23 3175 | MH23 3227 | MH23 3268 | | |
| <i>Bathypolaria</i> sp. | MNHN-IA-PNT-68 <i>Bathypolaria</i> sp. | MNHN-IA-PNT 68 | MH23 3193 | MH23 3241 | MH23 3280 | | |
| <i>Bruunilla</i> sp. | MNHN-IA-PNT-72 <i>Bruunilla</i> sp. | MNHN-IA-PNT 72 | MH23 3201 | MH23 3247 | MH23 3263 | | |
| <i>Levensteiniella iris</i> | <i>Levensteiniella iris</i> | | XXXX | XXXX | | XXXX | |
| <i>Levensteiniella kincaidi</i> | <i>Levensteiniella kincaidi</i> | | XXXX | | | XXXX | |
| <i>Levensteiniella plicata</i> | <i>Levensteiniella plicata</i> | SH101-381 | XXXX | XXXX | | XXXX | XXXX |
| <i>Levensteiniella raisae</i> | <i>Levensteiniella raisae</i> | SH092-136 | | XXXX | XXXX | XXXX | |
| <i>Macellicephala clarionensis</i> | MNHN-IA-TYPE-1829 <i>Macellicephala clarionensis</i> | MNHN-IA-TYPE 1829 | MH23 3184 | MH23 3234 | MH23 3270 | | |
| <i>Macellicephala parvafauces</i> | MNHN-IA-TYPE-1831 <i>Macellicephala parvafauces</i> | MNHN-IA-TYPE 1831 | MH23 3153 | | MH23 3282 | | |
| <i>Macellicephala</i> sp. | <i>Macellicephala</i> sp. spm MB210 | | KX867 370 | | KX867 446 | | |
| <i>Macellicephala</i> sp. | <i>Macellicephala</i> sp. spm MB36 | | KX867 378 | | KX867 448 | | |
| <i>Macellicephala</i> sp. | MNHN-IA-2010-303 <i>Macellicephala</i> sp. | MNHN-IA-2010-303 | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Pelagomacellicephala iliffei</i> | ZMUC_POL_2397 <i>Pelagomacellicephala iliffei</i> | ZMUC_POL_239 7 | KY454 428 | KY454 411 | KY454 443 | KW533 _28S | |
| <i>Pelagomacellicephala iliffei</i> | ZMUC_POL_2391 <i>Pelagomacellicephala iliffei</i> | ZMUC_POL_239 1 | KY454 414 | | KY454 430 | | |
| <i>Pelagomacellicephala iliffei</i> | ZMUC_POL_2392 <i>Pelagomacellicephala iliffei</i> | ZMUC_POL_239 2 | KY454 416 | KY454 405 | KY454 431 | | |

