



HAL
open science

Etude multi-locus des pressions évolutives s'exerçant sur les bivalves marins en situation d'hybridation : cas des moules hydrothermales du genre *Bathymodiolus* des dorsales médio-Atlantique et est-Pacifique

Jean-Baptiste Faure

► To cite this version:

Jean-Baptiste Faure. Etude multi-locus des pressions évolutives s'exerçant sur les bivalves marins en situation d'hybridation : cas des moules hydrothermales du genre *Bathymodiolus* des dorsales médio-Atlantique et est-Pacifique. Ecosystèmes. Montpellier II, 2008. Français. NNT : . tel-01112819

HAL Id: tel-01112819

<https://hal.sorbonne-universite.fr/tel-01112819>

Submitted on 3 Feb 2015

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.

Avertissement

Au vu de la législation sur les droits d'auteur, ce travail de thèse demeure la propriété de son auteur, et toute reproduction de cette oeuvre doit faire l'objet d'une autorisation de l'auteur. (cf Loi n°92-597; 1/07/1992. Journal Officiel, 2/07/1992)

T H E S E

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE MONTPELLIER II

Discipline :

Biologie des Populations et Ecologie

Ecole Doctorale :

Systèmes Intégrés en Biologie, Agronomie, Géoscience, Hydrosience, Environnement

présentée et soutenue publiquement

M. Baptiste FAURE

Le 19 septembre 2008

Etude multi-locus des pressions évolutives s'exerçant sur des bivalves marins en situation d'hybridation : cas des moules hydrothermales du genre *Bathymodiolus* des dorsales médio-atlantique et est-pacifique.

Devant le jury composé de :

Pr. Xavier VEKEMANS, Univ. Lille I

Rapporteur ; Président

Pr. Grant POGSON, Univ. Calif. Santa Cruz

Rapporteur

Dr. François BONHOMME, CNRS - Univ. Montpellier II

Directeur de Thèse

Dr. Didier JOLLIVET, CNRS - Station Biologique de Roscoff

Examineur

Dr. Nicolas BIERNE, CNRS - Univ. Montpellier II

Examineur

Dr. Pierre BOUDRY, Ifremer

Examineur

REMERCIEMENTS

Une thèse est un travail de longue haleine dont le manuscrit final ne représente, en réalité, que la partie émergée de trois longues années (...et quelques mois).

La réussite dans une telle épreuve (initiatique ?) est liée, en partie, à un peu de travail personnel (eh oui, il en faut quand même) mais aussi à la présence, à la sympathie et au soutien d'un grand nombre de personnes que je tiens à remercier aujourd'hui.

Dans un premier temps, je souhaite remercier mes deux rapporteurs : Xavier Vekemans venu de Lille et Grant Pogson, de Santa Cruz. Merci de l'intérêt que vous avez porté à mon travail, pour vos remarques et vos félicitations. Merci aussi d'être venus d'aussi loin.

Il faut commencer par préciser que ma thèse présentait une particularité géographique, elle s'est en effet déroulée dans des centres de recherches aux caractéristiques régionales très contrastées : la station méditerranéenne de l'environnement littoral à Sète et la station biologique de Roscoff. Un contexte idéal pour un bon co-encadrement... mais aussi pour développer un certain sens de la diplomatie... (c'est aussi ça une thèse !)

Je souhaite adresser un grand remerciement à mon directeur de thèse François Bonhomme, qui m'a accordé sa confiance pour ce travail. Merci à toi François pour le temps que tu as réussi à me consacrer lors de mes déplacements à Sète et ce, malgré ton emploi du temps plus que chargé. J'ai découvert, grâce à toi, que des discussions formelles pouvaient avoir lieu aussi bien dans un bureau qu'autour d'un verre, dans une voiture ou dans une crêperie ! Merci de m'avoir permis de découvrir ce magnifique village qu'est Montbazin et ses nombreuses spécialités : les barbecues, les travaux forcés dans les vignes (et les coups de soleil associés), le Pradinel dont la renommée est désormais nationale, le mystère de la reproduction des cacahouètes (parce que vous savez, vous, comment se reproduisent les cacahouètes ??) et la piscine très agréable (il faut cependant noter que même si les chats savent nager, ce n'est pas une très bonne idée de les mettre dans une piscine gonflable... sous peine d'utiliser des rustines pendant quelques jours). Merci aussi d'avoir partagé l'unique plongée que j'ai eu l'occasion de faire à Sète ! Merci aussi à Juliette pour sa gentillesse et ses tartes aux prunes et grosses bises à Alice.

Merci aussi à mon directeur breton Didier Jollivet. Merci pour ton aide, ton soutien et merci d'avoir supporté ma présence, chaque jour, dans le bureau. Le temps que tu consacres à l'encadrement et aux discussions scientifiques (quelle

que soit l'heure du jour... ou de la nuit !) font de toi un vrai directeur de thèse et je vais te poser cette question récurrente : à quand l'HDR ??!! Il serait trop long de faire ici la liste de toutes les choses que tu m'as apportées, mais je peux citer, entre autre, la rigueur scientifique, l'initiation aux techniques moléculaires, la découverte du milieu hydrothermal (si passionnant !), la plongée en Nautille lors d'une petite croisière assez sympathique, tes vinyles inclassables et les *baseball furious*. Merci donc de m'avoir consacré autant de temps et de m'avoir initié à la pêche aux ormeaux ! Te payer des cafés reste et restera toujours un plaisir. C'est agréable de pouvoir faire des pauses avec des gens qui ne parlent pas sans cesse de boulot ! J'ai aussi une pensée pour Sylvie et tes enfants ; je leur suis reconnaissant de t'avoir laissé libre de m'aider dans les moments un peu tendus (quelques nuits blanches et des échanges de mails très tardifs) tout ça parce qu'avec ma super organisation, je n'étais jamais capable de tenir les délais...

Merci aussi à Pierre Boudry alias Mr huître. Merci pour ta bonne humeur et les discussions sincères que nous avons pu avoir. Désolé cependant que le volet « huître » soit aussi peu présent dans le manuscrit final.

Je souhaite tout particulièrement remercier Myriam de m'avoir accueilli au sein de l'équipe EGPM et de m'avoir permis de partager un temps son bureau. Merci aussi aux autres EGPM-iens : Fred, Claire, Christophe, Arnaud, Denis et Dominique.

Une mention particulière pour Arnaud qui a souvent été mon colocataire lors des congrès : Sète, San Diego, Barcelone. Ce fut un plaisir de partager ces moments et ces verres de Manzana ! Merci pour tes conseils nombreux et avisés concernant les manips (tes deux mots d'ordre : efficacité et rentabilité) et merci aussi d'être toujours prêt à filer un petit coup de pattes lorsque l'on en a besoin. Une chose est sûre, quand tu es présent dans une pièce, on ne risque pas de s'ennuyer.

Merci à mes collègues thésards de l'équipe EGPM avec qui il était particulièrement plaisant de pouvoir batifoler aux quatre coins de la Bretagne (et par tous les temps) pour aller chercher des ophiures ou des pectinaires, compter et mesurer des *Undaria* ou des *Fucus*. Merci donc à vous Delphine, Taimour, Marie et Manue.

Merci aux différentes générations de thésards de l'équipe : Patrice, Taimour, Lise, Manue, Marie, Mèl (eh oui on pense encore à toi !), Delphine, Sabrina, Frantz, Sophie, Valeria, Daphné, Joao, Antonio, Leïla, Florence, Benoit, Stacy.

Merci aussi aux nombreux amis, collègues ou stagiaires passés dans l'équipe et désolé pour ceux qui seraient oubliés, croyez bien que ce n'est pas intentionnel :

Damien, Solenn, Dorothee, Hélène, Cédric, Elodie, Anne-Marie, Marie, Marion, Mathieu, Florentine, Cécile, Thomas, Nicolas T. (2 fois !)...

Merci aussi aux autres amis stationnaires avec qui j'ai partagé cafés et soirées, le clan des benthologues et tous les autres : Marjo, Caro, Gauthier & Alice, Sakina, Vincent, Claire, Régis, Céline, Guillaume, Aurélie, Damien, Fred, Natalie, PO, Andrès...

Merci aussi à l'AJC, ses soirées et ses t-shirts.

Une pensée aussi pour mes expatriées préférées : Marie-Laure, toujours prête à envoyer des cartes postales vraiment moches dès que l'occasion se présente (c'est toujours un plaisir de constater le talent que tu as pour dénicher tes cartes). Merci aussi pour les bouteilles de pisco, pour les magnifiques mouffles mais surtout pour ces magnifiques petits chaussons chiliens qui sentent encore le mouton !! La deuxième expatriée est Delphine (mais est-ce vraiment une expatriation la Réunion ??), merci donc pour tes commentaires sur les plongées tropicales, c'est toujours sympa de savoir que tu passes tes WE à plonger dans des eaux chaudes avec des dauphins et des tortues, quand je passe les miens dans une eau à 12°C en compagnie de quelques lieus et des oursins ! Merci aussi de savoir revenir pour les grands événements, serais-tu attirée par le bruit des fines bulles dans les verres ?

Merci à tous les anciens « Locataires YD », pour les quelques mois passés ensemble et qui resteront inoubliables. Même si pour la plupart vous êtes désormais devenus Rennais (Matt' et la GO Laeti'), bordelais (Tiph), ou pire... parisiens (Aurel', Ludo, Nico, Julie), heureusement que certains sont restés faire de la résistance dans notre irréductible village breton de Roscoff (Manon et Damien). Pour vous tous, je n'aurai que quelques mots : SBR power forever !

Et que dire de la colocation dans notre Maison du Bonheur à Taulé ? Comment cette maison, qui a vu passer autant de monde, est-elle encore debout ?? Alors merci à tous les gens qui y ont résidé pour des périodes plus ou moins longues et avec lesquels j'ai partagé tant de choses : les apéros ou les créations culinaires dignes des plus grands chefs, le mur d'escalade dans le garage (la classe !), les sacs de Silicagel devant la cheminée, les soirées Wanted interminables ou les soirées UNO (et le lave vaisselle !). Par ordre d'apparition, merci donc à Nico, Ludo, Manon, Laetitia, Camille, Marie, Sonny, Julie, Alexi, PO, Elodie, sans compter les nombreux invités toujours les bienvenus lors des soirées ou des vacances, les chats et les gerbilles !! Merci aussi à la poste, à la gentille factrice et à la voisine qui pensait à nous sortir les poubelles chaque semaine.

Mais les soirées bretonnes ne se résument pas uniquement à des soirées entre stationnaires ! Alors merci aussi à Mel & Yann, Roland & Sophie, et nos anciens voisins Plouzévédéens (dommage que cette période soit passée aussi vite) : Ronan et Agnès.

Je n'oublie pas le GPC et son CJP. Merci donc à tous les plongeurs de Carantec, Gilles, Marie, Manue, Delphine, Yann, Lise, Vincent, Aurélie, Sylvie, ainsi que tous les autres. C'est toujours un plaisir de vous retrouver le WE à Châteaulin ou pour partager des plongées en Baie de Morlaix où l'eau est souvent sombre et froide ! Les membres du groupe d'œnologie : Vincent, Gérard, Cédric, Roland et tous les autres. Quel plaisir de pouvoir déguster du vin le vendredi ! Et que dire des WE à Douarnenez... Mémorables !

Bravo à toute l'équipe du Gulf Stream pour les repas dignes d'un vrai resto tous les midis... et quel plaisir de pouvoir profiter du soleil en prenant son café sur le bord de la piscine !

Assez parlé des Bretons, il faut maintenant parler des gens du sud :

Merci à Nicolas Bierne, pilier indispensable de la station de Sète. Merci d'avoir toujours été présent, de m'avoir fait découvrir le monde magique et inquiétant de la génétique théorique et des logiciels récalcitrants. Merci aussi d'avoir été à mes côtés lors de la dernière ligne droite. C'est aussi grâce à toi que j'ai vu mes premiers flamands roses ! Merci pour les barbecues, la gestion de la logistique du midi : les commandes de kebab ou les résas dans les paillottes. Merci aussi de nous avoir prêté ton super appart (climatisé ☺), de nous avoir laissé le cochon d'Inde et tes bouteilles de whisky. Des bises aussi à Sandrine et à Tilia.

Un énorme merci à Matt', cousin sudiste dont l'accent ne trompe pas. Nous avons partagé un directeur, des sujets de thèse très proches, un bureau, l'organisation du PhD meeting et de très nombreuses autres soirées. Alors, comment te remercier de m'avoir fait découvrir le Lou Biou, les meilleurs bars et les paillottes de Sète ? J'espère que nous aurons l'occasion de nous revoir de nombreuses fois malgré la distance.

Merci à Sylvie pour m'avoir fait découvrir le festival d'Avignon, son succulent tajine et son jeu très énervant : les anagrammes !

Merci aussi à Stefano, Emilie, Yolande, Cathy, Jean-François et Michel dont la présence a participé à la bonne humeur toujours rencontrée lors de mes séjours à la SMEL.

Et Karin, toujours dans les bons plans ! De Sète à Bordeaux, en passant par Roscoff... et bien sûr à Marseille ! Merci à toi, à Jean-Bapt, à Pierre et à la clique des thésards sympas d'Endoume. Bon courage dans votre lutte contre le côté obscur de la station.

Merci à mes amis de longue date : Yann et Natacha, Nico et son Elo, Fred et Audrey.

Merci aussi à ma famille et en particulier mes parents, mon frère, Nicolas, mon cousin, Romain, mes grands-parents et tous les autres qui m'ont toujours soutenu, aidé et encouragé.

Un dernier et énorme merci à Elodie, qui a réussi à supporter mes horaires chargés et mes longues périodes d'absence lors de la rédaction. Merci de m'avoir encouragé pendant les derniers mois difficiles de la thèse. Je t'embrasse très fort. Merci à toi, Apolline d'avoir attendu sagement 9 mois, la fin de la thèse et la soutenance, avant de pointer le bout de ton nez.

A ma grand-mère

Sommaire

Introduction	- 1 -
1. Notion d'espèce et spéciation.....	- 1 -
2. Mécanismes d'isolement reproductif	- 1 -
3. Modes de Spéciation	- 5 -
4. Théorie de la coalescence.....	- 7 -
5. Comportement des gènes en condition d'hybridation :	- 9 -
6. Intérêt des bivalves marins pour la génétique	- 11 -
7. Les environnements marins réduits.....	- 14 -
8. Un milieu contrasté, riche en microhabitats et potentiellement favorable à la sélection diversifiante.....	- 19 -
9. Le peuplement associé aux sources hydrothermales.....	- 21 -
9.1. Les polychètes	- 24 -
9.2. Les bivalves.....	- 25 -
9.3. Le milieu hydrothermal: un milieu fragmenté et instable.....	- 29 -
10. Dispersion larvaire dans le milieu hydrothermal.	- 30 -
11. Zones d'étude et objectifs de la thèse.....	- 35 -
11.1. La dorsale est pacifique.....	- 35 -
11.2. La dorsale Atlantique	- 39 -
12. Objectifs de la thèse	- 42 -
13. Déroulement du manuscrit	- 42 -
Chapitre I. Matériel et Méthodes Générales	- 45 -
I.1. Stratégie d'échantillonnage.....	- 45 -
I.1.1. Préparation des échantillons.	- 46 -
I.1.2. Choix des marqueurs génétiques.	- 47 -
I.1.3. Informations sur les gènes étudiés	- 49 -
I.2. Amplification, marquage/recapture, clonage et séquençage.....	- 52 -
I.2.1. Amplification des marqueurs	- 54 -
I.2.2. Traitement des séquences avant analyse.....	- 57 -
I.2.3. Alignement et correction des séquences	- 57 -
I.2.4. Correction des jeux de données	- 58 -
I.3. Méthodes d'analyse	- 59 -
I.3.1. Méthodes utilisées dans la réalisation des arbres phylogénétiques	- 59 -
I.3.2. Indices de diversité	- 60 -
I.4. Tests moléculaires.....	- 61 -
I.4.1. Tests de neutralité	- 61 -
I.4.2. Mesures de différenciation génétique entre populations	- 63 -
I.5. Quelques questions techniques	- 63 -
I.5.1. Estimation et caractérisation des singletons	- 63 -
I.5.2. Première étude méthodologique	- 65 -
I.5.3. Deuxième étude : basée sur les attendus théoriques entre mutations synonymes et non-synonymes.....	- 66 -
I.6. Mise en évidence de gènes paralogues au niveau des arbres étudiés.....	- 70 -
I.6.1. Un exemple simple : le cas du GF1.	- 70 -
I.6.2. Deux exemples compliqués : le cas du lysozyme et du collagène.....	- 72 -

Chapitre II : Phylogénie des Bathymodiolinae.....	- 75 -
II.1. Introduction	- 75 -
II.2. Matériels et méthodes.....	- 77 -
II.2.1. Choix des marqueurs	- 77 -
II.2.2. Acquisition des séquences	- 78 -
II.2.3. Traitement des séquences	- 81 -
II.2.4. Tests de saturation sur les deux locus.....	- 82 -
II.2.5. Définition des standards de notation des noms d'espèce	- 83 -
II.2.6. Analyse sous PAML.....	- 83 -
II.3. Résultats.....	- 84 -
II.3.1. Reconstructions phylogénétiques	- 87 -
II.3.2. Phylogénie avec le marqueur COI.....	- 87 -
II.3.3. Phylogénie avec le marqueur ITS2.....	- 93 -
II.3.4. Analyse bayésienne et maximum de vraisemblance sur les données du COI. .	- 95 -
II.4. Discussion.....	- 97 -
II.4.1. Diversité génétique chez les Bathymodiolinae : relations avec l'habitat (profondeur, symbiose et milieux)	- 97 -
II.4.2. Causes probables à la diversité des Mytilidae des environnements réduits	- 99 -
II.4.3. Evolution du gène mitochondrial COI chez les Bathymodiolinae : évidence d'une convergence adaptative associée à la symbiose?	- 102 -
II.4.4. Histoire évolutive des Bathymodiolinae et colonisation des dorsales océaniques....	- 104 -
II.4.5. Evolution et histoire régionales des espèces de <i>Bathymodiolinae</i>	- 106 -
II.4.6. Polyphylie au niveau du marqueur ITS2 et possibilités d'hybridation chez certaines espèces	- 109 -
II.5. Conclusion	- 110 -
 Chapitre III. Contact secondaire en milieu profond : analyse multi-locus de la divergence et des flux géniques entre deux espèces proches de <i>Bathymodiolus</i>.	- 111 -
III.1. Résumé	- 111 -
III.2. Annexe du chapitre III: présentation de l'article près à être soumis: "Secondary contact zone in the deep sea: a multi-locus analysis of divergence and gene flow between two closely-related species of <i>Bathymodiolus</i> "	- 114 -
III.2.1. Introduction / introduction	- 114 -
III.2.2. Matériel et méthode / Materiel and methods.....	- 116 -
III.2.3. Résultats / Results	- 120 -
III.2.4. Discussion / Discussion.....	- 132 -
III.2.5. Conclusion / conclusion	- 135 -
Remerciements / Acknowledgments	- 135 -
 Chapitre IV. Structure génétique des populations de <i>Bathymodiolus thermophilus</i> et mise en évidence d'une zone hybride sur la Dorsale Est-Pacifique (EPR).....	- 137 -
IV.1. Introduction	- 137 -
IV.2. Matériel et méthodes.....	- 140 -
IV.2.1. Echantillons biologiques	- 140 -
IV.2.2. Extraction des protéines et visualisation sur gel d'amidon.....	- 141 -
IV.2.3. Séquençage des locus nucléaires et mitochondrial	- 142 -
IV.2.4. Descripteurs génétiques associés à l'analyse en fréquence des allozymes....	- 142 -
IV.2.5. Descripteurs génétiques associés à l'analyse des séquences.	- 144 -
IV.3. Résultats	- 144 -

IV.3.1. Structure génétique des populations de <i>Bathymodiolus thermophilus</i> au niveau de l'EPR sud.....	- 144 -
IV.3.2. Recherche d'allèles introgressés à partir d'une analyse de coalescence de huit gènes.....	- 154 -
IV.4. Discussion	- 166 -
IV.4.1. Structure génétique des populations	- 166 -
IV.4.2. Détection de zones de contact secondaire avec hybridation entre deux espèces cryptiques de <i>B. thermophilus</i> au sud et au nord de la dorsale EPR.	- 169 -
IV.4.3. Mise en évidence d'un processus de spéciation allopatrique avec formation d'un contact secondaire récent (le cas de la microplaque de l'île de Pâques).....	- 170 -
IV.4.4. Influence de l'introgession sur la différenciation génétique des populations de <i>B. thermophilus</i>	- 172 -
IV.4.5. Zone d'hybridation : contact secondaire entre espèces vicariantes ou colonisation de la dorsale EPR par une espèce issue du Pacifique Ouest ?.....	- 175 -
IV.4.6. Dynamique de la colonisation de la dorsale.....	- 177 -
IV.5. Conclusion	- 180 -
Chapitre V. Processus d'évolution moléculaire de quelques gènes.	- 183 -
V.1. Introduction	- 183 -
V.1.1. Evolution des gènes et de leurs fonctions.	- 183 -
V.1.2. Objectifs	- 186 -
V.2. Recherche d'un polymorphisme ancestral balancé au niveau du gène de la sulfotransférase.....	- 186 -
V.2.1. Présentation des données.....	- 186 -
V.2.2. Analyses de diversité.....	- 188 -
V.2.3. Test de détection d'un effet sélectif : le test de Mc Donald et Kreitman	- 190 -
V.2.4. Discussion sur l'évolution du gène de la sulfotransférase chez les bathymodioles. .	- 190 -
V.3. Mise en évidence d'un effet légèrement délétère du polymorphisme intronique sur le gène EF1 α chez le bivalve hydrothermal profond <i>Bathymodiolus</i>	- 191 -
V.4. Rôle de la duplication dans l'évolution de la protéine de choc thermique (Heat Shock Protein 70 : HSP70).....	- 194 -
V.4.1. Description de la famille des HSP 70.....	- 194 -
V.4.2. Présentation du jeu de données et des arbres de coalescence.....	- 195 -
V.4.3. Comparaison des pressions sélectives entre les clades.	- 199 -
V.4.4. Sélection positive ou relâchement des pressions de sélection ?	- 205 -
V.4.5. Conclusion sur contraintes de sélection concernant les lignées paralogues de HSP70.....	- 207 -
V.4.6. Analyse de la diversité nucléotidique au sein des clades du gène HSP70.	- 207 -
V.4.7. Discussion.	- 212 -
V.5. Recherche de sélection positive sur les lignées paralogues de la Ferritine 1 (GF1) - 215	-
V.5.1. Recherche de différences au niveau des pressions sélectives entre les lignées.- 217	-
V.5.2. Analyse des pressions sélectives sur les codons	- 218 -
V.5.3. Analyses de la diversité nucléotidique au sein des 8 lignées	- 221 -
V.5.4. Conclusion concernant l'évolution du gène GF1	- 222 -
V.6. Conclusion concernant les gènes dupliqués	- 224 -

Chapitre VI. Conclusion générale et perspectives sur les bivalves marins.....	- 225 -
VI.1. Synthèse des résultats obtenus sur <i>Bathymodiolus</i>	- 225 -
VI.2. Conclusions évolutives concernant le genre <i>Bathymodiolus</i>	- 227 -
VI.3. Comparaison inter-genres chez les bivalves marins	- 229 -
VI.3.1. Le modèle <i>Crassostrea</i>	- 230 -
VI.3.2. Comparaison de deux gènes communs	- 236 -
VI.3.3. Discussion générale sur la comparaison des genres et perspectives	- 240 -

ANNEXES

Introduction

1. Notion d'espèce et spéciation

L'espèce est l'unité de base reconnue par tous les biologistes. Cependant répondre à la question simple « qu'est ce qu'une espèce ? » pose toujours problème. Le concept biologique de l'espèce a été définie pour la première fois par Mayr en 1942 « un groupe de populations naturelles, effectivement ou potentiellement interfécondes, génétiquement isolées d'autres groupes similaires » (Mayr 1942) mais, le débat reste toujours d'actualité comme l'atteste la littérature récente à ce sujet (de Queiroz 2005, Hey 2006a). En fait, il n'existe aucun consensus universel sur cette définition, sauf qu'une majorité s'accorde pour reconnaître le critère d'interstérilité *in natura* comme suffisant, sinon nécessaire, pour preuve d'appartenance à des espèces différentes. A cet égard, l'une des principales préoccupations des biologistes de l'évolution depuis la publication de « L'origine des espèces » par Charles R. Darwin en 1859 (Darwin 1859), est de comprendre quels sont les mécanismes qui permettent d'aboutir *in fine* à cette unité reproductivement isolée qu'est l'espèce.

L'isolement reproductif est le résultat de la mise en place de barrières aux flux géniques entre deux espèces et ne résulte pas systématiquement de la mise en place de barrières physiques empêchant tout contact entre populations. L'isolement reproductif est un processus complexe affectant l'ensemble du génome. En effet, si deux populations divergentes continuent à partager un ensemble de gènes entre leurs génomes, ces populations ne peuvent être considérées comme étant reproductivement isolées (revue dans (Wu & Ting 2004)).

2. Mécanismes d'isolement reproductif

Les mécanismes conduisant à un isolement reproductif complet peuvent être distingués en fonction de leur nature pré- ou post-reproductive (Figure 0.1). Un isolement pré-reproductif qualifie l'absence de possibilités de fécondation interspécifique (et donc de production d'hybrides). Ceci peut être lié au fait par exemple, que les deux espèces présentent des caractéristiques écologiques ou éthologiques différentes. Ces caractéristiques peuvent

être : un décalage des périodes de reproduction ou de floraison (Rice 1987), une faible attractivité ou une non reconnaissance entre partenaires sexuels potentiels (chants différents chez les oiseaux : Irwin (2000) Podos (2007), des patrons de coloration différents chez les papillons Jiggins et al. (1997)). L'isolement post-reproductif complet ou partiel, entre deux espèces correspond à la production de zygotes hybrides, mais ces hybrides seront généralement affectés par une valeur sélective plus faible (léthalité des jeunes stades de développement, stérilité de la descendance, dépression hybride, maladaptation de certains génotypes).

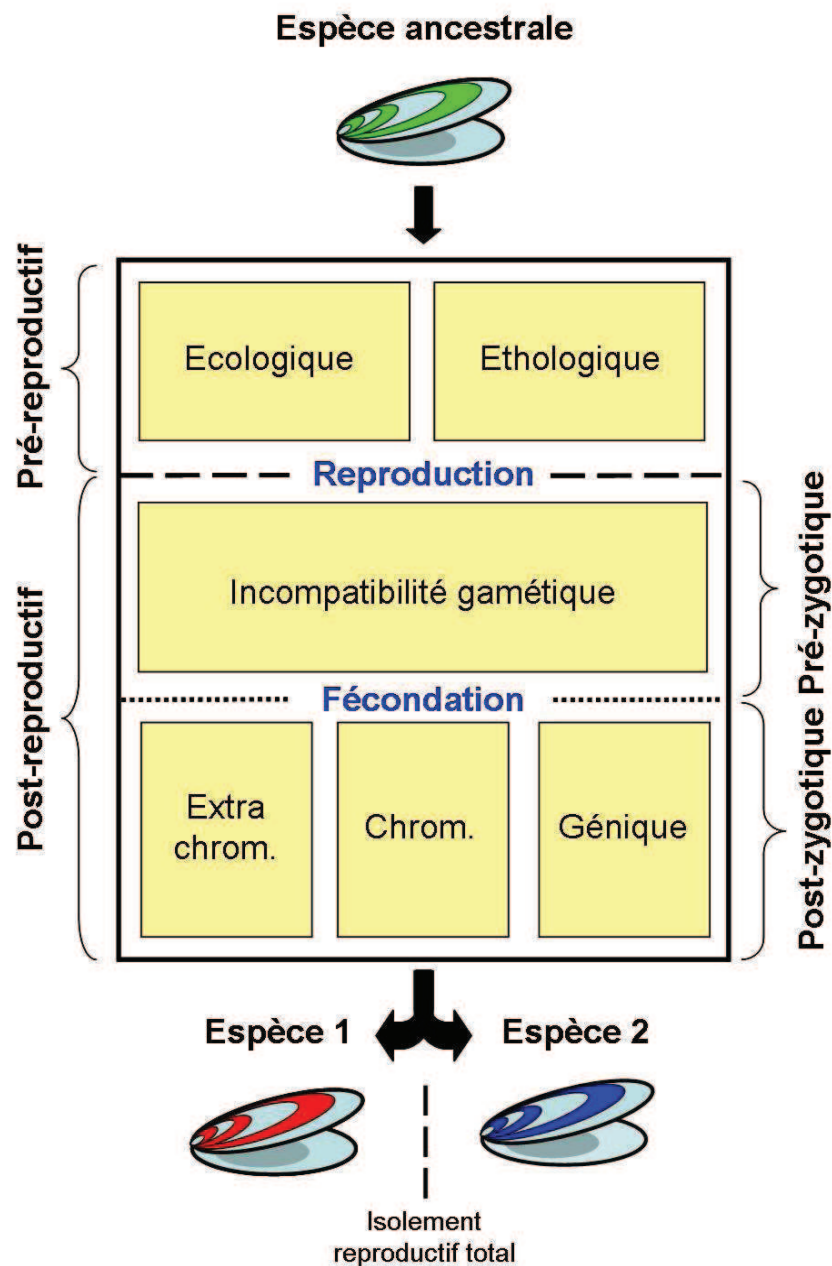


Figure 0.1: Schéma des mécanismes intervenant sur l'isolement reproductif. (Extra chrom : extra-chromosomique ; Chrom : chromosomique).

Deux classes de mécanismes d'isolement post-reproductif peuvent également être distinguées : les mécanismes pré-zygotiques qui interviennent avant la fécondation (incompatibilités gamétiques (Howard 1999)), et les mécanismes post-zygotiques qui se produisent après la fécondation. Ces mécanismes post-zygotiques sont à l'heure actuelle les mieux étudiés et l'on peut distinguer plusieurs catégories : (i) les causes extra-chromosomiques (occasionnées par des transferts horizontaux de matériel génétique entre les génomes, ces transferts pouvant être dus à une invasion virale, à la présence de symbiotes cytoplasmiques ou à la présence d'éléments transposables (Kazazian 2004), (ii) les remaniements chromosomiques (Rieseberg 2001) tels que la fusion Robertsonienne (Pascoe et al. 1996), ou polyploïdisation (Otto & Whitton 2000, Thiriot-Quievreux 2003), et (iii) les causes géniques (incompatibilité des gènes entre espèces divergentes (Presgraves 2003)).

Le modèle d'incompatibilité génique des hybrides le plus couramment invoqué se conforme au modèle à deux locus dit de Dobzhansky-Muller (Orr 1995) (Figure 0.2). Considérons deux locus et une population ancestrale possédant uniquement des individus diploïdes caractérisés par le génotype $AA BB$. Après séparation de la population ancestrale en deux sous-populations isolées, des allèles nouveaux vont se fixer dans chacune des populations : A mute pour a dans la première sous-population et B mute pour b dans la deuxième sous-population. Confrontées l'une à l'autre, ces deux mutations (a et b) peuvent être mutuellement incompatibles (c'est le cas par exemple si elles agissent de manière contradictoire sur l'expression d'un génotype) mais cela ne pose aucun problème car leurs populations respectives sont séparées. Lors de l'hybridation d'individus provenant de chacune des deux populations, les deux mutations a et b sont alors mises en contact et vont interagir (par interaction épistatique) dès la première génération. En fonction du niveau d'incompatibilité entre les deux mutations, la dépression hybride pourra engendrer une simple diminution de la valeur sélective (fitness) des hybrides, leur stérilité ou leur létalité.

L'isolement reproductif complet résulte d'interactions épistatiques négatives fortes entre les allèles de différentes espèces à plusieurs locus qui entraînent un dysfonctionnement et bloquent les échanges génétiques entre génomes par élimination des hybrides (Orr 1995). Ces locus d'isolement, appelés gènes de spéciation constituent les « verrous » des génomes car ils s'opposent à leur ré-homogénéisation complète, en cas d'hybridation et leur accumulation conduit inévitablement à l'isolement total, étape ultime de la spéciation (Figure 0.3).

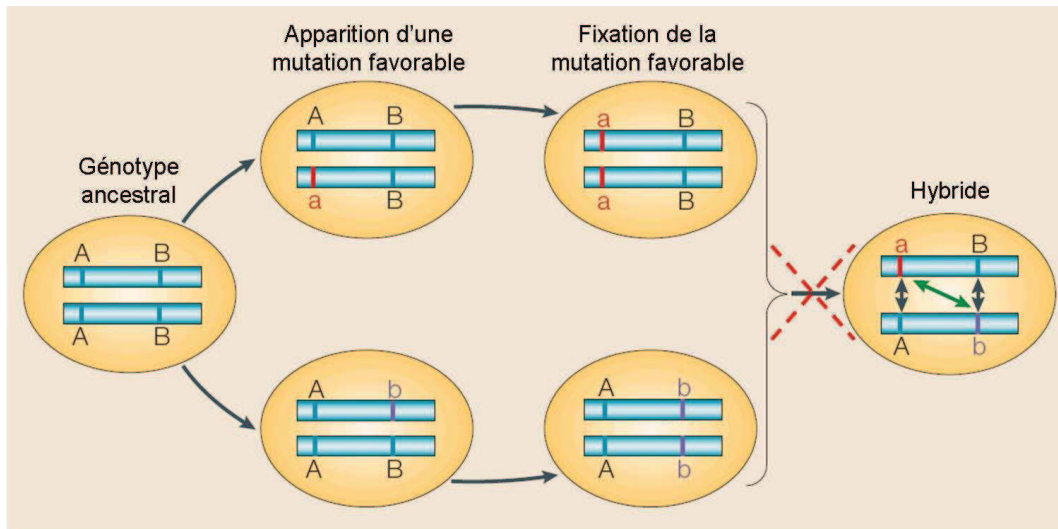


Figure 0.2. représentation schématique du modèle de Dobzhansky-Muller. A et B sont les génotypes ancestraux ; a et b (en rouge) sont les mutations favorables. Pour l'hybride, les doubles flèches noires représentent les loci divergents incriminés et la double flèche verte correspond aux interactions épistatiques négatives entre les allèles incompatibles. La croix rouge représente la létalité des hybrides lorsque l'incompatibilité est totale entre les allèles a et b. Modifié de (Wu & Ting 2004).

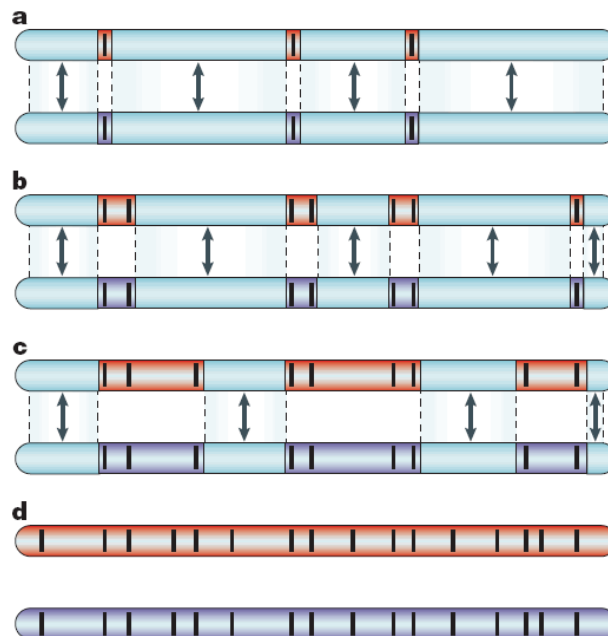


Figure 0.3. Importance des gènes « verrous » dans la spéciation. Les deux barres horizontales représentent le génome de deux populations divergentes. Au début (a), les allèles de trois loci (traits noirs) sont mutuellement incompatibles et les gènes de ces loci ne s'échangent plus entre les populations, (ce sont des gènes « verrous »). Les flux génétiques sont maintenus, pour le reste du génome (flèches) grâce à la recombinaison qui se produit chez des hybrides à fitness réduite mais pas nulle. Au fur et à mesure (b et c), la différenciation entre loci augmente et les flux de gènes se réduisent permettant la formation de régions complètement verrouillées (en rouge et violet). Finalement (d) les deux populations présentent un isolement reproductif complet et peuvent être considérées comme des espèces séparées. Issu de (Wu & Ting 2004).

3. Modes de Spéciation

Trois modes de spéciation basés sur des critères géographiques sont généralement décrits : en allopatrie, en parapatrie et en sympatrie.

La spéciation en allopatrie consiste en la création de nouvelles espèces après un isolement géographique des populations. Le type de barrière à l'origine de cet isolement peut être la création d'une montagne, l'isolement par un réseau hydrographique...) Dans ce modèle, il n'existe pas de flux géniques entre les populations divergentes et la barrière physique à la migration va permettre aux deux populations d'évoluer indépendamment du fait de la dérive génique et de l'adaptation locale (figure 0.4a, b, c, d).

Le mode de spéciation en parapatrie correspond à l'adaptation locale d'une population située en marge de l'aire de répartition de la population initiale. Ces populations restent connectées par des flux géniques résiduels pendant le processus de divergence (Wilding et al. 2001) (figure 0.4e, f, g h).

Le mode de spéciation en sympatrie, très controversé, décrit la formation de nouvelles espèces au sein d'une même aire géographique. Il s'agit d'une forme extrême de parapatrie, avec des flux géniques moins restreints et une différenciation entre les niches écologiques occupées par les populations divergentes. Cette spéciation peut se produire à une petite échelle spatiale (Barluenga et al. 2006, Savolainen et al. 2006, Nosil 2008) (figure 0.4i, j, k, l).

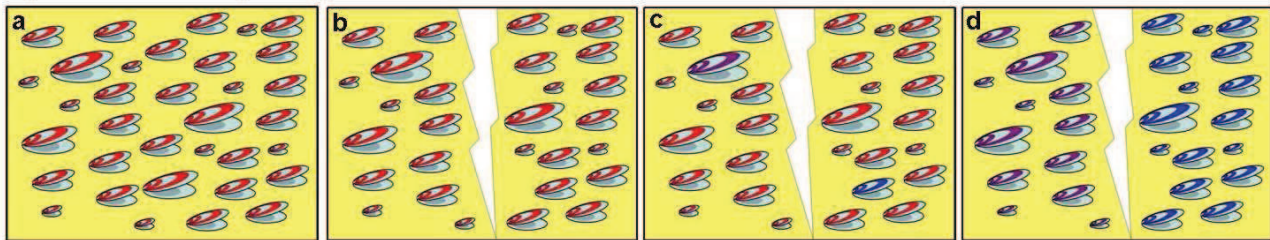
Contrairement au modèle allopatrique, les modes de spéciation en parapatrie et en sympatrie résultent de processus adaptatifs diversifiants et ne nécessitent aucune barrière physique. Il semble cependant que les mécanismes d'adaptation locale soient renforcés par une atténuation des flux de gènes liés à la mise en place de barrières perméables (orientation de flux migratoires sur l'estran (Engel et al. 2005); capacités dispersives nulles ou limitées, etc...).

Entre des populations interconnectées possédant des flux géniques importants et des espèces 'vraies' présentant une complète incompatibilité inter-locus, il existe un continuum d'intensité de flux géniques (Harrison 1990, Arnold 1997). Ces modes de spéciation avec flux géniques font depuis longtemps l'objet d'un débat (Wiens & Graham 2005 Bolnick 2004). Jusqu'ici, les modèles théoriques ne prédisent leur existence que sous des conditions très

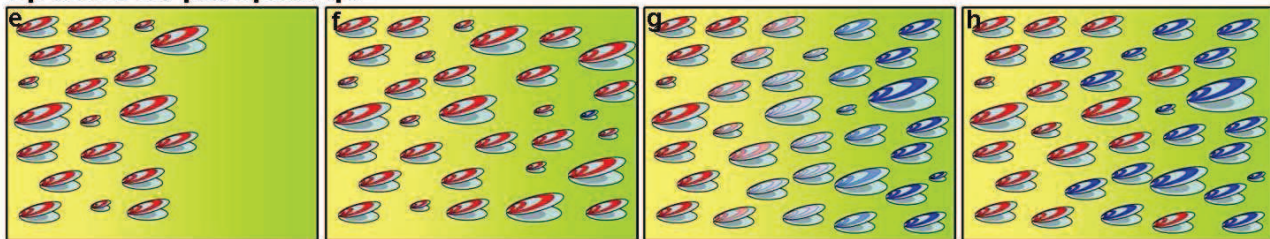
particulières et donc rares (Bolnick & Fitzpatrick 2007). Régulièrement, des exemples de spéciation avec flux génique sont pourtant mis en évidence (Via 2001, Nosil 2008).

L'analyse des séquences nucléotidiques sur un grand nombre de locus et d'individus est désormais facilement accessible par le développement récent des méthodes de séquençage 'haut débit'. L'acquisition de séquences entre et au sein d'espèces proches est particulièrement adaptée à l'étude des mécanismes de spéciation en présence de flux géniques. Ce type d'analyses repose sur la théorie de la coalescence.

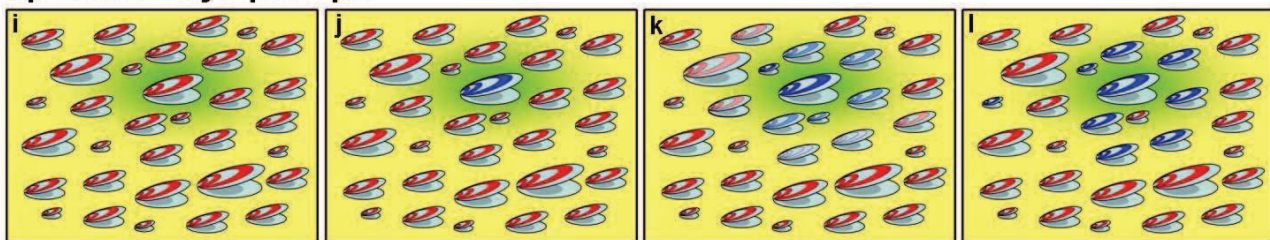
Spéciation allopatrique



Spéciation parapatrique



Spéciation sympatrique



Temps

Figure 0.4. Schéma illustratif des différents modes de spéciation. Spéciation en allopatrie : l'aire de distribution ancestrale (a) est fragmentée par une barrière totale aux flux géniques (b), de nouvelles mutations vont apparaître (c), les deux populations isolées vont diverger par dérive et créer deux espèces-sœurs. (d). Spéciation en parapatrie : en limite d'aire de distribution (e) colonisation d'une nouvelle niche par quelques individus puis une nouvelle mutation adaptative apparaît (f), les populations vont diverger en maintenant un certain niveau de flux géniques (g), et favoriser l'isolement d'une espèce-sœur en bordure de l'aire de répartition ancestrale (h). Spéciation en sympatrie : les deux futures espèces occupent la même aire de distribution présentant une mosaïque d'habitats (i), un polymorphisme adaptatif apparaît par sélection diversifiante (j), les populations vont progressivement diverger en conservant des flux géniques (k), jusqu'à l'isolement d'une espèce-sœur dans un micro habitat (l).

4. Théorie de la coalescence

L'approche classique de la génétique des populations est basée sur l'analyse des fréquences d'allèles dans les populations (Wright 1977). Il s'agit d'une approche prospective qui consiste à prédire l'écart du polymorphisme génétique à l'attendu d'Hardy-Weinberg dans une population, puis à rechercher la part des différentes forces évolutives en présence sous l'hypothèse d'un équilibre migration/dérive (Hartl 1988). La théorie de la coalescence (Kingman 1982) propose une approche complètement opposée car elle est rétrospective. Elle retrace l'histoire des gènes à partir de la description des processus de fusion (coalescence) des lignées phylogénétiques au sein de gènes orthologues dans une population ou une espèce (Figure 0.5 b). L'objectif est de retracer l'histoire évolutive de ces lignées alléliques jusqu'à leur plus proche ancêtre commun (MCRA). La coalescence des généalogies est décrite dans des modèles de population théoriques où le nombre de paramètres est réduit. Le modèle le plus simple est celui de Fisher (Fisher 1930) Wright (Wright 1931). Dans ce modèle, la taille de la population est constante et les générations ne sont pas chevauchantes. A chaque génération, tous les individus meurent. La génération suivante est ensuite générée à partir d'un tirage aléatoire des gamètes issus de la génération précédente (Figure 0.5 a). Seule la dérive modifie la fréquence des allèles.

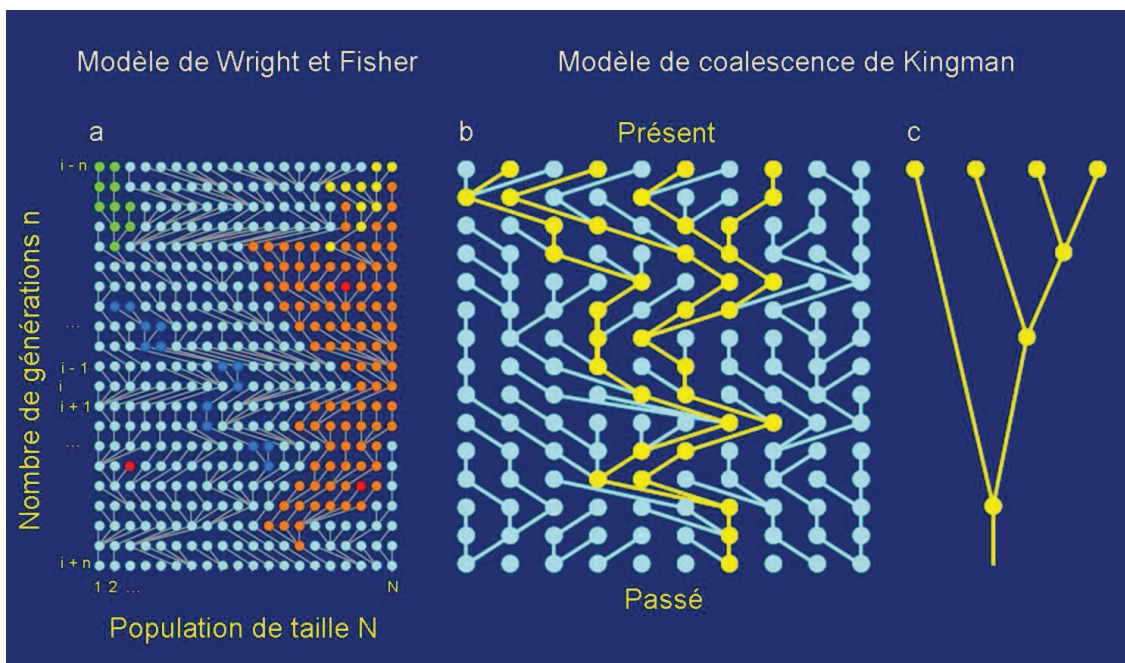


Figure 0.5. Modèle de base utilisé par la théorie de la coalescence, chaque point représente un individu et la couleur représente son génotype. Les lignes indiquent l'apparentement entre les génotypes à chacune des générations. (a) : représentation de l'évolution des génotypes dans une population sous le modèle de Fisher et Wright. (b) : représentation d'une lignée particulière sous le modèle de Fisher et Wright, et (c), sa correspondance en coalescence sous le modèle de Kingman. (modifié Beerli)

La théorie de la coalescence permet de calculer l'ensemble des indices de diversité génétique au sein des populations, mais surtout de simuler facilement les processus évolutifs affectant la diversité génétique en utilisant un minimum de descripteurs de l'état du système à chaque génération. Sous l'hypothèse d'une évolution neutre des allèles dans le polymorphisme, la diversité nucléotidique observée est directement reliée à la taille efficace de la population par la relation $\theta\pi = \theta_w = 4N_e\mu$ où $\theta\pi$ et θ_w sont des estimateurs de la diversité nucléotidique et μ est le taux de mutation au locus considéré. Dans un coalescent neutre et sous l'hypothèse d'un équilibre mutation/dérive avec un modèle de mutation en nombre infini d'allèles, le nombre de mutants croît en suivant une loi de puissance à chaque génération. En comparant les polymorphismes observés à ceux attendus sous l'hypothèse neutre, il est possible de tester différents scénarios évolutifs comme la mise en place d'une structure géographique, les effets démographiques passés (ou les effets sélectifs qu'ils soient diversifiants (sélection positive) ou stabilisants (sélection balancée)).

La coalescence est donc très puissante pour comparer la vraisemblance de scénarios évolutifs pouvant conduire à la divergence moléculaire constatée sur l'ensemble des gènes d'une population. Ainsi la topologie des arbres de coalescence peut être très informative sur le type de sélection affectant l'histoire d'un gène en particulier (Figure 0.6).

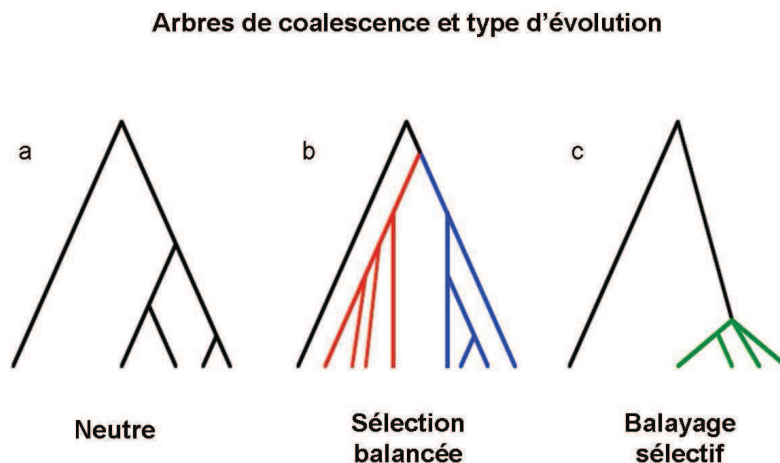


Figure 0.6. Topologies d'arbre de coalescence pouvant être observées selon une évolution neutre (a) ou sous contraintes sélectives (b, c).

Des approches récentes permettent également de tester l'existence de flux géniques éventuels au cours des phénomènes de spéciation (revue dans Hey (2006b)). Ces méthodes utilisent des données de polymorphismes/divergence provenant de plusieurs locus, et ce, pour chacune des deux espèces ou populations divergentes car les patrons de coalescence varient

d'un gène à l'autre. Les exemples de divergence avec flux géniques sont alors mis en évidence lorsque certains gènes ne montrent que très peu de divergence et des polymorphismes mélangés alors que d'autres présentent un fort niveau de divergence et des polymorphismes séparés. Cette variation de la divergence entre 2 espèces au niveau d'un grand nombre de gènes est alors plus importante que celle attendue sous un modèle sans flux génique (Wakeley & Hey 1997).

5. Comportement des gènes en condition d'hybridation :

A la différence d'une barrière complète aux flux géniques qui, par définition, est totalement hermétique pour l'ensemble des gènes, il existe des barrières dites « semi-perméables » qui vont permettre des échanges génétiques sous certaines conditions (Harrison 1990). Ce sont ces barrières semi perméables qui sont à l'origine des zones d'hybridation.

Une zone d'hybridation résulte de la confrontation entre deux forces antagonistes : la migration et l'isolement reproductif. La migration tend à homogénéiser les fréquences alléliques entre les populations (ou espèces) alors que l'isolement reproductif va privilégier les génotypes parentaux. Une zone d'hybridation peut ainsi être résumée par la création d'un gradient de fréquence (cline) des gènes neutres autour du point de rupture des fréquences d'allèles associées aux gènes verrous, impliqués dans l'isolement reproductif. L'intensité du cline sera variable selon les gènes (Kruuk et al. 1999) (Figure 0.7).

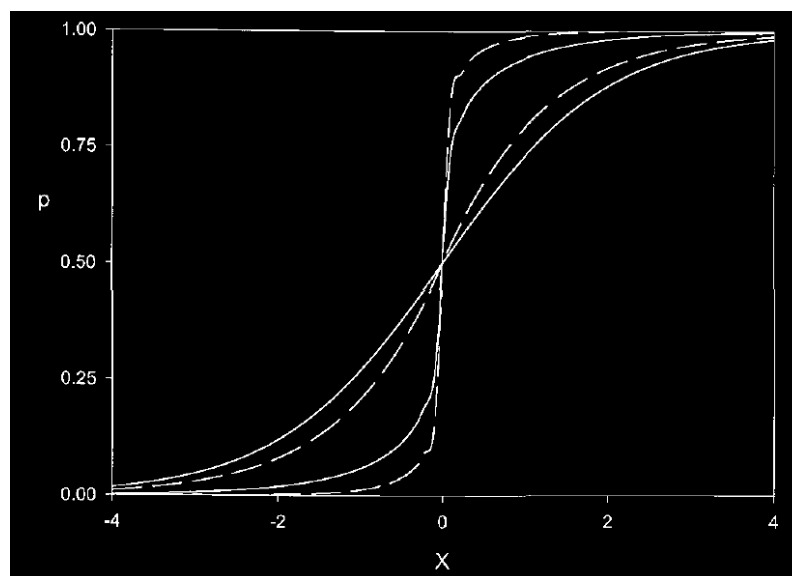


Figure 0.7. Cliniques de fréquences obtenus par sélection environnementale et dépression hybride. X est la distance de chaque côté du point de rencontre entre les 2 espèces. P est la fréquence de l'allèle pour un gène donné (Kruuk et al. 1999).

Le génome se révèle fortement hétérogène vis-à-vis de l'introgession. Une fois introgressés dans une nouvelle population, certains gènes adaptatifs pourront se trouver sous sélection positive et pourront être soumis à un balayage sélectif¹ leur permettant d'atteindre de fortes fréquences et de se fixer. Il s'agit là d'introgession adaptative dirigée, donnant lieu à des flux asymétriques entre les 2 génomes.

Le passage des gènes neutres à travers une barrière semi-perméable est théoriquement peu soumis à la sélection et donc symétrique entre les deux populations (espèces). Pourtant la fréquence des allèles propres à ces gènes et l'intensité de leur introgression vont être modulées en fonction des pressions de sélection associées aux gènes auxquels ils sont liés (Figure 0.8). En effet, si les gènes neutres sont liés à des gènes verrous, l'échange d'allèles à travers la barrière semi-perméable sera fortement ralenti. En revanche, s'ils sont liés à un gène sous sélection positive, ils profiteront de sa propagation par l'intermédiaire de l'effet d'auto-stop génique (Maynard Smith & Haigh 1974, Faure et al. 2008). La fréquence des allèles introgressés qui sera atteinte au niveau de ces gènes neutres dans les populations sera fonction de la distance de ces gènes avec le locus sous sélection et de l'intensité de la recombinaison. En effet, plus la distance est importante, plus les chances de casser la liaison entre le locus sélectionné et le locus neutre par recombinaison sont importantes et moins la fréquence maximale d'équilibre est élevée. Il ne faut cependant pas oublier que dans le cas complètement neutre (liés à rien), certains locus sont tout de même capables d'atteindre l'équilibre et l'équifréquence dans les 2 populations parentales.

Les moules du genre *Mytilus* sont connues pour former des zones d'hybridation en mosaïque, ce qui a permis d'entreprendre des études sur les flux géniques et l'introgession (Daguin et al. 2001, Bierne et al. 2003, Faure et al. 2008). Cet effet mosaïque semble être le résultat du fort pouvoir dispersif des larves de ces bivalves à méso-échelle (par exemple : Atlantique nord est). Cependant, cette caractéristique semble être plus largement l'apanage des Mytilidae au sens large puisque qu'une zone d'hybridation a été bien documentée chez les *Bathymodiulus* des zones profondes (O'Mullan et al. 2001, Won et al. 2003a).

¹ Balayage sélectif : diminution locale de la diversité génétique suite à la fixation d'une mutation avantageuse et des mutations qui lui sont liées.

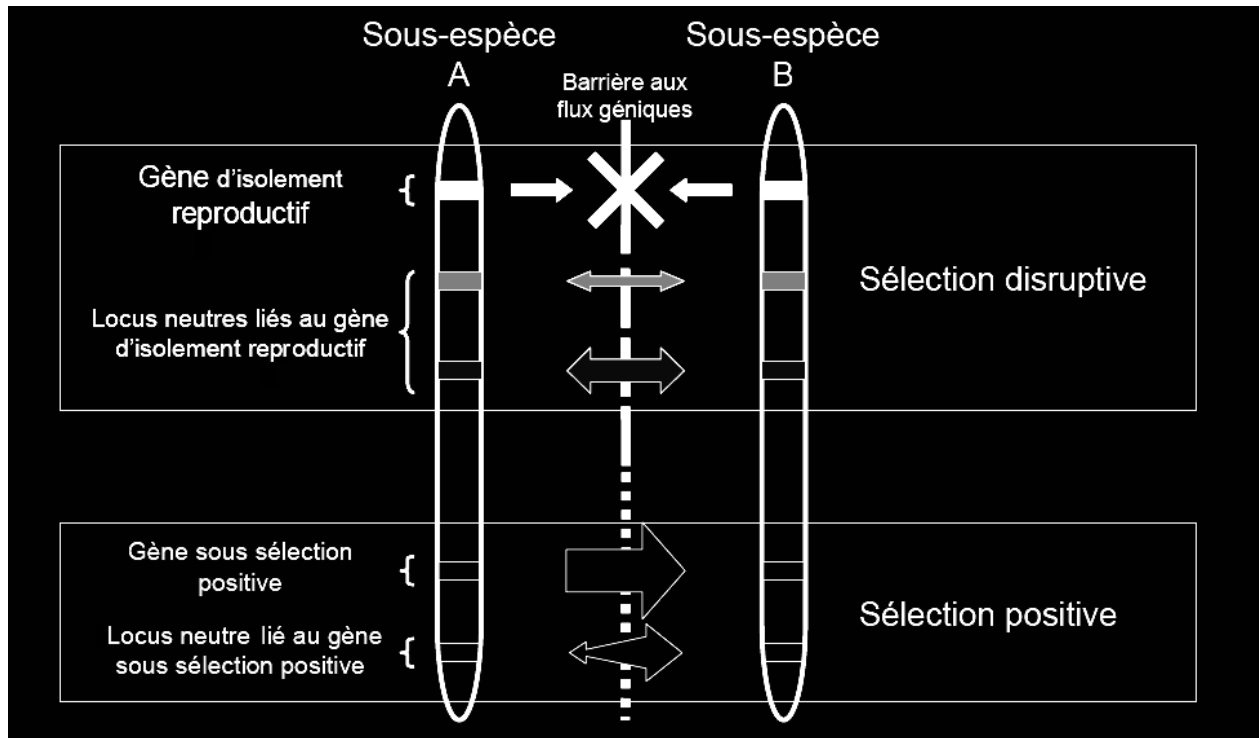


Figure 0.8. Schéma de deux pools génétiques (représentés par des chromosomes) séparés par une barrière semi-perméable aux flux génétiques. Les différentiels de déséquilibre de liaison et de régimes de sélection expliquent la variabilité d'introgression (Faure 2007).

6. Intérêt des bivalves marins pour la génétique

Les moules et les bivalves marins présentent en général, un très fort polymorphisme génétique. Cette diversité contraste fortement avec le très faible nombre de critères morphologiques diagnostiques présents au sein de ce taxon et de la forte plasticité phénotypique observée au niveau de leur coquille (Locard 1889). La diversité enzymatique observée chez les bivalves marins est la plus élevée des Eucaryotes (Solé-Cava & Thorpe 1991) et de fortes diversités ont également été rapportées avec des marqueurs microsatellites (Huvet et al. 2000, Launey & Hedgecock 2001) ou introniques (Daguin et al. 2001, Bierne et al. 2003). Ces données suggèrent que la taille efficace des populations (N_e) de bivalves marins est très grande et proche d'une taille 'infinie' préconisée dans les modèles théoriques de la génétique des populations.

Comme pour la plupart des espèces marines, les bivalves ont un cycle de vie benthopélagique, avec une phase adulte sessile présentant une très forte fécondité (plusieurs milliers de larves par individu) et une phase larvaire dispersive généralement longue, permettant des flux génétiques sur de longues distances (Thorson 1950).

Au moment de la reproduction, les gamètes sont émis dans le milieu et la fécondation a lieu dans la colonne d'eau (il existe cependant des exceptions : *Ostrea sp*). La phase larvaire de longue durée et la grande variété des habitats marins favorisent l'action de la sélection chez ces animaux. De plus, les très grands effectifs efficaces des bivalves permettent d'augmenter la fréquence d'apparition d'une mutation « adaptative » dans la population et, donc la récurrence de balayages sélectifs dans les populations.

Il convient néanmoins de relativiser l'importance de la taille efficace comme moteur de l'évolution chez les bivalves marins. Considérant le nombre élevé de reproducteurs dans une population et étant donné le grand nombre de larves émises dans le milieu naturel, il serait intuitif de penser que le « pool » larvaire renferme une diversité similaire à celle de la population initiale ; pourtant dans certains cas, seuls quelques parents participent majoritairement à la génération suivante. Cette variance du succès reproducteur a pu être mise en évidence chez l'huître américaine *Crassostrea virginica* en milieu artificiel (Hedgecock & Sly 1990, Hedgecock et al. 1992). La variation des contributions parentales est liée à la qualité des gamètes, à l'interaction sperme / ovocytes ou à un différentiel de viabilité entre les génotypes (Boudry et al. 2002). Ces différences génétiques au cours de la reproduction et de la phase larvaire constituent autant de phases au cours desquelles la sélection peut jouer. Dans un contexte d'élevage, ce type de phénomène va rapidement conduire à une perte de diversité génétique chez les reproducteurs et pourra entraîner de la consanguinité à plus ou moins long terme.

Dans le milieu naturel, la rencontre des gamètes va se faire de manière aléatoire. Cependant, le succès reproducteur dépend généralement du degré d'agrégation des individus et il n'est pas improbable que des gamètes issus d'individus apparentés se fécondent. Dans ce cas, une dépression de consanguinité peut être liée à la fixation ou à la ségrégation de locus enzymatiques (McGoldrick & Hedgecock 1997).

La phase larvaire longue distance peut entraîner la rencontre de gamètes provenant de deux espèces proches, encore interfécondes et conduire à des phénomènes d'hybridation. Ces exemples ont été très bien étudiés chez les espèces de Mytilidae (Sarver & Foltz 1993, Rawson et al. 1999, Bierne et al. 2002). Ces appariements inter-spécifiques produisent de la dépression hybride (contre-sélection des individus hybrides). Il a ainsi été démontré qu'en

présence de compétition gamétique entre *M. edulis* et *M. galloprovincialis*, les fécondations se faisaient préférentiellement au sein de chaque espèce (entre gamètes de la même espèce).

Compte tenu du nombre élevé de larves à chaque nouvelle génération, il est impossible que toutes parviennent à se fixer pour donner des adultes capables de se reproduire. Si l'on imagine un exemple théorique où toutes les larves seraient du même génotype (même valeur sélective) et seraient placées dans des conditions optimales de croissance l'ensemble de ces larves ne pourrait en aucun cas atteindre le stade adulte. En effet, les phénomènes de compétition, que ce soit pour la ressource trophique ou pour l'espace (lors du recrutement benthique) induiraient une mortalité aléatoire. Avec un tel exemple, il est aisé de comprendre que la moindre larve possédant un léger avantage sélectif sera favorisée et qu'au contraire, toutes celles exprimant la moindre déficience seront rapidement éliminées. C'est exactement ce qui se passe dans le milieu naturel. Compte tenu de la grande taille du « pool » larvaire (quasi illimité), le fait d'éliminer des individus n'a que peu de conséquences sur la taille ou le renouvellement de la population adulte. Le dernier argument concernant le bénéfice d'un nombre quasi-illimité de larves est que ce « pool » larvaire va renfermer une multitude de génotypes variés multipliant ainsi les chances d'être adapté à une micro-niche écologique. Cette multitude de mutants permet à la sélection naturelle de s'exercer sur le nuage larvaire afin de ne sélectionner que les génotypes les plus résistants ou les mieux adaptés.

L'étude du polymorphisme génétique des bivalves marins à partir de données de séquences nucléotidiques devrait donc permettre de mieux comprendre leurs dynamiques populationnelles et d'inférer plus précisément la contribution relative des forces évolutives agissant sur leurs génomes, que ce soit à l'intérieur d'une espèce ou en condition d'hybridation.

Les exemples de la littérature montrent que dans certains cas, les phénomènes d'adaptation locale et de spéciation peuvent être liés à l'apparition de mutations avantageuses pour un environnement particulier (Savolainen et al. 2006, Nosil 2008). Un milieu très fragmenté et constitué d'habitats variés, pouvant influencer la spécificité d'un génome serait donc particulièrement intéressant pour étudier les phénomènes d'adaptation, de sélection génique et les mécanismes de la spéciation. Il semble que les milieux profonds réduits correspondent parfaitement à ces exigences.

7. Les environnements marins réduits

Le milieu marin possède des environnements très particuliers aux conditions physico-chimiques extrêmes : les milieux réduits et anoxiques, côtiers ou profonds. En milieu profond, trois grands types d'environnement peuvent être distingués : les zones de suintements froids et les sources hydrothermales (Figure 0.9) et, de manière plus anecdotique (du point de vue de leur densité et de leur distribution), les cadavres de baleines et les bois coulés.

Les zones de suintements froids ont été décrites pour la première fois au niveau de saumures situées à plus de 3000 m de profondeur à la base de l'Escarpement de Floride (Paull et al. 1984) puis au niveau de zones à hydrocarbures du Golfe du Mexique (Kennicutt et al. 1985). D'autres sites ont ensuite été découverts au niveau du Pacifique Est sur des carbonates de la zone de subduction de l'Oregon (Kulm et al. 1986), puis au large du Japon au niveau de la zone de subduction Kaiko-Nankai par plus de 6000 m de fond (Ohta & Laubier 1987) ainsi que dans des zones moins profondes du Pacifique Ouest (Sagami Bay : 900 m, (Hashimoto et al. 1989)). Le nombre de sites depuis n'a cessé d'augmenter aux quatre coins du globe, notamment au niveau de la Barbade (Faugères et al. 1987), et plus récemment le long de la marge Africaine (zone REGAB : (Ondreas et al. 2005)) ainsi qu'au niveau de volcans de boue en mer Méditerranée (Olu-Le Roy et al. 2004), tous ces sites ayant été découverts à des profondeurs comprises entre 400 et 6000m. Ces suintements sont généralement associés à des marges actives ou passives. Ces habitats sont des environnements sédimentaires associés à des flux riches en méthane ayant une température proche de l'eau de fond. Le méthane peut posséder deux origines distinctes : thermogénique ou biogénique. L'origine thermogénique est issue de la transformation accélérée de la matière organique contenue dans l'eau interstitielle enfouie dans les sédiments profonds par les températures élevées de la compaction. Le méthanogénèse biogénique est obtenu par la transformation microbienne de la matière organique dans les couches anoxiques des sédiments. Dans certains cas, le méthane possède les deux origines (Kulm et al. 1986, Boulegue et al. 1987, Sibuet & Olu 1998). Dans les sédiments, la réduction des sulfates produit des sulfites qui jouent un rôle important dans la production énergétique de ces environnements. Ces zones, fortement stratifiées, présentent des variations dans la composition chimique, l'origine du méthane et le taux d'émission de leurs fluides, ce qui entraîne une hétérogénéité spatiale de l'habitat et influe sur les communautés biologiques (Sibuet & Olu 1998).

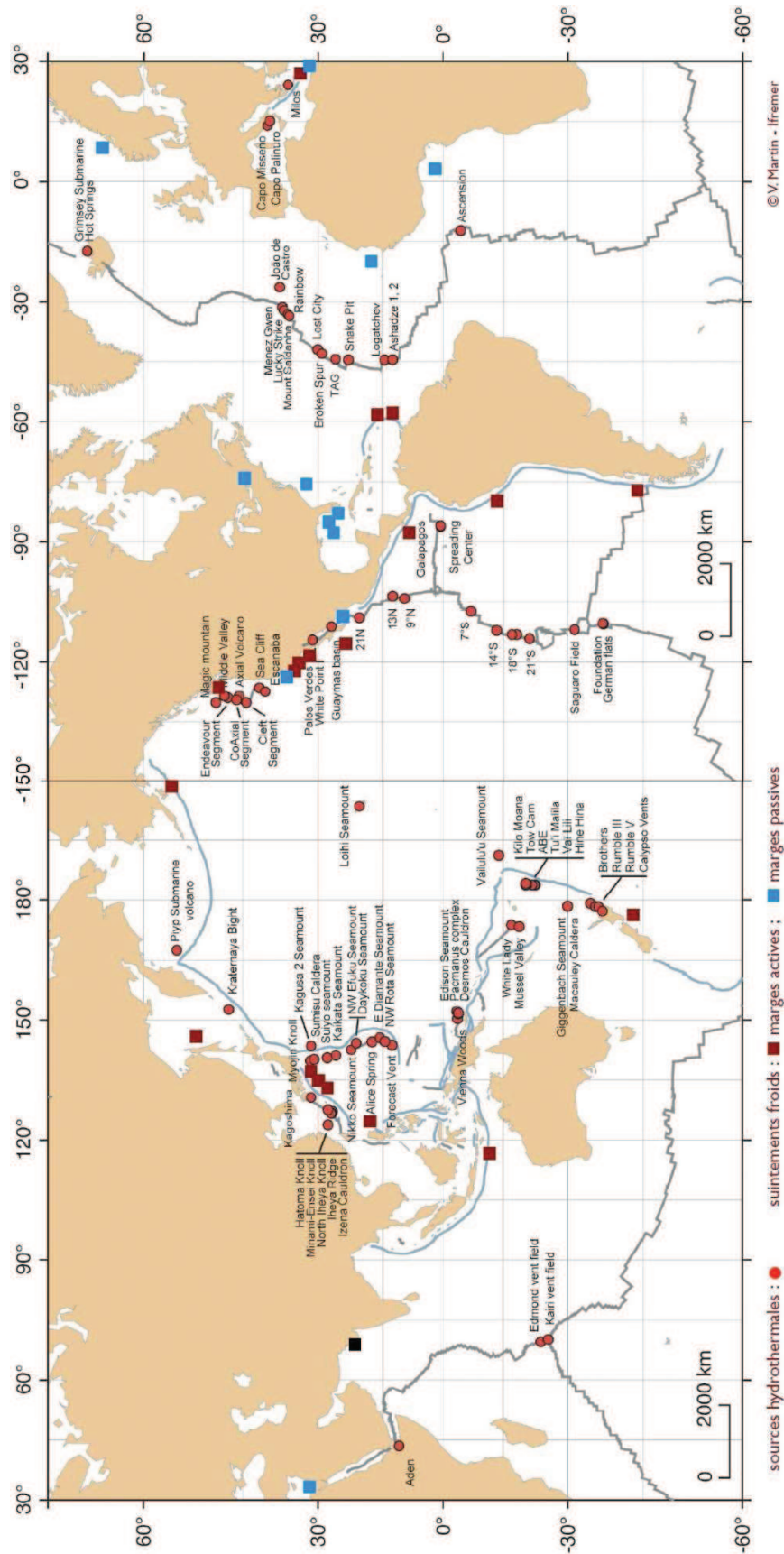
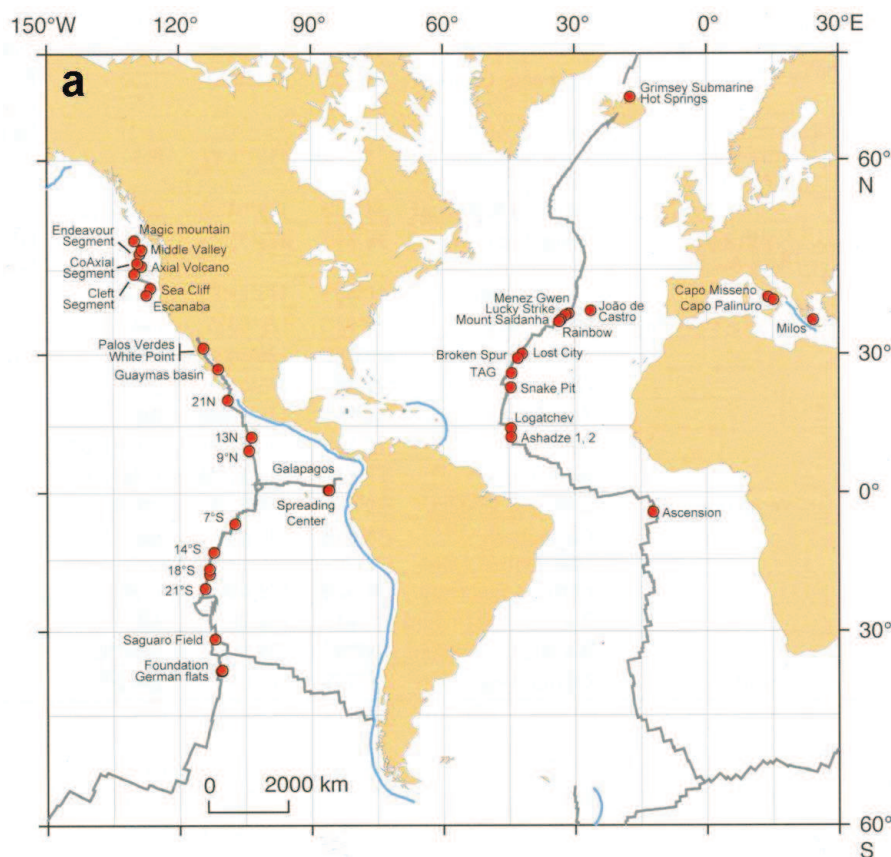


Figure 0.9. Distribution des zones de suintements froids (carrés bleus marges passives, carrés rouges marges actives) et des sources hydrothermales profondes (rond rouge) à l'échelle de l'océan mondial.

Les communautés associées à l'hydrothermalisme profond ont été décrites quelques années seulement auparavant par Lonsdale (1977) sur la ride des Galapagos, puis le long de la dorsale du Pacifique Est (EPR : Corliss et al. (1979)). Ces manifestations thermochimiques se manifestent généralement au niveau des zones d'accrétion formant les dorsales océaniques qui ceignent le globe. Ces émissions sont en effet présentes sur l'ensemble de l'EPR (Hessler et al. 1985, Renard et al. 1985), la dorsale du Pacifique nord (Tunnicliffe et al. 1985, Tunnicliffe et al. 1986), la dorsale médio-Atlantique (Rona 1986 ; Ondreas et al. 1997), les bassins arrière-arc du Pacifique ouest (Jollivet et al. 1989 ; Hessler & Lonsdale 1991) et la dorsale indienne (Van Dover et al. 2001). Depuis, la découverte de ces oasis de vie, caractérisés par de fortes biomasses, une faible richesse spécifique et un fort endémisme lié à la ressource trophique, les sources hydrothermales profondes n'ont eu de cesse d'attirer les scientifiques de toutes disciplines. Ainsi, taxonomistes, écologistes, physiologistes, microbiologistes, généticiens, physiciens et chimistes, collaborent afin d'étudier et comprendre ce milieu si particulier. Durant ces trois décennies, de nombreuses campagnes océanographiques dotées de véhicules subaquatiques (bathyscaphes habités, robots télé-opérés) ont été organisées pour explorer les dorsales océaniques et ont permis la découverte de nombreuses communautés hydrothermales dans le Pacifique, l'Atlantique, l'océan indien aussi bien que dans les bassins arrière-arc, où les arcs volcaniques du Pacifique (Figure 0.10 a, b).



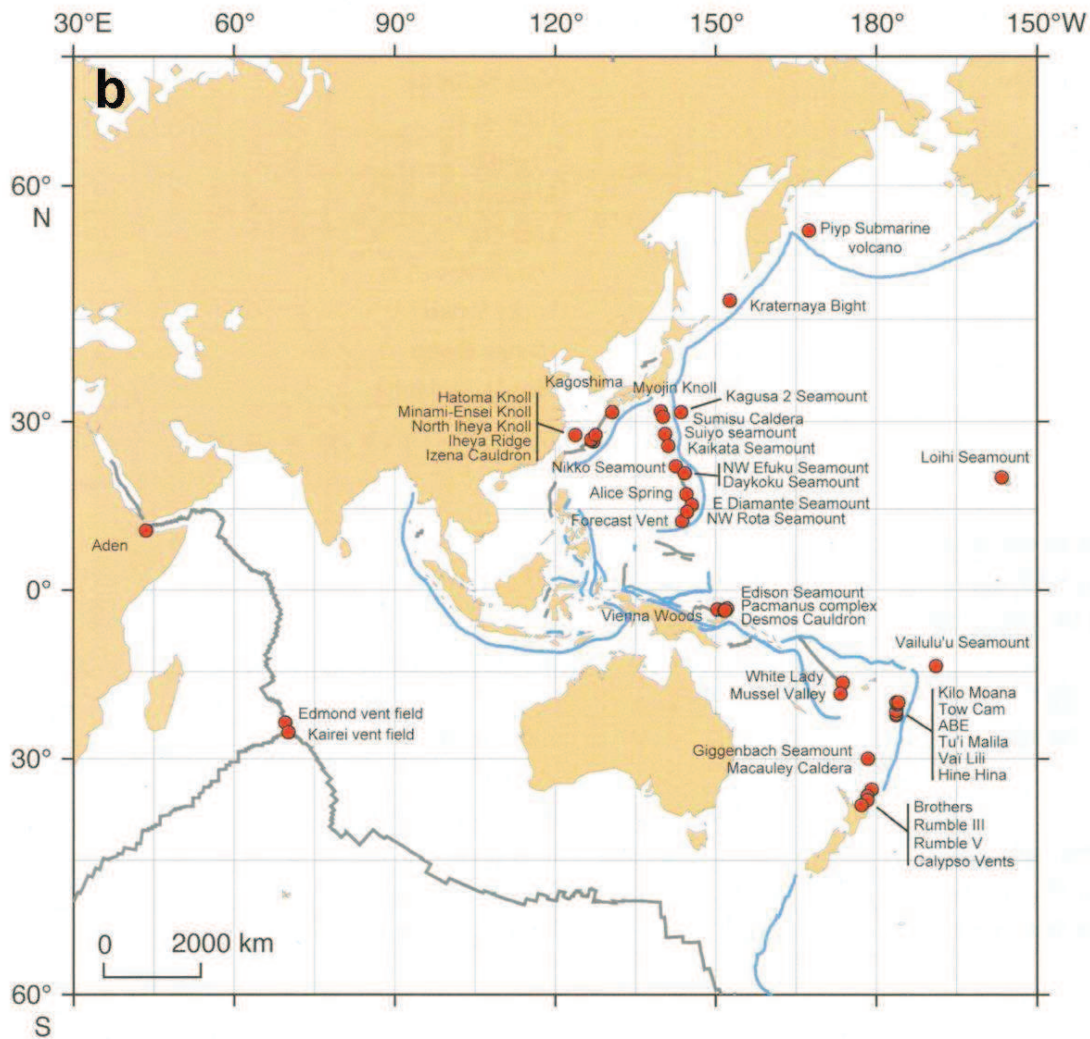


Figure 0.10. Localisation des principaux sites hydrothermaux connus à l'échelle de l'océan mondial : (a) Région Pacifique Est et Atlantique ; (b) Région Océan Indien et Pacifique Ouest.

La forte activité géologique rencontrée au niveau des dorsales entraîne la fracturation du plancher océanique et permet à l'eau de mer de s'infiltrer à travers les multiples fissures. Cette eau de mer de fond, froide et dense (2 à 4°C ; pH 7,8) va s'infiltrer jusqu'à la chambre magmatique au contact de laquelle elle va se réchauffer. Ce réchauffement va entraîner une diminution de sa densité et l'eau sous pression va alors se mettre à remonter à travers les roches. Cette remontée et l'augmentation de la température vont créer des interactions avec les roches environnantes et l'eau va progressivement se charger en éléments réduits tels que le H_2S , CH_4 , CO_2 , SiO_2 , NH_4^+ , et s'acidifier (pH 3,5). Cette acidification va permettre la dissolution de nombreux éléments métalliques (Mn^{2+} , Fe^{2+} , Li^+ , Cd^{2+} , Cu^{2+} , Zn^{2+}) donnant ainsi naissance au fluide hydrothermal. Lorsque ce fluide va jaillir hors de la croûte océanique, il va se créer une brusque zone de mélange entre l'eau de mer (2°C ; pH 7,8 ;

oxygénée et pauvre en métaux) et le fluide hydrothermal (350°C ; pH 3,5 ; réducteur; riche en métaux) entraînant la précipitation de nombreux composés et créant ainsi le panache caractéristique qualifié de « fumeur noir ». A terme, la précipitation du fluide en sulfures polymétalliques et sulfate de calcium (anhydre) entraîne la construction des édifices communément appelés « cheminées hydrothermales » (Figure 0.11).



Figure 0.11. Cheminée hydrothermale avec panache noir (type ‘fumeur noir’), structure caractéristique des sources hydrothermales desquelles jaillit le fluide hydrothermal.

Lorsque le fluide subît une dilution préalable au cours de sa remontée, la précipitation des éléments dissous se produit partiellement au sein de la croûte océanique (séparation de phases) puis le fluide, appauvri en éléments métalliques et enrichi en sulfures d’hydrogène, est émis à des températures plus faibles (variable en fonction de la dilution) à travers le réseau de fissures situé à l’interface croûte / eau de mer. Il forme alors une zone de diffusion (Lowell et al. 1995). La composition chimique du fluide hydrothermal dépend d’un très grand nombre de facteurs tels que la profondeur de la chambre magmatique, la nature des roches et leur état d’altération (Campbell et al. 1988, James et al. 1995), la présence ou non de sédiments qui apportent des constituants dérivés de la matière organique (You et al. 1994). L’ensemble de ces processus donne naissance à une gamme très étendue de manifestations hydrothermales, avec notamment des fluides de composition variée.

Les cadavres de baleine et les bois coulés constituent le troisième environnement réduit profond, riche en matières organiques et qui présente une faune relativement proche de celle des deux précédents (Smith et al. 1989, Kiel & Goedert 2006). Les bois coulés sont rencontrés à toutes les profondeurs, particulièrement autour des îles tropicales, avec une plus forte densité devant les estuaires et au niveau des bassins d’accumulation sédimentaire. Ces

substrats végétaux d'origine terrestre et, paraissent importants en termes de quantité et de surfaces disponibles pour la colonisation. Ils permettent le développement d'écosystèmes « durables » (Cayré & Richer de Forges 2002) et pourraient jouer un rôle majeur dans le domaine profond en termes de refuges. C'est la dégradation bactérienne des substrats *in situ* qui va créer un milieu réducteur favorisant leur colonisation par des organismes chimiosynthétiques. Les carcasses de baleines sont également une ressource trophique pour les animaux chimioautotrophes profonds (Butman et al. 1995, Jelmert & OppenBerntsen 1996), ces organismes tirent alors profit des dégagements sulfurés issus des tissus en putréfaction. Les os des squelettes de baleines sont extrêmement riches en lipides (plus de 60%) et leur décomposition va être la source de nombreux éléments réduits qui seront utilisés par des bactéries chimioautotrophes (Deming et al. 1997). Les os de baleine, répartis çà et là sur les fonds, vont pouvoir jouer le rôle de « pierres de gué » dans les bassins océaniques profonds et, favoriser ainsi la dispersion de nombreux organismes associés aux environnements réduits (Distel et al. 2000).

Les trois environnements « réduits » que nous venons de présenter sont à l'origine d'écosystèmes particuliers : les écosystèmes chimiosynthétiques. Dans ces écosystèmes, la production primaire est assurée par des procaryotes chimioautotrophes mais aussi par des animaux endosymbiotiques, soit à partir du méthane (méthanotrophie) soit à partir des sulfures d'hydrogène (thioautotrophie). La production secondaire sera assurée par des procaryotes hétérotrophes, des eucaryotes hétérotrophes, des suspensivores, des dépositivores, des carnivores, des nécrophages....

Dans notre étude, nous nous sommes intéressés plus particulièrement aux écosystèmes hydrothermaux car la variabilité des conditions physico-chimiques et l'instabilité du milieu, constitue un modèle idéal de fonctionnement en métapopulations. De plus, la linéarité de cet environnement (circonscrit à l'axe des dorsales océaniques) permet de simplifier notre perception de la dispersion et éventuellement des processus de spéciation s'y déroulant.

8. Un milieu contrasté, riche en microhabitats et potentiellement favorable à la sélection diversifiante

A l'échelle du globe, les environnements réduits sont rencontrés sur une vaste gamme de profondeurs. Des zones de suintements froids ont été observées depuis des profondeurs de

80 m (Kagoshima Bay : (Hashimoto et al. 1993)) jusqu'à plus de 6000 m (zone de subduction du Japon KAIKO-NANKAI) (Sibuet & Olu 1998). Les zones présentant une activité hydrothermale sont situées depuis quelques mètres de profondeur dans les zones côtières (Golfe de Naples) jusqu'à plus de 4200 m (Dorsale médio atlantique) (Desbruyères et al. 2006). A une échelle plus locale, les différences de profondeur restent très importantes, même entre des sites situés à quelques kilomètres de distance : au niveau du Golfe du Mexique, un gradient de 3100 m de profondeur sépare les sites situés sur la pente continentale de Louisiane (Louisiana Slope : 400 m) et de ceux trouvés au niveau de l'Escarpement de Floride (Florida Escarpment : 3500m) et, sur le dorsale médio Atlantique plus de 1500 m de dénivelé séparent les sites hydrothermaux proches du point triple des Açores, Menez Gwen (840m) et Rainbow (2350m). Les conditions environnementales et les communautés d'espèces varient fonction de la profondeur. La pression hydrostatique (pression de l'eau) pourrait exercer une pression de sélection sur les espèces car aucune espèce n'est adaptée à des amplitudes de pressions allant de quelques atmosphères jusqu'à plusieurs centaines d'atmosphères. Des travaux sur les poissons abyssaux ont pu en effet montré des clines de fréquences alléliques en relation avec la profondeur (Siebenaller & Somero 1978). La pression liée à la profondeur va également jouer sur la composition chimique des fluides en modifiant la solubilité de certains éléments comme par exemple celle du méthane. Cette modification de la composition chimique de l'environnement va avoir un impact direct sur l'adaptation des organismes, notamment au niveau des processus de détoxification, et, sur leur plasticité physiologique notamment en termes de symbioses.

Au niveau trophique, les apports nutritifs provenant de la production de surface vont se trouver nettement réduits à grande profondeur. Cela aura une influence directe pour certaines espèces (comme les bathymodiales) qui présentent une activité partiellement ou totalement suspensivore ou détritivore. De plus, l'atténuation progressive de la saisonnalité des apports nutritifs de surface aura des conséquences indirectes sur la rythmicité de la reproduction, les animaux des grandes profondeurs présentant souvent une forte asynchronie dans leur reproduction (Gage & Tyler 1991). De même, la réduction puis la disparition de la lumière avec la profondeur va limiter la production primaire photosynthétique aux couches euphotiques et donc limiter l'accès à cette ressource aux espèces de filtreurs les moins profondes.

La modification des paramètres physico-chimiques du milieu et l'adaptation des animaux à la profondeur va se traduire par des pressions de sélection sur certains paramètres physiologiques. Au niveau de la respiration par exemple, l'acquisition d'une symbiose thio-oxydante impose aux organismes hôtes d'être soumis à de fortes conditions anoxiques et, donc des adaptations respiratoires ont été nécessaires pour survivre dans ces milieux (Toulmond 1990). En effet l'hydrogène sulfuré (H_2S) très présent dans les milieux réduits est un antagoniste de l'oxygène concernant la fixation et le transport par l'hémoglobine. Chez les animaux côtiers, la présence de H_2S et sa fixation sur les molécules d'hémoglobine va entraîner une nette diminution de l'efficacité respiratoire et va donc perturber l'ensemble du fonctionnement métabolique. La plupart des invertébrés des milieux réduits présentent une hémoglobine capable de lier réversiblement l' O_2 nécessaire à leur propre respiration, mais aussi le H_2S nécessaire aux symbiotes (Arp & Childress 1981). Le milieu hydrothermal est également un milieu très riche en métaux lourds (Seyfried & Janecky 1985) et en éléments radioactifs (Kadko 1996) connus pour être des éléments mutagènes de l'ADN. A cet égard, Cherry et al. (1992) ont montré que les annélides hydrothermaux étaient capables d'accumuler dans leurs tissus des doses d'éléments radioactifs 1000 fois supérieures à celle trouvées chez les organismes vivant dans des zones côtières contaminées (e.g. rejets de la Hague). Pour permettre la survie des espèces, il leur est donc nécessaire de disposer d'adaptations physiologiques leur permettant d'acquérir une machinerie cellulaire très performante permettant de renforcer le contrôle du cycle cellulaire et de réparer plus efficacement l'ADN (Pruski & Dixon 2003). Toutes ces adaptations physiologiques résultent potentiellement de la sélection de mutations favorables au niveau des gènes impliqués dans ces fonctions. On peut donc logiquement en déduire que les micro-habitats et les conditions difficiles qu'impose le milieu hydrothermal, pourraient induire de fortes pressions de sélection, et peut-être favoriser une récurrence de balayages sélectifs pour des caractères ou des fonctions dont la majorité nous est encore inconnue. De tels processus devraient avoir un impact sur la diversité génétique de ces organismes.

9. Le peuplement associé aux sources hydrothermales

Le compartiment bactérien est très diversifié et constitué de Bacteria et d'Archaea. Des métabolismes très divers sont représentés, chimioautotrophes et hétérotrophes, anaérobies et aérobies. Les bactéries peuvent être libres dans la zone de mélange entre le fluide hydrothermale et l'eau de mer, fixées sur des substrats durs comme les roches basaltiques, les

coquilles ou des tubes d'invertébrés (Karl et al. 1980, Jannasch 1985, Jannasch & Mottl 1985, Prieur et al. 1989). Elles peuvent aussi vivre en symbiose avec certains invertébrés de l'écosystème. Le principal intérêt des endosymbioses chimioautotrophes est de réaliser un réseau trophique raccourci, augmentant ainsi son rendement et son efficacité (Childress & Fisher 1992). Ce phénomène de symbiose est, par définition, à bénéfice réciproque, pour les 2 parties (hôte et symbionte). Les bénéfices de l'hôte seront la nutrition « à domicile » et la détoxification, alors que les bénéfices pour le symbionte seront une protection aux conditions fluctuantes de l'environnement et un accès stabilisé aux substrats. Il existe 2 grands types de transmission des symbiontes : le mode vertical (héritabilité totale avec les génomes parentaux) et la transmission horizontale (par contamination environnementale). Dans les modes de transmission verticaux avec transfert des symbiontes par les ovocytes de l'hôte, il va se créer un déséquilibre nucléo-cytoplasmique et des phénomènes de sélection entre les gènes cytoplasmiques provenant de l'hôte et les génomes des symbiontes transmis (Sanchez et al. 2000, Sanchez & Hastings 2003, Hurst & Jiggins 2005). A ce stade, des gènes peuvent être transférés du symbiote vers l'hôte (réduction du génome du symbiote) dans la voie d'internalisation. C'est le cas des symbioses retrouvées avec les bivalves du genre *Calyptogena* (Cary 1994, Peek et al. 1998). Au contraire, lors des transmissions horizontales, le choix des symbiontes et la contamination est surtout liée à l'environnement (Won et al. 2003c) et n'entraîne pas nécessairement de phénomènes de déséquilibre nucléo-cytoplasmique ou de co-spéciation avec les symbiontes. Le mode de transmission horizontale le mieux décrit est celui observée chez le vestimentifère emblématique des sources, *Riftia pachyptila* où les symbiontes thio-oxydants envahissent la post-larve par voie mésodermique et induisent la perte du tube digestif de la larve). Ce mécanisme se traduit par une absence totale de co-évolution entre l'hôte et le symbionte (Nelson & Fisher 2000). Les espèces de bathymodiolineae possèdent un large spectre d'associations symbiotiques car la moitié d'entre elles possèdent uniquement des symbiontes sulfo-oxydants, un petit nombre d'espèces (3) ne possède que des symbiontes méthanotrophes et 6 espèces possèdent les 2 types de symbiontes (Won et al. 2008). Il a été montré chez la bathymodiole atlantique, *Bathymodiolus azoricus* qui possède une double symbiose, que l'acquisition des symbiontes méthanotrophes se faisait depuis l'environnement dans lequel elles vivent (Won et al. 2003c). Pourtant, l'observation des cellules branchiales de post-larves chez *B. azoricus* a permis d'observer des cellules bactériennes qui ont été identifiées comme des symbiontes sulfo-oxydants et méthanotrophes (Salerno et al. 2005), laissant penser à un transfert horizontal s'effectue à un stade précoce du développement, (Won et al. 2003c).

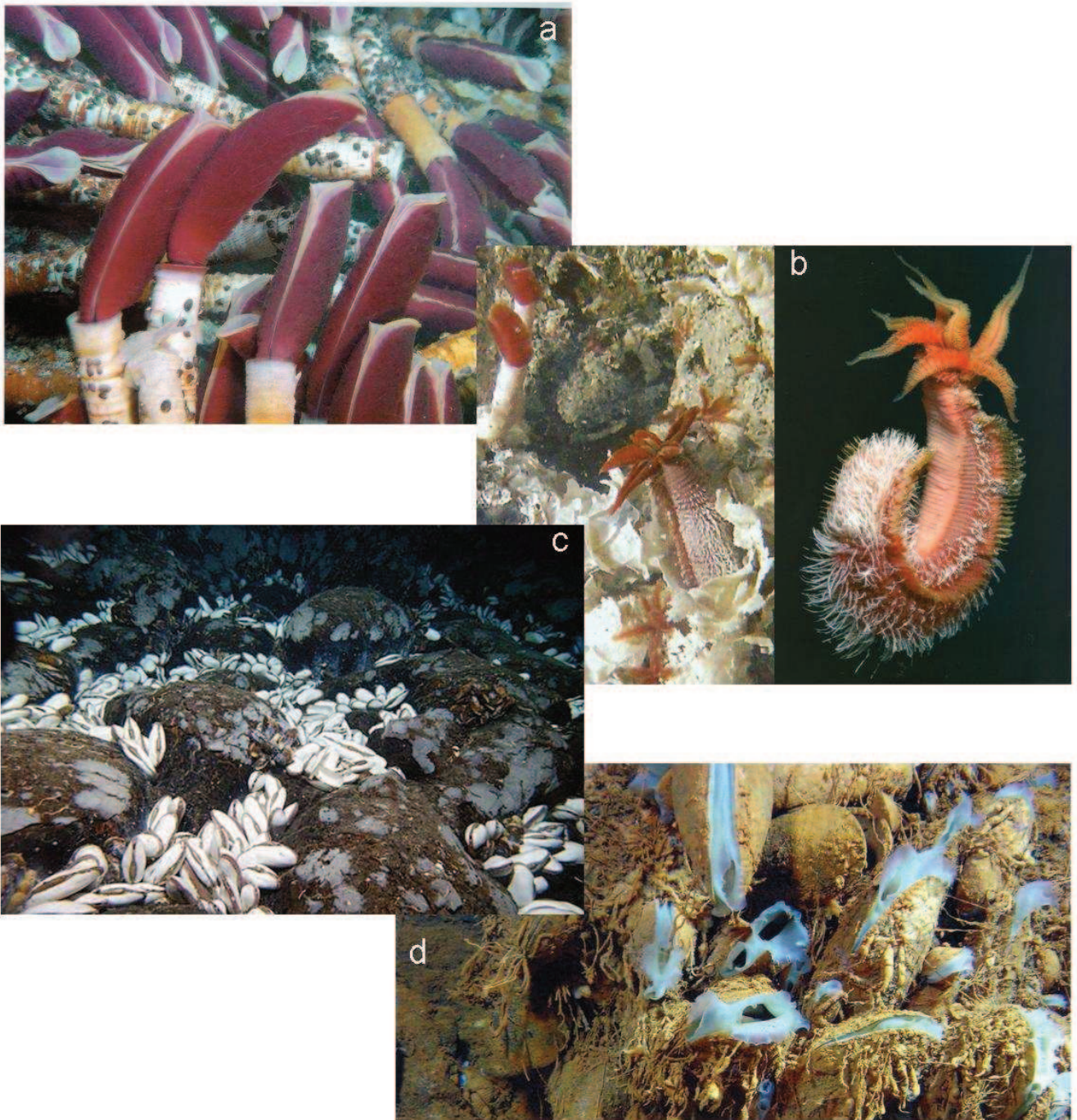


Figure 0.12. Illustration des 2 principaux groupes taxonomiques de la macrofaune associée à l'environnement hydrothermal : les polychètes et les bivalves. (a) *Riftia pachyptila* en bouquet (13°N, EPR, campagne Phare© Ifremer) ; (b) *Alvinella pompejana* ; à gauche : colonie *in situ* (13°N, EPR, campagne Phare© Ifremer) ; à droite : animal hors de son tube (par Dugornay, © Ifremer) ; (c) *Calyptogenia magnifica* dans une zone de diffusion (Ride des Galápagos, photo Tim Shank) ; (d) *Bathymodiolus azoricus* en colonie et recouvertes de sédiments (Rainbow, MAR, campagne Atos© Ifremer).

Une liste exhaustive et actualisée des espèces a été publiée par Desbruyères et collaborateurs en 2006 (Desbruyères et al. 2006). Deux grands groupes peuvent cependant être distingués par leur diversité et leur abondance : les polychètes et les mollusques. (gastéropodes + bivalves)

9.1. Les polychètes

Les vers polychètes avec 111 espèces décrites constituent 18-20% du nombre total d'espèces actuellement identifiées dans les communautés des sources hydrothermales (Desbruyères et al. 2006). La grande majorité de ces espèces (94%) sont considérées comme endémiques des sources hydrothermales, moins de 4% sont des espèces communes aux environnements réduits profonds et les derniers 2% sont des espèces abyssales « classiques ».

Les polychètes comprennent les vestimentifères, vers géants tubicoles emblématiques des sources hydrothermales : *Riftia pachyptila* Jones 1981. Ces vers appartiennent à l'ordre des Sabellida et à la classe des Siboglinidés. Les vestimentifères vivent dans un tube de nature chitineuse (Gaill & Hunt 1986) et forment de véritables « bouquets » dans les zones de diffusion sur l'EPR et les zones de suintements froids du Golfe du Mexique (Figure 0.12-a). Ils sont dépourvus de tube digestif et dépendent donc totalement de bactéries chimioautotrophes endosymbiotiques localisées dans les cellules du trophosome, organe spécialisé et vascularisé (Cavanaugh et al. 1981, Felbeck 1981). Ils maintiennent leur panache branchial dans le flux hydrothermal de façon à y prélever l'O₂, élément nécessaire à leur métabolisme ainsi que ceux nécessaires à leurs bactéries, tels que NH₄⁺, les sulfures et le CO₂ à l'aide de leurs hémoglobines extracellulaires en bi-couche hexagonale (Chabasse et al. 2006).

Un deuxième groupe de polychètes est emblématique des sources hydrothermales : il s'agit de la famille des Alvinellidae. Leur habitat, répartis dans un gradient thermique le long des cheminées hydrothermales est considéré comme extrême, et a contribué à dresser les espèces de cette famille au rang des espèces dites « modèles ». L'espèce la plus connue de la famille est sans conteste *Alvinella pompejana* Desbruyères et Laubier 1980, le fameux « ver de Pompei ». Cette espèce grégaire est aussi très étudiée, aussi bien du point de vue de son écologie (Jollivet 1993, Chevaldonné & Jollivet 1993), de sa reproduction (Faure et al. 2007; article en annexe 4) , sa résistance à la température (Cary et al. 1998) ou de sa structure

génétique (Jollivet et al. 1995a, Jollivet et al. 1998, Hurtado et al. 2004). *Alvinella pompejana* est également considérée comme le métazoaire le plus thermorésistant au monde puisse qu'il a été observé comme pouvant résister sans dommage à de brefs expositions à une température supérieure à 100°C (Chevaldonné et al. 1992, Cary et al. 1998). Les animaux de cette espèce vivent dans un tube parcheminé dont ils peuvent occasionnellement sortir (Figure 0.12-b), et qui au contraire de celui des vestimentifères ne contient pas de chitine. Les Alvinellidae passent le plus clair de leur temps dans leur tube avec les branchies et l'appareil buccal tendu vers l'extérieur du tube. Ils possèdent un régime alimentaire détritivore et ingèrent de nombreuses particules minérales mélangées à de la matière organique mais ils semblent qu'ils fassent des excursions régulières hors de leur tube pour aller se nourrir de bactéries filamenteuses libres qui se développent à la surface de la colonie (Alayse-Danet et al. 1985, Alayse-Danet et al. 1987). Ils possèdent également des épibiontes bactériens le long de soies dorsales dont la fonction est encore mal connue mais qui pourraient jouer un rôle dans la détoxification (Gaill et al. 1987, Gaill et al. 1988, Alayse-Danet et al. 1987).

9.2. Les bivalves

Les bivalves des environnements réduits se répartissent selon quatre grandes familles : Les Solemyidae, les Vesicomysidae, les Mytilidae et les Pectinidae (Desbruyères et al. 2006). Les Vesicomysidae et les Mytilidae sont les familles regroupant le plus grand nombre d'espèces.

La famille des Solemyidae est représentée dans les sources hydrothermales et les suintements froids par le genre *Acharax*. Les espèces de cette famille sont des bivalves de relativement grande taille (de 10 à 22cm) vivant profondément enfouis dans des sédiments meubles (zones hydrothermales à forte sédimentation comme le bassin des Guaymas ou les bassins arrière-arc du Pacifique ouest : (Métivier & von Cosel 1993). Certaines de ces espèces sont caractérisées par l'absence totale de tractus digestif et sont donc complètement dépendantes de leur symbiose avec des bactéries chimioautotrophes.

Pour le moment, la famille des Pectinidae des milieux profonds est uniquement représentée par 2 espèces : *Bathypecten vulcani* et *Sinepecten segonzaci*. La première a été observée en plusieurs points de la dorsale du Pacifique Est (13°N et 38°S) ainsi que sur la ride des Galápagos. La seconde espèce a été échantillonnée dans le bassin arrière-arc de Manus.

Considéré comme des fossiles vivants du Crétacé, peu de choses sont connues sur ces espèces en général sinon qu'il possède un tractus digestif fonctionnel et ne présentent pas d'endosymbioses particulières. En effet, la petite taille des ces bivalves (moins de 30mm) fait qu'ils sont extrêmement difficile à échantillonner et par conséquent à étudier.

La famille des Vesicomidae est la plus diversifiée, en taille, en morphologie et en nombre d'espèces : actuellement, 84 espèces ont été recensées mais il reste encore de nombreuses espèces actuellement en cours de description et la famille est très nettement polyphylétique suggérant une remise en question des critères morphologiques de classification (Vrijenhoek et al. 1994).

La famille comprend un petit nombre d'espèces inféodées au milieu profond mais n'est pas restreinte aux habitats réduits. En revanche un grand nombre d'espèces est confinée aux environnements réduits, sources hydrothermales et suintements froids. Certaines espèces sont même retrouvées à la fois dans les environnements hydrothermaux et dans les zones de suintements froids. Ces coquillages sont connues pour dominer les assemblages d'invertébrés trouvés dans les zone hydrothermales, et les zones de suintements froids à travers tous les océans du monde (Sibuet & Olu 1998)(Figure 0.12-c). Plus de 50 espèces ont été décrites à travers le globe et de nombreuses espèces cryptiques ont été identifiées par l'outil moléculaire (Goffredi et al. 2003, Vrijenhoek et al. 1994). Ces espèces sont toutes hautement adaptées à la vie symbiotique et dépendent complètement des bactéries chimioautotrophes qui se développent dans leurs branchies pour leur nutrition (Cavanaugh 1983, Fisher 1990). Les sulfures nécessaires au métabolisme bactérien sont absorbés au niveau du pied, puis transportés par le sang à l'aide d'une protéine enrichie en Zn encore non identifiée, tandis que l'O₂ et le CO₂ sont absorbés directement au niveau des branchies (Arp et al. 1984). Même si les Vesicomidae sont restreints aux substrats enrichis en sulfures, la plupart des espèces ne possèdent pas les mêmes caractéristiques écologiques. Les grandes différences existantes entre les espèces concernant leurs adaptations physiologiques vis-à-vis des sulfures, du méthane et de l'oxygène, celles-ci les ayant conduit vers une forte spécialisation à des microhabitats (Goffredi & Barry 2002)

Au sein de la famille des Mytilidae, 2 grandes sous-familles réunissent l'ensemble des modioles des sources hydrothermales et des zones de suintements. Le premier groupe est celui des Bathymodiolinae, et comprend les genres *Bathymodiolus* Kenk et Wilson 1985,

Gigantidas Cosel et Marshall 2003 et *Tamu* Gustafson et al 1998. La seconde sous famille est celle des Modiolinae avec les genres *Idas* Jeffrey 1876 , *Adipicola* Dautzenberg 1927 et *Idasola* Iredale 1939 (Tableau 0.1).

Tableau 0.1. Classification des Mytilidae associés aux environnements réduits profonds

Domaine	Eukaryota Whittaker & Margulis, 1978						
Royaume	Metazoa Linnée 1758						
Sous royaume	Bilateria Hatschek 1888						
Embranchement	Protostomia Grobben 1908						
Phylum	Mollusca Linnée 1758						
Classe	Bivalvia Linné 1758						
Sous classe	Pteriomorphia Beurlen 1944						
Ordre	Mytiloidea Ferussac 1822						
Super famille	Mytiloidea Rafinesques 1815						
Famille	Mytilidae Rafinesques 1815						
sous famille	Bathymodiolinae Kenk et Wilson 1985			Modiolinae keen 1958			
Genre	Bathymodiolus Kenk et Wilson 1985	Gigantidas Cosel et Marshall 2003	Tamu Gustafson et al 1998	Benthomodiolus Dell 1987	Idasola Iredale 1939	Adipicola Dautzenberg 1927	Idas Jeffrey 1876

La répartition et le mode de vie des espèces de mytilidés hydrothermaux sont très variables. Par exemple, le genre *Idas* regroupe des petites moules trouvées associées à des bois coulés ou à des os de baleine dans l'océan Pacifique et la mer Adriatique (Smith & Baco 2003, Pailleret et al. 2007) ainsi que des espèces associées aux suintements froids du Golfe du Mexique et de Méditerranée (Craddock et al. 1995a). Le genre *Bathymodiolus* quant à lui comprend des espèces de grande taille, qui sont distribuées à l'échelle mondiale, mais qui sont uniquement inféodées aux sources hydrothermales ou aux zones de suintements froids (Iwasaki et al. 2006, Jones et al. 2006)(Figure 0.10-d).

Si les moules Bathymodiolinae sont parmi les animaux qui dominent les communautés associées aux sources hydrothermales et aux suintements froids à l'échelle du globe, cela est dû en particulier à son régime mixotrophe associant une endosymbiose avec des bactéries symbiotiques chimiosynthétiques au niveau de leurs cellules branchiales avec un système de filtration fonctionnel (Fiala-Médioni et al. 1986, Fiala-Médioni & Le Pennec 1987). En effet dans ces environnements profonds, la matière organique en suspension dont se nourrissent les Mytilidae côtiers est particulièrement rare. Les Bathymodiolinae (vivant de 80m à 4500 m de profondeur) ont vraisemblablement conservé un régime mixotrophe, ce qui a certainement énormément contribué à leur succès écologique (Fisher et al. 1987).

Il a été proposé que les lignées des sources hydrothermales profondes dériveraient d'anciennes lignées originaires des zones de suintement froids et d'environnements peu profonds des bois coulés (Craddock et al. 1995a, Distel et al. 2000, Samadi et al. 2007). Cette colonisation des environnements profonds aurait pu faire suite aux extinctions massives ayant eu lieu lors de phases anoxiques à l'Eocène (Jacobs & Lindberg 2001). La recolonisation des sources hydrothermales aurait alors pu avoir lieu à partir d'espèces moins profondes ayant su tirer profit d'environnements réduits temporaires ou habitats relais, composés par les bois coulés et les os de baleine (Distel et al. 2000). Cependant, cette hypothèse s'appuie uniquement sur une phylogénie établie à partir d'un nombre limité d'espèces provenant principalement de l'océan Atlantique. D'autres études phylogénétiques montrent que certaines espèces du Pacifique Ouest pourraient être relictuelles (notamment en Nouvelle Zélande : (Smith & Baco 2003)) et ne permettent pas de définir un état ancestral et un état dérivé entre les environnements hydrothermaux et les zones de suintements froids (Miyazaki et al. 2004). De plus, des reversions d'habitats (hydrothermal vers suintements froids) ont pu se produire dans certaines lignées de bathymodiales Atlantiques (Won et al. 2002).

En contexte hydrothermal, les mytilidae du genre *Bathymodiolus* sont présents sur l'ensemble des champs hydrothermaux de la ride médio-Atlantique, sur une partie de la dorsale du Pacifique Oriental (de 13°N à 37°S), sur la ride des Galápagos, dans les bassins arrière-arc du Pacifique Ouest et dans l'océan Indien au niveau du point triple de Rodriguez.

Malgré un nombre d'études unique dans l'histoire de l'exploration de l'océan profond, les contraintes logistiques font que nous sommes encore bien loin de connaître la distribution exacte, la composition et le fonctionnement des écosystèmes hydrothermaux. En dépit du nombre encore trop grand de zones peu ou pas explorées (Atlantique Sud, dorsale Indienne, dorsales Arctique et Antarctique), certaines zones ont fait l'objet d'études approfondies et de suivis temporels (Dorsale du Pacifique Est, dorsale nord Pacifique (segment de Juan de Fuca), Point triple des Açores) qui, ont permis de mettre en évidence que le milieu hydrothermal est soumis à une très forte instabilité temporelle liée au volcanisme effusif (épanchement de laves couvrant de nombreuses zones actives : Haymon et al. (1991) et Tunnicliffe et al. (1997)) ou à une forte activité tectonique (Jollivet 1993). Cette instabilité de l'habitat et sa fragmentation à toutes les échelles d'espace affectent grandement le fonctionnement des populations et la composition spécifique de ces communautés.

9.3. Le milieu hydrothermal: un milieu fragmenté et instable

Les dorsales océaniques forment la plus grande chaîne de montagnes et le plus grand système volcanique actif du système solaire (Macdonald et al. 1991). Le réseau de dorsales est continu à l'échelle du globe sur près de 70 000 km (figure 9), mais l'activité hydrothermale n'est pas homogène car elle est liée à la présence de réservoirs magmatiques sous l'axe de la dorsale. Les champs hydrothermaux se répartissent donc sur la dorsale à la manière d'un collier de perles et peuvent être séparés par plusieurs centaines de kilomètres (Chevaldonné et al. 1997). De nombreuses ruptures d'axe (failles transformantes ou simples discontinuités géologiques) affectent également la linéarité de la dorsale et la circulation des masses d'eau le long de la vallée axiale (Cannon & Pashinski 1990, Mullineaux et al. 2002). Ces failles transformantes résultent de mouvements de cisaillement liés à l'activité tectonique et sont perpendiculaires à l'axe principal de la dorsale. Elles consistent en un décalage latéral de plusieurs kilomètres à quelques centaines de kilomètres et créent ainsi une réelle barrière physique à la dispersion des espèces hydrothermales. Les segments de dorsale, isolés les uns des autres, peuvent avoir des longueurs comprises entre 14 et 600 km (Macdonald et al. 1991) et pourraient être considérés comme des unités indépendantes concernant le fonctionnement de leurs communautés (dispersion et flux géniques limités).

A une échelle plus fine, c'est-à-dire au sein d'un champ hydrothermal (de l'ordre de plusieurs kilomètres) il existe également une forte hétérogénéité dans la distribution des sites capable d'influencer le fonctionnement des populations. En effet la durée de vie des cheminées dépend de la vitesse d'accrétion des dorsales océaniques et des événements volcaniques et tectoniques actuels : depuis leur découverte, plusieurs éruptions ont déjà détruit de nombreux sites en ensevelissant des pans entiers de champs hydrothermaux par des coulées de lave (Haymon et al. 1991, Tunnicliffe et al. 1997, Ferrini et al. 2007). Plus la vitesse d'accrétion est rapide, plus le milieu est instable dans le temps et dans l'espace. Des modèles de convection thermique à travers un milieu poreux montrent que l'activité se déplace le long de l'axe générant un perpétuel remaniement des sites au niveau d'un champ (Watremez & Kervévan 1990). De plus, les zones actives sont progressivement repoussées de part et d'autre de la zone d'accrétion (sur plusieurs millions d'années) et le flux hydrothermal se tarit jusqu'à s'arrêter complètement entraînant la disparition des communautés animales. Cette dynamique va donc entraîner un renouvellement régulier des habitats et des populations par un jeu d'épisodes récurrents d'extinction et de recolonisation (Jollivet et al. 1999). Le taux

de renouvellement des sites et les pressions sélectives associées sur les peuplements seront différents selon les océans et les dorsales étudiés. La dorsale médio-Atlantique, par exemple, est considérée comme relativement stable car elle possède une faible vitesse d'accrétion : 10 à 40 mm/an, (Macdonald et al. 1991). Les sites hydrothermaux ont une durée de vie de plusieurs milliers d'années (cf. Lalou et al. (1990), (Lalou et al. (1993)). La dorsale Est-Pacifique est au contraire considérée comme une dorsale ultra rapide avec des vitesses d'accrétion de 90 à 162 mm/an (Naar & Hey 1989, Demets et al. 1990). Le taux de renouvellement des sites et y sera nettement plus élevés que dans l'Atlantique avec une durée de vie moyenne des cheminées plus proche de la centaine d'années (Lalou et al. 1985).

10. Dispersion larvaire dans le milieu hydrothermal.

La majorité des Polychètes et Bivalves hydrothermaux possèdent un cycle de vie benthopélagique incluant une phase larvaire (McHugh 1989, Comtet & Desbruyères 1998, Thiébaud et al. 2002). Les expérimentations menées en laboratoire et *in situ* ont permis d'étudier les premiers stades de développement embryonnaire chez *Alvinella pompejana* (Pradillon et al. 2005a) et chez *Riftia pachyptila*, (Marsh et al. 2001). En revanche, aucune de ces études portant sur le développement n'a été menée jusqu'à l'obtention d'une larve compétente, une étape de la vie larvaire qui indique la fin de la vie pélagique et qui permettrait ainsi de déterminer avec précision la durée totale de la phase larvaire (Hadfield et al. 2001). Des études menées sur des échinodermes Antarctiques ont démontré que dans un milieu froid (-1,5°C), la durée de vie larvaire pouvait excéder 1 an (Marsh et al. 1999). Les larves des animaux hydrothermaux qui dispersent dans l'eau de fond (2°C) pourraient donc avoir une durée de phase larvaire du même ordre de grandeur. Une autre étude plus récente menée sur le développement d'une ascidie polaire montre que dans un milieu froid la durée de phase larvaire peut être jusqu'à 6 fois supérieure à celle des espèces tempérées (Strathmann et al. 2006), mais que cette durée est limitée par la dépense énergétique liée à la nage. La nutrition des larves pendant leur phase pélagique pourrait leur permettre d'allonger la durée de celle-ci. Cependant, le milieu abyssal baignant les sources/zones de suintements est oligotrophe et de nombreuses études ont montré que les larves d'espèces hydrothermales sont pour la plupart lecithotrophes (Tyler & Young 1999). C'est notamment le cas des larves du vestimentifère *lamellibrachia sp* (Young et al. 1996),. Une durée de vie larvaire à 38 jours a été estimée pour *Riftia pachyptila* (Marsh et al. 2001), mais à ce jour, la durée complète de cette phase larvaire en milieu hydrothermal demeure un mystère. Chez *Alvinella pompejana*,

la phase de dispersion ne se ferait pas pendant les stades larvaires, mais au niveau de l'ovocyte fécondé comme semblent le montrer les travaux de Pradillon et al. (2001).

La phase larvaire est une étape clef dans la dispersion. La dispersion passive de la larve peut se faire par le biais de la dispersion des panaches hydrothermaux le long des dorsales (Mullineaux et al. 1995) et par les courants de fond. Des hypothèses de remontée des larves dans la phase éclairée des océans ont été émises, notamment chez les espèces possédant des larves planctonotrophes comme les bathymodioles (Berger & Young 2006) ou les crustacés décapodes (Stevens et al. 2008). Il a en effet été montré que ces larves pouvaient se nourrir de la matière organique issue de la production de surface (Pond et al. 1997). La dispersion active par la nage de la larve semble négligeable (2 mètres par jour chez *R. pachyptila* (Marsh et al. 2001), mais, cette vitesse aussi faible soit-elle peut permettre aux larves de rejoindre les courants de fond, s'extraire des courants convectifs issus du mélange entre le fluide hydrothermal chaud avec la colonne d'eau froide dans laquelle il est injecté (Kim & Mullineaux 1998, Mullineaux et al. 2005). A l'inverse, les courants transverses (perpendiculaires à l'axe de la dorsale) peuvent être considérés comme des barrières à la dispersion. En effet, dans un milieu aussi linéaire que le milieu hydrothermal, toute larve expulsée hors de l'axe de la dorsale par un courant latéral n'a que peu de chance de trouver un habitat favorable à sa métamorphose. Les courants canalisés « dans l'axe » de la dorsale permettent une dispersion efficace sur de très longues distances (Mullineaux et al. 2005). Pourtant des études en laboratoire (Helfrich & Battisti 1991, Marsh et al. 2001) et sur le terrain (Speer & Marshall 1995, Joyce et al. 1998) montrent une faible dispersion des panaches hydrothermaux. Du fait des renverses successives du sens des courants de marée, une larve dont la durée de vie pélagique serait d'un mois n'aurait que très peu de chance d'être exportée à plus de 100 km de son point d'émission (Chevaldonné et al. 1997). Contre toute attente, il semble également que plus les larves passent de temps dans le courant plus elles ont de chances d'être balayées et exportée hors de la dorsale (Marsh et al. 2001). Des conclusions hâtives conduiraient à penser que le régime dynamique des sites et l'hydrodynamisme local exerceraient une sélection en faveur d'une phase larvaire courte afin de permettre à un maximum de larves d'être recrutées sur un habitat favorable. Il ne faut cependant pas oublier que la phase larvaire pélagique est le seul moyen qu'ont les espèces hydrothermales pour réussir à disperser sur un segment de dorsale se situant par exemple au delà d'une faille transformante. Il doit donc exister un compromis évolutif (trade-off) entre le fait de limiter la durée de la phase larvaire pélagique (et d'augmenter le nombre de

descendants recrutés à la future génération) et le fait de disperser loin pour réduire les risques d'extinction liés au régime volcano-tectonique des dorsales (Pechenik 1999).

Les études de génétique des populations constituent une approche indirecte pour permettre de tester les différentes hypothèses de dispersion en milieu hydrothermal. Jusqu'ici, deux grands modèles de structure de populations ont été testés : le modèle en « pas japonais » (stepping-stone model : (Kimura & Weiss 1964)) et le modèle en îles. (n-Island model : (Wright 1943), cf. Figure 0.13).

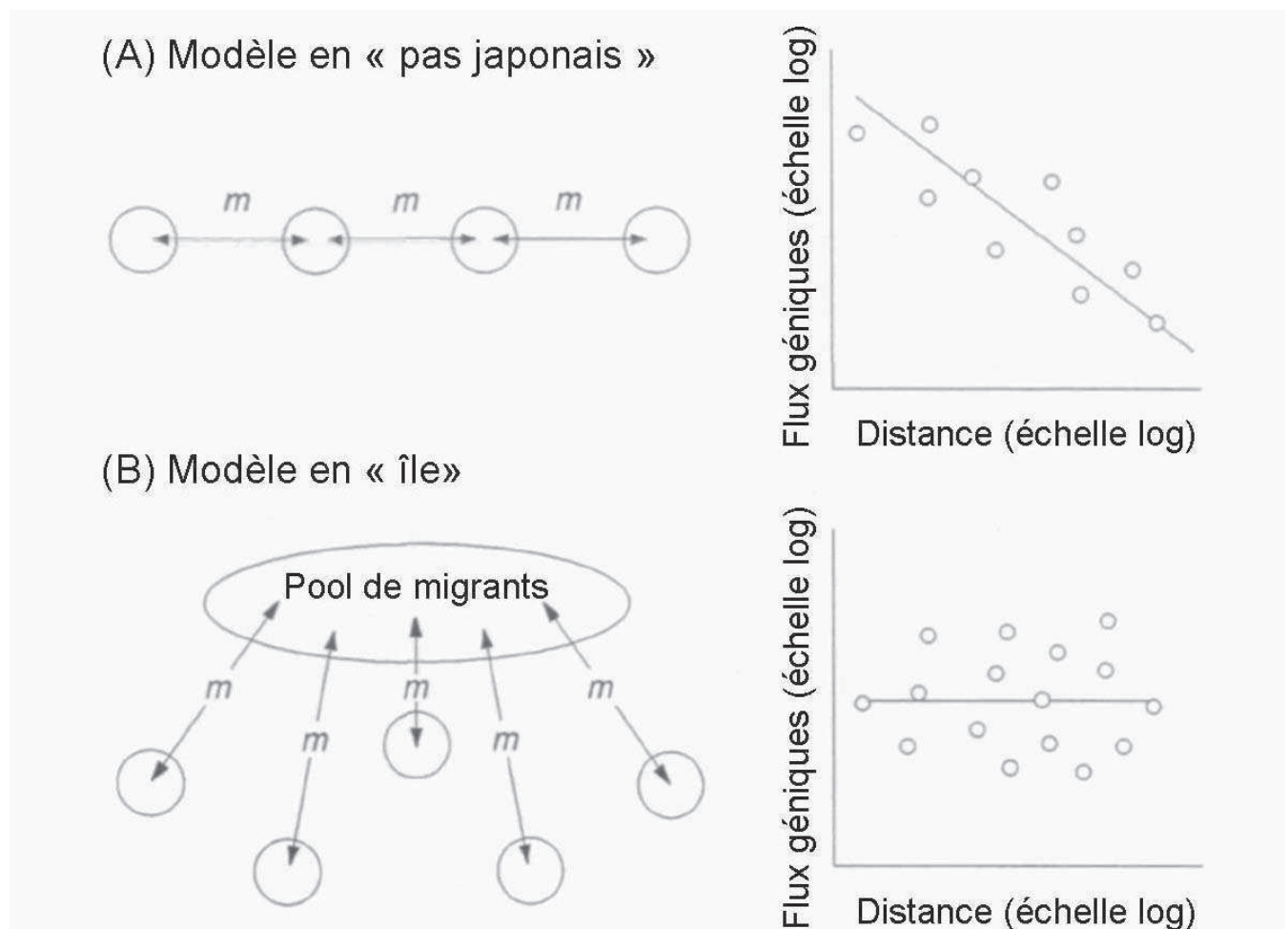


Figure 0.13. Modèles de structure de populations avec flux génétiques. (A) le modèle en « pas japonais » est approprié pour les espèces avec des capacités de dispersion limitées – Le niveau de flux génétiques entre les populations diminue avec l'augmentation de la distance géographique. (B) Le modèle en « îles » est approprié pour les espèces possédant de capacité de dispersion sur de longues distances – Le niveau de flux génétiques entre les populations est indépendant de la distance géographique et constant entre toutes les populations. (Schéma modifié de (Vrijenhoek 1997).

En d'autres termes, si aucune structure génétique n'est détectée ($F_{st} = 0$ (Wright 1951)), on peut en déduire que l'espèce possède une capacité de dispersion qui dépasse l'aire géographique échantillonnée, et, que l'espèce n'est pas sensible à la fragmentation de l'habitat (failles transformantes, courants, profondeur ne constituent pas de barrières aux flux de gènes) et/ou ne présente pas d'adaptations locales. En revanche, sous l'hypothèse migration/dérive, lorsqu'une différenciation génétique est observée entre 2 populations, elle peut être due à une caractéristique propre de l'espèce à coloniser de proche en proche (espèce incubante ou phase larvaire de courte durée), ou, elle peut être locus-spécifique et refléter d'une sélection différentielle sur les gènes. Les résultats obtenus sur le vestimentifère *Riftia pachyptila* (Bucklin 1988, Black et al. 1994, Vrijenhoek 1997, Shank & Halanych 2007) semblent être en faveur du modèle en pas japonais avec un fort isolement par la distance. En revanche, les études concernant la moule *Bathymodiolus thermophilus* sont beaucoup moins tranchées. Une première étude allozymique, (Grassle 1985) suggérait des différences génétiques entre individus selon leur zone de provenance (ride des Galápagos *versus* EPR) et leur taille, certains allèles étant fixés entre les sites. Une seconde étude allozymique (Craddock et al. 1995b) utilisant les mêmes sites échantillonnés ultérieurement concluait de façon opposée (aucune différenciation génétique à grande échelle) laissant supposer l'existence de fortes hétérogénéités temporelles dans la colonisation des sites. L'analyse du génome mitochondrial sur les mêmes individus confirmait l'absence de différenciation entre populations et privilégiait l'hypothèse de la dispersion sur de longues distances (Won et al. 2003b) Un deuxième bivalve, *Calyptogena magnifica* dont les œufs lécithotrophes de grande taille laissaient prédire une phase larvaire de courte durée (Lutz et al. 1986), ne présente pas d'isolement par la distance (Karl et al. 1996). Chez les polychètes Alvinellidae, Jollivet et al. (1995a) ont mis en évidence l'existence d'une différenciation par la distance à l'échelle d'un champ hydrothermal (quelques kilomètres) mais ne détectent pas de différences entre populations à une échelle spatiale plus grande (e.g. EPR nord), laissant supposer que des échanges existent entre populations distantes, soit en raison d'une stratégie dispersive à longues distances, soit en raison de la dynamique du système hydrothermal lui même (par un déplacement de l'activité favorisant le déplacement des champs. L'étude génétique de (France et al. 1992) sur l'amphipode *Ventrella sulfuris* à développement direct montre que la profondeur et l'orientation des courants peuvent avoir un effet très prononcé sur l'isolement des populations en créant des barrières à la dispersion. La nature éphémère des sources hydrothermales et l'instabilité du système conduisent à une récurrence d'extinction/recolonisations qui n'est pas sans conséquence sur la structure génétique des

espèces (Jollivet et al. 1999). Selon le mode de reproduction et la stratégie dispersive de l'espèce, ces processus ont un impact important sur le fonctionnement de la métapopulation et peuvent tantôt promouvoir les flux géniques longues distances (*Alvinella*), tantôt la différenciation génétique à faible échelle (*Bathymodiolus*). Jollivet et al. (Jollivet et al. 1999) ont en effet montré qu'il existait une fenêtre étroite des paramètres taux d'extinction/recolonisation et durée de vie larvaire permettant d'expliquer à la fois le maintien du polymorphisme et l'homogénéité génétique des populations à grandes distances. Ces facteurs joueront également un rôle non négligeable dans la micro-structuration des populations hydrothermales. En effet, l'hétérogénéité locale observée par (Grassle 1985) chez *B. thermophilus* a été interprétée comme étant un effet Wahlund temporel entre différentes classes de taille suggérant une colonisation récente du site, cet effet devant disparaître dès le premier épisode de reproduction « croisée » au sein du site. En revanche, une micro-structure génétique peut également correspondre à une sélection différentielle par l'habitat comme cela a pu être montré chez *Alvinella pompejana* sur les allèles de la phosphoglucomutase (Jollivet et al. 1995b; Piccino 2004). L'utilisation de nouveaux outils génétiques (type AFLP, génomic scans) plus informatif à l'échelle de l'individu, pourraient permettre une étude plus précise des différenciations locales. De telles méthodes ont permis de mettre en évidence la présence d'individus migrants de première génération entre populations éloignées de *Riftia pachyptila* (Shank & Halanych 2007). La structure génétique trouvée correspondait à celle des cohortes larvaires, qui aurait été maintenue grâce à une rétention hydrodynamique favorisant l'autorecrutement local et la différenciation entre populations, ceci en accord avec le modèle hydrodynamique développé par Marsh et al. (2001).

Il semble qu'il soit difficile de tirer des conclusions concernant la dispersion des espèces hydrothermales, on peut cependant proposer 3 conclusions principales : (1) la taille des œufs et le caractère planctonotrophe/lecithotrophe du développement ne constituent pas des outils prédictifs pour inférer une distance de dispersion, (2) toutes les espèces ne possèdent pas la même sensibilité au niveau de fragmentation de l'habitat hydrothermal, (3) l'environnement et les conditions physico-chimiques jouent clairement un rôle dans l'isolement des populations et permettent probablement des phénomènes l'adaptation locale à des microenvironnements.

11. Zones d'étude et objectifs de la thèse

11.1. La dorsale est pacifique

L'océan Pacifique comprend quatre dorsales actives : la dorsale du Pacifique-est, *East Pacific Rise* (EPR) (dorsale rapide à ultra-rapide), la dorsale Cocos-Nazca au large de l'Amérique centrale, la dorsale du Chili au sud de l'île de Pâques et au nord-est, la dorsale du Pacifique Nord composée des segments Gorda, Juan de Fuca, Endeavour et Explorer, le segment Gorda étant connecté à la terminaison nord de la faille de San Andreas. Notre zone d'étude est plus particulièrement centrée sur la Dorsale est Pacifique. L'axe de cette dorsale constitue un dôme régulier. Cette structure est liée au régime thermique de la dorsale et donc à l'épaisseur de la lithosphère (Lagabrielle 2005). Sur l'EPR comme pour l'ensemble des dorsales rapides, la lithosphère se trouve au-dessus du réservoir magmatique et la dorsale ne représente que la simple fente (à l'échelle de la plaque) d'où s'échappe le magma, résultant de la séparation des plaques. Il n'y a pas effondrement du dôme ce qui confère à l'EPR cet aspect régulier. Cependant 20% de la dorsale est caractérisé par une caldera plus ou moins profonde (effondrement consécutif à la vidange de la chambre magmatique sous jacente), donnant une zone de fracture (*graben*) tectonisée.

C'est entre la zone de Fracture « Garrett » (13°S) et la microplaque de l'île de Pâques (23°S) que le taux d'accrétion est le plus rapide au monde (entre 141 et 162 mm/an) (Naar & Hey 1989, Demets et al. 1990).

L'axe de l'EPR est très segmenté (Macdonald et al. 1991), il est découpé en tronçons séparés par des failles transformantes de grande ampleur ou de discontinuité de moindre importance (Figure 0.14). Lorsque les extrémités des axes de deux segments de dorsale se recouvrent partiellement (chevauchement à certaines latitudes), certains utilisent le mot anglais de OSC (*Overlapping Spreading Center*) (Macdonald & Fox 1983). (cf schéma explicatif sur Figure 0.14). C'est ce phénomène qui s'est produit à plus grande échelle lorsque les extrémités de deux axes actifs de la Dorsale du Pacifique est se sont rejoints pour isoler les microplaques de l'île de Pâques et de Juan Fernandez (Larson et al. 1992).

La dorsale du Pacifique est se révèle être un réseau hétérogène et des études sur la répartition des espèces et leur phylogéographie ont permis de mettre en évidence une

succession de barrières à la dispersion. Ainsi, du nord au sud, quatre grandes barrières ont pu être identifiées

(1) La barrière se situant au niveau de la zone de fracture de Riviera (cf Figure 0.14) a été déterminée par l'observation d'un cline de fréquences des haplotypes mitochondriaux chez le polychète *Oasisia alvinae* (Hurtado et al. 2004). Il s'agit également de la limite nord de répartition pour de nombreuses espèces comme l'annélide *Tevnia jerichonana* ou la modiole *Bathymodiolus thermophilus*. Des courants gyroïres forts au sein de cette région pourraient entraîner les larves de certaines espèces hors de l'axe de la dorsale.

(2) La barrière du Hess Deep : point triple avec la ride des Galápagos : la région entre la dorsale du Pacifique est et celle des Galápagos présente des caractéristiques hydrographiques et topographiques pouvant créer une barrière à la dispersion. En effet, certaines espèces (par exemple *Bythograea galapagensis*) sont endémiques des Galápagos et d'autres comme *Alvinella pompejana* ou le crabe *Cyanograea praedator* endémiques de l'EPR. Néanmoins, certaines espèces sont retrouvées sur les deux dorsales et présentent de faibles échanges génétiques entre populations (par exemple *Paralvinella grasslei* (Jollivet et al. 1995a) ou, au contraire, comme *Bathymodiolus thermophilus* d'importants flux géniques (Won et al. 2003b)

(3) Au niveau de l'équateur se trouve un fort courant profond qui génère un puissant gyre lorsqu'il traverse l'axe de l'EPR (Reid 1997). Ce gyre est soupçonné de bloquer la dispersion des larves pélagiques pouvant expliquer les limites de répartition de 2 espèces de crabes : *Bythograea microps* au nord (21-9°N) et *Bythograea laubieri* au sud (11-32°S). Chez *Alvinella pompejana*, les populations « nord » et « sud » sont également réciproquement monophylétiques au niveau des séquences du gène mitochondrial COI. En revanche, la distribution des espèces *Riftia pachyptila* (Hurtado et al. 2004) ou les bivalves *Calymene magnifica* (Hurtado et al. 2003) et *B. thermophilus* (Won et al. 2003b) ne présentent pas de discontinuités génétiques.

(4) La microplaque de l'île de Pâques, mise en place il y a environ 5 millions d'années, est également considérée comme une barrière à la dispersion. En effet les données mitochondriales indiquent une nette divergence entre les populations de *Riftia pachyptila* situées de part et d'autre de cette plaque. Cet effet vicariant est retrouvé chez de nombreuses

espèces associées à *R. pachyptila* (Crabes Bythograeidae, vestimentifère *Oasisia alvinae*, bivalve *Calyplogena magnifica*). De même chez les Bathymodiolinae, la divergence observée entre les populations de *Bathymodiolus thermophilus* du nord (jusqu'à 21°S) et du sud (à partir de 32°S) supporte l'hypothèse d'un complexe d'espèces cryptiques (Won et al. 2003b).

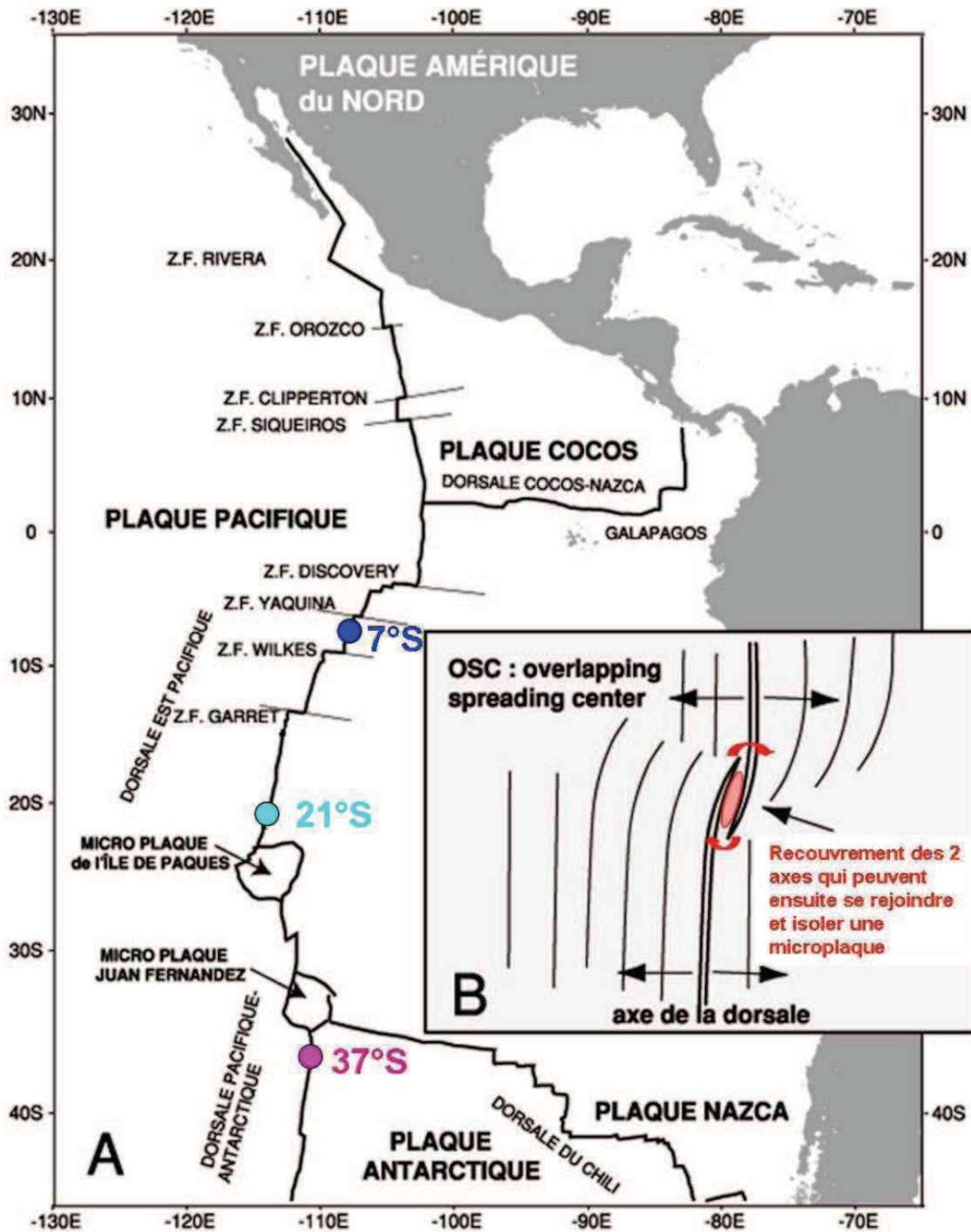


Figure 0.14. A : Géomorphologie générale de la dorsale est pacifique. B : Schéma d'un « Overlapping Spreading Center » (cf texte pour explications). Illustration modifiée de Lagabrielle (2005).

Les barrières que nous venons d'évoquer sont des barrières qui persistent depuis suffisamment longtemps pour avoir laissé leur empreinte dans l'évolution des espèces. Il ne faut cependant pas oublier que des zones de fracture comme celles de Garret ou de Yaquina plus récentes, (d'un point de vue géologique), peuvent atténuer les flux larvaires sans avoir encore avoir pu « marquer » l'histoire des espèces. Les outils génétiques et les marqueurs mitochondriaux utilisés nous permettent uniquement d'observer aujourd'hui le résultat de phénomènes passés. Ils ne nous permettent en aucun cas de réaliser une « photographie » des processus actuels de dispersion entre les champs de la dorsale.

Trois champs hydrothermaux de la dorsale EPR sud ont été échantillonnés pour notre étude : 7°25S, 21°33S et 37°S. (Figure 0.14) . Ces trois champs partagent la caractéristique d'avoir un temps de renouvellement rapide des sites.

Sur la base des données mitochondriales, l'espèce rencontrée à 7°25S à 21°33S est *Bathymodiolus thermophilus*. L'analyse de la composition faunistique des moulières montre que la plupart des espèces sont communes et largement distribuées entre ces 2 points, suggérant l'existence d'une province biogéographique unique pour l'ensemble de la dorsale est pacifique (Matabos 2007). Néanmoins, les communautés présentes à 21°33S se distinguent par la présence d'espèces nouvelles ou proches des bassins arrière-arc du Pacifique ouest (Jollivet et al. 2004).

Chez les bathymodiales, aucune structure génétique entre l'EPR nord et l'EPR sud n'a été mise en évidence, seule la présence de 2 clades distincts d'haplotypes mitochondriaux a été observée en mélange dans les populations. De plus, les études menées sur 26 locus enzymatiques (Craddock et al. 1995b) n'ont pas permis de mettre en évidence de barrière aux flux de gènes le long de la dorsale.

Le site 37°S est le site le plus méridional de la dorsale et se situe sur la dorsale Pacifique-Antarctique (PAR : Pacific-Antarctic Rise) située en deçà de la microplaque de l'île de Pâques. Cette zone n'a fait l'objet que d'une première campagne d'exploration (Campagne « Easter Microplate Expedition », R.C. Vrijenhoek & C.L. van Dover : co-chefs de mission). Peu de choses sont connues sur le fonctionnement et la diversité des communautés qui s'y trouvent. L'espèce de bathymodiale échantillonnée à 37°S n'est pas encore décrite mais

s'apparente aux modioles récoltées à 32°S présentant un haplotype mitochondrial très différent de ceux trouvés chez *B. thermophilus*.

11.2. La dorsale Atlantique

La dorsale médio Atlantique est limitée au nord par l'Islande. Il s'agit de l'émergence locale de la crête de la dorsale en relation avec une zone anormalement chaude du manteau sous-jacent (Point chaud).

L'axe de la dorsale médio Atlantique est constitué d'une vallée longitudinale presque continue, ou rift axial, large de 10 à 30 km, profonde de 2 km en moyenne et limitée par des failles normales (Lagabrielle 2005). Sur la dorsale médio-Atlantique, comme sur toutes les dorsales lentes, la vallée est affectée par la séparation des plaques sur une largeur plus importante que celle trouvée au niveau des dorsales rapides, et cela peut conduire à l'ouverture d'un fossé tectonique large.

11.2.a. Menez Gwen.

Le champ hydrothermal Menez Gwen est situé par 37°51'N sur les pentes d'un petit volcan qui s'est développé sur le plancher de la vallée axiale (Figure 0.15). Le site actif est situé à une profondeur comprise entre 860 et 842 mètres. Le fluide diffus dont la température est de l'ordre de 10-40°C est émis au niveau des dépôts hydrothermaux qui tapissent les pentes du volcan. Deux cheminées laissent échapper un fluide clair dont la température est voisine de 280°C (Donval et al. 1994).

La faune est dominée par l'espèce *Bathymodiolus azoricus* (large moulière de plusieurs centaines de milliers d'individus), les gastéropodes *Lepetodrilus atlanticus* et *Protolyra valvatoides* et de nombreux carnivores bathyaux (comme le crabe géryon, les poissons du genre *Béryx* ou les rascasses profondes) se nourrissant sur le site (Galkin et al. 2004).

11.2.b. Rainbow.

Le champ hydrothermal Rainbow est situé à 36°13'N par 2250 mètres de profondeur dans la vallée axiale de la dorsale (Figure 0.15).

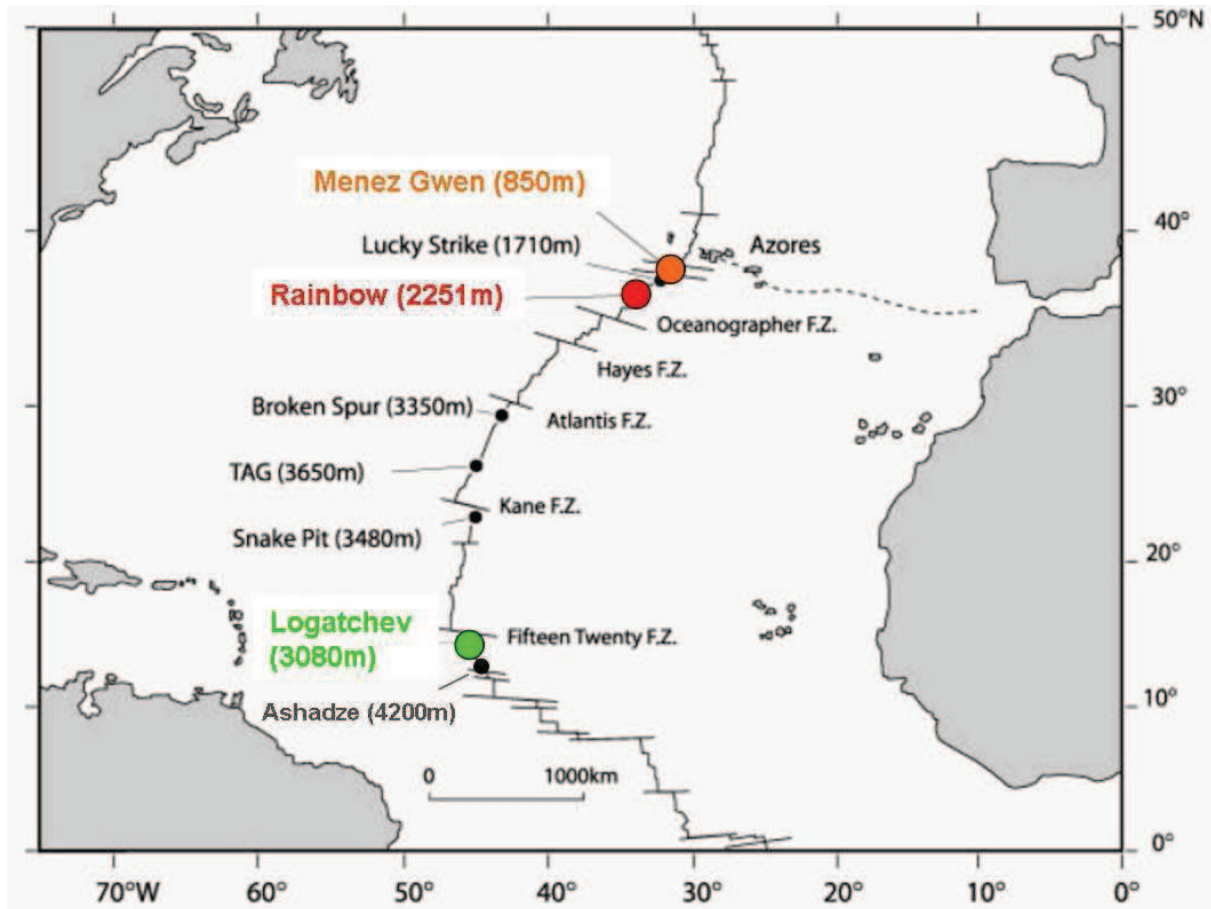


Figure 0.15: Géomorphologie générale de la Ride médio Atlantique Nord. (Carte modifiée par O'Mullan et al. (2001), d'après Needham (1996) avec détails de German & Parson (1998))

Il couvre une surface de 250 mètres de long sur 60 mètres de large. Une vingtaine de cheminées actives de grande taille (dizaines de mètres) ont été localisées dans la zone, avec des fluides noirs dont la température peut dépasser 350°C.

La particularité de ce site provient de l'affleurement de roches du manteau terrestre (Fouquet et al. 1997) provoquant un hydrothermalisme de type ultramafique. Ces roches sont la source d'un fluide très acide, fortement chargé en hydrogène et en méthane et, extrêmement riche en métaux lourds, notamment en Fe^{++} (Douville et al. 2002).

La faune est essentiellement dominée par des essaims de la crevette aveugle *Rimicaris exoculata*. L'espèce *Bathymodiolus azoricus* n'est présente qu'en faible nombre sur les parois

de quelques cheminées avec des individus de grande taille généralement ensevelis dans des sédiments de sulfures de fer oxydés.

11.2.c. Logatchev.

Le champ hydrothermal Logatchev (14°45'N) est le plus profond des 3 zones étudiées dans notre étude avec une profondeur de 3080m (Figure 0.15).

Il inclut 3 sites hydrothermaux aux caractéristiques différentes : un site composé de monticules de sulfures avec des cratères fumants, un site comprenant un réseau de cheminées actives type « fumeur noir » et, un troisième site composé de sédiments meubles d'où s'échappe un fluide diffus.

Les peuplements sont composés de bivalves Vesicomidae et d'essaims de crevettes *Rimicaris exoculata*. L'espèce *Bathymodiolus puteoserpentis* y forme un tapis dont la biomasse observée est la plus importante sur la dorsale médio-Atlantique (Gebruk et al. 2000), alors que cette espèce est plutôt rare (Zone du Snake Pit : 23°N/MAR) ou absente (zones Achadze (12°N/MAR) et TAG (21°N/MAR)) dans les autres champs hydrothermaux jusqu'ici découverts. Il convient néanmoins de noter que les moulières formées par *B. puteoserpentis* sont nettement moins grandes et denses que celles formées par *B. azoricus* au niveau du point triple des Açores.

La seule grande barrière à la dispersion connue sur la dorsale médio-Atlantique semble être la distance inter-site associée au manque de sites actifs entre Broken Spur (29°10'N) et Rainbow (36°13'N) vers le nord, et Broken Spur et Logatchev (14°45'N) vers le sud (cf Figure 0.10-a).

Des études menées sur les 2 espèces de *Bathymodiolus* de la dorsale Atlantique ont mis en évidence que ces espèces pouvaient s'hybrider au niveau du champ hydrothermal Broken Spur (O'Mullan et al. 2001, Won et al. 2003a), situé en limite des aires de répartition des deux espèces (*B. azoricus* au nord et *B. puteoserpentis* au sud).

12. Objectifs de la thèse

L'objectif premier de ce travail vise, à travers une approche multi locus sur plusieurs espèces par analyse du polymorphisme de séquences, à mieux comprendre les mécanismes d'évolution agissant chez des Mytilidae soumis à des pressions environnementales extrêmes : celles des milieux hydrothermaux profonds, et dans une moindre mesure, quand cela s'avère possible, de les comparer à ceux connus chez les bivalves côtiers (*Mytilus sp* et *Crassostrea sp*) pour tester certaines hypothèses sur une éventuelle réduction de la diversité génétique associée à la nature instable et contraignante de cet écosystème profond .

Au cours de ce manuscrit, les points suivants seront plus particulièrement analysés : (a) l'évolution de la famille des Bathymodiolinae, (b) rechercher l'existence d'écarts éventuels à la neutralité qui pourraient indiquer des gènes soumis à certains types de sélection (c) détecter les phénomènes d'introgession et l'hybridation au niveau de certaines zones de contact présentes à la fois sur la dorsale médio-Atlantique et la dorsale du Pacifique est.

13. Déroulement du manuscrit

Cette étude comporte 6 volets.

Pour commencer, une partie générale présente les échantillonnages et les outils d'analyses qui sont communs aux différents chapitres. Nous y abordons aussi quelques questions techniques concernant la détection des singletons et des gènes paralogues.

Dans la deuxième partie, une phylogénie complète des Bathymodiolinae a été reconstituée à l'aide du gène mitochondrial de la Cytochrome oxydase 1 (COI). Les données de séquences obtenues lors de nos travaux ont été ajoutées à celles précédemment publiées dans la littérature. Les relations phylogénétiques entre les espèces d'un même océan, les incohérences observées en fonction du marqueur utilisé, les interactions entre les espèces des zones de suintements froids et celles des sources hydrothermales sont discutées. Les hypothèses formulées au cours de cette première partie représentent le fondement des discussions abordées dans les chapitres suivants.

Dans une troisième partie, une étude d'isolement d'espèces avec flux de gènes a été réalisée sur la zone d'hybridation de la dorsale Atlantique à l'aide d'une approche multi-locus afin de (i) déterminer la structure génétique et le niveau d'introgession entre les espèces *Bathymodiolus azoricus* et *B. puteoserpentis*, (ii) d'estimer l'intensité et le sens des flux géniques entretenus entre les 2 espèces par l'utilisation d'une approche bayésienne, et (iii) rechercher des gènes dont la coalescence s'écarte des attendus neutres de l'évolution moléculaire en présentant des traces de sélection.

Le quatrième volet s'intéresse plus particulièrement au complexe d'espèces de la dorsale est-Pacifique. Notre étude démontre la présence d'une zone d'hybridation au niveau de l'EPR sud avec une espèce de la dorsale Antarctique-Pacifique qui reste pour l'instant non décrite. L'utilisation de données allozymiques et de nombreux marqueurs nucléaires nous permet d'obtenir des résultats qui contrastent avec ce qui a préalablement été décrit dans la littérature concernant l'homogénéité génétique des peuplements de *Bathymodiolus thermophilus* le long de la dorsale EPR et de caractériser plus finement cette zone.

La cinquième partie de notre étude porte sur l'étude des processus d'évolution moléculaire de quelques gènes qui au cours des précédents chapitres ont révélé un intérêt particulier : évolution par duplication ou des écarts à l'évolution neutre laissant supposer l'action d'une sélection directe ou indirecte.

Enfin, le sixième et dernier volet de notre étude dresse un bilan des principaux résultats obtenus lors de la thèse et cherche à établir une comparaison entre les principales caractéristiques génétiques des Bathymodiales et celles des bivalves côtiers des genres *Mytilus* et *Crassostrea* afin de savoir si l'instabilité et la nature 'extrême' du milieu hydrothermal peut avoir un impact sur la diversité génétique des espèces y résidant.

Chapitre I. Matériel et Méthodes Générales

Au cours de cette thèse, en dépit de la diversité des thèmes abordés et de l'originalité des approches utilisées, certains protocoles et outils d'analyse restent transversaux à l'ensemble des chapitres. Aussi, afin d'éviter les redites et alléger le manuscrit, la plupart des méthodes communes aux différents chapitres seront détaillées dans cette partie générale à laquelle nous ferons ensuite référence. Au cours de ce chapitre, certains résultats méthodologiques ayant trait aux protocoles de marquage-recapture des allèles et à la détection de gènes paralogues seront développés et discutés.

I.1. Stratégie d'échantillonnage

Les moules bathymodiolinae sont endémiques des milieux océaniques profonds et sont par conséquent difficiles d'accès. Ces échantillons faunistiques ont été acquis au cours de campagnes océanographiques de plongées en bathyscaphe mettant en œuvre d'importants moyens humains et des technologies océanographiques des plus perfectionnées. Ces campagnes hauturières mettent en œuvre les plus gros bateaux océanographiques des flottes scientifiques françaises et américaines (N/O L'Atalante, N/O Le Pourquoi Pas ?, R/V Atlantis II) car elles nécessitent l'utilisation d'appareils sous-marins pour la récolte des échantillons. Ces engins peuvent être habités (bathyscaphes français Nautille et américain Alvin) ou télécommandés depuis la surface (robots de type ROV² comme le ROV français Victor 6000) et les échantillons sont récoltés à l'aide de leurs bras télé-manipulateurs puis remontés jusqu'en surface à l'intérieur d'un panier isotherme (Figure I.1).

Les échantillons utilisés proviennent de 6 campagnes ayant eu lieu dans les océans Pacifique (les campagne françaises : HOT 96 sur l'EPR nord et Biospedo 2004 au niveau de l'EPR sud entre 7°25S et 21°33S, la campagne américaine 'Eastern Microplate Expedition' 2005 au niveau de la dorsale Pacifique-Antarctique entre 32 et 37°S) et Atlantique (les campagnes européennes ATOS 2001 et Momareto 2006 sur le point triple des Açores à 37°N et, la campagne franco-russe Serpentine 2007 sur la dorsale médio-Atlantique entre 12°50N et

² ROV : Remote Operated Vehicle (littéralement, véhicule commandé à distance).

14°N). Ces informations sont récapitulées dans le Tableau I.1. De plus amples détails seront fournis dans les différents chapitres (taille des échantillonnages, profondeur...).

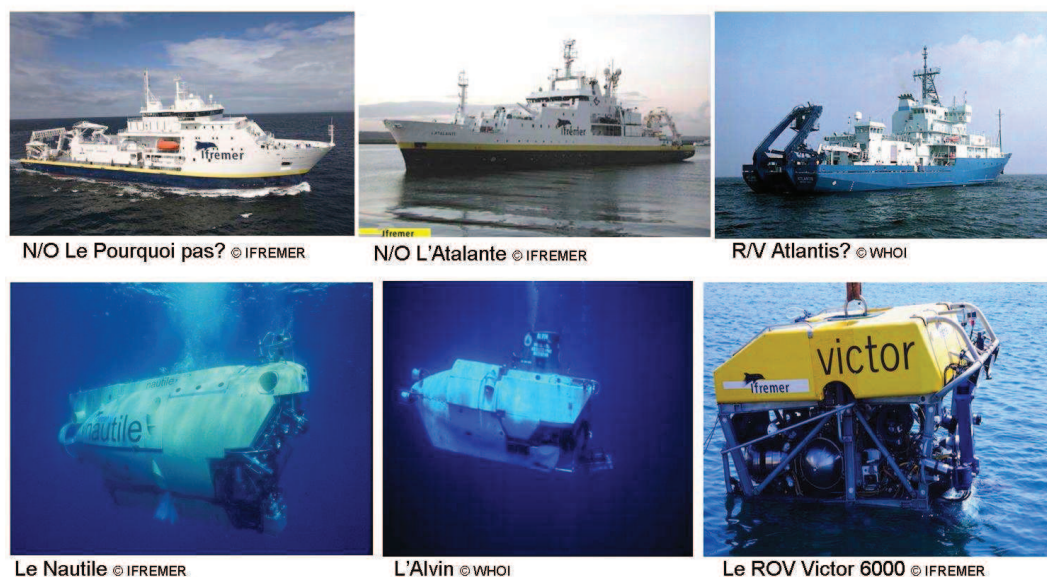


Figure I.1. Moyens à la mer utilisés pour effectuer les échantillonnages de *Bathymodiolus*.

Tableau I.1. Récapitulatif des échantillonnages.

Provenance géographique	Espèce	Nom de la population	Coordonnées géographiques	Campagne d'échantillonnage
Dorsale Est-Pacifique	<i>B. thermophilus</i>	Qvent/BioVent	9°50'N / 104°17' W	HOT 96
		Sarah's Spring	7°25'S / 107°47' W	Biospeedo (2004)
		Lucky Eric	13°59'S / 112°29' W	Biospeedo (2004)
		Oasis	17°25'S / 113°12' W	Biospeedo (2004)
		Rehu Marka	17°25'S / 113°12' W	Biospeedo (2004)
		Suzie	17°35'S / 113°15' W	Biospeedo (2004)
		Animal Farm	18°36'S / 113°24' W	Biospeedo (2004)
		Gromit	21°33'S / 114°18' W	Biospeedo (2004)
	<i>B. aff. thermophilus</i>	Sebastian's Teamer	37°47'S / 110°55' W	Eastern Microplate Expedition (2005)
Dorsale Médio-Atlantique	<i>B. azoricus</i>	Menez Gwen	37°51'N / 31°31'W	Momareto (2006)
		Rainbow	36°13'N / 33°54'W	ATOS (2001)
		<i>B. puteoserpentis</i>	Logatchev	14°45'N / 44°58'W

I.1.1. Préparation des échantillons.

Une fois les moules remontées à bord, les animaux sont sexés et la coquille de chaque individu est mesurée au niveau de sa longueur maximale, sa hauteur et sa largeur à la charnière à l'aide d'un pied à coulisse. Les échantillons sont ensuite disséqués sur un bloc

froid. et les tissus du pied et des branchies et font l'objet d'un double échantillonnage. Une partie des tissus est immédiatement congelée dans l'azote liquide, et l'autre partie conservée dans des tubes d'éthanol absolu jusqu'à l'extraction d'ADN qui a eu lieu au laboratoire (Figure I.2). L'ADN est extrait suivant 3 procédés : en utilisant le protocole d'extraction standard au Phénol-Chloroforme décrit par (Sambrook & Russel 2001), la méthode d'extraction à la résine Chelex[®] 100 ou encore la méthode d'extraction au CTAB/PVP 1% décrite dans (Jolly 2005). Les trois protocoles utilisés sont fournis en annexe 1

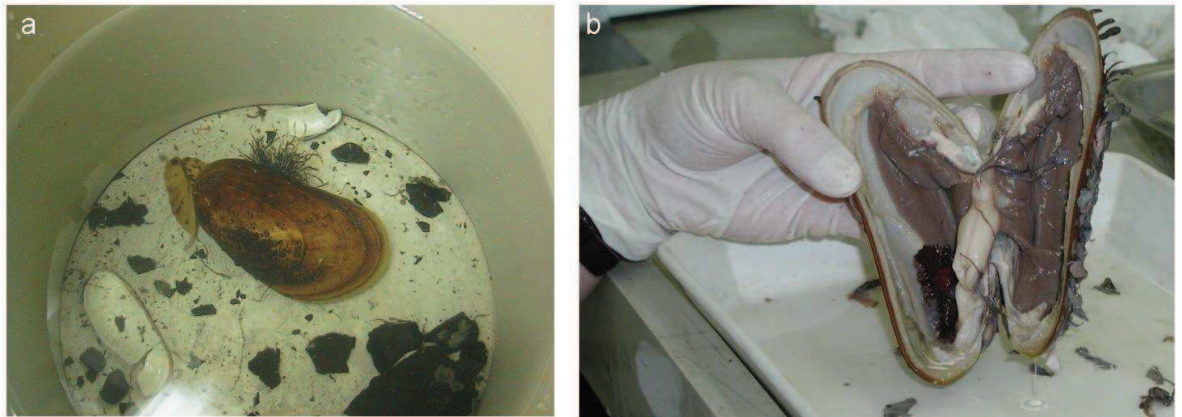


Figure I.2. Préparation des échantillons de *Bathymodiolus*. **(a)** récupération de 2 échantillons de *Bathymodiolus thermophilus* dans une boîte d'échantillonnage du bathyscaphe Nautile (en blanc : coquille brisée de *Bathymodiolus*, en noir : fragment de basalte ; **(b)** dissection des animaux dans le laboratoire humide de N/O L'Atalante (ici *Bathymodiolus thermophilus* et son polychète commensal *Branchipolynoe symmytilida*). Crédits Photo : (a) B. Faure, (b) C. Daguin ; pendant la campagne Biospedo 2004.

I.1.2. Choix des marqueurs génétiques.

L'étude des mécanismes évolutifs à partir des données de polymorphisme moléculaire nécessite l'utilisation de gènes nucléaires. En effet, ils sont les seuls à pouvoir fournir une information portant sur l'introgression de certains allèles entre espèces proches, le taux de recombinaison et le degré d'hétérozygotie dans les populations. Le choix des marqueurs et la définition des amorces utilisées pour les amplifier dans notre étude ont donc été établis selon quatre critères :

(1) Implications des gènes sélectionnés dans des fonctions physiologiques contrastées afin de constituer un échantillon représentatif de gènes présentant un panel de fonctions (respiratoire, digestive, reproductrice, lutte contre le stress thermique) potentiellement soumises à des pressions de sélection différentes selon l'espèce considérée. L'utilisation de

gènes indépendants permet également d'explorer un maximum de ces fonctions et augmente ainsi la probabilité de rencontrer un gène sous sélection diversifiante ou balancée.

(2) Présence de régions codantes et non codantes. En effet, le type d'approches utilisées permet, dans les régions codantes de comparer le polymorphisme synonyme (silencieux) au polymorphisme non synonyme pour détecter des effets sélectifs, et d'utiliser les régions non codantes (introns : polymorphisme silencieux) comme une référence neutre.

(3) Une taille maximale de l'insert de 1500 bp afin de pouvoir amplifier et cloner sans problème ces portions de gène, mais aussi de les séquencer en série.

(4) L'amplification de l'ensemble des individus d'une population, de l'ensemble des populations de la même espèce, et des individus provenant de plusieurs espèces proches doit être possible à partir du même jeu d'amorces, d'où la nécessité de définir les amorces sur des régions conservées (exons).

La sélection des gènes d'intérêt et la définition des amorces a été rendue possible grâce à la construction d'une banque d'ADNc pleine longueur non normalisée et le séquençage automatisé de 6000 clones (Tanguy et al. 2008). Cette banque a été construite à partir des ARN totaux extraits d'un mélange de branchies, muscles adducteurs, manteau et glande digestive de cinq spécimens de l'espèce *Bathymodiolus azoricus* provenant du site hydrothermal Menez Gwen (840 m de profondeur, 37°50,56'N de latitude et 31°31,27'W de longitude) sur la dorsale médio-atlantique au niveau du point triple des Açores pendant la campagne ATOS 2001 (P.-M. Sarradin, chef scientifique). La banque a été construite à l'aide d'un kit de synthèse SMART PCR cDNATM (Clontech) et les ARNm Poly(A⁺) ont été isolés des ARN totaux grâce au protocole d'isolation PolyATtract® (Promega, Madison, WI, USA). Un total de 2.10⁶ clones a été obtenu. Le séquençage de la banque a été réalisé au Génoscope (Evry, France ; <http://www.genoscope.cns.fr/>) avec un séquenceur automatique à capillaires ABI 3730 et un kit de séquençage ABI BigDye Terminator v.3.1.

La mise au point des amorces à l'aide du logiciel Primer Premier 5 (<http://www.premierbiosoft.com/primerdesign/>) et le choix des gènes d'intérêt présentant du polymorphisme nucléotidique ont été réalisés à Roscoff. L'obtention des séquences génomiques (introns/exons) a été effectué par amplification PCR sur des machines PE

GeneAmp 9700 et MJ PCT-200. Les amplifiats présentant la taille attendue ont été clonés individuellement à l'aide d'un kit de clonage pGEM-T® (Promega) après avoir été découpés préalablement sur gel d'agarose 2% et purifiés sur colonnes QIAQuick™ (QIAGEN). Le séquençage et la mise au point des amorces ont été facilités grâce à la présence locale de la plateforme de séquençage Genomer.

A cette collection de gènes nucléaires s'est ajouté le gène mitochondrial codant pour la Cytochrome c oxydase I, ce marqueur ayant été très largement utilisé dans la diagnose des bathymodiolinae (cf. Chapitre II).

I.1.3. Informations sur les gènes étudiés

I.1.3.a. Cytochrome c oxydase I (COI)

Le gène mitochondrial codant pour la sous-unité I de la Cytochrome-C Oxydase (COI) a déjà été utilisé pour décrire le statut phylogénétique des bivalves Bathymodiolinae (Iwasaki et al. 2006) et constitue le marqueur retenu pour effectuer l'inventaire du vivant en 'barcoding' (Hebert et al. 2003). Cette sous-unité joue un rôle essentiel dans la respiration mitochondriale en catalysant le transfert d'électrons entre le ferrocyclochrome *c* et les molécules d'oxygène et contribue à la synthèse d'ATP nécessaire aux cycles de Krebs et de la glycolyse (Malatesta et al. 1995). De précédentes études ont suggéré que le marqueur COI pouvait être sous forte sélection purifiante chez les organismes qui vivent dans des conditions anaérobiques comme c'est le cas pour les espèces hydrothermales (Powell & Somero 1986), ce qui pourrait expliquer le très faible taux d'évolution observé chez ces espèces (Chevaldonne et al. 2002, Johnson et al. 2006). D'autres travaux ont également montré que ce marqueur pouvait accumuler de nombreuses mutations délétères dans le polymorphisme (Nachman 1998, Lynch & Blanchard 1998, Muths et al. 2006) et pouvait subir une récurrence de balayages sélectifs chez les espèces présentant de larges tailles efficaces (Bazin et al. 2006, Eyre-Walker 2006).

I.1.3.b. Lysozyme

Cette protéine est une enzyme capable de couper les liaisons 1-4 entre l'acide *N*-acétylmurannique et le *N*-acétylglucosamine du peptidoglycane bactérien. Il a été suggéré que cette enzyme pouvait jouer un rôle dans la défense contre les infections bactériennes et

notamment dans le contrôle de la prolifération endosymbiotique, ainsi qu'un rôle dans la digestion des composés glucidiques chez la bathymodiole (Bachali et al. 2002).

1.1.3.c. Facteur d'Elongation 1 alpha (EF1 α)

Il s'agit d'une sous unité (alpha) du principal facteur de la machinerie traductionnelle, mais cette protéine a aussi des fonctions remarquables concernant la limitation des proliférations cellulaires et/ou l'apoptose (Ejiri 2002). Cette molécule a également été largement utilisée dans les études de phylogénie profonde, notamment chez les espèces des environnements profonds (McHugh 1997, Halanych & Janosik 2006) en raison de sa grande conservation (au moins dans sa partie codante) à l'échelle du vivant (Baldauf et al. 2000).

1.1.3.d. Collagène type XIV

La matrice extra-cellulaire forme un réseau moléculaire complexe qui détermine l'architecture complexe des tissus et dirige la polarisation, la migration, la prolifération, la survie et la différenciation des cellules. Les collagènes, comme les glycoprotéines non collagènes et les peptidoglycanes, appartiennent à la matrice extra-cellulaire et sont des protéines multi domaines exerçant de multiples fonctions biologiques (Ruehl et al. 2005). Le collagène utilisé dans notre étude (le collagène XIV ou DEC-1 d'après les résultats du BLASTX) est produit par les fibroblastes et les myofibroblastes et il est exprimé tardivement lors du développement embryonnaire.

1.1.3.e. Espaceur inter-génique ribosomique ITS2

Chez la plupart des Eucaryotes, les séquences matures d'ARNr sont produites dans le nucléole à partir d'un unique précurseur (pré-ARNr) contenant des séquences non codantes (introniques) appelées espaceurs. Ces espaceurs inter-géniques, dont l'ITS2 fait partie (Internal Transcribed Spacer 2) vont être supprimés dans une série d'étapes d'épissage afin de donner un ARNr mature et fonctionnel. La fonction biologique de ces espaceurs est mal connue, laissant supposer que ce marqueur puisse évoluer de façon neutre. Cependant, leur structure secondaire pourrait être sous contraintes sélectives, celle-ci ayant un rôle potentiel dans l'épissage car elle est fortement conservée chez les Eucaryotes (Joseph et al. 1999). Ce marqueur a largement été utilisé dans les études phylogénétiques effectuées à l'échelle du

genre ou de la famille (Schlötterer et al. 1994). Il appartient à la famille multigénique des ADN ribosomiques qui forme une série de copies répétées en tandem généralement soumises à l'action de l'évolution concertée (Dover 1982). Il existe cependant de nombreux exemples où cette famille peut évoluer en gènes paralogues, notamment lorsque les NORs (*Nucleolar Organiser Regions*) sont répartis sur plusieurs paires de chromosomes (Tang et al. 1996).

1.1.3.f. Protéine de stress HSP 70

Les protéines de choc thermique (HSP : Heat Shock Proteins) sont présentes à la fois chez les Procaryotes et chez les Eucaryotes. Elles possèdent une structure fortement conservée suggérant qu'elles jouent un rôle fondamental dans les processus cellulaires, notamment l'homéostasie (Kiang & Tsokos 1998). La surexpression de certaines de ces protéines HSP peut être induite en réponse à un stress environnemental, notamment la température (Tissiere. et al. 1974), l'augmentation en concentration de métaux lourds (Levinson et al. 1980), ou l'infection bactérienne (Kiang & Tsokos 1998). Les HSP sont une famille multigénique dont le nom correspond à leur poids moléculaire : les HSP 10, 23, 40, 60, 70, 90 et 110 kDa. Avec sa forme en tonneau caractéristique, la principale fonction de la HSP 70 est celle de protéine chaperonne, c'est-à-dire qu'elles servent à guider ou favoriser le maintien de la structure secondaire ou permet la re-naturation de certaines protéines du métabolisme. De part leur fonction, ces protéines ont un rôle très important dans l'adaptation des organismes aux fluctuations thermiques de l'environnement (Sorensen 2003), au moins du point de vue de leur variabilité d'expression entre individus (Dahlhoff & Rank 2000).

1.1.3.g. Sulfotransférase

Les enzymes appartenant à la famille des sulfotransférases catalysent des réactions de conjugaison des sulfates : il s'agit du transfert d'un groupe de sulfates depuis un donneur vers un accepteur. Cette réaction, appelée sulfonation, est une voie importante dans la biotransformation (activation ou désactivation) de nombreuses hormones, neurotransmetteurs, et composés xénobiotiques. Ce processus permet en effet d'augmenter la solubilité de ces molécules afin de faciliter leur excrétion (Weinshilboum et al. 1997).

I.1.3.h. Adénosine Homocystéinase

Cette enzyme dont le nom complet est la S-Adenosylhomocystéine-hydrolase (ou SAHH), catalyse la réaction d'hydrolyse de la S-Adenosylhomocystéine (SAH) en adénosine et homocystéine (De la Haba & Cantoni 1959). Elle constitue une enzyme clé dans les réactions permettant la formation d'adénosine, car cette molécule est impliquée dans tous les transferts d'énergie via le couplage adénosine triphosphate (ATP) /adénosine diphosphate (ADP), dans la transduction des signaux *via* l'adénosine monophosphate cyclique (AMPC) et, dans les fonctions immunitaires (Fox & Kelley 1978). Cette enzyme a déjà été utilisée dans une étude du polymorphisme dans les populations humaines (Kloor et al. 2006) et constitue un marqueur à fort taux d'évolution.

I.1.3.i. Ferritine GF1

Cette molécule dont le nom correspond à Gigas Ferritin 1 (GF1), car elle a été isolée chez *C. gigas*, possède deux fonctions principales : la détoxification et le stockage du fer. Ces protéines sont importantes dans l'homéostasie car elles stockent et régulent le fer au niveau cellulaire et peuvent fixer certains radio-éléments (Durand et al. 2004). Certaines études attribuent aux ferritines un rôle de protection contre les dommages pouvant être causés par les radicaux libres (Orino et al. 2001) ou suggèrent qu'elles puissent agir dans la régulation des gènes impliqués dans la prolifération cellulaire (Le & Richardson 2002). Il apparaît donc que ces protéines jouent un rôle important dans la protection cellulaire. Elles ont un rôle particulièrement important chez les espèces hydrothermales qui sont soumises à des concentrations extrêmement élevées en Fer (Roesijadi & Crecelius 1984, Van Dover et al. 2001), notamment au niveau des sources présentes sur des roches ultramafiques (site Rainbow (35°N/MAR) ou site Logatchev (14°N/MAR)).

I.2. Amplification, marquage/recapture, clonage et séquençage

Des amorces spécifiques ont été définies sur les régions conservées (COI, ITS2) ou sur les régions exoniques flanquant les introns (gènes nucléaires) à partir de séquences de *Bathymodiolus* issues de la collection d'ESTs³ de *B. azoricus*. Après la mise au point des marqueurs et lorsque cela a été possible, certaines amorces ont été définies sur des régions

³ EST : Expressed Sequence Tags

introniques conservées pour les gènes où l'intron était de trop grande taille (EF1 α). La structure des portions de gènes utilisés dans l'étude est représentée sur la Figure I.3.

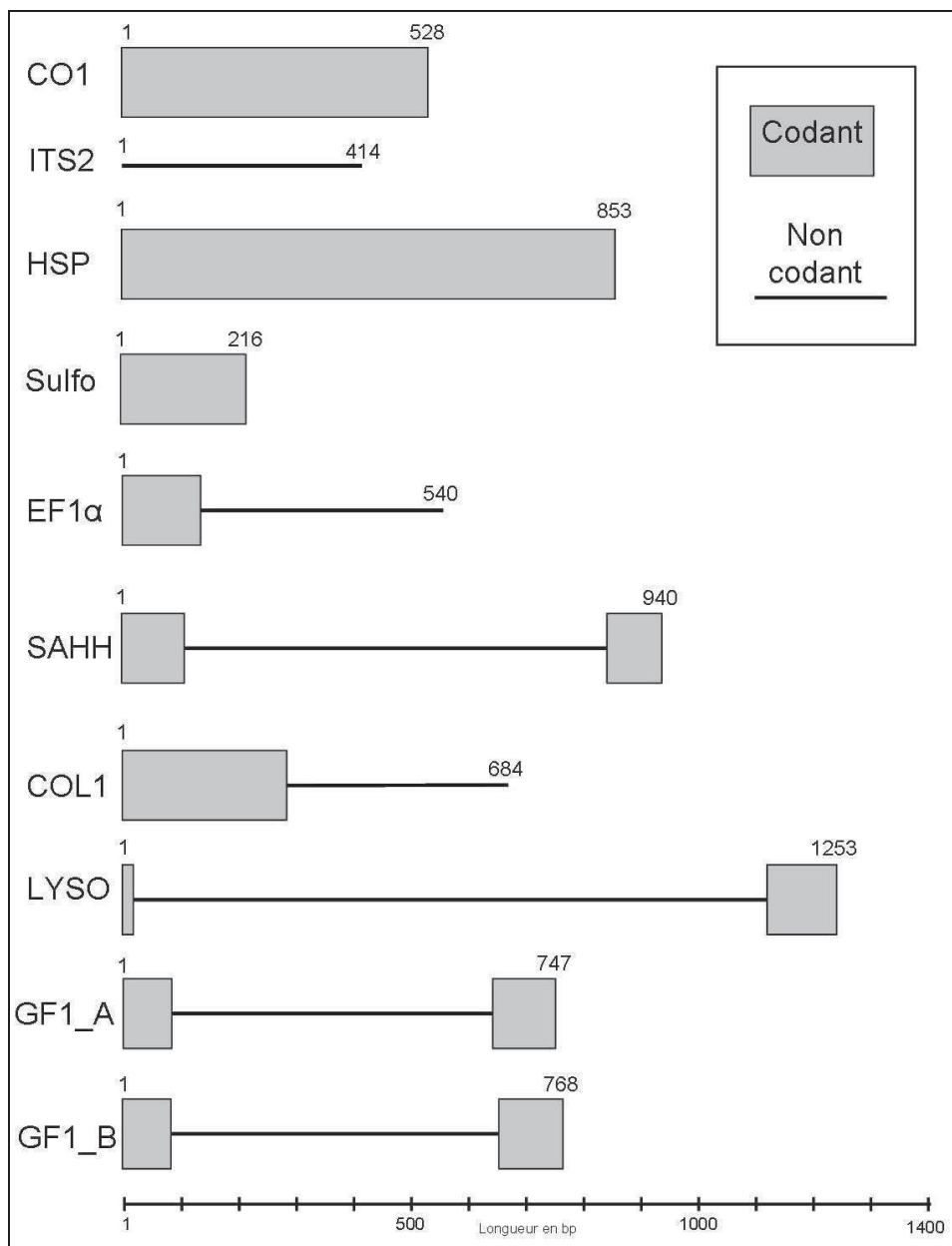


Figure I.3. Structure des portions de gènes utilisés dans la thèse. En boîte grisée : séquences codantes (exons) ; ltraits noirs : séquences non codantes (introns). Tous les gènes sont représentés à la même échelle (en pb).

Les couples d'amorces utilisées pour amplifier la sous-unité I du gène mitochondrial de la Cytochrome c oxydase (BatCOI), les amorces utilisées pour amplifier les gènes nucléaires utilisés dans l'étude (HSP70, EF1 α , SAHH, Col1, LYSO, SULFO, GF1 A et B, et ITS2) ainsi que la séquence des amorces définies sur le plasmide (BS1F et BS1R) pour tester les inserts dans les clones positifs par visualisation sur gel d'agarose 1% ou celles utilisées pour le séquençage (SP6 ; T7) sont indiquées dans le Tableau I.2.