| | | | | | | | |
|-------------------------------------|---|-------------------------------|--------------|--------------|--------------|--------------|--------------|
| <i>Pelagomacellicephala iliffei</i> | ZMUC_POL_2393 <i>Pelagomacellicephala iliffei</i> | ZMUC_POL_2393 | KY454 419 | KY454 407 | KY454 434 | | |
| <i>Pelagomacellicephala iliffei</i> | ZMUC_POL_2395 <i>Pelagomacellicephala iliffei</i> | ZMUC_POL_2395 | KY454 422 | | KY454 437 | | |
| <i>Pelagomacellicephala iliffei</i> | ZMUC_POL_2396 <i>Pelagomacellicephala iliffei</i> | ZMUC_POL_2396 | KY454 424 | | KY454 440 | | |
| <i>Yodanoë desbruyeresi</i> | MNHN-IA-TYPE-1843 <i>Yodanoë desbruyeresi</i> | MNHN-IA-TYPE 1843 | MH23 3156 | | MH23 3251 | | |
| <i>Yodanoë</i> sp. | MNHN-IA-PNT-73 <i>Yodanoë</i> sp. | MNHN-IA-PNT 73 | MH23 3195 | | MH23 3273 | | |
| Polaruschakovinae | | | | | | | |
| <i>Polaruschakov lamellae</i> | MNHN-IA-TYPE-1837 <i>Polaruschakov lamellae</i> | MNHN-IA-TYPE 1837 | MH23 3150 | MH23 3205 | MH23 3250 | | |
| <i>Polaruschakov omnesae</i> | MNHN-IA-TYPE-1841 <i>Polaruschakov omnesae</i> | MNHN-IA-TYPE 1841 | MH23 3155 | MH23 3213 | MH23 3283 | | |
| Polynoïnae | | | | | | | |
| <i>Acholoe astericola</i> | <i>Acholoe astericola</i> | | XXXX | XXXX | | XXXX | XXXX |
| <i>Acholoe astericola</i> | SMNH-73630 <i>Acholoe astericola</i> | SMNH73630 | | AY839 567 | AY839 576 | | |
| <i>Antarctinoë ferox</i> | MNHN-IA-2013-18-2 <i>Antarctinoë ferox</i> | MNHN-IA-2013- 18-2 | XXXX | XXXX | XXXX | | |
| <i>Antipathipolyeunoa</i> sp. | NSJC066-3760-2 <i>Antipathipolyeunoa</i> sp. | NSJC066_ 3760_2 | KU738 149 | KU738 169 | KU738 202 | KU738 184 | |
| <i>Austrolaenilla antarctica</i> | NHMUK-2014-21 <i>Austrolaenilla antarctica</i> | NHMUK_2014_2 1 | KJ676 606 | | KJ676 620 | | |
| <i>Austrolaenilla pelagica</i> | NHMUK-2014-26 <i>Austrolaenilla pelagica</i> | NHMUK_2014_2 6 | KJ676 607 | | KJ676 632 | | |
| <i>Barrukia cristata</i> | MNHN-IA-2013-43-2 <i>Barrukia cristata</i> | MNHN-IA-2013- 43-2 | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Brychionoe</i> sp. | NSJC66-113-S001 <i>Brychionoe</i> sp. | NSJC66_113_SO 01 | KU738 167 | KU738 182 | | KU738 200 | |
| <i>Bylgides elegans</i> | SMNH-118962 <i>Bylgides elegans</i> | SMNH118962 | JN852 890 | JN852 822 | JN852 924 | JN8528 52 | |
| <i>Bylgides promamme</i> | NUNAV-0268 <i>Bylgides promamme</i> | NUNAV_0268 | | | HQ02 4273 | | HQ024 496 |
| <i>Bylgides sarsi</i> | SMNH-118961 <i>Bylgides sarsi</i> | SMNH118961 | JN852 891 | JN852 823 | JN852 925 | JN8528 53 | |
| <i>Enipo torelli</i> | 07-PROBE-05308 <i>Enipo torelli</i> | 07PROBE05308 | | | HQ02 3480 | | HQ023 831 |
| <i>Eunoe brunnea</i> | MNHN-IA-2015-539 <i>Eunoe brunnea</i> | MNHN-IA-2015- 539 | XXXX | XXXX | | XXXX | XXXX |
| <i>Eunoe</i> cf. <i>oerstedii</i> | SWU41305 <i>Eunoe</i> cf. <i>oerstedii</i> | SWU41305 | KF006 980 | KF006 979 | | KF006 981 | |
| <i>Eunoe nodosa</i> | SMNH-118963 <i>Eunoe nodosa</i> | SMNH118963 | JN852 892 | JN852 824 | JN852 926 | JN8528 54 | |
| <i>Eunoe</i> sp. | NSJC66_496_S001 <i>Eunoe</i> sp. | NSJC66_496_SO 01 | KU738 168 | KU738 183 | KU738 214 | KU738 201 | |
| <i>Eunoe</i> sp. | MNHN-IA-2013-122 <i>Eunoe</i> sp. | MNHN-IA-2013- 122 | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Gattyana ciliata</i> | <i>Gattyana ciliata</i> | | | AY894 297 | AY894 312 | | |
| <i>Gattyana cirrhosa</i> | SMNH-118965 <i>Gattyana cirrhosa</i> | SMNH118965 | JN852 894 | JN852 826 | JN852 928 | JN8528 56 | |
| <i>Gorekia</i> sp. | MNHN-IA-2015-391 <i>Gorekia</i> sp. | MNHN-IA-2015- 391 | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Gorgoniapolynoe caeciliae</i> | NSJC66-4277-1 <i>Gorgoniapolynoe caeciliae</i> | NSJC66_4277_1 | KU738 153 | KU738 172 | KU738 205 | KU738 188 | |