Tableau I.2. Séquences des amorces utilisées dans cette étude.

locus	5'-[taille tag] Amorce Forward -3'	5'-[taille tag] Amorce Reverse -3'
BatCO1	ATA AAA AGA TGT ATT RAA RTG ACG	TGT GGT CTG GAA TAA TTG GAA
Hsp 70	[3bp] GGG AAA GTT GAC ATT ATT GCC AAT G	[3bp] ATT CAT AAA TTC TGT CAA CAT TTT CTG T
Ef1 α	[3bp] ACG CCT GGG TAT TGG ACA AAC T	[3bp] CCA AGA GGG GTC GTA CAA ATT A
Sahh	[4bp] TAA ATC TTG GTT GTG CTC ATG GTC ATC CA	[4bp] TTT GAA TGG TCC TTC TTT AGG TAG AC
Col1	[4bp] TTA CCA AAT CAC AGG TCG GGG TCA TG	[4bp] TCT CCT TCG CTA TTT TTG TGG G
Lyso	[4bp] GTT TCC CCA AAA TGT ATG AGC TGT	[4bp] TAA TCT CCG CCT GGA CTA CCA CAA TC
Sulfo	[4bp] TCT TTA AAG TCA GGA TCA CAT TGG	[4bp] TAA GGC AAA GTG GAA CAA CGA GAC CGC
Gf1A et B	[4bp] ATG TCT CAA AGT CAA CCT CGT CAA AAT	[4bp] TTG ATG TCC TGA AGA ACG ATT CGT CCT CC
ITS2	ACA TTG CGG CTT TGG GTC AC	GCT TAA ATT CAG CGG GTA CT
BS1	AAA GGG GGA TGT GCT GCA AGG CG	GCT TCC GGC TCG TAT GTT GTG TG
T7	GTA ATA CGA CTC ACT ATA	Fonctionne avec SP6
SP6	Fonctionne avec T7	CAT TTA GGT GAC ACT ATA G

I.2.1. Amplification des marqueurs

Le gène mitochondrial COI a été amplifié dans un volume de 25 μ l selon les conditions suivantes: 1X tampon PCR (fourni avec la Taq polymérase), 2,5mM MgCl₂, 0,25 μ M de chaque dNTP, 0,5 μ M de chaque amorce, 0,04U de *Taq* polymérase Thermoprime Plus (ABgene), 25ng ADN génomique. Les paramètres du cycle d'amplification sont : 94°C pendant 2 min, 5 cycles de 3 étapes (94°C pendant 35s, 46°C pendant 35s, 72°C pendant 70s); suivis de 30 cycles de 3 étapes (94°C pendant 35S, 50°C pendant 35s, 72°C pendant 70s) ; puis une élongation finale à 72°C pendant 10 min. Les produits PCR ont été purifiés avant séquençage avec les plaques «Multiscreen 96-plate purification kit™» (Millipore) puis séquencés dans les deux sens pour chaque individu en utilisant le kit BigDye Terminator Chemistry (Perkin Elmer) et le séquenceur à capillaires ABI 3100 de la plateforme de Roscoff, suivant le protocole du fournisseur.

Les gènes nucléaires ont été séquencés en utilisant une nouvelle méthode appelée la « méthode de clonage par marquage/recapture » (Bierne et al. 2007) (cf article en annexe 5), qui permet de cloner en une seule fois tous les produits PCR issus de l'amplification individuelle de chaque individu. Cette méthode consiste à marquer individuellement les séquences pendant la PCR en utilisant des amorces « marquées », c'est-à-dire qui possèdent au niveau de leur partie 5' une petite queue de nucléotides n'intervenant pas dans l'hybridation. (Figure I.4)

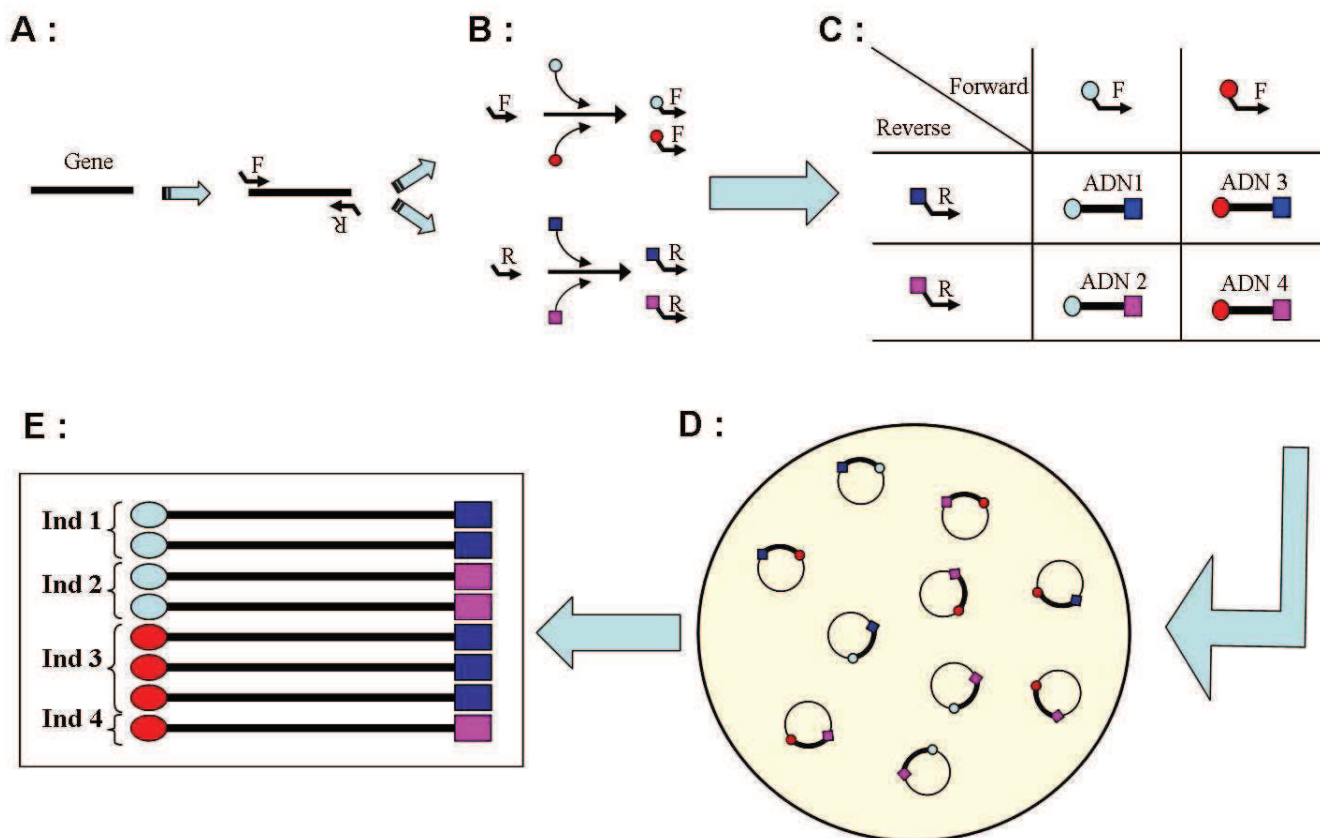


Figure I.4. Schéma de fonctionnement de la méthode du clonage par marquage-recapture.

A : définition d'amorces classiques pour amplifier le gène cible.

B : Ajout d'une petite queue poly-nucléotidique à chaque extrémité 5'. Cette étiquette ne va pas interagir avec la matrice d'ADN pendant la réaction de PCR.

C : Dans une plaque PCR, combinaison des amorces 'forward' et 'reverse' afin d'obtenir des paires d'amorces uniques pour chacun des individus à amplifier (ADN 1 à 4). Chaque réaction de PCR va se faire indépendamment dans chaque puits (= pour chaque individu).

D : Les produits de PCR sont mélangés dans des proportions équimolaires pour une purification et un clonage unique

E : Séquençage des clones avec des amorces universelles du plasmide (SP6 et T7) dans des régions proches du site d'insertion. Les étiquettes poly-nucléotidiques sont également séquencées et permettront l'identification a posteriori de l'individu duquel proviennent les séquences.

Les gènes nucléaires sont amplifiés dans un volume de 20µl contenant 1X de tampon PCR (fourni avec la Taq polymérase), 2mM MgCl₂, 0,25µM de chacun des dNTPs, 0,4µM de chaque amorce, 0,025 U de *Taq* polymérase Thermoprime Plus (ABgene), 25ng d'ADN génomique. Les paramètres du cycle d'amplification sont : 94°C pendant 5 min, 35 cycles de 3 étapes (94°C pendant 45s, 60°C pendant 1min, 72°C pendant 2min) suivis d'une élongation finale de 10 minutes à 72°C. Chaque individu est amplifié indépendamment puis les produits de PCR sont regroupés et purifiés au sein de chaque population pour un clonage unique selon

la méthode de marquage recapture. Le vecteur de clonage utilisé est le pGEM-T en utilisant des kits de clonage pGEM-T (Promega, Madison, WI, USA) et les bactéries sont des DH5 α T1 phage résistantes (Invitrogen). Le clonage est effectué selon le protocole fourni en annexe 2. Le clonage avec la méthode de marquage et recapture permet d'obtenir et de distinguer les deux allèles des individus hétérozygotes, ce qui n'est pas possible dans la majorité des autres protocoles. L'effort de capture et de séquençage est de deux, c'est-à-dire que le nombre de clones séquencés est égal à deux fois le nombre d'individus amplifiés qui constituent la population étudiée (généralement 48 individus, donc 96 clones séquencés dans les deux sens). Après chaque réaction de clonage, un certain nombre de clones subissent un « contrôle qualité » par une amplification PCR (avec les amorces BS1F et BS1R (Tableau I.2) afin de vérifier s'ils possèdent bien un insert de la taille attendue pour chacun des gènes (Figure I.5). Les clones sont ensuite envoyés au Génomscope d'Evry pour y être séquencés avec les amorces universelles de plasmide SP6 et T7 (cf Tableau I.2). Les paramètres du cycle d'amplification de la réaction de séquence sont : 96°C pendant 5 min, 50 cycles de 3 étapes (96°C pendant 30s, 56°C pendant 30s, 60°C pendant 4min), la réaction de séquence est ensuite maintenue à 4°C.

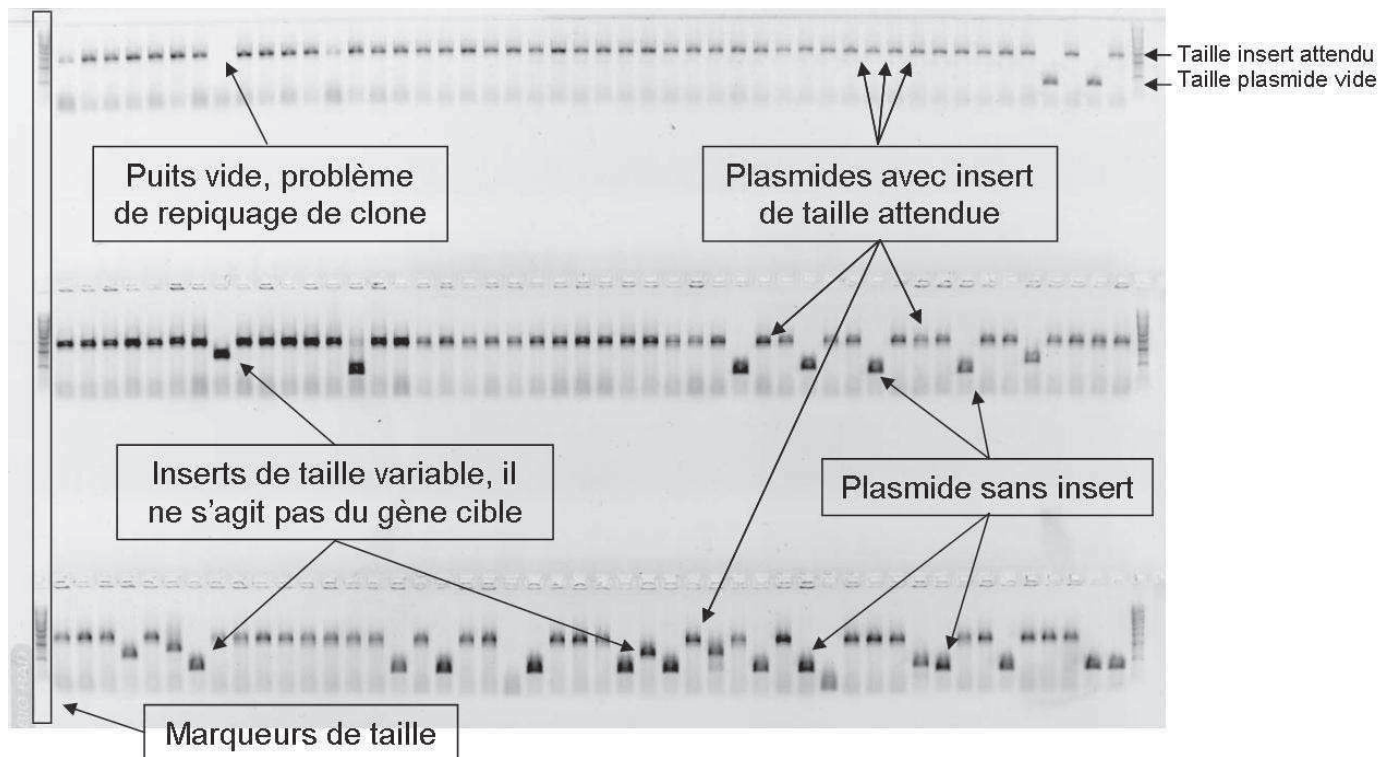


Figure I.5. Exemple de contrôle avant séquençage de la qualité du clonage de la HSP 70 sur gel d'agarose. L'amplification des clones contrôlés a été réalisée avec les amorces BS1f / BS1r.

I.2.2. Traitement des séquences avant analyse

Après récupération des séquences, celles-ci sont alignées, corrigées puis ré-attribuées à leur individu de provenance. Du fait de la méthode utilisée, de fortes disparités sont observées au niveau du nombre total d'individus ré-échantillonnés (càd. certains individus ne sont pas systématiquement ré-échantillonnés au niveau de leurs allèles), mais aussi au niveau du nombre d'allèles re-capturés plusieurs fois pour un individu donné. Lorsque le mélange des produits de PCR a été réalisé de façon équimolaire, la méthode de clonage par marquage-recapture permet d'obtenir des séquences provenant de 36 individus sur 48 (soit 75%) ou 27 sur 33 individus (82%). En revanche, un léger déséquilibre dans le mélange des produits de PCR ou des différences au niveau des qualités d'ADN génomique conduit à un clonage préférentiel de certains allèles entraînant un net déséquilibre dans les recaptures. Ainsi, dans certains cas heureusement rares, le nombre d'individus recapturés a été assez faible (22 individus sur 48). Dans ce cas, de nouveaux clones ont été séquencés afin d'obtenir des séquences d'individus supplémentaires. Dans d'autres cas, un grand nombre d'individus a été recapturés mais avec un net déséquilibre dans le nombre de recaptures par individu : nombreux individus à une séquence contre un individu à 15 séquences !

I.2.3. Alignement et correction des séquences

Pour chacun des gènes, la qualité des séquences a été vérifiée visuellement sur les chromatogrammes en utilisant le logiciel Chromas 2.22 (Technelysium Pty. Ltd., Helensvale, Australie). Les séquences ont été alignées avec le programme ClustalW (Thompson et al. 1994) inclu dans le logiciel Bioedit (Hall 1999) ou avec Multalin (Corpet 1988) directement depuis le site internet (<http://bioinfo.genopole-toulouse.prd.fr/multalin/multalin.html>). Les séquences ont ensuite été corrigées avec le logiciel BioEdit (Hall 1999).

Les corrections appliquées aux données et le séquençage réalisé dans les deux sens, permettent d'éliminer certains problèmes liés au séquençage comme l'incertitude des pics à certains sites, la dégradation de la qualité du signal sur les longues séquences, les problèmes liés à la présence d'un élément répété... Après corrections, il ne devrait donc rester qu'une seule séquence chez les animaux diploïdes homozygotes et deux séquences au maximum chez les hétérozygotes. Il s'agit là de la théorie car dans les faits, certains individus possèdent plus de deux séquences (souvent liés à un polymorphisme de singletons). On constate deux grands types de multi allélisme intra-individuel : (i) Le premier type, rare, est celui d'individus

présentant au moins trois types de séquences très divergentes. Dans la plupart des cas, on peut facilement démontrer grâce aux « étiquettes poly-nucléotidiques » qu'il s'agit d'un problème d'attribution de la séquence à un individu donné (mutation artéfactuel sur le tag, base manquante sur le tag, ou recombinaison *in vitro* pendant la PCR ou le clonage). Dans certains cas, en revanche, aucun problème d'attribution n'est constaté et le multi allélisme ne peut s'expliquer que par des phénomènes de gènes présentant des copies multiples pour lesquels l'évolution concertée se fait mal ou des gènes paralogues. (ii) le deuxième grand type de multi allélisme correspond à des allèles très proches, qui diffèrent entre eux par des singletons. Il n'est pas possible *a priori* d'éliminer ces singletons car ils sont bien réels sur les chromatogrammes. Ces singletons peuvent avoir trois origines distinctes : polymorphisme intra-individuel (dû à des mutations somatiques accumulées entre différentes cellules de l'organisme), ou artéfactuelles liées à un problème de fidélité de la *Taq* ou à un problème de réplication des bactéries pendant le clonage.

I.2.4. Correction des jeux de données

Il a donc été nécessaire de définir un protocole de correction des séquences afin de pouvoir les utiliser dans les analyses de polymorphisme sans induire de biais sur les estimateurs de diversité. Nous avons donc défini trois types de jeux de séquences correspondant à des niveaux différents de qualité.

Le tout premier niveau de correction correspond à l'ensemble des séquences obtenues et corrigées en fonction des chromatogrammes. Les séquences identiques trouvées plusieurs fois pour un même individu sont supprimées pour n'en laisser qu'une seule, et le nombre de recaptures est indiqué avec la séquence gardée.

Dans le 2^e niveau de correction, l'objectif est de réduire le nombre de séquences avec singletons considérées comme artéfactuelles. Sont considérés comme artéfactuels, les singletons qui ne sont retrouvés qu'une seule fois dans les captures multiples d'un individu et dont la mutation n'est pas présente dans le polymorphisme de l'espèce. Cela revient pratiquement à réaliser une séquence consensus correspondant à celle recapturée plusieurs fois sans la mutation incriminée pour l'individu considéré. Au final, le jeu de données issu de ce deuxième niveau de correction conserve un nombre limité de séquences clairement distinctes pour chaque individu. Ces allèles correspondent à la fois aux deux allèles attendus

chez les individus hétérozygotes ou à des copies différentes d'un même gène. Dans le cas de figure d'un gène présentant 2 copies paralogues, il est théoriquement possible d'obtenir au moins 4 séquences différentes pour un individu hétérozygote. Cependant, compte tenu de l'effort de recapture, il est souvent difficile de conclure qu'un individu à un allèle constitue un vrai homozygote et cela induit nécessairement un biais dans les analyses.

Le troisième niveau de correction permet de minimiser ce biais en fournissant un jeu de données appelé aussi jeu de données de **Haute Qualité (HQ)**. Afin d'homogénéiser le nombre de séquences recapturées entre différents individus, seule une séquence sera conservée. Ce sont les séquences les plus recapturées qui seront conservées dans le jeu de données HQ. Par convention, lorsque deux séquences ont été recapturées le même nombre de fois, ce sont préférentiellement les allèles nouveaux qui sont gardés, si les deux allèles sont déjà présents, c'est la première séquence du fichier d'alignement qui sera conservée. La population de *Bathymodiolus aff. thermophilus* de 37°S sur la dorsale Pacifique-Antarctique pour laquelle nous n'avons que deux individus a fait l'objet de clonages individuels. Dans ce cas particulier, l'effort de séquençage a été conséquent ($\times 8$ le nombre d'individus (n), au lieu de $2n$ dans le protocole de clonage avec marquage/recapture, soit un total 16 clones pour les 2 individus), ceci afin de pouvoir récupérer la totalité des allèles de ces deux individus et pouvoir les conserver dans le jeu de donnée HQ.

C'est le jeu de données HQ, avec un nombre plus faible de séquences mais constituant un jeu non biaisé, qui a été utilisé pour les analyses réalisées au cours de ce travail.

I.3. Méthodes d'analyse

I.3.1. Méthodes utilisées dans la réalisation des arbres phylogénétiques

Les arbres ont été réalisés avec le logiciel MEGA 4 (Kumar et al. 2004), selon la méthode de regroupement des plus proches voisins (Neighbour-Joining ; (NJ)) (Saitou & Nei 1987). Cette méthode est une bonne méthode heuristique⁴ pour estimer l'évolution minimale entre les allèles et elle est couramment employée car elle est rapide. Pour les arbres nucléotidiques, le calcul des distances utilise le modèle de substitutions défini par Tamura et Nei (Tamura & Nei 1993). Ce modèle prend en compte les différences entre probabilité de

⁴ Méthode rapide pour résoudre des problèmes difficiles, mais pas nécessairement optimale.

passage entre les 4 bases nucléotidiques en distinguant les taux de transition (mutation de base purine (A ou G) en base purine ou base pyrimidine (C ou T) en pyrimidine) des taux de transversion (base purine vers base pyrimidine et inversement). Pour la reconstitution phylogénétique basée sur les séquences en acides aminés, la méthode utilisée pour calculer les distances est celle basée sur la Matrice de Jones (Jones et al. 1992). Cette matrice décrit les probabilités de remplacement d'un acide aminé par un autre à partir de l'observation de 59190 mutations sur 16300 protéines chez la drosophile.

I.3.2. Indices de diversité

Les analyses et le calcul des indices de diversité ont été réalisés avec le logiciel DNAsp 4.20.2 (Rozas et al. 2003).

Le premier indice est celui du nombre de sites polymorphes. Il s'agit du nombre de sites polymorphes dans un échantillon de séquences (population ou espèce). Cet estimateur, noté S , dépend de la longueur de la séquence et de la taille de l'échantillon. Il est donc difficilement comparable entre locus et entre échantillons.

Le deuxième indice de diversité nucléotidique, π (π) est le nombre moyen de différences nucléotidiques par site et par paire de séquences (Nei 1987), il est calculé de la façon suivante :

$$\pi = \sum_{ij} x_i x_j \pi_{ij}$$

Avec x_i et x_j les fréquences des haplotypes i et j respectivement et π_{ij} le nombre de différences nucléotidiques entre les haplotypes i et j .

Le troisième indice de diversité calculé est le θ de Watterson (θ_w), basé sur le nombre de sites variables S dans notre échantillon et rapporté à la longueur de l'alignement (Watterson 1975).

I.4. Tests moléculaires

I.4.1. Tests de neutralité

Ces tests sont des méthodes indirectes pour détecter l'effet de la sélection naturelle sur les gènes. Le principe de base est de comparer le niveau d'ajustement entre la diversité observée pour certains gènes, et des valeurs théoriques attendues sous l'hypothèse de l'évolution sous un modèle neutraliste (à l'équilibre mutation dérive).

I.4.1.a. *D* de Tajima

L'estimateur d'un écart à l'évolution neutre défini par Tajima (Tajima 1989a) permet de tester l'écart au modèle neutre sous l'hypothèse d'équilibre mutation dérive, à partir des données de polymorphisme nucléotidique d'une espèce. Cet estimateur, noté *D*, est basé sur la comparaison des deux mesures de diversité nucléotidique énoncées dans le paragraphe précédent : π et θ_w . Sous l'hypothèse de neutralité des mutations et de l'équilibre entre mutation et dérive, ces deux mesures de la diversité sont des estimateurs du même paramètre $\theta=4N_e\mu$, ils doivent être identiques et le *D* de Tajima, qui est la différence entre les deux divisée par la variance attendue, égal à zéro. On peut alors tester si la différence observée entre les deux estimateurs s'écarte significativement de zéro soit en utilisant la variance de cette différence entre les sites nucléotidiques (si la séquence considérée est suffisamment longue pour fournir une distribution normale des différences observées : test *Z*), soit en effectuant un ré-échantillonnage aléatoire des séquences par bootstraps pour obtenir un nombre suffisant de pseudo-valeurs de cette différence permettant d'effectuer le test *Z*.

Un excès de mutations en faible fréquence (singletons) va être préférentiellement détecté par l'estimateur θ_w et la valeur de *D* sera alors négative. Cet excès de mutations, lorsqu'il est significatif, peut être le résultat (i) d'un goulot d'étranglement suivi d'une expansion démographique, (ii) d'un balayage sélectif (perte de la diversité nucléotidique puis début de re-diversification), ou (iii) d'un effet légèrement délétère des mutations polymorphes observées. Au contraire, une valeur positive du *D* indique un excès de mutations en fréquences intermédiaires pouvant s'expliquer comme (i) la marque d'une structure populationnelle, (ii) le résultat d'un goulot d'étranglement récent et suffisamment faible pour avoir permis la subsistance de plusieurs lignées ou (iii) l'action de la sélection balancée. Les effets démographiques vont affecter de manière semblable l'ensemble des gènes des

populations concernées, alors que les effets sélectifs ne vont en affecter qu'un nombre limité. Ainsi, l'une des solutions pour faire la part des choses est d'utiliser une approche multi-locus qui nous permet de trancher au final pour l'une des hypothèses.

1.4.1.b. Test de Hudson-Kreitman-Aguadé

Ce test, nommé test HKA (Hudson et al. 1987), compare les niveaux de variation intra et inter-spécifiques pour un gène d'intérêt et un autre locus supposé neutre. Il est également possible de mener cette analyse au sein des différentes régions des gènes (comme comparaison des variations entre régions codantes et régions introniques, bien qu'idéalement les deux régions comparées doivent avoir des généalogies indépendantes).

L'idée du test HKA est d'utiliser la divergence qui ne dépend que du taux de mutation et le polymorphisme qui ne dépend que de la taille efficace et du taux de mutations sous le modèle neutre. Ainsi, un gène ayant subi un balayage sélectif ou soumis à une contrainte sélective aura un ratio polymorphisme / divergence inférieur à celui des locus neutres.

1.4.1.c. Test de McDonald et Kreitman

Ce test, appelé test MK (McDonald & Kreitman 1991) vérifie si les ratios entre mutations non-synonymes et mutations synonymes diffèrent entre les données de polymorphisme interspécifique (θ_N/θ_S) et les données de divergence entre espèces (K_N/K_S). Un simple test d'homogénéité permet de détecter l'écart à l'hypothèse de neutralité. Si les deux classes de mutations sont neutres, alors leur taux de fixation est déterminé uniquement par la force de la sélection purifiante et n'a pas de raison de différer entre polymorphisme et divergence.

Si $(K_N/K_S) > (\theta_N/\theta_S)$, cela signifie qu'il y a eu de la sélection positive dans au moins l'une des branches conduisant à l'une ou l'autre des espèces. Au contraire, si $(K_N/K_S) < (\theta_N/\theta_S)$, cela indique un maintien d'un excès de polymorphisme non-synonyme à l'intérieur des espèces par rapport à la fraction qui est finalement fixée. Cela pourrait être le cas pour des mutations légèrement délétères. Ce test suppose que le taux de mutation neutre est partagé entre les deux catégories de mutations (synonymes et non-synonymes). Ce test nécessite aussi l'hypothèse d'une taille de population constante, mais pour des raisons différentes de celles

des tests de sélection indirecte (comme le D de Tajima ou le test HKA). L'effet d'un changement de taille efficace va jouer sur l'efficacité de la sélection et l'intensité de la contrainte sélective. Si la taille efficace a été plus faible pendant la divergence, certaines mutations délétères ont pu se fixer alors que d'autres mutations tout aussi délétères ne sont plus observées dans le polymorphisme actuel d'une population de plus grande taille (Eyre-Walker et al. 2006).

I.4.2. Mesures de différenciation génétique entre populations

L'estimateur de différenciation génétique par paire de populations (Φ_{st}) ainsi que les probabilités associées d'écart à zéro ont été estimés avec le logiciel Arlequin 3.11 (Excoffier et al. 2005). Cet estimateur se calcule selon la formule :

$$\Phi_{st} = \frac{\pi_t - \pi_s}{\pi_t}$$

Avec : π_t : la diversité nucléotidique globale estimée sur l'ensemble des populations ;
 π_s : la diversité nucléotidique moyenne des populations.

L'écart des valeurs de Φ_{st} à la valeur zéro (hypothèse de non différenciation) est testé à l'aide de 1000 permutations.

I.5. Quelques questions techniques

I.5.1. Estimation et caractérisation des singletons

La technique de clonage par marquage-recapture est une méthode quantitative. Les allèles séquencés sont tirés au hasard parmi l'ensemble des allèles de la population (en théorie, tous les allèles sont présents lors du clonage). Cependant, l'un des effets collatéraux de cette méthode est le nombre variable d'individus recapturés d'un clonage à l'autre dont les allèles ont été séquencés et le nombre variable de recaptures pour un allèle donné au sein de chaque individu. Cette méthode nous permet donc d'avoir accès à des individus dont les allèles ont été séquencés plusieurs fois. Ces recaptures mettent en évidence le caractère « artéfactuel » de certaines mutations révélant un polymorphisme que l'on pourrait appeler « intra-individuel ».

Le taux d'erreurs induites par la méthode de PCR (Polymerase Chain Reaction), a déjà été estimé précédemment (Saiki et al. 1988, Lundberg et al. 1991). Ce taux d'erreurs associé à des mutations artéfactuelles a pu être estimé pour différentes polymérases et il varie de 2.0×10^{-5} pour la *Taq* polymérase à 1.6×10^{-6} pour la *Pfu*. Ce phénomène se traduit par l'apparition de singletons artéfactuels et a été souvent négligé dans les études de génétique des populations ou de phylogénie car l'erreur induite est inférieure au niveau de divergence moyen qui peut être observé entre deux haplotypes de la même espèce (Kwiatowski et al. 1991). Ces singletons ont pour conséquence d'augmenter la diversité allélique observée ou l'hétérozygotie observée mais ce biais reste négligeable (Kwiatowski et al. 1991, Palumbi & Baker 1994).

Un autre type d'erreur peut être induit par l'amplification par PCR, il s'agit de la recombinaison *in vitro* (Cronn et al. 2002). Ces recombinants sont néanmoins facilement identifiables par notre méthode grâce à l'identification des allèles par les 'queues' nucléotidiques. En effet, ces recombinants forment une troisième classe d'allèles issue des 2 allèles initiaux qui peuvent être facilement identifiables lorsque la recapture est conséquente. Ils peuvent alors être supprimés de notre jeu de données.

Enfin, il est possible que les singletons observés correspondent à des mutations somatiques induites par les conditions extrêmement mutagènes du milieu hydrothermal (fort niveau de radioactivité naturelle dans les tissus des organismes : (Cherry et al. 1992, Kadko 1996), concentrations cellulaires élevées en métaux lourds (Cosson et al. 2008), nombreux radicaux libres générés par l'oxydation des composés soufrés : (Pruski & Dixon 2003, Marie et al. 2006).

Le nombre très élevé de singletons retrouvé dans le polymorphisme des modioles et l'allélisme intra-individuel nous a donc naturellement conduit à déterminer si les singletons observés dans nos séquences sont des artefacts (erreurs de *Taq*, de réplication pendant le clonage ou des problèmes liés au séquençage) ou s'ils constituent une réalité biologique (mutations somatiques).

I.5.2. Première étude méthodologique

Nous avons écarté les biais d'amplification et de clonage en utilisant une population « test » (pop 37°S, n=2 individus). Les gènes ont été amplifiés en utilisant une *Taq* haute fidélité type *Pfu* et le clonage a été réalisé de manière individuelle en utilisant une souche de bactéries compétentes neuve et garantie non recombinante.

Après clonage, nous avons appliqué un taux de recapture de 4n, c'est-à-dire bien supérieur à celui habituellement utilisé dans la technique de marquage/recapture (2n) (Bierne et al. 2007). Le séquençage a été effectué dans les deux sens permettant de corriger directement les séquences entre elles. Les mutations restantes ont ensuite été vérifiées sur le chromatogramme évitant ainsi les erreurs générées par l'annotation automatique du séquenceur (par exemple lorsque deux pics apparaissent peu décalés ou superposés). Les résultats ne sont présentés que pour 7 gènes (GF1_A, GF1_B, HSP, COL1, SULFO, LYSO et SAHH) car les séquences des gènes ITS2, COI et EF1 α , avaient déjà été obtenues selon le protocole 'classique' (*Taq* non *Pfu*, ni de bactéries neuves) et ne pouvaient donc pas être incluses dans cette étude.

Après avoir réattribué les séquences à chaque individu, celles-ci ont été classées dans quatre catégories selon si elles étaient strictement identiques, ou si elles possédaient, 1, 2 ou plus de 2 mutations. Les résultats sont consignés dans le Tableau I.3. On peut constater que malgré le protocole visant à éliminer les mutations artéfactuelles ; il persiste des séquences avec des singletons pour l'ensemble des locus. Il est attendu de trouver des allèles divergents pour les individus hétérozygotes. Dans ce cas et si l'on suppose que le gène n'est présent qu'en une copie unique, on s'attend à ne trouver que 2 groupes de séquences maximum. Les gènes HSP et LYSO illustrent bien des individus hétérozygotes car 2 groupes de séquences identiques sont retrouvés (deux paires de séquences identiques pour les individus avec le HSP, et pour le LYSO : une paire plus un groupe de 4 séquences identiques). La présence de certaines séquences ayant accumulé plus de deux mutations (jusqu'à 5 pour HSP), peut révéler l'existence de copies multiples ou de paralogues. Les séquences du gène SAHH obtenues pour l'individu n°2 sont révélatrices de l'excès de mutation qui persiste malgré nos précautions. En effet, même si 5 séquences sont rigoureusement identiques, 8 séquences obtenues présentent des mutations. Ces huit séquences, toutes différentes les unes des autres ne peuvent donc pas correspondre au deuxième allèle d'un hétérozygote. N'ayant pas d'excès

de mutations entre séquences (maximum deux) la présence d'une lignée divergente semble peu probable.

Les résultats obtenus ne sont pas différents de ceux trouvés pour les clonages en population utilisant une Taq polymérase moins fidèle pour l'amplification. Ces résultats nous permettent donc de conclure que les mutations observées sont sans doute bien réelles au sein des individus analysés. Il est d'ailleurs à noter que la technique de marquage-recapture appliquée à d'autres bivalves, notamment *Crassostrea gigas* et *Mytilus edulis* fournit des valeurs de singletons 'artéfactuels' moindre avec les mêmes outils moléculaires (données non montrées).

Tableau I.3. Bilan des recaptures réalisées sur les 2 individus de 37°S après amplifications avec de la *Taq Pfu* et des bactéries non recombinantes. **n** : nombre de séquences de qualité obtenues après séquençage.

gène	individu de 37°S	n	seq identiques	1 mutation	2 mutations	mutations > 2
GF1_A	n°1	1	1	.	.	.
	n°2	4	3	.	1	.
GF1_B	n°1	1	1	.	.	.
	n°2	6	3	2	1	.
HSP	n°1	7	2 paires	2	1	1 (5 mutations)
	n°2	5	2 paires	.	.	.
COL	n°1	5	2	.	1	2
	n°2	3	0	.	.	3
SULFO	n°1	6	5	1	.	.
	n°2	5	3	1	.	1
LYSO	n°1	0		.	.	.
	n°2	15	4, plus une paire	4	3	2
SAHH	n°1	3	0	3	.	.
	n°2	13	5	4	4	.

I.5.3. Deuxième étude : basée sur les attendus théoriques entre mutations synonymes et non-synonymes

Afin de savoir si les singletons observés dans le polymorphisme intra-individuel constituent des artéfacts méthodologiques ou des mutations somatiques, nous avons voulu tester l'hypothèse suivante : les mutations artéfactuelles doivent ségréger indifféremment sur les sites synonymes et non-synonymes dans les parties codantes des gènes testés alors que les mutations somatiques délétères doivent nécessairement être purgées d'une génération à l'autre et donc être moins fréquentes sur les sites non synonymes. Nous avons donc décidé de nous

intéresser au nombre et à la position (site synonyme ou non-synonyme) des singletons. Pour réaliser l'étude la plus robuste possible, nous avons choisi le gène avec la plus grande séquence codante possible, soit le gène codant pour la sulfotransférase qui est constitué de 216 bp de codant. Le HSP possède une plus grande séquence codante mais il possède des copies multiples et ne peut donc pas être utilisé ici.

Le jeu de séquences comprend l'ensemble des séquences recapturées plusieurs fois au niveau de chaque population auquel nous avons soustrait les séquences rigoureusement identiques au sein de chaque individu (dans ce cas, une seule séquence est gardée et le nombre de séquences identiques recapturées est noté). Certains individus se retrouvent donc avec un grand nombre de séquences, souvent uniquement différenciées par des singletons.

Notre étude, ne concerne que les mutations 'artéfactuelles' uniques à l'origine d'un polymorphisme intra-individuel. Le terme 'singleton' n'est pas adapté pour décrire ces mutations car il définit habituellement des allèles hétérozygotes du polymorphisme vrai et pourrait donc porter à confusion. Nous avons donc défini un nouveau nom pour définir ces mutations intra-individuelles : les « MUR » pour Mutation Unique de Recapture. Dans la suite de la discussion, nous distinguerons donc les MUR et nous réserverons le terme singletons pour le polymorphisme vrai.

Dans notre étude il était important de ne travailler que sur les MUR et d'éliminer tous les singletons. Nous avons donc appliqué une correction très stricte du jeu de donnée, en ne conservant que les MUR qui n'étaient retrouvées ni dans la population de l'individu étudié, ni dans aucune autre population de l'espèce. De cette manière le risque de conserver des vrais singletons est extrêmement réduit.

Notre analyse consiste à repérer les MUR au sein de chaque population, à déterminer si elles sont synonymes ou non synonymes, et à les rapporter au nombre théorique de sites potentiellement synonymes et non synonymes. Nous avons utilisé DnaSP pour déterminer le nombre moyen théorique de sites synonymes et non synonymes (Tableau I.4). Ce type de comptage fait débat car il n'est pas possible prendre en compte les doubles dégénération sur les différentes position des codons (Bierne & Eyre-Walker 2006). On a cependant considéré que le comptage des sites était une estimation fiable pour notre étude.

Nous avons ensuite rapporté le nombre de mutations observées pour chaque classe au nombre de sites potentiels estimés pour chaque classe afin de comparer les fréquences d'apparition de chaque type de singletons (Tableau I.5).

Tableau I.4. Nombre de singletons et de sites potentiels pour les mutations synonymes et non synonymes, observé pour chacune des populations de *Bathymodiolus* sur le gène de la Sulfotransférase

	Nombre de séquences	MUR Syn	MUR Non Syn	Sites Syn	Sites Non Syn
<i>B. azoricus</i> (MG)	60	8	14	45,8	167,2
<i>B. azoricus</i> (Rb)	53	6	11	46,61	169,39
<i>B. puteoserpentis</i>	41	7	10	46,86	169,14
<i>B. thermophilus</i> (7'S)	41	1	13	43	161
<i>B. thermophilus</i> (21'S)	49	11	11	46,63	169,37
<i>B. thermophilus</i> (37'S)	11	1	2	46,64	169,36

Tableau I.5. Diversité par site (nombre de MUR observées / nombre de sites potentiels moyen pour ce type de mutation) (calculée d'après le Tableau I.4).

ratios	Nombre d'individus	Nombre de séquences recapturées	Nombre moy seq/ indiv	Syn	Non Syn
<i>B. azoricus</i> (MG)	14	60	4.29	0.1747	0.0837
<i>B. azoricus</i> (Rb)	13	53	4.08	0.1287	0.0649
<i>B. puteoserpentis</i>	12	41	3.42	0.1494	0.0591
<i>B. thermophilus</i> (7'S)	10	41	4.10	0.0233	0.0807
<i>B. thermophilus</i> (21'S)	9	49	5.44	0.2359	0.0649
<i>B. thermophilus</i> (37'S)	2	11	5.50	0.0214	0.0118

Ces rapports ont ensuite été reportés sur un graphe en fonction du nombre moyen de séquences recapturées par individu, et ce, pour chacune des populations (Figure I.6)

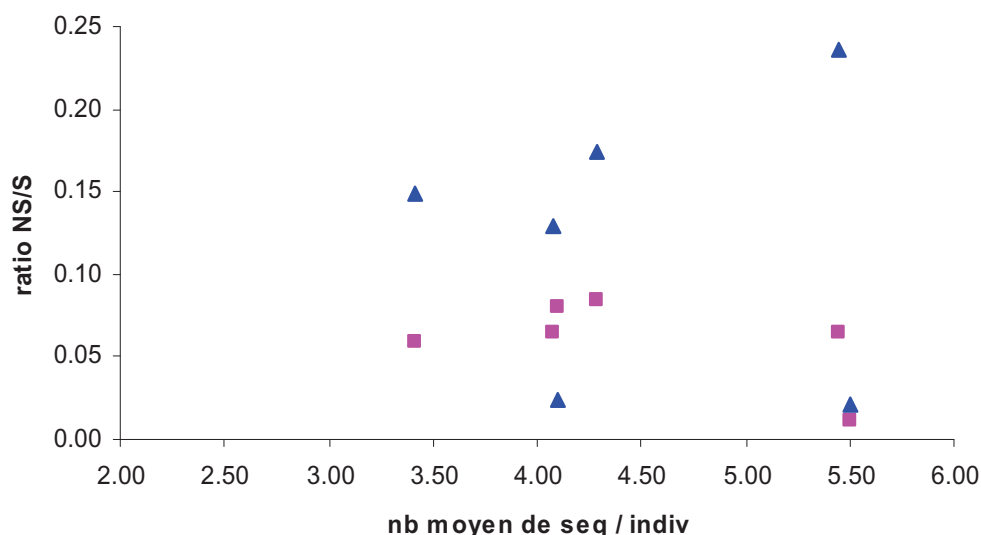


Figure I.6. Evolution du nombre de MUR synonymes et non synonymes rapportés à leur nombre maximum de sites respectifs en fonction du nombre moyen de séquences capturées dans chaque population. Triangles bleus : ratio mutations synonymes ; Carrés roses : ratio mutations non synonymes.

La Figure I.6 montre que quel que soit le nombre de séquences recapturées par individu, le taux de mutations non synonymes est relativement constant. En revanche, le taux de mutations synonymes varie considérablement entre les sites.

Ces différences ont été testées à l'aide d'un test-t pour données appariées et à l'aide d'un test non paramétrique Mann-Whitney (Tableau I.6). Dans les deux cas, les différences observées sont significatives.

Tableau I.6. Résultats des 2 tests statistiques effectués pour comparer les nombres de singletons synonymes et non synonymes ajustés

test-t données appariées:		
Ecart-moyen		0,085
DDL		5
t		4,128
p		0,0091
test non paramétrique:		
p		0,031

Ce décalage entre les nombres de MUR synonymes et non-synonymes constitue une preuve de la réalité biologique de ces mutations créant un polymorphisme intra-individuel. En effet, si ces singletons résultaient d'un artefact méthodologique (effet *Taq* ou autre), les mutations S et NS auraient été affectées de la même manière et leur valeurs seraient confondues. Le décalage observé peut être interprété comme étant le reflet d'une forte sélection purificatrice (contre-sélection des singletons non-synonymes, légèrement délétères). Le niveau constant des mutations non-synonymes constitue le fardeau génétique des individus.

Nos résultats peuvent s'expliquer comme étant le reflet de la nature extrêmement mutagène de l'environnement hydrothermal générant de nombreux dommages à l'ADN. Ce fort potentiel mutateur de l'environnement hydrothermal profond a déjà été noté par Cherry et al. (1992) sur les polychètes Alvinellidae et par Pruski & Dixon (2003) chez les Alvinellidae et les Bathymodiolinae.

Il serait intéressant de compléter nos observations en réalisant cette étude sur les parties codantes de l'ensemble des gènes de *Bathymodiolus* ou en comparant nos résultats

avec ceux que nous pourrions acquérir sur d'autres bivalves côtiers comme *Mytilus edulis* ou l'huître *Crassostrea gigas* afin de déterminer s'il s'agit bien d'un effet environnemental.

Pour compléter cette étude et réduire le nombre de singletons artificiels induits par la PCR, il faudrait utiliser une polymérase *Pfu*. Cependant l'utilisation de la méthode de marquage/recapture avec l'utilisation d'une *Taq* haute fidélité semble être un bon compromis pour obtenir des séquences de bonne qualité.

I.6. Mise en évidence de gènes paralogues au niveau des arbres étudiés

Certains arbres peuvent présenter une topologie qui s'écarte clairement de l'attendue neutre et correspondre à l'apparition d'une divergence issue de l'évolution d'un gène par duplication. Les soupçons concernant la présence de gènes paralogues au sein de certains marqueurs proviennent de deux types d'observations complémentaires. La première de ces observations correspond au nombre d'allèles différents observés au sein d'un même individu lorsque ces allèles sont recapturés après le clonage. En effet, lorsque plus de deux allèles sont fréquemment recapturés chez les individus d'une même population pour le même gène, ces résultats conduisent à une forte présomption de paralogie. La deuxième observation repose sur la topologie des arbres phylogénétiques car eux seuls permettent de retracer la généalogie des allèles et donc de faire la différence entre des clades d'allèles très divergents ou des gènes paralogues obtenus par duplication. Le problème qui peut se poser avec les arbres de coalescence, c'est que lorsque les gènes dupliqués ne sont pas de la même taille, il peut y avoir une amplification préférentielle des allèles pour le gène le plus court. Dans l'arbre de coalescence final, le gène le plus long ne sera alors représenté que par un petit nombre d'allèles et formera un clade rare et divergent qu'il est possible de confondre avec un clade formé par de vrais lignées divergentes non balayées provenant d'un polymorphisme ancestral.

I.6.1. Un exemple simple : le cas du GF1.

Au cours de ce travail, l'exemple le plus frappant de gènes paralogues est celui observé au gène codant pour la ferritine GF1 car les deux gènes ont été co-amplifiés avec le même succès. L'arbre réalisé à partir de ces séquences permet sans ambiguïté de distinguer deux gènes paralogues chez *Bathymodiolus* (Figure I.7).

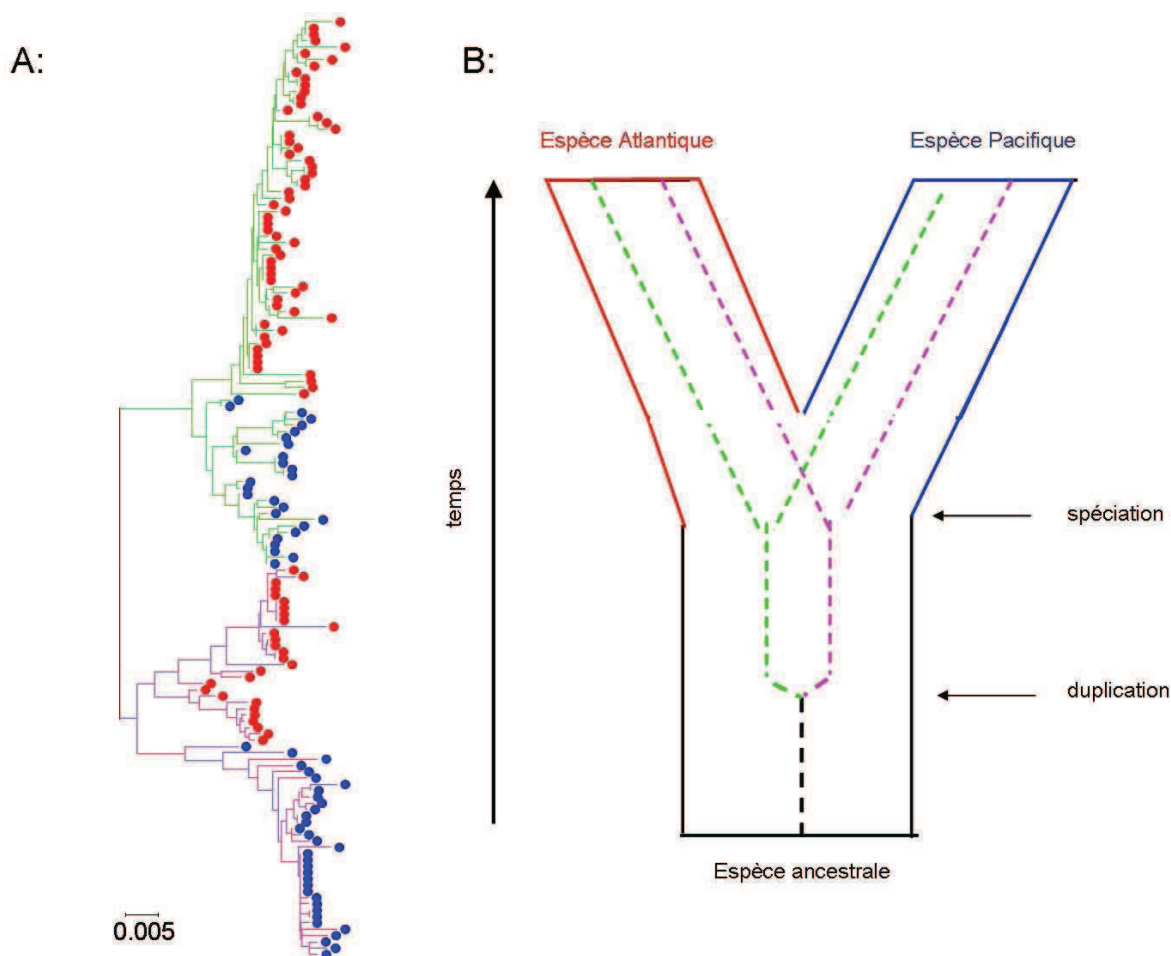


Figure I.7. (A) arbre en NJ obtenu sur la totalité des séquences du gène GF1 révélant la présence de deux gènes paralogues chez les espèces *B. azoricus* (Atlantique) et *B. thermophilus* (Pacifique). (B) Schéma représentant l'effet d'un événement de spéciation faisant suite à un événement de duplication sur la coalescence des allèles tels qu'observé pour le gène GF1 chez les bathymodiales. En bleu : séquences provenant de l'océan pacifique ; en rouge : séquences provenant de l'océan atlantique. Les pointillés verts et violets représentent les deux gènes dupliqués, les pointillés noirs représentent le gène ancestral.

Ici, les gènes se sont dupliqués avant la spéciation des espèces car chacun des gènes paralogues présente une dichotomie en fonction de la provenance géographique des séquences. En revanche les deux espèces atlantiques (*Bathymodiolus azoricus* et *B. puteoserpentis*) et les deux espèces pacifiques (*B. thermophilus* et *B. aff. thermophilus_37S*), sont issues d'événements de spéciation beaucoup plus récents car la coalescence ne permet pas de les distinguer au sein de chaque clade paralogue.

Les gènes paralogues du GF1 seront appelés GF1_A et GF1_B par la suite et feront l'objet d'une étude indépendante dans la suite de ce travail.

I.6.2. Deux exemples compliqués : le cas du lysozyme et du collagène

Pour ces deux gènes, la paralogie est également suspectée car plusieurs groupes d'allèles divergents mélangeant les espèces atlantique et pacifique sont observables sur les arbres. Cependant le faible nombre de séquences échantillonnées à certains clades et la topologie de l'arbre ne nous permettent pas de trancher entre les hypothèses de paralogie ou d'allèles ancestraux (Figure I.8).

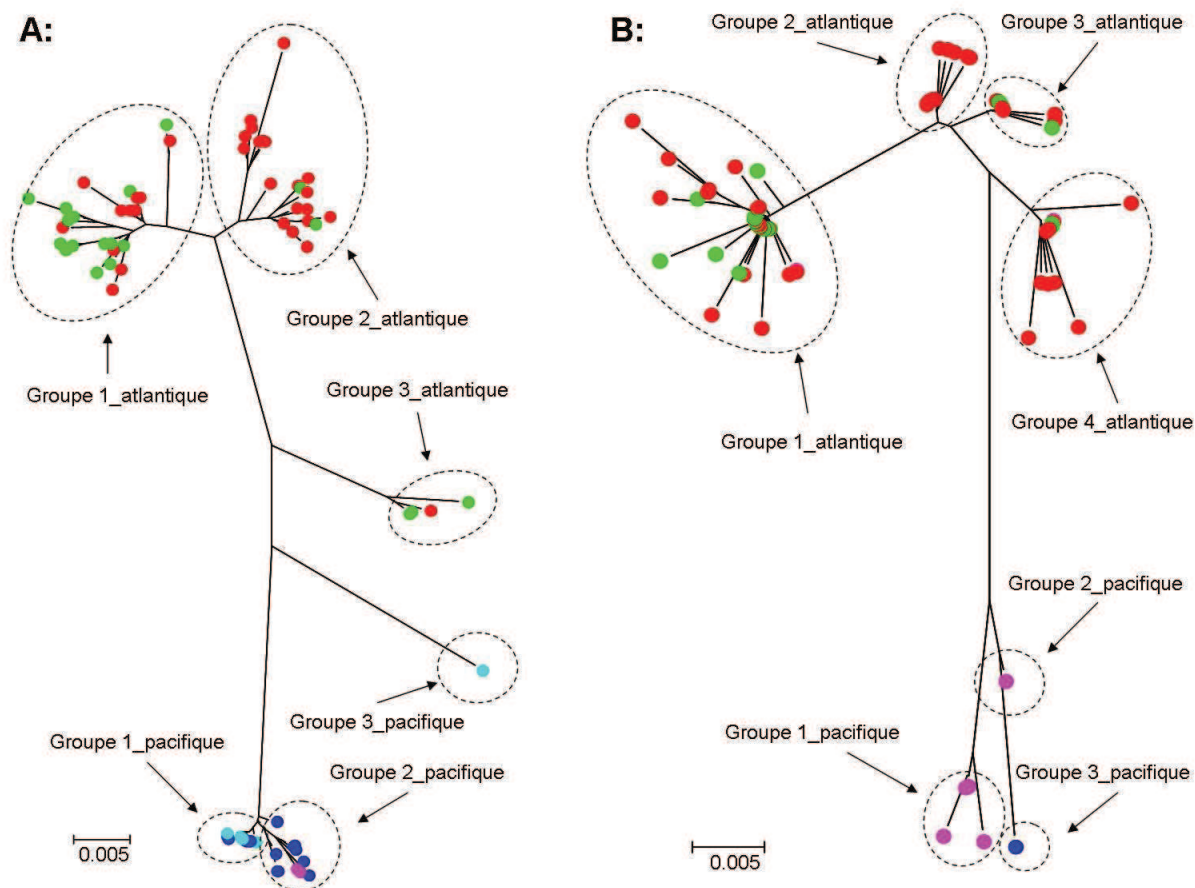


Figure I.8. Arbres NJ non-racinés sous la forme en étoile obtenus à l'aide du jeu de données HQ. (A) Arbre des séquences du Lysozyme. (B) Arbre des séquences du Collagène XIV. Séquence Atlantique : point rouge : *Bathymodiolus azoricus* ; point vert : *B. puteoserpentis*. Séquence Pacifique : point bleu (foncé et clair) : *B. thermophilus* (7°S et 21°S) ; point violet : *B. aff. thermophilus_37°S*. Les cercles en pointillés représentent les différents clades identifiés pour chaque gène à partir des jeux de données 1 et 2 (plus de séquences).

Lorsque l'on regarde l'arbre dans son ensemble, chaque individu devrait avoir au maximum deux allèles. En revanche si certains clades représentent des paralogues, les individus doivent présenter au moins une séquence dans chacun des clades divergents (maximum deux allèles par clade pour les individus hétérozygotes). Le clonage individuel et le séquençage d'un grand nombre d'allèles pour l'ensemble des individus pourraient être une

solution. Cette méthode coûteuse en temps et en argent n'est pas non plus fiable à 100%. En effet, si les amplifications se font de manière préférentielle ou si le nombre de clones séquencés est trop faible, les interprétations seront également erronées.

Pour pouvoir trancher, nous avons donc employé une technique moléculaire utilisant des enzymes de restrictions (RFLP). Le principe de cette technique est de trouver une combinaison d'enzymes de restriction diagnostique des allèles de chacun des clades. Après migration sur gel d'agarose 2% (TBE 1X), les profils de digestion propres à chaque clade permettent de déterminer quels « allèles » sont présents au sein de chaque individu.

Si tous les « allèles » sont présents pour l'ensemble des individus, il s'agit de gènes paralogues. En revanche si chaque individu est constitué d'un maximum de 2 types de séquences (clades), l'hypothèse de paralogie peut être écartée au profit de celle de l'allélisme.

Pour réaliser les profils de digestion, les produits PCR correspondant aux portions de gènes amplifiés sont digérés par un cocktail d'enzymes de restriction préalable repérer sur les cartes de restriction de chaque classe allélique.

Pour le Lysozyme, deux enzymes de restriction ont été utilisées de manière séparée : *ItaI* (incubation du produit de PCR à 37°C pendant 1 heure) et *TaqI* (65°C, 1h). La comparaison des profils obtenus pour l'une et l'autre des digestions enzymatiques permet de classer les allèles présents selon les différents clades observés.

Pour le collagène, deux enzymes de restriction ont également été utilisées: *DraI* et *MaeIII*, mais de manière conjuguée. Pour ce gène, les digestions ont été réalisées de manière successive. Une première digestion a été réalisée avec *DraI* (37°C, 1h), puis ce premier produit de digestion est digéré une nouvelle fois avec l'enzyme *MaeIII* (55°C, 1h), ceci afin de respecter un gradient de salinité croissante pour les tampons de digestion. C'est donc un profil de restriction unique qui permet de classer les « allèles » présents au sein de chaque individu dans les différents clades observés.

Cette approche nous a permis de déterminer que sur les trois « clades » observés pour le Lysozyme, l'un d'entre eux (le clade 3 pour l'Atlantique et le Pacifique) était constitué d'un paralogue et était diagnostiqué pour tous les individus. Les séquences de ce gène

paralogue, capturées en faibles fréquences dans notre jeu de données ont été supprimées des analyses.

Au contraire, le collagène XIV, montre des profils de restriction variables permettant d'affirmer qu'aucun des « groupes » n'est présent pour l'ensemble des individus (Figure I.9-A). Au contraire, il a même été possible de diagnostiquer des individus homozygotes pour chaque groupe d'allèle (Figure I.9-B), confirmant ainsi que même si les clades observés dans l'arbre semblent divergents, il s'agit de véritables allèles. Les séquences provenant de chacun des clades ont donc été conservées pour les analyses.

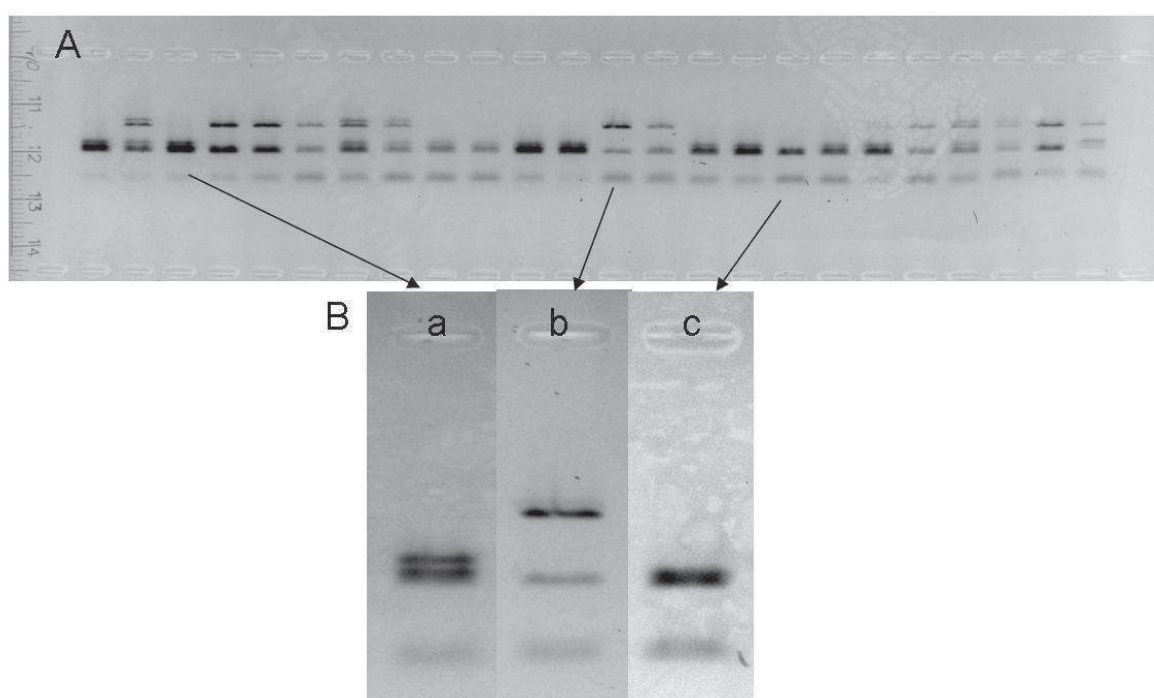


Figure I.9. Exemple de profils de restrictions obtenus par l'action successive de deux enzymes de restriction sur le gène COL. (A) : profils de restriction pour 24 individus ; (B) : détails d'individus « homozygotes » pour chacun des trois groupes suspectés être des paralogues (a, b, c).

Chapitre II : Phylogénie des Bathymodiolinae

II.1. Introduction

Au niveau des sources hydrothermales profondes situées le long des dorsales actives comme sur les zones de suintements froids présentes le long des zones de subduction, il existe un écosystème biologique original et unique où la production primaire n'est pas basée sur la photosynthèse, mais sur la chimiosynthèse bactérienne. Ces environnements particuliers peuvent être regroupés sous le vocable d'environnements réduits profonds. Les moules du genre *Bathymodiolus* (Kenk & Wilson 1985) constituent l'un des taxons le plus fréquemment rencontrés dans ces milieux où elles dominent le peuplement. A ce titre, elles peuvent être considérées comme des espèces-ingénieurs des environnements réduits au même titre que les vestimentifères ou les polychètes Alvinellidae (Jollivet 1996). Ces espèces forment donc en général des moulières denses autour des points d'émission de méthane ou de sulfures d'hydrogène ou elles servent d'habitats à de nombreuses espèces de la petite macrofaune. Plus de vingt espèces ont été recensées en 2008 (Won et al. 2008) et de nombreuses espèces sont encore en cours de description. Les espèces décrites se répartissent dans toutes les mers du monde : parmi celles-ci, neuf espèces proviennent des bassins arrière-arc de l'océan Pacifique Ouest avec une richesse spécifique toute particulière des eaux japonaises (6 des 9 espèces) (Von Cosel et al. 1994, Hashimoto & Okutani 1994, Von Cosel & Marshall 2003, Okutani et al. 2004) ; une espèce provient de l'Océan Indien (Hashimoto 2001) ; sept espèces ont été recensées dans l'océan Atlantique, dont 4 dans le Golfe du Mexique (Von Cosel et al. 1994; Von Cosel & Olu 1998; Gustafson et al. 1998; Von Cosel et al. 1999; Von Cosel 2002) ; et enfin, une espèce est localisée au niveau de la dorsale du Pacifique Est (Kenk & Wilson 1985). De nombreuses espèces de modioles récoltées sur des zones hydrothermales/suintements froids découvertes récemment sont encore en cours de description (Von Cosel 2002, McKiness et al. 2005) et sont souvent désignées par leur localité (par exemple *B. sp.*_Bassin de Manus ; *B. sp.*_Est Antarctique-Pacifique, *B. sp.*_ Juan de Fuca). Ces modioles se distribuent en 2 grandes sous-familles : les Bathymodiolinae et les Modiolinae (Samadi et al. 2007). Au sein des Bathymodiolinae, trois genres ont été recensés : *Bathymodiolus*, *Gigantidas* et *Tamu*. La grande majorité des espèces décrites appartiennent au genre *Bathymodiolus* bien que cette classification soit uniquement basée sur des critères

morphologiques. Deux espèces ont été décrites dans le genre *Gigantidas* (Von Cosel & Marshall 2003; Hashimoto & Yamane 2005) au niveau des sources hydrothermales de Nouvelle Zélande et des îles Bonin (Kaikata Seamount, Sud Japon) et une seule espèce est attribuée au genre *Tamu* (Gustafson et al. 1998) au niveau des zones de suintements froids de faible profondeur du Golfe du Mexique.

L'autre grande sous-famille de Mytilidae représentée dans les communautés des environnements réduits profonds est celle des Modiolinae. Cette sous-famille présente des lignées extrêmement divergentes (Samadi et al. 2007) à la base de l'arbre des Modiolinae. Trois genres sont actuellement bien connus et présentent l'originalité de regrouper des espèces qui, en plus des environnements réduits classiques (hydrothermaux et suintements froids), colonisent les bois coulés et les carcasses de baleine. Le genre *Benthomodiolus* a été décrit dans le Pacifique Ouest (Nouvelle-Zélande) (Dell 1987), le genre *Idas* dans les zones de suintements froids de l'Atlantique (Golfe du Mexique), de la Méditerranée et du Pacifique Est (Gustafson et al. 1998, Dell 1987) et, le genre *Adipicola*, dans les zones de suintements froids du Pacifique Ouest (Okutani et al. 2004).

En dépit de nombreuses études taxonomiques et moléculaires sur le sujet, la phylogénie des Bathymodiolinae n'est toujours pas résolue (Craddock et al. 1995, Miyazaki et al. 2004, Iwasaki et al. 2006). Ceci est dû en partie au manque de résolution des marqueurs ribosomiques au niveau des branches profondes et à l'effet de saturation des séquences mitochondriales (Samadi et al. 2007). Les principales explications évoquées à la forte diversité spécifique des *Bathymodiolinae* dans la littérature sont : la diversité des habitats réduits (hydrothermaux suintements froids, os de baleine ou bois coulés) (Won et al. 2002, Jones et al. 2006, Samadi et al. 2007), la profondeur (O'Mullan et al. 2001, Miyazaki et al. 2004), et la nature des symbioses au niveau des zones branchiales (Won et al. 2008).

Les approches et les outils utilisés pour étudier la phylogénie et les relations entre espèces de *Bathymodiolus* sont presque aussi nombreux que les espèces elles-mêmes, avec des approches allozymiques (Craddock et al. 1995), le couplage allozymes/gène mitochondrial (O'Mullan et al. 2001, Smith et al. 2004), le couplage protéines/gène mitochondrial (Miyazaki et al. 2004), ainsi que des approches multi-locus (Iwasaki et al. 2006, Jones et al. 2006, Olu-Le Roy et al. 2007, Samadi et al. 2007, Won et al. 2008).

Malgré l'évolution des techniques et le nombre sans cesse croissant de séquences disponibles dans les banques, la phylogénie des bathymodiales et, plus particulièrement celle concernant les taxons hautement divergents, reste incertaine.

La phylogénie proposée dans cette thèse est innovante car elle utilise conjointement deux locus (le locus mitochondrial COI et le locus nucléaire ITS2) sur le clade le moins divergent : celui des Bathymodiolinae. Cette approche n'avait pour l'instant été utilisée que pour une étude à l'échelle régionale (Olu-Le Roy et al. 2007). Cette analyse porte sur l'ensemble des séquences disponibles pour les espèces de *Bathymodiolinae* en utilisant quelques *Modiolinae* peuplant les environnements réduits profonds comme groupes externes. Les informations obtenues pour chacun de ces deux locus sont confrontées et des conclusions sont tirées quant à l'histoire évolutive des bivalves profonds.

Ce chapitre peut être considéré comme un préambule au corps de la thèse. Il définit le cadre actuel des connaissances scientifiques dans lequel nous allons évoluer et permet aussi de présenter et d'expliquer en détail les principaux résultats sur lesquels se basent les hypothèses qui seront évoquées au cours de la thèse.

Les données phylogénétiques sont utilisées comme support principal permettant d'étudier des domaines aussi variés que (i) la notion d'espèce chez les Bathymodiolinae, (ii) leur histoire évolutive et (iii) étayer les principaux scénarios de colonisation des dorsales.

Les effets pouvant brouiller le signal phylogénétique (effets démographiques, introgression et sélection) seront également discutés.

II.2. Matériels et méthodes.

II.2.1. Choix des marqueurs

Le premier marqueur, le locus mitochondrial COI, est classiquement utilisé dans les analyses de «barcoding» (Hebert et al. 2003, Marshall 2005) et constitue un des marqueurs les plus couramment utilisés en phylogénie moléculaire, notamment celles concernant les Bathymodiolinae. Il est en général très résolutif à l'échelle du genre ou de la famille (Hebert et al. 2003). En revanche, son évolution n'est pas neutre car ce gène, codant, est soumis à

forte sélection purifiante et aurait pu au cours de son évolution subir de nombreux balayages sélectifs, notamment chez les invertébrés marins: organismes présentant de grandes tailles efficaces de population (Bazin et al. 2006). Le second marqueur correspond à la région intergénique ribosomique ITS2 (internal transcribed Spacer 2) située entre les régions 5,8S et 28S de l'ADN ribosomique. Il s'agit d'une séquence traduite mais non directement codante, permettant l'accumulation de polymorphisme informatif d'un point de vue phylogénétique à l'échelle du genre voire de la famille.

Le gène mitochondrial COI est un élément indispensable de la chaîne respiratoire mitochondriale et permet la production d'une grande partie de l'énergie nécessaire à la cellule. De part sa fonction, cette protéine va être soumise à de grandes pressions de sélection purifiante, notamment chez les espèces vivant dans les milieux réduits souffrés (Powell & Somero 1983, Vetter et al. 1991). Les mutations pouvant se fixer seront essentiellement synonymes. L'espaceur ITS2, à l'inverse, est une séquence non codante transcrite. Cependant, l'espaceur a un rôle essentiel dans la production des pré-rRNA (Michot et al. 1999) entrant dans le cycle de la synthèse des ARN ribosomiaux. La fonction biologique des ITS2 n'est pas réellement connue, mais il semble que la structure secondaire ait un rôle important dans l'épissage des gènes ribosomiques (Michot et al. 1999; Joseph et al. 1999). Cette structure est conservée à travers l'ensemble des eucaryotes et ce, malgré une forte divergence entre les séquences (décrite comme « dramatique » par ces mêmes auteurs !). Les ITS2 sont présents en copies multiples dans le génome et peuvent présenter de la variation intra génomique individuelle (Harris & Crandall 2000) lorsque l'évolution concertée ne s'effectue pas bien entre les différentes régions chromosomiques des NORs (Nucleolar Organizer Regions). Ces variations peuvent perturber les relations entre allèles à échelle intra populationnelle et intra spécifique, mais en font un bon marqueur pour l'étude des relations phylogénétiques entre espèces (Schlötterer et al. 1994).

II.2.2. Acquisition des séquences

La plupart des séquences du gène mitochondrial de la Cytochrome Oxydase 1 sont actuellement disponibles dans les banques de données pour les sous-familles des Bathymodiolinae et Modiolinae des environnements réduits, et ont été récupérées pour réaliser cette reconstruction phylogénique (les numéros d'accèsion sont fournis dans l'annexe 3).

Quelques séquences additionnelles ont été obtenues au cours de la thèse par une approche PCR au niveau des espèces plus particulièrement étudiées dans ce travail et, pour une espèce pouvant avoir un rôle clef dans les échanges trans-panaméens passés. Ces séquences ont été amplifiées avec les amorces suivantes :

COI sens : 5'-ATA AAA AGA TGT ATT RAA RTG ACG-3'

COI anti-sens : 5'- TGT GGT CTG GAA TAA TTG GAA-3'

Après amplification et contrôle sur gel d'agarose 1% (TBE X1, visualisation avec du bromure d'éthidium), les produits PCR ont été purifiés sur membrane Millipore (X96), récupérés dans 50 µL d'eau milliQ puis séquencés de manière directe à l'aide du kit de séquençage BigDye TerminatorTM (PerkinElmer) sur un séquenceur automatique à capillaires ABI 3100. Le séquençage a été effectué dans les deux sens par la plateforme GENOMER, à la Station Biologique de Roscoff (les conditions d'amplifications sont détaillées dans le chapitre matériel et méthodes générales de la thèse).

Au niveau de la reconstruction phylogénétique des Bathymodiolinae à l'aide du marqueur COI, nous avons préféré restreindre le nombre de séquences par taxon pour éviter de donner plus de poids aux taxons plus spécifiquement analysés dans cette étude. Les séquences ont été choisies selon le critère suivant : ne sélectionner que les haplotypes différents trouvés au sein de chaque population (le nombre de séquences à choisir reste limité compte tenu de la faible diversité mitochondriale chez les Bathymodiolinae). Concernant les deux localités pour lesquelles seuls deux individus étaient en notre possession (sud EPR 37°S et zone de subduction du Costa Rica), nous avons conservé les deux séquences dans le jeu de données même si elles étaient identiques. Un récapitulatif des espèces utilisées dans cette analyse est indiqué dans le Tableau II.1. Ce récapitulatif regroupe les données de classification morphologique, la provenance géographique, l'habitat (Hydro : sources hydrothermales ; Suintements : suintements froids) ; et le type d'endosymbionte des modioles (S : sulfo-oxydant ; M : méthanotrophe ; SM : sulfo-oxydant et méthanotrophe).

Tableau II.1. Liste des espèces utilisées dans notre étude. Pour le type de symbionte, S : sulfo-oxydant ; M : méthanotrophe ; ? : non déterminé

Famille	Sous famille	Genre	Espèce	Provenance	Habitat	symbionte		
Mytilidae	Bathymodiolinae	Bathymodiolus	Bathy_thermophilus	Pacifique Est	Hydro	S		
			Bathy_aff_thermophilus_Sp37S	Pacifique Est	Hydro	S		
			Bathy_septemdiarium = mariana	Pacifique Ouest	Hydro	S		
			Bathy_brevior	Pacifique Ouest	Hydro	S		
			Bathy_marisindicus	Indien	Hydro	S		
			Bathy_brooksi	Atlantique_GoM	Suintements	SM		
			Bathy_boomerang	Atlant. Prisme Barbades	Suintements	SM		
			Bathy_aff_boomerang	Atlant. Prisme Barbades	Suintements	SM		
			Bathy_heckerae	Atlantique GoM	Suintements	SM		
			Bathy_puteoserpentis	Atlantique_MAR	Hydro	SM		
			Bathy_azoricus	Atlantique_MAR	Hydro	SM		
			Bathy_sp._Lau_A	Pac. Sud-Ouest (Lau)	Suintements	?		
			Bathy_sp._Lau_B	Pac. Sud-Ouest (Lau)	Suintements	?		
			Bathy_sp._New Zealand	Pac. Sud-Ouest (NZ)	Suintements	?		
			Bathy_sp._manusensis	Pacifique Ouest (Manus)	Hydro	S		
			Bathy_aduloides	Pacifique Ouest (Japon)	Hydro. et saint.	S		
			Bathy_hirtus	Pacifique Ouest (Japon)	Suintements	?		
			Bathy_sp._(nouveau_clade_papua)_A	Pacifique Ouest	Hydro	?		
			Bathy_sp._(nouveau_clade_papua)_B	Pacifique Ouest	Hydro	?		
			Bathy_sp._Kikaijima	Pacifique Ouest	?	?		
			Bathy_japonicus	Pacifique Ouest (Japon)	Hydro. et saint.	M		
			Bathy_securiformis	Pacifique Ouest (Japon)	Suintements	M		
			Bathymodiolus_tangaroa	Pacifique Ouest (NZ)	Suintements	S		
			Bathy_platfrons	Pacifique Ouest (Japon)	Hydro. et saint.	M		
			Bathy_aff_childressi	Atlant._Prisme Barbades	Suintements	M		
			Bathy_childressi	Atlantique_GoM	Suintements	M		
			Bathy_mauritanicus	Atlant. Marge Gabon	Suintements	M		
			Bathy_elongatus	Pacifique Ouest (Nord Fidji)	Hydro	S		
			Bathy_sp_CostaRica	Pacifique Est (CostaRica)	Suintements	?		
			Bathy_Sp_A-zaire (proche Aff_boom)		Suintements	SM		
			Bathy_Sp_B-nigeria (proche mauritanicus)		Suintements	?		
			Bathy_sp._3-NZ3	Pacifique Ouest (NZ)	Hydro	S		
				Gigantidas		Gigant_gladius	Pacifique Ouest (NZ)	Hydro
		Gigant_sp._Aitape_Papua	Pacifique Ouest (Manus)			Hydro	?	
		Gigant_sp._Ashizuri	Pacifique Ouest (Japon)			?	?	
		Gigant_horikoshii	Pacifique Ouest (Japon)			Hydro	?	
		Tamu		Tamu_fisheri	Atlant._GoM	Suintements	?	
		Modiolinae	Benthomodiolus	Bentho_sp.	Pacifique Ouest (Japon)	Hydro	?	
				Bentho_lignicola	Pacifique Ouest (NZ)	Bois / baleine	?	
				Bathy_sp._JDF-2005	Pacifique Est (JdF)	Hydro	?	
			Idas		Idas_macdonaldi	Atlantique_GoM	Suintements	S
					Idas_washingtonia	Pacifique Est (Mont. Bay)	Bois / baleine	S
			non classé ou idasola		Idasola_japonica	Pacifique Ouest (Japon)	Os de baleine	?
			Adipicola		Adipi_iwaotakii	Pacifique Ouest (Japon)	Suintements	?
					Adipi_crypta	Pacifique Ouest (Japon)	Os de baleine	?
					Adipi_pacifica	Pacifique Ouest (Japon)	Os de baleine	?

Les séquences du locus nucléaire ITS2 utilisées dans notre étude sont issues de la publication de Olu-Le Roy et al. (2007). Ces données ont été complétées par des séquences supplémentaires obtenues sur les espèces hydrothermales provenant de la dorsale médio-atlantique, de la dorsale du Pacifique Est et de la zone de subduction du Costa Rica à partir d'amplifications par PCR et clonage.

Le locus ITS2 a été amplifié à l'aide d'amorces définies spécifiquement pour les bathymodiales :

ITS2 sens : 5'-ACA TTG CGG CTT TGG GTC AC-3'

ITS2 anti-sens : 5'-GCT TAA ATT CAG CGG GTA CT-3'

Les amplifications ont été réalisées à l'aide d'un robot cycler Perkin-Elmer GenAmp 9700 dans un volume réactionnel de 25 mL contenant 1X tampon PCR, 2 mM MgCl₂, 0,12 mM dNTP, 0,2 mM de chaque amorce, 0,5 U Thermoprime Taq polyméraseTM (ABGene) et 25 ng d'AND cible. Les conditions d'amplification utilisées étaient : une phase de dénaturation à 96°C pendant 3 minutes suivie de 35 cycles d'amplification de 75 secondes à 50°C, 60 secondes à 72°C et 60 secondes à 96°C puis une phase de complémentation de 10 minutes à 72°C (ajout de A aux extrémités du produit de PCR). Les produits de PCR ont ensuite été purifiés par utilisation de mini-colonnes QIAQuickTM PCR purification et, ligaturés dans un vecteur BlueScriptTM T-vector (par ajout de TopoisoméraseI) pour être ensuite clonés dans une souche bactérienne compétente DH5 α . Les clones positifs ont été ensuite criblés par PCR et séquencés par la technologie du Big Dye TerminatorTM (PerkinElmer) sur séquenceur automatique à capillaires ABI 3100

II.2.3. Traitement des séquences

Le jeu complet de séquences (données banques plus séquences additionnelles) comprenait des séquences présentant une forte hétérogénéité de longueur. Les séquences ont donc été alignées puis coupées à l'aide du logiciel BioEdit (Hall 1999) afin de trouver le meilleur compromis entre le nombre de taxons étudiés et le plus grand nombre de bases analysées.

Les arbres phylogénétiques ont été construits à l'aide du logiciel MEGA 4 (Tamura et al. 2007), selon la méthode des proches voisins, Neighbor-Joining (Saitou & Nei 1987), avec 1000 bootstraps, et une distance évolutive calculée selon la méthode de Tamura & Nei (1993)

Le locus COI a également été traduit en acides aminés (code génétique Invertébré mitochondrial). L'arbre issu des séquences protéiques a été construit selon la méthode Neighbor-Joining (Saitou & Nei 1987), et la distance évolutive minimale a été calculée selon la matrice de substitution d'acides aminés de Jones-Taylor-Thornton (JTT) (Jones et al. 1992). Cet arbre a été enraciné à l'aide de séquences de Mytilidae côtiers issues de GenBank (*Mytilus edulis*, *M. galloprovincialis*, *M. trossulus*, *Trichomya hirsuta* (Mytilidae tropicale) et *Brachidontes pharaonis*). Les numéros d'accession de ces séquences sont fournis dans l'annexe 3.

II.2.4. Tests de saturation sur les deux locus.

Nous avons également procédé à une analyse portant sur la troisième base des codons (données COI) et sur le rapport entre le nombre de transitions et de transversions (données ITS2) afin de savoir si les locus utilisés présentaient (ou non) une saturation du signal phylogénétique entre les lignées les plus divergentes. L'effet de saturation existe lorsqu'un marqueur utilisé est mal adapté au niveau taxonomique de l'étude. Dans ce cas, le nombre de mutations accumulées dans la divergence risque d'être tel que certains sites auront été l'objet de plusieurs mutations successives (pouvant induire de l'homoplasie), les rendant ainsi non informatifs du point de vue des méthodes de reconstruction généralement utilisées. La technique que nous avons employée ici consiste à comparer les corrélations entre les valeurs de divergence calculées selon deux méthodes présentant des sensibilités différentes à l'homoplasie.

Sur le COI, nous avons donc choisi 47 séquences, représentatives de l'ensemble des espèces pour mener cette analyse. Dans un premier temps, la distance K2P (Kimura 2 paramètres (Kimura 1980)) a été calculée avec la longueur totale des séquences. Dans un deuxième temps, cette distance a été recalculée à partir des deux premières positions de chaque codon (suppression de la troisième base, mutations synonymes). Une représentation graphique est ensuite réalisée à partir des deux estimateurs appariés de la divergence, sachant qu'une saturation du signal entraînera un effet asymptotique de la régression (lorsque tous les

sites placés en troisième position de codon sont mutés alors que les sites placés aux deux premières positions peuvent encore accumuler de la divergence). Lorsque l'on réalise une régression comparant les valeurs de divergence calculées entre des espèces trop éloignées, la saturation se traduira donc par l'obtention d'un plateau.

Pour l'espaceur ITS2, locus traduit mais non codant, il n'est pas possible d'utiliser la même méthode. Nous avons donc utilisé une méthode basée sur l'accumulation des transitions. En effet, au lieu de comparer la distance K2P calculée sur l'ensemble de la séquence avec celle calculée à partir des deux premières positions de chaque codon, nous avons comparé la distance K2P globale avec le nombre de transitions observées entre chaque paire d'espèces. Cette méthode repose sur le fait que les transitions s'accumulent plus rapidement que les transversions dans la divergence et vont donc conduire également à un effet plateau lorsque l'on compare des espèces génétiquement trop éloignées.

II.2.5. Définition des standards de notation des noms d'espèce

Dans notre étude, nous utiliserons le terme de "clade *thermophilus*" pour parler du clade monophylétique qui regroupe l'espèce *B. thermophilus* et les espèces de *Bathymodiolus* présentant au moins le symbiote sulfo-oxydant. De même, le clade paraphylétique auquel appartient *Bathymodiolus childressi* (Jones & Vrijenhoek 2006) et dans lequel se trouve aussi le genre *Gigantidas* sera désigné dans la suite de l'étude comme le "clade *childressi*". Enfin, pour les espèces de Bathymodiolinae non décrites mais dont les séquences forment des clades distincts, le nom '*Bathymodiolus*' sera placé entre apostrophes et suivi de la provenance géographique, par exemple '*Bathymodiolus*' *sp.*_Costa Rica.

II.2.6. Analyse sous PAML

Un arbre NJ effectué sous PHYLIP avec la distance F84 (Felsenstein 1984) a été utilisé comme arbre d'ajustement pour des modèles de substitution de codon pour le logiciel PAML 3.14 (Yang 2000). Un alignement sans indel de 124 codons, sur 54 taxons représentant l'intégralité des Bathymodiolinae et Modiolinae des environnements réduits profonds a été utilisé pour cette analyse.

Le modèle de substitution de codon est celui dérivé de Goldman & Yang (1994). Trois simulations ont été effectuées en utilisant le code génétique mitochondrial pour les invertébrés : (1) modèle supposant un dN/dS unique pour toutes les branches et pour tous les codons (dit modèle 0) avec fixation d'une horloge moléculaire et un module permettant d'effectuer une reconstruction des séquences ancestrales aux différents nœuds internes de l'arbre testé, (2) modèle 1 (dit presque neutre) supposant un dN/dS unique pour toutes les branches mais catégorisant deux classes de mutations au niveau des codons, l'une pour les mutations délétères (dN/dS = 0), l'autre pour les mutations neutres (dN/dS = 1), et, (3) un modèle dit sous sélection positive supposant un dN/dS unique pour toutes les branches mais catégorisant trois classes de mutations au niveau des codons (délétère : dN/dS = 0, neutre : dN/dS = 1 et sélection positive : dN/dS > 1).

II.3. Résultats

Au total, 208 séquences ont été obtenues dans les banques de gènes pour le locus mitochondrial COI. Cela représente l'ensemble des séquences disponibles pour les espèces vivant dans des environnements réduits et dont les genres appartiennent aux sous familles des *Bathymodiolinae* et *Modiolinae* (c.f. annexe 3 pour consulter la liste détaillée des séquences avec leur N° d'accession). A ces données ont été ajoutées 43 séquences COI supplémentaires (cf Tableau II.2).

Pour le marqueur intergénique ITS2, 45 séquences proviennent de l'article de Olu-Le Roy et al. (2007) auxquelles s'ajoutent 104 séquences supplémentaires (cf. Tableau II.2), soit un total de 149 séquences.

Ces séquences supplémentaires ont un double intérêt: (i) augmenter le nombre de séquences dans des familles peu ou pas représentées dans les banques (peu de séquences étaient disponibles pour l'espèce *B. puteoserpentis* et aucune séquence provenant des zones de suintements froids des marges du Pacifique Est n'était disponible), et (ii) la qualité, l'identité (nom d'espèce), mais surtout la provenance exacte de ces séquences sont connues avec exactitude. Dans une étude de phylogénie, il ne s'agit pas uniquement d'avoir des séquences, mais il importe également d'en connaître la provenance. L'acquisition de séquences nouvelles a permis de compléter certains renseignements, manquants ou trop succincts, indiqués pour les séquences déposées dans les banques. Ainsi, il a été possible de

déterminer que les individus *B. aff. thermophilus* de la dorsale Pacifique-Antarctique, sans indication de coordonnées géographiques précises étaient identiques aux nôtres provenant de la localité 37°S sur l'EPR.

Tableau II.2. Nombre de séquences supplémentaires obtenues au cours de cette thèse et ajoutées aux séquences issues des banques pour le gène mitochondrial COI et l'espaceur intergénique nucléaire ITS2.

Localité	nombre de séquences par locus	
	COI	ITS2
Océan Pacifique		
<i>B. thermophilus</i> _9°50N	4	8
<i>B. thermophilus</i> _7°S	8	9
<i>B. thermophilus</i> _21°S	5	16
<i>B. aff. thermophilus</i> _37°S	2	16
<i>B. sp.</i> _ Costa Rica	2	2
Océan Atlantique		
<i>B. azoricus</i> _Rainbow	6	15
<i>B. azoricus</i> _Menez Gwen	10	14
<i>B. puteoserpentis</i>	6	24

Effet de la saturation sur les marqueurs

L'analyse de la divergence mitochondriale utilise les espèces de *Bathymodiolus* du clade « *thermophilus* », du clade « *childressi* » et une d'une espèce de Juan de Fuca proche du genre *Benthomodiolus* (Modiolinae). Quelque soit le niveau de divergence observé entre deux espèces, il ne présente pas d'effet de saturation (Figure II.1A). Le marqueur COI présente donc un signal phylogénétique fiable sur l'ensemble du groupe des modioles profondes. Ce résultat contredit les travaux de Miyazaki et al. (2004) qui concluaient à une saturation de la 3^e paire de base des codons à partir d'une divergence de 20%.

A l'inverse, les données obtenues avec l'ITS2, montrent un net effet de saturation dès lors que l'on compare les groupes des Bathymodiolinae et des Modiolinae (Figure II.1B). Les distances calculées entre l'espèce *Tamu fisheri* (Modiolinae) et les autres espèces du groupe des Bathymodiolinae s'écartent très nettement de la droite de régression estimée pour les autres points. Le nombre de transitions est par contre parfaitement corrélé à la distance K2P pour les comparaisons intra-clades et inter-clades « *thermophilus* » et « *childressi* ». Le marqueur ITS2 est donc parfaitement adapté à l'étude phylogénétique de la sous-famille Bathymodiolinae mais conduira à des biais de regroupements au niveau des comparaisons

Modiolinae/Bathymodiolinae, par homoplasie ou attraction des longues branches dans les arbres phylogénétiques.

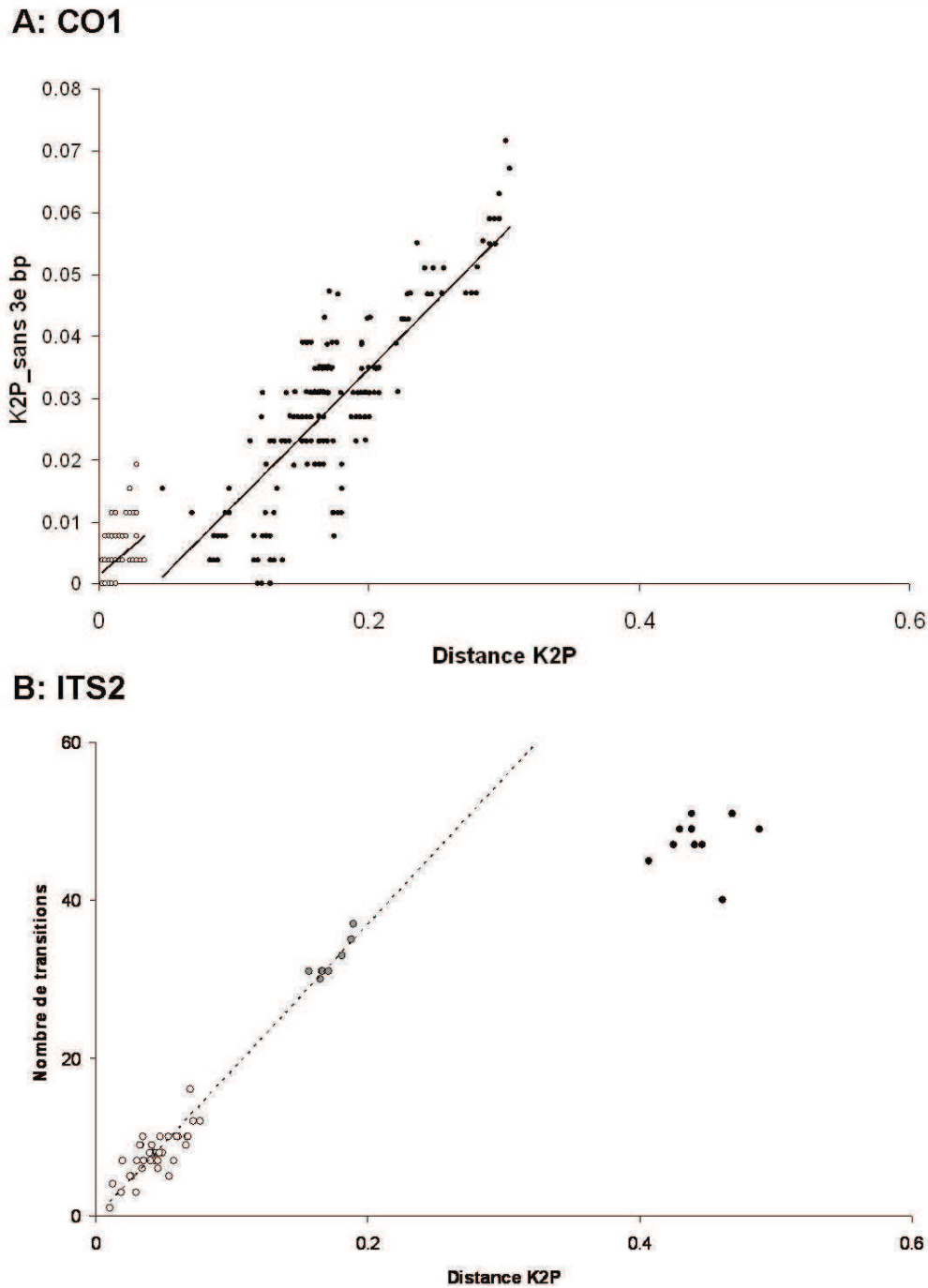


Figure II.1. Graphes présentant l'effet de saturation sur les deux marqueurs moléculaires utilisés.