| | | | | | | | |
|--|---|-------------------------|--------------|--------------|--------------|--------------|--------------|
| <i>Gorgoniapolynoe corralophila</i> | NSJC66-133-S001 <i>Gorgoniapolynoe corralophila</i> | NSJC66_133_S001 | KU738 154 | KU738 173 | KU738 206 | KU738 189 | |
| <i>Gorgoniapolynoe corralophila</i> | NSJC66-3219-S001 <i>Gorgoniapolynoe corralophila</i> | NSJC66_3219_S001 | KU738 155 | | KU738 207 | KU738 190 | |
| <i>Gorgoniapolynoe corralophila</i> | MNHN-IA-2010-308 <i>Gorgoniapolynoe corralophila</i> | MNHN-IA-2010-308 | XXXX | XXXX | XXXX | XXXX | |
| <i>Harmothoe acuminata</i> | MNHN-IA-2015-541i | MNHN-IA-2015-541 | XXXX | XXXX | | XXXX | XXXX |
| <i>Harmothoe clavigera</i> | <i>Harmothoe clavigera</i> | | XXXX | XXXX | | XXXX | XXXX |
| <i>Harmothoe crosetensis</i> | <i>Harmothoe crosetensis</i> | | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Harmothoe fragilis</i> | BIOUG<CAN>:BAMPOL0417 <i>Harmothoe fragilis</i> | BIOUG<CAN>:BAMPOL0417 | | | HM47 3407 | | HM473 223 |
| <i>Harmothoe fuligineum</i> | <i>Harmothoe fuligineum</i> | | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Harmothoe fuligineum</i> | <i>Harmothoe fuligineum</i> spm MB91 | | KX867 264 | | KX867 415 | | |
| <i>Harmothoe fuligineum</i> | NHMUK-2014-29 <i>Harmothoe fuligineum</i> | NHMUK:2014.29 | KJ676 609 | | KJ676 634 | | |
| <i>Harmothoe glabra</i> | SMNH-118967 <i>Harmothoe glabra</i> | SMNH118967 | JN852 896 | JN852 828 | JN852 929 | JN8528 58 | |
| <i>Harmothoe imbricata</i> | BAMPOL0345 <i>Harmothoe imbricata</i> | BAMPOL0345 | | | HM47 3426 | | HM473 225 |
| <i>Harmothoe imbricata</i> | BIOUG-BAMPOL0025 <i>Harmothoe imbricata</i> | BIOUG_BAMPOL0025 | | | HM47 3412 | | HM473 224 |
| <i>Harmothoe imbricata</i> | SMNH-73635 <i>Harmothoe imbricata</i> | SMNH_73635 | AY340 463 | AY340 434 | AY839 580 | AY340 400 | |
| <i>Harmothoe impar</i> | SMNH-118968 <i>Harmothoe impar</i> | SMNH118968 | JN852 897 | JN852 829 | JN852 930 | JN8528 59 | |
| <i>Harmothoe oculinarum</i> | <i>Harmothoe oculinarum</i> | | | AY894 299 | AY894 314 | | |
| <i>Harmothoe oculinarum</i> | SMNH-118969 <i>Harmothoe oculinarum</i> | SMNH118969 | JN852 898 | | | JN8528 60 | |
| <i>Harmothoe rarispina</i> | BIOUG<CAN>:BAMPOL0432 <i>Harmothoe rarispina</i> | BIOUG<CAN>:BAMPOL0432 | | | HM47 3428 | | HM473 226 |
| <i>Harmothoe rarispina</i> | BIOUG<CAN>:NUNAV-0187 <i>Harmothoe rarispina</i> | BIOUG<CAN>:NUNAV-0187 | | | HQ02 4360 | | HQ024 512 |
| <i>Harmothoe rarispina</i> | <i>Harmothoe rarispina</i> spm KW406 | | KY657 641 | KY657 611 | KY657 659 | KY657 624 | |
| <i>Harmothoe sp.</i> | <i>Harmothoe extenuata</i> | | XXXX | XXXX | XXXX | XXXX | |
| <i>Harmothoe sp.</i> | <i>Harmothoe sp.</i> spm G1_P1 | | XXXX | XXXX | XXXX | XXXX | |
| <i>Harmothoe sp.</i> | NSJC66-793-S001 <i>Harmothoe sp.</i> | NSJC66_793_S001 | KU738 164 | KU738 179 | | KU738 197 | |
| <i>Harmothoe sp.</i> | <i>Harmothoe sp.</i> spm SH061-066 | SH061-066 | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Harmothoe sp.</i> | <i>Harmothoe sp1</i> | | XXXX | XXXX | XXXX | XXXX | |
| <i>Malmgrenia ljunghmani</i> | <i>Malmgrenia ljunghmani</i> | | XXXX | XXXX | | XXXX | XXXX |
| <i>Malmgrenia sp.</i> | <i>Malmgrenia sp.</i> | | XXXX | | | XXXX | XXXX |
| <i>Malmgreniella mcintoshii</i> | SMNH-118976 <i>Malmgreniella mcintoshii</i> | SMNH118976 | JN852 904 | JN852 834 | JN852 935 | JN8528 66 | |
| <i>Malmgreniella nigralba</i> | BIOUG<CAN>:BAMPOL0034 <i>Malmgreniella nigralba</i> | BIOUG<CAN>:BAMPOL0034 | | | HM47 3462 | | HM473 237 |
| <i>Melaenis loveni</i> | SMNH-118977 <i>Melaenis loveni</i> | SMNH118977 | JN852 905 | JN852 835 | JN852 936 | JN8528 67 | |
| <i>Neopolynoe paradoxa</i> | SMNH-118978 <i>Neopolynoe paradoxa</i> | SMNH118978 | JN852 906 | JN852 836 | JN852 937 | JN8528 68 | |

| | | | | | | | |
|-------------------------------------|---|-------------------------|--------------|--------------|--------------|--------------|------|
| Paradyte sp. | Paradyte sp. spm GR-C1_P1 | | XXXX | XXXX | XXXX | XXXX | XXXX |
| Paradyte sp. | Paradyte sp. spm GR-D1_P1 | | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Paralepidonotus ampulliferus</i> | SMNH-118980 <i>Paralepidonotus ampulliferus</i> | SMNH118980 | JN852 908 | | JN852 939 | | |
| Pettibonesia furcosetosa | Pettibonesia furcosetosa | | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Polyeunoa laevis</i> | NSJC66-477-S001 <i>Polyeunoa laevis</i> | NSJC66_477_S001 | KU73 8160 | KU73 8176 | KU73 8212 | KU738 193 | |
| <i>Polyeunoa laevis</i> | NSJC66-962-S001 <i>Polyeunoa laevis</i> | NSJC66_962_S001 | KU73 8161 | KU73 8177 | KU73 8213 | KU738 194 | |
| Polyeunoa laevis | Polyeunoa laevis | | XXXX | | XXXX | | |
| <i>Polynoe scolopendrina</i> | SMNH-118981 <i>Polynoe scolopendrina</i> | SMNH118981 | JN852 909 | JN852 839 | JN852 940 | JN852 870 | |
| Robertianella sp. | Robertianella sp. | | XXXX | XXXX | | XXXX | XXXX |
| Subadyte pellucida | Subadyte pellucida | | XXXX | XXXX | | XXXX | XXXX |
| NA | | | | | | | |
| Gen. sp. | MNHN-IA-2015-77 Gen. sp. | MNHN-IA-2015-77 | XXXX | XXXX | XXXX | XXXX | XXXX |
| Gen. sp. | NHMUK-2014-28 Gen. sp. | NHMUK_2014_28 | KJ676 610 | | KJ676 636 | | |
| Gen. sp. | Gen. sp. spm SH062-278 | SH062-278 | XXXX | XXXX | XXXX | XXXX | |
| Gen. sp. | Gen. sp. spm MCR-1185 | | KM97 9543 | | KJ566 959 | | |
| Gen. sp. | Gen. sp. spm MCR-1288 | | KM97 9542 | | KJ566 958 | | |
| Gen. sp. | Gen. sp. spm 1E8-St95 | Polychaeta1E8_St95 | KF713 467 | | KF713 380 | | |
| Gen. sp. | 1G7-St95 Gen. sp. | Polychaeta1G7_St95 | KF713 462 | | KF713 382 | | |
| Gen. sp. | Gen. sp. spm H10-St95 | Polynoidae_H10_St95 | KF713 465 | | KF713 378 | | |
| Gen. sp. | A1_P1 Gen sp. | | XXXX | XXXX | XXXX | XXXX | XXXX |
| Gen. sp. | MNHN-IA-2013-185 Gen. sp. | MNHN-IA-2013-185 | XXXX | XXXX | | XXXX | XXXX |
| Gen. sp. | MNHN-IA-2013-120 Gen. sp. | MNHN-IA-2013-120 | XXXX | XXXX | XXXX | XXXX | XXXX |
| Gen. sp. | MNHN-IA-2013-121 Gen. sp. | MNHN-IA-2013-121 | XXXX | XXXX | XXXX | XXXX | XXXX |
| <i>Hodor anduril</i> | MNHN-IA-TYPE-1826 <i>Hodor anduril</i> | MNHN-IA-TYPE 1826 | MH23 3191 | MH23 3240 | MH23 3288 | | |
| <i>Hodor hodor</i> | MNHN-IA-TYPE-1825 <i>Hodor hodor</i> | MNHN-IA-TYPE 1825 | MH23 3189 | MH23 3238 | MH23 3257 | | |
| Podarmus sp. | Podarmus sp. | | XXXX | | | XXXX | XXXX |
| | | | | | | | |
| SIGALIONIDAE | | | | | | | |
| Fimbriosthenelais sp. | Fimbriosthenelais sp. | | XXXX | | | XXXX | XXXX |
| Mayella sp. | Mayella sp. | | XXXX | XXXX | | XXXX | XXXX |
| <i>Neoleanira tetragona</i> | SMNH-118984 <i>Neoleanira tetragona</i> | SMNH118984 | JN852 911 | | | JN852 872 | |
| <i>Neoleanira tetragona</i> | SMNH-73638 <i>Neoleanira tetragona</i> | SMNH73638 | | AY839 570 | AY839 582 | | |
| <i>Psammolyce sp.</i> | SMNH118989 <i>Psammolyce sp.</i> | SMNH118989 | JN852 916 | JN852 843 | | JN852 877 | |
| <i>Sigalion spinosus</i> | <i>Sigalion spinosus</i> | | | AY894 304 | AY894 319 | | |
| <i>Sthenelais boa</i> | SMNH-73636 <i>Sthenelais boa</i> | SMNH73636 | DQ77 9635 | AY839 575 | | DQ779 711 | |
| <i>Sthenelanelia uniformis</i> | <i>Sthenelanelia uniformis</i> | | | AY894 306 | AY894 322 | | |