A. COI : cercles vides représentent les distances calculées entre individus de la même espèce ; les cercles pleins représentent les distances calculées pour chacune des paires d'espèces.

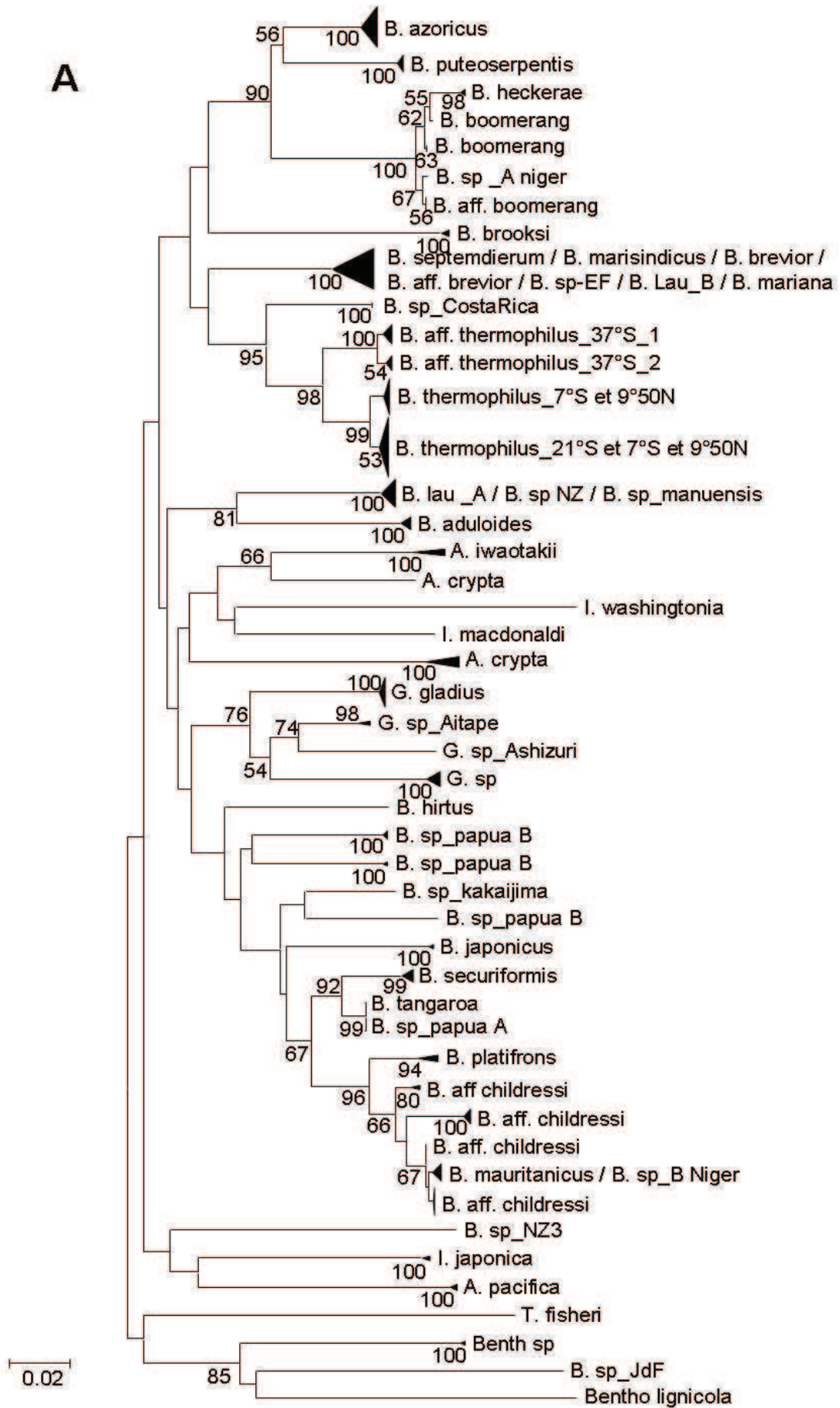
B. ITS2 : cercles vides représentent les distances calculées entre les espèces *Bathymodiolus* du clade « *thermophilus* » ; les cercles grisés représentent les distances entre les clades « *thermophilus* » et « *childressi* », cercles noirs correspondent aux distances calculées entre les espèces Bathymodiolinae et les Modiolinae.

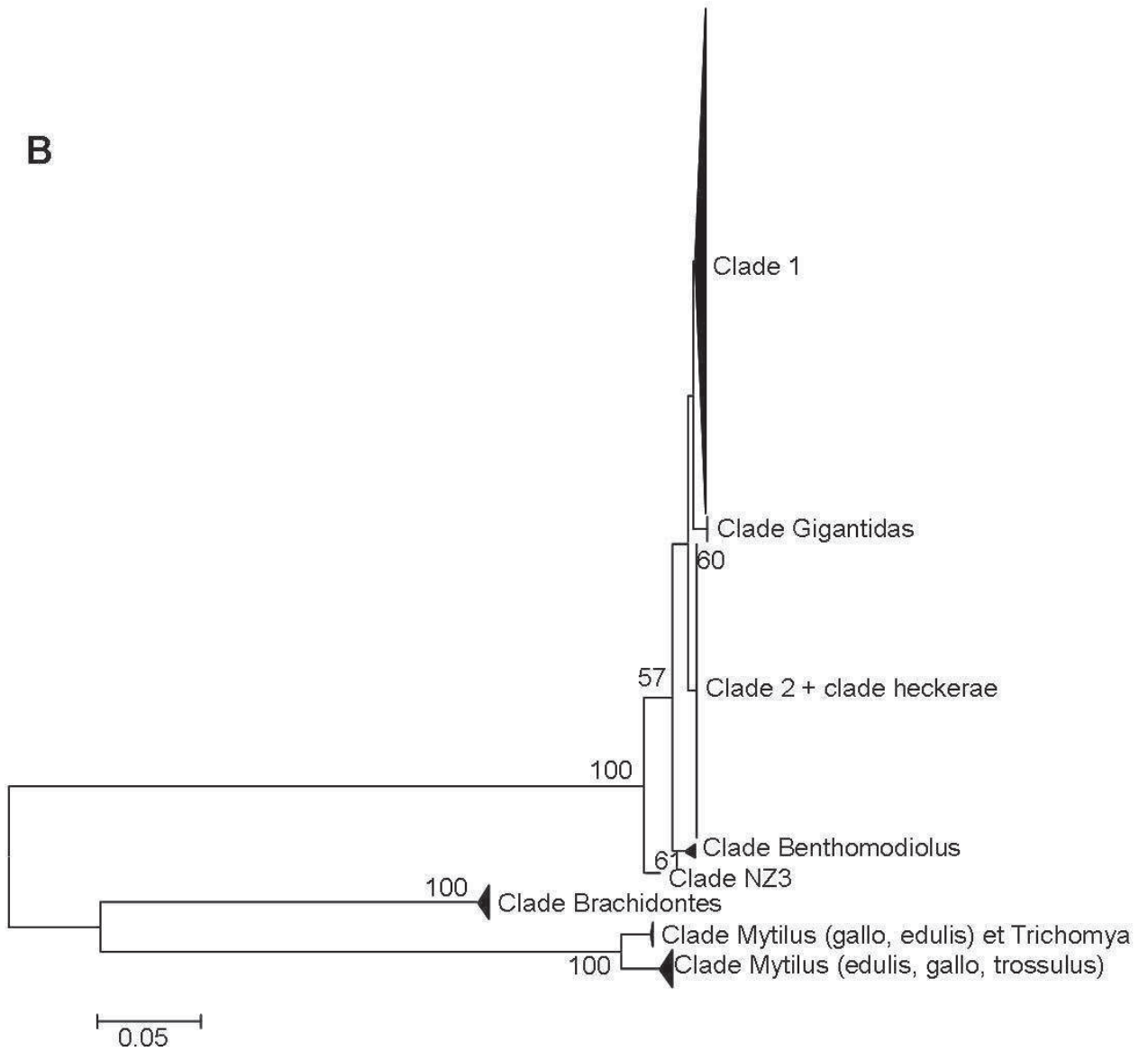
II.3.1. Reconstructions phylogénétiques

L'analyse a été menée sur l'ensemble des Mytilidae des environnements réduits pour le marqueur COI et seulement au niveau des Bathymodiolinae pour le marqueur ITS2. En général, les valeurs de bootstraps sont élevées et attestent de la grande solidité des regroupements effectués par la méthode du Neighbor-Joining.

II.3.2. Phylogénie avec le marqueur COI

Dans l'arbre de COI réalisé avec les séquences nucléotidiques, les 41 taxons géographiques utilisés se répartissent en 30 lignées phylogénétiques bien définies (Figure II.2 A). Ces lignées se regroupent distinctement en un clade '*thermophilus*' et un clade '*childressi*'. Au sein du clade '*thermophilus*', on peut distinguer un groupe « est-pacifique » formé de l'espèce *B. thermophilus* de la dorsale EPR, de celle trouvée sur la dorsale Pacifique-Antarctique à 37°S, et de celle récoltée au niveau de la marge active du Costa Rica. De la même façon, on peut distinguer un groupe « atlantique » formé par les espèces de la Ride Médio-Atlantique, celles du Golfe de Guinée et du Golfe du Mexique (à l'exception de *Tamu fisheri* et des espèces affiliées à *B. childressi*). Un troisième groupe constitué des nombreuses espèces du Pacifique Ouest et de l'espèce de la dorsale indienne (*B. marisindicus*) occupe une place intermédiaire entre les groupes 'Atlantique' et 'Pacifique Est' au sein de ce clade. Le second clade, dénommé '*childressi*' regroupe quant à lui trois groupes distincts : un groupe constitué d'espèces à symbiotes sulfo-oxydant (*B. aduloides*), un groupe constitué du genre *Gigantidas* (symbiote sulfo-oxydant) et enfin un groupe constitué d'espèces méthanotrophes strictes (*B. japonicus*, *B. childressi*, *B. platifrons*). Il convient de noter que la base du clade est uniquement constituée d'un ensemble d'espèces originaires du Pacifique Ouest (*B. manuensis* et *B. platifrons*), et que les espèces atlantiques telle que l'espèce «*Bathymodiolus*» *childressi*, endémique du Golfe du Mexique, ou *Bathymodiolus mauritanicus* (Golfe de Guinée) apparaissent comme étant dérivées. Les divergences observées entre taxons au sein de ce clade sont inférieures ou proches de celles observées entre les deux espèces de la dorsale médio-atlantique (*B. azoricus* et *B. puteoserpentis*). Cette absence de différences fortes entre distances d'espèces prises deux à deux indépendamment de leur éloignement géographique suggère une radiation relativement récente de ce groupe.





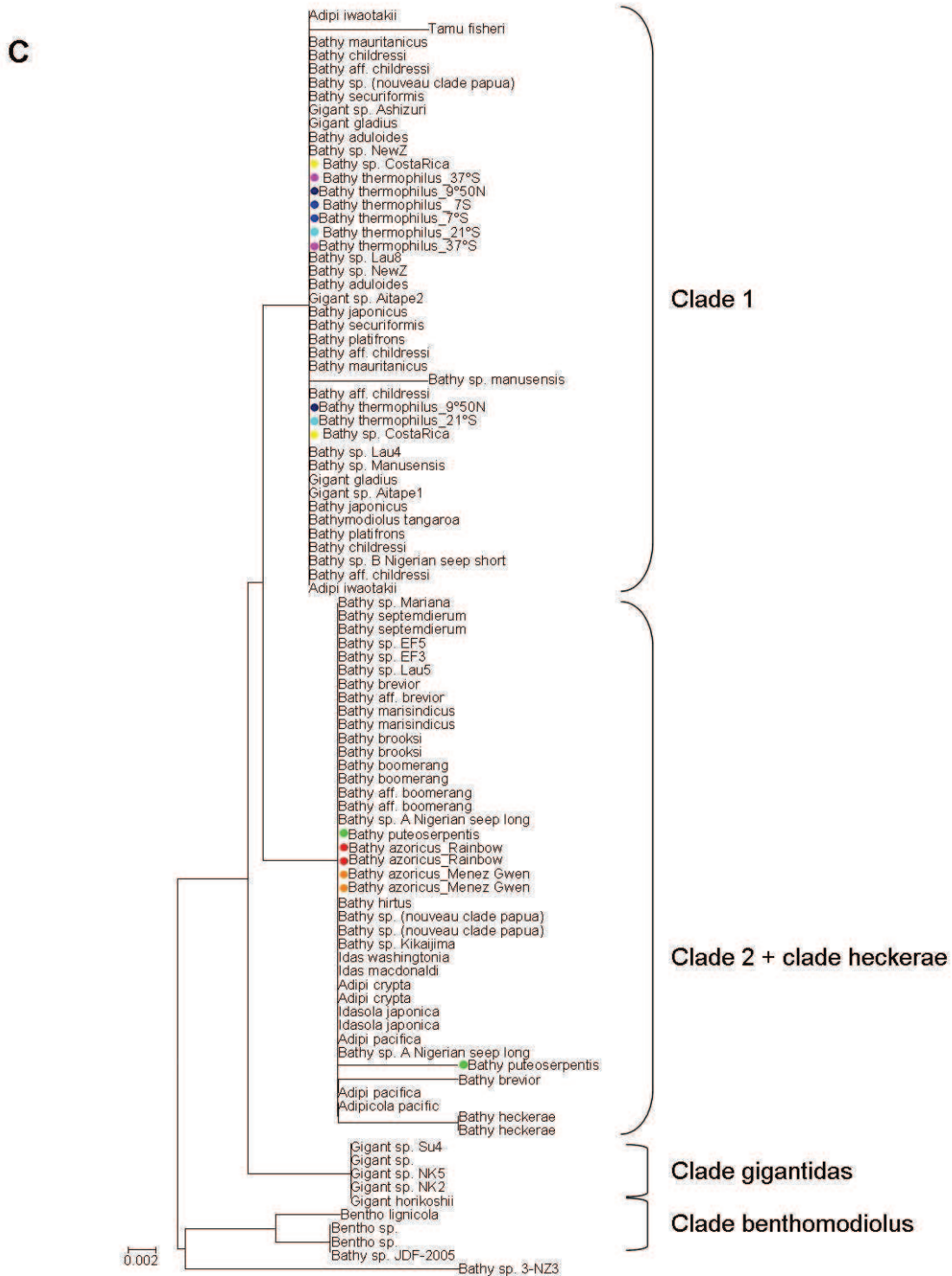


Figure II.2. Phylogénies réalisées à l’aide du marqueur COI. (A) Arbre NJ obtenu sur les trois bases de codon des séquences nucléotidiques (376 bp) à l’aide de la distance Tamura/Nei avec un rééchantillonnage de 1000 bootstraps ; (B) Arbre NJ obtenu en acides aminés à l’aide de la matrice de Jones-Taylor-Thornton (JTT) avec un rééchantillonnage de 1000 bootstraps. (C) Arbre NJ obtenu sur les séquences primaires du COI des Bathymodiolinae avec une sélection préalable de deux séquences par espèce et par clade. Les étiquettes colorées correspondent aux espèces qui sont utilisées dans cette thèse.

L'arbre obtenu à partir de l'ensemble des séquences traduites (Bathymodiolinae et Mytilidae) n'apporte pas d'information supplémentaire sur la résolution de la phylogénie profonde. En revanche, il met en évidence l'extrême conservation de la structure primaire de la protéine : la totalité des espèces des environnements profonds se répartissent en effet sur quatre haplotypes majoritaires, dont deux seulement apparaissent diagnostiques d'un genre (*Benthomodiolus* et *Gigantidas*). En effet, notre étude porte sur la totalité des espèces peuplant les environnements profonds et les séquences proviennent de 41 espèces ayant accumulé déjà un grand nombre de mutations synonymes (Tableau II.1). Pourtant lorsque la séquence est traduite en protéine, il ne subsiste que quatre grands clades : celui des *Benthomodiolus*, celui des *Gigantidas*, et deux autres grands clades regroupant l'ensemble des autres genres, (*Idas*, *Idasola*, *Adipicola*, et toutes les *Bathymodiolus* (Figure II.2 B). Les *Bathymodiolus heckerae* forment un sous ensemble distinct à l'intérieur du clade 2 et la ' bathymodiole' _sp_NZ des sources hydrothermales de Nouvelle Zélande forme une branche divergente à la base de l'arbre. On peut également noter que la divergence protéique séparant ces quatre clades reste inférieure à celle rencontrée entre espèces proches dans les familles côtières proches (par exemple au sein des espèces *Mytilus galloprovincialis* et *Mytilus edulis*) Figure II.2 B. Les remplacements en acides aminés ont été cartographiés et replacés dans l'arbre nucléotidique (Figure II.3). Une analyse avec PAML a été réalisée afin de connaître les mutations ancestrales pour chacun des nœuds. Les probabilités postérieures bayésiennes (BEB) associées à chaque mutation ancestrale ont été estimées et sont toutes comprises entre 0,95 et 1 (valeurs non indiquées dans l'arbre).

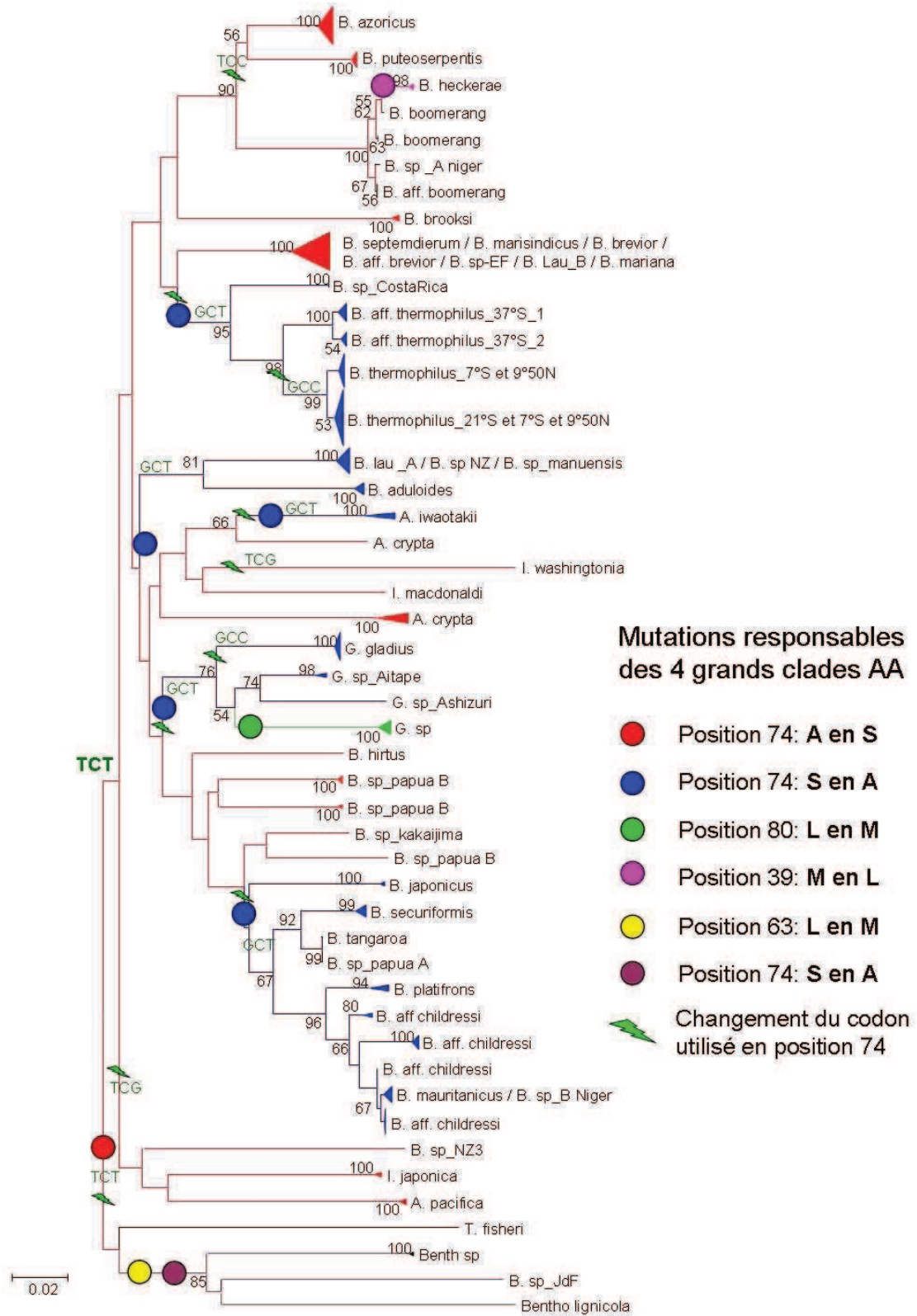


Figure II.3. Répartition et évolution des remplacements en Acides Aminés. Les mutations indiquées sont fixées et correspondent aux clades de la (Figure II.2). En **Bleu** : mutations et lignées qui correspondent aux espèces du clade 1 ; en **rouge** : Clade 2 (avec en **violet** : clade heckerae); en **vert** : Clade Gigantidas ; en **prune** : le clade Benthomodiolus. Le codon utilisé en position 74 (S ou A) est indiqué sur l'arbre. La cartographie des mutations ancestrales a été déterminée avec PAML à partir des séquences nucléotidiques du CO1.

II.3.3. Phylogénie avec le marqueur ITS2

A première vue, l'arbre obtenu à l'aide des séquences nucléotidiques du marqueur ITS2 présente une topologie similaire à celle observée pour le marqueur mitochondrial. Les regroupements sont cependant moins robustes avec des valeurs de bootstraps plus faibles. (Figure II.4).

L'arbre se divise en quatre grands clades: un clade 'Atlantique' (*B. azoricus*, *B. puteoserpentis*, *B. boomerang*, 'B.'_Zaire); un clade 'Pacifique Est' (toutes les espèces *B. thermophilus* et affiliées (37°S), plus 'B.'_CostaRica); un clade mal défini regroupant les espèces du Pacifique Ouest (*B. brevior* et *B. elongatus*), et enfin un clade regroupant les espèces du groupe « *childressi* » ('B.'_Barbades et *B. childressi*). *B. brooksi* (Golfe du Mexique) ne forme pas de clade mais se regroupent en polytomie à la base de l'arbre. *B. heckerae* associée à *B. boomerang* et 'B'_Zaire (espèces de suintements froids) constituent un groupe frère des espèces hydrothermales de la ride médio-atlantique.

Ce marqueur semble plus résolutif que le COI pour distinguer les espèces de la dorsale Est-Pacifique. Même si l'on n'observe pas de divergence profonde entre les espèces de 37°S et celles plus au nord, il est cependant possible de distinguer trois « clades » certes mal définis, mais permettant de distinguer les populations les plus au nord (9°50N et 7S) de celles de 21°S et de 37°S. Cette distinction en trois clades n'avait encore jamais été observée à l'aide des marqueurs allozymiques et mitochondriaux. Les nouvelles séquences obtenues chez la bathymodioline du Costa Rica sont clairement intermédiaires entre le clade 'Pacifique Est' et le clade 'Atlantique'.

Il est également intéressant de remarquer que l'espèce *B. puteoserpentis* est la seule à présenter un grand nombre de lignées divergentes au sein du groupe « Atlantique » suggérant une polyphylie au sein de ce taxon.

Enfin, quelques séquences, pourtant de bonne qualité sur les chromatogrammes, apparaissent mal classées et pourraient constituer soit des lignées ancestrales non éliminées, soit la possibilité d'une évolution non concertée entre certaines copies de l'ADN ribosomique. Ainsi, une séquence de *B. puteoserpentis* se retrouve à la base du clade « Pacifique » et une séquence issue de *B. azoricus* se retrouve au niveau du clade « *childressi* ».

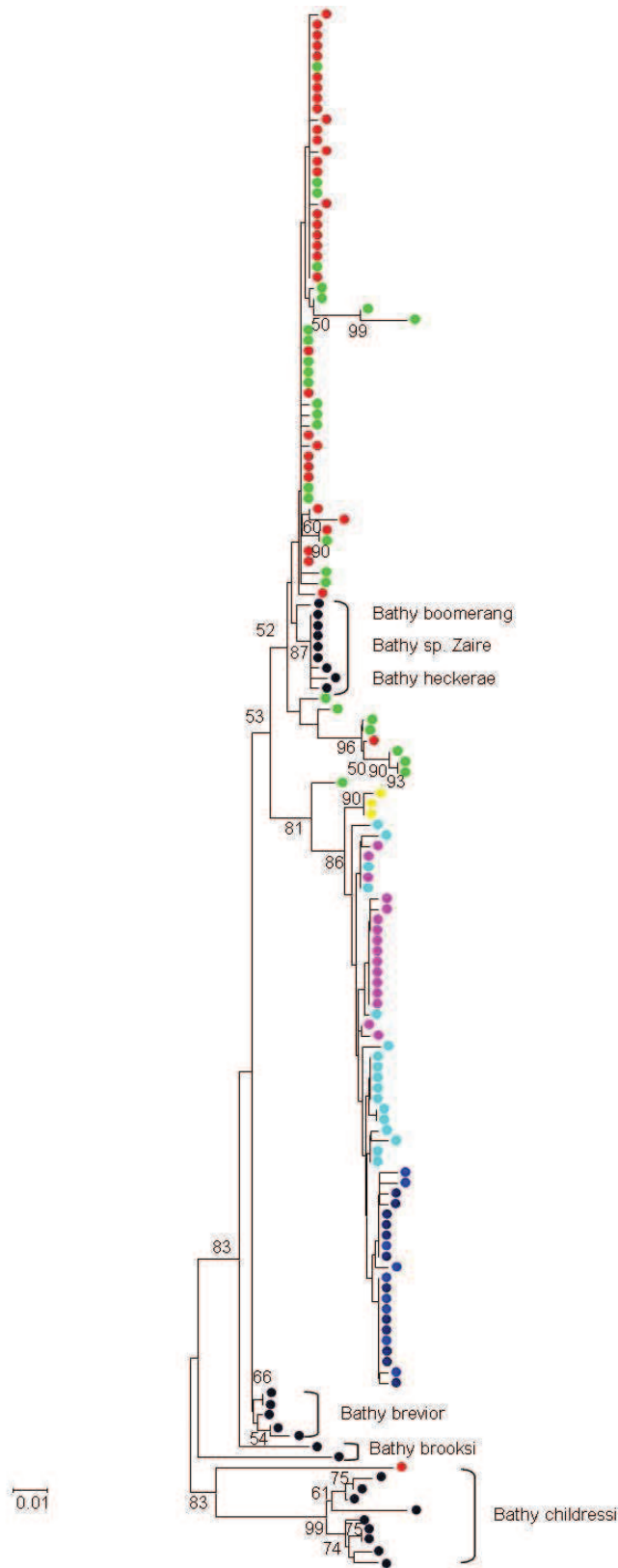


Figure II.4. Arbre NJ réalisé avec le marqueur ITS2 (460bp) à l'aide de la distance K2P et avec un ré-échantillonnage des données de 1000 bootstraps. Bleu : *Bathymodiolus thermophilus* (dorsale Est Pacifique) avec des variations en fonction de la latitude (plus foncé au nord, plus clair au sud) ; Violet : *Bathymodiolus aff. thermophilus* de 37°S ; Rouge : *B. azoricus* ; Vert : *B. puteoserpentis* ; Jaune : *B. sp. Costa Rica*.

II.3.4. Analyse bayésienne et maximum de vraisemblance sur les données du CO1.

La longueur des arbres et les valeurs de vraisemblance d'ajustement entre les données et l'arbre de CO1 ont été calculées pour chacun des modèles et sont indiquées dans le Tableau II.3.

Le premier modèle a été utilisé pour réaliser un arbre pour lequel les distances sont ajustées au modèle de Goldman & Yang (1994) (Figure II.5). Cet arbre illustre une radiation explosive et récente chez les Bathymodiolinae. Les mutations ancestrales ont également été cartographiées avec ce modèle et ont été ajoutée dans l'arbre phylogénétique (Figure II.3). Les modèles neutres et avec sélection ont été utilisés pour tester les pressions de sélection s'appliquant sur chacun des codons. On peut constater que quel que soit le modèle utilisé, il présente un fort biais dans le Kappa (κ) (excès de transitions).

L'étude des pressions de sélection a été réalisée en comparant deux modèles dont les données sont consignées dans le Tableau II.4. Avec chacun des deux modèles, 100% des sites (proportion =1) sont classés dans la classe de $d_N/d_S = 0$. C'est-à-dire que la totalité du gène est sous forte sélection purifiante (toutes les mutations sont fortement délétères). Aucun codon n'est présent dans la classe de la neutralité ou de la sélection positive. L'estimateur (ω) est très faible et identique pour les deux modèles comparés.

Tableau II.3. Descriptif des paramètres et valeurs de Log likelihood pour le CO1. **L** : longueur de l'arbre estimée en nombre de substitutions nucléotidiques par codon ; **lnL** : probabilité d'ajustement des données à l'arbre sous le modèle choisi ; **κ** : Ratio entre les taux de (mutation) transition / transversion ; **ω** : Ratio entre les taux de mutation synonymes et non-synonymes (= d_N/d_S)

gène	nb codon	modèle	L	lnL	κ	ω
CO1	124	Horloge moléculaire	103, 24	-4059,08	4,40	0,0023
		Neutre et sélection positive	66,13	-3994,87	3,96	0,0037

Tableau II.4. Estimation des paramètres sous les différents modèles pour le gène CO1

Modèle		$d_N/d_S = 0$	$d_N/d_S = 1$	$d_N/d_S < 1$
neutre	Proportion	1	0	-
	ω	0,0037	1	-
sélection positive	Proportion	1	0	0
	ω	0,0037	1	1

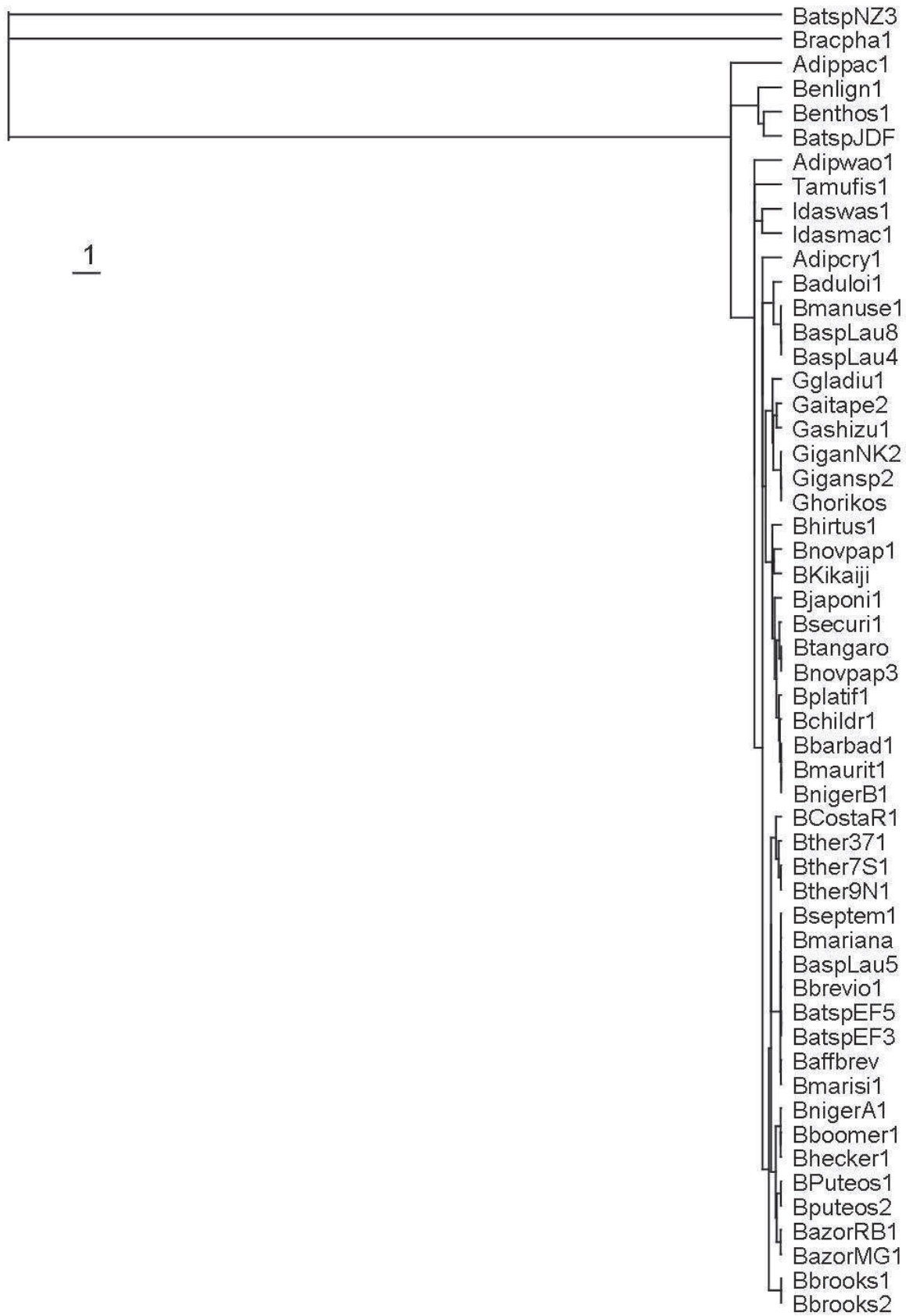


Figure II.5. Arbre ajusté sur les valeurs de distances calculées par PAML selon le modèle de Goldman & Yang (1994).

II.4. Discussion

II.4.1. Diversité génétique chez les Bathymodiolinae : relations avec l'habitat (profondeur, symbiose et milieux)

Les premières descriptions d'espèces de Bathymodiolinae datent de 1985 (Kenk & Wilson 1985: description de *B. thermophilus*) et étaient uniquement basées sur des caractéristiques morphologiques de la coquille ou l'emplacement des muscles adducteurs (Von Cosel et al. 1994, Von Cosel & Olu 1998). Ces études ont permis de décrire la plupart des espèces hydrothermales et de définir plus particulièrement les différents genres connus à l'heure actuelle (*Bathymodiolus*, *Tamu*, *Gigantidas* : cf Gustafson et al. (1998), Von Cosel & Marshall (2003)) par rapport à ceux déjà connus dans le milieu profond (*Idasola* Iredale 1915 ; *Adipicola* Dautzenberg 1927 ; Dell, 1987). Pourtant, de nombreux problèmes d'identification vont alors se poser en raison de la très forte plasticité phénotypique de la coquille (Olu-Le Roy et al. 2007). Comment savoir alors si, malgré toutes les précautions d'usage, les espèces décrites uniquement d'un point de vue morphologique ne sont pas constituées d'un ensemble d'espèces ? Certaines espèces décrites ainsi que de nombreux isolats géographiques ont alors fait l'objet d'une identification moléculaire systématique (*barcoding*) basée sur le gène mitochondrial codant pour la cytochrome oxydase I qui a révélé l'existence d'une forte diversité génétique au sein de ce groupe (Won et al. 2003b, Miyazaki et al. 2004, Jones & Vrijenhoek 2006). En effet, un grand nombre d'isolats géographiques, semblables d'un point de vue morphologique, se sont avérés distincts d'un point de vue génétique. De même, certaines populations de modioles récoltées en un point précis du globe se sont avérées être en fait un mélange d'espèces difficilement reconnaissables du point de vue de leur coquille, notamment au niveau du Pacifique Ouest où la diversité des modioles est la plus forte. Ainsi, certains taxons regroupés sous un même nom « régional » sont en fait constitués d'espèces génétiquement différenciées ('*Bathymodiolus*'_spI, spII, spIII_dans le bassin de Lau ou celui de Manus (Miyazaki et al. 2004). La présence d'espèces cryptiques, phénomène décrit comme étant commun dans le milieu marin (Knowlton 1993, Knowlton 2000), apparaît comme extrêmement courant chez les invertébrés des environnements hydrothermaux, notamment chez les crustacés (Guinot & Hurtado 2003), les polychètes (Hurtado et al. 2004), les Vesicomydés (Vrijenhoek et al. 1994), les Lepetodrilidae (Johnson et al. 2008) et les Bathymodiolinae (Won et al. 2003b). Dans la plupart des cas, ce type de spéciation semble découler de processus tectoniques ayant séparé les faunes lors de la subduction et/ou la fossilisation de certaines dorsales (Tunnicliffe & Fowler 1996, Johnson et

al. 2006) ou la mise en place de points triple ou de microplaque (Hurtado et al. 2004, Won et al. 2003b, Matabos 2007). A l'inverse, l'identification moléculaire des Bathymodiolinae a également permis de montrer que certains isolats géographiques pouvaient être vraisemblablement mis en synonymie. Il semble en effet, si l'on se réfère aux noms d'espèces trouvés dans la littérature (indiqués dans le Tableau II.1), qu'il existe un manque de concertation dans la ré-appellation périodique d'espèces, celle-ci ayant conduit à la situation où une même espèce « génétique » peut posséder plusieurs noms synonymes (*B. septendierum*, *B. brevior*, *B. marisindicus*) selon sa provenance géographique. Cette difficulté à définir clairement les contours géographiques d'une espèce est particulièrement vraie au sein des complexes d'espèces *B. boomerang* et *B. childressi* au niveau de l'Atlantique (Olu-LeRoy et al. 2007).

La phylogénie décrite dans cette thèse montre l'ampleur de cette diversité génétique révélée en utilisant de nombreuses séquences provenant d'isolats géographiques encore non décrits. La plupart de ces séquences apparaissent clairement divergentes et souvent mal classées au sein d'un genre, donnant lieu à une polyphylie du genre *Bathymodiolus*. Les études précédentes dénombraient une vingtaine d'espèces de Bathymodiolinae. Si l'on exclut les espèces appartenant aux Modiolinae ainsi que les espèces attribuées de façon erronée au genre '*Bathymodiolus*' mais qui branchent à l'extérieur du groupe (par ex. '*Bathy*'_sp NZ, '*Bathy*'_sp Juan de Fuca) et, que l'on se base sur les divergences généralement observées entre les espèces décrites proches, le nombre d'espèces de Bathymodiolinae atteindrait une valeur de 30 espèces génétiquement différentes. Ce décompte est largement supérieur à ceux effectués par Iwasaki et al. (2006) : 19 espèces décrites, ou par Won et al. (2008) : 20 espèces décrites. La faible diversité des habitats réduits (sources hydrothermales, suintements, carcasses, bois coulés), l'extrême pauvreté des caractéristiques morphologiques diagnostiques et le grand nombre d'espèces cryptiques trouvées mais non décrites pose la question des mécanismes évolutifs ayant mis en place une telle diversité et de la pertinence d'utiliser le seul gène COI comme critère d'identification. De nombreux auteurs ont tenté de mieux comprendre ces mécanismes et, notamment quel pouvait être le rôle de l'endosymbiose, la profondeur des sites et la nature des émissions réduites sur la spéciation des bathymodioles.

II.4.2. Causes probables à la diversité des Mytilidae des environnements réduits

Les premières études phylogénétiques concernant l'environnement et les Bathymodiolinae mettent en parallèle l'évolution des espèces et la préférence d'habitats en cherchant à savoir quel était l'habitat ancestral de ces lignées particulières. L'hypothèse principale est celle selon laquelle les anciennes lignées des environnements réduits auraient occupé dans un premier temps les zones de suintements froids dans les environnements côtiers (McLean 1985, Craddock et al. 1995). Cette hypothèse est basée sur le fait que l'océan profond a connu des phases d'anoxie complète au cours des périodes du Cénomaniens/Turonien et de la fin du Paléocène ayant entraîné la quasi-extinction de la faune abyssale (Jackson 1986). Ces périodes d'extinctions massives seraient à la base de la radiation récente et explosive des Bathymodiolinae à partir de quelques lignées de Bathymodioliformes survivantes dans les eaux de surface (Jones et al. 2006).

La colonisation des environnements profonds (sources hydrothermales, suintements) aurait alors été permise par le biais de l'utilisation des bois coulés ou des carcasses de baleine (Distel et al. 2000). Il convient néanmoins de noter que certaines lignées ancestrales profondes auraient pu survivre localement, de nombreux travaux en paléontologie faisant état de faunes fossiles dominées par des bivalves bathymodioliformes dans des ophiolites du Dévonien ou du Crétacé inférieur (Haymon et al. 1984, Banks 1985, Little & Cann 1999). Ce type d'évolution devrait conduire à l'observation de topologies présentant une hiérarchie selon laquelle les espèces de suintements froids seraient basales, les espèces des bois coulés ou colonisant les os de baleine seraient intermédiaires et les espèces des sources hydrothermales apparaîtraient comme des espèces dérivées.

Cette hypothèse n'apparaît que peu crédible eu égard aux phylogénies présentées dans ce chapitre et confirment les tests d'hypothèses effectués par Jones et al. (2006), celui-ci concluant à l'équiprobabilité des différents milieux en termes de berceau des bathymodioles. Certes, les espèces des bois coulés appartenant à la sous-famille des Modiolinae ont une position généralement basale et présentent souvent des niveaux de divergence beaucoup plus élevés que ceux trouvés chez les Bathymodiolinae. Cependant, l'espèce considérée comme étant la plus proche de la racine est l'espèce hydrothermale NZ3 récoltée en Nouvelle Zélande (Samadi et al. 2007). En revanche, si l'on considère la sous famille des Bathymodiolinae, les espèces des sources hydrothermales apparaissent comme étant dérivées dans les groupes

« Atlantique » et « Pacifique » appartenant au clade « *thermophilus* » (exclusion des « *Bathymodiolus* » du clade '*childressi*'). Jones et al. (2006) qui ne possédaient des séquences que pour quelques espèces de zones de suintements froids « Atlantiques » (*B. heckerae*), avaient décrit cette particularité comme une réversion d'habitat locale (colonisation d'une zone de suintements par les espèces hydrothermales de la dorsale médio-Atlantique) (Figure II.6 b). L'analyse des séquences provenant des bathymodiales des zones de suintements froids du Pacifique Est ('B.'_CostaRica), du Golfe du Mexique (*B. brooksi*, *B. heckerae*), des Barbades (*B. boomerang*) et du Golfe de Guinée ('B'. sp. Zaire) montre clairement que la diversification du clade s'est effectuée à partir des zones de suintements, possiblement le Golfe du Mexique avant la fermeture de l'Isthme de Panama. La situation apparaît plus complexe au sein du clade « *childressi* » pour lequel la méthanotrophie est le mode dominant d'acquisition de l'énergie chez les symbiotes. Dans le Pacifique Ouest, on rencontre en effet souvent les mêmes espèces dans les deux types d'environnement : *B. aduloides*, *B. japonicus*, *B. platifrons* (Miyazaki et al. 2004).

La profondeur semble également un élément déterminant pour expliquer les relations phylogénétiques chez les Bathymodiolinae. Il semble que la distribution d'un certain nombre d'espèces soit effectivement fonction de la profondeur et pourrait expliquer la partition des bathymodiolinae dans les 2 principaux clades. Ainsi le clade « *thermophilus* » est principalement observé à des profondeurs dépassant 2000 mètres tandis que les espèces du clade « *childressi* » affectionnent plus volontiers de plus faibles profondeurs (Miyazaki et al. 2004, Jones et al. 2006) (Figure II.6 a). Cependant, les exceptions sont nombreuses et la profondeur ne peut être considérée comme un critère fiable. La répartition des espèces en fonction de la profondeur est souvent considérée de manière indépendante dans les études phylogénétiques précédemment réalisées, pourtant ce paramètre est étroitement lié à l'environnement. En effet les zones de suintements froids sont souvent situées en deçà (Sagami bay, Louisiane) ou au delà (Escarpe de Floride, Kaiko-Nankai, Golfe de Guinée) de celles des sources hydrothermales (2000-3000 mètres). Ainsi les Bathymodiales Atlantiques du groupe *B. azoricus* semblent faire exception à leur groupe en vivant à faible profondeur (850m). En fait, cette espèce est la seule à avoir colonisé des sources hydrothermales exceptionnellement peu profondes. Dans les eaux japonaises, en revanche, il est plus difficile de distinguer un tel patron. Les sources hydrothermales comme les zones de suintements froids se répartissent sur un large panel de profondeurs et les espèces qui y vivent sont capables de peupler ces deux environnements (Miyazaki et al. 2004, Iwasaki et al. 2006).

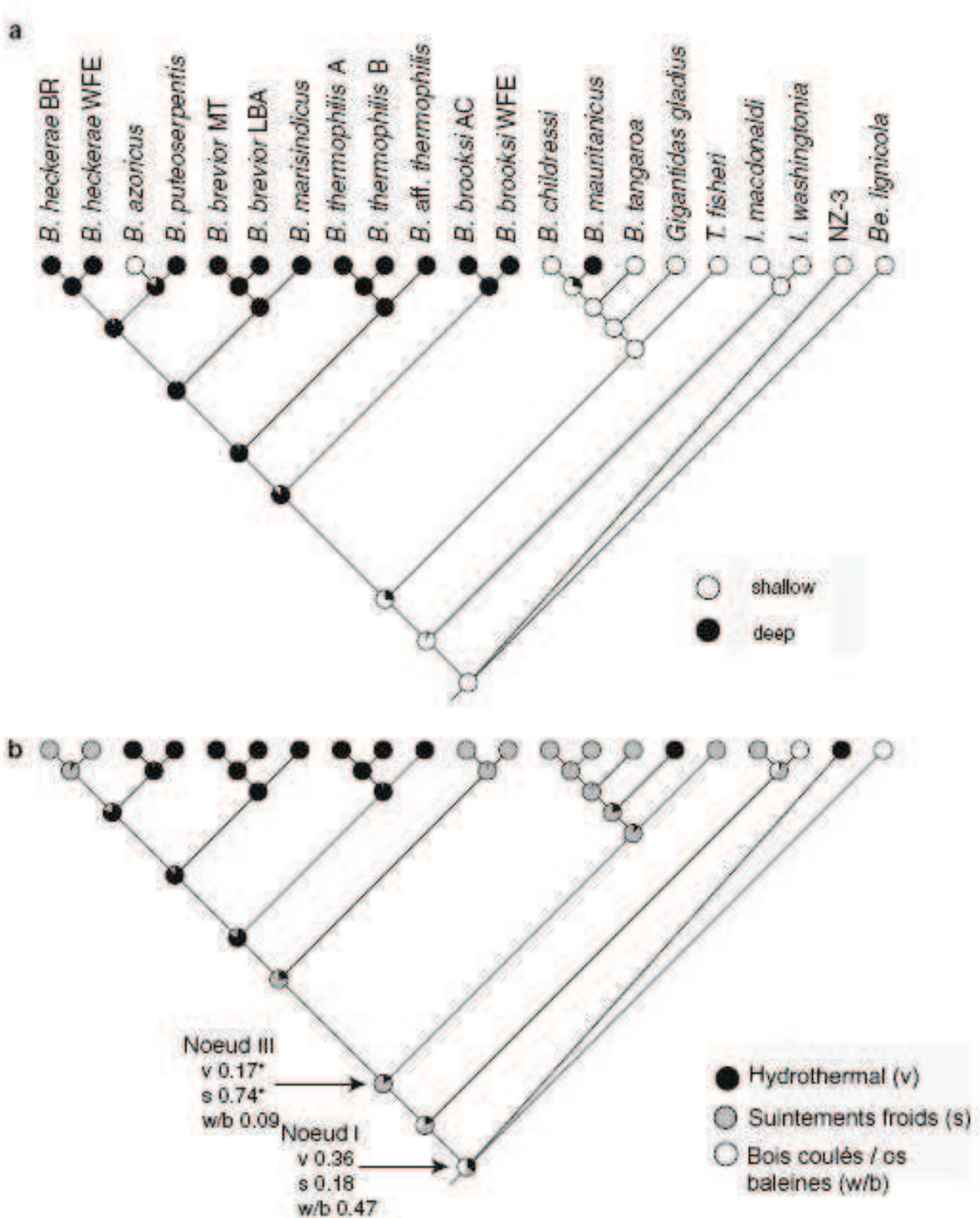


Figure II.6. Cladogrammes bayésien basés sur 3 gènes CO1, ND4, 28S avec des estimateurs en maximum de vraisemblance pour estimer le caractère ancestral de (a) la profondeur et de (b) l’habitat. La robustesse du caractère ancestral est indiquée pour chacun des nœuds avec un camembert. Les valeurs pour les 3 types d’habitats (v : hydrothermal ; S : suintement froids ; w/b bois coulés et carcasses de baleines) sont données pour les nœuds 1 et 3. Illustration modifiée de (Jones et al. 2006)

Une autre piste à la diversification des Bathymodiolinae est la diversité des endosymbioses observées au sein des tissus branchiaux. Toutes les Bathymodiolinae sont connues pour posséder des bactéries épi- ou endo-symbiotiques de 2 types : les méthanotrophes et les sulfo-oxydantes (Childress et al. 1986, Fisher et al. 1987). Dans les environnements réduits, certains bivalves possèdent une transmission verticale de leur symbiontes (Krueger et al. 1996, Peek et al. 1998) induisant une parfaite co-spéciation entre l'hôte et le symbionte (Peek et al. 1998). Cependant il a été démontré que l'acquisition des bactéries se faisait depuis l'environnement (transmission horizontale) chez les bathymodiales (Won et al. 2003c), ceci entraînant une absence de co-spéciation entre l'hôte et le symbionte (Won et al. 2008). Chez les bathymodiolinae, le clade « *thermophilus* » présente des espèces sulfo-oxydantes strictes dans le Pacifique et une double symbiose sulfo-oxydant-méthanotrophe dans l'Atlantique alors que le clade « *childressi* » présente surtout des espèces méthanotrophes strictes (Fiala-Medioni et al. 2002, Duperron et al. 2005, Duperron et al. 2006). Ces résultats ne s'opposent cependant pas complètement à l'hypothèse d'une transmission verticale car il suffit de rares épisodes de perte et d'acquisition des symbiontes pour que toute relation de co-spéciation disparaisse (Shoemaker et al. 2002). De la même façon, il a été montré que les bathymodiolinae présentant une symbiose utilisant des bactéries sulfo-oxydantes peuvent toutes abriter dans leur cavité palléale un polychète commensal du genre *Branchipolynoe* (Jollivet et al. 1998) qui pourrait avoir un rôle important dans la transformation de l'H₂S en thiosulfate (Martineu 1997). Ces bivalves possèdent un arsenal leur permettant d'exploiter tantôt le sulfure d'hydrogène, tantôt le méthane, tantôt les deux sources d'énergie. Ce type d'adaptation n'est sans doute pas sans conséquence sur l'évolution des protéines de la chaîne respiratoire mitochondriale de ces espèces, et, donc pourrait avoir un impact sur les phylogénies utilisant le gène mitochondrial COI.

II.4.3. Evolution du gène mitochondrial COI chez les Bathymodiolinae : évidence d'une convergence adaptative associée à la symbiose?

Il est intéressant de comparer la topologie des arbres réalisés en NJ à partir des données nucléotidiques et acides aminés de COI. L'arbre réalisé à partir des nucléotides distingue clairement de nombreux clades divergents dont la plupart correspondent à une espèce morphologique particulière. L'arbre met en évidence les groupes majoritaires déjà décrits chez les Bathymodiolinae (notamment les clades « *thermophilus* » et « *childressi* ») mais également des groupes associés aux genres *Gigantidas* ou *Idas*. La topologie de l'arbre

réalisé à partir des séquences traduites en acides aminés est toute autre. Au lieu d'un arbre diversifié présentant des branches profondes comme c'était le cas pour l'arbre nucléotidique, l'arbre se subdivise en quatre grands clades en râteau, séparés chacun par une seule mutation non synonyme ($M \rightarrow L$ et $S \rightarrow A$). Il est vraiment surprenant que la diversité non-synonyme spécifique et taxonomique des Modiolinae et Bathymodiolinae se résume à quatre clades mélangeant indifféremment les différents genres des milieux réduits profonds. Il existe cependant deux particularités : (i) les *Gigantidas gladius* sont groupées dans le clade 1 alors que les autres espèces du même genre (*Gigantidas horikoshii* et les espèces régionales) forment un clade bien distinct. (ii) chez les espèces « géographiques » de *Bathymodiolus* des bassins arrière-arc du Pacifique Ouest dont deux types ont été caractérisés ('Bathy'sp._Bassin de Lau et 'Bathy'sp._Papouasie-Nouvelle Guinée), se retrouvent dans des clades « acides aminés » différents. Il s'agit d'espèces géographiques et morphologiques proches vivant en sympatrie mais, qui, compte tenu de leur divergence génétique, appartiennent certainement à des lignées phylogénétiques distinctes. L'arbre réalisé par méthode de maximum de vraisemblance à partir du modèle de substitutions des codons de Goldman & Yang (1994) sur 54 taxons indépendants nous indique que ce marqueur est fortement contraint par la sélection purifiante (124 codons variables, longueur de l'arbre en dN = 0,37, longueur de l'arbre en dS = 156,06, ω moyen fixé = 0,0023). Ces quatre clades ne correspondent ni à des caractéristiques spécifiques des modioles, ni à l'habitat, ni à la profondeur ni à une aire géographique particulière. Cependant, les deux types de changements en acides aminés semblent présenter des réversions selon les clades incriminés, notamment au niveau du clade paraphylétique '*childressi*', laissant supposer qu'une convergence évolutive pourrait exister entre certains taxons. A ce titre, il est intéressant de noter que le remplacement $S \rightarrow A$ affecte systématiquement tous les taxons présentant une symbiose à un seul symbiote, que celle-ci soit avec le symbiote sulfo-oxydant (*Gigantidas*) ou avec le symbiote méthanotrophe (*B. japonicus/securiformis/platifrons/childressi*). Les reconstructions de séquences ancestrales par approche bayésienne montrent que les acides aminés sérine (pos 74) et leucine (pos 80) correspondent aux caractères ancestraux (probabilité = 0,999-1,000) des Modiolinae et devaient être présents chez l'ancêtre commun à tous les Bathymodiolinae. Cette quasi-absence de remplacements non-synonymes peut s'expliquer de deux façons : (1) compte tenu de la sélection purifiante, il n'existe qu'un très faible nombre de codons covariants libre de muter (càd une très faible proportion de codons suivant le "*covarion model*" de Fitch & Markowitz (1970)), et le polymorphisme en acides aminés observé entre taxons pourrait alors être presque neutre où (2) ces deux sites proches (pos 74 et 80) constituent un 'point chaud'

ou la sélection positive peut s'exercer permettant de fixer tantôt un acide aminé, tantôt l'autre, selon les pressions locales de sélection (convergence adaptative plus ou moins liée au mode de symbiose employé). Un argument allant dans le sens de la deuxième hypothèse est le fait que nous n'observons que deux possibilités de changements en acides aminés pour ces sites alors que d'autres types mutationnels auraient pu exister (par exemple, S→T ou S→P). De plus, l'analyse bayésienne effectuée à l'aide de PAML3.14 (Yang 2000) recherchant les sites sous sélection positive (modèles permettant de comparer différentes classes de dN/dS entre codons) montre sans aucune ambiguïté que tous les codons ($w_0 = 0.0037$, $p = 1,000$) sont sous forte sélection purifiante que l'analyse soit faite avec le modèle neutre (2 classes de dN/dS : mutation délétère, mutation neutre) ou le modèle de sélection positive (3 classes de dN/dS : délétère, neutre et positive). Les mutations non-synonymes trouvées entre espèces ne constituent donc pas une classe de mutations ayant un comportement neutre dans la phylogénie des Bathymodiolinae.

Cet effet sélectif est renforcé par le fait que la diversité mitochondriale est beaucoup plus élevée chez les Mytilidae côtiers. Chez les espèces côtières proches (*Mytilus*), le niveau de divergence observé en intra-spécifique est en effet supérieur à celui observé au sein des quatre clades chez les Mytilidae profonds. En généralisant, on peut donc considérer qu'il n'existe que trois grands clades mitochondriaux « profonds » : (i) celui comprenant tous les *Bathymodiolus*, *Tamu* et les Modiolinae *Idas*, *Idasola* et *Adipicola* ; (ii) celui des *Gigantidas* et (iii) celui des *Benthomodiolus*.

II.4.4. Histoire évolutive des Bathymodiolinae et colonisation des dorsales océaniques

La mise en évidence d'un aussi grand nombre d'espèces colonisant les environnements réduits nous pousse à réfléchir sur les processus biogéographiques ayant conduit à une telle expansion. La présence de la sous famille des Modiolinae à la base du groupe des Bathymodiolinae semble être un élément convaincant permettant d'affirmer que les environnements profonds réduits aient pu être colonisés par des espèces moins profondes ayant utilisé les bois coulés ou les os de baleine comme zones-relais (Craddock et al. 1995, Distel et al. 2000, Samadi et al. 2007). En revanche, les relations entre zones de suintements froids et sources hydrothermales sont nettement plus difficiles à déterminer. En effet, il semble que le scénario consistant en une colonisation des sources à partir des peuplements des

zones de suintements froids, celles-ci étant temporellement plus stables, soit vraiment réducteur. D'après les arbres phylogénétiques obtenus, la colonisation de ces deux types d'habitat a dû se faire de manière concomitante ou multiple car il n'existe pas de distribution hiérarchisée des espèces de ces deux environnements.

Une question générale restant en suspens est de pouvoir retracer les routes ayant permis la colonisation des dorsales. A l'échelle globale, on peut se demander si toutes les espèces actuelles découlent d'une espèce ancestrale unique venant de la surface, ou si au contraire la colonisation des dorsales a fait suite à une récurrence d'invasions issues d'environnements réduits contrastés. La faible diversité observée au niveau des séquences COI en acides aminés laisse à penser que la diversité initiale était faible et que l'histoire évolutive des Bathymodiolinae fait suite à une radiation récente. La faible diversité protéique observée peut cependant s'expliquer par des processus sélectifs masquant la diversité initiale et n'est donc pas en contradiction avec notre hypothèse d'un scénario à colonisation multiple.

La question des relations phylogéographiques entre communautés hydrothermales à l'échelle des grandes « provinces » biogéographiques (Atlantique, Pacifique Est, Pacifique Ouest, Océan Indien) a déjà l'objet d'un grand nombre de débats (Tunnickliffe 1991, Tunnickliffe & Fowler 1996, Van Dover et al. 2002, Bachraty et al. 2008). En fait, il existe deux grands scénarios de colonisation. Le premier repose sur une communication entre les espèces atlantiques et celles du Pacifique via les bassins arrière-arc du Pacifique Ouest et l'Océan Indien. Dans ce cas, les séquences des espèces indiennes/ouest-pacifiques devraient se trouver en position intermédiaire dans les arbres phylogénétiques, entre les groupes 'Atlantique' et 'Pacifique Est'.

Le deuxième repose sur une relation directe entre le Pacifique Est et l'Atlantique à travers l'Isthme de Panama, pouvant faire suite à des relais au niveau des zones de suintements froids du Golfe du Mexique. Ce scénario trans-panaméen est nécessairement antérieur à la fermeture de l'isthme (3,5 MA généralement utilisée pour les espèces côtière, mais on peut considérer 10 MA comme un minimum pour les espèces profondes (Ross & Scotese 1988). Sous cette hypothèse, les arbres devraient présenter les espèces des suintements du Golfe du Mexique comme intermédiaires entre les espèces du Pacifique et celles de la Ride Médio-Atlantique. Pour mieux évaluer ces deux scénarios, il nous faut

considérer les clades '*thermophilus*' et '*childressi*' séparément car ils présentent des histoires évolutives radicalement différentes.

Quel que soit le marqueur utilisé, au niveau du clade nommé « *thermophilus* » par Jones et al. (2006), les espèces des suintements froids (*B. brooksi*, *B. sp* Costa Rica, *B. boomerang* & *B. heckerae*) situées de part et d'autre de l'Isthme de Panama présentent une position centrale par rapport aux espèces hydrothermales, ce qui permet de corroborer l'hypothèse de la colonisation des environnements hydrothermaux de chaque océan à partir d'espèces communes de suintements froids. Cette communication directe est cependant rendue moins évidente, par la présence centrale du clade composé de certaines espèces hydrothermales du Pacifique Ouest et de la dorsale indienne, leur présence privilégiant alors l'hypothèse de la colonisation par la voie de la dorsale indienne.

Au niveau du clade « *childressi* », le scénario le plus parcimonieux apparaît plus évident puisque toutes les espèces atlantiques dérivent directement de celles du Pacifique Ouest par paraphylie. Les deux clades se partagent l'ensemble des environnements réduits profonds, sans qu'il soit possible de faire une distinction entre leur type d'habitat, leur profondeur ou leur type de symbiose. Cela laisserait penser que la colonisation des milieux profonds s'est effectuée à partir des deux grandes lignées ancestrales divergentes qui se seraient distribuées sur l'ensemble des dorsales et auraient donné naissance aux deux grands groupes que l'on connaît aujourd'hui.

II.4.5. Evolution et histoire régionales des espèces de *Bathymodiolinae*.

A une échelle plus régionale, comment expliquer la présence d'espèces différentes sur une même dorsale comme c'est le cas sur la dorsale du Pacifique Est et la Ride Médio Atlantique ? Quels sont les mécanismes de spéciation ayant permis leur apparition et leur maintien ?

Si l'on s'intéresse à l'histoire du Pacifique Est en particulier, l'ajout de séquences provenant d'espèces de suintements froids dans cette région (Costa Rica) constitue un apport considérable aux connaissances permettant d'étayer les hypothèses de colonisation de la dorsale. Dans un premier temps, les séquences COI mettent en évidence l'antériorité des moulières des suintements froids par rapport à celles des sources hydrothermales, qu'elles

proviennent de l'EPR ou de la Ride Médio-Atlantique. La phylogénie établie à partir du marqueur ITS2 renforce cette hypothèse et fournit des informations supplémentaires quant aux modalités de la colonisation de la dorsale EPR. En effet, ce gène est particulièrement résolutif à l'échelle du complexe d'espèces *B. thermophilus* et permet de distinguer trois « ensembles ». On peut ainsi déduire des données que la colonisation s'est déroulée de l'Isthme de Panama via les marges actives de l'Amérique du Sud (fosse du Pérou) au niveau du sud de la dorsale (plus fort apparemment entre les séquences des suintements froids du Costa Rica et celles de 37°S/EPR), puis que celle-ci est remontée vers le Nord. En effet, les séquences provenant de 21°S, 7°S et 9°50N forment les branches dérivées du groupe de 37°S/EPR.

Deux grands mécanismes de spéciation peuvent être invoqués pour rendre compte de l'extrême diversité génétique des modioles à l'échelle locale : la spéciation parapatrique (Tunnicliffe et al. 1993, Jollivet et al. 1995, Bates et al. 2005) et la spéciation vicariante de type allopatrique (Tunnicliffe 1991, Tunnicliffe & Fowler 1996, Johnson et al. 2006, Matabos 2007). Dans le premier cas, les espèces actuelles se seraient formées suite à des événements de spéciation en parapatrie découlant d'adaptation à des microenvironnements résultant de l'extrême variabilité locale de l'environnement hydrothermal. Dans le second cas, une spéciation par vicariance pourrait s'être produite après l'instauration d'une barrière physique aux flux géniques (création d'une faille transformante, mise en place d'une microplaque, formation d'un courant transverse ou subduction d'une dorsale). Ce type de modification tectonique aurait comme résultat de diviser l'aire de répartition de l'espèce ancestrale entraînant la création de deux nouvelles espèces-sœurs. Cette hypothèse a été mise en avant pour la première fois par Tunnicliffe (1991) pour expliquer le grand nombre d'espèces jumelles ou sous-espèces trouvées de part et d'autre de la faille de St Andreas (pointe de Californie) entre les communautés de l'EPR et Juan de Fuca. Ce scénario colle aussi très bien avec ce qui a été observé sur la dorsale du Pacifique Est. Cette dorsale, relativement récente (- 18 MA: (Mammerickx et al. 1980)) est le résultat de la subduction de la 'Mathematician ridge' sous la plaque américaine. La portion sud (Antarctique-Pacifique : 32-38°S) a été isolée des autres par le nord lors de la formation de la microplaque de l'île de Pâques, il y a 5,9 MA (Naar & Hey 1991). A ce niveau, les flux géniques entre le nord et le sud ont été complètement interrompus entraînant une spéciation de type allopatrique observable aujourd'hui sous la forme de la divergence entre l'espèce *B. thermophilus* (nord) et l'espèce sud située entre 32°S et 37°S. L'hypothèse d'une barrière tectonique d'une telle ampleur ne

semble pas cependant fonctionner pour la dorsale médio-atlantique sur laquelle plusieurs espèces sont présentes sans qu'aucune barrière stricte aux flux géniques ne soit observée.

Une autre hypothèse repose sur une vision plus dynamique de l'évolution des espèces, et notamment que celles-ci, au regard de leur fort potentiel dispersif peuvent utiliser de nombreux environnements relais pour coloniser le milieu où elles sont actuellement observées. En effet, les bivalves sont connus pour avoir une très forte fécondité (Davaine 1853), une phase larvaire qui leur confère un fort potentiel dispersif, et une très forte plasticité phénotypique. Ces caractéristiques auraient pu leur permettre d'utiliser les zones de suintements froids comme pierres de gué. On trouve une excellente illustration de ce phénomène en étudiant les espèces *Bathymodiolus* des suintements froids rencontrées de part et d'autre de l'Atlantique sur les marges africaine et américaine. Pour expliquer l'absence de différenciation génétique de ces populations, Olu et collaborateurs (2007) ont ainsi évoqué l'hypothèse d'une communication trans-océanique des faunes africaines et américaines à l'aide des courants profonds. De même, la proximité génétique d'espèces du Pacifique Ouest (à proximité des côtes japonaises : *B. septemdiarum*) avec celle trouvée au milieu de l'Océan Indien (*B. marisindicus*) renforcent l'idée d'une dispersion sur de très grandes distances.

Ces phénomènes de dispersion 'longue distance' confortent l'hypothèse selon laquelle les milieux réduits profonds ont pu être colonisés par un petit nombre d'espèces ancestrales qui se sont dispersées pour coloniser au final l'ensemble des environnements réduits. Cependant, il est parfois difficile de comprendre les mécanismes permettant le maintien de deux espèces proches au sein d'une même aire géographique et ne présentant pas de barrière physique évidente, sans prendre en compte des mécanismes d'isolement génétique tels que le renforcement, comme c'est le cas avec les deux espèces présentes dans le nord de la dorsale médio-atlantique (*B. azoricus* et *B. puteoserpentis*).

D'après nos arbres phylogénétiques, il ne fait peu de doute que ces deux espèces descendent des espèces des suintements froids rencontrés dans l'Atlantique (complexe d'espèces *B. boomerang*). Dans ce cas particulier, il ne s'agit pas d'une colonisation depuis l'Océan Indien *via* l'Atlantique Sud. Les deux espèces ont donc une origine 'suintements atlantiques' qui peut suivre deux scénarios évolutifs : (1) une colonisation de la dorsale pas à pas avec effets de fondation répétés et sélection environnementale (spéciation parapatrique)

ou (2) deux événements indépendants de colonisation de la dorsale avec formation d'une zone de contact secondaire (spéciation allopatrique avec renforcement).

II.4.6. Polyphylie au niveau du marqueur ITS2 et possibilités d'hybridation chez certaines espèces

La famille des Mytilidae a souvent été étudiée en milieu côtier pour sa capacité d'hybridation entre espèces proches (Bierne et al. 2003). L'hybridation entre certaines espèces proches de Bathymodiolinae peut entraîner des biais dans l'interprétation des résultats phylogénétiques selon que le marqueur est nucléaire ou mitochondrial. Le marqueur ITS2 est reconnu pour être un marqueur fiable à un niveau inter-spécifique mais pas à un niveau intra-spécifique ou l'homogénéisation de la famille multi-copies de l'ADN ribosomique dépend des taux de mutation et d'évolution concertée (Li & Graur 1991). Ainsi, si les espèces s'hybrident, la barrière intra/inter-espèce peut s'estomper à cause des phénomènes d'introgession. Les phylogénies réalisées alors ne représentent plus les relations phylogénétiques strictes entre les espèces mais plutôt un faisceau de relations entre espèces proches générant de la polytomie (râteau évolutif) ou de la polyphylie. C'est très clairement ce que nous observons ici avec l'espaceur intergénique au niveau des espèces *B. puteoserpentis* et *B. azoricus* (séquences retrouvées à plusieurs endroits de la phylogénie). Cette potentialité d'hybridation permettrait d'expliquer la présence d'allèles très divergents dans les populations étudiées et leur regroupement au sein de taxons différents (ici *B. azoricus* ou *B. childressi*). Ces allèles peuvent être interprétés comme le reflet d'hybridation récente, ou alors d'une introgession ancienne au cours de laquelle certains allèles ancestraux auraient été conservés de part la nature multi-copies du marqueur. Il a été démontré dans la littérature que les représentants « profonds » de cette famille possèdent la capacité de s'hybrider entre espèces proches. Les premiers à avoir formulé cette hypothèse sont Jollivet et al. (1998) au sujet des espèces *B. azoricus* et *B. puteoserpentis* le long de la dorsale médio-atlantique. O'Mullan et al. (2001) puis Won et al. (2003a) ont confirmé ces propos en démontrant l'existence d'une zone d'hybridation au niveau du champ hydrothermal Broken Spur (29°N/MAR) située à la confluence des aires géographiques de chacune des espèces. Ces derniers auteurs ont également supposé que d'autres zones d'hybridation pouvaient également exister, notamment sur la dorsale Est Pacifique, de part et d'autre de l'Ile de Pâques car les espèces allopatriques trouvées possèdent entre elles un niveau de divergence similaire à celui trouvé entre les espèces atlantiques hydrothermales (Won et al. 2003b). Si l'on considère

qu'un niveau de divergence inférieur à 6% puisse être un indicateur de la capacité d'hybridation entre deux espèces (il s'agit du niveau de divergence qui sépare les espèces *B. azoricus* et *B. puteoserpentis* pour lesquelles l'hybridation a été démontrée (O'Mullan et al. 2001), de très nombreuses espèces de *Bathymodiolus* décrites ont alors un réel potentiel pour s'hybrider.

II.5. Conclusion

Notre étude phylogénétique a mis en évidence une forte diversité spécifique chez les modioles hydrothermales (supérieure à celle déjà décrite dans la littérature). L'apport de nouvelles séquences et la confrontation de deux marqueurs moléculaires nous ont permis de montrer que les espèces des zones de suintements froids pouvaient être à l'origine des espèces hydrothermales pacifiques et atlantiques. Cela nous a également permis d'établir des scénarios de colonisation et de dispersion qui seront testés dans la suite de notre étude. Nous avons aussi observé que des contraintes sélectives très fortes (sélection purificatrice notamment) pouvaient s'exercer sur le gène mitochondrial COI.

Chapitre III. Contact secondaire en milieu profond : analyse multi-locus de la divergence et des flux géniques entre deux espèces proches de *Bathymodiolus*.

Ce chapitre correspond à un article préparé pour la revue *Molecular Biology and Evolution*. Le manuscrit près à être soumis est présenté en annexe de ce chapitre après un résumé en français.

III.1. Résumé

La compréhension et la reconstitution de l'histoire et des flux géniques entre les espèces proches est l'une des préoccupations majeure actuelle des biologistes de l'évolution. (Hey et Machado 2003). L'analyse du polymorphisme nucléotidique associée à des nouvelles méthodes statistiques basées sur l'utilisation de la coalescence, fournissent des méthodes élégantes pour tester l'hypothèse de l'existence de flux géniques pendant les processus de divergence (Hey 2006). Cette approche, appliquée à un certain nombre d'espèces proches (Won 2005, Niemiller 2008), a permis dans la plupart des cas de proposer que des flux géniques jusque là insoupçonnés se sont produits pendant les phases de divergence conduisant éventuellement à la spéciation (Nosil 2008). Ce constat semblerait indiquer que le mode de spéciation en parapatrie est peut être plus répandu que ce que l'on peut imaginer. Il n'est cependant pas évident que les flux géniques détectés impliquent des échanges constants tout au long de la période de divergence. En effet, il est difficile de distinguer une situation de contact secondaire avec introgression, d'une situation avec une différenciation primaire en cours avec flux géniques (càd parapatrie) car la variabilité des processus de coalescence peut conduire à des situations très contrastées depuis une monophylie réciproque jusqu'à un fort partage des allèles (Machado 2003). Dans tous les cas, la sélection naturelle est suspectée d'empêcher les flux géniques entre les régions du génome qui sont devenues incompatibles

du fait par exemple d'une adaptation différentielle de la part de l'une ou l'autre des espèces. Dans d'autres régions du génome, les flux géniques seront maintenus car les gènes seront soit complètement neutres ou incompatibles, soit même verront leur transfert dans d'autres compartiments favorisés car inconditionnellement avantageux. Les flux de ces différents types de gènes sont contrôlés par un équilibre entre migration, sélection, et recombinaison. Cette hétérogénéité des flux géniques entre les différentes régions du génome est appelée barrière semi-perméable aux flux géniques. (Harrison 1990).

L'étude présentée ici concerne le genre *Bathymodiolus*. Les deux espèces proches, *Bathymodiolus puteoserpentis* et *B. azoricus* vivent dans ce qui semble être le théâtre idéal pour étudier les phénomènes de spéciation : le milieu hydrothermal. Les populations et les espèces vivant dans ce milieu profond, n'ont été que très peu sensibles aux grandes perturbations climatiques qui ont affecté les milieux terrestres et côtiers et ne semblent donc pas avoir subi de longues périodes d'isolement géographique. De plus, la grande variabilité des conditions environnementales (température, conditions chimiques, tectonique et volcanisme) semblent être autant d'éléments qui peuvent conduire à de l'adaptation locale ou de la spéciation écologique. Notre étude, couplant sept locus nucléaires et un locus mitochondrial a été menée sur les populations des deux espèces proches, qui sont situées de part et d'autre d'une zone d'hybridation déjà décrite sur la ride médio-atlantique Figure III.1.

L'analyse des réseaux (Figure III.3) nous présente une forte variabilité du signal phylogénétique selon les locus. En effet pour certains d'entre eux tous les allèles sont partagés et aucune divergence n'est observable, alors qu'au contraire deux locus (CO1 et SAHH) présentent une nette divergence. Le locus SAHH est un peu particulier car sa généalogie est en étoile et les tests indiquent que ce locus aurait pu faire l'objet d'un balayage sélectif.

Une analyse avec le logiciel IMA (*Isolation with Migration* Hey et Nielsen 2007) nous a permis de montrer l'existence d'un flux génique, asymétrique, entre les populations qui étaient pourtant considérées comme étant situées en dehors de la zone hybride (Figure III.7). Le niveau de cette introgression est très variable d'un locus à l'autre et la question se pose de savoir si ces résultats sont liés à une spéciation en cours (parapatricie) ou au contraire s'il s'agit d'un contact secondaire entre les espèces.

L'analyse du spectre des fréquences alléliques des différents locus indique un excès de mutations en fréquences intermédiaires qui peut être interprétés comme la preuve d'une évolution distincte de deux lignées indépendantes qui auraient été remises en contact (Figure III.11).

Le locus GF1B présente deux clades d'allèles majoritaires pouvant correspondre aux deux lignées (ou espèces) ancestrales. Les réseaux (Figure III.3) mettent en évidence l'existence de séquences intermédiaires que l'on peut considérer sans ambiguïté comme des recombinants naturels formés à partir d'une séquence de chacun des clades divergents. La présence d'allèles recombinés en fréquences suffisante pour être capturés dans notre étude est un solide argument en faveur de l'hypothèse d'un contact secondaire. Il semble donc que contrairement aux présupposés concernant le milieu hydrothermal, les mécanismes de spéciation allopatriques sont à l'œuvre.

Un possible balayage sélectif a été détecté sur le locus nucléaire le plus discriminant : le SAHH. De plus, une valeur négative et significative du D de Tajima a été obtenue pour le seul locus qui soit différentiellement fixé entre les deux espèces, le locus mitochondrial COI. Des épisodes de balayages sélectifs récurrents sont susceptibles de rajeunir les temps de coalescence des espèces qui possèdent une grande taille de population comme chez le bivalves marins (Bazin et al 2006). On peut donc supposer que c'est la sélection qui participe à l'accélération de l'évolution de nos deux locus vers des monophylies réciproques et qui leur confère le statut de marqueurs diagnostiques. Le balayage sélectif détecté pour le locus SAHH est peut être la conséquence de la sélection positive sur un gène de spéciation auquel il est lié.

III.2. Annexe du chapitre III: présentation de l'article près à être soumis: "Secondary contact zone in the deep sea: a multi-locus analysis of divergence and gene flow between two closely-related species of *Bathymodiolus*"

B. Faure^{1,2}, D. Jollivet^{1,2}, A. Tanguy^{1,2}, F. Bonhomme^{3,4}, N. Bierne^{3,4}

1. UPMC, Station Biologique de Roscoff, BP.74, Place Georges Teissier, 29682, Roscoff cedex, France
2. CNRS UMR 7144, Station Biologique de Roscoff, BP.74, Place Georges Teissier, 29682, Roscoff cedex, France
3. Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier, France
4. CNRS UMR 5554, Institut des Sciences de l'Evolution, Station Méditerranéenne de l'environnement littoral, 1 quai de la daurade, 34200 Sète

Mots clé:

Bivalve, hydrothermal vent, nuclear genes, hybrid zone, introgression, IMA,

III.2.1. Introduction / introduction

Reconstructing the history of divergence and gene flow between closely-related organisms is one difficult task of evolutionary genetics that has recently received much attention (Hey and Machado 2003). The analysis of DNA sequence polymorphism at multiple loci associated with new statistical coalescence-based inferences now provides an elegant approach to test for the existence of gene flow during the process of divergence (Hey 2006). This approach has been applied to a number of closely related species (Won and Hey 2005, Niemiller, Fitzpatrick, and Miller 2008)). Somewhat unanticipated, gene flow has been detected in most cases (Nosil 2008). Taken at face value, these observations would suggest parapatric speciation as a more common mode of speciation than previously thought. However, it is not clear that the detection of gene flow necessarily means genes continuously exchanged during the whole divergence period. Secondary introgression can be difficult to separate from ongoing primary

differentiation with gene flow (i.e. parapatry), because the high stochasticity of the coalescence process often leads to a full range of situations, from reciprocal monophyly to extensive allele sharing (Machado and Hey 2003). Furthermore, in both cases natural selection is expected to prevent gene flow in regions of the genome where species-specific adaptations operate while gene flow can still be substantial in regions lacking isolation genes, resulting in a semi-permeable barrier to gene flow (Harrison 1990).

We here address the issue of the history of gene flow during differentiation in an original environment, hydrothermal vents in the deep ocean. Oceanic ridges are motivating places to study mechanisms by which speciation occurs. Ridges are nearly all connected into a one-dimensional network that has spread over the globe following plate tectonics. Contrary to what prevails in surface, quaternary climatic oscillations have not perturbed so much the deep ocean and it is unclear whether long episodes of geographic isolation could have occurred in the recent history of

species inhabiting ridges. In addition, the high variability of hydrothermal vent conditions (steep thermal gradients and chemistry heterogeneity) could induce perfect conditions for ecological speciation. For example, because of the ultramafic nature of the oceanic crust beneath the Mid-Atlantic Ridge (Bougault et al. 1993), hydrothermal vent emissions from the deepest sites are naturally enriched in methane but nearly depleted in sulphides whereas it is the reverse in the shallower sites (Charlou et al. 2002). Depth, hydrothermal fluid composition in heavy metals, sulphides and methane, temperature and symbiotic associations are highly fluctuant parameters, which are likely to induce local adaptation and possibly ecological speciation. On the other hand, the subduction of ancient rifts and the emergence of new ones might have led to either isolation of vent fields at a large spatial scale on the long term or to transient events of spatial isolation on the short term, both phenomena being thought to have deeply affected the hydrothermal vent fauna (Tunnicliffe 1991, Van Dover et al. 2002). The constant modification of the hydrothermal vent landscape due to the fluid displacement led to recurrent 'crash and flush' in populations that might likely promote speciation and cause species crypticism (Matabos et al. 2008, Genio et al. 2008).

Mussels from the genus *Bathymodiolus* Kenk and Wilson 1985 are one of the most widespread mollusc dominating deep-sea hydrothermal vents and cold-seeps communities throughout the world. Interestingly, *Bathymodiolus* mussels still possess a dispersive larval stage, a biological feature that has long questioned the possibility of geographic isolation in marine species (Palumbi 1992). The genus *Bathymodiolus* has diversified a few millions of years ago following an explosive radiation leading to a large number of geographic species (Jones et al. 2006, Olu-Le roy et al. 2007). Two distinct

species of *Bathymodiolus*, *B. azoricus* and *B. puteoserpentis*, were morphologically described from the northern part of the Mid-Atlantic Ridge. Previous studies using both rDNA ITS2 intergenic spacer (Jollivet et al. 1998) and a combination of mitochondrial DNA sequences and allozymes (O'Mullan et al. 2001) demonstrated the hybridizing potential of these two species on a very restricted area where the distribution of the two species meet. Concordant narrow clines in allele frequencies have been described (O'Mullan et al. 2001) together with strong cytonuclear disequilibria in intermediate populations (Won et al. 2003a). The history of divergence and gene flow, however, remains to be settled. Indeed, the hydrothermal vent complex of *Bathymodiolus* species in the northern Atlantic extends over more than 3000 km along the ridge axis, with fields separated by large transform faults, which offset the rift over tens to hundreds kilometres (McDonald et al. 1986, Cannat et al. 1995). The distribution of *B. azoricus* is limited to the north because of the rise of the rift up to Iceland and its southern limit is the hybrid zone centred on the Broken Spur site (29°N). Conversely, the northern range limit of *B. puteoserpentis* is the hybrid zone while the southern limit is unknown. *Bathymodiolus* phylogeny clearly indicates that *B. azoricus* and *B. puteoserpentis* are sister species of the *Bathymodiolus boomerang* species that inhabit the deep Atlantic cold seeps situated both on the American and African margins (Olu-Le roy et al. 2007). Won et al. (2003a) therefore proposed that the Broken Spur vent field (29°N) hybrid zone corresponds to the dual arrival of migrants from more extreme source populations. Actually, this situation may suggest that either (i) *B. azoricus* and *B. puteoserpentis* have recently colonised the Mid-Atlantic Ridge via two independent entrance points and subsequently met again to hybridise in a secondary contact (hybrid zone hypothesis), or (ii) that ongoing speciation

is occurring between the two species, in which case one species would derive from the other along an environmental gradient (parapatric speciation hypothesis).

DNA sequence polymorphism data from one mitochondrial and seven nuclear genes were here gathered in order to assess genetic diversity, population subdivision and gene flow within and between both Atlantic vent species, and to perform divergence/polymorphism analyses using the outgroup species *Bathymodiolus thermophilus* (Pacific Ocean). The goals of this study were: (1) to infer the existence of gene flow in the process of divergence by using the Isolation with Migration (IM) model (Hey, 2004 #6259), (2) to estimate the genome-wide level of gene flow and attempt to analyse its possible variation among genes, and (3) to determine whether genetic variation follows the neutral mutation-drift expectations.

III.2.2. Matériel et méthode / Materiel and methods

III.2.2.a. Echantillonnage / Data collection

The three *Bathymodiolus* species were collected during the oceanographic cruises ATOS 2001, BioSpeedo 2004 (both on board the N/O L'Atalante), Momareto 2006 and Serpentine 2007 (both on board of the N/O Le Pourquoi Pas?). The first species, *Bathymodiolus azoricus*, was sampled from the sites Rainbow (2500m; 36°13'N-33°54'W) and Menez Gwen (850m, 37°50'N-31°31'W) at the Azorean triple junction. The second species, *Bathymodiolus puteoserpentis*, was sampled at the site Logatchev (3080m; 14°45'N-44°58'W) on the Mid-Atlantic Ridge (MAR) (Figure III.1). These two first species were collected using the

telemanipulated arm of the ROV Victor6000. The third species, *Bathymodiolus thermophilus*, was sampled at 7°25S vent field (2450m; 7°25'S-107°47'W) on the East-Pacific Rise (EPR) using the telemanipulated arm of the manned submersible Nautilie, and at the site Sebastian's Steamer (2216m, 37°47S-110°55'W) using the manned submersible Alvin. Samples were brought back to the surface alive inside an insulated basket and, subsequently dissected on board. Foot tissue and gills were immediately frozen and kept in liquid nitrogen until nucleic acid extraction back in our laboratory. DNA was extracted using the classical phenol-chloroform protocol (Sambrook and Russel 2001). A total of 44 and 52 individuals were used for the Rainbow and Menez Gwen *B. azoricus* populations respectively, 33 individuals for the Logatchev population of *B. puteoserpentis*, and 4 and 2 individuals for the outgroup *B. thermophilus* at the 7°S sites and 37°S respectively.

III.2.2.b. Amplification PCR, clonage et séquençage / PCR amplification, cloning and sequencing

DNA sequences were obtained for 7 nuclear loci and one mitochondrial gene (Figure III.2). Six are protein coding genes, Elongation Factor 1 α (EF1 α); S-adenosylhomocysteine hydrolase (SAHH); Collagen type XIV (COL), Lysozyme (LYZ), Ferritin GF1 (with two distincts paralogues: GF1_A and GF1_B). The two other loci are the mitochondrial Cytochrome-C Oxydase subunit I (CO1) and the ribosomal internal spacer 2 (ITS2). Targeted nuclear genes were obtained from a collection of *B. azoricus* ESTs (full-length cDNA library, (Tanguy et al. 2008)).

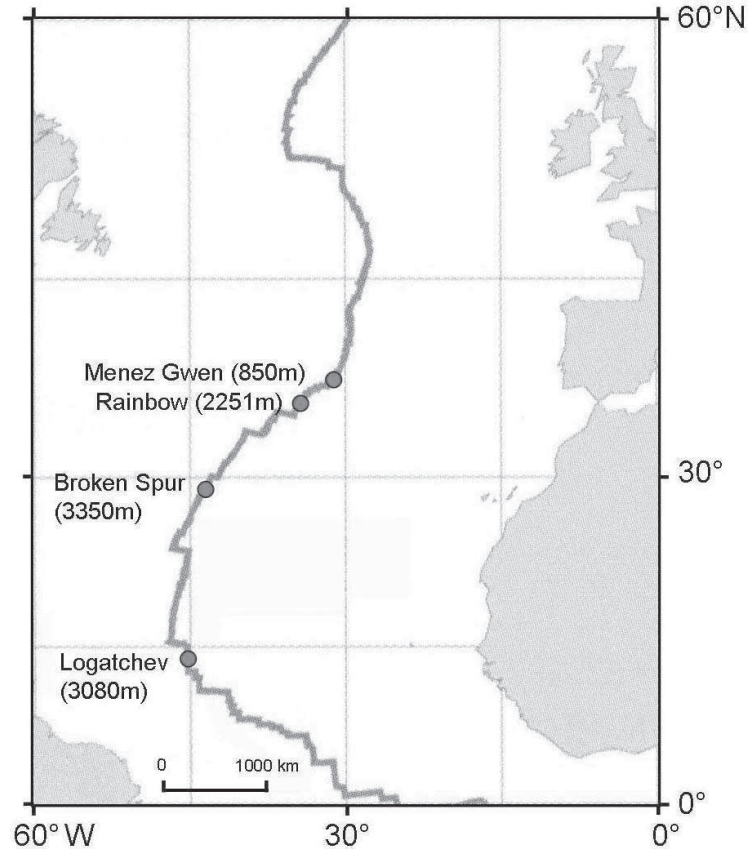


Figure III.1. Champs hydrothermaux de la dorsale médio atlantique / Mid-Atlantic Ridge hydrothermal vent fields.