Annexe 2 Information sur le génome mitochondrial de *Segonzacia mesatlantica*. Le type de gène (CDS : séquence codante ; tRNA ; RC : Région de contrôle) ; début et fin du gène ; Lg : longueur du gène ; brin à partir duquel est transcrit la séquence, + brin lourd / - brin léger) ; séquences de l'anticodon et du codon associé dans le cas des ARNt, NI : nombre de nucléotides intergéniques non codants ; codons start et stop dans le cas de gènes codant pour des protéines ; pourcentage de GC. Les gènes sont rangés selon leur ordre dans le mitogénome.

| Gène | Type | Début | Fin | Lg | Sens | Anticodon | Codon associé | NI | start | stop | % GC |
|-----------------|------|--------|--------|------|------|-----------|---------------|----|-------|------|------|
| COX1 | CDS | 1 | 1 534 | 1534 | + | | | 0 | ATG | T- | 38.1 |
| tRNA Leu | tRNA | 1 535 | 1 599 | 65 | + | UAA | UUA | 0 | | | 29.2 |
| COX2 | CDS | 1 620 | 2 303 | 684 | + | | | 20 | ATG | TAA | 35.8 |
| tRNA-Lys | tRNA | 2 309 | 2 376 | 68 | + | UUU | AAA | 5 | | | 39.7 |
| tRNA-Asp | tRNA | 2 376 | 2 438 | 63 | + | GUC | GAC | -1 | | | 25.4 |
| ATP8 | CDS | 2 439 | 2 597 | 159 | + | | | 0 | GTG | TAG | 31.4 |
| ATP6 | CDS | 2 591 | 3 265 | 675 | + | | | -7 | ATT | TAA | 34.5 |
| COX3 | CDS | 3 265 | 4 054 | 790 | + | | | -1 | ATG | T- | 36.6 |
| tRNA-Gly | tRNA | 4 055 | 4 118 | 64 | + | UCC | GGA | 0 | | | 28.1 |
| NAD3 | CDS | 4 119 | 4 472 | 354 | + | | | 0 | ATC | TAA | 33.3 |
| tRNA-Ala | tRNA | 4 475 | 4 538 | 64 | + | UGC | GCA | 2 | | | 32.8 |
| tRNA-Arg | tRNA | 4 539 | 4 603 | 65 | + | UCG | CGA | 0 | | | 35.4 |
| tRNA-Asn | tRNA | 4 604 | 4 670 | 67 | + | GUU | AAC | 0 | | | 28.4 |
| tRNA-Ser | tRNA | 4 675 | 4 740 | 66 | + | UCU | AGA | 4 | | | 30.3 |
| tRNA-Glu | tRNA | 4 746 | 4 810 | 65 | + | UUC | GAA | 5 | | | 26.2 |
| tRNA-His | tRNA | 4 828 | 4 891 | 64 | - | GUG | CAC | 17 | | | 31.1 |
| tRNA-Phe | tRNA | 4 895 | 4 958 | 64 | - | GAA | UUC | 3 | | | 23.4 |
| NAD5 | CDS | 4 958 | 6 685 | 1728 | - | | | -1 | ATG | TAA | 32.5 |
| NAD4 | CDS | 6 733 | 8 067 | 1335 | - | | | 47 | GTG | TAG | 30.3 |
| NAD4L | CDS | 8 061 | 8 363 | 303 | - | | | -7 | ATG | TAA | 32.0 |
| tRNA-Thr | tRNA | 8 366 | 8 429 | 64 | + | UGU | ACA | 2 | | | 21.9 |
| tRNA-Pro | tRNA | 8 430 | 8 495 | 66 | - | UGG | CCA | 0 | | | 27.3 |
| NAD6 | CDS | 8 498 | 9 004 | 507 | + | | | 2 | ATC | TAA | 32.5 |
| CytB | CDS | 9 004 | 10 138 | 1135 | + | | | -1 | ATG | T- | 36.6 |
| tRNA-Ser | tRNA | 10139 | 10 203 | 65 | + | UGA | UCA | 0 | | | 16.9 |
| NAD1 | CDS | 10 210 | 11 148 | 939 | - | | | 6 | ATG | TAA | 31.3 |
| tRNA-Leu | tRNA | 11 175 | 11 242 | 68 | - | UAG | CUA | 26 | | | 27.9 |
| 16S rRNA | rRNA | 11 247 | 12 589 | 1343 | - | | | 4 | | | 26.7 |
| tRNA-Val | tRNA | 12 589 | 12 661 | 73 | - | UAC | GUA | -1 | | | 32.9 |
| 12S rRNA | rRNA | 12 661 | 13 492 | 832 | - | | | -1 | | | 28.6 |
| Région contrôle | RC | 13 493 | 14 116 | 624 | | | | 0 | | | 21.5 |
| tRNA-Ile | tRNA | 14 117 | 14 184 | 68 | + | GAU | AUC | 0 | | | 30.9 |
| tRNA-Gln | tRNA | 14 182 | 14 249 | 68 | - | UUG | CAA | -3 | | | 36.8 |
| tRNA-Met | tRNA | 14 250 | 14 318 | 69 | + | CAU | AUG | 0 | | | 30.4 |
| ND2 | CDS | 14 319 | 15 323 | 1005 | + | | | 0 | ATG | TAG | 34.1 |
| tRNA-Trp | tRNA | 15 322 | 15 389 | 68 | + | UCA | UGA | -2 | | | 25.0 |
| tRNA-Cys | tRNA | 15 389 | 15 455 | 67 | - | GCA | UGC | -1 | | | 28.4 |
| tRNA-Tyr | tRNA | 15 456 | 15 521 | 66 | - | GUA | UAC | 0 | | | 30.3 |

Annexe 3 Données de séquençage, sélection des données brutes et données aux différentes étapes de l'assemblage et du traitement post assemblage. Nombre de reads brutes de séquençage ; pourcentage de ces reads ayant été sélectionnés après Trimmomatic ; nombre de contigs et le nombre total de bases assemblés pour chacune des trois étapes de l'assemblage : (A) l'assemblage primaire, (B) l'assemblage filtré et (C) après le traitement post-assemblage. Les données (B) sont manquantes dans le cas des données récupérées sous forme déjà assemblées. Les trois espèces grisées forment le groupe externe. Les espèces avec un astérisque proviennent de nos données de séquençage.

| Espèce | Nombre read brut | % après Trim. | Nombre contigs | | | Nombre total de base | | |
|--|------------------|---------------|----------------|---------|---------|----------------------|-------------|------------|
| | | | A | B | C | A | B | C |
| <i>Antecaridina lauensis</i> | 58 974 040 | | 112 329 | | 77 964 | 91 369 515 | | 27 163 339 |
| <i>Athanas nitescens</i> * | 98 420 753 | 91,4 | 160 618 | 65 810 | 57 336 | 100 969 682 | 52 117 975 | 22 113 415 |
| <i>Atyopsis moluccensis</i> * | 113 733 324 | 91,0 | 113 631 | 24 134 | 19 830 | 101 398 966 | 40 659 328 | 15 345 631 |
| <i>Caridina multidentata</i> | 16 646 448 | 90,8 | 78 390 | 30 313 | 29 199 | 46 289 537 | 34 773 891 | 19 250 981 |
| <i>Caridina rubella</i> | 41 610 710 | | 20 216 | | 15 514 | 13 046 436 | | 6 355 949 |
| <i>Caridion steveni</i> * | 91 236 366 | 88,3 | 117 058 | 88 910 | 77 406 | 111 042 140 | 67 472 062 | 28 954 238 |
| <i>Chorismus antarcticus</i> | 102 119 756 | 98,8 | 275 284 | | 239 512 | 277 625 659 | | 98 792 932 |
| <i>Crangon crangon</i> * | 103 495 285 | 87,6 | 110 312 | 35 529 | 30 705 | 87 346 439 | 43 469 280 | 18 098 963 |
| <i>Halocaridina rubra</i> | 8 603 841 | 100,0 | 68 295 | | 50 843 | 75 503 607 | | 26 855 029 |
| <i>Halocaridinides trigonophthalma</i> | 56 910 600 | | 123 892 | | 82 994 | 106 142 218 | | 32 541 837 |
| <i>Heterocarpus aff. dorsalis</i> * | 113 836 022 | 87,6 | 157 538 | 34 013 | 29 629 | 114 056 102 | 42 219 455 | 15 295 669 |
| <i>Macrobrachium nipponense</i> | 175 130 000 | | 58 871 | | 50 635 | 60 535 022 | | 29 779 952 |
| <i>Macrobrachium olfersi</i> | 25 636 097 | 93,6 | 88 388 | 71 460 | 66 577 | 74 583 766 | 64 681 436 | 28 960 431 |
| <i>Macrobrachium rosenbergii</i> | 41 121 163 | 72,3 | 111 237 | 104 037 | 98 088 | 71 051 852 | 67 014 122 | 37 433 111 |
| <i>Metabetaeus lohena</i> | 57240274 | | 161 396 | | 116 143 | 115 006 916 | | 38 201 351 |
| <i>Metabetaeus minutus</i> | 56 501 920 | | 40 067 | | 29 081 | 21 764 267 | | 10 179 283 |
| <i>Neocaridina denticulata</i> | 29 987 459 | 86,7 | 115 304 | 91 061 | 82 477 | 101 127 207 | 87 107 748 | 41 905 773 |
| <i>Oplophorus gracilirostris</i> * | 110 179 501 | 91,5 | 136 720 | 28 143 | 25 100 | 79 263 397 | 28 057 848 | 12 782 369 |
| <i>Palaemon carinicauda</i> | 54 207 796 | 95,6 | 101 843 | 80 237 | 73 046 | 64 192 321 | 54 008 940 | 33 415 361 |
| <i>Palaemon serratus</i> | 38 867 738 | 79,4 | 87 586 | 66 459 | 62 435 | 81 040 231 | 67 227 731 | 32 453 616 |
| <i>Palaemon argentinus</i> | 50 625 940 | 90,3 | 23 732 | 18 958 | 18 288 | 11 957 949 | 10 303 964 | 6 149 146 |
| <i>Pandalus latirostris</i> | 58409677 | 69,7 | 83 056 | 75 728 | 70 590 | 53 766 863 | 49 684 983 | 26 499 815 |
| <i>Paratya australiensis</i> | 24 518 416 | 90,8 | 94 454 | 82 216 | 76 890 | 79 733 736 | 71 646 845 | 33 380 574 |
| <i>Periclimenes brevicarpalis</i> * | 106 475 923 | 88,2 | 265 779 | 70 618 | 66 241 | 129 139 417 | 52 347 904 | 22 090 732 |
| <i>Rimicaris exoculata</i> * | 104 041 410 | 91,6 | 156 894 | 42 904 | 37 798 | 102 198 445 | 41 803 179 | 17 438 244 |
| <i>Rimicaris sp.</i> | 22 964 656 | 91,7 | 233 735 | 163 103 | 154 696 | 136 552 117 | 100 371 653 | 58 651 901 |
| <i>Hyas araneus</i> | 98 508 658 | - | 19 199 | | 16 800 | 13 549 811 | | 7 060 052 |
| <i>Penaeus vannamei</i> | 400 228 040 | 99,7 | 110 474 | | 97 286 | 125 657 935 | | 56 992 557 |
| <i>Petrolisthes cinctipes</i> | - | - | 97 806 | | 33 144 | 59 220 567 | | 11 771 148 |