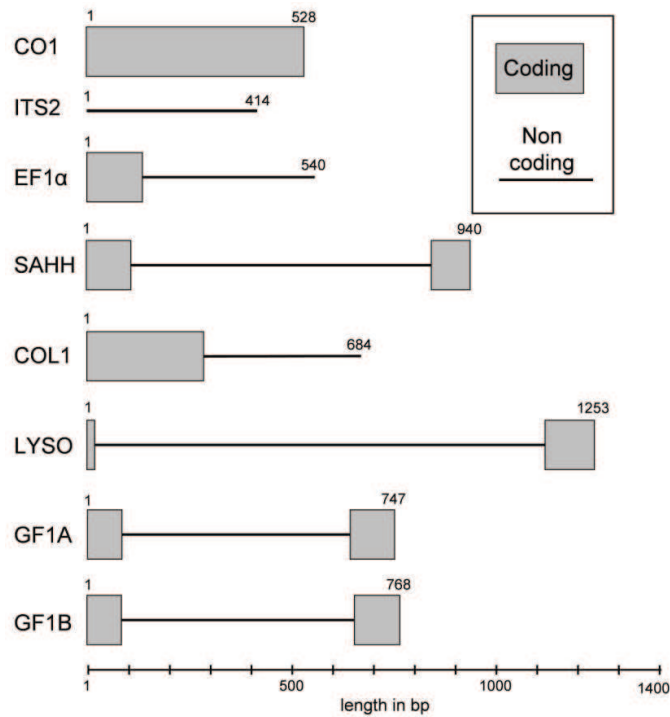


Figure III.2. Comparaison de la taille et de la structure des huit genes étudiés / Comparison of size and constitution of the eight gene fragments studied

PCR amplifications were done with specific sets of primers designed from the conserved exonic regions of *B. azoricus* EST sequences, for which an intron was

Table III.1. Primer sets and conditions used for COI, EF1 α and for Lysozyme have been described elsewhere (Faure et al. 2007). For new loci, genes were amplified and cloned following the Mark-Recapture (MR) method described in (Bierne et al. 2007 see also Faure et al. 2007, Faure et al. 2008) with 4-nucleotides 5'-tailed primers (4 bp polynucleotides tag at their 5' ends). Reactions were performed in a 20 μ l volume containing 1X PCR buffer (supplied with polymerase enzyme), 2nM MgCl₂, 0.25 μ M of each dNTPs, 0.4 μ M of each primer, 0.025 U of Thermoprime Plus *Taq* polymerase (ABgene), and 20ng genomic DNA. Cycling parameters were 94°C for 5 min, 35 cycles of 94°C for 45s, 60°C for 1 min, 72°C for 2 min; and a final elongation at

known to occur in other taxa (Bierne et al. 2000). The sequences of the primer sets are summarised in

72°C for 10 min for each gene. Every individual was PCR-amplified separately with different sets of 5'-tagged primers. Because we used MR cloning protocol, PCR products were mixed together (equimolar) and cloned into a pGEM-T vector by using the pGEM-T cloning kit (Promega, Madison, WI, USA). We used a capture effort of two in MR-cloning: 96 clones per population and per gene were sequenced at the Genoscope (Evry) (<http://www.genoscope.cns.fr/>) with the universal primer sp6 or T7 flanking the insert on the plasmid. A total of 4300 initial sequences was used for this study (direct sequencing for COI in both strands, and 96 clones sequenced in both strands, for 3 populations for the 7 nuclear genes).

Table III.1. Amorces utilisées pour les amplifications / Amplification primers

locus	5'-[size of the tag] primer Forward -3'	5'-[size of the tag] Primer Reverse -3'
COI	ATA AAA AGA TGT ATT RAA RTG ACG	TGT GGT CTG GAA TAA TTG GAA
EF1 α	[3bp] ACG CCT GGG TAT TGG ACA AAC T	[3bp] CCA AGA GGG GTC GTA CAA ATT A
SAHH	[4bp] TAA ATC TTG GTT GTG CTC ATG GTC ATC CA	[4bp] TTT GAA TGG TCC TTC TTT AGG TAG AC
COL1	[4bp] TTA CCA AAT CAC AGG TCG GGG TCA TG	[4bp] TCT CCT TCG CTA TTT TTG TGG G
LYSO	[4bp] GTT TCC CCA AAA TGT ATG AGC TGT	[4bp] TAA TCT CCG CCT GGA CTA CCA CAA TC
GF1A and GF1B	[4bp] ATG TCT CAA AGT CAA CCT CGT CAA AAT	[4bp] TTG ATG TCC TGA AGA ACG ATT CGT CCT CC
ITS2	ACA TTG CGG CTT TGG GTC AC	GCT TAA ATT CAG CGG GTA CT

III.2.2.c. Analyses des données / Data analyses

For all genes, sequences were checked and edited using the Chromas 2.22 computer program (Technelysium Pty. Ltd., Helensvale, Australia). The sequences from each dataset were initially aligned with the program ClustalW (Thompson, Higgins, and Gibson 1994) in

BioEdit Sequence Alignment Editor (Hall, 1999), improved manually. Artefactual mutations and *in vitro* recombinants were discarded when detected from alleles sequenced several times in multiple-recaptures. In order to prevent from sampling bias, only one sequence per individual was kept. In the specific case of heterozygote individuals, when both allelic forms were captured, the most frequently

captured allele was retained. We used the two populations of *B. azoricus* as distinct samples (never pooled) in order to test whether a “site effect” occurred between Menez Gwen (850m) and Rainbow (2500m) and to get partial replicates when comparing *B. azoricus* and *B. puteoserpentis*.

As intragenic recombination was often detected, gene trees do not necessarily reflect gene genealogies. To represent an overview of the genetic relationships among alleles we inferred a network with the NeighborNet algorithm (Bryant and Moulton 2004) in the software SplitsTree4 (Huson and Bryant 2006). In addition, we paid a special attention to reconstructing gene genealogies for two loci which hold useful information. For the SAHH locus, recombination was not detected and the single possible gene tree was reconstructed with Genetree (Bahlo and Griffiths 2000). For the *Gfl-B* locus, in which a few recombination breakpoints and a few recombinant alleles were easily identified, we reconstructed the more parsimonious genealogy without recombinants and added recombinants afterwards with the help of the ancestral recombination graph reconstruction made possible with the Beagle software (Lyngso, Song, and Hein 2005).

For each gene, the number of segregating sites (S), nucleotide diversities (θ_w (Watterson 1975) and θ_π (Nei 1987)), Tajima’s D statistic (Tajima 1989a), minimum number of recombination events (R_m) (Hudson and Kaplan 1985) and absolute divergences (D_{XY}) were estimated using the DnaSP version 4.10 software package (Rozas et al. 2003). Genetic differentiation between populations was estimated with the ϕ_{st} statistic analysis carried out with the Arlequin software (Excoffier, Laval, and Schneider 2005), using a matrix of Tamura-Nei’s distances between individuals; with 1000 permutations. Statistical significance of Tajima’s D tests for each locus and sample was assessed by coalescent simulations

with the software HKA (J. Hey’s web page:

<http://lifesci.rutgers.edu/~heylab/HeylabSoftware.htm#HKA>). Although the Tajima’s D test relies on polymorphism data, coalescence simulations are conditioned on Theta estimates that incorporate the information (e.g. mutation rate) contained in the divergence with the outgroup. A multilocus version of the HKA test (Hudson, Kreitman, and Aguadé 1987), another test of gene departure from neutral evolution that compare divergence and polymorphisms at several loci, was performed on the 8 loci with the HKA software in order to test whether differences in tree topologies reflect differing levels of evolutionary rates or selective processes. In this latter test, a locus under positive selection is supposed to have a polymorphism/divergence ratio smaller than that of a neutral one. Finally, when possible we also used the coalescence-based maximum-likelihood method of Galtier et al. (Galtier, Depaulis, and Barton 2000). This method is designed to detect a distortion in the shape of gene genealogies generated by a diversity-reducing event (hitchhiking or bottleneck). The likelihood of a model in which a drop in effective size of strength S occurred at time T in the past is compared to the likelihood of a constant-size model.

III.2.2.d. Analyse *IMa* / *IMa* analysis

To discriminate between the relative effects of divergence and gene flow on the speciation process, we analysed our data set under the Isolation with Migration analytic program (Hey and Nielsen 2007). The model assumes that an ancestral population splits into two descendant populations that may continue to exchange genes after separation. The method estimates posterior probability distributions for both ancestral and actual population sizes, directional migration rates between the two populations, and the time elapsed since population splitting. A MCMC

approach is used to sample the best coalescence tree from the space of genealogies and divergence, and to estimate population parameters. The posterior distributions of migration rates and population sizes are derived analytically from the sampled genealogies (Nielsen and Wakeley 2001, Hey and Nielsen 2004). Convergence by the Markov chain simulations toward the true stationary distribution is checked by monitoring multiple independent chains started at different starting points and by assessing the autocorrelation of the parameter values over the course of the run. We also used the procedure for swapping among multiple heated chains (Metropolis coupling) to further ensure that the distributions we obtained actually reflected the stationary distributions (Geyer 1992). Each locus was assigned an inheritance scalar, to adjust for its relative expected effective population size: 1.0 for autosomal locus, 0.25 for mtDNA. The IMA method assumes no recombination and allows two mutation models: infinite sites (IS) and HKY with back mutation (Hasegawa, Kishino, and Yano 1985). We assumed the HKY model and used the largest fragment without recombination for the IMA analyses.

We conducted analyses separately for each gene and then in a combined fashion. The 90% highest posterior density (HPD) interval that is the shortest span that includes 90% of the probability density was recorded as credibility interval for each parameter. Conversion into effective population size estimates were done from the relationship $N_e = \theta/4u$. Migration parameters of the model were used to calculate the number of migrants exchanged between the two hybridising species (i.e., $M = 2Nm$, the product of the effective number of gene copies and the per gene copy migration rate). Thus, $M = 2N_e m = (4N_e u \times m/u)/2 = \theta \times m/2$ (Hey and Nielsen 2004).

III.2.3. Résultats / Results

III.2.3.a. Variations génétiques / Genetic variation

To avoid a sampling bias in estimates and tests that assume random sampling, we used one multiple captured allele per individual, this allele being the most frequently captured in our mark-recapture cloning procedure (Bierne et al. 2007). From both Atlantic species and the outgroup *B. thermophilus*, a total of 559 allelic sequences have been analysed at one mitochondrial locus and 7 nuclear loci, leading to a global number of 437 756 bases sequenced (745 bp per sequence in average). Allelic and nucleotide diversities as well as divergence to the outgroup are summarized in Table III.2 for the three *Bathymodiolus* species. The weighted mean of Watterson's θ estimators (Watterson 1975) across autosomal loci (i.e. without COI), were 0.0076 0.0091, 0.0101 for *B. azoricus* (Rainbow), *B. azoricus* (Menez Gwen), *B. puteoserpentis*, respectively. The average divergence to the pacific species, *B. thermophilus*, at autosomal loci was 3.6%. The divergence observed at the mtDNA locus was much higher (11.8%) while the diversity was lower than the nuclear average, in accordance with a higher rate of evolution and a smaller effective population size of the mitochondrial genome. Pairwise Φ_{st} estimates between the two *B. azoricus* populations were never significantly different from zero. As a consequence the two *azoricus* samples will be pooled in the following analyses. However, the between-species comparison (*B. azoricus* / *B. puteoserpentis*) detected significant differences at five loci: COI ($P= 0.01$), SAHH ($P= 0.01$), LYSO ($P= 0.01$), EF1 α ($P= 0.05$), and ITS2 ($P= 0.05$).

Table III.2. Résumé des statistiques du polymorphisme nucléotidique des séquences mitochondriales et nucléaires. / Summary statistics of nucleotidic polymorphism for mitochondrial and nuclear sequences

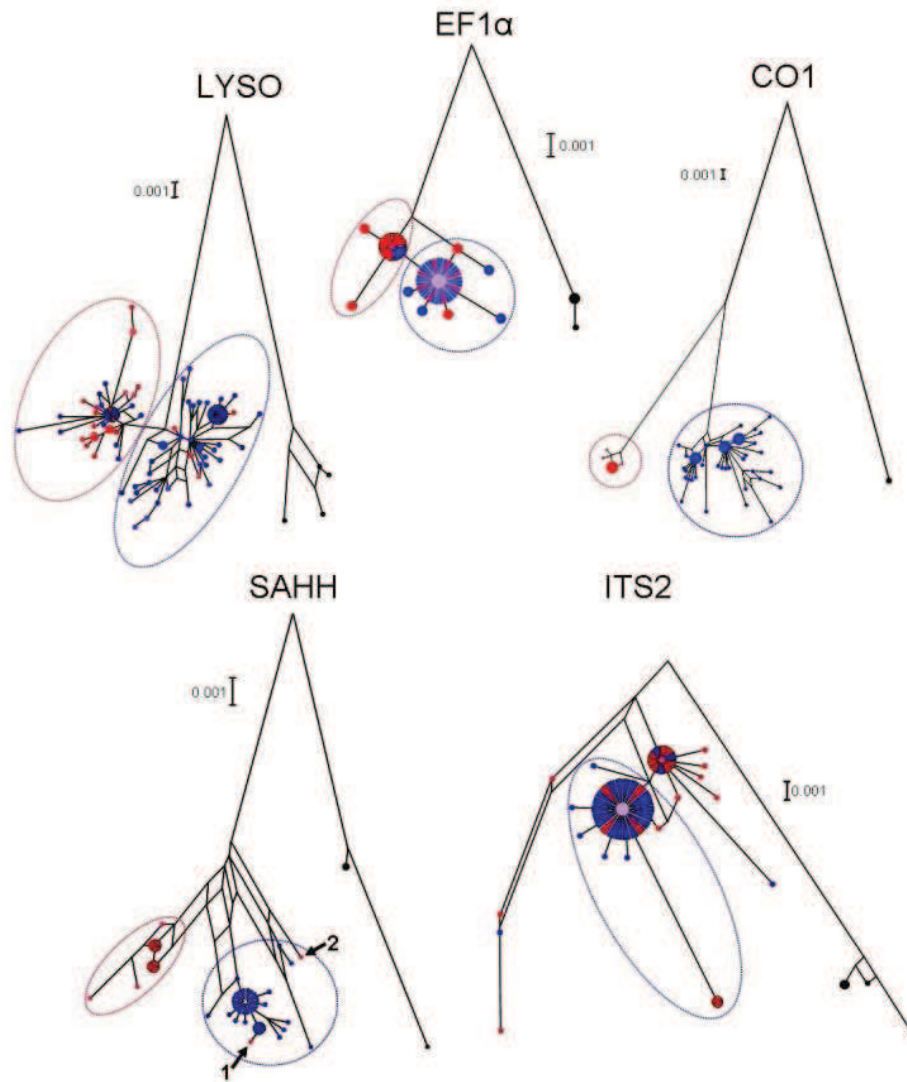
locus	Species	N	L	S	θ	π	Tajima's D	Div. (vs. 7S)	Div, Ba/Bp	Div. Rb/MG	Rm	Φ_{st}	
												MG	Bp
COI	B. azoricus_Rb	44	528	26	0.0113	0.0058	-1.6176*	0.1181	0.0611	0.0048	4	0.0217	0,9447***
	B. azoricus_MG	52	528	18	0.0075	0.0036	-1.6411* ^a	0.1178	0.0609		4		0,9627***
	B. puteoserpentis	33	528	3	0.0014	0.0006	-1.3758	0.1190			0		
	B. thermophilus 7S	4	528	0	0.0000	0.0000							
EF1 α	B. azoricus_Rb	14	540	5	0.0030	0.0013	-1.8893*	0.0258	0.0019	0.0009	0	0.0090	0,0833*
	B. azoricus_MG	18	540	2	0.0011	0.0004	-1.5078*	0.0254	0.0014		0		0,0907*
	B. puteoserpentis	15	540	6	0.0035	0.0021	-1.3326	0.0251			0		
	B. thermophilus 7S	4	540	1	0.0010	0.0009							
SAHH	B. azoricus_Rb	13	940	18	0.0064	0.0048	-1.0545	0.0396	0.0134	0.0050	0	-0,0404	0,5553***
	B. azoricus_MG	13	940	22	0.0079	0.0055	-1.3315	0.0405	0.0144		0		0,5648***
	B. puteoserpentis	11	940	26	0.0101	0.0075	-1.1897	0.0384			2		
	B. thermophilus 7S	4	940	14	0.0088	0.0100							
COL1	B. azoricus_Rb	18	683	37	0.0162	0.0116	-1.1564	0.0405	0.0106	0.0110	2	-0,0160	0.0482
	B. azoricus_MG	31	683	42	0.0154	0.0109	-1.0809	0.0407	0.0099		4		0.0102
	B. puteoserpentis	24	683	32	0.0125	0.0087	-1.1690	0.0408			1		
	B. thermophilus 7S	1	683	/	/	/							
LYSO	B. azoricus_Rb	32	1297	58	0.0123	0.0120	-0.0418	0.0587	0.0141	0.0120	11	-0,0035	0,2020***
	B. azoricus_MG	35	1297	59	0.0129	0.0124	-0.1372	0.0581	0.0142		17		0,1935***
	B. puteoserpentis	25	1297	50	0.0111	0.0101	-0.3588	0.0586			9		
	B. thermophilus 7S	4	1297	16	0.0071	0.0068							
GF1A	B. azoricus_Rb	24	748	22	0.0079	0.0061	-0.8371	0.0213	0.0072	0.0070	2	0.0280	0.0274
	B. azoricus_MG	20	748	25	0.0098	0.0074	-0.9462	0.0229	0.0074		5		-0,0324
	B. puteoserpentis	18	748	27	0.0105	0.0079	-1.0133	0.0228			5		
	B. thermophilus 7S	4	748	9	0.0066	0.0065							
GF1B	B. azoricus_Rb	14	768	25	0.0106	0.0137	1.2203	0.0318	0.0152	0.0179	3	0.0591	0.0202
	B. azoricus_MG	13	768	39	0.0174	0.0201	0.6576	0.0316	0.0172		2		-0,0536
	B. puteoserpentis	13	768	27	0.0117	0.0161	1.6281*	0.0300			3		
	B. thermophilus 7S	4	768	5	0.0036	0.0033							
ITS2	B. azoricus_Rb	14	413	6	0.0049	0.0030	-1.3944	0.0326	0.0082	0.0042	0	0.0231	0.0493
	B. azoricus_MG	14	413	13	0.0107	0.0053	-2.0740*	0.0318	0.0094		0		0,0617*
	B. puteoserpentis	22	413	28	0.0210	0.0127	-1.5021	0.0326			1		
	B. thermophilus 7S	4	413	3	0.0041	0.0042							

N : Number of sequences ; L : Average length (bp) of the sequences from each species; S: Number of polymorphic sites; θ : Estimate of $2N_f\mu$ (*mtDNA*) (N_f is the effective number of females) or $4N_e\mu$ per base pair using the number of polymorphic sites ((Watterson 1975)); π : Estimate of $2N_f\mu$ (*mtDNA*) or $4N_e\mu$ using the average number of nucleotide difference per site (Nei 1987); **Tajima's D**: Tajima's statistic ((Tajima 1989a); **Div .** Average number of nucleotidic substitutions per site between populations, from each taxon and the sequences of *B. thermophilus* ; **Φ_{st}** : population differentiation, tested by 1000 permutations with the Arlequin software

**III.2.3.b. Congruence du signal
phylogénétique entre les arbres des
différents locus / Congruence of the
phylogenetic signal across gene trees**

All loci displayed a large number of phylogenetically informative sites (77 per locus on average when excluding the outgroup). Networks inferred with the NeighborNet algorithm depicts the genetic relationship among alleles for the 8 genes in Figure III.3. The mitochondrial locus COI is the only locus that displayed complete reciprocal monophyly of alleles sampled in both species. For the nuclear genes SAHH and LYSO two well defined groups of alleles were apparent, each of which preponderantly composed of genes sampled in the same species with only a

few genes sampled in the other species. To a lesser extent, the ribosomal spacer ITS2 and EF1 α tended to fit fine with the species dichotomy *B. azoricus* and *B. puteoserpentis*. The three other genes COL1, GF1_A and GF1_B displayed highly reticulated networks and an absence of association with geography (species). Although one could have attempted to define two groups of alleles at loci GF1_A and GF1_B, there was no possibility to attribute these groups to a given species. The position of the outgroup and the shape of the network obtained at locus COL1 did not produce phylogenetic relationships which easily fit with a long period of divergence between two gene pools.



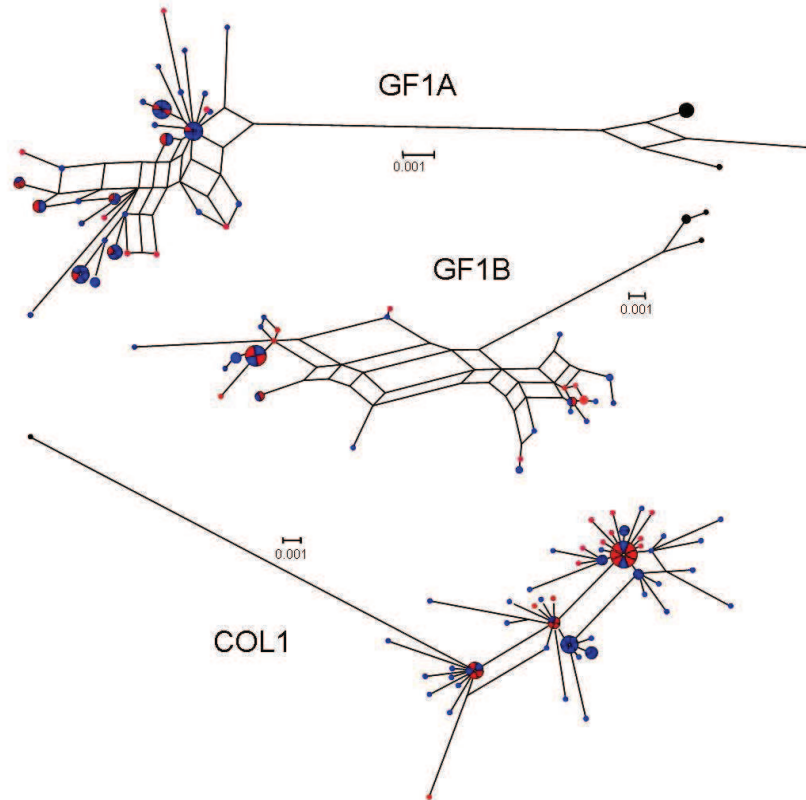


Figure III.3. Réseaux réalisés avec l’algorithm NeighborNet pour les 8 locus / Networks inferred with the NeighborNet algorithm for the 8 loci

The five loci which retained a probable footprint of differentiation were used to define two groups of alleles (ellipses in Figure III.3) and the frequency of such defined synthetic alleles was reported in Table III.3.

Variations in phylogenetic patterns across loci can be readily summarized by comparing the levels of species divergence across loci. We computed the net nucleotide divergence d (Nei and Li 1979), which is the average divergence between *B. azoricus* and *B. puteoserpentis* minus the average sequence divergence within both species $(\pi_{\text{puteo}} + \pi_{\text{azo}})/2$. Without inter-species gene flow, d should be positive, while under a model in which *B. azoricus* and *B. puteoserpentis* have exchanged genes since the separation from their common ancestry, d may fall down to zero and even become negative. The quantity d was positive for loci COI, SAHH and close to zero for the remaining loci (Figure III.4). d can then be compared

to the divergence obtained between the two Atlantic species and their Pacific outgroup. In the absence of gene flow, a positive relationship is expected between evolutionary rates. However, the relationship between the *B. azoricus* / *B. puteoserpentis* divergence and the Pacific / Atlantic divergence was not significant. Five loci (SAHH, COL1, LYSO, GF1_A, ITS2) out of the eight displayed similar ratios (Figure III.5). Three genes seemed to depart from the genomic trend. COI and GF1_B exhibited higher values, while EF1 α displayed a lower ratio.

Tajima’s D tests (Tajima 1989a) were used to measure the deviation from the mutation/drift equilibrium under neutrality (Table III.2). In the three mussel populations, Tajima’s D was mostly negative and sometimes significantly so (COI, EF1 α , ITS2). However, the LYSO loci displayed a D close to zero and the GF1B locus displayed a significantly positive D value. As

population structure has an important impact on Tajima's D, the average D was plotted against Φ_{ST} in Figure III.6. The higher Tajima's D observed at loci LYSO and GF1B could likely be the consequence of secondary introgression of divergent alleles while low gene flow at the CO1 and SAHH loci prevented many divergent alleles to segregate within the same

populations. However, it remained puzzling why the D was so low at the remaining four loci with low population structure. To complement Tajima's D tests, a multi locus HKA test has also been performed. This test is presented in Table III.4 and indicates that the ratio of polymorphism to divergence did not vary significantly among loci.

Table III.3. Distribution des alleles / Distribution of alleles

		number of alleles per clades		proportion of alleles per clades	
		"Bazori"	"Bputeo"	"Bazori"	"Bputeo"
CO1	<i>B. azoricus_Rb</i> (▼)	44	0	100.00	0.00
	<i>B. azoricus_MG</i> (▲)	52	0	100.00	0.00
	<i>B. puteoserpentis</i> (□)	0	33	0.00	100.00
SAHH	<i>B. azoricus_Rb</i> (▼)	13	0	100.00	0.00
	<i>B. azoricus_MG</i> (▲)	13	0	100.00	0.00
	<i>B. puteoserpentis</i> (□)	2	9	18.18	81.82
LYSO	<i>B. azoricus_Rb</i> (▼)	24	8	75.00	25.00
	<i>B. azoricus_MG</i> (▲)	25	10	71.43	28.57
	<i>B. puteoserpentis</i> (□)	5	20	20.00	80.00
ITS2	<i>B. azoricus_Rb</i> (▼)	10	4	71.43	28.57
	<i>B. azoricus_MG</i> (▲)	12	2	85.71	14.29
	<i>B. puteoserpentis</i> (□)	8	14	36.36	63.64
GF1B	<i>B. azoricus_Rb</i> (▼)	10	7	58.82	41.18
	<i>B. azoricus_MG</i> (▲)	6	7	46.15	53.85
	<i>B. puteoserpentis</i> (□)	7	6	53.85	46.15
GF1A					
EF1 α		more complexe coalescence			
COL1					

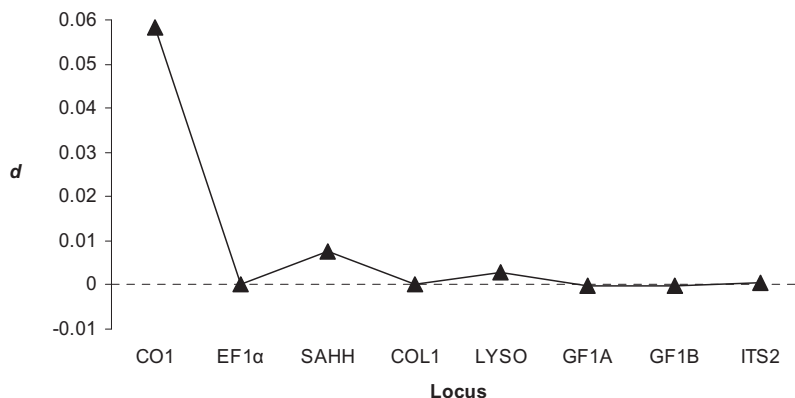


Figure III.4. Différence des divergences de séquences entre *B. azoricus* et *B. puteoserpentis*. Difference in sequence divergence between *B. azoricus* and *B. puteoserpentis*. Each point represents the value of absolute divergence d for the locus shown on the x-axis: $d = d_{azo/puteo} - (\pi_{puteo} + \pi_{azo})/2$ where $d_{azo/puteo}$ is the mean interspecific sequence divergence between *B. azoricus* and *B. puteoserpentis*, and $(\pi_{puteo} + \pi_{azo})/2$ is the mean population diversity between *B. azoricus* and *B. puteoserpentis*. Under a strict isolation model consistent with the traditional phylogeny (strict divergence azoricus/puteoerpentis), the value of d is expected to be greater from zero for most loci.

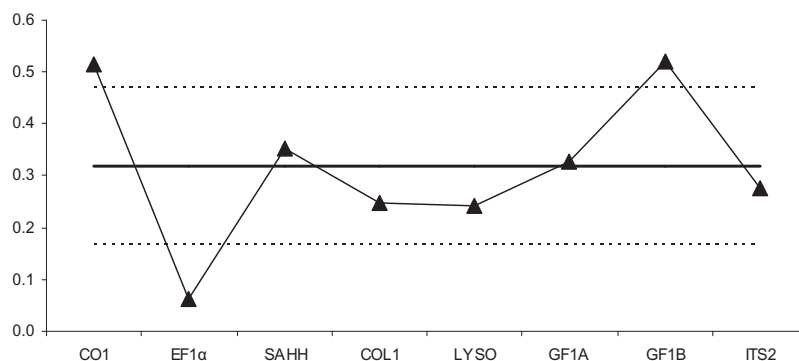


Figure III.5. Taux $D_{azi/puteo} / D_{ATL/PAC}$ pour les 8 loci étudiés. Rate $D_{azi/puteo} / D_{ATL/PAC}$ for the 8 loci studied. The horizontal line is the average divergence between the Atlantic and the Pacific. Dotted lines represent the standard deviation.

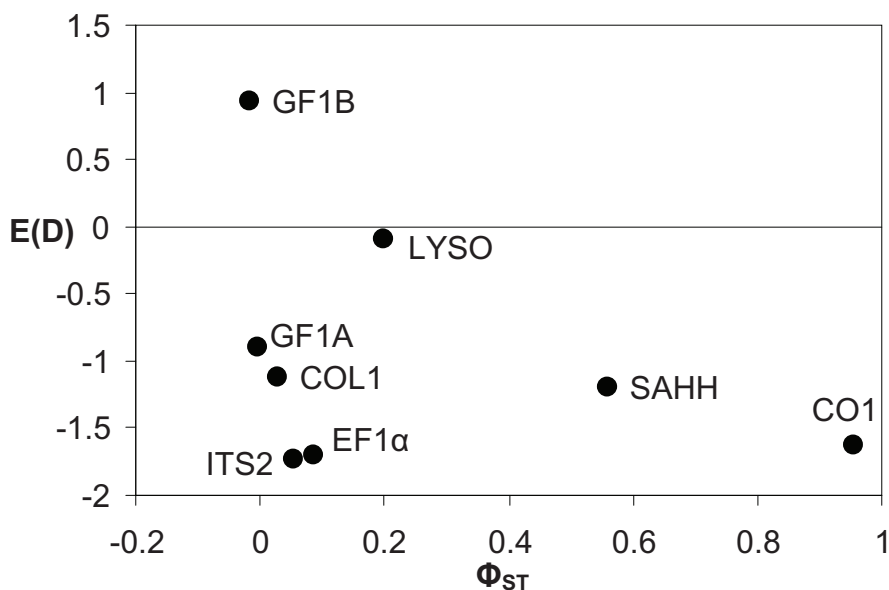


Figure III.6. Représentation de la valeur du D de Tajima moyen en fonction du Φ_{PhiST} . Representation of the average Tajima's D value against the Φ_{PhiST} .

Table III.4. Résultats du test HKA / Results of the HKA test***B. azoricus* Rb vs. *B. thermophilus* 7S**Sum of deviations: 3.9587 (df= 7, $P = 0.7845$)***B. azoricus* MG vs. *B. thermophilus* 7S**Sum of deviations: 7.7511 (df = 7; $P = 0.3550$)***B. puteoserpentis* vs. *B. thermophilus* 7S**Sum of deviations: 13.3811 (df = 7; $P = 0.0633$)**III.2.3.c. Analyses IMA / IMA analyses**

Repeated runs of the IMA program revealed unambiguous marginal posterior probability distributions of the parameters for the two actual species and the ancestral one. The peaks of the six parameter estimates were confined to fairly narrow ranges with corresponding credibility intervals illustrated in the Figure III.7. In order to convert parameter estimates in more easily interpreted units, we estimated substitution rates for each locus from the divergence between the species of *B. thermophilus* from 21°S and its sister-species from 37°S, under the hypothesis that deep-sea species have been separated during the formation of the Eastern Island microplate (at 25°S) about 5.9 millions years ago (Naar and Hey 1991). The Eastern Island microplate was created on the axis of the ancestral Pacific ridge, and is thus responsible of the split of the South Pacific hydrothermal fauna north and south of this island. A geometric mean of the 21°S-37°S DNA sequence divergence across loci was calculated and subsequently used to convert the time since separation parameter, t , to divergence in years. The mutation rate can be estimate as $r = D / 2T$, where D is the pairwise sequence divergence and T is the time of divergence, multiplied by 2 to account for the age of each lineage. The geometric divergence for the mitochondrial and the seven nuclear genes is 1.07%. This divergence gave a substitution rate of 0.09

% per MY. These values are in average twice as low as previous evolution rates estimated by (Chevaldonne et al. 2002) and (Johnson et al. 2006) for other hydrothermal invertebrates using the mtCOI gene. Moreover, this substitution rate is nearly equal to the rate of 0.10% per MYrs found by (Johnson et al. 2006) on a nuclear locus. Taking into account the length of each gene (**Table III.2**), we estimated the number of mutations per locus per year from 0.25×10^{-6} for ITS2, to 2.07×10^{-6} for the CO1). The geometric mean, $u = 0.63 \times 10^{-6}$ was used to rescale the IMA parameter estimates from the combined analysis. We used this value to directly convert the estimate t (the number of substitutions) to a number of years since population splitting.

To convert estimates of the population mutation rate parameters (θ_1 , θ_2 , and θ_A) to estimates of effective population size (N_1 , N_2 , and N_A , respectively), we estimated the generation time of *Bathymodiolus* from previous estimations of the mussel individual growth rate: 0.6 to 1.1cm per year (Rhoads et al. 1981, Comtet 1998) and the size at which individuals reach their sexual maturity (2 years: the minimum size observed for individuals able to reproduce is 19.7 cm for male and 31.9 cm for female (unpublished data from the BIOSPEEDO 2004 cruise). Calculations yielded an estimated number of mutations per generation per locus of 1.26×10^{-6} .

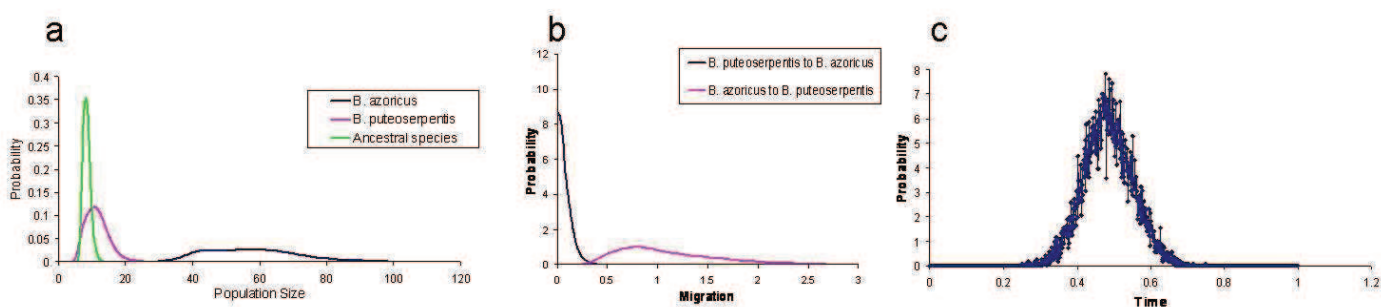


Figure III.7. Distribution des probabilités postérieures marginales des paramètres du modèle. The marginal posterior probability distribution for model parameters (scaled by the neutral mutation rate). Curves are shown for the analysis with and for all *B. azoricus* and *B. puteoserpentis* (g) (h) (i).

Table III.5. Estimation des valeurs de maximum de vraisemblance (MLE) et de l'intervalle des plus fortes valeurs de densité postérieure (HPD) des paramètres démographiques.

Maximum-Likelihood Estimates (MLE) and the 90% Highest Posterior Density (HPD) Intervals of Demographic Parameters

Comparison	θ_1	θ_2	θ_A	m_1	m_2	t	N_1	N_2	N_A	$M_1 = 2N_1m_1$	$M_2 = 2N_2m_2$	t (years)
Bazori x Bputeo												
MLE	57.936	10.829	8.224	0.001	0.792	0.477	11,497,577	2,149,008	1,632,117	0.014	4.291	758,090
lower 90% HPD	35.844	5.998	6.558	0.001	0.372	0.377	7,113,292	1,190,341	1,301,395			599,328
higher 90% HPD	78.225	16.545	10.326	0.165	1.827	0.600	15,523,971	3,283,425	2,049,147			953,368

Maximum-likelihood effective population size estimates are shown in Table III.5. The estimate of the *B. azoricus* population size was five times greater than that of *B. puteoserpentis* and the ancestral one, which suggests population expansion in *B. azoricus*. The marginal posterior probability distribution of the divergence parameter, t , revealed a sharp peak at 0.477 (Figure III.7), which converted into a divergence time of 1.032 MYA with the 90% HPD interval ranged from 0.599 to 0.953 MYA (Table III.5). The multi-locus estimates of migration revealed an asymmetric gene flow from *B. azoricus* to *B. puteoserpentis* with a maximum around ($2Nm = 4.291$) (Figure III.7, Table III.5), while the reverse flow from *B. puteoserpentis* to *B. azoricus* did not appear to significantly depart from zero.

Although the multi-locus estimates benefited from the statistical power of the full information contained in the eight loci

the single locus analysis of population differentiation (Table III.2) suggested that gene-flow could vary between genome regions as expected in the model of semi-permeable barriers to gene flow (Harrison 1990). A monolocus analysis of divergence with gene-flow under the IM model was therefore conducted and results are shown in Figure III.8 as a series of single-locus probability distributions. Estimating demographic parameters from single locus result in a lack of precision. Often likelihood surfaces were flat or an asymptotic plateau was obtained. However, some useful trend can nonetheless be gained from this analysis. SAHH, GF1_B, ITS2 and EF1 α displayed an asymmetric gene flow from *B. azoricus* to *B. puteoserpentis*, whereas GF1_A and COL1 were not informative. However, one locus, LYSO, showed an opposite asymmetry to the genomic trend (i.e. from *B. puteoserpentis* to *B. azoricus*), which still represents one locus on the six

informative ones (17%). In addition, despite the strong stochastic variance, differences between extreme values showed up. For instance, loci ITS2 and EF1 α showed distinguishable probability distribution to locus CO1, with the latter suggesting evidence of gene flow while the

former did not significantly depart from a no-migration model. Divergence times estimated with single locus tended to be congruent with what was found from the multilocus analysis although probability distributions were often very flat.

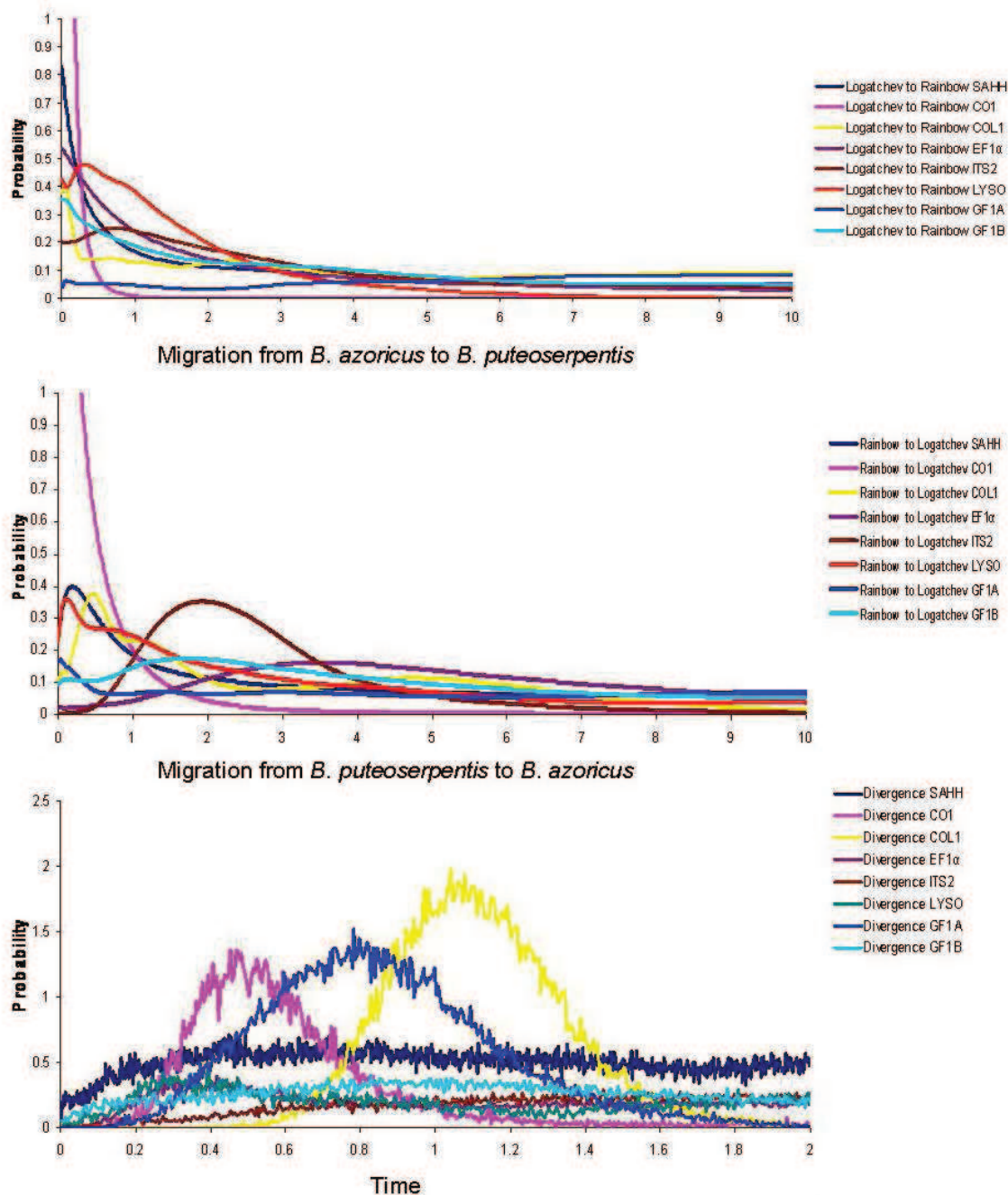


Figure III.8. Analyses IMA mono locus pour chacun des 8 gènes. / IMA single locus analyses for each of the 8 genes.

III.2.3.d. Preuve d'un balayage sélectif au locus SAHH / Evidence of a selective sweep at the SAHH locus

SAHH is the only nuclear locus exhibiting a strong differentiation between the two species and a high $d/D_{ATL/PAC}$ ratio. The hypothesis that selection could have played a role in the evolution of this locus was therefore hypothesised. Although Tajima's D and HKA tests were not significant, the Fay and Wu's H test which is designed to be sensible to an excess of high frequency derived variants was significantly different from zero ($H=-8.52$,

$P=0.04$). The SAHH data were compatible with the infinite mutation model and the single gene tree is reconstructed in Figure III.9. The gene tree shows the partially star-shaped genealogy observed at this locus which is characteristic of selective sweeps (Fay and Wu 2000) We therefore used the coalescence-based maximum-likelihood method of Galtier et al. (Galtier, Depaulis, and Barton 2000). Results of the runs indicated that the model with a selective sweep was significantly better supported than the model without diversity reducing event ($\theta=13.2$, $S=0.7$, $T=0.22$, Likelihood ratio tests: $P=0.006$).

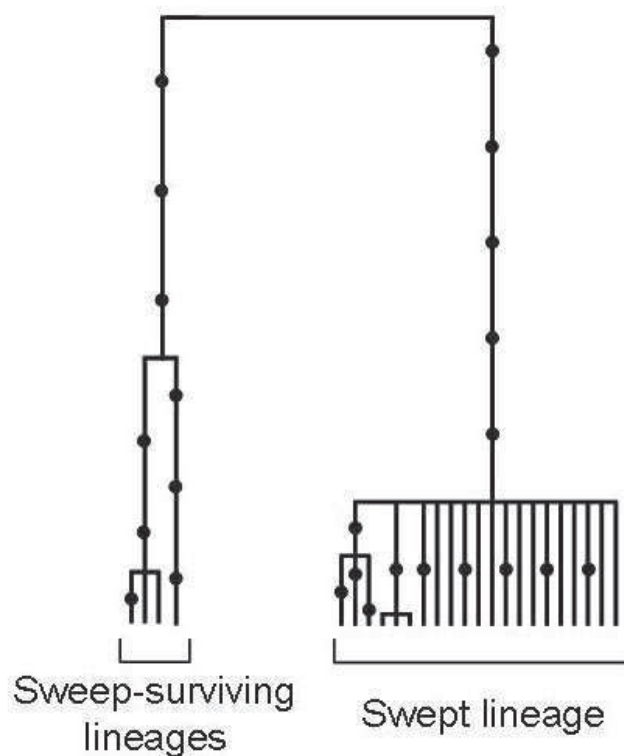


Figure III.9. Reconstruction de la généalogie de la SAHH / Reconstructed genealogy of the SAHH locus.

III.2.3.e. Preuve d'un contact secondaire avec introgression à partir de l'observation d'allèles recombinés inter-clade au locus SAHH / Evidence for secondary introgression from interclade recombinant alleles at the GF1B locus

The GF1_B locus displayed positive Tajima's D values (Table III.2), a fast rate of evolution in the Atlantic species

when compared to the divergence to the outgroup *B. thermophilus* (Table III.2), and evidence of recombination between divergent alleles. Although highly reticulated, a closer look at the phylogenetic network in Figure III.3 revealed two groups of alleles and intermediate sequences with small terminal branches. One expects ancestral lineages to exhibit long terminal branches and these

sequences seemed better explained by recombination. We conducted a detailed analysis of recombination the results of which are synthesised in the ancestral recombination graph presented in Figure III.10. Four distinct recombination events between well distinct and divergent alleles have been identified. The interesting result of this analysis was that the diversity

observed at this locus allowed to infer with little ambiguity that recombination between divergent alleles occurred recently after a long period during which recombination did not happened. This finding is in agreement with the hypothesis of secondary introgression of divergent alleles after a period of divergence in complete allopatry.

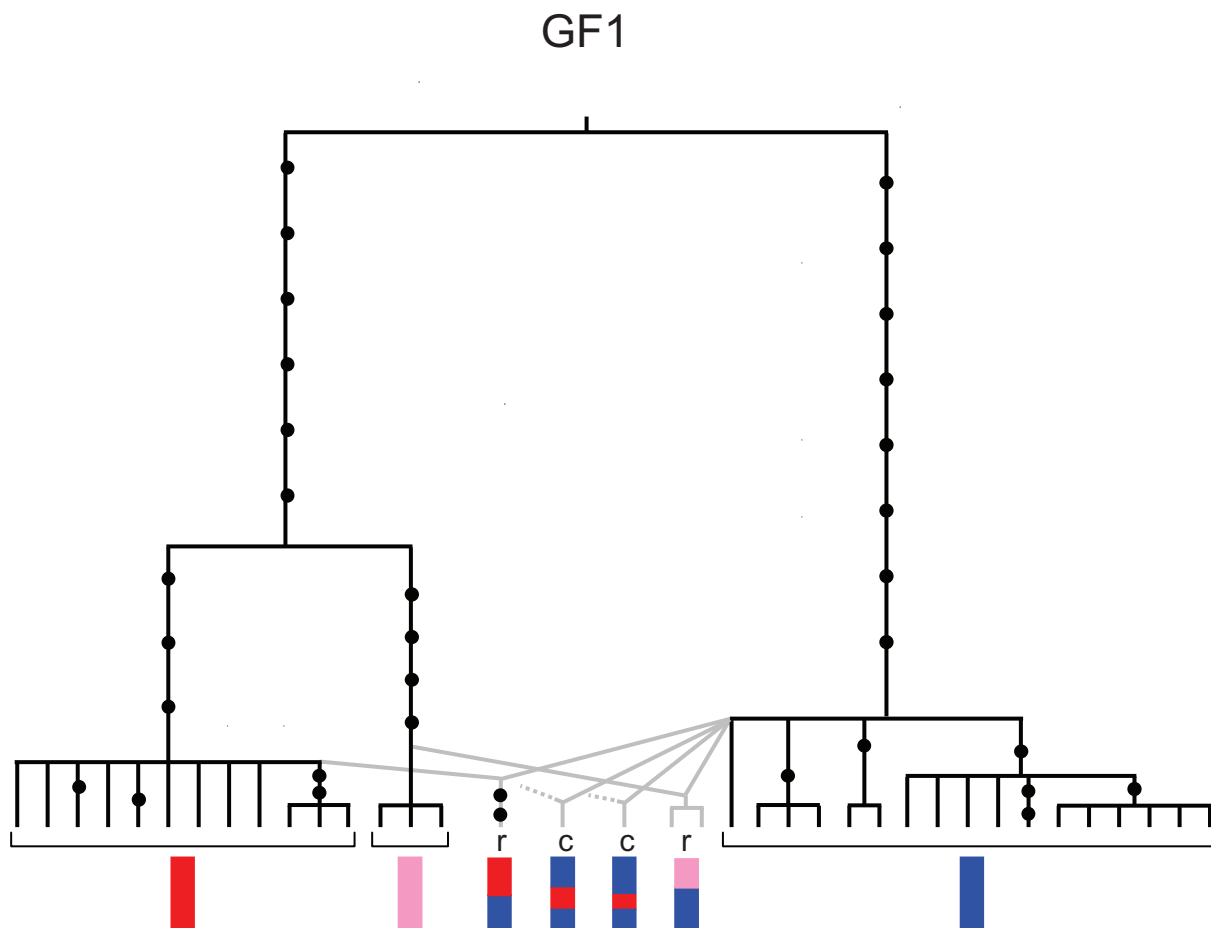


Figure III.10. Dédudition de l'histoire des alleles naturels recombinés du locus SAHH. / Predicted history of the GF1_B natural recombinant. Alleles 'r' are recombinant from alleles present in the populations; 'c' are recombinant alleles between a present and an unsample alleles.

III.2.3.f. Spectre des fréquences alléliques: recherche de l'empreinte de l'interruption des flux géniques dans l'histoire des genes. / *Allele frequency spectra: searching for the footprint of interrupted gene-flow in the gene history*

As the GF1B locus provided compelling evidence of secondary

introgression, we searched for the footprint of interrupted gene flow at other loci. The idea was that intragenic recombination can erase the divergence produced by a period of allopatry as it was partly the case for the GF1B locus. Substitutions that were fixed during allopatry would then be dispatched among chromosomes by recombination but if introgression was balanced enough and the number of substitutions was high

enough, an excess of mutations in intermediate frequency should be observed. In Figure III.11 is presented the distribution in the two Atlantic species of the average frequency of the derived variants (as oriented by the outgroup). Logically, the CO1 and SAHH loci showed an excess of mutations with an intermediate average frequency which are indeed the fixed differences between the two species. The GF1B locus also

displayed a clear excess of mutations with an intermediate average frequency. For the other five genes, the main effect was an excess of rare variants which can be explained by demography or a slightly deleterious effect of polymorphic mutations. However, once the excess of rare variants was accounted for, an excess of mutations with an intermediate frequency was still detected at the locus LYSO (Fisher's exact test: $p=0.01$).

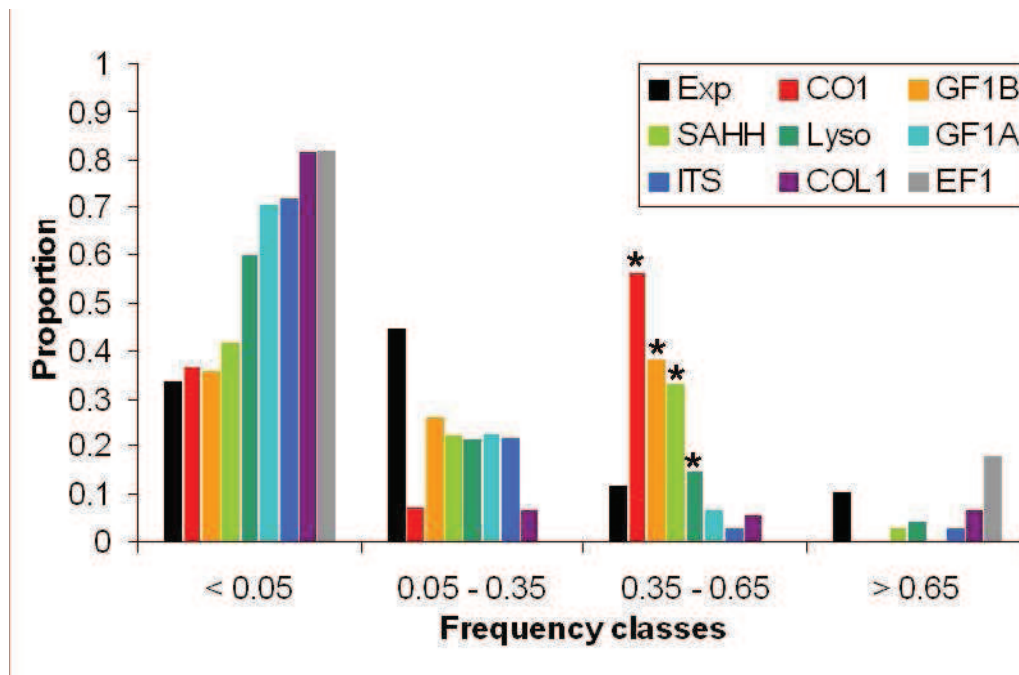


Figure III.11. Distribution des fréquences moyennes des variants dérivés des deux espèces atlantiques. / Distribution of the average frequency of the derived variant for the two Atlantic species

III.2.4. Discussion / Discussion

The two species of *Bathymodiolus* discovered along the Mid-Atlantic Ridge (Ségonzac 1992; Van Dover et al. 1996) were first described from their morphologies and erected to the species names of *B. azoricus* and *B. puteoserpentis* by (Von Cosel, Métivier, and Hashimoto 1994). O'Mullan et al. 2001 and Won et al. 2003a, however, reported that these two species were able to hybridise at the Broken Spur vent field (29°10'N/43°10'W; 3350m depth) situated between the two species' ranges. These authors have interpreted local hybridisation between the two Atlantic species as the result of an

accidental secondary contact between two divergent species (Won et al. 2003a) which is the traditional interpretation in the hybrid zone framework (Barton and Hewitt 1985). However, it is also well-known in the hybrid zone literature that it is extremely difficult to tell apart secondary contact zone from parapatric primary differentiation (Endler 1977) and the former hypothesis was preferred because the latter is more unlikely to happen. Recently, a new burst of data relevant to this question lead to a re-evaluation of the easiness of speciation with gene flow (Nosil 2008). In the case of *B. azoricus* and *B. puteoserpentis*, the geographic context is not easily reconcilable with the

opportunity for a secondary contact as the Northern Mid-Atlantic Ridge is a dead-end limited to the north by Iceland. This is why our eight-locus analysis of DNA sequence polymorphism was aimed to reconstruct the history of divergence and gene flow between these two closely-related organisms in the original environment of hydrothermal vents, in order to attempt to accumulate evidence for one of the two scenarios, secondary contact or parapatric speciation. Our data are in better agreement with the former one.

III.2.4.a. Déduction à partir du modèle d'isolement avec flux géniques / Inferences from the isolation with gene-flow model

As recently done for an increasing number of pair of closely related species (Hey 2006, Niemiller, Fitzpatrick, and Miller 2008, Nosil 2008), the multi-locus fit to an isolation with gene-flow model produced useful inferences to understand how isolation happened between *B. azoricus* and *B. puteoserpentis*. First, divergence is estimated to date back to half a million years. Secondly, we estimated that the effective population size of *B. azoricus* was 2-5 times greater than that of *B. puteoserpentis* and the ancestral population suggesting population expansions in *B. azoricus*. This finding is in agreement with census population sizes observed for both species –dense mussel beds in the case of *B. azoricus* (Comtet and Desbruyères 1998) versus scattered individuals in the case of *B. puteoserpentis* (Gebruk et al. 2000)-. Finally, evidence for gene-flow was unambiguous although mainly asymmetrical from the *B. azoricus* to *B. puteoserpentis*. The asymmetry of gene flow can have several explanations. (i) Hybrid zone theory predicts asymmetry of a genetic barrier to gene flow according to the density of populations (Barton 1986). Both population genetics inferences and observations in the field are in accordance for a denser and bigger

population of *B. azoricus*, which would agree well with the asymmetry of the barrier to gene flow. (ii) Hybrid zone theory also predicts that moving hybrids zones are expected to leave a footprint in the form of asymmetrical introgression (Barton and Hewitt 1985, Buggs 2007). The inference of an asymmetric gene flow could therefore also be in accordance with a northward displacement of the hybrid zone. The recent review of Buggs (2007) indeed suggests that hybrid zone movements could be more frequent than previously thought. (iii) Finally, asymmetric gene flow could be a consequence of an asymmetry of hybridisation. Distinguishing among these various explanations is difficult and would require experiments in the lab hardly achievable with *Bathymodiolus* mussels.

III.2.4.b. L'histoire du flux génique: quelles preuves pour une période d'interruptions de flux? / The history of gene-flow: what evidence for a period of no gene-flow?

When we undertook the analysis of DNA sequences in *Bathymodiolus* we had in mind that we might be able to distinguish secondary contact from parapatric speciation. However, it is not that simple and although evidence of gene flow was straightforward, it is unclear whether it means genes were continuously exchanged during the whole divergence period or whether it remains compatible with a period without gene flow before secondary contact. One idea is to closely look at the branching of easily identified introgressed alleles which are visible when divergence is strong enough and gene flow low enough. This was the case for loci SAHH and LYSO which exhibited fairly clear examples of introgression of *B. azoricus* alleles in the *B. puteoserpentis* background. We would like here to emphasise that this observation cannot be the result of the sampling of rare *B. azoricus* individual migrants in a *B.*

puteoserpentis population because we sampled far away from the identified hybrid zone (O'Mullan et al. 2001) and none of the introgressed individuals displayed a hybrid signature. At the SAHH locus, we identified two alleles which are likely results of introgression (see the two arrays in Figure III.3), one (labelled 1 in Figure III.3) coalesced at the tip of the network and was in accordance with recent secondary introgression while the other (labelled 2 in Figure III.3) coalesced more deeply and could well indicate ancient gene flow. At the LYSO locus, both kind of branching were also observed.

Another useful information, although less studied (but see Machado et al. 2002), is intragenic recombination, because recombination can only occur between alleles that are segregating together within the same populations and genomes. The reconstructed ancestral recombination graph of the GFB locus (Figure III.10) allowed us to identify the presence of recent recombinant alleles between two divergent clades of alleles. The recent take again of recombination at the GF1B locus is a strong argument for the hypothesis of secondary introgression. However, it also suggests that secondary introgression operates for a sufficient period for newly formed recombinants to reach a high enough frequency to be sampled.

Secondary contact and subsequent introgression therefore explain our results well. Because no species can enter the Atlantic ridge further north of Iceland, the only scenario of the MAR colonisation that allows the formation of a secondary contact zone is a first colonisation by the species *B. azoricus*. Its geographic distribution was then challenged by *B. puteoserpentis* located further south. Such a scenario implies that the Mid Atlantic Ridge would have been independently colonised twice. Because both vent species seem to be related to the *B. boomerang* complex of species (sister group) located at cold seeps from the deep-sea Caribbean region (Olu-Le Roy et al. 2007),

geographic isolates originated from either Barbados accretion prism or the Florida escarpment were likely to colonise the rift separately and met subsequently to hybridise.

III.2.4.c. Impact de la sélection sur les processus de différenciation / The impact of selection on the differentiation process

The effect of a possible selective sweep at the most discriminant nuclear locus SAHH has been detected. In addition, negative and significant value of the Tajima's D was found at the mitochondrial gene COI which remains the only locus differentially fixed between the two species. Recurrent selective sweeps are indeed likely to rejuvenate the coalescence time of the mitochondrial genome in species with big populations such as marine mollusks (Bazin, Glémin, and Galtier 2006). One can therefore argue that selection helped to accelerate the evolution toward monophyly at these two loci, explaining their status of diagnostic markers. However, at least two other loci, LYSO and GF1B, also registered a strong divergence between two groups of alleles that are interpreted to have introgressed secondarily after a stage of reciprocal monophyly. Therefore, SAHH and COI also need to be in genomic regions influenced by isolation genes responsible for the genetic barrier to gene-flow. One can then hypothesise that the sweep detected was the consequence of positive selection at a speciation gene (Orr, Masly, and Presgraves 2004). Some other ecological elements like the symbiont composition of the *Bathymodiolus* gills can also act as selective constraints. Both species indeed yield two kinds of symbionts, one being a sulfo-oxidizing bacteria and the other a methanogen coccus. Both symbionts co-exist in the mussel gills (Distel, Lee, and Cavanaugh 1995, Fiala-Medioni et al. 2002) but can be easily lost depending on the fluid

composition (Kadar et al. 2005) or selectively acquired by each species in the environment (Won et al. 2003b).

Another possible effect of selection we already discussed in a previous paper is slight purifying selection on polymorphic mutations. We suspected such a deleterious effect of polymorphic mutations at the EF1 α locus in *Bathymodiolus* (Faure et al. 2007). Purifying selection at some locus can inflate the contrast observed between loci with some having registered well the period of allopatry with multiple fixations (e.g. COI, SAHH, LYSO) while other loci having evolved two slowly (e.g. EF1 α , ITS2). However, despite low evolutionary rate, the species structure is still observable and alleles segregate into two distinct clades. For some other loci (COL1, GF1A), the situation seems more complex. Other confounding effects can be a high rate of intragenic recombination that artificially makes younger the divergence (Schierup and Hein 2000) such as observed at GF1A, or the complete replacement of one lineage by introgression. Such processes were reported in *Drosophila* to

explain complex genealogy situations between *D. permilis*, *D. pseudoobscura* and *D. p. bogotana* (Hilton, Kliman, and Hey 1994; Stephan et al. 1998).

III.2.5. Conclusion / conclusion

The two closely-related species of *Bathymodiolus* from the Mid-Atlantic Ridge were known to hybridise locally within the narrow tension zone of Broken Spur (O'Mullan et al. 2001; Won et al. 2003a). The analysis of populations located at both extremities of the mussel range by using a multilocus approach indicated that the divergence started half a million years ago and that gene flow occurs although mainly asymmetrically from *B. azoricus* to *B. puteoserpentis*. Finally, we obtained some evidence that gene-flow is due to secondary introgression after a period of allopatry, mainly because intragenic recombination at one locus, GF1B, seems to have recently show on again after a period during which it was blocked.

Remerciements / Acknowledgments

Authors gratefully acknowledge the captain and crew of the R/V “L’Atalante”, R/V Le Pourquoi Pas?”, and “R/V Atlantis”; and the pilots and co-pilots of the ROV Victor 6000, “Nautilie”, and “Alvin” teams for collecting material. We are particularly grateful to Dr. P-M Sarradin, (chief scientist of the ATOS 2001 cruise), to Dr. Y. Fouquet (chief scientist of the Serpentine 2007 cruise) and Dr. Robert Vrijenhoek and Dr. Cindy Van Dover (from the Eastern Microplate Expedition) who allowed us to conduct this work. Special thanks to Dr Nicole Dubilier who provide us the 37°S samples. We are also truly indebted to the sequencing genomic platforms GENOMER (Roscoff) and GENOSCOPE (Evry) for sequencing our clone collection. This work was partly funded by the French national programme GIS “Génomique Marine” via the “Bivalvomix” project (coord N. Bierne) and Marine Genomics Europe NoE (Fish & Shellfish node).

Chapitre IV. Structure génétique des populations de *Bathymodiolus thermophilus* et mise en évidence d'une zone d'hybridation sur la dorsale est-pacifique

IV.1. Introduction

Comme nous venons de le voir dans le chapitre précédent, les études visant à comprendre les mécanismes de spéciation chez *Bathymodiolus* ont majoritairement été menées sur les deux espèces présentes sur la dorsale médio-atlantique (Jollivet et al. 1998, O'Mullan et al. 2001, Won et al. 2003a). Ces études ont mis en évidence que les espèces proches *B. azoricus* et *B. puteoserpentis* étaient capables de s'hybrider et d'échanger des allèles. De plus, les niveaux d'introgession sont parfois très importants, laissant penser que la mise en place de la zone de contact est relativement ancienne.

Les études de génétique concernant les populations de bathymodiales de la dorsale est-pacifique sont plus anciennes (Grassle 1985, Craddock et al. 1995b, Won et al. 2003b) et sont principalement focalisées sur l'étude des flux de gènes et des barrières à la dispersion de l'espèce pacifique *B. thermophilus* en utilisant des marqueurs allozymes ou mitochondriaux.

A l'inverse de la dorsale médio-atlantique, la dorsale est-pacifique (ou EPR East Pacific Rise) constitue une zone où l'activité hydrothermale est relativement continue et où les failles transformantes limitant les segments sont beaucoup moins marquées. En général, la distance inter-sites ou inter-champs est relativement faible (Chevaldonné et al. 1997) laissant présumer des échanges beaucoup plus importants entre populations (Chevaldonné et al. 1997, Jollivet et al. 1995, Karl et al. 1996, Black et al. 1998). La dorsale est-pacifique présente également des caractéristiques géologiques particulières ; en effet cette dorsale est considérée comme ultra rapide (jusqu'à 16 cm d'accroissement par an) contrairement à la dorsale médio-atlantique (Demets et al. 1990, Naar & Hey 1991). La vitesse d'expansion élevée de l'EPR génère un décalage de l'activité hydrothermale et induit une instabilité environnementale

(renouvellement rapide des sites actifs) qui a un impact direct sur les peuplements hydrothermaux (Baker 1995) en causant des extinctions massives lors de phases éruptives avec épanchements de laves (Haymon et al. 1991). Au cours du temps, ces déplacements d'habitats peuvent favoriser les échanges entre champs éloignés ('bursts' de colonisation) et induisent une récurrence d'extinctions/recolonisations (Jollivet et al. 1999). A l'inverse, les champs hydrothermaux atlantiques sont séparés par de grandes distances (> 1000 km), par des failles transformantes prononcées et caractérisées par une quasi-absence de déplacement d'activité le long de la dorsale générant une activité sur site de longue durée (> 100 000 ans : Lalou et al. (1993a), Lalou et al. (1993b)) Il est donc supposé que les populations étudiées sur la dorsale du Pacifique Est sont plus sujettes aux effets démographiques que les populations de la dorsale atlantique. Au niveau de l'EPR elle-même, le taux d'expansion est variable en fonction de la latitude : 12cm/an à l'équateur jusqu'à 16cm/an aux alentours de 30°S (Naar & Hey 1991). Cette variabilité engendre donc des pressions environnementales différentes, en termes d'instabilité tectonique, entre les populations du nord et du sud de la dorsale.

Il convient donc de savoir si ce régime d'activité, très différent de celui observé sur la dorsale médio-atlantique peut avoir un impact différent sur le mode de spéciation des espèces hydrothermales. Les études précédentes ont permis de déterminer que les bathymodiales pouvaient disperser leurs propagules sur de grandes distances et que les seules populations nettement différenciées sur la dorsale étaient celles situées au sud de la microplaque de l'Ile de Pâques, à 31 et 32°S de latitude. Ces populations du sud présentent une divergence élevée (4,4% sur le marqueur mitochondrial COI) avec celle situées au nord de la dorsale (Won et al. 2003b). Ce niveau de divergence est similaire à celui observé entre les deux espèces proches de la dorsale médio-atlantique, laissant présager que l'espèce située au sud de l'Ile de Pâques pourrait constituer une espèce à part entière. La séparation des faunes hydrothermales au cours de la mise en place de la microplaque de l'île de Pâques, il y a environ 5 millions d'années (Won et al. 2003b, Hurtado et al. 2004) expliquerait facilement une longue période d'isolement géographique à l'origine de la spéciation. Cependant, est-il possible que ce type de spéciation par vicariance soit le plus fréquent le long des dorsales océaniques même lorsque l'échelle spatiale est relativement restreinte ? Une étude récente a montré qu'un épisode de spéciation allopatrique avec formation d'une zone de contact secondaire directement lié à l'histoire tectonique de la dorsale avait eu lieu chez le gastéropode *Lepetodrilus elevatus* présent sur les moulières de l'EPR (Matabos et al. 2008).

Il est donc vraisemblable qu'il puisse exister des zones de contact secondaire entre des espèces proches de la dorsale est-pacifique et que cela se traduise par l'existence d'une zone d'hybridation si les espèces sont suffisamment proches, comme cela a été observé entre *B. azoricus* et *B. puteoserpentis* le long de la dorsale médio-atlantique (cf chapitre précédent). Une telle hybridation n'a jamais été détectée chez *B. thermophilus* avec les outils moléculaires utilisés dans les précédentes analyses (locus mitochondrial COI et allozymes) mais il faut souligner que seul un petit nombre de populations a été étudié au sud de l'EPR, où une zone d'hybridation pourrait exister. La comparaison des populations de *Bathymodiolus* provenant de différents sites répartis le long de l'EPR sud permettrait de détecter d'éventuelles zones de contact avec cette faune vicariante située plus au sud sur la dorsale pacifique-antarctique et de quantifier le niveau d'hybridation si celui-ci existe. La zone de l'EPR sud comprise entre 7°25S et 21°3S constitue en effet une zone de transition biogéographique entre les faunes hydrothermales du Pacifique Est et du Pacifique Ouest (Bachraty et al. 2008).

Ces observations générales concernant la dorsale est-pacifique et ses populations de *Bathymodiolus thermophilus* nous ont amené à nous poser quatre questions essentielles, auxquelles nous allons répondre dans ce chapitre :

_ Les populations de bathymodiales situées au sud de la dorsale (comprise entre 7°25S et 21°33S/EPR) sont elles génétiquement structurées ?

_ Y a-t-il la trace d'une divergence allopatrique entre les populations de bathymodiales situées de part et d'autre de la microplaque de l'île de Pâques ?

_ Peut-on détecter de l'hybridation entre ces 'espèces' géographiques et est-elle limitée au sud de la dorsale ?

_ Les données génétiques permettent-elles d'inférer des scénarios de colonisation de la dorsale sud EPR ?

Une première approche utilisant des marqueurs allozymiques porte sur sept populations situées le long de la dorsale entre 7°25S et 21°33S. Elle a pour but de mettre en évidence la structure génétique à fine échelle des populations de bathymodiales le long de la dorsale. Une seconde approche utilisant des données de séquences de gènes nucléaires et mitochondriaux, est réalisée sur les populations les plus extrêmes de 7°25S et de 21°33S afin de déterminer si la séparation avec l'espèce proche *B. aff. thermophilus* présente au sud de l'île de Pâques est totale et permet d'expliquer une éventuelle structure géographique. La

population de 7°25S, éloignée de la zone de contact présumée, devrait permettre de mesurer l'étendue de l'introggression si l'isolement entre les espèces n'est pas total. L'estimation de la différenciation génétique entre les populations les plus éloignées de l'EPR sud à partir du polymorphisme de séquences de plusieurs gènes devrait également permettre de confirmer la présence d'une éventuelle structure géographique et d'obtenir des informations précieuses sur l'histoire de la colonisation de la dorsale.

IV.2. Matériel et méthodes

IV.2.1. Echantillons biologiques

Les échantillons de l'espèce *Bathymodiolus thermophilus* ont été récoltés pendant les campagnes océanographiques HOT96 à 9°50N/EPR et Biospeedo 2004 sur la portion sud de la dorsale du Pacifique Est à l'aide de la pince télémanipulateur (à godet) du Nautilite sur les sites Sarah's Spring (7°25S), Lucky Eric (13°50S), Oasis (17°25S), Rehu Marka (17°25S), Suzie (17°34S), Animal Farm (18°33S), et Gromit (21°33N ; cf. Tableau IV.1). Les échantillons de l'espèce proche *B. aff. thermophilus* ont été prélevés pendant la campagne américaine Eastern Microplate Expedition 2005 à l'aide du bathyscaphe Alvin au niveau du site Sebastian's Steamer (37°S) et fournis grâce à l'aimable assistance de N. Dubilier (MPI Department of Microbiology). Plus de 1700 km séparent les deux sites les plus éloignés de l'espèce *B. thermophilus* (7°25S et 21°33S), l'espèce *B. aff. thermophilus*, récoltée sur le site 37°S à 1830 km au sud du site 21°33S est quant à elle distante de près de 3400 km du champ 7°25S.

Tableau IV.1. Echantillons récoltés sur la dorsale du Pacifique Est (EPR).

Localité	Latitude / Longitude	espèce	profondeur (m)	Echantillonnage	
				Allozymes	Séquences
Qvent/BioVent	9°50'N / 104°17' W	<i>B. thermophilus</i>	2530	0	45
Sarah's Spring	7°25'S / 107°47' W	<i>B. thermophilus</i>	2735	11	48
Lucky Eric	13°59'S / 112°29' W	<i>B. thermophilus</i>	2623	39	0
Oasis	17°25'S / 113°12' W	<i>B. thermophilus</i>	2575	78	0
Rehu Marka	17°25'S / 113°12' W	<i>B. thermophilus</i>	2575	84	0
Suzie	17°35'S / 113°15' W	<i>B. thermophilus</i>	2590	24	0
Animal Farm	18°36'S / 113°24' W	<i>B. thermophilus</i>	2636	43	0
Gromit	21°33'S / 114°18' W	<i>B. thermophilus</i>	2800	126	27
Sebastian's Steamer	37°47'S / 110°55' W	<i>B. aff. thermophilus</i>	2216	0	2

IV.2.2. Extraction des protéines et visualisation sur gel d'amidon

Les échantillons utilisés pour le génotypage allozymique ont été immédiatement congelés à bord après dissection et conservés dans l'azote liquide jusqu'à leur extraction protéique. Des morceaux de tissu provenant du pied et du manteau ont également été placés dans des tubes d'éthanol absolu afin de réaliser les extractions d'ADN.

Les extractions protéiques ont été réalisées à partir d'un broyage Ultra-Turax® sur de la glace dans une solution d'extraction composée de TRIS (0,01M) ; EDTA (2mM) ; NADP (0,5mM), β -Mercaptoethanol (0,05%), de PMSF (0,1 mM) et de sucrose (0,25M) suivant les indications de Pasteur et al. (1987). Les extraits protéiques bruts sont utilisés pour imbiber des papiers Whatman n°3 qui sont eux-mêmes placés dans un gel d'amidon à 12% contenant un tampon Tris-Citrate et soumis à une électrophorèse à 80 mA (100-250 W). Deux systèmes de tampons d'électrophorèse ont été utilisés : le système continu TC8 (Tris-Citrate pH 8,0) et le système discontinu TC6,7 (Tris-Citrate pH 6,3/6,7 pour les tampons d'électrophorèse et de gel, respectivement). Après migration électrophorétique, les gels sont découpés en quatre tranches pour révéler les systèmes enzymatiques à partir des méthodes histo-chimiques présentées dans le manuel d'électrophorèse des protéines de Pasteur et al. (1987). Neuf locus polymorphes ont été sélectionnés : l'Aconitase (*AcoI*), la Glucose-6-Phosphate-Isomerase (*Gpi*), l'Hexokinase (*Hk*), la Leucine-amino-peptidase cytosolique (*Lap*), les locus cytosolique et mitochondrial de la Malate Déshydrogénase (*Mdh1* et *Mdh2*), la Mannose-phosphate-isomérase (*Mpi*), l'Octopine Déshydrogénase (*Opdh*), et la Phosphoglucomutase (*Pgm*)⁵. Ces locus ont déjà été utilisés dans de précédentes études sur les Mytilidae (*Lap*, *Opdh*, *Pgm*, *Gpi*) (Gardner & Skibinski 1990), (*Gpi*, *HK*, *Lap*, *Mdh*, *Pgm*) (Blot et al. 1987), et chez les Bathymodiolinae (*Aco*, *Gpi*, *Lap*, *Mdh*, *Mpi*, *Opdh*) (Craddock et al. 1995a). Les locus sont numérotés de l'anode vers la cathode et les allèles sont numérotés en fonction de leur mobilité relative par rapport à l'allèle le plus fréquent chez *B. azoricus* (espèce Atlantique).

⁵ Le génotypage allozymique a été réalisé par Dominique le Guen à la Station Biologique de Roscoff.

IV.2.3. Séquençage des locus nucléaires et mitochondrial

Pour obtenir les données de séquences, 7 gènes ont été amplifiés à l'aide d'amorces spécifiques ; COI ; EF1 α , SAHH, LYSO, GF1-A et GF1-B, et ITS2 (cf tableau 1 du chapitre Matériels et Méthodes). Les techniques d'amplification et clonage selon la méthode de clonage par marquage recapture (Bierne et al. 2007) et séquençage sont détaillées dans le chapitre « Matériels et Méthodes ».

IV.2.4. Descripteurs génétiques associés à l'analyse en fréquence des allozymes

Pour chaque population, les fréquences alléliques, le nombre moyen d'allèles par locus (N_{all}) ; l'hétérozygotie observée (H_o) et l'hétérozygotie attendue (H_{NB}) ont été estimés avec le logiciel Genetix (Belkhir et al. 2004) d'après les formules suivantes :

Le nombre moyen d'allèles :

$$N_{all} = \frac{1}{n} \sum_{i=1}^n a_i$$

avec 'n' le nombre de locus et a_i le nombre d'allèles par locus.

L'hétérozygotie observée (fréquence moyenne des hétérozygotes dans l'échantillon) par locus :

$$H_o = 1 - \sum_{i=1}^k X_{ii}$$

avec X_{ii} , la fréquence des homozygotes pour l'allèle i .

L'hétérozygotie attendue (fréquence moyenne des hétérozygotes attendue sous l'équilibre de Hardy-Weinberg et corrigée selon Nei (Nei 1987) pour éviter d'induire un biais d'échantillonnage par locus:

$$H_{nb} = \frac{2N(1 - \sum_{i=1}^k x_{i^2})}{(2N - 1)}$$

avec x_i^2 ; la fréquence attendue pour l'homozygote au $i^{\text{ème}}$ allèle avec k égal au nombre d'allèles dans un locus et N , le nombre d'individus dans l'échantillon.

Nous avons utilisé le logiciel Fstat (Goudet 2002), pour calculer les richesses alléliques. La richesse alléliques (R_s), est une mesure permettant de comparer les nombres d'allèles entre des échantillons de taille indépendante. Elle estime le nombre d'allèles attendus dans un échantillon de $2n$ gènes sachant que $2N$ gènes ont été échantillonnés (El Mousadik & Petit 1996).

Les statistiques de Wright (1965), ont été calculées à l'aide d'estimateurs reposant sur une analyse de variance des fréquences alléliques, selon les formules de Weir & Cockerham (1984). Ces statistiques ont également été calculées avec le logiciel Genetix⁶.

L'estimateur F_{is} , qui compare le nombre observé d'hétérozygotes et le nombre attendu d'hétérozygotes sous l'hypothèse de l'équilibre de Hardy-Weinberg est calculé pour chaque locus et pour l'ensemble des locus de chaque population selon la formule :

$$f = 1 - \frac{\sum (C_i)}{\sum (B_i + C_i)}$$

Où, pour le $i^{\text{ème}}$ allèle, B_i est la composante de la variance des fréquences alléliques entre individus à l'intérieur de chaque population; et C_i la composante de la variance des fréquences alléliques entre gamètes prises au sein d'un même individu.

L'estimateur F_{st} mesure la différenciation entre un échantillon de populations. Il est obtenu en mesurant l'écart entre le nombre attendu d'hétérozygotes pour l'ensemble des populations et le nombre attendu d'hétérozygotes dans chacune des populations (sous l'hypothèse de l'équilibre de Hardy-Weinberg). Il est calculé selon la formule suivante :

$$\hat{\theta} = \frac{\sum (A_i)}{\sum (A_i B_i C_i)}$$

Où pour $i^{\text{ème}}$ allèle, A_i est la composante inter-population de la variance des fréquences alléliques, B_i : la composante de la variance des fréquences alléliques entre individus à l'intérieur de chaque population; et C_i : composante de la variance des fréquences alléliques au sein de chaque individu.

⁶ Les formules utilisées sont issues de : <http://www.genetix.univ-montp2.fr/genetix/FST.htm>

Les valeurs de F_{is} et de F_{st} ont été testées par rapport à la valeur zéro (attendue pour l'hypothèse nulle de non-différenciation génétique des populations) avec un rééchantillonnage de 1000 permutations. Les tests par permutations consistent à comparer les valeurs observées aux valeurs d'une distribution théorique de l'estimateur générée par la permutation des allèles entre individus à l'intérieur des populations pour les F_{is} et par la permutation des génotypes entre les populations pour les F_{st} . Il est alors possible de déterminer la probabilité de s'écarter de cette distribution théorique. Celle-ci correspond à un test unilatéral avec $p = \text{Nb de classes} \geq \text{à la valeur observée} / \text{nombre total de permutations} + 1$.

IV.2.5. Descripteurs génétiques associés à l'analyse des séquences.

L'alignement, la correction des séquences et le traitement du polymorphisme nucléotidique et de la divergence entre espèces est décrite en détail dans le chapitre 'Matériels et Méthodes'.

Les indices de diversité nucléotidique θ_w (Watterson 1975), et θ_π (Nei 1987), le nombre de sites ségrégeants, les estimateurs D de Tajima (Tajima 1989a), le nombre minimum d'événements de recombinaisons (R_m) (Hudson & Kaplan 1985) et la distance absolue (D_{XY}) sont estimés à partir du jeu de donnée Haute Qualité (cf M&M) en utilisant le logiciel DnaSP, version 4.10 (Rozas et al. 2003). Les tests d'écarts à l'évolution neutre sont les mêmes que ceux utilisés précédemment.

IV.3. Résultats

IV.3.1. Structure génétique des populations de *Bathymodiolus thermophilus* au niveau de l'EPR sud.

Les modioles ont été échantillonnées sur l'ensemble des champs répertoriés sur l'EPR sud entre 7°25S et 21°33S. Le nombre d'individus échantillonnés varie de 11 (7°25S) à 126 (21°33S). Les données allozymiques ont été utilisées pour estimer la diversité génétique au sein de chaque population, d'estimer les déséquilibres de liaison entre locus et en hétérozygotes pour chaque population et le niveau de différenciation génétique entre les Sarah Spring, Gromit, Oasis, Animal Farm, Suzie, Rehu Marka et Lucky Eric. Les fréquences alléliques pour les 9 locus utilisés sont présentées dans le Tableau IV.2.

Tableau IV.2. Fréquences alléliques calculées à 9 locus enzymatiques et pour chacune de 7 populations. N : taille de l'échantillon. Surligné en rouge : allèles privés ; surligné en jaune : locus monomorphe.

		Sarah	LuckyEric	Oasis	RehuMar	Suzie	AnimalF	Gromit
Pgm haut	(N)	11	39	78	80	24	39	126
	64	-	-	-	0,006	-	-	-
	74	0,045	-	0,013	0,006	-	0,051	0,020
	78	-	-	-	-	-	-	0,004
	81	0,682	0,577	0,500	0,550	0,542	0,551	0,492
	85	-	-	0,013	-	-	0,013	0,036
	89	0,273	0,410	0,436	0,413	0,396	0,372	0,429
	95	-	0,013	0,039	0,025	0,063	0,013	0,020
Mpi	(N)	11	39	78	83	24	43	126
	100	0,046	0,039	0,026	0,006	0,042	-	0,008
	120	0,955	0,949	0,974	0,988	0,938	1,000	0,980
	140	-	0,013	-	-	0,021	-	0,012
	160	-	-	-	0,006	-	-	-
Odh	(N)	11	39	78	83	24	43	125
	60	-	0,064	0,006	0,006	-	0,023	0,008
	80	1,000	0,936	0,955	0,964	1,000	0,965	0,968
	100	-	-	0,039	0,030	-	0,012	0,024
Lap	(N)	11	39	77	80	22	42	126
	75	-	0,013	-	-	-	-	0,004
	82	0,909	0,731	0,714	0,744	0,750	0,798	0,698
	90	-	0,013	0,020	0,000	-	0,024	0,012
	100	0,091	0,231	0,253	0,225	0,227	0,143	0,262
	110	-	0,013	0,013	0,031	0,023	0,036	0,024
Gpi	(N)	11	39	77	84	24	42	112
	88	-	0,013	0,013	0,024	0,042	-	-
	100	-	-	-	-	-	-	0,005
	112	0,864	0,949	0,935	0,952	0,896	0,952	0,946
	116	0,046	-	-	-	-	0,012	-
	120	0,091	0,039	0,052	0,024	0,063	0,036	0,049
Mdh bas	(N)	11	39	78	84	24	43	126
	74	1,000	1,000	1,000	0,988	1,000	1,000	0,996
	100	-	-	-	0,012	-	-	0,004
Mdh haut	(N)	11	39	78	84	24	43	126
	100	-	0,167	0,039	0,107	0,021	0,221	0,083
	120	1,000	0,833	0,962	0,893	0,979	0,779	0,917
Hk	(N)	10	39	78	84	24	43	126
	70	0,050	-	-	0,006	0,021	-	-
	85	-	-	-	0,006	-	-	-
	100	0,900	1,000	1,000	0,988	0,979	1,000	1,000
	130	0,050	-	-	-	-	-	-
Aco	(N)	11	25	59	35	24	12	114
	80	0,136	0,220	0,170	0,271	0,188	0,375	0,158
	100	0,818	0,600	0,585	0,557	0,375	0,542	0,491
	117	-	-	0,059	-	0,083	-	0,018
	134	0,046	0,120	0,161	0,157	0,354	0,083	0,281
	147	-	0,060	0,025	0,014	-	-	0,053

De la même façon, le nombre d'allèles 'privés' est relativement faible avec un maximum de trois pour la population de Rehu Marka (*Pgm*, *Mpi*, *Hk*). Ces allèles privés sont toujours en faible fréquence (fréquence < 0,05) et toutes les populations possèdent des allèles rares présents dans au moins deux populations.

Les richesses alléliques ont été estimées pour chaque locus pour l'ensemble des populations. Les résultats sont consignés dans le Tableau IV.3.

Tableau IV.3. Valeur de richesse allélique (**Rs**) calculées par Fstat (Goudet 2002). Ces valeurs ont été estimées pour une taille minimum d'échantillons de 10 individus diploïdes. **Total pop** : valeur de Rs calculée pour le locus sur l'ensemble des populations

	Sarah's Spring 7°25	LuckyEric 13°59	Oasis 17°25	RehuMarka 17°25	Suzie 17°35	AnimalFarm 18°36	Gromit 21°33	total pop
Pgm haut	2,91	2,26	3,05	2,67	2,81	3,22	3,29	3,00
Mpi	1,91	1,85	1,43	1,24	2,08	1,00	1,37	1,42
Odh	1,00	1,78	1,70	1,60	1,00	1,65	1,55	1,61
Lap	2,00	2,77	2,59	2,49	2,45	2,96	2,69	2,62
Gpi	2,91	1,85	1,92	1,80	2,48	1,80	1,74	1,88
Mdh bas	1,00	1,00	1,00	1,23	1,00	1,00	1,08	1,07
Mdh haut	1,00	1,99	1,57	1,91	1,42	2,00	1,84	1,87
Hk	3,00	1,00	1,00	1,24	1,42	1,00	1,00	1,12
Aco	2,91	3,75	4,13	3,27	3,89	2,98	3,96	3,91

Les valeurs de Rs ont été utilisées dans la Figure IV.1 pour illustrer la variation de la richesse allélique en fonction de la latitude.

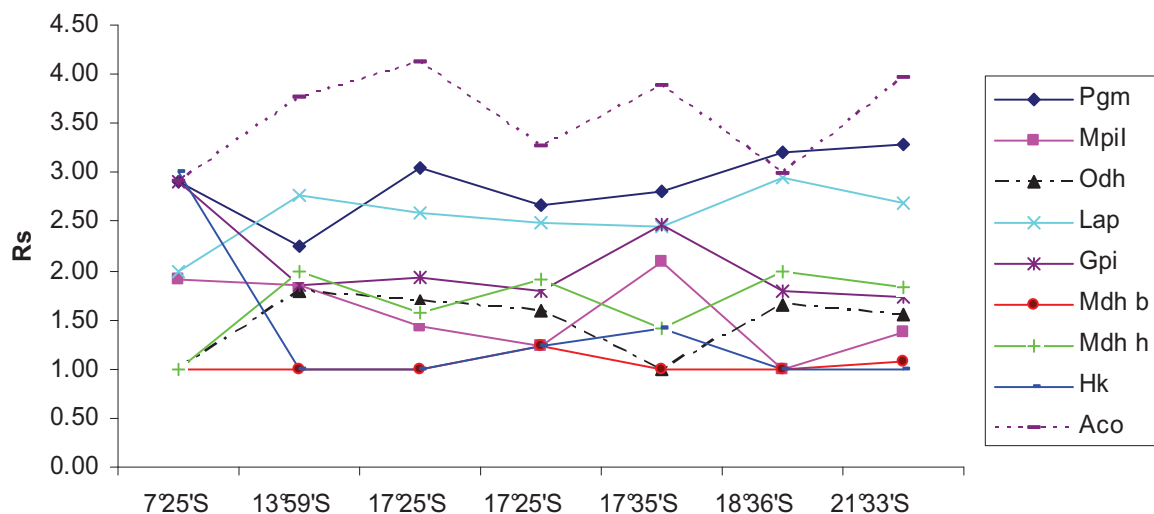


Figure IV.1. Richesse allélique représentée le long du gradient latitudinal des sites de l'étude.

Sur l'ensemble des sites, la diversité génétique de chaque locus est assez constante, en revanche, il existe une certaine variabilité de la richesse allélique selon les locus (MDHb : $R_s = 1,07$; ACO : $R_s = 3,91$).

Il n'est pas possible de distinguer de tendance pour l'évolution de la valeur de R_s en fonction de la latitude.