Annexe 4 Caractéristiques métriques N50, longueur médiane et longueur moyenne des assemblages pour les espèces utilisées lors de cette étude, pour chacun des trois étapes de l'assemblage : (A) l'assemblage primaire, (B) l'assemblage filtré et (C) après le traitement post-assemblage. Les données (B) sont manquantes dans le cas des données récupérées sous forme déjà assemblées. Les trois espèces grisées forment le groupe externe. Les espèces avec un astérisque proviennent de nos données de séquençage.

| Espèce | N50 | | | Med | | | Moy | | |
|--|-------|-------|-------|-----|---------|-------|---------|---------|-------|
| | A | B | C | A | B | C | A | B | C |
| <i>Antecaridina lauensis</i> | 1 425 | | 417 | 411 | | 215,0 | 813,4 | | 348,4 |
| <i>Athanas nitescens</i> * | 936 | 1 284 | 597 | 365 | 464,0 | 216,0 | 628,6 | 792 | 385,7 |
| <i>Atyopsis moluccensis</i> * | 1 859 | 2 824 | 1 599 | 395 | 1 141,0 | 318,0 | 892,4 | 1 684,7 | 773,9 |
| <i>Caridina multidentata</i> | 1 318 | 1 998 | 1 391 | 332 | 655,0 | 270,0 | 706,4 | 1 147,2 | 659,3 |
| <i>Caridina rubella</i> | 937 | | 504 | 361 | | 252 | 645,4 | | 409,7 |
| <i>Caridion steveni</i> * | 996 | 1 222 | 555 | 374 | 438 | 213 | 652,5 | 758,9 | 374,1 |
| <i>Chorismus antarcticus</i> | 2 260 | | 804 | 418 | | 207 | 1 008,5 | | 412,5 |
| <i>Crangon crangon</i> * | 1 457 | 2 076 | 1 359 | 394 | 765 | 249 | 791,8 | 1 223,5 | 589,4 |
| <i>Halocaridina rubra</i> | 2 209 | | 1 113 | 511 | | 238 | 1 105,6 | | 528,2 |
| <i>Halocaridinides trigonophthalma</i> | 1 534 | | 572 | 452 | | 222 | 856,7 | | 392,1 |
| <i>Heterocarpus aff. dorsalis</i> * | 1 286 | 2 116 | 1 083 | 366 | 787 | 249 | 724 | 1 241,3 | 516,2 |
| <i>Macrobrachium nipponense</i> | 1 294 | | 1 002 | 675 | | 306 | 1 028 | | 588,1 |
| <i>Macrobrachium olfersi</i> | 1 673 | 1 816 | 792 | 396 | 417 | 225 | 843,8 | 905,1 | 435 |
| <i>Macrobrachium rosenbergii</i> | 980 | 984 | 558 | 352 | 358 | 214 | 638,7 | 644,1 | 381,6 |
| <i>Metabetaeus lohena</i> | 1 223 | | 407 | 363 | | 198 | 712,6 | | 328,9 |
| <i>Metabetaeus minutus</i> | 700 | | 396 | 321 | | 240 | 543,2 | | 350 |
| <i>Neocaridina denticulata</i> | 1 727 | 1 836 | 996 | 412 | 466 | 246 | 877,1 | 956,6 | 508,1 |
| <i>Oplophorus gracilirostris</i> * | 830 | 1 644 | 1 047 | 328 | 607 | 240 | 579,8 | 997 | 509,3 |
| <i>Palaemon carinicauda</i> | 890 | 979 | 771 | 380 | 408 | 249 | 630,3 | 673,1 | 457,5 |
| <i>Palaemon serratus</i> | 1 916 | 2 097 | 1 121 | 416 | 448 | 238 | 925,3 | 1 011,6 | 519,8 |
| <i>Palaemon argentinus</i> | 617 | 693 | 414 | 342 | 373 | 220 | 503,9 | 543,5 | 336,2 |
| <i>Pandalus latirostris</i> | 1 011 | 1 018 | 503 | 362 | 371 | 222 | 647,4 | 656,1 | 375,4 |
| <i>Paratya australiensis</i> | 1 691 | 1 730 | 802 | 389 | 404 | 222 | 844,2 | 871,5 | 434 |
| <i>Periclimenes brevicarpalis</i> * | 559 | 1 228 | 417 | 306 | 415 | 192 | 485,9 | 741,3 | 333,5 |
| <i>Rimicaris exoculata</i> * | 1 035 | 1 653 | 963 | 353 | 565 | 225 | 651,4 | 974,3 | 461,4 |
| <i>Rimicaris sp.</i> | 776 | 872 | 465 | 348 | 352 | 240 | 584,2 | 615,4 | 379,1 |
| <i>Hyas araneus</i> | 730 | | 510 | 593 | | 327 | 705,8 | | 420,2 |
| <i>Penaeus vannamei</i> | 2 701 | | 1 440 | 429 | | 246 | 1 137,4 | | 585,8 |
| <i>Petrolisthes cinctipes</i> | 695 | | 425 | 674 | | 261 | 605,5 | | 355,2 |