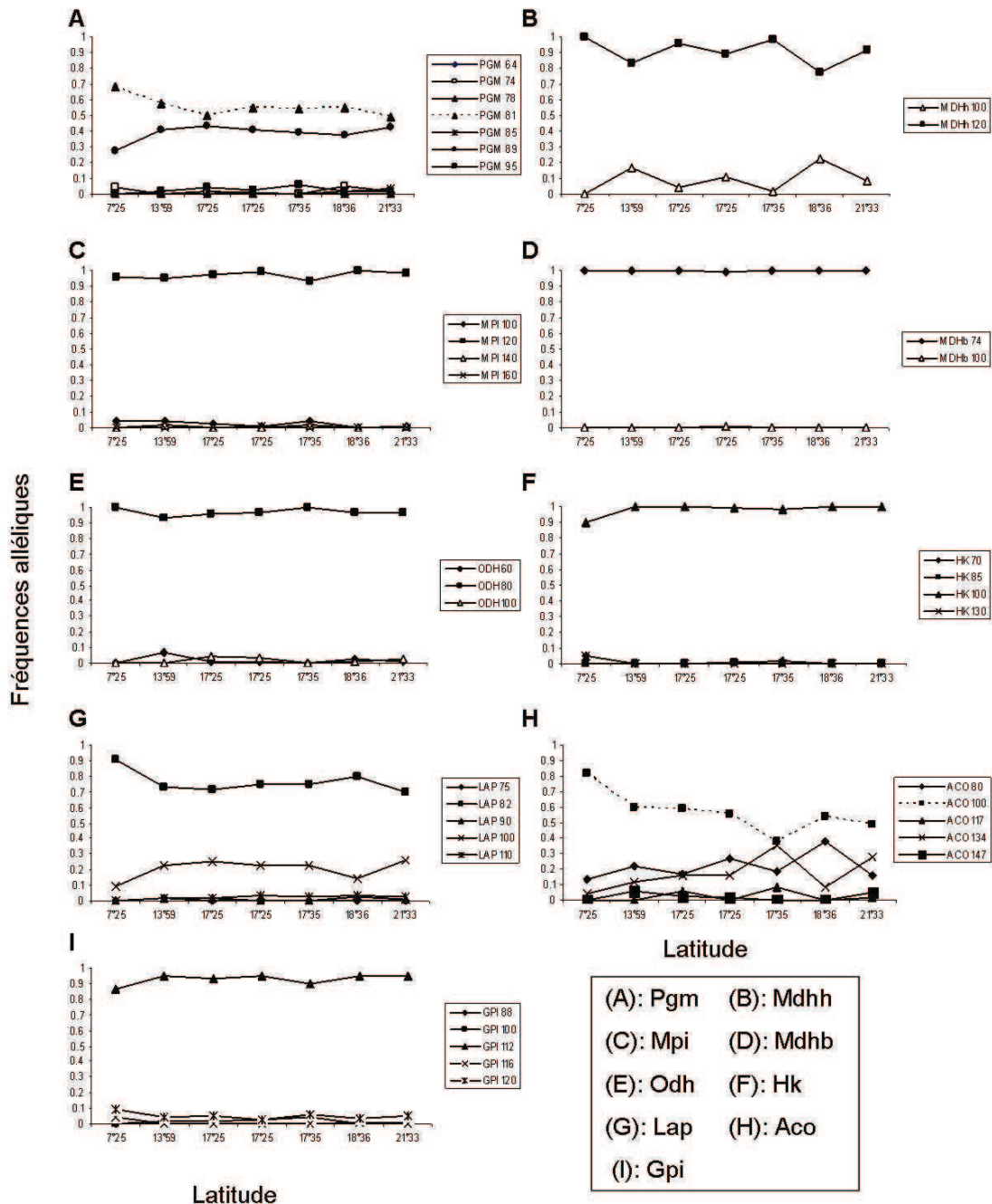


Figure IV.2. Distribution latitudinale des fréquences alléliques pour les 9 locus enzymatiques le long de la dorsale EPR.

La distribution des fréquences alléliques le long de la dorsale EPR (Figure IV.2) démontre une homogénéité des fréquences pour l'ensemble des populations de modioles situées entre 7° et 21°33S. Aucune inversion de fréquence ni aucun effet clinal fort n'est observé, cependant la population de 7°25S (Sarah's Spring) semble présenter une légère différenciation génétique vis-à-vis des autres populations pour les locus *Pgm*, *Hk*, *Lap*, *Aco*, *Odh* et *Gpi*. Cependant, ces différences pourraient être dues au faible effectif récolté à 7°25S (11 individus). Six des locus (*Mdh* haut, *Mpi*, *Mdh* bas, *Odh*, *Hk*, et *Gpi*) sont caractérisés par la présence d'un allèle majoritaire aux fréquences supérieures à 90% avec un cortège d'allèles rares. La *Pgm* et la *Lap* présentent deux allèles majoritaires aux fréquences plus ou moins équivalentes dans les populations : 50/50 (*Pgm*) et 80/20 (*Lap*). L'aconitase est nettement plus variable que les autres locus, tant au niveau du nombre d'allèles dont les fréquences sont supérieures à 10%, que pour la variabilité inter-sites de la diversité allélique. Par exemple, l'allèle majoritaire *Aco*_100 varie en fréquence de 82% pour Sarah's Spring (7°25S) à 37% pour la population de Suzie (17°35S).

Tableau IV.4. Tableau récapitulatif des valeurs de variabilité génétique. **N_{all}** : nombre moyen d'allèles ; **H_{obs}** : hétérozygotie observée ; **H_{NB}** ; Hétérozygotie attendue corrigée selon Nei 1987) et valeurs de **F_{is}** obtenues pour chacun des locus à chaque population. (* p< 0,05 ; ** p<0,01, *** p<0.0001)

	Sarah's Spring	LuckyEric	Oasis	RehuMarka	Suzie	AnimalFarm	Gromit
Nall	2.111	2.667	2.889	3.111	2.444	2.556	3.333
Hobs.	0.164	0.228	0.197	0.2	0.179	0.151	0.187
Ecart type	0.175	0.206	0.214	0.204	0.2	0.176	0.195
HNB	0.169	0.235	0.214	0.217	0.229	0.222	0.227
Ecart type	0.166	0.221	0.245	0.231	0.263	0.241	0.258
F _{is} Pgm	-0.143	-0.118	0.042	0.009	-0.047	0.273*	0.227**
Mpi	0	-0.031	-0.02	-0.003	0.319	-	0.394*
Odh	-	-0.056	-0.035	-0.026	-	-0.016	-0.023
Lap	-0.053	0.08	-0.002	0.026	0.197	0.245*	0.018
Gpi	-0.071	-0.031	-0.051	0.228	-0.065	-0.028	-0.048
Mdh bas	-	-	-	-0.006	-	-	0
Mdh haut	-	-0.188	-0.034	-0.114	-	-0.273	-0.087
Hk	-0.029	-	-	-0.003	0	-	-
Aco	0.452	0.252*	0.244**	0.240*	0.534***	0.862***	0.345***
total locus	0.032	0.027	0.082*	0.080*	0.222**	0.327***	0.177***

Les valeurs de Fis obtenues pour chacun des locus pour chaque population sont indiquées dans le Tableau IV.4. Seule la population de Sarah's Spring ne présente pas d'écart significatif à l'équilibre de Hardy-Weinberg. Les six autres populations présentent des locus avec un déficit en hétérozygotes (*Pgm*, *Mpi* et *Aco* pour Gromit ; *Aco* pour Oasis ; *Pgm*, *Lap2* et *Aco1* pour Animal Farm ; *Aco1* pour Suzie, Rehu Marka et Lucky Eric). Le locus *Aco1* présente un déficit en hétérozygotes pour 6 des 7 populations.

Un déficit en hétérozygotes sur l'ensemble des locus est détecté pour 6 populations sur 7 (Gromit, Oasis, Animal Farm, Suzie, Rehu Marka et Lucky Eric).

La diversité génétique des sites peut être comparée à partir des estimateurs d'hétérozygotie. L'hétérozygotie observée (H_{obs}) montre une légère augmentation de la diversité vers le sud (0,164 à 7°25S jusqu'à 0,228 à 21°33S). Cette tendance est moins nette si l'on utilise les valeurs attendues (H_{NB}) mais les deux populations extrêmes 7°25S et 21°33S conservent respectivement les valeurs de diversité la plus faible et la plus élevée

Les valeurs de déséquilibre de liaison observées sont indiquées dans le Tableau IV.5. Des déséquilibres sont observés au niveau de chaque population et sur l'ensemble des populations pour chaque paire de locus. Les corrections de Bonferroni séquentielles ont été calculées en considérant dans un premier temps le niveau de significativité de 0,05 divisé par le nombre de paires de locus comparées ($p < 0,05/36 = p < 0,0014$). Les paires de locus significatives sont notées puis l'analyse est reconduite en ne considérant que le nombre de paires de locus restant. Dans notre cas, trois valeurs étaient significatives avec le premier seuil de correction, mais aucune nouvelle ne l'a été lorsque nous avons considéré le second seuil de 0,0015 (= $0,05/(36-3)$). Lorsque l'on regarde la variation du nombre de déséquilibres de liaison significatifs entre les différentes populations, on peut constater que ce nombre diffère selon les sites échantillonnés : Sarah's Spring, Lucky Eric et Animal Farm (0), Suzie et Gromit (1) Oasis (2), Rehu Marka (3). Cependant, ces valeurs sont essentiellement liées au locus *Aco* pour lequel des difficultés de mise au point du génotypage ont certainement entraîné des imprécisions de lectures (Le Guen, D., communication personnelle). Les populations de 17°S (Rehu Marka, Oasis et dans une moindre mesure Suzie) montrent un nombre de déséquilibres de liaison plus élevé que celle des autres sites (et qui ne concernent pas seulement le locus *Aco*). Ce résultat ne peut être lié à une dépendance physique entre locus, et serait plutôt lié à : (i) la structuration génétique interne des populations avec

l'existence d'au moins deux entités distinctes ou (ii) des effets sélectifs pouvant être liés à de la sélection contre les hétérozygotes ou de la sélection diversifiante. Cependant, dans un contexte de zone hybride, des déséquilibres de liaison sont attendus pour tous les locus d'une population, ce qui n'est pas le cas de nos données. Nous pouvons donc écarter cette hypothèse.

La valeur de la différenciation globale calculée entre les sept populations est significativement différente de zéro ($F_{st} = 0,0095$; 1000 permutations, $P = 0,005$). Pour déterminer si certains sites ou locus particuliers étaient responsables de cette différenciation, nous avons utilisé une approche de rééchantillonnage 'Jackknife', en testant les sites (ou les locus) un par un (cf Tableau IV.6). De cette manière nous nous sommes rendu compte que la différenciation persistait quel que soit le site ou le locus supprimé de l'analyse. La valeur de différenciation est cependant tirée par la population de Sarah's Spring et par le locus de la Mdh. En effet, lorsque ceux-ci ne sont plus inclus dans le calcul des F_{st} , le test de cette valeur flirte avec les niveaux de non-significativité. Nous avons également réalisé cette analyse en éliminant simultanément les populations extrêmes (Sarah's spring (7°25S) et Gromit (21°33S)). Dans ce cas, les populations situées entre 13°59S et 18°33S ne présentent plus de différenciation (données non présentées).

Les valeurs de différenciation par paires de populations ont également été calculées et sont regroupées dans le Tableau IV.7. Seule, la population de Rehu Marka (localisation centrale au niveau de l'EPR sud) n'est pas différente des autres populations, alors que la population de Sarah's Spring diffère significativement des trois populations situées plus au sud. Aucune barrière géographique nette n'est mise en évidence et la différenciation des populations ne semble pas liée à la distance géographique. La population de 21°33S par exemple est différente de celle de 18°35S alors qu'elle ne l'est pas de la population de 13°59. Il est cependant intéressant de noter que les 3 populations de 17°S (Oasis, Rehu marka et Suzie) ne présentent pas de différenciation pour les F_{st} par paires.

Afin de détecter un isolement potentiel par la distance, nous avons calculé un indice de distance à partir du F_{st} : $F_{st}(1-F_{st})$. Cette distance est indiquée dans la demi-matrice au dessus de la diagonale du Tableau IV.7. Ces distances sont comparées aux distances géographiques calculées entre les sites et sont indiquées sous la diagonale du tableau Tableau IV.7.

Tableau IV.5. Tableau récapitulatif des déséquilibres de liaison par paire de locus estimés à partir de la méthode de Garnier & Dillman sous Genetix. Le chiffre entre parenthèses correspond au nombre d'allèles impliqués dans les déséquilibres de liaison significatif sein de chaque population. Le niveau de significativité indiqué correspond à celui après corrections de Bonferroni. * : significatif après corrections; NS : non significatif, tirets '-' : aucun déséquilibre de liaison détecté.

Paires de locus		Sarah' Spring	Lucky Eric	Oasis	Rehu Marka	Suzie	Animal Farm	Gromit	Populations regroupées
Pgm	Mpi	-	-	* (2)	-	-	-	-	-
Pgm	Odh	-	-	-	-	-	-	-	-
Pgm	Lap	-	-	-	-	-	NS (1)	-	-
Pgm	Gpi	-	-	-	* (1)	-	-	-	NS (1)
Pgm	Mdh b	-	-	-	-	-	-	-	-
Pgm	Mdh h	-	-	-	-	-	-	-	-
Pgm	Hk	-	-	-	-	-	-	-	-
Pgm	Aco	-	-	-	-	-	-	-	* (1)
Mpi	Odh	-	-	-	-	-	-	-	-
Mpi	Lap	-	NS (1)	-	-	* (3)	-	-	NS (1)
Mpi	Gpi	-	NS (2)	-	-	-	-	-	-
Mpi	Mdh b	-	-	-	-	-	-	-	-
Mpi	Mdh h	-	-	-	-	-	-	-	-
Mpi	Hk	-	-	-	-	-	-	-	-
Mpi	Aco	-	-	-	-	-	-	-	-
Odh	Lap	-	-	-	-	-	-	-	-
Odh	Gpi	-	-	-	-	-	-	-	-
Odh	Mdh b	-	-	-	-	-	-	-	-
Odh	Mdh h	-	-	-	-	-	-	-	-
Odh	Hk	-	-	-	-	-	-	-	-
Odh	Aco	-	-	-	-	-	-	-	-
Lap	Gpi	-	NS (2)	-	-	-	-	-	-
Lap	Mdh b	-	-	-	-	-	-	-	-
Lap	Mdh h	-	-	-	NS (2)	-	-	-	-
Lap	Hk	-	-	-	-	-	-	-	-
Lap	Aco	-	-	-	-	-	-	* (1)	* (1)
Gpi	Mdh b	-	-	-	-	-	-	-	-
Gpi	Mdh h	-	-	-	-	-	-	-	-
Gpi	Hk	-	-	-	-	-	-	-	-
Gpi	Aco	-	NS (1)	-	-	-	-	-	NS (1)
Mdh b	Mdh h	-	-	-	-	-	-	-	-
Mdh b	Hk	-	-	-	-	-	-	-	-
Mdh b	Aco	-	-	-	* (2)	-	-	-	-
Mdh h	Hk	-	-	-	-	-	-	-	-
Mdh h	Aco	-	-	* (2)	* (2)	-	-	-	* (1)
Hk	Aco	-	-	-	-	-	-	-	-

Tableau IV.6. Valeurs de différenciation (**Fst**) calculées par stratégie Jackknife pour chacune des populations et chacun des locus. Les valeurs de Fst sont testées avec 1000 permutations.
* : $P < 0,05$; ** : $P < 0,01$

stratégie Jackknife		valeurs Fst	P=
population supprimées	locus supprimé		
aucune	aucun	0,0095**	0.005
Sarah's spring	aucun	0,0063*	0.048
Gromit	aucun	0,0112*	0.016
Oasis	aucun	0,0119*	0.011
Animal Farm	aucun	0,0073*	0.026
Suzie	aucun	0,0089*	0.018
Rehu marka	aucun	0,0126**	0.003
Lucky Eric	aucun	0,0109**	0.009
aucune	Pgm	0,0143**	0.004
aucune	Mpi	0,0097*	0.011
aucune	Odh	0,0097**	0.005
aucune	Lap	0,0116**	0.008
aucune	Gpi	0,0101**	0.005
aucune	Mdh b	0,0096**	0.008
aucune	Mdh h	0,0065*	0.05
aucune	Hk	0,0094*	0.01
aucune	Aco	0,0049*	0.037

Tableau IV.7. Matrice de différenciation et de distance géographique. Au dessus de la diagonale et dans chaque case : valeur du haut : **Fst** par paire, valeur du bas : estimateur **Fst(1- Fst)**. En dessous de la diagonale **distances géographiques** (en Km) séparant les paires de sites.

	Sarah's Spring	Lucky Eric	Oasis	Rehu Marka	Suzie	Animal Farm	Gromit
Sarah's Spring		0,029 0,028	0,032* 0,031	0,034 0,033	0,063** 0,059	0,04 0,038	0,054** 0,051
Lucky Eric	892		0,002 0,002	-0,007 -0,007	0,021* 0,021	-0,007 -0,007	0,007 0,007
Oasis	1257	389		0,000 0,000	0,009 0,009	0,021* 0,021	0,002 0,002
Rehu Marka	1257	389	0		0,011 0,011	-0,004 -0,004	0,004 0,004
Suzie	1276	409	19	19		0,031 0,030	-0,001 -0,001
Animal Farm	1384	523	133	133	114		0,023** 0,022
Gromit	1722	863	474	474	455	341	

Les relations entre les distances génétiques et les distances géographiques sont représentées sur le graphique de la Figure IV.3.

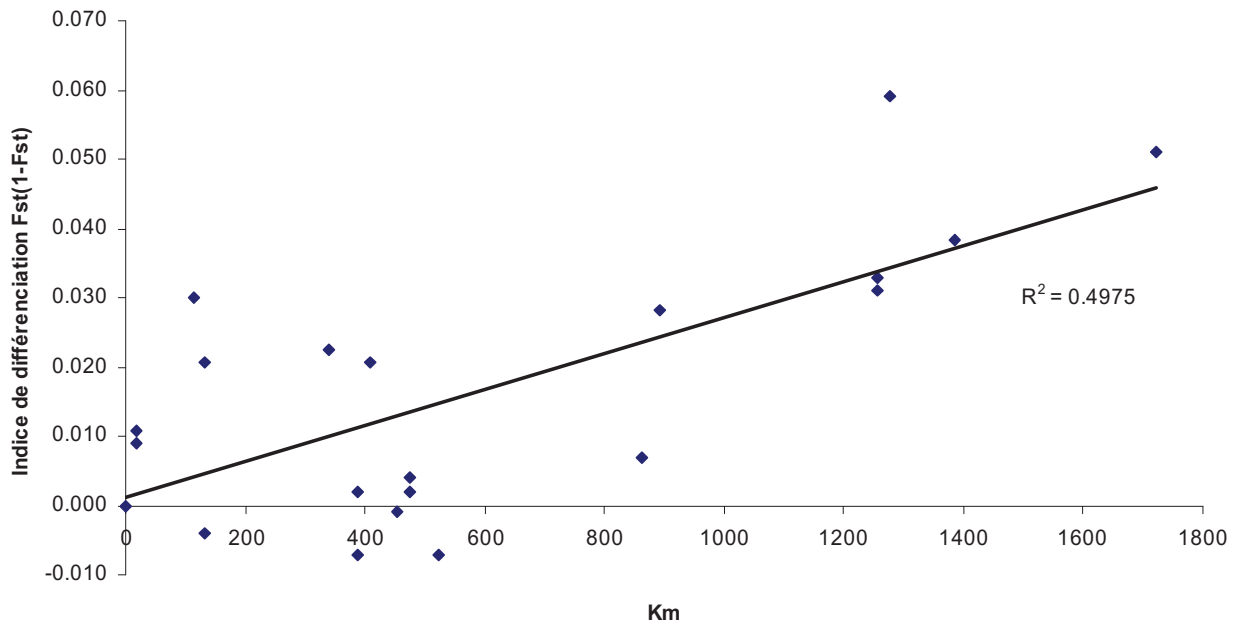


Figure IV.3. Illustration de la distance génétique en fonction de la distance géographique.

La représentation graphique met en évidence une tendance à l'augmentation de la distance génétique avec la distance géographique. Trois éléments pourraient expliquer cette tendance : (1) Isolement par la distance à l'équilibre migration/dérive de proche en proche, (2) sélection diversifiante associée à la géographie, (3) une zone hybride.

La corrélation entre les distances génétiques et géographiques ont été testées par un test de Mantel (Mantel 1967) sous Genetix (Belkhir et al. 2004). La valeur du test est de $p=0,139$ et donc non significative. Cependant, la probabilité d'erreur en rejetant l'hypothèse nulle, H_0 , d'indépendance des distances génétiques et géographiques des populations, n'est que de 14%. L'hypothèse de l'isolement par la distance des populations reste donc la plus probable.

IV.3.2. Recherche d'allèles introgressés à partir d'une analyse de coalescence de huit gènes

IV.3.2.a. Analyse de la différenciation génétique entre les populations cibles de B. thermophilus de part et d'autre du gradient latitudinal

L'analyse de la différenciation génétique des populations situées aux extrémités de la distribution latitudinale de *B. thermophilus* (plus les individus de *B. aff. thermophilus* 37°S) a été réalisée sous Arlequin.

Les valeurs de Φ_{st} obtenues par paire de populations sont indiquées dans le Tableau IV.8. Les deux individus de 37°S se différencient de manière significative des autres populations et ce, pour l'ensemble des gènes à l'exception des gènes SAHH et SULFO pour la population 21°33S. Les résultats des tests de différenciation concernant les 3 autres populations de *B. thermophilus* sur l'EPR sont contrastés. Cinq gènes montrent une structure génétique entre les populations de 7°25S et 21°33S (COI, EF1 α , SAHH, SULFO et ITS2) alors que les populations 7°25S et 21°33S ne sont pas génétiquement différenciées aux gènes LYSO, GF1A et GF1B. Aucune différenciation n'est détectée entre les populations de 9°50N et de 7°25S (résultats basés sur 3 gènes uniquement).

Les données de diversité nucléotidique et les tests d'écarts à l'évolution neutre obtenus pour les 8 gènes sont résumés dans le Tableau IV.9. La population de 9°50N dont les ADN étaient de moins bonne qualité n'a pu être amplifiée que pour quelques individus aux locus nucléaires, EF1 α et ITS2. Les deux populations situées aux extrêmes du gradient latitudinal sud de l'espèce *B. thermophilus* (7°25S et 21°33S) ont été choisies pour étudier plus en détail les coalescences d'allèles pour 7 locus nucléaires et un locus mitochondrial. Le nombre de séquences est similaire entre les deux populations pour chacun des gènes étudiés mais il existe une forte disparité du nombre de séquences entre les différents gènes (de 9 et 10 séquences pour SULFO, à 27 et 48 séquences pour COI).

La population de 21°33S montre une diversité nucléotidique (θ_w) supérieure à celle estimée pour les deux autres populations (de 1,5 à 7 fois supérieure) pour l'ensemble des gènes nucléaires à l'exception du locus SULFO. En revanche la diversité nucléotidique du marqueur COI est deux fois moindre que celle des populations du nord de la dorsale. Cette

tendance est cependant nettement moins marquée en ce qui concerne l'estimateur π ou l'on observe même une inversion de tendance au locus SAHH.

Tableau IV.8. Différenciation estimée par le calcul du Φ_{st} par paires de populations. Les valeurs ont été testées avec 1000 permutations. * : $P < 0,05$; ** : $P < 0,01$

		7°25S	21°33S	37°S
COI	<i>Bt_9°50N</i>	0,02757	0,32736**	0,85791**
	<i>Bt_7°25S</i>		0,16185**	0,85079**
	<i>Bt_21°33S</i>			0,94134**
EF1	<i>Bt_9°50N</i>	-0,13965	-0,10648	1.0000
	<i>Bt_7°25S</i>		0,06001*	0,97902**
	<i>Bt_21°33S</i>			0,7201*
SAHH	<i>Bt_7°25S</i>		0,81388**	0,78662**
	<i>Bt_21°33S</i>			0,03282
LYSO	<i>Bt_7°25S</i>		0,04172	0,20614*
	<i>Bt_21°33S</i>			0,29428*
GF1_A	<i>Bt_7°25S</i>		-0,00531	0,26218*
	<i>Bt_21°33S</i>			0,28413**
GF1_B	<i>Bt_7°25S</i>		-0,00531	0,26218*
	<i>Bt_21°33S</i>			0,28413*
SULFO	<i>Bt_7°25S</i>		0,00098**	0,04785*
	<i>Bt_21°33S</i>			0,99902
ITS2	<i>Bt_9°50N</i>	0,03572	0,40885**	0,7363**
	<i>Bt_7°25S</i>		0,26469**	0,54534**
	<i>Bt_21°33S</i>			0,30417**

La divergence avec l'espèce atlantique *B. azoricus* est similaire pour chacune des populations au sein d'un même gène, en revanche il existe de fortes disparités de divergence entre les différents gènes : la moyenne géométrique de la divergence sur l'ensemble des gènes et des populations est de 3,45%. La divergence trouvée au niveau des gènes nucléaires est environ quatre fois inférieure à celle observée pour le gène mitochondrial COI (12%). Le gène présentant la divergence Pacifique/Atlantique la plus faible est la SULFO avec une divergence moyenne de 1,1%.

Si l'on considère la divergence entre les deux espèces cryptiques de la dorsale du Pacifique Est (*B. thermophilus* (9°50N, 7°25S, 21°33S), et *B.aff. thermophilus* 37°S), celle-ci

est de 4,6% pour le gène mitochondrial quelle que soit la population '*thermophilus*' considérée. La divergence entre les populations 7°25S et 21°33S vis-à-vis de 37°S est à peu près la même entre les locus. En revanche pour les locus SAHH et SULFO, la divergence entre 21°33S/37°S est quasi nulle alors qu'elle est très marquée entre 7°25S et 37°S.

Le test de Tajima a permis d'utiliser les données de polymorphisme afin de tester si l'évolution des gènes présentait un écart à l'attendu sous l'hypothèse de neutralité et d'équilibre mutation/dérive.

Les valeurs du D pour la population de 21°33S sont significativement négatives pour quatre gènes (COI, EF1 α , SAHH et GF1_B). Pour la population de 7°25S, seuls EF1 α et ITS2 présentent des valeurs significativement négatives. Au contraire, le gène SAHH présente une valeur significativement positive au niveau de la population de 7°25S. En fait, seulement trois gènes ne s'écartent pas de l'attendu neutre sur l'ensemble des populations : le Lysozyme, la Sulfotransférase et le GF1_A. Le gène EF1 α est quant à lui le seul à présenter des valeurs de D négatives et significatives pour les deux populations. Le résultat le plus surprenant est celui observé pour le gène SAHH qui présente des valeurs opposées et significatives pour chacune des deux populations. Même si celles-ci ne sont pas significatives, il faut noter que quatre gènes présentent des valeurs du D positives pour la population de 7°25S, alors que toutes les valeurs obtenues pour la population de 21°33S sont inférieures ou très proches de -1.

Des valeurs négatives à tous les locus auraient pu être la conséquence d'un effet démographique en relation avec une phase d'expansion des populations. Cependant, le nombre limité de valeurs négatives significatives observées dans les deux populations semble contredire l'hypothèse démographique. La première explication des valeurs négatives de D est évidemment l'excès de mutations en faible fréquences, cependant des valeurs négatives peuvent aussi être interprétées comme une trace de sélection purifiante ou d'un balayage sélectif à certains locus. La seule valeur positive significative observée pour la population de 7°25S au locus SAHH ne peut s'expliquer par une réduction récente de la taille de la population (*bottleneck*) pour les mêmes raisons et pourrait être le signe d'une sélection balancée ou d'introgession à ce locus.

Tableau IV.9. Récapitulatif des analyses réalisées sur les 7 gènes nucléaires et le gène mitochondrial. L'espèce atlantique *B. azoricus* (*B. azo*) sera utilisée comme groupe externe dans la suite du chapitre. Les valeurs de significativité sont indiquées par : * : $P < 0,05$. Les tirets '-' indiquent des valeurs impossibles à calculer par manque de polymorphisme.

locus	Espèce	N	L	S	θ	π	D de Tajima	Div. vs. Rb	Div, vs 37S	Rm
CO1	<i>B. thermo_9°50N</i>	45	528	15	0,0065	0,0068	0,1581	0,1159	0,0458	0
	<i>B. thermo_7°25S</i>	48	528	20	0,0085	0,0072	-0,4883	0,1165	0,0461	4
	<i>B. thermo 21°33S</i>	27	528	10	0,0049	0,0026	-1,5126*	0,1173	0,0462	1
	<i>B. thermo 37°S</i>	2	528	4	0,0076	0,0076		0,1226		0
	<i>B. azo Rainbow</i>	4	528	7	0,0072	0,0066	-0,8173			0
EF1	<i>B. thermo_9°50N</i>	4	527	0	0,0000	0,0000	-	0,0254	0,0135	-
	<i>B. thermo_7°25S</i>	25	527	2	0,0010	0,0003	-1,5140*	0,0255	0,0137	0
	<i>B. thermo 21°33S</i>	16	540	12	0,0070	0,0040	-1,6634*	0,0236	0,0129	0
	<i>B. thermo 37°S</i>	2	529	0	0,0000	0,0000		0,0270		-
	<i>B.azo Rainbow</i>	4	533	1	0,0010	0,0009	-0,6124			0
SAHH	<i>B. thermo_7°25S</i>	24	859	16	0,0050	0,0071	1,5189*	0,0401	0,0335	0
	<i>B. thermo 21°33S</i>	19	813	29	0,0102	0,0050	-2,0105*	0,0308	0,0051	2
	<i>B. thermo 37°S</i>	3	900	7	0,0052	0,0052		0,0380		0
	<i>B.azo Rainbow</i>	4	880	11	0,0068	0,0063	-0,8374			0
LYSO	<i>B. thermo_7°25S</i>	25	1231	23	0,0050	0,0057	0,5264	0,0572	0,0071	5
	<i>B. thermo 21°33S</i>	18	1231	35	0,0085	0,0059	-1,2655	0,0577	0,0082	2
	<i>B. thermo 37°S</i>	2	1231	7	0,0057	0,0057		0,0589		0
	<i>B.azo Rainbow</i>	4	1231	26	0,0115	0,0123	0,7160			0
GF1_A	<i>B. thermo_7°25S</i>	13	698	19	0,0088	0,0090	0,0931	0,0179	0,0120	2
	<i>B. thermo 21°33S</i>	16	698	33	0,0147	0,0115	-0,9030	0,0195	0,0149	3
	<i>B. thermo 37°S</i>	3	698	9	0,0086	0,0086		0,0209		0
	<i>B.azo Rainbow</i>	4	698	5	0,0039	0,0036	-0,7968			0
GF1_B	<i>B. thermo_7°25S</i>	16	741	11	0,0045	0,0048	0,2583	0,0283	0,0041	2
	<i>B. thermo 21°33S</i>	15	741	32	0,0133	0,0074	-1,8773*	0,0294	0,0050	0
	<i>B. thermo 37°S</i>	2	741	3	0,0041	0,0041		0,0270		0
	<i>B.azo Rainbow</i>	4	741	22	0,0162	0,0171	0,5708			2
SULFO	<i>B. thermo_7°25S</i>	10	216	7	0,0115	0,0092	-0,8501	0,0157	0,0110	0
	<i>B. thermo 21°33S</i>	9	216	4	0,0068	0,0064	-0,2288	0,0103	0,0045	0
	<i>B. thermo 37°S</i>	4	216	1	0,0025	0,0023	-0,6124	0,0077		0
	<i>B.azor Rainbow</i>	4	216	2	0,0050	0,0046	-0,7099			0
ITS2	<i>B. thermo_9°50N</i>	8	397	1	0,0010	0,0014	1,1665	0,0399	0,0089	0
	<i>B. thermo_7°25S</i>	10	397	9	0,0089	0,0058	-1,5345*	0,0410	0,0098	0
	<i>B. thermo 21°33S</i>	16	397	12	0,0091	0,0058	-1,3817	0,0380	0,0071	1
	<i>B. thermo 37°S</i>	16	397	6	0,0046	0,0032		0,0370		0
	<i>B.azo Rainbow</i>	4	376	5	0,0073	0,0067	-0,7968			0

L'hétérogénéité des populations pour les différents locus peut être observée à partir de la distribution des différences par paires (courbes de mismatch) (Figure IV.4). Nous avons choisi de regrouper les populations de 7°S et 21°S avec celle de 37°S, car la différenciation génétique entre ces deux espèces proches (*B. thermophilus* et *B. aff. thermophilus*) ne concerne pas l'intégralité des locus. La présentation de la courbe de mismatch est un excellent moyen de comparer le niveau de différenciation des deux espèces entre les locus. En revanche, la courbe représentant les fréquences attendues doit être considérée avec précautions car elle est établie sous l'hypothèse d'une population homogène, or pour certain locus, les allèles de *B. thermophilus* et *B. aff. thermophilus* sont nettement divergents.

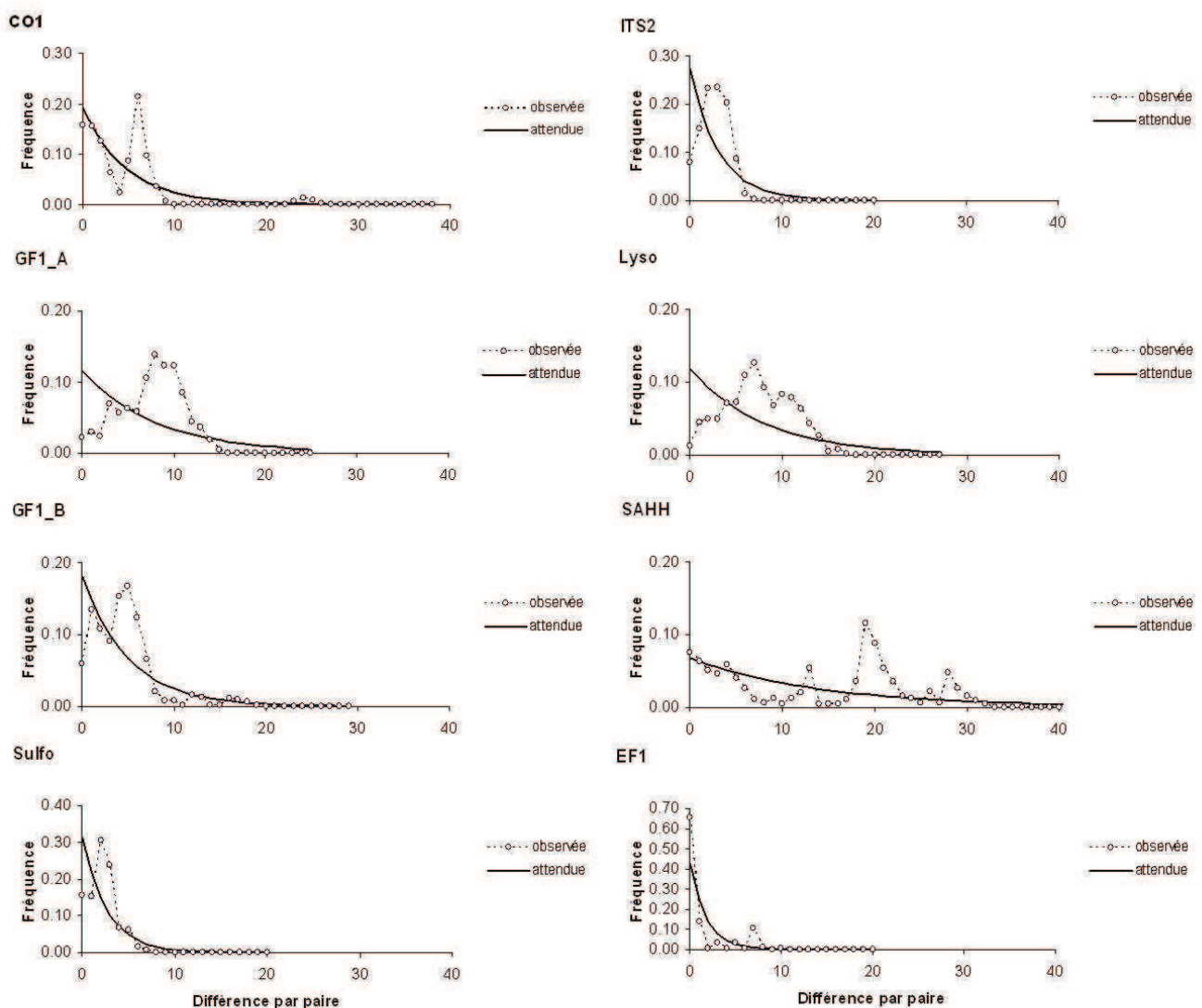


Figure IV.4. Courbes de mismatch pour les 8 locus étudiés. Cette distribution des fréquences du nombre de différences par paire de séquence à été réalisée à partir de l'ensemble des séquences du Pacifique : 9°50N (uniquement pour COI, EF1 α , ITS2), 7°25S, 21°33S et 37°S

La représentation en histogrammes de mismatch permet d'observer pour certains locus la formation de pics, caractéristiques de la présence d'une structure généalogique. Ces pics observés à partir de la totalité des séquences de chaque locus permettent d'identifier le nombre de 'groupes' de séquences divergentes présents sur la dorsale pacifique. La présence d'allèles divergents peut être liée à des populations structurées sur la dorsale, ou à une structuration intra-pop avec la présence de deux lignées divergentes (lignées ancestrales, sélection balancée, introgression...). A l'exception de l'EF1, tous les locus s'écartent de l'attendu neutre à l'équilibre. Le nombre de pics observés et donc de groupes d'allèles divergents est très variable selon les locus. Pour la majorité des locus (CO1, ITS2, GF1A GF1B) on observe un pic principal pour une divergence de cinq à une dizaine de substitutions, signe de l'existence d'au moins deux grands ensembles d'allèles. Le lysozyme présente deux pics conjoints et le SAHH présente au moins quatre pics de fréquence. Pour nos données, on s'attend à ce que les deux groupes correspondent aux populations *B. thermophilus* et *B. aff. thermophilus* de 37°S. Les groupes supplémentaires reflètent certainement une structuration des populations de *B. thermophilus* (entre 7°25S et 21°33S).

Les gènes ITS, SULFO et EF1 présentent un excès de mutations rares alors que les autres gènes montrent des excès de mutations en fréquence intermédiaire, scénario en faveur de la différenciation des populations ou de l'introgression d'allèles.

Un test HKA multi-locus a également été conduit pour les 2 populations de 21°33S et 7°25S en considérant la divergence avec l'espèce atlantique *B. azoricus* en groupe externe. Les résultats de ces tests sont présentés dans le Tableau IV.10. Ce test montre que toutes les mutations se sont accumulées de manière neutre dans le polymorphisme des *B. thermophilus*.

Tableau IV.10. Résultats des tests HKA réalisés sur les 8 gènes.

***B. thermophilus* 7S avec *B. azoricus* Rainbow**

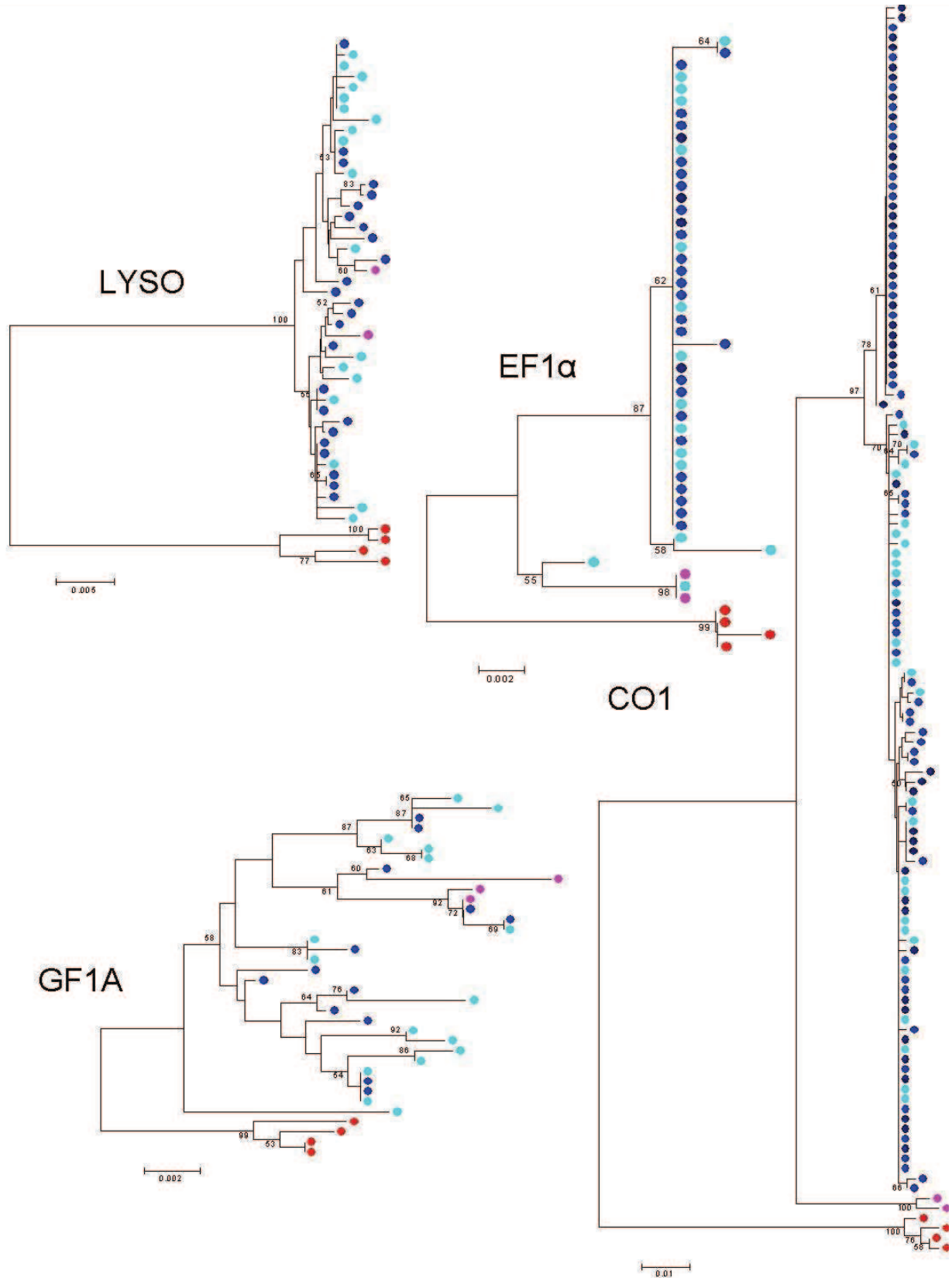
Sum of deviations: 21,4828 (df= 14, $P = 0.0899$)

***B. thermophilus* 21S avec *B. azoricus* Rainbow**

Sum of deviations: 12,0794 (df= 14, $P = 0.5999$)

IV.3.2.b. Arbres de coalescence des 8 gènes étudiés

Un total de 430 séquences issues des jeux de données HQ ont été conservées dans l'analyse et ont permis la réalisation des 8 arbres phylogénétiques représentés sur la Figure IV.5.



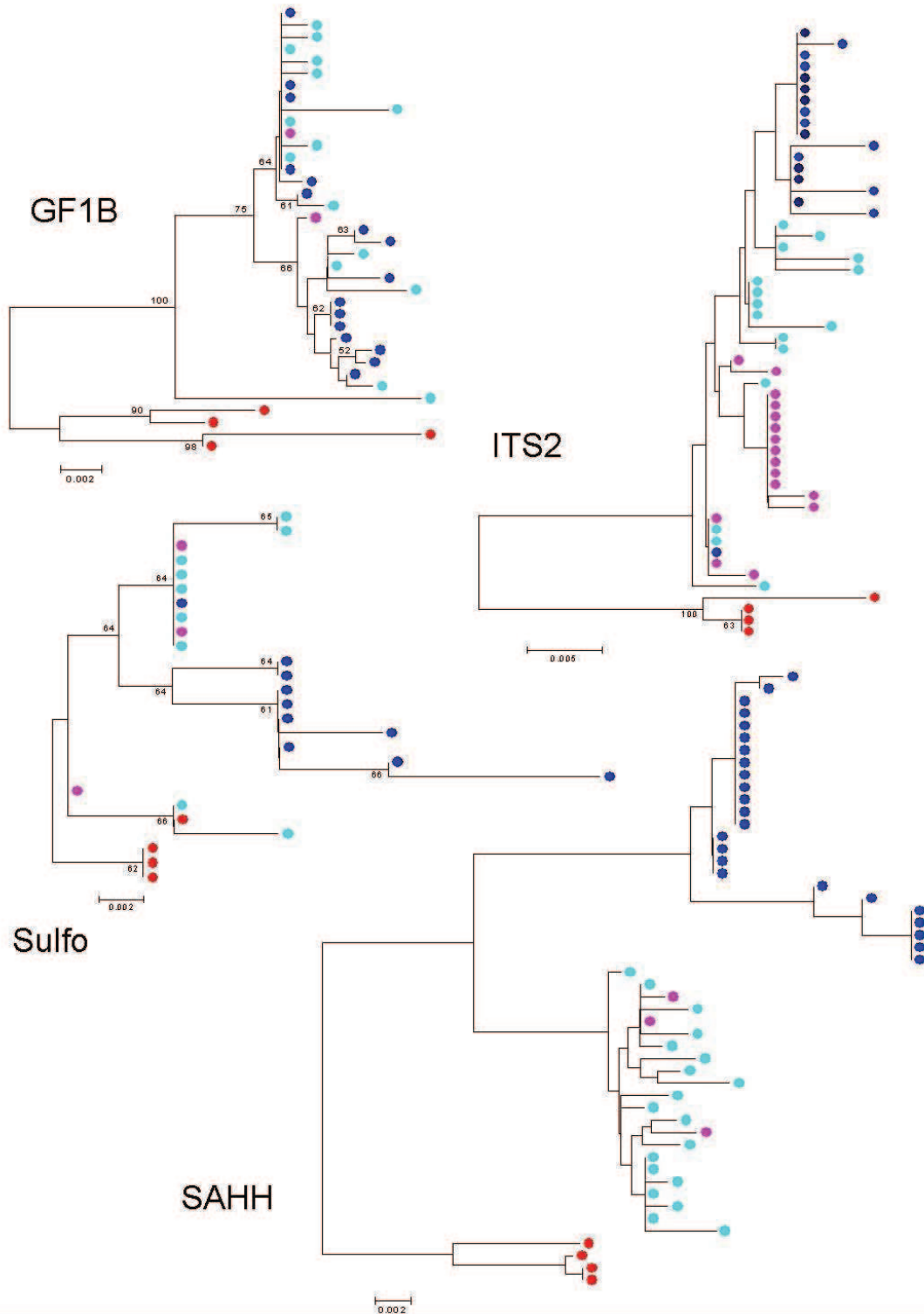


Figure IV.5. Arbres phylogénétiques des allèles réalisés par Neighbour-Joining avec des distances calculées selon la méthode de Tamura et Nei; Noir (EF1 α , ITS2 et COI) *Bathymodiolus thermophilus* population 9°50N; Bleu foncé: *B. thermophilus* population 7°25S, Bleu clair: *B. thermophilus* population 21°33S; Violet: *B. aff. thermophilus* 37°S. Rouge: séquences de *B. azoricus* utilisées pour enraciner l'arbre. Les valeurs de bootstrap indiquées au dessus des branches sont les pourcentages obtenus après 1000 ré-échantillonnage du jeu de données.

Seuls deux gènes (COI et EF1 α) correspondent à l'attendu selon lequel l'espèce *B. thermophilus* (présente à 7°25S et à 21°33S) serait différente de *B. aff. thermophilus* (37°S). En effet, ces deux gènes sont les seuls à présenter une nette divergence séparant les deux espèces 'vicariantes'. Le gène SAHH présente aussi une nette dichotomie des séquences en deux clades divergents mais celle-ci sépare les populations les plus au sud (21°33S et 37°S) de la population nord (7°25S). Pour ce locus, il convient également de noter que la population 7°25S présente 2 types alléliques nettement divergents (à l'origine de la valeur significative positive du test de Tajima) que l'on pourrait rapprocher de la situation observée au locus COI (présence de 2 haplotypes divergents uniquement trouvés en mélange à 7°25S et 9°50N). Les arbres des autres gènes ne présentent pas de divergence aussi marquée alors que des clades peuvent être distingués pour la plupart d'entre eux. Ainsi, deux clades se détachent pour les gènes du Lysozyme, du GF1A et du GF1B. Pour ces deux derniers gènes, il convient de noter la présence d'un allèle extrêmement divergent pouvant être considéré comme un clade à part entière : 1,5 % de divergence avec les populations *B. thermophilus* (5 mutations dans l'intron) pour GF1_A ; et près de 2% de divergence (8 mutations introniques, 2 silencieuses et 2 non synonymes) pour l'allèle divergent de GF1_B). Ces allèles isolés appartiennent à deux individus distincts dans la population de 21°33S. Les gènes de la sulfotransférase et de l'ITS2 présentent une structure plus complexe de type paraphylétique du sud vers le nord. La sulfotransférase présente une divergence entre les individus de 7°25S et ceux du sud (21°33S et 37°S). Un clade peut être considéré comme clade 'Sud' et réunit la majorité des séquences de 21°33S et 37°S. Il inclut également une séquence de 7°25S. En revanche, un problème se pose car une séquence du groupe externe *B. azoricus* possède un clade commun avec des séquences de 21°33S. L'évolution lente de ce locus déduite de la faible divergence avec le groupe externe est probablement liée à sa particularité d'être entièrement constitué d'un exon. L'arbre observé pour le gène de la SULFO ne peut en aucun cas être obtenu sous l'hypothèse d'une évolution lente et neutre. En effet, sous condition de la neutralité, il n'est pas possible que certaines lignées ne présentent pas de divergence avec l'Atlantique alors que certaines branches, longues et différenciées dans le Pacifique, montrent bien que ce gène peut évoluer plus rapidement. Des allèles recombinés auraient pu avoir cette position 'basale' dans l'arbre, mais il est difficile d'imaginer cette hypothèse pour des allèles provenant d'océans différents.

La sélection purifiante qui agit sur les mutations non-synonymes des régions codantes peut expliquer cette évolution plus lente. Il est cependant surprenant d'observer une divergence entre les populations de 7°25S et celle du sud de la dorsale. Une seule mutation

non-synonyme proche sépare les deux clades (Méthionine (M) pour 7°25S et Valine (V) pour 21°33S et 37°S).

La structure paraphylétique observée pour le gène ITS2 est mieux résolue et il est possible d'identifier quatre « ensembles » : (1) populations de 9°50 N et 7°25S, (2) la population 21°33S, (3) la population 37°S et (4) un dernier groupe constitué d'un mélange de séquences provenant des 3 populations.

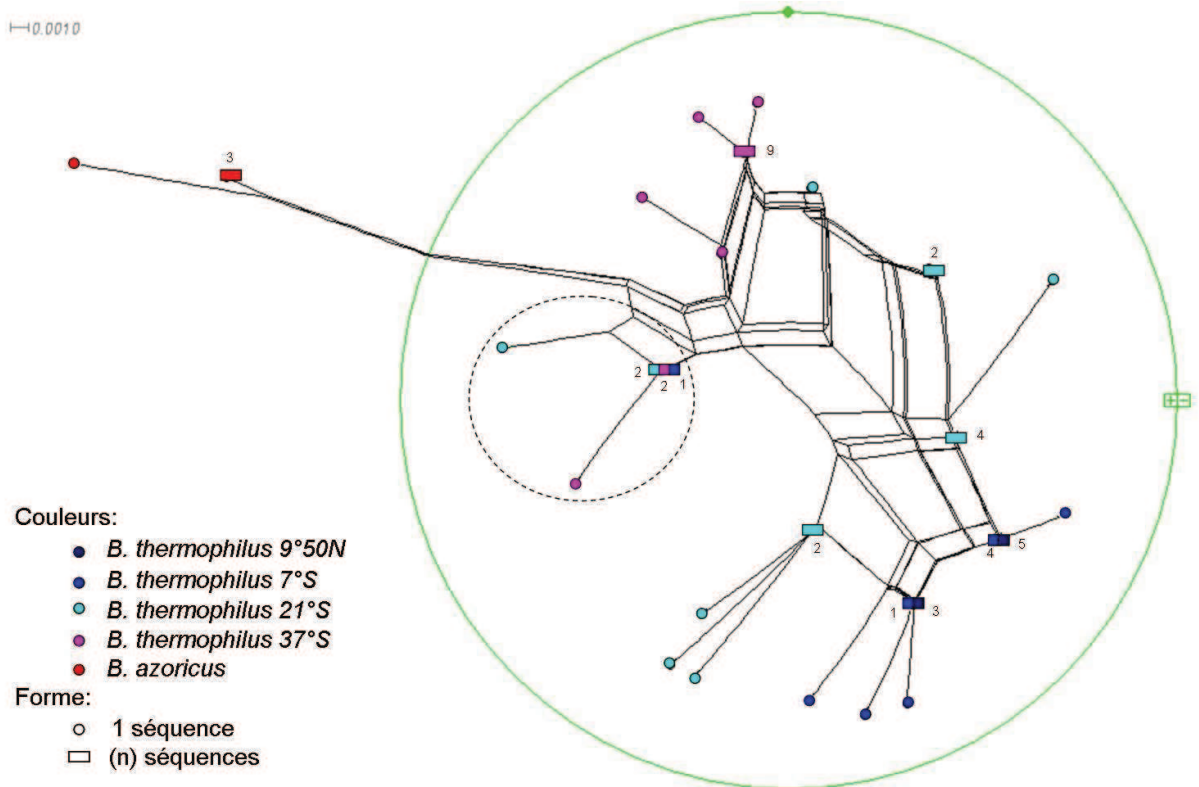


Figure IV.6. Représentation des séquences d'ITS2 dans un réseau en NeighborNet réalisé avec le logiciel Splitstree (Huson & Bryant 2006). Les séquences sont les mêmes que celle utilisées pour l'arbre de coalescence de la Figure IV.5. La légende des étiquettes est indiquée sur la figure. Le grand cercle vert correspond à la fonction 'loupe' de Splitstree permettant de zoomer localement sur le réseau. Le cercle en pointillés correspond au clade basal dans lequel sont retrouvées des séquences de trois populations du Pacifique (7°25S, 21°33S et 37°S).

La structure observée dans le réseau pour le gène ITS2 (Figure IV.6), nous permet de d'observer que le groupe constitué des séquences 'basales' de l'arbre n'est pas situé en position intermédiaire entre les clades et n'est donc pas issu de séquences recombinantes.

Il est intéressant de noter qu'il s'agit d'un groupe d'allèles qui est conservé entre les différentes populations de l'EPR (7°25S, 21°33S et 37°S). Il peut s'agir d'une lignée ancestrale qui n'aurait pas été perdue.

Le gène de l'EF1 α présente lui aussi une séquence qui branche à un endroit assez particulier dans l'arbre. Cette séquence se situe de manière intermédiaire entre les séquences de 37°S et celle de *B. thermophilus*. La position de cette séquence fait penser à un recombinant entre les allèles des deux espèces (position intermédiaire et branche courte) mais celui-ci n'est pas détectée par les logiciels DNAsp et RDP (Recombination Detection Software) (Martin & Rybicki 2000) (données non montrées). Nous avons donc observé chacune de ces mutations pour connaître son statut (recombinant ou lignée ancienne). La séquence de cet allèle (appelé « recombinant ») a été comparée aux séquences consensus de chacune des trois espèces : *B. thermophilus* (9°50N, 7°25S et 21°33S car pas de divergence entre les 2 populations), *B. aff. thermo* 37°S, et *B. azoricus*. Les mutations observées ont été classées en cinq classes selon qu'elles étaient caractéristiques d'une ou plusieurs populations (cf Figure IV.7).

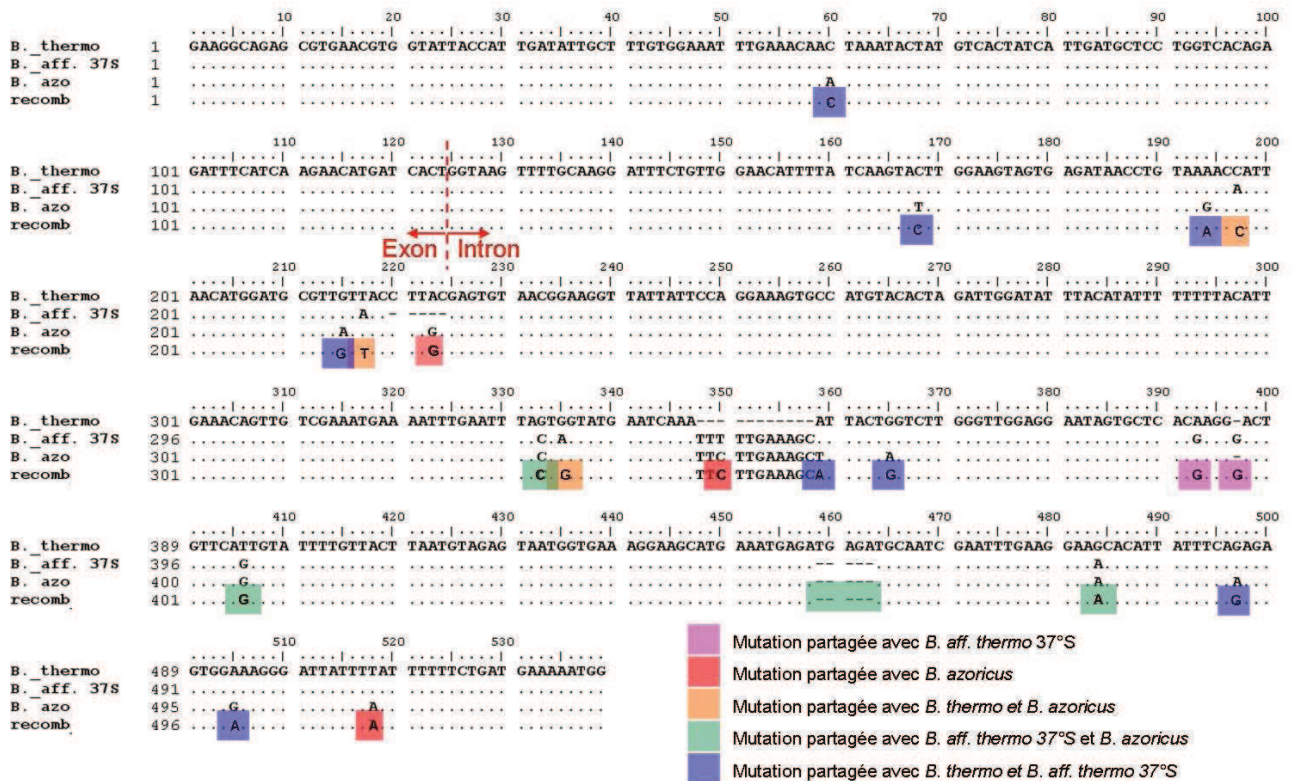


Figure IV.7. Comparaison entre l'allèle (appelé 'recomb') situé en position intermédiaire dans la coalescence du gène EF1 α et les séquences consensus de chacune des 3 espèces (*B. thermophilus*, *B. aff. thermophilus* et *B. azoricus*). Le nucléotide présent pour chaque site ségrégeant est indiqué.

La structure de cet allèle n'est pas caractéristique d'un recombinant simple (formé par uniquement deux allèles). Cet allèle ne possède pas de mutation diagnostique (non retrouvée

dans les autres populations ou espèces), mais présente des mutations caractéristiques de chacun des grands ensembles étudiés (*B. thermophilus*, *B. aff. thermophilus* et *B. azoricus*). A l'exception d'une mutation (synonyme, à 124bp), toutes les autres mutations de l'allèle divergent sont situées dans l'intron. Il est étonnant d'observer dans cet allèle du pacifique des mutations qui semblent caractéristiques de l'espèce atlantique (*B. azoricus*) et qui ne sont retrouvées dans aucune autre population du Pacifique (position 223, 350, 518).

Seuls les locus SAHH, SULFO et ITS2 permettent de distinguer une structure géographique nette entre les populations de l'espèce *B. thermophilus*. Le marqueur mitochondrial COI présente quant à lui une structure un peu particulière avec deux grands clades dans lesquels se répartissent l'ensemble des haplotypes mitochondriaux. Le premier clade regroupe des séquences provenant de l'ensemble des populations de l'EPR (9°50N, 7°25S, et 21°33S) alors que le deuxième clade est composé uniquement d'haplotypes provenant des populations « nord » (9°50 et 7°25S). Les deux clades observés ne présentent pas le même niveau de diversité nucléotidique : le clade « nord » présente une très faible diversité : 4 haplotypes uniques forment le polymorphisme ($\theta = 0,0018$, et $\pi = 0,0004$), alors que le 2^e clade est plus ramifié et diversifié (de 5 à 8 fois plus diversifié : $\theta = 0,0092$, et $\pi = 0,0031$). Cette distribution suggère qu'un balayage sélectif récent a pu avoir lieu au niveau de l'haplotype trouvé uniquement au nord de la dorsale.

Inversement, les deux clades peu divergents observés pour les gènes LYSO, GF1A et GF1B, ne résultent pas d'une structure géographique car les allèles provenant des populations 7°25S et 21°33S sont distribués équitablement entre les clades. En considérant le cas particulier du COI, cette divergence à trois locus nucléaires pourrait cependant suggérer qu'un début d'isolement allopatrique suivi d'un contact secondaire ait pu avoir lieu entre l'EPR nord et l'EPR sud.

Si l'on observe la profondeur des coalescences au niveau de certains arbres, il faut noter que le gène EF1 α présente une topologie en étoile ou en râteau regroupant les trois populations de *B. thermophilus*. Cette topologie caractérise une faible diversité allélique. Ce constat est conforme à celui décrit comme un effet légèrement délétère des mutations sur l'intron de ce gène puisque ce type de topologie est également décrit sur les espèces atlantiques (Faure et al. 2007).

IV.4. Discussion

L'objectif de cette étude était double. Le premier de ces objectifs était de déterminer si les populations de *B. thermophilus* présentaient une structure géographique le long de la dorsale. En effet, de précédentes études (réalisées par Lutz et al. 1980, Craddock et al. 1995b), ont évoqué l'hypothèse que les larves planctoniques de *Bathymodiolus thermophilus* étaient capables de se disperser sur de grandes distances (> 2370 km). Dans le cas de Lutz et al. (Lutz et al. 1980), cette hypothèse est basée sur la taille des prodissoconques I et II sur les coquilles des individus adultes et partiellement vérifiée sur l'espèce atlantique *B. azoricus* pour laquelle Dixon et al. (Dixon et al. 2006) ont montré que la ponte et le recrutement étaient séparés par plus de quatre mois. Dans le cas de Craddock et al. (1995b), les auteurs ont utilisé deux types de marqueurs (allozymes et haplotypes mitochondriaux) pour suggérer qu'il n'y avait pas de barrière aux flux géniques entre les populations du nord de l'EPR et les populations de la dorsale des Galápagos. En revanche, les travaux plus récents de Won et al. (2003b) couplant eux aussi des allozymes aux séquences du gène mitochondrial COI mettent en évidence un isolement partiel de la population de 17°S/EPR avec celles situées plus au nord, mais révèlent aussi l'existence d'une barrière à la dispersion entre les sites 17°S et 32°S. Cette barrière a ensuite été confirmée pour l'ensemble de la faune hydrothermale par Guinot & Hurtado (2003) et Hurtado et al. (2004).

Le deuxième objectif était d'allier une étude de fréquences alléliques à une étude plus novatrice de coalescence pour déterminer si les espèces *B. thermophilus* et *B. aff. thermophilus_37°S* sont capables de s'hybrider et si oui, savoir jusqu'où l'introgession des gènes peut remonter vers le nord (7°25S et 9°50N).

IV.4.1. Structure génétique des populations

Les données de diversités alléliques et les variations de fréquences des allèles dans les populations sont de très bons outils pour décrire la structure des populations le long de la dorsale EPR et constituent parfois dans le cadre d'un polymorphisme neutre, une bonne estimation indirecte mais informative de la dispersion (Kinlan & Gaines 2003).

L'analyse des fréquences alléliques entre populations de *B. thermophilus* montre clairement qu'il n'y a pas de barrière nette aux flux géniques le long de la dorsale et confirme en partie les résultats obtenus par Craddock et al. (1995b). La distribution des fréquences

allozymiques en fonction de la situation géographique des sites sur la dorsale ne présente pas de clines de fréquences comme ce pourrait être le cas pour une zone de tension entre espèces s'hybridant (Barton & Hewitt 1985) ou pour des locus sous sélection (Oakeshott et al. 1982). Aucune inversion de fréquence (pouvant être associés à des phénomènes d'hybridation) n'est détectée comme cela avait été observé au niveau de la zone de tension de Broken Spur (29°N) sur la dorsale médio-atlantique (O'Mullan et al. 2001). En revanche, l'analyse des indices de différenciation, qu'ils soient estimés sur l'ensemble des populations ou calculés entre paires de populations révèle l'existence d'une structure génétique significative associée à des différences observées au niveau des populations les plus extrêmes du gradient latitudinal. La littérature fait référence à un courant transverse à l'axe de la dorsale pacifique au niveau de l'équateur et qui pourrait affecter la dispersion larvaire de bivalves *Calymene magnifica* et *B. thermophilus* (Won et al. 2003b, Hurtado et al. 2004). Cette hypothèse a été évoquée pour expliquer la légère différenciation génétique observée par les auteurs entre les populations du nord et la population de *B. thermophilus* située plus au sud à 17°S. L'histoire tectonique de la dorsale EPR peut également expliquer un isolement allopatrique des populations nord et sud, notamment avec la mise en place de la microplaque de Bauer entre 15°S et l'équateur, il y a 5-6 millions d'années. Cette hypothèse a été mise en avant par Matabos et al. (2008) pour expliquer la présence d'espèces cryptiques sud et nord ayant été remises en contact à 9°50N. Nos analyses suggèrent que le flux génique semble relativement homogène au niveau des populations situées au centre de la dorsale EPR sud et, que seules les populations les plus extrêmes pourraient être différenciées. Ce constat est cependant à nuancer car la population de 7°25S n'a été que faiblement échantillonnée et pourrait donc biaiser les estimations. Les populations potentiellement différenciées présentent soit une diversité allélique plus faible (cas de 7°25S), soit une diversité allélique plus forte (cas de 21°33S) qui pourraient traduire des comportements locaux de la dynamique populationnelle de *B. thermophilus* (apports de migrants génétiquement différenciés ou effets démographiques liés aux processus d'extinction/recolonisation). Les différences trouvées au niveau de la population de 7°25S pourraient également s'expliquer par un ralentissement du flux génique au niveau de la faille transformante de Yaquina située entre 7° et 13°58S. Cependant, l'existence d'une faille transformante de même ampleur (faille de Garret) entre 13°58S et 17°25S ne semble pas constituer un obstacle majeur aux échanges de gènes entre ces dernières populations. Au contraire, les populations provenant de sites géographiquement proches peuvent présenter des niveaux de différenciation plus élevés (7°25S et 13°25) que des populations beaucoup plus éloignées géographiquement (7°25S et 18°36S). Il faut cependant noter que le seul site ne

présentant de différence avec aucun autre site est situé de manière centrale par rapport à l'ensemble des sites échantillonnés (Rehu Marka, 17°25S). De plus, les sites vraiment proches (Oasis, Rehu Marka et Suzie) sont génétiquement homogènes et la différenciation génétique semble augmenter avec la distance. Ces éléments sont en faveur d'une dispersion larvaire de proche en proche suivant un modèle dit des pas japonais. L'isolement par la distance attendu par une dispersion de proche en proche est certainement contrarié par la dynamique spatio-temporelle de l'environnement hydrothermal comme la récurrence des épisodes d'extinction et de recolonisation des sites. Hellberg (1994) a listé plusieurs conditions (depuis la sélection naturelle jusqu'au choix des marqueurs) pour expliquer la difficulté de mettre en évidence un isolement par la distance. Une des conditions concerne particulièrement le milieu hydrothermal : les flux géniques entre populations et la dérive génétique n'ont pas le temps de s'équilibrer. En effet, les sites hydrothermaux, et plus particulièrement ceux situés sur la dorsale est pacifique (dorsale ultra rapide), ne sont jamais réellement à l'équilibre migration/dérive, mais pourraient être au contraire toujours en phase d'expansion géographique ou suivre un modèle de type « pool de migrants » dans lequel le nombre de colonisateurs (recolonisation d'habitats vides) est important et d'origine multiple (contrairement à une dispersion de proche en proche), connus pour gommer la différenciation génétique (Slatkin 1977, Wade & McCauley 1988, Whitlock 1992).

La colonisation des nouveaux sites actifs peut se faire en fonction des larves présentes dans la colonne d'eau au moment opportun et ne sont pas forcément représentatives de la diversité générale de l'espèce. En effet, si l'on suppose que les larves sont exportées à la faveur des panaches hydrothermaux (Mullineaux et al. 1992), les larves présentes au sein de ces panaches vont constituer un sous-échantillonnage de la diversité d'une source unique (un champ hydrothermal donné). Ces panaches vont donc former des ensembles larvaires distincts qui vont être exportés par les courants.

Sachant que les processus de colonisation des sites hydrothermaux suivent une chronologie bien définie liée à l'évolution des conditions environnementales (Tunnicliffe 1991) et aux interactions biotiques (Johnson et al. 1988), le laps de temps permettant la colonisation par des larves de *Bathymodiolus* se trouve être limité. Les larves à même de coloniser un site proviennent donc d'un nombre limité de nuages larvaires présents au bon moment. La saisonnalité de la reproduction larvaire chez les *Bathymodiolus* a été décrite par (Comtet & Desbruyeres 1998) et la discontinuité du recrutement larvaire peut être vérifiée à partir des données biométriques réalisées sur les moulières de la dorsale EPR dans la thèse de

M. Matabos (Matabos 2007). En effet, ces analyses réalisées sur les mêmes moulières que celles d'où proviennent nos échantillons (échantillonnées lors de la campagne Biospedo) montrent une forte disparité dans les tailles d'animaux (absence des petites classes de taille ou des plus grandes cohortes selon les sites) et des recrutements larvaire d'intensité différente entre les sites (3 % de la population est constituée de juvéniles à Oasis et près de 70% de juvéniles sont retrouvées à Susie).

Les conditions que nous venons d'évoquer (fragmentation de l'habitat, taille réduite des populations, extinctions locales et recolonisation, migrations potentiellement orientées...) pourraient expliquer pourquoi les résultats de la différenciation génétique des modioles ne suit pas un attendu théorique en accord avec les modèles d'IBD ou en îles mais forme plutôt une structure en mosaïque le long de la dorsale. Cependant, une telle structure en mosaïque peut aussi être expliquée par des phénomènes d'hybridation et d'introgession, voire par de la sélection sur les allozymes (Oakeshott et al. 1982, Barton & Hewitt 1985)

IV.4.2. Détection de zones de contact secondaire avec hybridation entre deux espèces cryptiques de *B. thermophilus* au sud et au nord de la dorsale EPR.

Suite à l'exploration du segment de dorsale pacifique-antarctique entre 32° et 38S, Won et al. (2003b), sont les premiers à avoir séquencé le gène mitochondrial COI à un niveau populationnel chez le bivalve *Bathymodiolus thermophilus*. Ils ont décrit un haplotype mitochondrial divergent de près de 4% de celui connu pour *B. thermophilus*, au niveau des moulières situées au sud de la microplaque de l'île de Pâques, et concluent à l'existence d'une nouvelle espèce cryptique. Cette divergence sur le gène mitochondrial, couplée à la divergence allozymique supérieure à celle observée entre les deux espèces de la dorsale médio-atlantique (*B. azoricus* et *B. puteoserpentis*), a conduit à penser que les populations de *B. aff. thermophilus* de l'extrême sud pouvait être élevées au rang d'espèce à part entière mais que celles-ci pouvaient avoir gardé la capacité de s'hybrider.

Compte tenu des capacités de dispersion larvaire des bathymodioles et sachant que les espèces de ce genre peuvent s'hybrider localement entre espèces proches (O'Mullan et al. 2001, Olu-Le Roy et al. 2007 et chapitre précédent de la thèse), nous avons cherché à savoir si la structure génétique observée à l'échelle de l'EPR sud et l'augmentation de diversité allélique observée à 21°33S ne pouvaient pas refléter de l'introgession entre ces deux

espèces, ce malgré la barrière quasi-étanche que constitue la microplaque de l'île de Pâques. Il est en effet admis qu'à ce niveau, l'habitat hydrothermal est extrêmement discontinu et restreint à un chapelet de volcans sous-marins tels que le volcan PITO (23°20S-111°38W) : (Naar et al. 2004).

Afin de valider cette hypothèse, nous avons privilégié une approche de coalescence des allèles sur plusieurs gènes nucléaires avec des données de séquences. Pour pouvoir détecter une éventuelle introgression d'allèles à l'aide de séquences, il faut réunir plusieurs conditions : (1) que la séparation des espèces en allopatrie soit suffisamment ancienne pour générer une divergence nette entre les localités de 21°33S et 37°S pouvant confirmer celle obtenue à partir des données mitochondriales et allozymiques, (2) qu'il existe bien une zone d'hybridation entre les deux dorsales permettant un échange d'allèles entre espèces, et (3) pouvoir définir correctement des clades d'allèles spécifiques de chaque espèce pour estimer l'intensité de l'introgression (en fréquence et en distance géographique) de manière incontestable à partir de la zone d'hybridation.

L'acquisition de séquences provenant des sites les plus au nord (9°50N et 7°25S) va également nous permettre de tester l'hypothèse d'une zone de contact secondaire avec d'autres espèces de *Bathymodiolus* pouvant provenir de la dorsale de Galápagos ou de Juan de Fuca. En effet, de nombreux auteurs (Hurtado et al. 2002, Hurtado et al. 2004, Won et al. 2003b, Matabos et al. 2008) ont montré la présence d'espèces vicariantes dans cette région et suspectent la présence d'une barrière au flux génique permettant la formation d'espèces en allopatrie. Même si cette barrière a engendré une divergence plus faible entre les espèces que celle créée de part et d'autre de l'île de Pâques, nos données devraient permettre d'en confirmer la présence.

IV.4.3. Mise en évidence d'un processus de spéciation allopatrique avec formation d'un contact secondaire récent (le cas de la microplaque de l'île de Pâques).

L'arbre obtenu à partir des séquences du gène COI et, dans une moindre mesure, ceux des gènes EF1 α et ITS, montrent clairement des monophylies réciproques entre la population de 37°S et celle de *B. thermophilus* à 7°25S et 21°33S. En se basant sur ces trois gènes, il semble donc que la population située sur la dorsale pacifique-antarctique soit bien à une espèce géographique distincte mais proche de *B. thermophilus*.

Cependant les arbres de plusieurs gènes nucléaires regroupent dans les mêmes clades des allèles provenant des trois populations et des deux espèces. Ce type de topologie peut être interprété de deux manières : il peut s'agir (1) d'un gène qui présente un faible taux d'évolution et ne permet pas de discriminer correctement les allèles des deux espèces, ou (2) d'une introgression d'allèles entre espèces suite à de l'hybridation. Dans l'hypothèse du faible taux d'évolution pour certains gènes, cette évolution ne peut être considérée comme neutre que si aucune structure ni branche longue n'existe à ce locus. A partir du moment où au moins une lignée est divergente ou qu'une structure populationnelle apparaît, l'hypothèse neutre n'est plus valable. L'hypothèse neutre n'est donc pas valable pour expliquer nos observations mais la manque de divergence observé pourrait résulter d'effets sélectifs. Pour valider la deuxième hypothèse concernant l'introgression, il est nécessaire d'utiliser un groupe externe et de comparer les divergences entre les différents gènes. Si cette divergence est identique pour les gènes à monophylie réciproque et pour les gènes qui présentent des mélanges d'allèles entre espèces, il est possible de conclure à de l'introgression. Les gènes GF1_A, GF1_B et LYSO sont dans la situation où il n'est pas possible de distinguer de clades entre *B. thermophilus* et *B. aff thermophilus* « 37°S » car les séquences des trois populations sont réparties de manière homogène entre les clades. Les divergences qui existent entre les séquences « pacifiques » et les séquences de *B. azoricus* (1,9%, 2,9% et 5,7% ; respectivement pour GF1_A, GF1_B et Lyso) encadrent celle obtenue pour l'EF1 α (2,4%) pour lequel deux clades sont nettement divergents entre les espèces « thermophilus » et « 37°S », suggérant ici que l'explication la plus parcimonieuse est celle de l'introgression d'allèles entre les deux espèces.

Pour le gène EF1 α , il est facile de distinguer qu'un allèle diagnostique de 37°S a introgressé la population de 21°33S, ce qui nous permet d'affirmer que l'introgression depuis le sud vers le nord est possible. L'arbre obtenu pour le gène SAHH permet d'affiner l'histoire évolutive des deux espèces et confirme l'hypothèse d'une orientation de l'introgression vers le nord. Pour ce gène en effet, les séquences provenant des deux populations de l'espèce *B. thermophilus* appartiennent à des clades clairement divergents, alors que les séquences provenant de l'espèce *B. aff thermophilus* (37°S) se regroupent uniquement avec les séquences de la population de 21°33S.

Dans ce cas, on peut émettre l'hypothèse que le clade regroupant les séquences de 21°33S et de 37°S soit un clade totalement introgressé par des allèles provenant de l'espèce *B. aff thermophilus* 37°S. En effet, cette hypothèse permet d'expliquer la divergence observée entre les populations 7°25S et 21°33S alors que le taux d'évolution de ce gène n'est pas plus

élevé que celui des autres gènes (3,6% de divergence avec *B. azoricus*). De plus, la divergence observée de 2,9 % entre les 2 clades « 7°25S » et « 21°33S » correspond plutôt aux valeurs de divergence observées entre *B. thermophilus* et *B. sp_37°S*, plutôt qu'à une valeur de différenciation entre les clades d'une même espèce.

Le scénario selon lequel un allèle de 37°S aurait complètement balayé la variabilité présente dans la population de 21°33S semble difficile à expliquer avec une introgression neutre car aucun des autres locus étudié ne présente ce phénomène. Il pourrait donc y avoir un avantage sélectif à l'allèle de 37°S lui permettant d'envahir complètement la population de 21°33S. Cette hypothèse semble confirmée par le résultat négatif et significatif du test du D de Tajima à ce locus. Cependant, la profondeur des coalescences montre que ce clade a déjà regagné de la diversité car il ne s'agit pas d'un râteau net comme suite à un balayage sélectif récent (Galtier et al. 2000).

Peu d'éléments nous permettent de confirmer la symétrie des flux géniques à partir des arbres de coalescence. En effet pour conclure à de l'introgression, il faut dans un premier temps observer la présence de clades caractéristiques de chacune des deux espèces. Les seuls arguments pouvant soutenir l'hypothèse d'une introgression du nord vers le sud sont donnés par les arbres des GF1_A et GF1_B. Dans un premier temps, il faut admettre que les séquences très divergentes (trouvées à 21°33S) représentent le clade *B. aff. thermophilus* (introgression sud vers nord). Dans ce contexte, les séquences de 37°S retrouvées dans les clades *B. thermophilus* peuvent être considérées comme des séquences introgressées de *B. thermophilus* vers la population de *B. aff. thermophilus* (introgression du nord vers le sud).

IV.4.4. Influence de l'introgression sur la différenciation génétique des populations de *B. thermophilus*

La plus forte diversité nucléotidique observée dans la population de 21°33S par rapport aux populations de 7°25S et de 9°50N est probablement liée à la présence d'allèles introgressés provenant de l'espèce de 37°S (au moins pour les locus EF1 α , GF1_A, GF1B et ITS). De même, l'effet de l'introgression de quelques allèles de 37°S dans la population de 21°33S conduit à l'obtention d'un nombre de sites ségrégeants nettement plus élevé que dans les autres populations.

En revanche, il est étrange que les valeurs du D de Tajima soient plus élevées (et majoritairement positives) dans la population de 7°25S alors qu'elles sont systématiquement négatives dans la population de 21°33S. En effet d'un point de vue démographique, des valeurs positives de ce test et l'excès d'allèles en fréquence intermédiaire sont liés à l'existence de populations structurées (Tajima 1989a). On aurait donc eu tendance à attendre de telles valeurs pour des populations présentant de l'introgession d'allèles comme celle de 21°33S. En effet si l'on observe les courbes de mismatch des deux populations (7°25S et 21°33S) pour les gènes présentant le plus de différences entre les valeurs obtenues pour le D de Tajima (LYZO, GF1_A et GF1_B), les populations de 21°33S semblent les plus structurées Figure IV.8. Pour chacun de ces gènes, on observe un pic de fréquence pour un nombre de différences plus élevé. C'est donc la population de 21°33S qui possède les allèles les plus divergents, vraisemblablement des allèles introgressés de 37°S. Même pour des locus dont les valeurs du D de Tajima sont similaires ($EF1\alpha$), les allèles de la population de 21°33S sont plus différenciés et l'on remarque des pics de fréquence correspondant à l'allèle introgressé (à 7bp) et à la séquence intermédiaire (à 5bp) que nous avons diagnostiqués précédemment. Les gènes GF1_A, GF1_B et Lyso de 7°25S présentent un pic de fréquence pour des valeurs de différenciation intermédiaires mettant en évidence une différenciation dans les allèles qui les composent.

L'existence d'une population structurée à 7°25S pourrait s'expliquer par l'existence d'une deuxième zone de contact secondaire avec les populations situées sur l'EPR nord et, séparées de l'EPR sud par la barrière bathymétrique et hydrodynamique du point triple à l'équateur (*The Hess Deep*). Cette barrière a d'ailleurs été largement évoquée pour expliquer les niveaux de divergence observés sur le marqueur COI entre les populations sud et nord chez les polychètes *Alvinella pompejana* et *Branchiopolynoe symmytilida* (Hurtado et al. 2004) et *Bathymodiolus thermophilus* (Won et al. 2003b). A cet égard, il existe également une zone de contact secondaire avec hybridation dans cette région entre deux espèces cryptiques du gastéropode patelliforme *Lepetodrilus elevatus* (Matabos et al. 2008, Plouviez et al. en préparation). L'existence d'une zone de contact secondaire entre des populations qui auraient été isolées beaucoup moins longtemps que celles situées de part et d'autre de la microplaque de l'île de Pâques pourrait expliquer pourquoi nous trouvons une richesse allélique plus grande et un isolement génétique plus élevé à 7°25S au niveau des allozymes, la présence de deux haplotypes mitochondriaux divergents et la présence des deux lignées divergentes au gène SAHH à l'origine du D significativement positif dans cette même population.

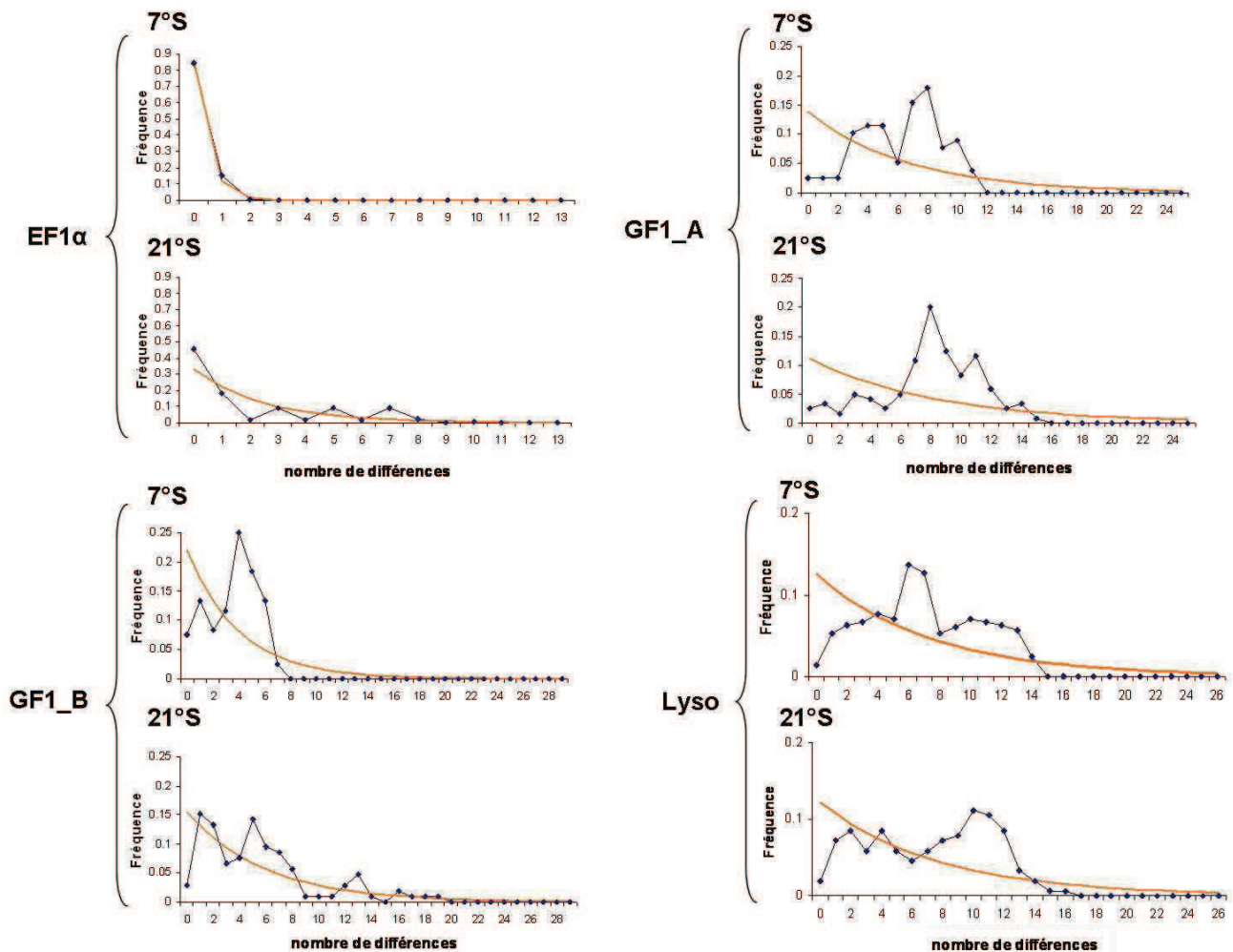


Figure IV.8. Comparaison des distributions de mismatch entre les populations de 7°S et 21°S pour les gènes EF1, GF1_A, GF1_B et Lyso. Courbe bleue avec des marques : distribution observée ; courbe orange : distribution attendue théorique.

Du côté de la population 21°S, 8 gènes présentent des valeurs négatives et 4 de ces valeurs s'écartent significativement de l'attendu neutre. Même s'il a été montré que l'intron de l'EF1 α était sous sélection purifiante (Faure et al. 2007) il est difficile de penser que les 7 gènes (avec des introns pour la plupart) puissent être tous sous sélection purifiante. En revanche, les effets démographiques peuvent conduire à ce genre de valeurs en affectant la majorité des locus. Cet effet pourrait être d'autant plus fort si l'on considère que l'introgression d'allèles en provenance de *B. aff. thermophilus* devrait s'opposer à l'effet démographique en augmentant le nombre d'allèles (sites ségrégeants) en fréquences intermédiaires pour certains gènes (cas du locus EF1 α par exemple).

Les valeurs de différenciation obtenues par le calcul des Φ_{st} par paires présentent une différence significative entre les populations de 21°33S et de 7°25S. Ces données sont en accord avec les résultats de différenciation estimée à partir des données allozymiques. Les valeurs de différenciation obtenues pour la comparaison entre les populations de 21°33S et de 37°S, montrent une différence significative pour l'ensemble des locus à l'exception du SAHH, (le gène potentiellement introgressé totalement au sud). Il faut cependant rester précautionneux dans la prise en compte de ces résultats car les allèles de 37°S étaient très nettement sous-représentés (maximum quatre provenant de deux individus seulement) et il n'est pas exclu, voire très probable que les populations de la dorsale pacifique/antarctique soient aussi introgressées à partir d'allèles issus de *B. thermophilus*.

IV.4.5. Zone d'hybridation : contact secondaire entre espèces vicariantes ou colonisation de la dorsale EPR par une espèce issue du Pacifique Ouest ?

Il y a 5,9 MA, au niveau de la latitude actuelle de l'île de Pâques, la dorsale est-pacifique était constituée de deux dorsales distinctes séparées par une forte discontinuité géologique ayant pu jouer un rôle de barrière géographique (Figure IV.9). A l'inverse, la mise en place progressive de la microplaque générée par la rotation de ces tronçons est à l'origine de la formation d'un arc volcanique ayant pu favoriser la remise en contact des faunes par des pierres de gué (Naar & Hey 1991).

Dans l'hypothèse d'un point de rencontre entre deux dorsales n'ayant pas d'histoire géologique commune, il est possible d'imaginer qu'une espèce située plus au sud puisse être à même de coloniser la dorsale EPR à travers le passage que constitue l'arc volcanique récemment mis en place. Dans ce cadre, la microplaque constituerait une zone de tension entre une espèce localement implantée et une espèce colonisatrice. Cependant, cette hypothèse est peu vraisemblable pour deux raisons : (1) l'analyse génétique d'au moins cinq espèces distinctes (vestmentifères, polychètes, bivalves et crabes) indique que les peuplements trouvés de part et d'autre de la microplaque ont une histoire commune et ont été obtenus par un effet vicariant (Guinot & Hurtado 2003, Hurtado et al. 2004, Won et al. 2003b) et (2) en prenant en compte la phylogénie sur l'ensemble des bathymodiolinae les espèces *B. thermophilus* et *B. aff thermophilus* constituent des espèces sœurs, génétiquement et morphologiquement parlant. Il semble donc que deux lots de populations appartenant à une population ancestrale commune aient pu diverger en allopatrie par la mise en place d'une

barrière physique de type faille transformante et que la formation de la microplaque ait permis leur remise en contact secondairement et la mise en place de flux géniques inter-espèces. De telles situations ont été décrites à de nombreuses occasions dans le milieu marin (Denis et al. 1993, Palumbi 1994, Knowlton 2000).