Résumé :

De par leur originalité, certains organismes de la faune hydrothermale ont été classés dans des nouveaux groupes taxonomiques de haut rang. De précédentes études moléculaires ont permis d'en réassigner certains dans des lignées connues, occasionnant parfois d'importantes réductions de rangs taxonomiques. Par ailleurs, en phylogénie moléculaire, il est difficile d'optimiser à la fois l'échantillonnage taxonomique et le nombre de marqueurs. Cette thèse illustre cette limitation, mais fournit des avancées sur la compréhension de l'origine et l'évolution de trois groupes hydrothermaux. Chez les vers Polynoidae, l'approche multigène appliquée sur un grand nombre d'espèces suggère au moins deux événements de colonisation de ce milieu. Toutefois, le manque de résolution des marqueurs pour les nœuds profonds entrave la compréhension de l'histoire de ces colonisations. Une limitation similaire avait empêché de replacer les familles de crevettes Alvinocarididae et de crabes Bythograeidae dans leur infra-ordre respectifs (Caridea et Brachyura). Deux approches de recherche et d'identification de marqueurs sont donc testées pour ces groupes. La première, basée sur le séquençage de génome mitochondrial (facilement généralisable), résout les relations profondes des Brachyura et place les espèces de Bythograeidae disponibles proches des Xanthidae. La seconde, basée sur le séquençage de transcriptomes permet d'identifier des marqueurs suffisamment conservés pour résoudre les relations inter-familles chez les Caridea. Cette approche est moins généralisable, mais les marqueurs identifiés pourront a posteriori être recherchés dans un échantillonnage large par capture à l'aide de sonde.

Mots-clés : Phylogénie ; Sources Hydrothermales ; NGS ; Sanger ; Polynoidae ; Alvinocarididae ; Bythograeidae

Summary:

The originality of the hydrothermal vents fauna led to the classification of some organisms under new high taxonomic ranks. However, previous molecular studies reassigned them to known lineages, leading to major reductions in such ranking. Classically in phylogenetic studies, optimizing both taxonomic sampling and molecular markers is challenging. This Ph.D project illustrates this limitation, but still provides breakthroughs in the understanding of the origin and evolution of three hydrothermal taxa. In Polynoidae worms, the multigenic approach, led on a large taxonomic and ecological sampling, indicates at least two colonization events of hydrothermal vents. However, the limited resolution of these markers for deep nodes prevented the clear understanding of such events. A similar limitation was previously encountered for Alvinocarididae shrimp and Bythograeidae crabs families in their respective infra-orders (Caridea and Brachyura). Here, two approaches aiming to search and identify markers were tested on these groups. The first one, based on the sequencing of the mitochondrial genome (easily generalizable), resolves deep nodes in Brachyura, and places the available Bythograeidae species near the Xanthidae. The second, based on transcriptome sequencing, allows the identification of molecular markers conserved enough to resolve inter-familial relationships in Caridea. Although this approach is less generalizable, the identified markers could be targeted *a posteriori* on a wide taxonomic scale using marker-specific probes.

Key words: Phylogeny; Hydrothermal vents; NGS; Sanger sequencing; Polynoidae; Alvinocarididae; Bythograeidae.