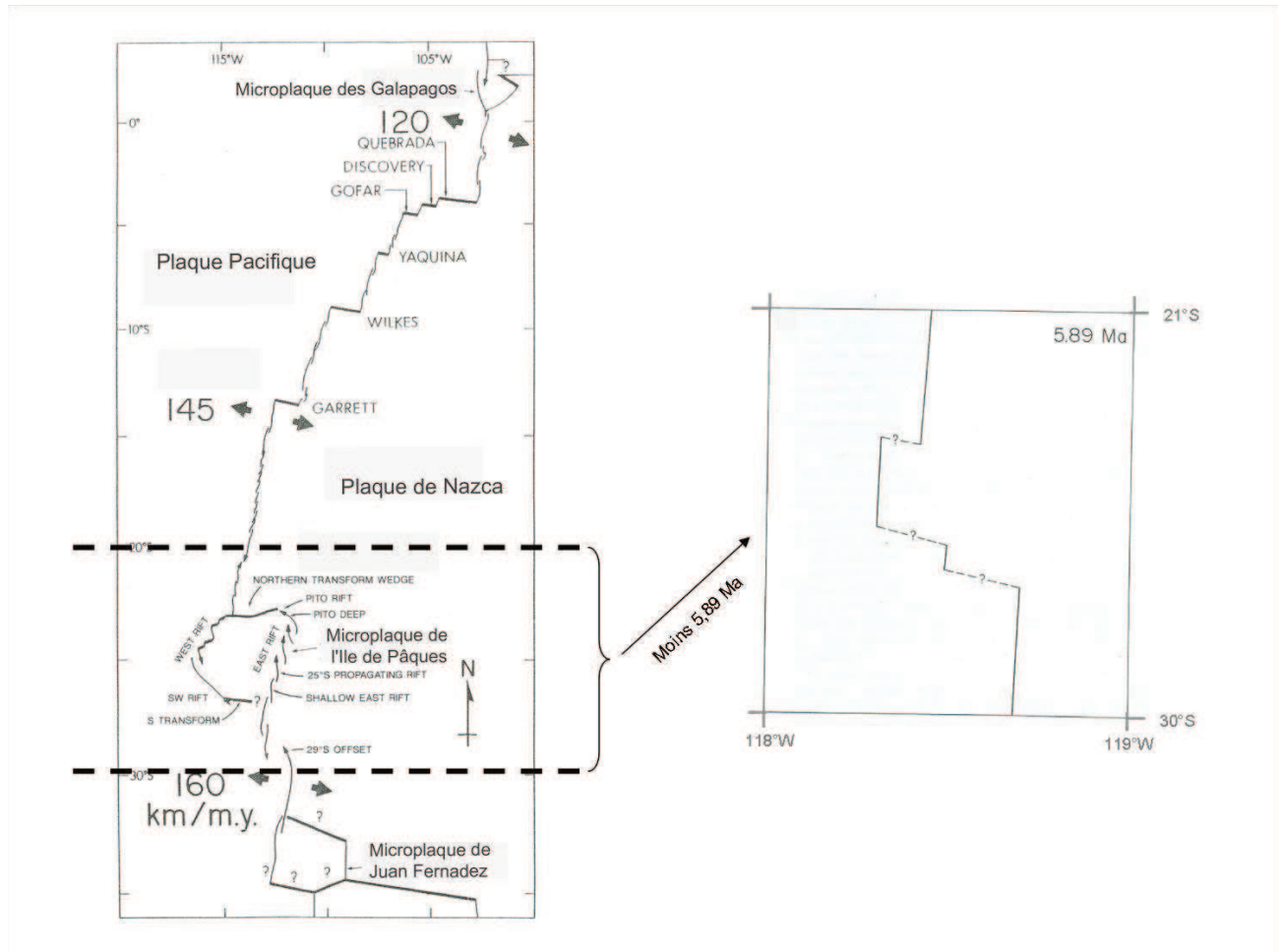


Figure IV.9. Evolution tectonique de la région de l'Île de Pâques. A gauche, dorsale actuelle et à droite, structure supposée de la dorsale il y a 5,89 Ma. Les traits correspondent aux dorsales actives, les chiffres correspondent aux vitesses d'expansion en km par million d'années, les pointillés sur le dessin de droite correspondent à des anomalies magnétiques détectées, pouvant correspondre à des failles. Illustrations modifiées de (Naar & Hey 1991).

Dans l'hypothèse d'un contact secondaire, l'espèce ancestrale a été divisée en deux populations qui ont évolué en allopatrie. Les niveaux de divergence observés dans les coalescences (de l'ordre de 1,5% en moyenne pour l'ensemble des loci et l'importance des flux de gènes entre les deux espèces sont en accord avec l'hypothèse d'une période relativement courte d'isolement géographique et nous permettent de retenir l'hypothèse de spéciation par vicariance.

IV.4.6. Dynamique de la colonisation de la dorsale.

Les arbres de coalescences obtenus pour les huit gènes étudiés peuvent également être utilisés pour retracer l'histoire de la colonisation sur la dorsale EPR entre 9°50N et 21°33S par l'espèce *Bathymodiolus thermophilus*.

Les hypothèses à confronter sont la dispersion sur de grandes distances, la dispersion en pas japonais (de proche en proche), avec isolement par la distance, ou une hypothèse alternative de différenciation lié au 'surf' de certains allèles lors d'une phase d'expansion (Excoffier & Ray 2008).

Dans le cas de la première hypothèse favorisant le modèle en îles, c'est-à-dire avec une dispersion sur de grandes distances, les populations doivent s'homogénéiser sous l'influence des apports larvaires répétés provenant théoriquement de l'ensemble des populations. Cet attendu est en accord avec les données allozymiques qui ont montré qu'il n'existait pas de barrière physique à la dispersion entre les populations nord et sud de la dorsale. L'argument principal que nous pouvons opposer à cette hypothèse, est cependant que les populations sont faiblement structurées aux deux extrêmes du gradient latitudinal, cette structure pouvant être due à la dispersion des allèles introgressés provenant de la zone de tension située au niveau de la microplaque de l'île de Pâques. En effet, sous l'hypothèse d'un flux génique à longue distance, les allèles introgressés caractéristiques de la population de 37°S pourraient se retrouver dans les populations éloignées. Aucun allèle introgressé n'est retrouvé à 9°50N, ni même à 7°25S pour les gènes pour lesquels ces populations sont différenciées de celle du sud de la dorsale (EF1, SAHH, ITS, CO1 et SULFO). Pour certains gènes cependant, (LYSO, GF1_A et GF1_B) les allèles provenant des trois populations de la dorsale sont mélangés, ce qui semble refléter une non différenciation des populations à ces gènes plutôt qu'une introgression complète à grande échelle.

Le fait que certains locus mélangent des allèles provenant de populations distantes alors que d'autres locus sont divergents pour ces mêmes populations n'est pas un argument suffisant pour exclure l'introgression. En effet la théorie prédit l'existence de gènes verrous dans le génome des espèces proches. Dans ce cas, l'introgression des gènes sera d'autant plus faible qu'ils sont proches des gènes verrous et que la zone d'hybridation est récente. Cette

hypothèse serait donc en accord avec l'hétérogénéité de l'introgession observée entre les différents locus.

Dans le cas de l'hypothèse d'une dispersion de proche en proche, nous devrions observer un isolement par la distance. Cet isolement par la distance, même s'il n'est pas mis en évidence de manière statistique pour les *Bathymodiolus* de dorsale EPR, ne peut cependant pas être exclu. Sous cette hypothèse, on s'attend à ce que les populations les plus distantes (21°33S contre les populations de 9°50N ou 7°25S soient différenciées). Cependant, lorsque l'on observe nos données de séquences, quatre gènes (LYSO, EF1, GF1_A et GF1_B) ne présentent pas de structure nette entre le nord et le sud. En revanche, les gènes COI, ITS2 et SAHH présentent une structuration nette entre le nord et le sud en accord avec une dispersion limitée de l'espèce. Dans le cas du gène SAHH, nous avons évoqué la possibilité que la divergence observée entre les populations 7°25S et 21°33S reflète une introgession totale des allèles issus de 37°S dans cette dernière population. Cependant, pour les deux autres gènes (COI et ITS2), les individus de la population de 37°S constitue un clade d'allèles nettement divergent qui n'est pas retrouvé dans les populations EPR. En conséquence, la différenciation génétique observée entre les populations nord et sud de l'EPR ne semble pas être influencée par l'introgession à ces locus. Il convient d'ailleurs de noter que la paraphylie observée pour les séquences ITS oriente la colonisation du sud vers le nord. A partir de ces données, nous pouvons donc émettre l'hypothèse qu'il existe trois « ensembles » de Bathymodiolinae sur la dorsale. Le premier de ces ensembles est constitué des populations du nord (9°50 et 7°25S) et se distingue des populations du sud au niveau des gènes du COI, ITS2 et SAHH. Le deuxième ensemble est composé de la population de 21°33S qui, en fonction des gènes, présente soit une homogénéité avec les populations du nord, soit une forte homogénéité avec la population de 37°S (cf. SAHH). Le troisième ensemble correspond à la population de 37°S, qui peut être considérée comme une espèce à part entière de part le degré de divergence observé avec *B. thermophilus* aux locus COI, EF1 α , ITS2.

L'hypothèse alternative pouvant influencer les fréquences alléliques dans nos populations de *B. thermophilus* est celle proposée par (Excoffier & Ray 2008). Ces auteurs montrent que les fréquences de certains allèles minoritaires peuvent augmenter jusqu'à se fixer lors des épisodes d'expansion par simple effet de dérive (*gene surfing*) (Figure IV.10). Un tel processus d'expansion conduit à la formation de populations différenciées complexes pouvant mimer un pattern de processus adaptatif.

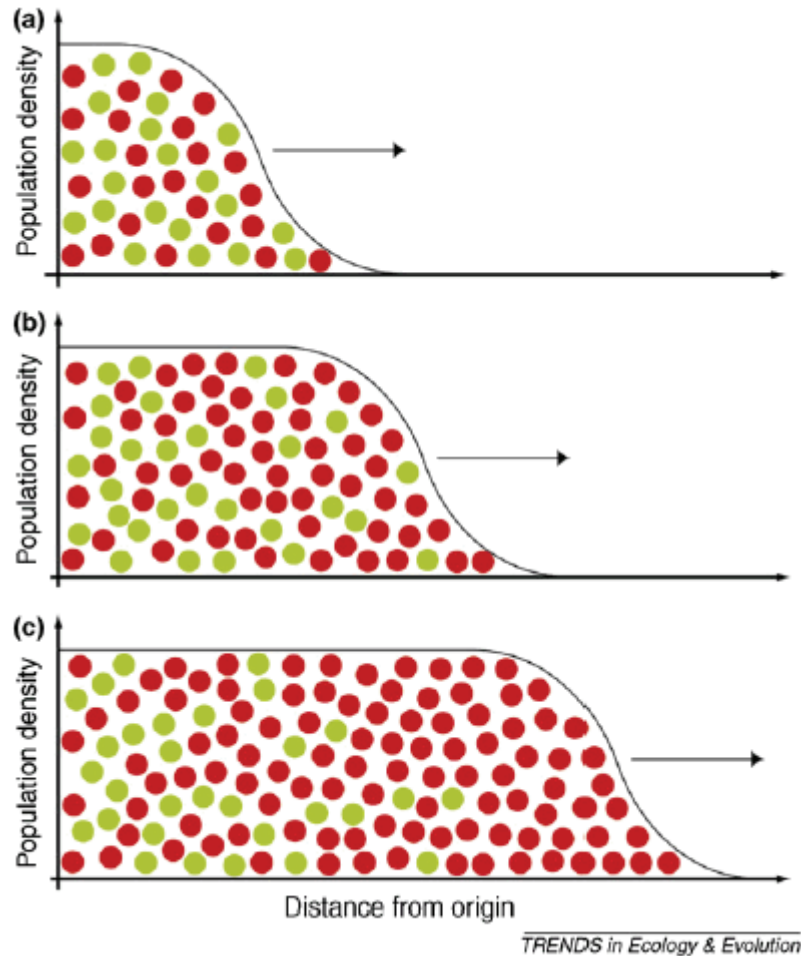


Figure IV.10. Dérive génétique se produisant sur la vague frontale d'une population en expansion, entraînant un changement dans les fréquences alléliques et son glissement ('surf'). (a) Conditions initiales avec équi-proportions des 2 allèles (rouges et verts) dans une population en expansion géographique. (b) l'allèle rouge se trouvant par chance au sommet de la vague en (a) augmente en fréquence à la faveur de la dérive locale. (c) L'allèle rouge se fixe par dérive sur le front de la vague et colonise les populations qui ne présenteront plus que cet allèle (tiré de Excoffier & Ray 2008).

Ces observations nous conduisent à proposer plusieurs scénarios de colonisation de la dorsale du Pacifique Est :

A partir du COI qui présente 2 haplotypes au nord mais un seul au sud, trois hypothèses peuvent être dégagées : (i) la première hypothèse est que les deux haplotypes du nord sont des haplotypes ancestraux mais que seul l'un des deux a réussi à coloniser le sud de la dorsale. Ce type d'hypothèse nécessite des pressions de sélections différentielles entre les deux haplotypes, (peu probable car les mutations fixées entre les deux haplotypes sont synonymes) ou un effet de dérive par surf de cet allèle lors d'une phase de colonisation (ii) la deuxième hypothèse est celle d'une colonisation du sud vers le nord. L'haplotype 'Sud' celui rencontré actuellement sur toute la dorsale aurait colonisé le nord de la dorsale et l'haplotype

'Nord' serait apparu récemment. Cette hypothèse a le mérite d'expliquer la faible diversité observée pour le l'haplotype 'Nord', mais il est toujours difficile d'expliquer comment cet allèle a pu se fixer en sympatrie. (iii) la troisième et la plus vraisemblable des explications est celle d'une double colonisation. L'haplotype 'Nord' est apparu dans les populations du nord suite à un isolement géographique de type péripatrique autour de l'équateur (la faible diversité allélique serait liée à la plus petite taille de pop ancestrale après isolement) alors que l'haplotype 'Sud' aurait évolué dans le sud. A la faveur d'une (re)colonisation du nord de la dorsale, l'haplotype majoritaire se serait retrouvé en mélange dans les populations suite à cette remise en contact. Ce scénario permettrait d'expliquer la présence des deux lignées divergentes au gène SAHH au sein de la population 7°25S.

IV.5. Conclusion

Cette étude combinant une approche allozymique avec une analyse de séquence nous a permis de mettre en lumière ce nombreux éléments incertains ou inconnus concernant les populations de *Bathymodiolus sp* de la dorsale Est Pacifique.

_Les populations de la dorsale sont structurées au niveau local mais la tendance à un isolement par la distance n'est pas significative. La structuration des populations résulterait plutôt de l'interaction entre une dispersion sur une distance limitée et les conditions environnementales locale liées à l'hydrothermalisme (potentialité d'extinction et de re-colonisation des sites)

_La présence d'un courant transversal à 15°S, si elle est avérée, ne perturbe en rien la dispersion et les flux géniques de part et d'autre de celle latitude.

_Les effets démographiques (expansion) semblent affecter la population de 21°33S située dans la région la plus instable de la dorsale rapide qu'est l'EPR.

_La population de 37°S est bien une espèce à part entière.

_Des phénomènes d'hybridation et d'introgession existent entre l'espèce *B. thermophilus* et l'espèce de *B. sp*_37°S mais ne semblent pas s'étendre jusqu'à 7°25S.

_L'hybridation rencontrée entre les deux espèces dans le sud de la dorsale résulte d'un contact secondaire entre des espèces proches qui se sont formées par spéciation en allopatrie à la suite de la formation de l'île de Pâques.

_l'introggression semble se faire préférentiellement du sud vers le nord.

_Les hypothèses de colonisation de la dorsale EPR les plus vraisemblables nécessitent une colonisation de la dorsale depuis le sud (au moins 21°33S) vers le nord. Il est probable qu'une espèce vicariante ait existé (ou existe toujours) dans le nord de la dorsale.

Chapitre V. Processus d'évolution moléculaire de quelques gènes.

V.1. Introduction

V.1.1. Evolution des gènes et de leurs fonctions.

V.1.1.a. Duplication des gènes

Les duplications de gènes sont souvent considérées comme un moyen nécessaire pour engendrer des nouveautés évolutives que ce soit en termes de structure ou de fonction de la molécule (Sidow 1996). Les duplications peuvent intervenir à différentes échelles sur le génome. En effet, elles peuvent concerner une portion de gène, un gène complet, tout ou une partie d'un chromosome ou même le génome complet. Les processus qui nous intéressent particulièrement dans cette thèse sont les duplications de gène et plus particulièrement les duplications dites « en tandem ». Les processus par lesquels les gènes évoluent après une première duplication complètement redondante jusqu'à l'acquisition d'une nouvelle fonction, sont encore assez méconnus. La théorie suggère trois alternatives au devenir des gènes dupliqués : (1) une copie va devenir silencieuse en accumulant des mutations et en dégénéralisant (pseudogénéralisation ou non-fonctionnalisation), (2) une copie va conserver sa fonction originelle et sera conservée par la sélection naturelle, tandis que l'autre copie va acquérir une nouvelle fonction ou affiner celle-ci (néo-fonctionnalisation), (3) les fonctionnalités des deux copies vont être restreintes par l'accumulation de mutations mais la fonctionnalité totale sera analogue de celle du gène ancestral mono-copie (sous-fonctionnalisation) (Lynch & Conery 2000).

Le temps estimé entre l'apparition du gène dupliqué et sa désactivation par accumulation de mutations délétères est considéré comme assez court (quelques millions d'années voire même beaucoup moins). Il suffit en effet d'une mutation entraînant l'apparition d'un codon stop pour que l'activité du gène soit cassée et qu'il finisse par être éliminé (Watterson 1983, Lynch & Force 2000). Cependant, chez les bathymodiales, comme

en général chez tous les bivalves marins qui possèdent une très grande taille efficace, le temps d'attente avant l'apparition d'une mutation légèrement favorable dans l'une des deux copies est réduit et celles-ci vont pouvoir se fixer plus fréquemment dans les populations. En revanche, si une copie a atteint une certaine fréquence par dérive, elle engendrera par mutations des copies défectueuses dont l'élimination sera d'autant plus longue que la taille efficace est grande. Il est donc attendu que le nombre de copies défectueuses (pseudogènes non-fixés) soit plus important chez nos espèces du fait de leur forte taille efficace. L'étude de Lynch & Conery (2000) met en évidence un taux de duplication très élevé de 1% / gène / million d'années. Au tout début de leur histoire, la plupart des gènes dupliqués présentent une courte phase où les pressions de sélection sont relâchées et, certains d'entre eux évoluent même de manière presque neutre pendant une plus ou moins longue période. Cependant, la grande majorité des duplicats non soumis à la sélection vont disparaître en quelques millions d'années (demi-vie d'un pseudogène estimée à 4 millions d'années) (Lynch & Conery 2000). L'acquisition de nouvelles fonctions à partir d'une duplication liée à un relâchement de la sélection ou à l'acquisition d'une mutation adaptative semble assez rare, ce qui fait qu'une duplication est un processus rarement invoqué dans l'évolution adaptative d'un gène. Cependant, le séquençage et l'analyse des génomes permet de détecter de plus en plus d'événements de ce type (Zhang et al. 1998, Zhang 2003). D'autres processus comme l'« *exon shuffling* » ou la recombinaison inter-locus sont plus fréquemment documentés comme vecteurs permettant à une fonction d'évoluer vers une autre. Il apparaît cependant intéressant de considérer que chez les espèces à très grandes tailles efficaces, les mutations potentiellement adaptatives sont plus fréquentes et la durée de vie des duplicats plus longue. Cette double constatation nous amène à penser que les duplications pourraient avoir un rôle beaucoup plus important dans l'évolution des gènes, notamment en termes de néofonctionnalité (Duda & Palumbi 1999, Bailly et al. 2003).

L'accumulation de polymorphisme et les pressions de sélection qui sont discutées ici et permettent l'acquisition de nouvelles fonctions ou au contraire qui conduisent à la formation de pseudo-gènes, concernent uniquement les mutations qui agissent sur les parties codantes (exoniques) des gènes. Pourtant les régions codantes ne sont pas les seules régions à pouvoir être affectées par les phénomènes de sélection. Les exemples de régions qui ne soient pas codantes pour des acides aminés, mais qui sont soumises à la sélection (intronique ou régulatrice, par exemple) sont de plus en plus nombreux.

V.1.1.b. Pressions de sélections dans les régions introniques

La recherche de polymorphisme 'adaptatif' à partir de l'analyse de séquences a débuté, il y a plus d'une vingtaine d'années, avec l'étude du locus de l'alcool-déshydrogénase chez *Drosophila melanogaster* (Kreitman 1983). Jusqu'à aujourd'hui, le débat à propos de la signification adaptative du polymorphisme observé n'a jamais réellement cessé (Lewontin 1991). Il est en effet très difficile de faire la part des choses entre des effets démographiques passés, l'empreinte de la sélection indirecte sur le polymorphisme neutre par un balayage sélectif et l'effet direct des pressions de sélection agissant sur certains polymorphismes dans des séquences codantes ou non-codantes (Glinka et al. 2003, Haddrill et al. 2005b, Li & Stephan 2006). L'idée est que les processus démographiques affectent le génome dans son intégralité alors que les processus sélectifs agissent sur les locus de manière spécifique (Hudson et al. 1987, Galtier et al. 2000). L'effet d'un allèle sous sélection positive est connu pour modifier la fréquence des mutations qui lui sont liées mimant ainsi l'effet d'un goulot d'étranglement de la population (Fu & Li 1993). La magnitude de cet effet appelé auto-stop génique (Maynard Smith & Haigh 1974) dépend de l'intensité de la sélection sur la mutation, du taux de recombinaison avec les sites neutres adjacents et dans une moindre mesure de la taille efficace de la population (Kaplan et al. 1989). Une perte de diversité génétique peut aussi indiquer un effet direct de la sélection dirigée contre les mutations délétères, aussi appelé background selection (Tajima 1989b, Charlesworth et al. 1993). Dans les régions exoniques par exemple, la fréquence des mutations non-synonymes qui affecte la nature de la protéine sera inférieure à celle des mutations synonymes (Fay & Wu 2001). De ce fait, la plupart des polymorphismes non-synonymes (notamment lorsque les mutations sont des singletons) sont légèrement délétères (Eyre-Walker et al. 2006) puisqu'une mutation « adaptative » sera rapidement fixée dans la population. Il a également été mis en évidence que de légers effets de sélection pouvaient affecter les mutations synonymes (Akashi 1995) et les régions non codantes et/ou introniques (Leicht et al. 1995, Jaruzelska et al. 1999). En effet, les introns sont connus pour posséder de nombreuses régions régulatrices, entraînant leur conservation par sélection purifiante (Hardison 2000, Bergman & Kreitman 2001, Haddrill et al. 2005a).

Dernièrement, Andolfatto (2005) a analysé l'accumulation des mutations dans le polymorphisme de *Drosophila melanogaster* à l'aide du D de Tajima pour différentes catégories de mutations (synonymes, non synonymes et silencieuses dans des régions non

codantes variées), et a suggéré qu'une partie non négligeable du polymorphisme non codant était sous sélection légèrement négative.

V.1.2. Objectifs

Contrairement à l'approche multi-locus abordée précédemment, ce chapitre a pour objectif d'analyser, de manière plus précise, certaines portions de gènes d'intérêt présentant une évolution par duplication (*Hsp70*, *GF1*) ou des écarts à l'évolution neutre laissant supposer l'action d'une sélection directe non attendue ou indirecte comme l'autostop génique (intron 2 du gène *EF1 α*). Les deux premières portions de gènes étudiées, celles des gènes HSP 70 et SULFO, présentent une histoire évolutive complexe (phylogénie non résolue entre les espèces proches), trop ambiguë pour que nous puissions utiliser ces gènes sans risque de biaiser nos estimations de flux de gènes inter-spécifiques. Le troisième gène sur lequel nous avons porté une attention particulière est le gène GF1 en l'analysant d'une manière plus globale en regroupant les deux paralogues GF1_A et GF1_B. Le dernier gène ayant retenu notre attention est l'intron 2 du gène *EF1 α* . Dans ce cas précis, nous avons pu mettre en évidence que celui-ci est sous sélection purifiante et que les mutations qui y apparaissent sont légèrement délétères.

V.2. Recherche d'un polymorphisme ancestral balancé au niveau du gène de la sulfotransférase

V.2.1. Présentation des données

Ce gène ne présente pas de divergence inter-spécifique nette entre les espèces atlantiques d'une part et les espèces pacifiques d'autre part. Il semble même présenter un niveau de divergence intra-spécifique supérieur à celui observé entre les espèces pacifiques et atlantiques.

Toutes les séquences retenues pour l'analyse du gène de la sulfotransférase (SULFO) sont des séquences issues de captures multiples pour lesquelles l'allélisme ne fait aucun doute. Cependant, ce nombre d'allèles 'robustes' est relativement limité. Pour conduire des analyses détaillées, il est préférable de posséder le jeu de données le plus conséquent possible. Pour se faire, tous les allèles provenant d'individus hétérozygotes ont été conservés lorsqu'ils avaient été recapturés au moins deux fois. Au total, 67 séquences provenant de 53 individus

ont été utilisées dans les analyses (soit une moyenne de 1,26 séquences par individu) (cf. Tableau V.1 et Figure V.1).

Tableau V.1. Séquences utilisées dans l'analyse du gène de la sulfotransférase

Espèces	Population	nombre d'individus recapturés	Nombre de séquences corrigées conservées
<i>B. azoricus</i>	Menez Gwen	12	15
	Rainbow	12	17
<i>B. puteoserpentis</i>	Logatchev	11	13
<i>B. thermophilus</i>	7°S	8	10
	21°S	8	9
<i>B. aff. thermophilus</i>	37°S	2	3
total:		53	67

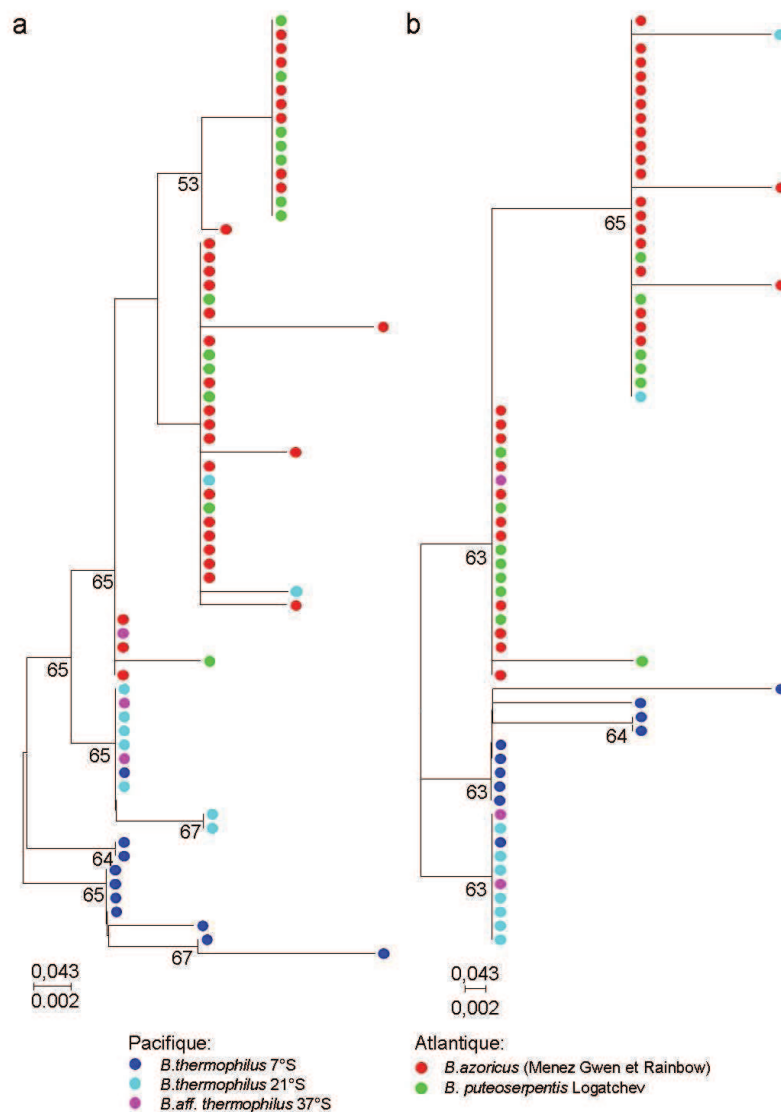


Figure V.1. Arbres obtenus par Neighbour-Joining avec 1000 bootstraps de ré-échantillonnages en bootstrap. **(a)** : Séquences nucléotidiques regroupées à partir de la distance de Tamura & Nei (1993); **(b)** : Séquences traduites en acides aminés et regroupées selon la matrice de JTT. Pour l'échelle : les chiffres au dessus de la ligne correspondent au nombre de mutations ; en dessous de la ligne distance calculée d'après les tables de JTT.

V.2.2. Analyses de diversité

Comme aucun site n'apparaît fixé entre les espèces d'un même océan, les analyses de diversité ont été réalisées en considérant toutes les populations d'un même océan comme une seule et même espèce. Les analyses de diversité ont également été réalisées en distinguant le polymorphisme synonyme et non synonyme sur l'exon analysé (séquence codante = 100% de la séquence SULFO) (Tableau V.2).

Tableau V.2 : Descripteurs de la diversité nucléotidique, obtenus à l'aide du logiciel DNAsp sur les séquences de la sulfotransférase. **N** : nombre de séquences, **L** : longueur des séquences utilisées, **S** : nombre de sites polymorphes, **Hd** : diversité allélique, π : estimateur de $4N_e\mu$ à partir du nombre moyen de différences nucléotidiques par site Nei (1987), θ_w : estimateur du $4N_e\mu$ par paire de bases en utilisant le nombre de sites polymorphes (Watterson 1975), **D** de Tajima : estimateur de Tajima (1989a). Le niveau de significativité est indiqué par : * ($P < 0,05$) ; ** ($P < 0,01$) ; *** ($P < 0,001$), **K** : différence nucléotidique moyenne entre séquences. π_s et π_N : respectivement : diversité nucléotidique pour les polymorphismes synonyme et non-synonyme.

	N	L	S	Hd	π	θ_w	D de Tajima	k	Synonymes		Non synonymes			
									nb sites	nb mut.	π_s	nb sites	nb mut.	π_N
Atlant.	45	216	7	0,658	0,0055	0,0074	-0,6936	1,190	46,69	2	0,0112	169,31	5	0,0040
Pac.	22	216	10	0,844	0,0099	0,0127	-1,0304	2,15	46,29	3	0,0173	169,71	8	0,0079

Les analyses de diversité nous indiquent qu'en dépit d'un plus faible nombre de séquences pour l'Océan Pacifique (23 pour 45 dans l'Atlantique), la diversité nucléotidique de l'espèce pacifique est près de deux fois supérieure à celle de l'espèce Atlantique. La valeur du D de Tajima obtenue au niveau de l'espèce pacifique est négative et inférieure à celle trouvée chez l'espèce atlantique, sans pour autant être significative. La différence nucléotidique moyenne (k) est également près de deux fois plus élevée pour l'espèce du Pacifique. Tous ces résultats sont en accord avec une diversité plus élevée dans le Pacifique qui pourrait être le reflet d'un effectif efficace plus important ou d'un mélange de populations différenciées. Une autre hypothèse que nous pouvons évoquer est celle du maintien d'un polymorphisme ancestral entre les deux espèces.

Les distributions des valeurs de la diversité nucléotidique π et du D de Tajima sont présentées dans la Figure V.2.

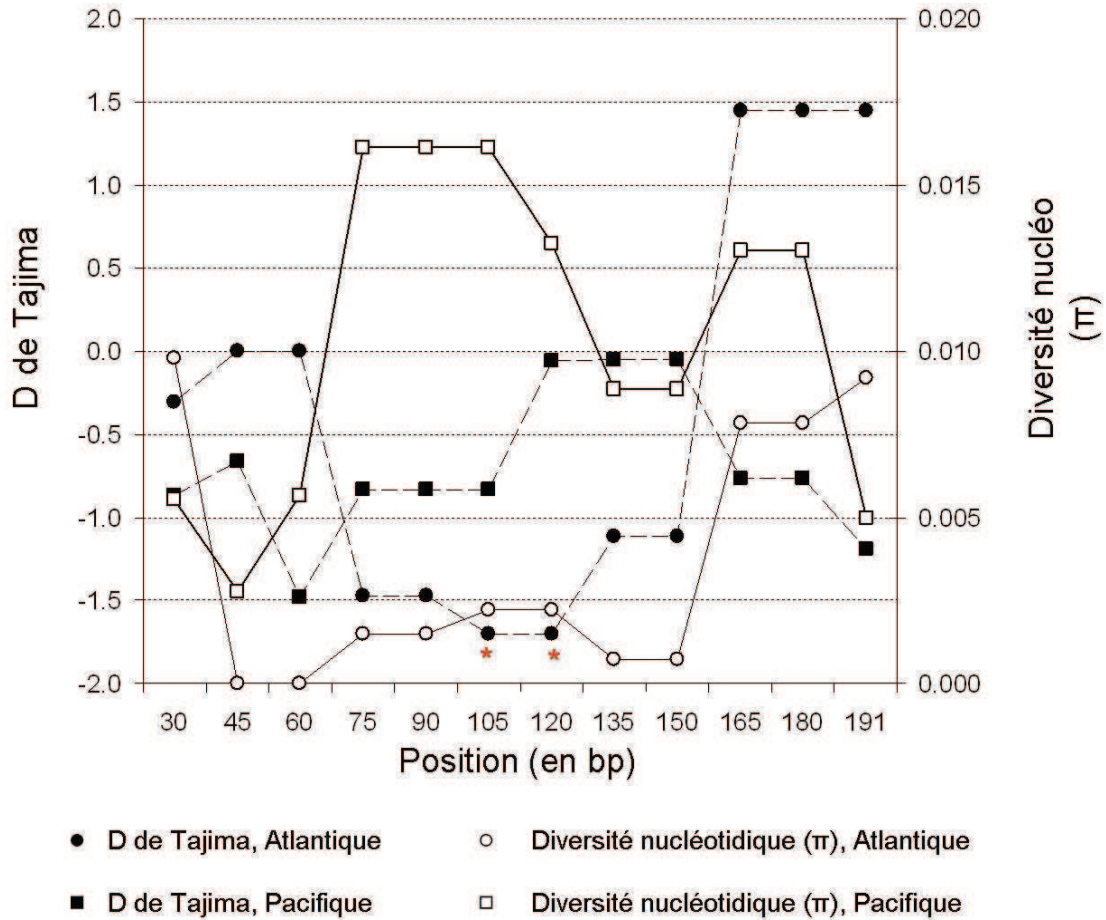


Figure V.2. Evolution des valeurs du D de Tajima et de la diversité nucléotidique π le long du gène de la sulfotransférase entre les deux espèces de Bathymodiolus à l'aide d'une fenêtre coulissante de 60 bp échantillonnées avec un pas de 15 bp. * : niveau de significativité de D ($P < 0,05$)

La limite de ce type de représentation est que les fenêtres coulissantes de petite taille peuvent induire de manière artificielle les distributions chaotiques que l'on observe. De même il faut considérer avec réserve les valeurs négatives et significatives pour D. En effet, afin de valider de manière statistique cet écart à l'évolution neutre, il faudrait utiliser des tests plus précis (pour des données non indépendantes). Compte tenu des éléments précédant, les interprétations qui suivent sont donc à considérer avec précaution.

Ce graphique montre que les indices de diversité semblent proposer une évolution opposée entre les deux espèces (la diversité est relativement élevée le long des séquences dans le Pacifique alors qu'au contraire, elle est très faible dans l'Atlantique pour n'augmenter qu'à la fin du gène). De même pour le D de Tajima, les valeurs significativement négatives augmentent fortement sur la fin de la séquence pour être fortement positives dans l'Atlantique ($D = 1,45$) alors que les valeurs de D diminuent progressivement le long du gène pour le

Pacifique ($D=-1,19$). Cette augmentation des valeurs de π et de D à la fin du gène pour l'espèce atlantique pourrait constituer un premier indice sur le maintien d'un polymorphisme ancestral en fin de gène.

V.2.3. Test de détection d'un effet sélectif : le test de Mc Donald et Kreitman

Ce test a pour principe de comparer l'accumulation de mutations synonymes et non-synonymes dans le polymorphisme et la divergence de deux espèces. Sur notre jeu de données, aucune mutation n'est fixée dans la divergence. Néanmoins, un site apparaît quasiment fixé entre les deux espèces. A l'inverse, un nombre non négligeable de mutations ségrègent dans le polymorphisme pouvant laisser supposer qu'un polymorphisme balancé pourrait exister (deux sites polymorphes non-synonymes partagés entre les espèces). Nous avons donc réalisé un test MK en éliminant trois séquences : deux séquences de 21°S et une de 37°S, afin de créer artificiellement une divergence non-synonyme entre les deux espèces (Tableau V.3). En toute rigueur cette analyse n'aurait pas pu être faite à cause du nombre d'observations trop faible (et donc du manque de divergence). Elle est cependant présentée ici pour illustrer la démarche scientifique utilisée lors de la recherche d'évènements sélectifs.

Tableau V.3. Résultats du test de Mc Donald et Kreitman sur les données du gène de la Sulfotransférase

	S	NS
divergence	0	1
polymorphisme	5	8

test exact de Fisher: 1,000 (NS)

Le test de Mc Donald et Kreitman est non significatif. Bien que nos données indiquent qu'il y a beaucoup plus de mutations accumulées dans le polymorphisme des espèces qu'entre les deux espèces, les deux lignées semblent donc avoir évolué de manière neutre.

V.2.4. Discussion sur l'évolution du gène de la sulfotransférase chez les bathymodiales.

Les analyses effectuées ne permettent pas de statuer sur l'existence d'un polymorphisme ancestral maintenu par sélection balancée comme l'aurait pu laisser soupçonner (1) l'absence de divergence inter-spécifique, (2) la présence de sites polymorphes non-synonymes entre espèces et (3) l'accumulation de mutations uniquement dans le

polymorphisme des espèces. L'analyse des données confirme cependant la plus forte diversité génétique observée dans l'Océan Pacifique. L'évolution du gène n'a pas été complètement homogène entre les océans et montre des profils opposés du point de vue de la distribution de la diversité le long du gène (sous réserve de la robustesse des analyses en fenêtre coulissante).

Le niveau de conservation du gène entre les quatre espèces de *Bathymodiolus*, constitue un point remarquable. Il est en effet étrange de constater que ce gène est extrêmement peu conservé chez les eucaryotes au point que nous n'avons pas réussi à trouver de groupe externe satisfaisant pour notre étude. Malgré l'utilisation de séquences du même gène (DOPA/ tyrosine sulfotransférase) chez *Lotia gigantea*, un mollusque gastéropode dont le génome a été séquencé, la structure primaire est tellement peu conservée que des indels apparaissent dans l'alignement de séquence, et ce, malgré le fait qu'il s'agisse d'une séquence codante (données non montrées). On peut émettre l'hypothèse que le fort niveau de conservation du gène au travers des différentes espèces de Bathymodiales reflète soit (1) de la sélection purifiante forte, soit (2) une spéciation très récente des espèces atlantiques et pacifiques. Cette dernière hypothèse n'est toutefois pas en accord avec les données obtenues sur les autres gènes : divergence nette et profonde entre les espèces. Une autre hypothèse aurait pu être évoquée pour expliquer le manque de divergence entre les espèces atlantiques et pacifiques : celle de l'introgession des espèces. Cependant compte tenue de leur habitat (dans des océans séparés), cette hypothèse a donc été écartée.

V.3. Mise en évidence d'un effet légèrement délétère du polymorphisme intronique sur le gène EF1 α chez le bivalve hydrothermal profond *Bathymodiolus*.

Cette étude a fait l'objet d'un article (présenté dans les annexes de la thèse). Ici, seul un résumé détaillé sera présenté.

Résumé

Les analyses multi-locus réalisées lors de la thèse ont fait apparaître un comportement particulier du gène EF1 α (faible diversité allélique et nucléotidique, topologie des clades de l'arbre en râteau) et ce, alors même qu'il s'agit d'une séquence intronique. Une étude particulière a donc été réalisée sur ce gène.

Cette étude a été menée en comparant le gène EF1 α à deux autres gènes : le gène mitochondrial de la Cytochrome-c Oxydase 1 et le gène nucléaire du Lysozyme. Deux espèces ont été utilisées pour cette analyse : *B. azoricus* (population de Rainbow) dans l'océan Atlantique et *B. thermophilus* (population de 7°25S) dans l'océan Pacifique.

Le polymorphisme observé pour les trois gènes est très variable, mais le locus EF1 α se distingue par un polymorphisme de 3 à 5 fois inférieur à celui des autres gènes. De plus, chez ce gène, la diversité nucléotidique (π) est très similaire entre la région intronique (412 bp) et la région exonique (127 bp).

Les données de divergence entre les deux espèces varient aussi entre les gènes, de 0,02 pour l'EF1 α à 0,12 pour le COI. Cette divergence est composée pour l'essentiel de mutations silencieuses (introniques) pour les gènes EF1 α et LYSO, et de mutations synonymes (80%) pour le gène COI.

La faible divergence observée pour le gène EF1 α nous indique que la vitesse d'évolution est trois fois inférieure à celle du lysozyme et six fois inférieure à celle du COI. Cette faible vitesse d'évolution est observée même lorsque l'on utilise uniquement l'intron, écartant ainsi tout effet sélectif sur la protéine elle-même.

La topologie de l'arbre du locus EF1 α est également très particulière avec des branches courtes qui coalescent rapidement vers l'ancêtre commun, alors que les deux autres gènes présentent des arbres plus ramifiés comprenant plusieurs lignées alléliques.

Les valeurs obtenues pour le D de Tajima sont négatives et très significatives pour les deux espèces de *Bathymodiolus* (-2,776, $P < 0,001$, et -2,679, $P < 0,001$). Ces valeurs restent significatives même en analysant uniquement la région intronique. Il est surprenant que les valeurs de D soient aussi fortement significativement négatives et avec des valeurs aussi proches alors que les espèces ont évolué de manière indépendante au niveau de leur généalogie. Les valeurs de significativité du test de Tajima ont également été testées à l'aide du logiciel HKA (disponible sur la page web de Jody Hey : <http://lifesci.rutgers.edu/~hey/HeylabSoftware.htm>) et sont toujours très nettement significativement différentes de zéro.

Les résultats obtenus pourraient refléter trois phénomènes : un effet démographique, un effet indirect de la sélection (balayage sélectif), ou un effet de sélection purifiante sur l'intron.

- L'effet démographique serait le résultat d'un goulot d'étranglement au niveau des populations des 2 espèces et aurait dû affecter l'ensemble des locus de la même manière. Dans ce cas, le locus mitochondrial qui possède une taille efficace plus faible (haploïde) et qui ne recombine pas (Ballard & Rand 2005, Bazin et al. 2006) devrait être plus sensible aux effets démographiques (Hudson & Turelli 2003). Il n'en est rien. L'absence d'écarts significatifs à l'accumulation neutre des mutations dans le polymorphisme des deux espèces pour le gène mitochondrial et le gène *LYSO* permet donc d'écarter l'hypothèse d'un effet démographique.

- L'effet indirect de la sélection correspond au résultat d'un balayage sélectif sur les allèles. Ce genre de processus affecte les locus indépendamment. Pour expliquer nos données, il faudrait que deux balayages sélectifs se soient produits de manière synchrone et récemment dans les 2 espèces, lesquelles vivent dans des océans différents. Cette situation est fortement improbable du fait de la rareté des épisodes de balayage sélectif.

- La dernière hypothèse considère que le faible taux d'évolution observé au niveau de l'intron pourrait être lié à de la sélection purifiante. En effet, nos résultats ne sont pas en accord avec l'attendu selon lequel la diversité nucléotidique intronique est plus élevée que celle de l'exon du fait du relâchement des pressions de sélection (Li & Graur 1991). Cette dernière hypothèse nous semble la plus parcimonieuse. Il a déjà été montré, dans de nombreux cas, que le faible taux d'évolution des introns pouvait être dû à des contraintes sélectives en rapport avec une fonction ou une structure spécifique de la zone non codante (Jaruzelska et al. 1999, Young & dePamphilis 2000, Parsch 2003).

Nous pouvons donc conclure que le second intron du gène *EF1 α* est sous sélection légèrement purifiante. Ces observations réalisées à partir de 3 gènes lors de l'écriture de l'article semblent encore d'actualité même après avoir acquit et analysé les séquences d'une dizaine de gènes. L'effet démographique est définitivement écarté et aucun autre gène ne présente la structure en râteau telle que celle observée pour l'*EF1 α* et aucun autre locus ne présente de niveau de diversité nucléotidique aussi faible que celui observé pour l'intron de l'*EF1 α* .

V.4. Rôle de la duplication dans l'évolution de la protéine de choc thermique (Heat Shock Protein 70 : HSP70)

Ce gène n'a pas été utilisé dans les analyses des chapitres précédents car il appartient à une famille multi-génique pouvant présenter une évolution par duplication en tandem (Requena et al. 1988). Ce type d'évolution est à même de 'brouiller' le signal phylogénétique entre espèces. Les premières séquences obtenues ont fait apparaître un manque d'évolution concertée laissant supposer que certaines lignées aient pu évoluer de manière séparée. Le très faible niveau de divergence entre lignées paralogues et l'absence d'une phylogénie d'espèces claire au sein de chaque lignée paralogue ne nous a pas permis de conserver ce gène pour l'analyse multi-locus des flux géniques inter-espèces. Ce gène a donc fait l'objet de l'analyse spécifique présentée ci-dessous, dans l'espoir d'éclaircir la situation. Dans cette partie nous n'aborderons pas les questions de duplication intra-individuelle ni la variation du nombre de copies entre les individus, les populations ou les espèces. Ce type d'approche nécessiterait une analyse génomique plus fouillée et des approches de PCR quantitative, southern blot...

V.4.1. Description de la famille des HSP 70.

Les HSPs sont connues pour être une famille de protéines dont la numérotation est liée au poids moléculaire (en KDa) de la molécule (HSP 10, 23, 40, 60, 70, 90, 110). Les séquences de HSP 70 de nombreuses espèces de mollusques (dont celles du génome séquencé de l'Aplysie) et de quelques autres phylums ont été obtenues à partir des banques de gènes en BlastX. L'alignement de ces séquences en acides aminés et leur regroupement sous la forme d'un arbre en Neighbor-Joining (matrice JTT) nous ont permis de montrer que dans notre jeu de séquences, les espèces de *Bathymodiolus*, comme l'Aplysie, présentent au moins deux gènes distincts dans la famille des HSP 70 : HSP 70A et HSP 70B (sans compter le HSC 70 obtenu chez l'Aplysie) (Figure V.3). Cet arbre ne vise pas à étudier la paralogie du gène HSP 70 à travers les espèces, mais uniquement à positionner les deux séquences obtenues chez les espèces de *Bathymodiolus*. Les séquences des deux gènes sont bien distinctes entre elles et présentent des fixations d'acides aminés suffisantes pour éviter de les confondre dans l'analyse d'un polymorphisme.

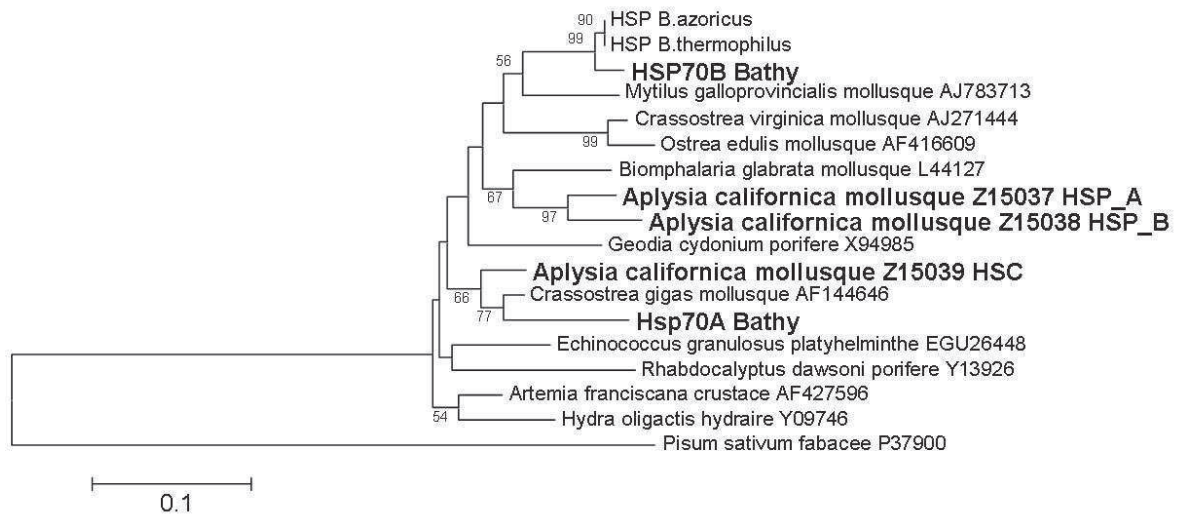


Figure V.3. Arbre en NJ réalisé avec les séquences traduites de HSP 70 récupérées dans les banques. Les distances sont calculées à partir de la matrice de JTT. L'arbre a été testé par 1000 bootstraps.

Des amorces ont été redéfinies afin de n'amplifier que le gène HSP 70 dit 'B'. C'est ce gène qui a été utilisé par la suite dans l'étude du polymorphisme car il présente lui-même une diversification allélique par duplication au sein des espèces de bathymodiales sans que l'on puisse savoir avec exactitude si ces événements sont postérieurs ou antérieurs à la spéciation Atlantique/Pacifique. Les séquences présentées dans l'arbre correspondent aux séquences des ADN complémentaires issus de la banque ESTs de *Bathymodiolus azoricus* (Menez Gwen). A ces séquences ont été ajoutées deux séquences (HSP *B. azoricus* et *B. thermophilus*) amplifiées au cours de la thèse sur l'ADN génomique d'un individu de l'espèce *B. thermophilus* et d'un individu de l'espèce *B. azoricus*, ces séquences faisant partie de notre jeu de données sur ce gène. De part leur position dans l'arbre, ces séquences valident la spécificité des amorces utilisées.

V.4.2. Présentation du jeu de données et des arbres de coalescence

Dans un premier temps, un arbre a été réalisé à partir de l'ensemble des séquences obtenues (361 séquences corrigées) et a été enraciné avec la séquence de *Mytilus galloprovincialis* (AJ783713) la plus proche de nos séquences (Figure V.4a). De nombreuses séquences ont souvent été recapturées pour un même individu (jusqu'à 12 séquences pour deux individus de la population de *B. azoricus* sur le site Rainbow), et permettent de distinguer 8 « clades » au sein du gène (Figure V.4b) (seuls les clades IV, V et VIII sont supportés par des valeurs supérieures à 50% après 1000 bootstraps). Dans tous les cas, un

nombre maximal de 4 clades a été retrouvé pour les individus recapturés un grand nombre de fois, laissant supposer que ce gène a évolué récemment par duplication. La répartition des allèles dans les différents clades a été étudiée au sein de chaque individu pour la totalité des individus (126 pour lesquels des séquences ont été recapturées) toutes espèces confondues. Les résultats sont résumés sous la forme d'un graphique (Figure V.5). Sur 126 individus, 48 ont été recapturés une seule fois. Ces séquences se répartissent aléatoirement dans chacun des clades et sont donc peu informatives. En ce qui concerne les individus pour lesquels plusieurs séquences ont été recapturées, on constate que ces séquences se répartissent dans les 4 clades pour chaque espèce. Pour au moins 4 séquences recapturées, plus de deux clades sont toujours trouvés. L'existence de 4 clades différents est avérée sur deux individus, l'un provenant du site Rainbow et, l'autre, du site Logatchev car les 4 séquences récupérées sont clairement distinctes et tombent chacune dans l'un des 4 clades préalablement observés. Cette tendance est confirmée par les individus pour lesquels 10, 11 ou 12 séquences ont été récupérées et qui, elles aussi, se répartissent dans les 4 différents clades. Les deux individus de 37°S qui ont fait l'objet d'un clonage individuel se répartissent en 3 clades à partir de 7 et 8 séquences récupérées (respectivement pour les individus 1 et 2).

Après avoir soigneusement relevé et noté la répartition des allèles, une seule séquence par clade et par individu est conservée (au final, 225 séquences). Ces séquences ont été représentées dans un arbre pour lequel on distingue les 8 clades (4 pour le Pacifique et 4 pour l'Atlantique) (Figure V.6).

A partir de ce nouveau jeu de séquences, nous avons défini une séquence « consensus » par clade en utilisant la règle du consensus majoritaire (50%) pour analyser plus finement la divergence entre lignées paralogues à partir des distributions des mutations synonymes et non-synonymes. L'approche avec la méthode des consensus possède une limite, elle confond en une séquence les informations de divergence fixée et de polymorphisme. Cependant, même si certains sites polymorphes en haute fréquence sont fixés par la méthode des consensus, les groupes ont été définis au préalable à partir des mutations fixées. Les consensus représentent des groupes réels et non pas la fixation artificielle de sites polymorphes. L'analyse a été effectuée à partir de ces 8 séquences « consensus » plus celle du groupe externe *Mytilus galloprovincialis*. La position dans l'arbre de chacune des 8 séquences consensus est représentée sur la Figure V.6.

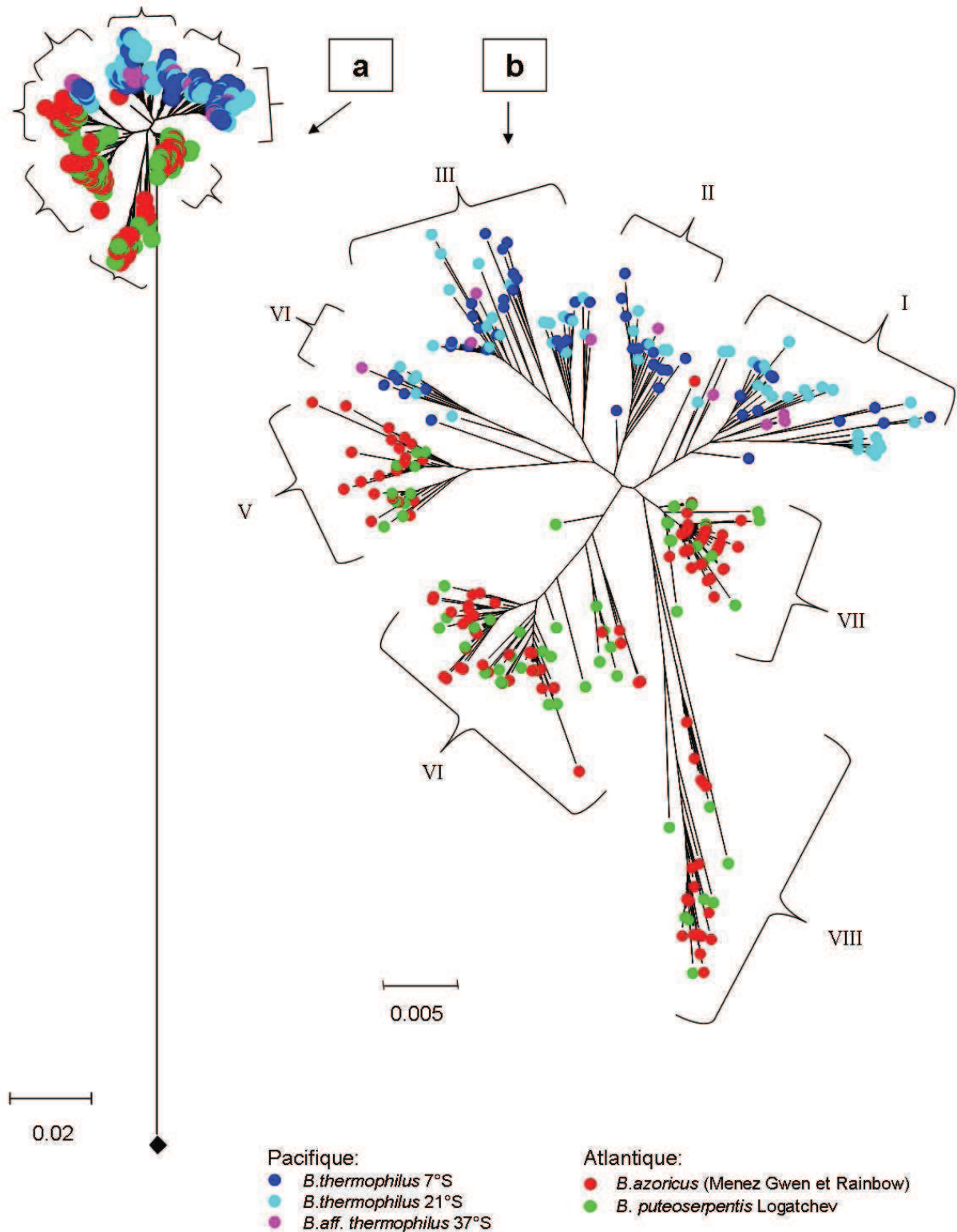


Figure V.4. Arbres en NJ, réalisés à partir de l'ensemble des séquences recapturées et corrigées. **(a)**: arbre raciné avec la séquence Hsp70 de *Mytilus galloprovincialis* AJ783713 : séquence la plus proche trouvée dans les banques ; **(b)** : détail des regroupements de séquences des quatre espèces de *Bathymodiolus* et numéro des différents groupes observés.

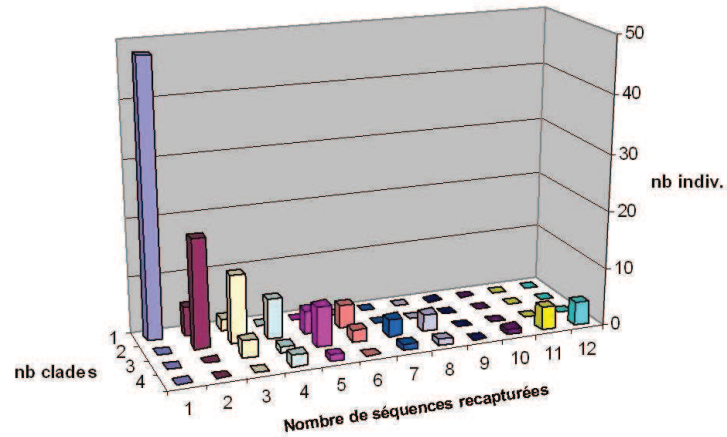


Figure V.5. Représentation du nombre de clades dans lesquels se répartissent les allèles au sein de chaque espèce en fonction du taux de recapture. Il s'agit du cumul des recaptures individuelles pour le gène HSP70 à partir des populations 7°S et 21°S (*Bathymodiolus thermophilus*), 37°S (*B. aff. thermophilus*), Rainbow et Menez Gwen (*B. azoricus*) et Logatchev (*B. puteoserpentis*) ; soit 126 individus et 361 séquences).

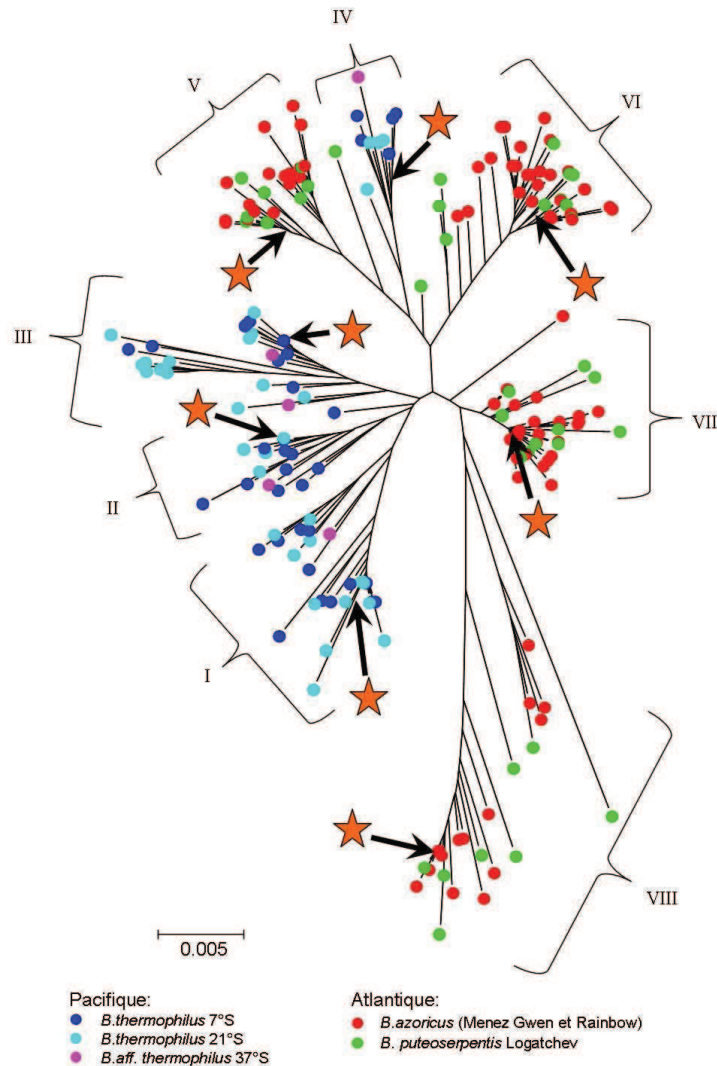


Figure V.6. Arbre obtenu en Neighbour-Joining à partir des séquences sélectionnées (1 par clade et par individu) à l'aide de la distance Tamura Nei. Les étoiles orange correspondent à l'emplacement des 8 séquences consensus utilisées dans l'analyse.

V.4.3. Comparaison des pressions sélectives entre les clades.

La topologie de l'arbre des HSP70 est assez homogène cependant certains clades sont plus ramifiés que les autres et semblent refléter des taux d'évolution différents. Nous nous sommes donc demandé si cette diversification du gène était liée à des événements de duplications successifs ou à des effets sélectifs sur certaines lignées paralogues.

Pour tester l'hypothèse d'une évolution non-neutre des lignées paralogues (diversification adaptative du gène), nous avons utilisé le programme CodeML du logiciel PAML en comparant différents modèles emboîtés de substitutions de codons (Yang 1997).

Pour tester les modèles de sélection sur le gène HSP70, il est nécessaire de fournir un arbre de base au logiciel. La première difficulté est donc de déterminer la topologie de l'arbre qui correspondait le plus à nos données, nous permettant donc d'obtenir des résultats fiables.

Deux méthodes de reconstruction d'arbres ont donc été comparées : le Neighbour-Joining et le maximum de vraisemblance.

En fonction des modes de calculs utilisés pour représenter les arbres, les topologies obtenues sont très différentes. Aucun de ces arbres ne présente de topologie facilement interprétable permettant de déterminer aisément l'évolution du gène. Seuls les groupes II et III trouvés au sein de l'espèce pacifique semblent être monophylétiques et associés de façon robuste. Les séquences qui se regroupent à l'extérieur de l'arbre sont les séquences VIII et VII, selon les représentations.

Les arbres phylogénétiques sont présentés sur la Figure V.7 et sont peu robustes. L'arbre vrai ne peut être déterminé car les lignées pourraient être affectées par des phénomènes de duplication et de spéciation auxquels peuvent s'ajouter de l'homoplasie, de la recombinaison, de la conversion génique, de l'évolution concertée, ou encore des effets sélectifs.

Les arbres étant peu soutenus, il nous a semblé préférable de tester nos données à partir d'un arbre correspondant à un scénario évolutif *a priori* le plus vraisemblable. L'arbre qui a été utilisé dans l'analyse a donc été sélectionné à partir de modèles théoriques ne prenant pas en compte l'évolution concertée mais combinant deux types d'événements

(duplication et spéciation). Cet arbre présenté dans la Figure V.6 a été choisi, sachant : (1) qu'au temps présent, 8 clades existent et qu'il sont répartis de manière équitable entre les deux océans (4 clades pacifiques et 4 clades atlantiques) et (2) que le clade comprenant les groupes IV (pacifique) et V (atlantique) et les groupes II et III sont robustes, ceux-ci étant regroupés quel que soit le mode de reconstruction utilisé (cf arbres de la Figure V.7).

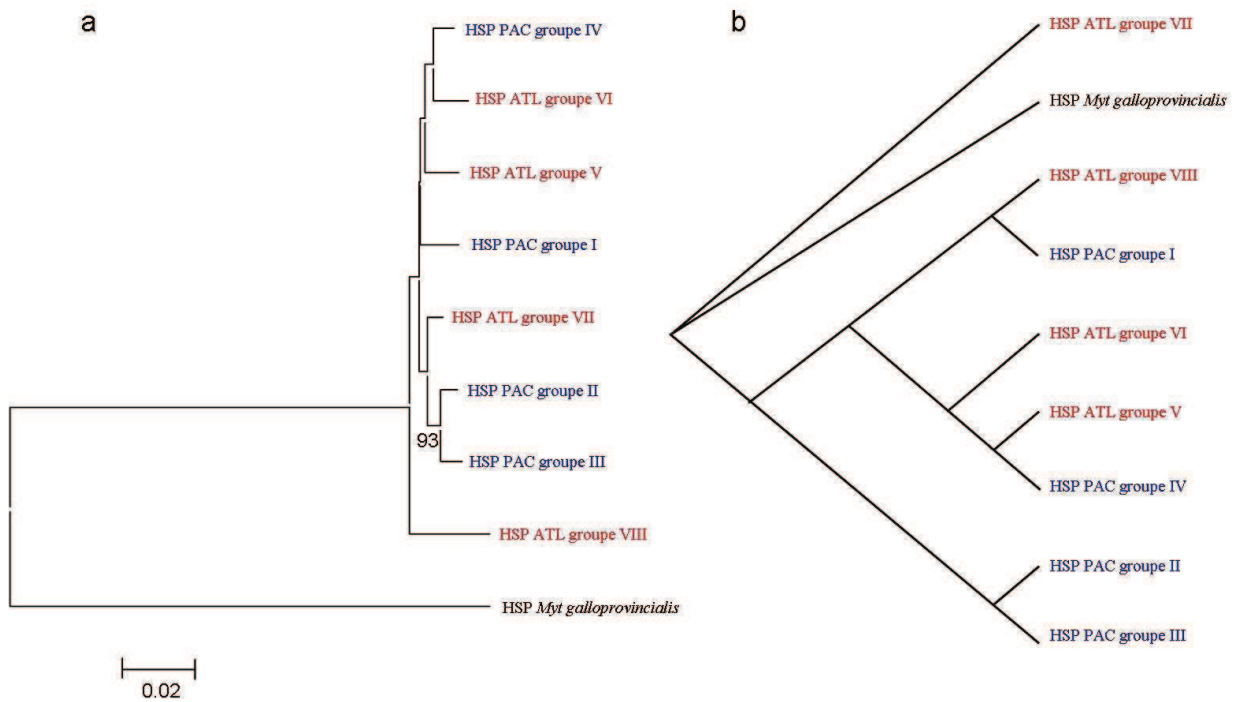


Figure V.7. Topologie des arbres obtenus selon les méthodes de (a) Neighbour Joining (1000 bootstraps) et (b) maximum de vraisemblance (recherche exhaustive du meilleur arbre à l'aide du programme DNAML du logiciel PHYLIP).

Après comparaison des topologies résultant des scénarios d'évolution du gène de la HSP70 (**Figure V.8**), c'est le modèle 'c' qui a été retenu car il correspond apparemment le mieux à la topologie observée **Figure V.6**. Ce modèle explique la formation d'un clade distinct (correspondant aux groupes IV et V dans nos données), et la monophylie réciproque de chaque océan colle bien avec les deux grands groupes observés dans les données réelles (clades bleu et rouge).

Les modèles de sélection seront donc testés à partir d'un arbre défini selon ce scénario théorique.

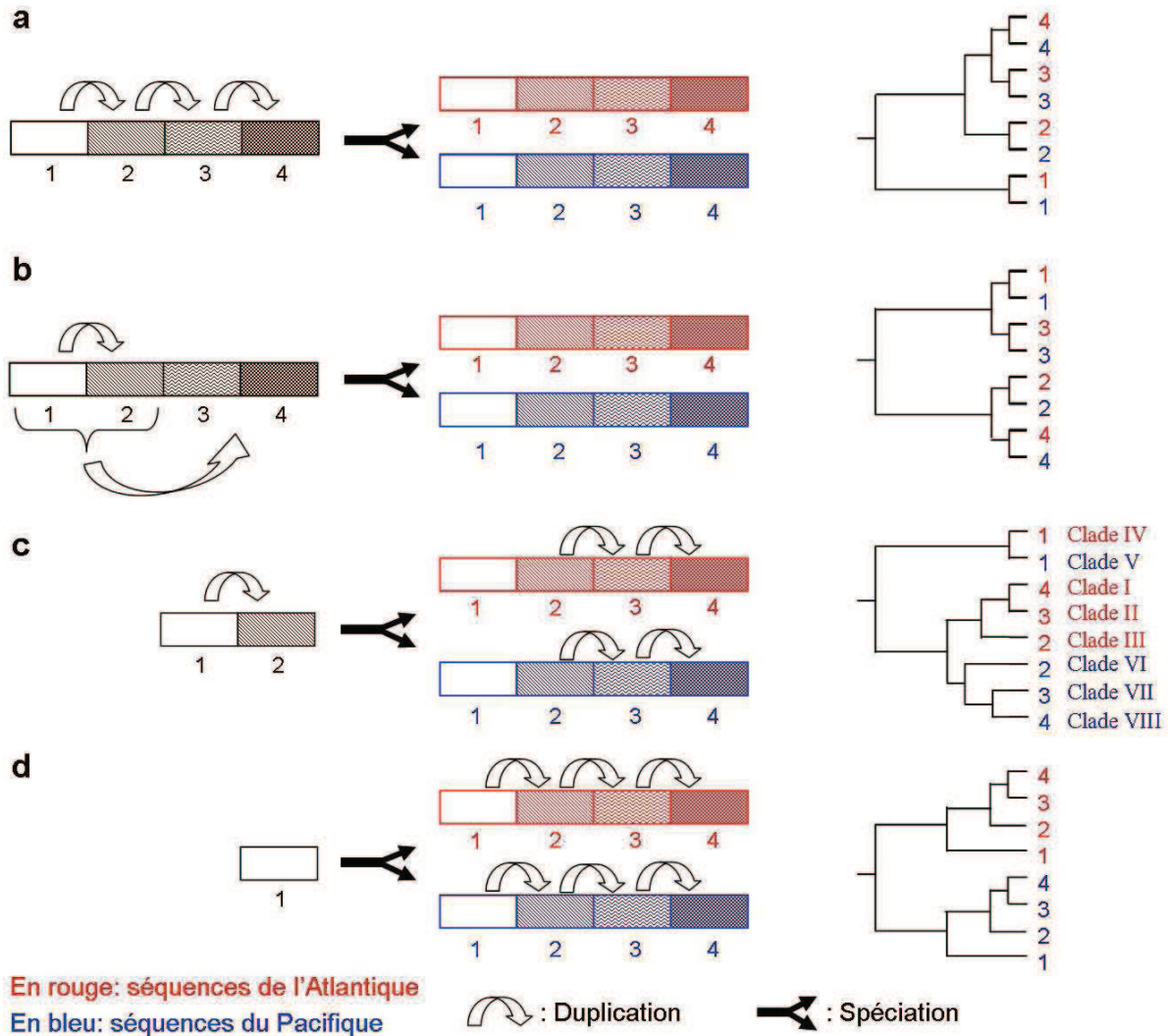


Figure V.8. Modèles théoriques d'évolution conduisant à la formation de 4 clades dans chacun des océans. Ces modèles combinent les phénomènes de spéciation et de duplication. Chaque rectangle correspond à une copie du gène. Les gènes en noir et gris : composition de ce gène chez l'espèce ancestrale. Les gènes en 'rouge' : clades actuels atlantiques, gènes en 'bleu' : clades actuels pacifiques. L'arbre théoriquement obtenu est indiqué à droite de chaque modèle. Les chiffres déterminent chaque copie de gène. **(a)** 3 duplications avant spéciation, **(b)** une duplication (1 puis 2) suivie d'une nouvelle duplication en tandem (1 et 2 donnent 3 et 4) avant spéciation; **(c)** : une seule duplication pré-spéciation puis deux épisodes de duplication après spéciation dans chacune des espèces ; **(d)** les 3 duplications sont post-spéciation. Les clades observés pour les données *Bathymodiolus* sont indiqués à côté de l'arbre théorique le plus vraisemblable.

L'arbre de distances, réajusté par vraisemblance avec le modèle de substitutions des codons de Goldman & Yang (1994) a été obtenu à partir de nos données et permet de montrer l'évolution réelle des clades en tenant compte du modèle défini et la contrainte topologique. Cet arbre a été réalisé à l'aide du programme CodeML du logiciel PAML3.14 et est représenté dans la Figure V.9.

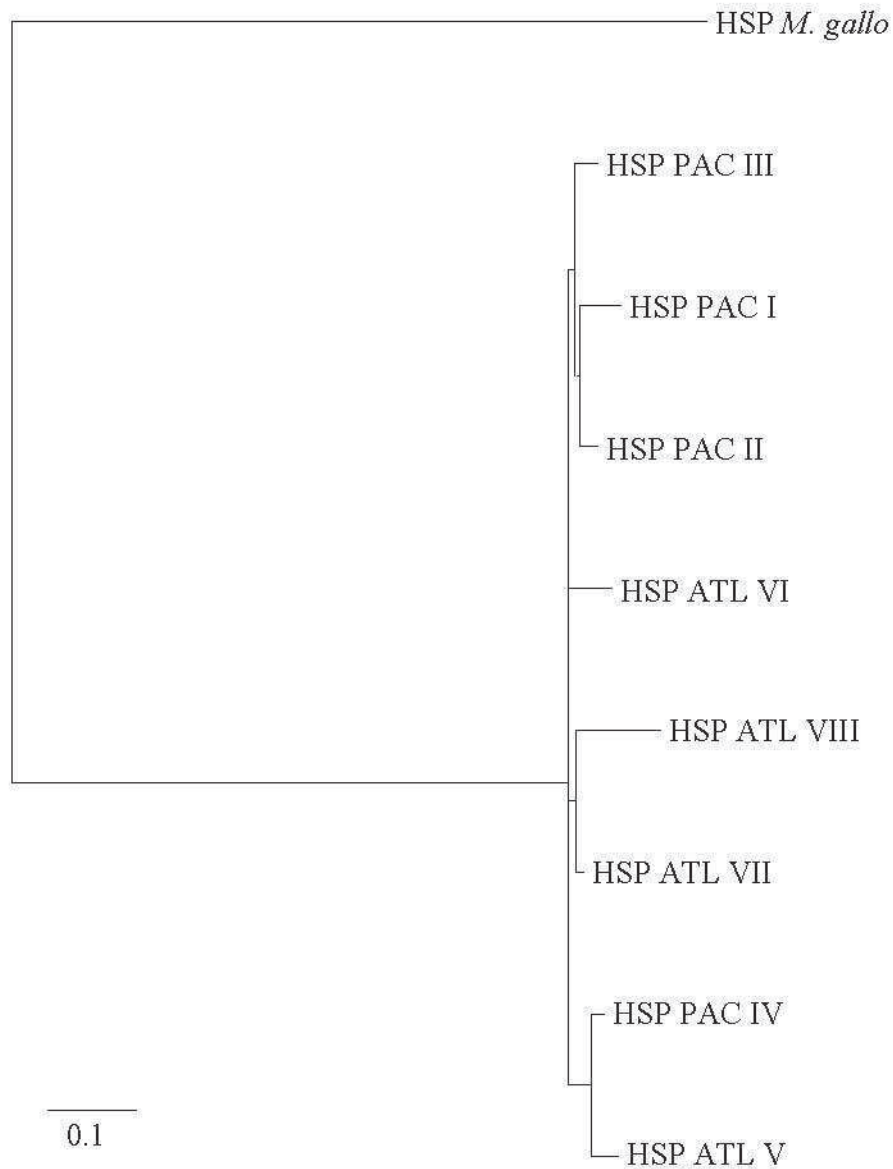


Figure V.9. Arbre ML de distances estimé à l'aide du modèle de substitutions des codons de Goldman & Yang (1994) à partir de la topologie théorique proposée.

La longueur des séquences utilisées est de 852 bp, ce qui correspond à 284 codons. Si les pressions de sélection sont supposées identiques entre les différents clades, le paramètre ω (ratio d_N/d_S) sera constant dans chacune des branches. En revanche, si certains clades évoluent de manière particulière, ω sera différent selon les branches. En général, un événement de duplication conduit souvent à un relâchement des pressions sur l'un des deux paralogues formés (Page & Holmes 2000), ce qui permet au dit paralogue de pouvoir évoluer plus rapidement vers une autre fonction (cf. Bailly et al. 2003). Ce sont ces hypothèses qui sont

testées avec PAML (modèles dit « branches ») en considérant un modèle M0 où ω ne varie pas selon la branche considérée avec un modèle M1 où ω est libre de varier pour toutes les branches. Cette analyse est conduite en deux temps : (i) estimation des valeurs de maximum de vraisemblance pour chacun des modèles d'évolution (avec ω constant et avec ω variable) ; (ii) test de la double différence des valeurs de vraisemblance obtenues avec un test de χ^2 (à n degré de liberté : n = différence entre les nombres de paramètres estimés dans les 2 modèles - 1) pour déterminer quel modèle d'évolution est le plus approprié pour expliquer nos données sous contrainte de la topologie observée.

Tableau V.4. Comparaison des 2 modèles d'évolution sous PAML. **L** : longueur de l'arbre en nombre de substitutions nucléotidiques par codon ; **kappa** : rapport transitions / transversions, **lnL** : valeur d'ajustement des données à l'arbre sous le modèle choisi, test de vraisemblance : **2*lnL (ddl)** : 2 fois la différence entre les deux lnL ; **ddl** : degré de liberté utilisé pour la comparaison dans la table de χ^2 . *** : différence significative pour $p < 0,001$.

	dN/dS par branche	L	dN total arbre	dS total arbre	kappa	lnL	2*ΔlnL (ddl)
modèle 0 :	constant	1,771	0,102	2,416	2,614	-2088,74	44*** (15)
modèle 1 :	variable	2,804	0,099	4,112	2,804	-2066,74	

Le modèle M0, qui utilise une valeur moyenne unique du paramètre ω pour l'arbre entier, donne une valeur de $\ln L_0 = -2088,74$. Le modèle M1 où ω est libre de varier, qui utilise 15 valeurs de ω (une par branche : arbre = 15 branches), fournit une valeur de $\ln L_1 = -2066,74$. La différence entre des deux valeurs de log likelihood, $2\Delta\ln L = 2(\ln L_0 - \ln L_1) = 44$. Cette valeur peut être comparée à une valeur seuil de la distribution de χ^2 avec 14 degrés de liberté (nombre de branches dans l'arbre moins 1). La différence entre les deux modèles est hautement significative ($P < 0,001$), indiquant que des différences entre le rapport d_N/d_S n'a pas été constant au cours de l'évolution (Tableau V.4).

L'estimation des paramètres du modèle de substitution des codons sous contrainte de la topologie observée pour le gène Hsp70 montre que quelques branches ont pu subir un relâchement des pressions de sélection liées à la structure et à la fonction de la protéine. Seules trois branches présentent des valeurs de $N*dN$ supérieures à celles estimées pour $S*dS$ (entourée en rouge) Figure V.10. Ce résultat peut être interprété de deux manières : il peut s'agir d'une forte sélection positive sur certains des codons ou alors du relâchement de la pression de sélection sur l'ensemble du gène, permettant une accumulation sensiblement égale des mutations synonymes et non synonymes. Pour pouvoir trancher entre ces 2 hypothèses,

une deuxième série d'analyses a été conduite en recherchant s'il existe une fraction de codons pouvant être sous sélection positive.

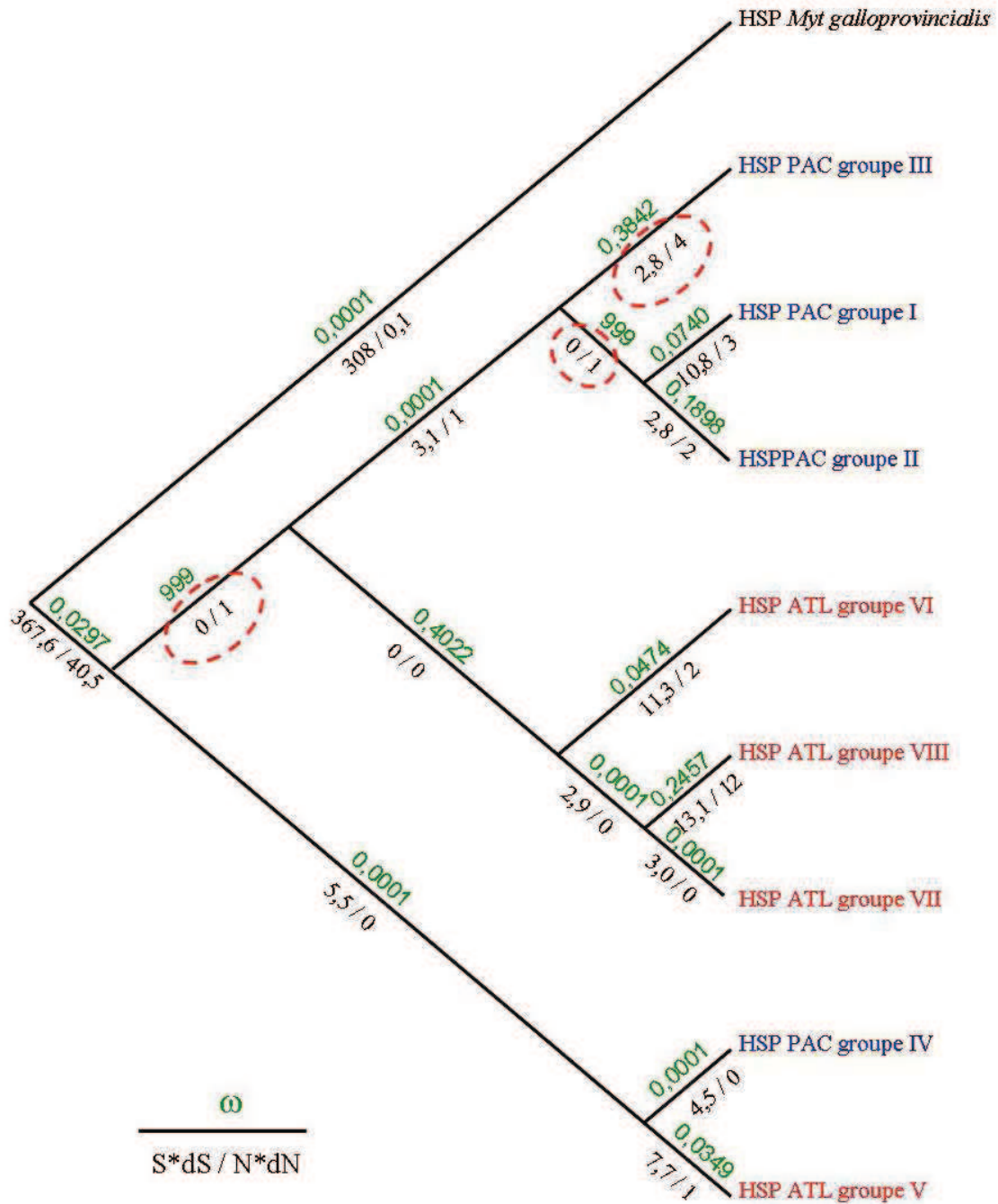


Figure V.10. Distribution des mutations synonymes (S) et non synonymes (N) rapportées à leurs nombres de sites potentiels respectifs dans les différentes branches de l'arbre théorique du gène *Hsp70*. Les valeurs indiquées le long de chaque branche sont les nombres de mutations non synonymes et synonymes (dans l'ordre S*dS / N*dN) pour la branche indiquée sous le modèle d'évolution libre du dN/dS. Entourée en rouge : branche avec une valeur de $S < N$. Les valeurs indiquées en vert au dessus des branches correspondent au ω . Le numéro des clades correspond à celui indiqué sur la figure Figure V.6. En bleu : clades « pacifiques » ; en rouge : clades « atlantiques ».

V.4.4. Sélection positive ou relâchement des pressions de sélection ?

Pour connaître laquelle des hypothèses de sélection positive ou de relâchement des pressions de sélection permet d'expliquer l'accumulation de mutations non-synonymes dans certaines branches, il est nécessaire de mener les analyses directement sur les codons. Dans l'hypothèse d'un relâchement des pressions de sélection, tous les codons sont considérés comme « presque neutres », alors que dans le cas d'une sélection positive sur l'une des lignées, certains codons doivent nécessairement avoir évolué sous sélection positive ($d_N/d_S \gg 1$).

Ces deux hypothèses peuvent être testées avec le programme CodeML du logiciel PAML3.14 en comparant un modèle d'évolution « presque neutre » (deux classes de mutations : mutations neutres ou légèrement délétères) et un modèle avec sélection positive (trois classes de mutations : mutations neutres, légèrement délétères et positives). Afin de déterminer lequel des deux modèles explique le mieux nos données sous contrainte d'une topologie donnée, il faut tester l'une contre l'autre les valeurs de vraisemblance fournies par les deux modèles emboîtés.

Dans le premier modèle, dit « presque neutre », les codons sont classés selon deux catégories : les codons neutres (dont la valeur du d_N/d_S approxime 1 ; notée par convention classe $d_N/d_S=1$), et les codons sous sélection purifiante (dont la valeur du d_N/d_S approxime zéro ; appelée arbitrairement classe $d_N/d_S=0$).

Dans le modèle avec sélection positive, les deux catégories de mutations établies pour le modèle « presque neutre » sont conservées, mais une troisième catégorie est ajoutée : celle des codons sous sélection positive (appelée classe $d_N/d_S > 1$). Les résultats des analyses sont présentés dans le Tableau V.5.

Les modèles emboîtés utilisés fournissent la même valeur de vraisemblance pour l'ajustement des données séquences sous contrainte de la topologie observée. Le modèle avec sélection positive ne permet pas d'expliquer mieux l'évolution des séquences HSP70 dans l'arbre observé. La proportion de sites sous sélection purifiante est la même dans les deux modèles. En revanche, les codons considérés comme 'presque neutres' dans le modèle neutre ($d_N/d_S = 1$; 0,0646), sont répartis entre les classes 'presque neutre' ($d_N/d_S = 1$; 0,0369) et 'sélection positive' ($d_N/d_S > 1$; 0,0277). Deux codons classés dans la classe ($d_N/d_S > 1$)

présentent des probabilités postérieures d'appartenance à cette classe (BEB) supérieures à 0,7 : il s'agit de I (isoleucine) en position 115 et de R (Arginine) en position 275 (Figure V.11). Le résultat obtenu pour la valeur ω de la classe de mutation $d_N/d_S > 1$ est étrange. Cette valeur correspond normalement au ω moyen calculé entre les codons situés dans cette classe. Dans notre cas la moyenne des oméga obtenu pour les codons 115 ($\omega = 1,944$) et 275 ($\omega = 2,392$) n'est pourtant absolument pas égale à 1. Il s'agit certainement d'une bizarrerie du logiciel.

Tableau V.5. Comparaison des modèles de sélection sur les codons. **L** : longueur de l'arbre en nombre de substitutions nucléotidiques par codon ; **κ** (Kappa) : rapport transitions / transversions ; **LnL** : valeur d'ajustement des données à l'arbre sous le modèle choisi ; **Proportion** : fraction des codons classés dans chacune des catégories ; **ω** : Omega moyen des codons classés dans chacune des catégories.

Modèle		$d_N/d_S = 0$	$d_N/d_S = 1$	$d_N/d_S > 1$	L	κ	LnL
neutre	Proportion	0,9354	0,0646	-	2,154	2,979	-2065,988
	ω	0,0188	1	-			
sélection positive	Proportion	0,9354	0,0369	0,0277	2,154	2,979	-2065,988
	ω	0,0188	1	1			

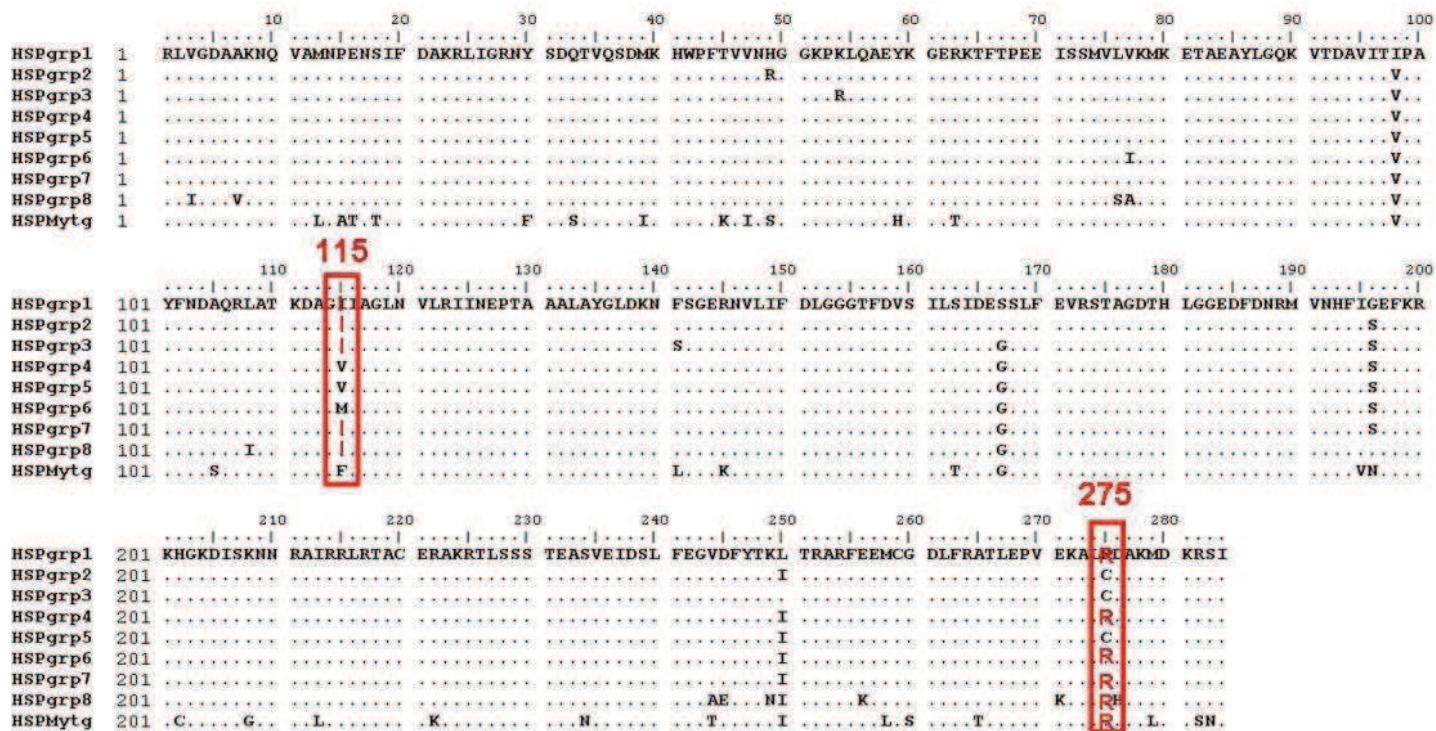


Figure V.11. Aligment en acides aminés pour des 9 séquences utilisées dans nos analyses (8 séquences consensus et séquence du groupe externe). Encadrés en rouge : la position des codons sous sélection positive ($\omega > 1$).

V.4.5. Conclusion sur contraintes de sélection concernant les lignées paralogues de HSP70

Tous les clades du gène de la HSP 70 n'évoluent pas de la même manière et à la même vitesse. Les contraintes de sélection purifiante qui agissent sur les branches des clades IV (pacifique) et VII (atlantique) sont très fortes ($\omega = 0,0001$) alors qu'elles apparaissent relâchées et permettent l'accumulation de mutations non synonymes pour les groupes 3 (pacifique) et 8 (atlantique) ($\omega = 0,3842$ $\omega = 0,2457$).

Les analyses indiquent que le modèle de sélection positive n'est pas responsable des accumulations de mutations non synonymes même si 2 codons présentent des valeurs de ω positives. Pour expliquer une accumulation de mutations non synonymes, l'hypothèse la plus vraisemblable est qu'au moins un des événements de duplication ayant conduit à la topologie actuelle du gène Hsp70 ait donné lieu à un relâchement des pressions de sélection sur certaines lignées.

Ces résultats sont robustes car la même analyse a été menée en parallèle en fixant une topologie différente. Cette 2^e topologie testée est celle obtenue en NJ avec les distances de Tamura et Nei (1993), représentée en 'a' sur la Figure V.7). Les résultats (résultats non présentés) obtenus avec les différents modèles de Yang étaient similaires et les codons 115 et 275 étaient également diagnostiqués comme étant sous sélection positive.

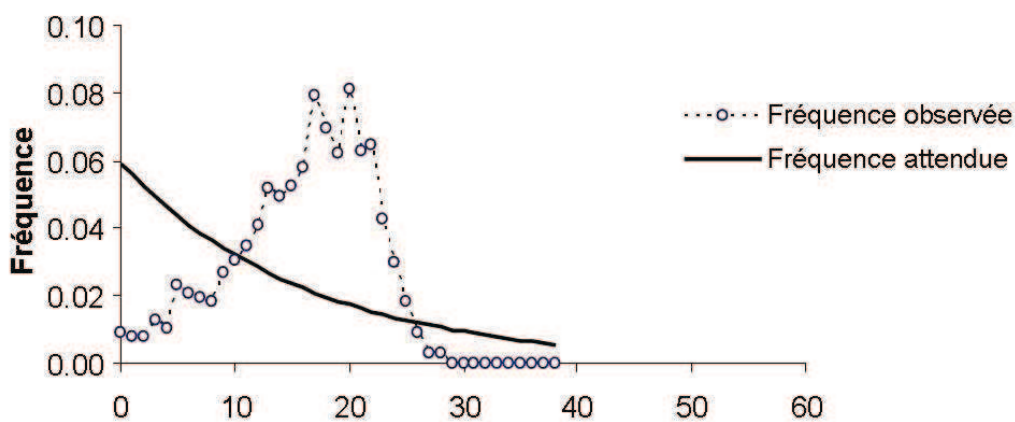
V.4.6. Analyse de la diversité nucléotidique au sein des clades du gène HSP70.

Dans cette deuxième partie, nous nous intéressons plus particulièrement à la diversité globale des séquences nucléotidiques de ce gène au sein de chaque clade. Les analyses sont conduites sur le jeu de données HQ dans lequel nous avons conservé un allèle par individu et par groupe. Les analyses ont été réalisées avec le logiciel DNAsp 4.0 (Rozas et al. 2003).

Dans un premier temps, nous avons voulu savoir si l'accumulation des mutations dans le polymorphisme s'effectuait d'une façon neutre pour le gène *Hsp70* dans chacun des deux groupes d'espèces présentant une monophylie réciproque à tous les locus (Pacifique contre Atlantique) en observant la distribution des fréquences du nombre de différences nucléotidiques par paires de séquences. Ces distributions sont comparées de manière graphique à la distribution des fréquences attendues sous le modèle d'une population à

l'équilibre mutation/dérive, ce sont les courbes de distribution de « mismatch » (Figure V.12).

Pacifique



Atlantique

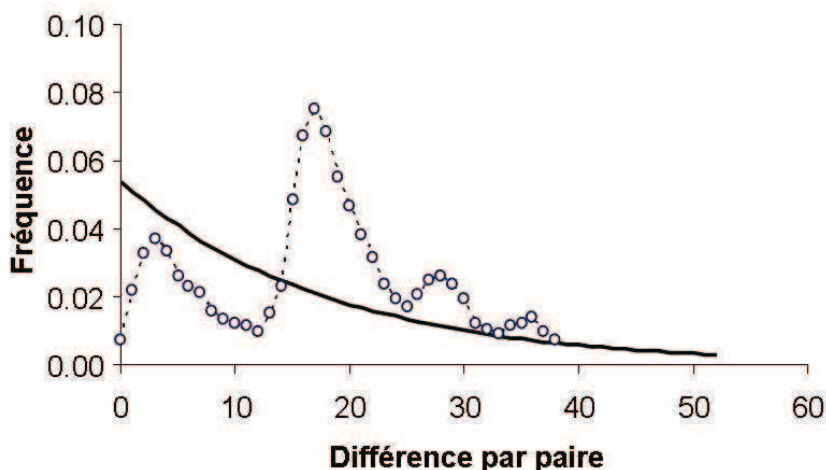


Figure V.12. Distribution du nombre de différences nucléotidiques par paire de séquences et comparaison par rapport à l'attendu théorique. En haut : séquences de HSP70 du Pacifique ; en bas : séquences de l'Atlantique.

Les courbes observées s'écartent de manière flagrante de l'attendu théorique avec un pic caractéristique centré autour de vingt différences par paire de séquences (càd. augmentation de fréquence des allèles présentant une divergence intermédiaire). Cet écart à l'attendu neutre peut s'expliquer de plusieurs façons : (1) un fort excès de divergences intermédiaires dû à la recombinaison, (2) un excès d'allèles présentant une forte divergence due à un mélange de populations spatialement structurées et, (3) un excès d'allèles présentant une forte divergence due à un mélange de gènes paralogues. Cependant, les distributions sont légèrement différentes entre les espèces des deux océans. Dans le Pacifique, le pic de fréquence aux alentours de vingt nucléotides reflète la structure de séquences en clades, soit :

(1) en deux grands ensembles de séquences pour lesquels la différence moyenne est de vingt nucléotides, soit (2) plus de deux clades mais pour lesquels la divergence est la même quelle que soit la comparaison effectuée (de 20 différences). Si l'on regarde plus précisément la distribution des fréquences, on constate que le « pic » à 20 différences est en fait constitué de deux pics proches situés à 17 et 20 mutations de différences, et au moins un autre pic est visible à 5 mutations de différences.

Concernant les espèces atlantiques, la distribution montre 4 pics distincts, pour les valeurs de 3, 17, 28 et 36 différences, démontrant de la sorte qu'il existe au moins deux groupes de séquences différenciées.

Ces résultats montrent que même si les séquences du gène HSP70 présentent des groupes différenciés dans chacun des océans, ces groupes ne sont pas symétriques du point de vue de leur divergence relative entre les différentes copies. Dans l'Atlantique, les quatre groupes semblent très distincts alors que ce pattern est moins visible pour les séquences du Pacifique. Cela pourrait accréditer la thèse qu'au moins un événement de duplication ait pu avoir lieu après l'événement de spéciation ayant conduit à la séparation des espèces atlantiques et pacifiques. En effet, dans ce cas, les duplications post-spéciation n'ont pas de raison d'être apparues simultanément.

Des analyses portant sur la diversité nucléotidique ont ensuite été réalisées pour chacun des clades. Les résultats sont consignés dans le Tableau V.6

Le nombre de sites polymorphes est très élevé dans chacun des groupes pour les deux complexes d'espèces. Le rapport entre le nombre de sites polymorphes et le nombre de séquences recapturées est constant pour l'ensemble des groupes à l'exception du groupe 8 qui possède un excès de mutations par rapport aux autres groupes (diversification allélique ?), et le groupe 7 qui au contraire semble être moins polymorphe que les autres groupes (sélection purifiante ?). Cette diversité plus importante retrouvée dans le groupe 8 (Atlantique), et à un niveau plus modeste dans le groupe 3 (Pacifique) pourrait être corrélée au taux de recombinaison qui est maximum pour ces clades.

Tableau V.6. Descripteurs de la diversité nucléotidique pour les différents clades du gène HSP70. **N**: Nombre de séquences ; **L**: longueur des séquences utilisées ; **S** : nombre de sites polymorphes ; **θ_w** : Estimateur du $4N\mu$ par paire de base en utilisant le nombre de sites polymorphes (Watterson 1975). ; **π** : Estimateur du $4N\mu$ à partir du nombre moyen de différences nucléotidiques par site (Nei 1987) ; **D** de Tajima : estimateur de Tajima (1989a). Le niveau de significativité est indiqué par : * ($P < 0,05$) ; ** ($P < 0,01$) ; *** ($P < 0,001$) ; **Rm** : Nombre minimum de recombinants détectés par la méthode de Hudson & Kaplan (1985)

Océan	Séquences utilisées	<i>N</i>	<i>L</i>	<i>S</i>	θ_w	π	D	Rm	Rm / <i>N</i>
Pac.	<i>Groupe 1</i>	30	853	57	0,0178	0,0102	-1,5898	7	0,23
Pac.	<i>Groupe 2</i>	16	853	37	0,0138	0,0094	-1,3478	4	0,25
Pac.	<i>Groupe 3</i>	25	853	52	0,0174	0,0122	-1,1626	11	0,44
Pac.	<i>Groupe 4</i>	10	853	34	0,0141	0,0100	-1,4085	3	0,30
Atl.	<i>Groupe 5</i>	31	853	65	0,0194	0,0100	-1,8148*	5	0,16
Atl.	<i>Groupe 6</i>	40	853	69	0,0198	0,0074	-2,2682**	6	0,15
Atl.	<i>Groupe 7</i>	49	853	51	0,0140	0,0038	-2,5324***	3	0,06
Atl.	<i>Groupe 8</i>	23	853	84	0,0273	0,0173	-1,4667	14	0,61
tout Pacifique		81	853	130	0,0335	0,0186	-1,4759	16	0,20
tout Atlantique		143	853	209	0,0471	0,0206	-1,8338*	25	0,17

Les indices de diversité estimés pour les groupes 1 à 6 sont du même ordre de grandeur avec des valeurs de π comprises entre 0,0074 et 0,0122. Les groupes 7 et 8 présentent quant à eux des diversités respectivement plus faibles (0,0038) et plus fortes (0,0173). Pour le groupe 8, ceci s'explique par la présence de nombreuses lignées très divergentes, celles-ci pouvant être liées à une accumulation plus importante de mutations entre les espèces *B. azoricus* et *B. puteoserpentis*.

Les indices de diversité calculés pour chaque complexe d'espèces sont approximativement deux fois plus élevés que ceux obtenus pour chaque groupe pris séparément. Cela confirme l'existence d'une forte hétérogénéité du polymorphisme à l'échelle de l'espèce pour ce gène. En effet, si ces groupes n'avaient pas eu d'existence réelle, la diversité inter-groupe (c'est-à-dire à l'échelle de l'espèce), n'aurait pas été plus forte que la diversité observée au sein des groupes.

Il faut également noter que la diversité totale des espèces atlantiques est plus forte que celle des espèces de l'Océan Pacifique. Cette plus forte diversité correspond à une divergence moyenne entre les groupes plus élevée dans l'Atlantique.

Ecarts à l'attendu neutre.

Dans chacun des groupes définis, les séquences proviennent de populations et d'espèces différentes : *B. thermophilus* et *B. aff. thermophilus*, pour les groupes 1, 2, 3 et 4 ; et *B. azoricus* et *B. puteoserpentis* pour les groupes 5, 6, 7 et 8. Ces regroupements peuvent engendrer de la structure à l'intérieur des différents clades. Cependant, en dépit d'une potentielle structure géographique, les valeurs obtenues pour le D de Tajima sont négatives pour l'ensemble des groupes, cela permet de confirmer que ces derniers ne sont que peu structurés entre espèces proches (flux de gènes inter-espèces importants au sein des deux océans pour ce gène).

Les valeurs de D se comportent différemment en fonction des deux complexes d'espèces : les groupes 5, 6, 7 formés avec des séquences « atlantiques » possèdent des valeurs de D négatives très significativement différentes de 0, alors qu'aucune valeur observée pour les espèces du Pacifique n'est significative. Ceci, couplé à la diversité nucléotidique plus élevée (liée à la divergence moyenne plus élevée dans l'Atlantique) pour les espèces atlantiques peut indiquer une sélection purifiante plus forte au sein de chaque clade.

Un nombre important de recombinaisons est également retrouvée au sein de chaque groupe. Le groupe 7 est celui qui présente le moins de recombinaison ($R_m=3$ pour 49 séquences), et le groupe 8, celui qui montre le plus grand nombre de recombinants ($R_m=14$ pour 23 séquences). Dans le Pacifique, le groupe 3 a une proportion R_m/N (nombre de recombinants / nb de séquences) plus élevée que les autres. L'indice R_m/N semble indiquer que ce nombre de recombinants pourrait être plus élevé dans le Pacifique que dans l'Atlantique. Ces valeurs semblent en adéquation avec les branches de l'arbre pour lesquels un relâchement des pressions de sélection a été détecté (branches allant aux groupes 3 (Pacifique) et 8 (Atlantique)).

La divergence globale estimée entre les deux complexes d'espèces est de 0,023 (non montrée). Cependant, en prenant en compte l'hypothèse du modèle « Duplication/Spéciation/Duplication » (fournissant 8 groupes), il semble que la valeur de divergence la plus juste soit celle estimée entre les groupes 4 (Pacifique) et 5 (Atlantique) ou entre les clades (1,2 & 3) et (6, 7 & 8). Dans le premier cas, la valeur de divergence est de

0,0203 et dans le second cas, elle peut être approximée à 0,0264 (valeur moyenne entre les groupes) Tableau V.7.

Tableau V.7. Divergences estimées entre les différents groupes de séquences. **A** : au sein des espèces pacifiques, **B** : au sein des espèces atlantiques, **C** : entre les deux complexes d'espèces. *Les valeurs de divergence indiquées correspondent au nombre moyen de substitutions par site nucléotidique, non corrigé pour les substitutions multiples.

A) Pacifique	Grp 2	Grp 3	Grp 4	B) Atlantique	Grp 6	Grp 7	Grp 8
Grp 1	0.0209	0.0224	0.0236	Grp 5	0.0234	0.0202	0.0334
Grp 2		0.0173	0.0246	Grp 6		0.0206	0.0393
Grp 3			0.0241	Grp 7			0.0296

C) inter océans	Grp 5	Grp 6	Grp 7	Grp 8
Grp 1	0.0228	0.0255	0.0193	0.0341
Grp 2	0.0270	0.0240	0.0158	0.0378
Grp 3	0.0282	0.0238	0.0183	0.0392
Grp 4	0.0203	0.0196	0.0224	0.0361

V.4.7. Discussion.

L'évolution du gène HSP70 est très complexe et fait vraisemblablement intervenir des duplications antérieures et postérieures à la séparation des complexes atlantiques et pacifiques, des niveaux de pression de sélection différents entre paralogues et certains codons pourraient même être sous sélection positive. Il ne faut cependant pas écarter l'hypothèse de l'évolution concertée qui pourrait intervenir entre les différentes lignées de HSP. Sous cette hypothèse, il est possible que l'ensemble de lignées observées soient antérieures à la spéciation et que l'évolution concertée soit responsable de la très faible divergence entre les lignées et de la non robustesse de l'arbre.

Cependant, le gène de la HSP est connu chez les eucaryotes pour avoir évolué par duplications successives et il existe déjà plusieurs copies du gène bien décrites au niveau des génomes déjà séquencés (forme induite Hsp vs forme constitutive Hsc : Boutet et al. (2003) ; formes mitochondriales vs formes cytosoliques : Kiang & Tsokos (1998). Requena et al. (1988) et Evgen'ev et al. (2004) ont d'ailleurs décrit avec précision le mécanisme de duplication en tandem à l'origine de la diversification de ce gène chez *Trypanosoma cruzi* et chez *Drosophila viridis*, respectivement. Bettencourt & Feder (2001) et Maside et al. (2002) ont également montré que la fixation d'éléments transposables (S-elements) dans la partie 5' du gène Hsp70 aurait pu être à l'origine de la diversité élevée trouvée pour ce gène chez

Drosophila melanogaster. Néanmoins, dans cette étude, il convient de noter qu'une telle diversité de paralogues n'avait encore jamais été décrite au sein des formes classiquement décrites à partir des génomes.

Si l'on accepte l'hypothèse d'une horloge moléculaire, il semble que les groupes de paralogues se soient différenciés moins rapidement dans l'Océan Pacifique que dans l'Océan Atlantique, laissant supposer que les duplications postérieures à la spéciation n'aient pas eu lieu en même temps. Il faut mettre cela en relation avec le taux de recombinaison plus élevé détecté chez les espèces pacifiques. Il est cependant difficile de déterminer si la divergence entre les groupes est plus faible du fait de la recombinaison, ou si au contraire, les recombinaisons sont favorisées en plus grand nombre dans le Pacifique de part un lien physique plus fort entre les groupes. Une autre explication pourrait être associée au mode de duplication lui-même, si celui-ci est intrinsèquement lié à une rétrotransposition active comme c'est le cas chez *Drosophila melanogaster* (Maside et al. 2002). En effet, il est connu que le processus de rétrotransposition peut être induit par la température (Li et al. 1999). L'extrême diversité des Hsp 70 rencontrée pourrait être de ce fait en partie due à la nature extrêmement instable du milieu hydrothermal en termes de fluctuations thermiques, et notamment sur la dorsale Est Pacifique où ces conditions sont beaucoup plus instables.

L'action de la recombinaison entre des allèles légèrement divergents conduit à augmenter la valeur du D de Tajima et, de ce fait, pourrait expliquer que les valeurs de D observées pour le Pacifique sont moins négatives (et non significatives) que celles observées pour les groupes atlantiques.

Les séquences issues des populations atlantiques sont également remarquables par le fait qu'elles forment deux groupes très divergents de tous les autres : les groupes 7 et 8. Le groupe 7 pourrait être le plus contraint : son rapport d_N/d_S (paramètre ω) est le plus faible de toutes les branches, sa diversité nucléotidique est la plus faible de tous les groupes, le nombre de recombinants est également le plus faible et le D de Tajima est très significativement négatif ($P < 0,001$), laissant supposer que ce paragraphe pourrait être sous forte sélection purifiante pour maintenir la fonction ancestrale. A l'inverse, le groupe 8 semble le moins contraint avec : la plus forte valeur du rapport d_N/d_S (ω), la plus forte diversité nucléotidique et de nombreux recombinants.

Compte tenu de l'ensemble des analyses sur le gène HSP70, nous pouvons proposer un scénario d'évolution complexe au niveau de ce gène avec deux séries de duplications et un événement de spéciation qui permettrait d'expliquer les observations.

Une duplication dans l'espèce ancestrale a conduit à la formation de deux lignées de gènes paralogues. Après spéciation, ces paralogues ont évolué de nouveau par duplication dans les deux nouvelles espèces, générant une famille multigénique. Après duplication, les pressions de sélection qui agissent sur chacune des copies sont souvent très différentes, l'un des paralogues évoluant sous forte sélection purifiante (maintien de la fonction), l'autre pouvant présenter un relâchement des pressions de sélection voire une sélection positive pour évoluer vers une nouvelle fonction ou une fonction plus spécifique (néo-fonctionnalisation). Ce mécanisme constitue l'un des moteurs de l'évolution et, permet d'expliquer pourquoi certaines protéines peuvent évoluer vers de nouvelles fonctions (Sidow 1996).

Plusieurs hypothèses peuvent expliquer la présence de nombreuses copies du gène chez *Bathymodiolus* : (1) la grande taille efficace des populations ne permet pas l'élimination rapide des gènes non fonctionnels (Watterson 1983, Lynch & Force 2000), ou (2) un taux de duplication plus élevé dans ce gène lié à une induction de la duplication par stress thermique.

Il est difficile de faire un choix entre les 3 modes d'évolution évoqués en introduction : néo fonctionnalité, sous-fonctionnalité, ou non fonctionnalité.

Aucun élément ne nous permet d'affirmer que l'une des copies ait pu acquérir une nouvelle fonction.

Dans le cas de l'hypothèse de sous-fonctionnalité, les duplicats auraient évolué de la manière conjointe en accumulant des mutations chacun de leur côté.

La dernière hypothèse est celle de la non-fonctionnalité. Elle correspond à la perte de fonction de l'une des copies de gène. Il est difficile de savoir si les allèles amplifiés possèdent une réelle activité protéique ou si alors il s'agit de copies « nulles » c'est-à-dire qui ont perdu leur fonction. On peut cependant éliminer la non fonctionnalisation des hypothèses concernant HSP car tous les oméga obtenus sur ce gène vont dans le sens de la sélection purificatrice et donc un maintien de la fonction.

En regardant l'ensemble des séquences recapturées, 9 d'entre elles possédaient un codon stop après corrections (séquençage dans les 2 sens et vérification sur le chromatogramme). Ces séquences ne présentent pas de signature caractéristiques de pseudo-gènes (elles ne présentent pas d'accumulation de mutations non-synonymes aberrantes) mais pourraient être des allèles nuls (transcrits mais non traduits). Ces copies avec codon stop constituent des allèles dont les fonctions sont cassées et représentent le fardeau génétique de *Bathymodiolus*. Ceci est d'autant plus probable que comme les gènes sont encore proches les uns des autres, une copie cassée n'est probablement pas très délétère.

V.5. Recherche de sélection positive sur les lignées paralogues de la Ferritine 1 (GF1)

Ce gène est constitué de deux gènes paralogues (mis en évidence au chapitre I, M&M), qui ont été traités indépendamment dans les précédents chapitres. Ici, l'objectif est d'étudier le gène dans sa globalité afin de mieux comprendre les mécanismes qui ont pu affecter son évolution après duplication. Il est d'ailleurs intéressant de noter que le gène GF1 n'est pas dupliqué chez l'huître *Crassostrea gigas*. Les analyses précédentes ont montré que les deux gènes étaient peu résolutifs dans la diagnose des espèces proches (i.e. *B. puteoserpentis* vs *B. azoricus* ; *B. thermophilus* vs *B. aff. thermophilus*) et qu'ils pouvaient présenter des coalescences profondes et des recombinants entre espèces. De plus, le gène GF1_B montrait des valeurs du D de Tajima significativement positives laissant supposer que des effets de sélection diversifiante étaient éventuellement possibles.

Les 167 séquences de haute qualité provenant des 2 gènes GF1_A et GF1_B ont été regroupées pour produire un arbre global en Neighbour-Joining (Figure V.13)

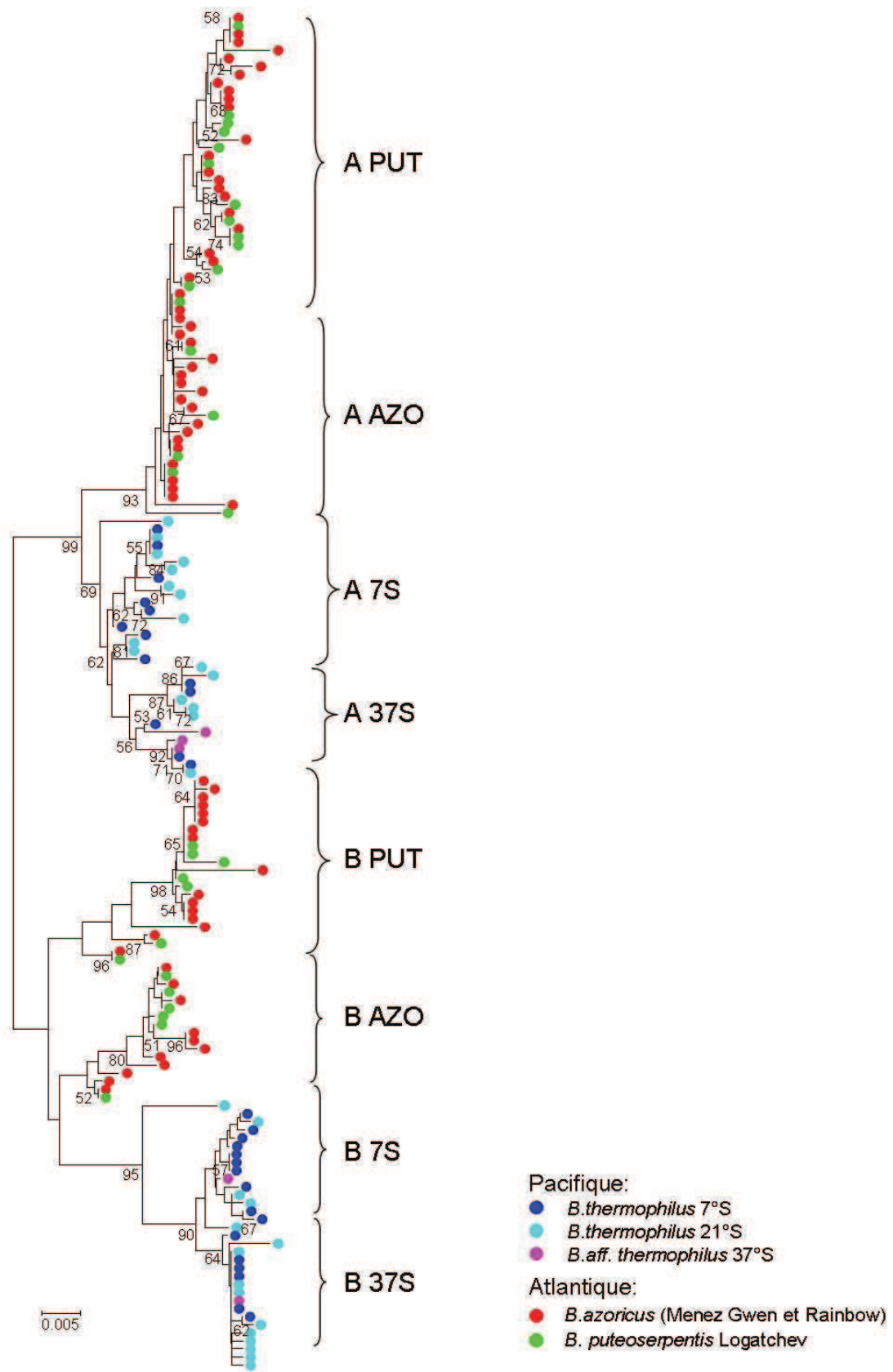


Figure V.13. Arbre en NJ réalisé à partir de la totalité des séquences haute qualité du GF1. Distance de Tamura et Nei, 1000 bootstraps. Les groupes définis correspondent aux séquences utilisées pour réaliser les 8 séquences consensus.

V.5.1. Recherche de différences au niveau des pressions sélectives entre les lignées

Le jeu de données complet (167 séquences) était bien trop lourd pour être utilisé dans une recherche de sélection positive entre lignées à l'aide du logiciel PAML3.14. Afin de réduire le temps d'analyse et faciliter l'interprétation des résultats, nous avons donc effectué les analyses sur des séquences consensus représentant les 8 clades principaux identifiés sur l'arbre NJ (Figure V.13). Il n'existe pas de divergence fixée, les espèces s'hybrident au sein d'un même océan. Aussi, le nom donné aux séquences consensus est purement arbitraire et ne correspond donc pas à une espèce donnée, mais plutôt à l'allèle dominant caractérisant le groupe utilisé pour faire la séquence consensus. Une séquence de GF1 de *Crassostrea gigas* récupérée sur NCBI (AJ564625) a été utilisée pour enracer les arbres. Les analyses d'évolution sont réalisées sur les codons et ne concernent donc que la région exonique de notre gène (195 bp). L'arbre utilisé par PAML avec l'optimisation de ses longueurs de branches sont illustrés sur la Figure V.14. Il faut noter ici que la vitesse d'évolution des lignées atlantiques et pacifiques n'est pas la même pour les 2 gènes GF1_B et GF1_A (OTUs atlantiques près de la racine et longueur de branches extrêmement réduites).

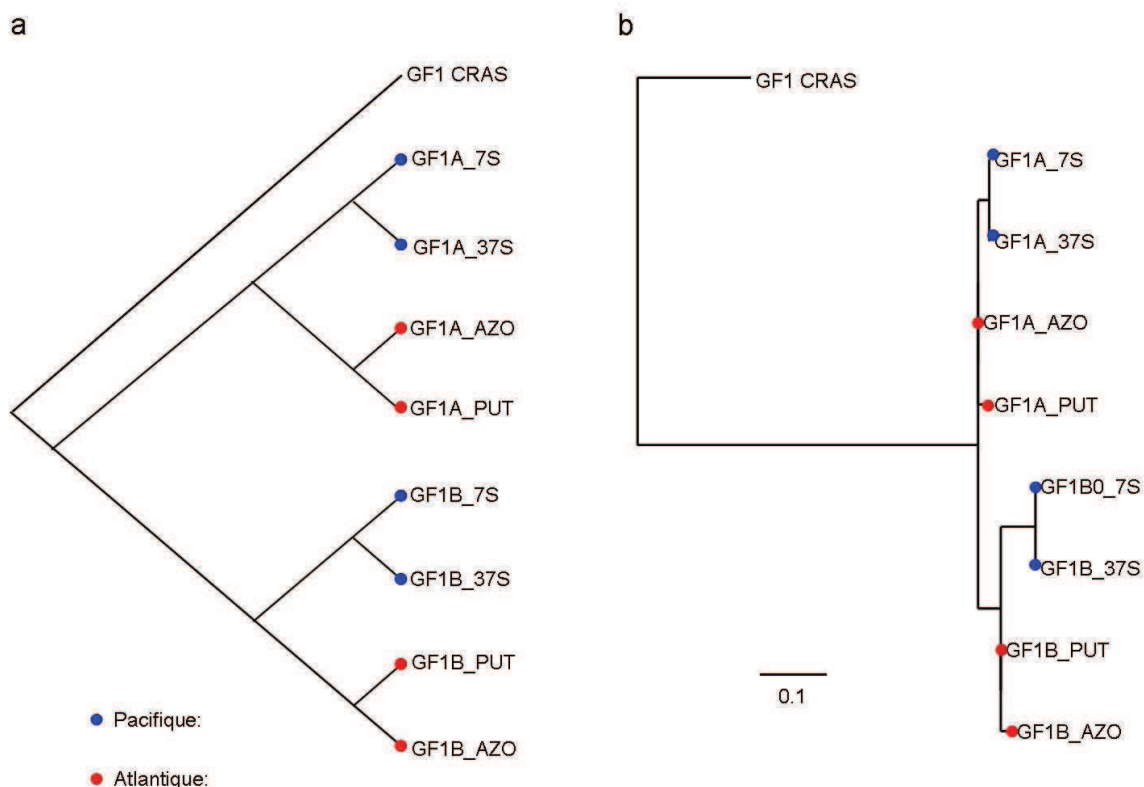


Figure V.14. Arbres avec les séquences consensus et le groupe externe utilisés par PAML. 'a' : topologie théorique de l'arbre utilisée montrant les événements de duplication et de spéciation; 'b' arbre ML optimisant les longueurs de branches à l'aide du modèle de substitution des codons de Goldman & Yang (1994) sous la topologie de 'a'.

L'arbre consensus est testé sous 3 modèles de substitutions des codons à l'aide du programme CodeML de PAML3.14 afin de déterminer le modèle d'évolution qui caractérise le mieux le jeu de données sous contrainte de notre arbre théorique. Le modèle 0 correspond à une évolution constante du rapport dN/dS (ω) dans toutes les lignées (branches) ; le modèle 1 correspond à une valeur de ω variable pour chaque branche de l'arbre ; enfin le modèle 2 correspond à une valeur de ω fixée pour l'ensemble de l'arbre sauf une branche choisie pour laquelle ω est libre de varier. Le choix de la branche à tester est fonction des résultats obtenus par le modèle 1 et permet de confirmer si un effet sélectif existe vraiment sur une lignée donnée. Les valeurs obtenues sont consignées dans le Tableau V.8.

Tableau V.8. GF1 : Comparaison des trois modèles d'évolution sous PAML. **L** : longueur de l'arbre en nombre de substitutions nucléotidiques par codon ; **kappa** : rapport transitions / transversions, **lnL** : Valeur d'ajustement des données à l'arbre sous le modèle choisi, test de vraisemblance : **2*lnL (ddl)** : 2 fois la différence entre le modèle choisi et le modèle 0 ; **ddl** : degré de liberté utilisé pour la comparaison dans la table de χ^2 (égal à la différence dans le nombre de paramètres estimés pour chaque modèle) * : différence significative à $p < 0,05$.

	dN/dS par branche	L	dN total arbre	dS total arbre	kappa	lnL	2* Δ lnL (ddl)
modèle 0 :	constant	0,816	0,072	1,150	1,022	-392,69	
modèle 1 :	variable	0,931	0,071	1,330	1,096	-386,88	11,62 NS (14)
modèle 2 :	2 branches variables	0,855	0,071	1,213	1,047	-389,58	6,22 * (2)

Les modèles 0 et 1 ne sont pas significativement différents l'un de l'autre. Il semble donc que le taux d'évolution du dN/dS soit similaire entre les branches. Cependant, bien que non-significativement différent du premier modèle, le modèle 1 fournit 2 branches (GF1_A PUT, GF1_B AZO) pour lesquelles la valeur de ω est supérieure à 1. Lorsque que ces branches sont sélectionnées pour varier librement, la comparaison entre le modèle 0 et le modèle 2 révèle une différence significative (6,22 ; $P < 0,05$). Il faut noter que dans les modèles 1 et 2, deux branches terminales de GF1_A PUT et GF1_B AZO, présentent des valeurs de dN/ dS très élevées ($\omega = 999$ puisque la seule mutation fixée est non-synonyme : S=0 et N=1). En fait ces branches ne possèdent pas suffisamment de mutation pour affirmer quoi que ce soit.

V.5.2. Analyse des pressions sélectives sur les codons

Dans cette deuxième partie de l'analyse, nous nous intéressons aux pressions de sélection qui s'appliquent sur les 65 codons afin de déterminer si l'évolution de chaque codon est neutre, sous sélection purifiante, ou au contraire sous sélection positive. Pour cela trois

types de modèles ont été testés : un modèle presque neutre de référence (« nearly neutral » model) pour lequel les codons sont classés en 2 catégories (délétère : ω proche de zéro et neutre $\omega = 1$) et testés par rapport à l'ensemble des branches de l'arbre (modèle NS1), un modèle de sélection positive pour lequel les codons sont classés en 3 catégories (délétère : ω proche de zéro, neutre : $\omega = 1$ et sélection positive : $\omega > 1$) et testés par rapport à l'ensemble des branches de l'arbre, et enfin un modèle testant les codons d'une branche en particulier par rapport à l'ensemble des branches restantes (modèle BS1 : Branch-Site 1 branche) dans lequel les codons sont également classés en 3 catégories.

Tableau V.9. Comparaison des modèles de sélection sur les codons (modèles NS1 et NS2). **L** : longueur de l'arbre en nombre de substitutions nucléotidiques par codon ; **κ** (Kappa): Rapport transitions / transversions ; **lnL** : Valeur d'ajustement des données à l'arbre sous le modèle choisi ; **Proportion** : fraction des codons classés dans chacune des catégories ; **ω** : valeur moyenne de Omega pour chaque classe de codons.

Modèle		$d_N/d_S=0$	$d_N/d_S=1$	$d_N/d_S>1$	L	κ	lnL
Presque neutre	Proportion	0,9528	0,0472	-	0,884	0,890	-390,71
	ω	0,0351	1	-			
sélection positive	Proportion	0,9593	0,00000	0,0407	0,899	0,889	-390,69
	ω	0,0367	1	1,3223			

Le premier modèle à avoir été testé est le modèle NS2 avec le modèle NS1 comme modèle de référence pour l'hypothèse nulle d'évolution presque neutre. Les 2 modèles sont comparés avec une ddl de 1 (un paramètre de plus estimé par le modèle NS2 : la classe de $\omega > 1$). Les résultats des analyses sont regroupés dans le Tableau V.9. La différence des LnL entre les deux modèles est minimale. Quelle que soit l'hypothèse d'évolution testée, (neutre ou avec une fraction des codons sous sélection positive), la majorité des codons (plus de 95%) est considérée comme étant sous sélection purifiante. En revanche, les 5% de codons restant considérés comme étant sous sélection neutre dans la première hypothèse, sont finalement classés comme étant sous sélection positive dans le modèle NS2. Les codons identifiés sous sélection positive sont les codons 3 et 22. En ce qui concerne le codon 22, il s'agit d'une mutation d'une valine V (invariante chez *Bathymodiolus*) en tyrosine Y dans la branche de l'outgroup *C. gigas*. En revanche la situation du codon 3 est plus intéressante. Le codon indiqué comme étant sous sélection positive code pour une isoleucine I, qui est mutée en leucine L pour la lignée *C. gigas* et en phénylalanine F dans la lignée GF1_B_Azoricus. Ces codons sont significatifs (>95%) avec le test NEB (Naive Empirical Bayes) qui teste si le codon est sous sélection positive par rapport à un « background » sous sélection négative (ω

proche de 0), mais pas le test BEB (Bayes Empirical Bayes) ($p=40\%$) qui teste si le codon est sous sélection positive par rapport à un « background » neutre ($\omega = 1$).

Le deuxième modèle d'évolution testé (le modèle BS1) compare les valeurs sélectives de chaque codon dans une branche donnée contre le « background » de l'arbre. Dans cette analyse, les codons sont répartis en 4 classes : 0, 1, 2a (ω « foreground » > 1 contre ω « background » proche de zéro) et 2b (ω « foreground » >1 contre ω « background » = 1). Les valeurs de LnL sont testées contre les valeurs du modèle NS1 ($2*\Delta\ln L$) selon une table de χ^2 . Les résultats de PAML sont présentés dans le Tableau V.10

Tableau V.10. Tableau : résultats des modèles BS1A (branche testée = GF1_A PUT) et BS1B (branche testée = GF1_B AZO).

Modèle		$d_N/d_S=0$	$d_N/d_S=1$	$d_N/d_S > 1$ (2a=avec ω arbre = 0)	$d_N/d_S > 1$ (2b=avec ω arbre = 1)	L	κ	lnL
BS1A (GF1_A PUT)	Proportion	0,951	0,049	0	0	0,914	0,879	-388,69
	Background ω	0,028	1	0,028	1			
	Foreground ω	0,028	1	66,812	66,812			
BS1B (GF1_B AZO)	Proportion	0,953	0,047	0	0	0,884	0,890	-390,71
	Background ω	0,035	1	0,035	1			
	Foreground ω	0,035	1	1	1			

Aucun de ces deux modèles ne présente de valeur de lnL statistiquement différente de celle obtenue avec le modèle NS1 ($2*\Delta\ln L = 4,06$ et $2*\Delta\ln L= 0$; respectivement pour les modèles BS1A et BS1B).

Un seul site présente une probabilité postérieure d'appartenir à la classe sous sélection positive ; il s'agit du site 15, M (BEB $> 97\%$). Ce codon correspond à la mutation d'une leucine L en méthionine M, dans branche terminale de la lignée GF1A_PUTEO.

V.5.3. Analyses de la diversité nucléotidique au sein des 8 lignées

La première partie de l'analyse du GF1 était focalisée sur les pressions évolutives affectant la région codant du gène. Dans cette deuxième partie, nous avons comparé les niveaux de diversité nucléotidique sur la totalité du locus (régions exoniques et introniques) entre les différents groupes les résultats sont consignés dans le Tableau V.11.

Tableau V.11. Principaux indices de la diversité nucléotidique mesurés sur l'ensemble des séquences correspondant aux 8 groupes consensus précédemment définis (N=175 seq). **N**: Nombre de séquences, **Hd** : diversité haplotypique, π : estimateur du $4Ne\mu$ à partir du nombre moyen de différences nucléotidiques par site (Nei 1987), θ_w : estimateur du $4Ne\mu$ par paire de bases en utilisant le nombre de sites polymorphes (Watterson 1975), **K** : nombre moyen de différences nucléotidiques, **Nb sites** : nombre de sites potentiel pour chaque type de mutation, **Nb mut** : Nombre de mutation, π_s : diversité nucléotidique dans chacune des régions du gène (synonyme, non-synonyme et intronique).

Diversité générale					
Lignée	N	Hd	π	θ_w	k
GF1_A_7S	14	0,934	0,0065	0,0086	4,51
GF1_A_37S	20	0,974	0,0094	0,0097	6,58
GF1_A_puteo	40	0,958	0,0071	0,0120	5,25
GF1_A_AZO	24	0,949	0,0049	0,0079	3,66
GF1_B_7S	16	0,967	0,0041	0,0057	3,03
GF1_B_37S	19	0,789	0,0044	0,0104	3,22
GF1_B_puteo	21	0,890	0,0050	0,0083	3,68
GF1_B_AZO	21	0,967	0,0111	0,0108	8,39

Analyse par région fonctionnelle									
Lignée	Synonyme			Non synonyme			Intronique		
	nb sites	nb mut.	π_s	nb sites	nb mut.	π_s	nb sites	nb mut.	π_s
GF1_A_7S	38,62	0	0	156,38	2	0,0018	541,62	17	0,0078
GF1_A_37S	38,70	4	0,0203	156,30	1	0,0006	541,70	23	0,0120
GF1_A_puteo	39,02	2	0,0038	155,98	3	0,0042	587,02	36	0,0078
GF1_A_AZO	38,92	0	0	156,08	1	0,0031	588,92	21	0,0054
GF1_B_7S	39,17	1	0,0032	155,87	2	0,0016	585,17	12	0,0047
GF1_B_37S	39,20	4	0,0108	155,8	3	0,0020	585,20	25	0,0050
GF1_B_puteo	38,52	1	0,0025	156,48	2	0,0012	584,52	20	0,0060
GF1_B_AZO	38,25	0	0	156,75	2	0,0027	596,25	28	0,0134

Le nombre de séquences composant chacune des lignées est assez équilibré car sept groupes comprennent de 14 à 24 séquences. Le groupe de GF1_A puteo est le mieux représenté avec 40 séquences. La diversité haplotypique est très élevée (Hd 0,949 pour la majorité des groupes). Les indices de diversité nucléotidique sont homogènes entre les différentes lignées (pas de différence significative entre les espèces des deux océans ou entre les deux paralogues).

Si l'on observe le polymorphisme en fonction des différentes régions fonctionnelles du gène, c'est le groupe GF1 37°S qui présente le maximum de diversité nucléotidique et cela pour chacun des deux paralogues. Chaque groupe présente un polymorphisme non synonyme (de 1 à 3 mutations), mais pas forcément de mutations synonymes sur l'exon. En revanche, tous possèdent un fort polymorphisme intronique. Ces données sont en accord avec ce que l'on peut attendre d'une accumulation neutre des mutations dans le polymorphisme des espèces, et ce pour les 2 lignées paralogues.

Aucune des lignées de GF1 ne présente d'excès de mutations non-synonymes. La plus forte diversité observée pour les groupes appelés '37°S' provient certainement du regroupement d'allèles provenant des populations différenciées de *Bathymodiolus aff. thermophilus* et *B. thermophilus*. En effet, même si ces populations ne présentent pas de divergence stricte, il est possible que leurs allèles soient en moyenne plus divergents que ceux des autres groupes.

V.5.4. Conclusion concernant l'évolution du gène GF1

Nos premières analyses sur les dN/dS mettent en évidence que certaines lignées accumulent des mutations non-synonymes plus rapidement que les autres (GF1_A_puteo et GF1_B_Azo). Un tel résultat peut être expliqué de deux manières : (1) par un relâchement des pressions de sélection, ou (2) par de la sélection positive sur certains codons. Il faut cependant relativiser ce résultat car il semble que le groupe externe, très divergent, ait pu entraîner une certaine imprécision dans les estimations des pressions sélection concernant les lignées de *Bathymodiolus*.

Dans le cas où le relâchement des pressions de sélection serait réel, on s'attend à ce que les branches concernées présentent une accumulation de mutations dans le polymorphisme sous-jacent, augmentant ainsi la diversité nucléotidique observée. Pourtant nos analyses de la diversité de chacune des lignées ne mettent pas en évidence ce type d'évolution. Les lignées possèdent certainement des changements d'acides aminés pouvant avoir une valeur adaptative. Cela a été mis en évidence par les tests d'évolution réalisés sur les codons. Ces résultats sont résumés dans la Figure V.15.

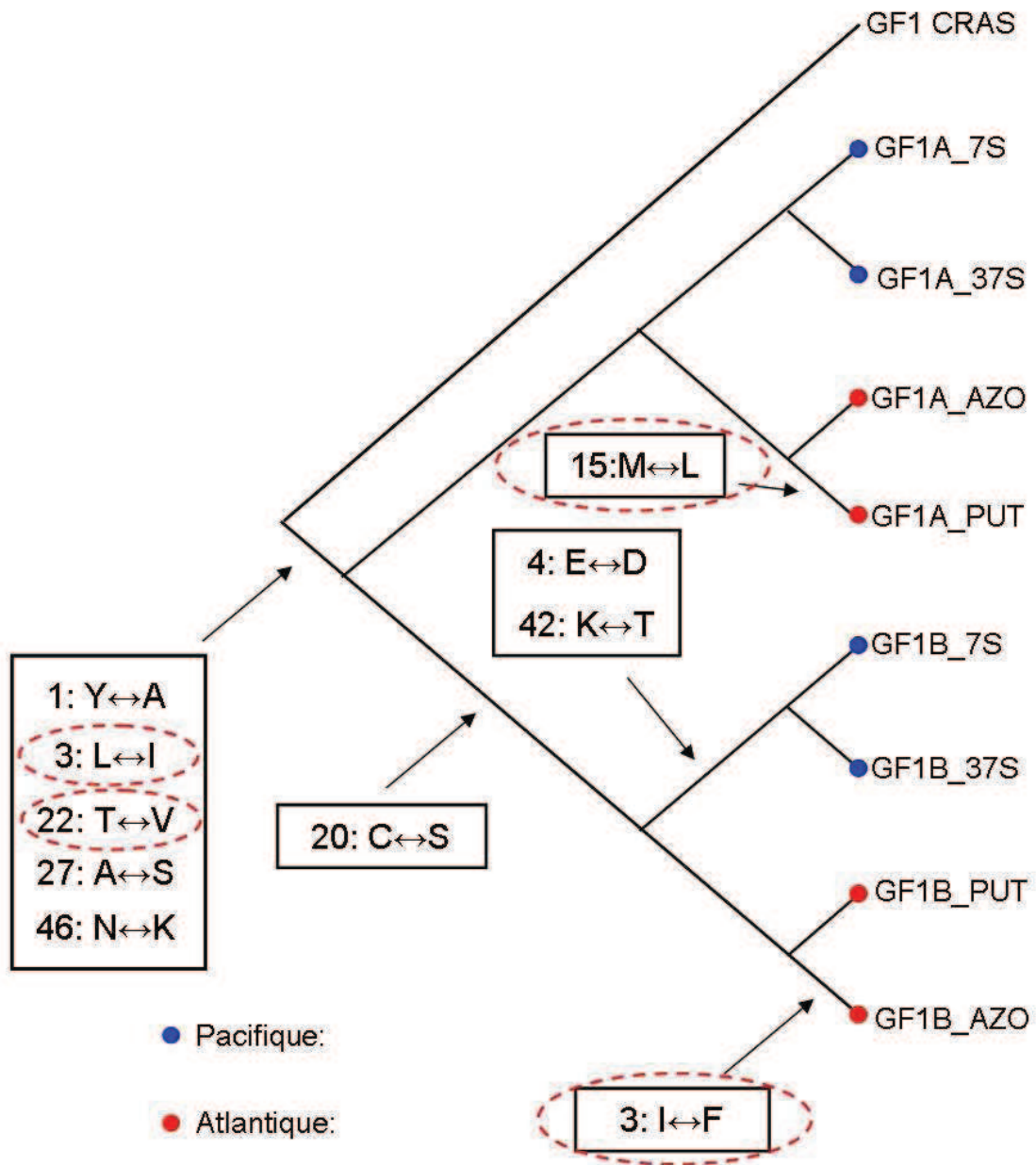


Figure V.15. Arbre schématisé représentant les changements d'acides aminés ayant eu lieu au cours de l'évolution du gène GF1 chez les bathymodioles. Encadrées en noir : mutations des acides aminés observées pour les codons dont la position est indiquée au début de la ligne. Flèches : branche de l'arbre concernée par les remplacements. Entourés en rouge : remplacements pouvant être sous sélection positive.

Cette figure illustre les codons pour lesquels les analyses ont mis en évidence de la sélection positive. Deux codons concernés impliquent la branche *Crassostrea gigas*, mais deux autres concernent directement les branches terminales de deux lignées du gène GF1 chez les bathymodioles. Toutes les deux sont situées au niveau des espèces atlantiques et affectent l'un et l'autre des dupliquats.

Ces résultats sur les codons 3 et 15 diagnostiqués respectivement par les tests BS1A (branche site 1) et NS2 sont confirmés par le résultat du modèle M2 qui a mis en évidence que les lignées GF1_A_Puteo et GF1_B_Azo présentent une évolution différente des autres lignées (avec des valeurs de ω très élevées). Cependant, ces modèles de sélection n'apparaissent pas meilleurs que le modèle presque neutre pour expliquer les données.

V.6. Conclusion concernant les gènes dupliqués

La duplication des gènes est un processus important dans l'évolution d'une famille multigénique parce qu'il s'agit de la première étape dans l'acquisition de nouvelles fonctions. Une fois le gène dupliqué, un certain nombre de choses peuvent lui arriver. Lorsqu'une seule copie du gène est nécessaire, la nouvelle copie devient automatiquement redondante. Dans ce cas, le nouveau gène dupliqué est libre de contraintes sélectives et va pouvoir accumuler des mutations de façon neutre. Il pourra éventuellement devenir un pseudo-gène ou encore être complètement supprimé du génome (Page & Holmes 2000). Aucun des gènes de notre étude ne semble présenter ce type d'évolution. Même si plusieurs allèles recapturés présentaient des codons stop dans leurs séquences, ces allèles ne présentaient pas d'excès de polymorphisme permettant d'en conclure qu'il puisse s'agir de pseudo-gènes et de plus ont été supprimés des analyses. Ces copies dont la fonction est cassée par les codons stop illustrent le fardeau génétique lié à la grande taille efficace des *Bathymodiolus*.

Une autre possibilité concernant le devenir des nouvelles duplications est l'acquisition de nouvelles fonctions, mais cette alternative est cependant rare (Lynch & Conery 2000). Selon l'hypothèse de Ohta (1988), la sélection positive Darwinienne permettant d'accumuler suffisamment de mutations favorables pour permettre au nouveau gène d'exercer une fonction modifiée ou pourquoi pas une nouvelle fonction nécessite souvent plusieurs mutations successives à un même codon. Selon cette même théorie, les gènes dont la fonction est fixée depuis longtemps auraient tendance à évoluer uniquement par dérive génétique. Si l'on applique ce principe à nos données, on peut imaginer que pour les gènes GF1 et HSP70, les deux lignées pour lesquelles des codons sont sous sélection positives sont certainement celles qui sont les plus récemment dérivées.

Chapitre VI. Conclusion générale et perspectives sur les bivalves marins

VI.1. Synthèse des résultats obtenus sur *Bathymodiolus*

Ce travail de thèse visait à étudier les mécanismes d'évolution affectant les espèces vivant dans des milieux considérés comme extrêmes : les espèces de *Bathymodiolus* du milieu hydrothermal profond.

Une approche multi-locus utilisant des généalogies d'allèles et des méthodes de coalescence a été mise en œuvre sur deux complexes d'espèces provenant pour l'un, de la dorsale médio-atlantique, et, pour l'autre, de la dorsale du Pacifique Est. Cette approche nous a permis de nous intéresser à l'évolution des gènes de ce bivalve particulier dans un contexte de spéciation par vicariance et hybridation secondaire. Au cours de cette thèse, nous nous sommes plus spécifiquement intéressés aux domaines suivants : (1) l'histoire évolutive de la famille des *Bathymodiolinae*, (2) l'hybridation et l'introgession de gènes de part et d'autre d'une zone d'hybridation décrite dans l'Atlantique, (3) l'évolution de la colonisation de la dorsale du Pacifique Est et la détection d'une zone d'hybridation entre deux espèces géographiques, et (4) le rôle éventuel de la sélection dans l'évolution de plusieurs gènes utilisés dans notre analyse.

Quatre conclusions majeures peuvent ainsi être dégagées de cette étude :

(1) La phylogénie complète de la famille des *Bathymodiolinae* pour le gène mitochondrial CO1 à l'aide de séquences nouvelles nous a permis de montrer que 30 espèces génétiquement différenciées composent le genre *Bathymodiolus*. En revanche, les nœuds profonds associés à cette phylogénie sont peu robustes et ne permettent pas de dégager une histoire globale des *Bathymodiolinae* et *Modiolinae*. Aucun paramètre écologique (profondeur, type d'environnement réduit, ou type de symbiose) ne permet d'expliquer de manière certaine, les regroupements d'espèces. Chez les *Bathymodiolinae sensu stricto*, nous avons néanmoins pu proposer que les espèces des suintements froids étaient probablement à

l'origine des espèces hydrothermales atlantiques et pacifiques. L'étude de l'évolution du gène mitochondrial à travers les différentes lignées de *Bathymodiolinae* nous a amené à penser que ce locus pouvait faire l'objet d'une très forte sélection purificatrice (aucun codon n'évoluant de façon neutre).

(2) En utilisant une approche multi-locus sur les deux espèces proches de *Bathymodiolus* peuplant la dorsale médio-atlantique, nous avons mis en évidence que l'introggression d'allèles affectait ces espèces bien au-delà de la zone d'hybridation précédemment décrite et que l'hybridation devait être relativement ancienne. Le génome des deux espèces semble introgressé sur l'ensemble de l'aire de répartition et les flux géniques sont majoritairement orientés du nord vers le sud (de *B. azoricus* vers *B. puteoserpentis*). Nous avons montré dans cette étude que la zone hybride correspond très probablement à un contact secondaire.

(3) L'étude menée sur les populations de *Bathymodiolus* de la dorsale du Pacifique Est a mis en évidence l'existence d'une nouvelle espèce au sud de l'île de Pâques (site de 37°S). Cette espèce est également capable de s'hybrider avec l'espèce *B. thermophilus* à la suite d'un contact secondaire qui semble récent et le niveau d'introggression est très variable selon les locus. Contrairement à ce qui a été publié par de précédents auteurs, aucune barrière à la dispersion ne semble exister le long de la dorsale sud EPR même si une tendance pour un isolement par la distance semble exister. En revanche plusieurs locus mettent en évidence de la divergence entre les populations du nord (9°50N et 7°25S) et les populations situées plus au sud (de 17°S à 21°33S).

(4) Quatre des gènes utilisés dans notre étude ne se comportent pas de manière strictement neutre. Le gène de la sulfotransférase, extrêmement variable chez les eucaryotes, semble au contraire très conservé entre les espèces de *Bathymodiolus* étudiées. Son évolution est certainement lente, mais marquée dans la coalescence. Il semble également qu'il existe des traces d'introggression entre les deux océans. Le deuxième intron du gène du Facteur d'Elongation 1 alpha subit également l'action de la sélection purifiante, ce qui n'est pas forcément attendu pour des séquences *a priori* non-codantes. Deux autres gènes (GF1 et HSP70), semblent présenter des duplications spécifiques du génome des *Bathymodiolus* et présentent des évolutions plus complexes avec des lignées (ou codons) présentant des

évolutions plus rapides dues soit à des relaxations de pressions de sélection post-duplication, soit à de la sélection positive ou à de l'évolution concertée.

VI.2. Conclusions évolutives concernant le genre *Bathymodiolus*

De manière générale, nos *a priori* concernant le milieu hydrothermal sont que les conditions environnementales supposées « extrêmes » auxquelles sont confrontées les *Bathymodiolus* (pression hydrostatique, température, radicaux libres issus de l'oxydation du soufre (RSS), métaux lourds (Seyfried & Janecky 1985), radioactivité (Kadko 1996) allaient induire chez ces espèces des adaptations particulières capables d'engendrer de la spéciation écologique associée à des effets sélectifs forts (Jollivet et al. 1995). Pourtant nos résultats n'ont pas identifié de signature claire d'accumulation de mutations liées à de la sélection positive dans les branches, entre les différentes espèces. Bien évidemment, le faible nombre de locus testés ne permet pas d'exclure cette hypothèse car les gènes 'verrous' pouvant être associés à un tel mode de spéciation sont peu fréquents. Il semble néanmoins que l'existence de zones d'hybridation à la fois dans l'Océan Pacifique et l'Océan Atlantique associées à de l'introggression relativement ancienne soit plutôt en faveur d'événements de spéciation par vicariance avec formation de zones de contact secondaire comme cela a pu être abondamment décrit dans la littérature sur les milieux terrestres (Boursot et al. 1993), (Latta & Mitton 1999), dulçaquicole (Lu et al. 2001) et côtiers (Palumbi 1994).

Une autre idée reçue est que le milieu hydrothermal engendre de nombreux déséquilibres démographiques par l'intermédiaire des épisodes récurrents d'extinction/recolonisation. De tels épisodes ont en effet été bien décrits sur le terrain et suivent des événements tectoniques majeurs ou des épanchements de lave à la suite d'épisodes de volcanisme effusif (Fustec et al. 1987), (Haymon et al. 1991), (Jollivet 1993), (Tunnicliffe et al. 1997). Ces épisodes d'extinction des populations sont également révélés par les nombreux cimetières de coquilles qui jonchent les sites inactifs (Grassle 1985).

Les épisodes d'extinction et de recolonisation devraient se traduire par des goulots d'étranglement dans les populations et donc par des diversités plus faibles et des coalescences d'allèles rajeunies (arbre en râteau avec une faible structure). Il n'en est rien car nos résultats montrent au contraire une très forte diversité nucléotidique et haplotypique chez les bathymodioles. On peut donc conclure que les effets démographiques des épisodes de colonisation et de recolonisation n'ont pas l'effet que nous aurions pu en attendre, même pour

les sites à durée de vie très courte de la dorsale EPR, considérée comme ultra rapide. Pour comprendre pourquoi les phénomènes d'extinction/recolonisation n'ont qu'un effet limité sur la taille efficace des populations de bathymodiales, il nous faut considérer que les flux de gènes entre populations même éloignées doivent être très conséquents, et ce malgré la nature extrêmement fragmentée du milieu hydrothermal. Les modèles théoriques de métapopulations prédisent en effet que la diversité génétique ne peut être maintenue avec un fort taux d'extinction que si l'espèce suit un modèle de type « pool de migrants » dans lequel le nombre de colonisateurs (recolonisation d'habitats vides) est important et d'origine multiple (contrairement à une dispersion de proche en proche). (Slatkin 1977), (Whitlock & McCauley 1990), (Pannell & Charlesworth 1999). Ceci implique l'existence de nombreuses populations le long de l'axe des dorsales et des échanges non limités entre les différents champs hydrothermaux pour assurer la recolonisation rapide et efficace des nouvelles zones d'émission. Des travaux ont montré que la durée de vie larvaire des bathymodiales pouvait être longue (> 4 mois) et que le taux de recrutement pouvait être localement très important (Comtet & Desbruyères 1998).

La non-différenciation des populations de *Bathymodiolus* le long de la dorsale du Pacifique Est démontre que les différentes failles et les courants ne jouent pas le rôle de barrière. En revanche, nos données sur trois gènes (CO1, ITS2 et SULFO), laissent penser qu'il existe au moins trois ensembles de populations le long de la dorsale pacifique : au nord (9°5N-7°S), au centre (21°S) et au sud (à 37°S). Ces ensembles pourraient refléter des épisodes passés d'isolement géographique vicariant qui auraient fait suite des événements à tectoniques majeurs, comme la création d'une importante faille transformante ou d'une microplaque.

Les conditions environnementales extrêmes (température, pression...) et la sélection ne semblent pas être les moteurs principaux de l'évolution des espèces de *Bathymodiolus*. L'évolution de ces espèces semble au contraire contraint par l'histoire tectonique de la dorsale et plus particulièrement par l'apparition récurrente de failles transformantes et autres phénomènes disruptifs (microplaques). L'apparition de ces failles transformantes et le décalage de l'axe des dorsales cisailent l'aire de répartition des espèces ancestrales, favorisent les phénomènes de spéciation en allopatrie et génèrent ainsi de nombreuses espèces proches (vicariantes) ou cryptiques (vésicomidés (Goffredi et al. 2003) ; moules (Won et al. 2003) ; crustacés (Guinot & Hurtado 2003) ; polychètes (Hurtado et al. 2002, Hurtado et al.

2004) ; gastéropodes (Matabos et al. 2008)). A ce titre, le réseau formé par les dorsales océaniques peut être considéré comme une véritable machine à créer des espèces par isolement géographique. Les espèces nouvellement créées sont autant d'espèces proches qui pourront potentiellement s'hybrider à la faveur de contacts secondaires.

Un autre moteur de l'évolution permettant d'augmenter la diversité génétique d'une espèce et ses potentialités évolutives à long terme a été mis en évidence au cours de cette thèse, il s'agit de la duplication des gènes. Notre étude a révélé que sur 9 gènes nucléaires étudiés trois présentent un profil multi-copies chez les espèces de *Bathymodiolus* : les gènes lysozyme (cf. le matériel et méthodes), GF1 et HSP70. L'espaceur ribosomique intergénique ITS2 (Odonnell & Cigelnik 1997) et la HSP 70 (Bettencourt & Feder 2001, Evgen'ev et al. 2004) avaient quant à eux déjà été décrits comme possédant des duplications en tandem.

Les gènes en copies multiples sont considérés comme des éléments moteurs de l'évolution (Ohta 1988). La grande taille efficace des bivalves permet à la sélection de trouver plus rapidement les mutations favorables qui peuvent par exemple conduire à de nouvelles fonctions. Pour les deux gènes multi-copies étudiés en détail (HSP et GF1), nous avons pu mettre en évidence que certaines lignées avaient subi un relâchement des pressions de sélection et pouvaient éventuellement présenter de l'évolution concertée ou des codons sous sélection positive.

On peut désormais se demander si cette haute fréquence des gènes en copies multiples est spécifique des *Bathymodiolinae*, des espèces hydrothermales ou alors s'il est fréquent chez les bivalves, mais nous ne disposons pas de suffisamment de données comparatives pour répondre à cette question. Il faudrait aussi savoir si toutes les copies observées sont fonctionnelles ou si leurs fonctions sont cassées et qu'elles ne représentent plus qu'un fardeau génétique.

VI.3. Comparaison inter-genres chez les bivalves marins

Au cours des précédents chapitres, nous avons détaillé les résultats obtenus pour un genre de bivalve particulier, celui des *Bathymodiolus*. Par conséquent, les observations et les résultats ne peuvent pas être généralisés à l'ensemble des bivalves marins. Afin de pouvoir distinguer ce qui est propre au milieu hydrothermal et aux espèces de bathymodioles, ou au

contraire caractéristique des bivalves en général, il est nécessaire d'effectuer des comparaisons avec des espèces de bivalves côtiers.

En fait, le sujet de cette thèse s'inscrivait dans le cadre d'un projet plus global piloté par N. Bierne dont l'objectif était de comparer les pressions évolutives entre trois genres de bivalves : *Bathymodiolus*, *Crassostrea* et *Mytilus*. Pour chacun de ces genres, deux espèces proches capables de s'hybrider ont été étudiées : *Crassostrea gigas* et *C. angulata* (données acquises entre les équipes de la Station Biologique de Roscoff et de l'Institut Universitaire Européen de la Mer à Brest : A. Tanguy & D. Moraga), et *Mytilus galloprovincialis* et *M. edulis* (données acquises à la Station Méditerranéenne de Sète et ayant fait l'objet de la thèse de M. Faure (Faure 2007)). Nous allons donc présenter ici quelques données acquises en parallèle sur *Crassostrea* afin d'effectuer une comparaison de descripteurs génétiques multi-locus entre les différents genres de bivalves.

VI.3.1. Le modèle *Crassostrea*

Cette partie a pour objectif de dresser un premier tableau du travail qui a été réalisé sur le modèle *Crassostrea*. Les résultats présentés ici ont été obtenus pour dix locus nucléaires et un locus mitochondrial pour deux espèces proches : *Crassostrea gigas* Thunberg (1793) et *C. angulata* Lamarck (1819). Ces espèces originaires du Pacifique ont été introduites en Europe où elles existent désormais en populations naturelles. La distinction entre ces deux espèces n'est pas aisée car elles présentent de fortes similitudes morphologiques et aucune barrière à la reproduction n'a été observée (Huvet et al. 2002). D'un point de vue génétique, seul le locus mitochondrial COI permet de mettre en évidence une divergence nette entre les deux espèces (Boudry et al. 1998) alors que les allozymes (Buroker et al. 1979) et les microsatellites (Huvet et al. 2001) ne révèlent qu'une faible différenciation génétique.

Notre approche multi-locus portant sur les généalogies d'allèles permet donc d'aborder cette problématique selon un angle nouveau, celui des analyses de séquences et des patrons de coalescence.

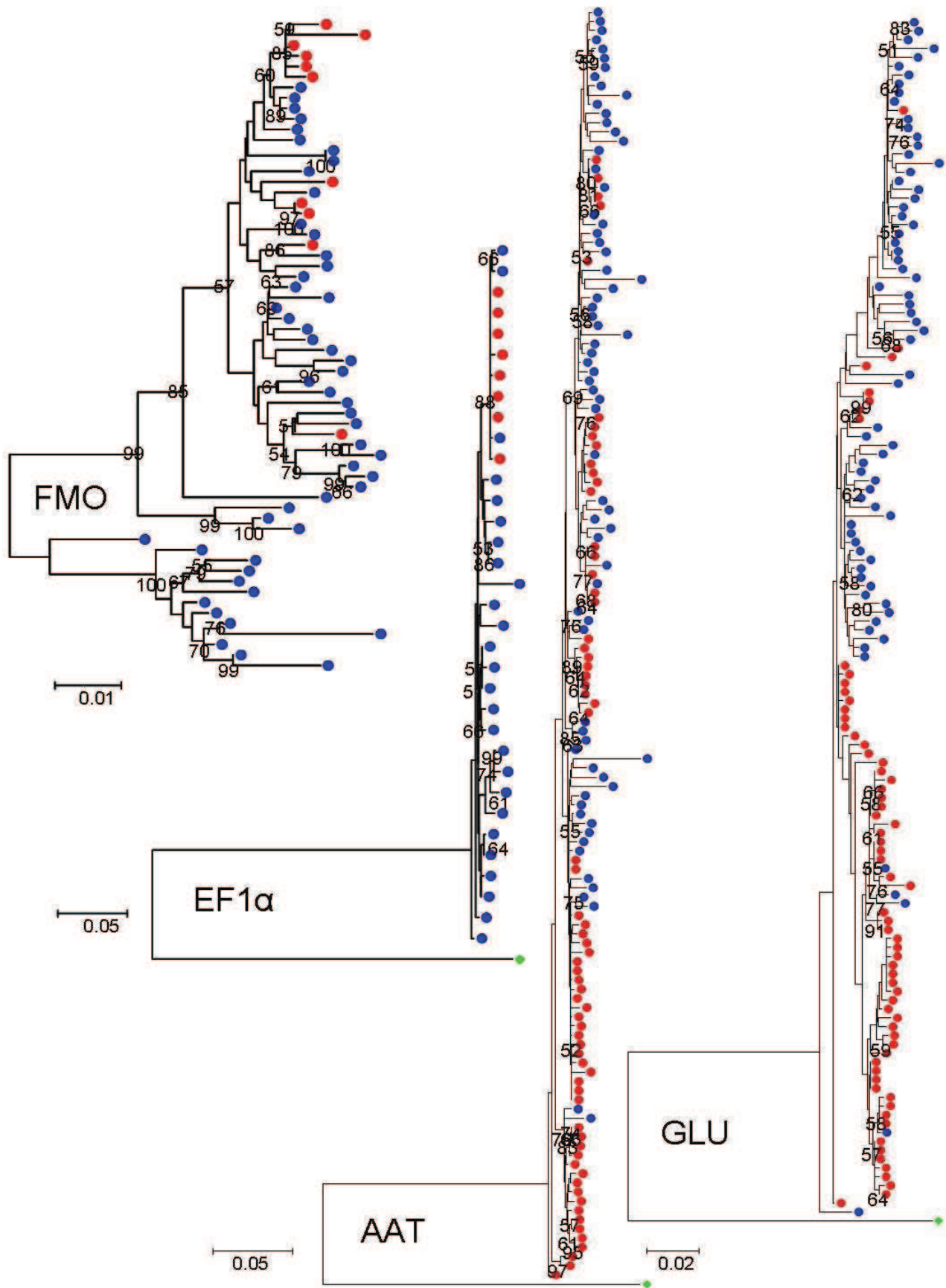
Les résultats concernant le genre *Crassostrea* seront analysés de manière indépendante hors du cadre de cette thèse. Cependant les premières analyses nous ont permis d'obtenir des estimations pour un certain nombre de descripteurs synthétiques (taille d'introns, diversité génétique, écart à l'attendue neutre et recombinaison) et sont présentées ici afin d'être

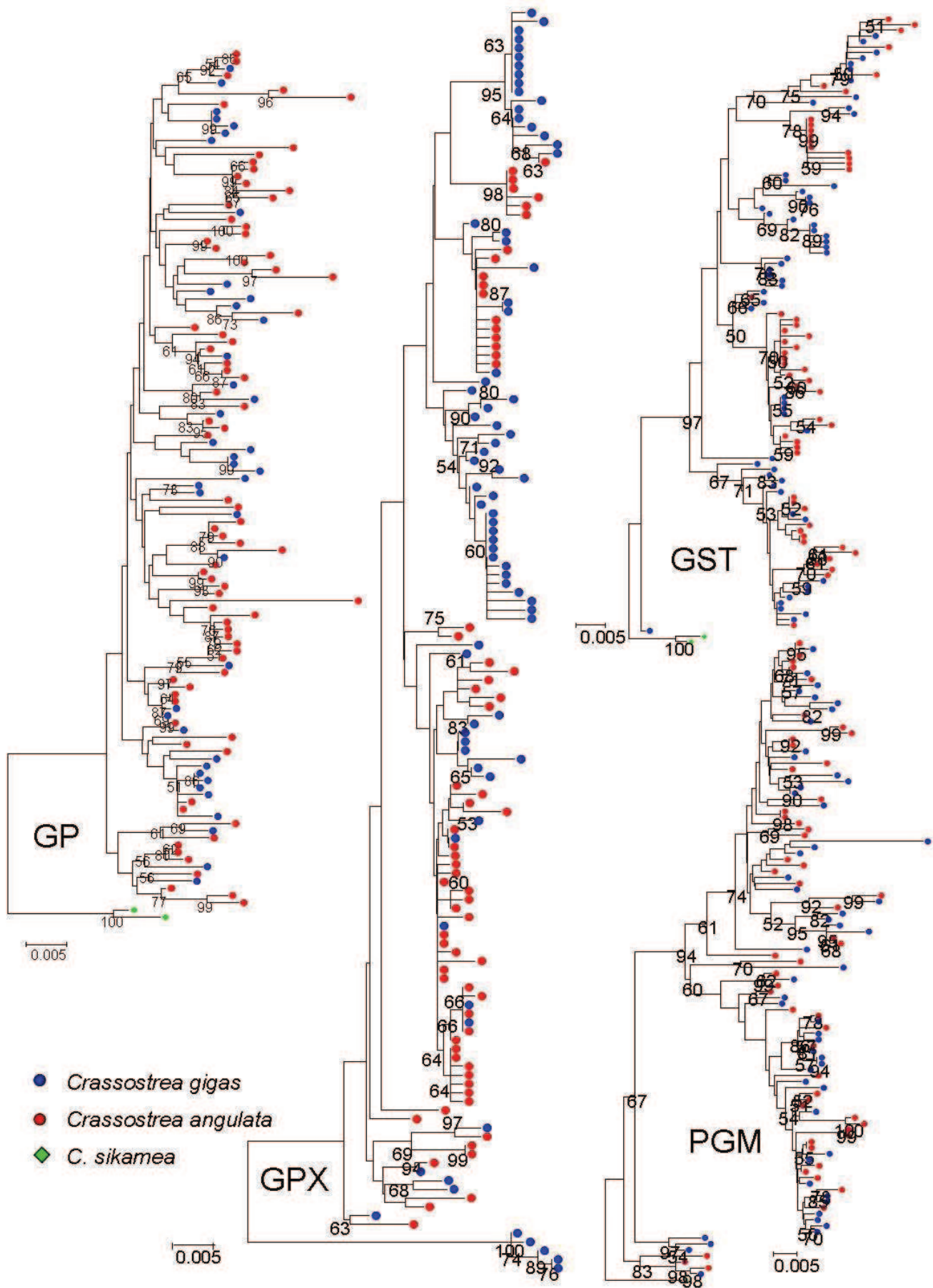
comparées à celles obtenues pour les espèces de *Bathymodiolus*. Les 11 locus pour lesquels les résultats sont présentés sont : l'Aspartate-Amino-Transférase (AAT), le Facteur d'élongation 1 α (EF1 α), la Flavine Monooxygénase (FMO), la Glutamine Synthétase (GLU), la Glycogène Phosphorylase (GP), la Glutathione Peroxydase (GPX), la Glutathione S-transférase (GST), le cytochrome P450 (P450), la PhosphoEnolPyruvate-Carboxy-Kinase 2 (PEPCK) et la PhosphoGlucoMutase (PGM). A ces dix premiers locus nucléaires a été ajouté le locus mitochondrial du Cytochrome-c Oxydase 1 (COI). Seuls deux gènes sont communs avec ceux étudiés chez *Bathymodiolus* : COI et EF1 α .

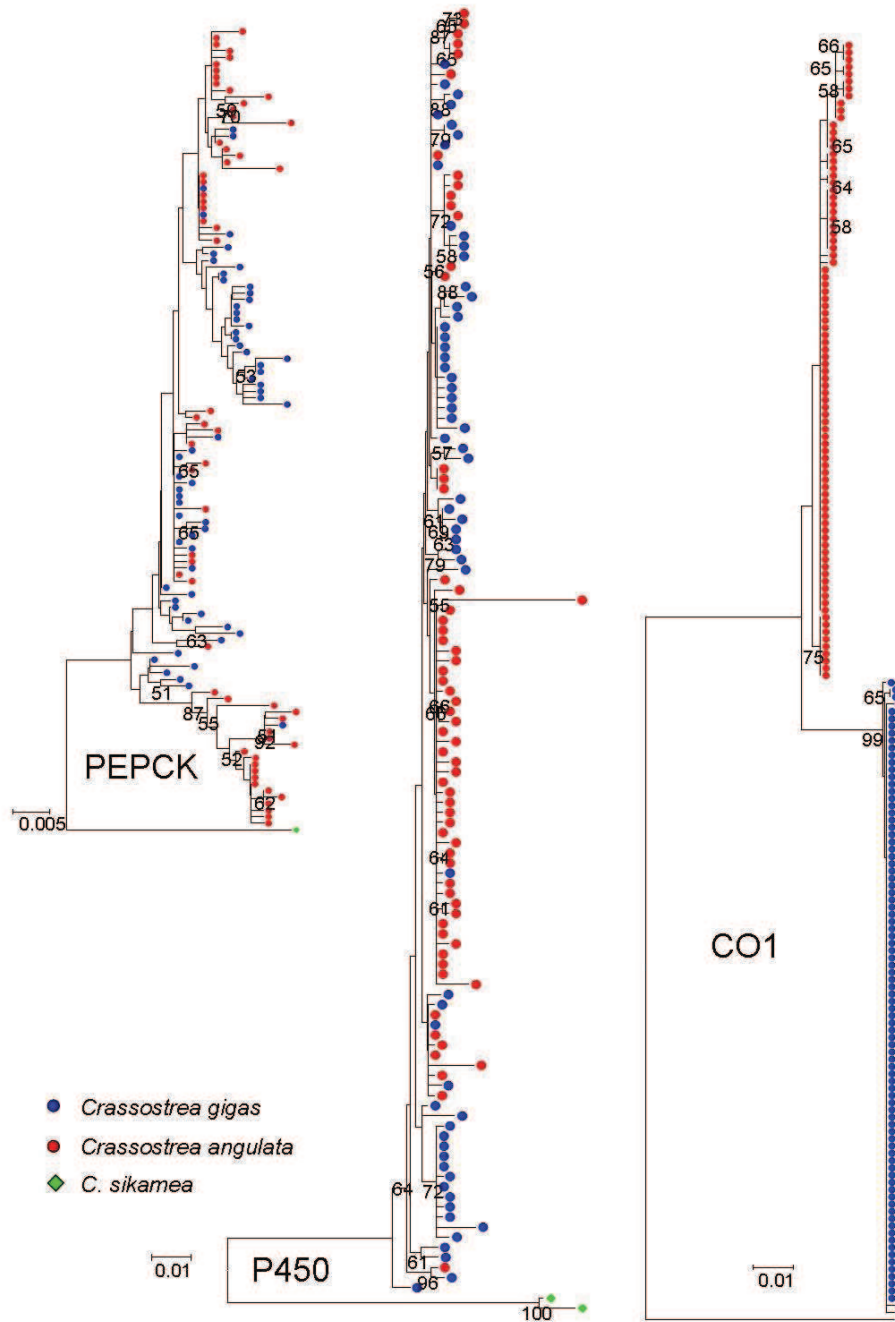
L'espèce *Crassostrea sikamea* (appelée huître de Kumamoto) a été choisie comme groupe externe, et des séquences ont pu être obtenues pour huit locus (AAT, EF1, GLU, GP, GST, P450, PEPCK, PGM et CO1). Les séquences du CO1 sont issues du travail de thèse de Frederico Batista (Batista 2007). Les séquences du groupe externe ont été obtenues à Roscoff, à l'exception de la séquence de CO1 de *C. sikamea* qui a été récupérée dans les banques de gènes (NCBI, numéro d'accèsion : EU007480).

Des arbres phylogénétiques ont été réalisés à partir d'un jeu de séquences de Haute Qualité analogue à ceux obtenus chez les bathymodioles. Ces arbres, réalisés en NJ sont présentés dans la Figure VI.1.

Figure VI.1. Arbres réalisés en NJ, avec des distances de Tamura-Nei et 1000 Bootstraps







Les arbres d'au moins trois locus (EF1 α , GLU, GPX) présentent une topologie fortement équilibrée entre les espèces *Crassostrea gigas* et *C. angulata*. Ces observations confirmeraient donc la structure observée pour le locus mitochondrial CO1 qui présente deux espèces distinctes. On peut aussi remarquer que les arbres des différents locus peuvent être caractérisés selon deux grands types de topologie : CO1, EF1 α , AAT, GLU et P450 sont en râteau alors que les arbres des locus FMO, GP, GPX, GST, PGM et PEPCK présentent des branches (et des feuilles) plus profondes. On peut également suspecter la présence de lignées paralogues pour le locus PEPCK.

Dans le cadre de la thèse, ces arbres phylogénétiques sont uniquement présentés pour illustrer la diversité des locus sur lesquels nous avons réalisé nos analyses. Les résultats de ces premières analyses sont réunis dans le Tableau VI.1. Le jeu de données utilisé pour réaliser ces analyses est l'équivalent du jeu de donnée Haute Qualité utilisé pour les Bathymodiales.

Tableau VI.1. Résultats des principaux descripteurs synthétiques obtenus pour les 11 gènes du genre *Crassostrea*. **n** : nombre d'allèles ; **Hd** : diversité haplotypique ; **π** : diversité nucléotidique calculée selon la méthode de Nei ; **D** : estimateur de du D de Tajima ; **Rm** : nombre minimum de recombinant détectés par la méthode de Hudson & Kaplan (1985)

Gène	Espèce	n	Hd	π	D (Tajima)	Rm	Rm / n
CO1	<i>C. gigas</i>	88	0,132	0,0003	-1,8775 (<0,05)	0	0.000
	<i>C. angulata</i>	88	0,681	0,0027	-1,1117 (NS)	0	0.000
	<i>C. sikamea</i>	1					
AAT	<i>C. gigas</i>	71	1	0.0257	-2.3189 (<0.01)	19	0.268
	<i>C. angulata</i>	67	0.994	0.0196	-1.3437 (ns)	11	0.164
	<i>C. sikamea</i>	1					
EF1	<i>C. gigas</i>	26	0,997	0,0162	-1,8603 (<0,05)	2	0.077
	<i>C. angulata</i>	8	0,643	0,0163	-1,5347 (NS)	0	0.000
	<i>C. sikamea</i>	1					
FMO	<i>C. gigas</i>	51	0,999	0,0420	-1.0811 (ns)	25	0.490
	<i>C. angulata</i>	11	1	0,0172	-1.2295 (ns)	6	0.545
GLU	<i>C. gigas</i>	70	0.999	0.0235	-1.8586 (<0.05)	29	0.414
	<i>C. angulata</i>	66	0.982	0.0133	-1.7051 (ns)	4	0.061
	<i>C. sikamea</i>	1					
GP	<i>C. gigas</i>	78	0.973	0.0198	-1.988 (<0.05)	19	0.244
	<i>C. angulata</i>	71	0.999	0.0239	-1.658 (ns)	23	0.324
	<i>C. sikamea</i>	1					
GPX	<i>C. gigas</i>	77	0.981	0.0206	-1.6473 (ns)	9	0.117
	<i>C. angulata</i>	67	0.987	0.0142	-1.8107 (<0.05)	8	0.119
GSTpi	<i>C. gigas</i>	67	0.992	0.0243	-0.8298 (ns)	11	0.164
	<i>C. angulata</i>	53	0.980	0.0274	-1.0309 (ns)	5	0.094
	<i>C. sikamea</i>	1					
P450	<i>C. gigas</i>	62	0.992	0.0109	-2.2545 (<0.01)	6	0.097
	<i>C. angulata</i>	65	0.951	0.0083	-2.517 (<0.001)	4	0.062
	<i>C. sikamea</i>	1					
PEPCK	<i>C. gigas</i>	61	0.972	0.0084	-2.1302 (<0.05)	6	0.098
	<i>C. angulata</i>	61	0.978	0.0149	-1.3217 (ns)	7	0.115
	<i>C. sikamea</i>	1					
PGM	<i>C. gigas</i>	53	0.998	0.0353	-1.4632 (ns)	15	0.283
	<i>C. angulata</i>	53	0.996	0.0353	-1.2879 (ns)	17	0.321
	<i>C. sikamea</i>	1					

Les valeurs de diversité haplotypique observées chez *Crassostrea* sont très élevées, près de 99% des haplotypes sont différents. La diversité nucléotidique moyenne entre les locus (0,0144) est le double de celle observée pour les espèces de *Bathymodiolus* (0,0056) Tableau VI.2.

Tableau VI.2. Comparaison des valeurs de diversité nucléotidique et de recombinaison moyennes calculées pour l'ensemble des locus et des populations. π : diversité nucléotidique calculée selon la méthode de Nei ; **Rm** : nombre de recombinants minimum, détectés par la méthode de Hudson & Kaplan (1985) et rapporté au nombre de séquences.

	genre	
	Crassostrea	Bathymodiolus
π moyen	0,0144	0,0056
Rm / nb séq.	0,184	0,144

Les valeurs de diversité nucléotidique moyennes de *Crassostrea* restent plus élevées que celle de *Bathymodiolus* même lorsque l'on ne prend pas en compte les locus pour lesquels nous pourrions avoir des suspicions de paralogie (PGM, PEPCK). En revanche, la diversité mitochondriale est très faible et les valeurs sont quasi-identiques à celles obtenues chez *Bathymodiolus* (entre 0,0003 et 0,0027 chez les *Crassostrea* et entre 0,0006 et 0,0072 pour les *Bathymodiolus*). Les valeurs du D de Tajima sont également très négatives et à la limite du seuil de significativité pour la majorité des locus, comme cela a déjà été constaté chez *Bathymodiolus*.

L'un des points qui semble caractéristique du génome de *Crassostrea* et qui n'est pas retrouvé chez *Bathymodiolus* est le nombre très élevé de recombinants (respectivement 0,184 ; et 0,144 recombinaisons moyenne par allèle sur l'ensemble des locus).

VI.3.2. Comparaison de deux gènes communs

Deux locus pour lesquels les données de séquence ont été acquises sont communs aux deux genres de bivalves : le locus mitochondrial CO1 et le locus nucléaire EF1 α . Ces locus ont été utilisés pour réaliser des comparaisons plus fines entre les genres de bivalves.

VI.3.2.a. Le CO1 :

Quel que soit le genre ou l'espèce étudiés, ces locus présentent toujours une très faible diversité nucléotidique. Nous avons utilisé ce gène codant pour comparer les valeurs de polymorphisme synonymes et non-synonymes entre les genres. Les résultats sont indiqués

dans le Tableau VI.3. La population de *B. thermophilus* de 21°S a été différenciée car les analyses réalisées dans le chapitre III ont montré qu'elle était génétiquement différente de celles du nord (9°50N et 7°S)

Tableau VI.3. Synthèse des diversités Synonymes et Non synonymes calculées sur le gène du CO1. *B. thermo*: *Bathymodiolus. thermophilus*; *B. azo*: *B. azoricus*; *B. puteo*: *B. puteoserpentis*, *C. angul*: *Crassostrea. Angulata*; π_S : diversité nucléotidique synonyme; π_{NS} : diversité nucléotidique non-synonyme; **nb S**: nombre de mutations synonymes; **nb NS**: nombre de mutations non synonymes

population	nombre de séquences	nombre de sites S	nombre de sites NS	π_S	nb S	π_{NS}	nb NS
<i>B. thermo</i> 9°50N + 7S	93	139,08	385,92	0,02586	21	0,00071	6
<i>B. thermo</i> 21°S	27	139,04	385,96	0,01006	10	0	0
Total <i>B. thermo</i>	120	139,07	385,93	0,02474	25	0,00055	6
<i>B. azo</i>	96	131,34	393,66	0,01809	33	0,00035	1
<i>B. puteo</i>	33	131,69	393,31	0,00179	2	0,00015	1
<i>C. angul</i>	88	134,26	397,74	0,00870	10	0,00063	2
<i>C. gigas</i>	88	136,69	385,31	0,00050	3	0,00018	2

Les niveaux de diversité non-synonyme sont similaires entre les espèces : 0,00015 à 0,00071 pour les *Bathymodiolus* (à l'exception de *B. thermo* de 21°S qui ne présente aucun polymorphisme non-synonyme) et de 0,00018 à 0,00063 pour les *Crassostrea*. En revanche, (et à l'exception de *B. puteoserpentis*), les niveaux de polymorphisme synonyme sont d'un ordre de grandeur plus élevé chez les espèces de *Bathymodiolus* que chez les *Crassostrea*. Cet excès de mutations synonymes pour le locus mitochondrial chez les espèces de *Bathymodiolus* pourrait être lié à l'influence du milieu hydrothermal et de ses éléments mutagènes. En effet sous cette hypothèse le taux de mutation devrait être plus élevé mais seules les mutations synonymes ne vont pas être éliminées par la sélection purifiante qui agit sur les régions codantes. Il est cependant nécessaire de réaliser cette comparaison avec d'autres gènes et avec d'autres espèces afin de déterminer si ce résultat est lié au milieu hydrothermal ou uniquement aux *Bathymodiolus*.

VI.3.2.b. Le Facteur d'Elongation (EF1 α) :

Ce locus a été séquencé pour trois genres de bivalves : *Bathymodiolus*, *Mytilus* et *Crassostrea*.

Une séquence composée de trois exons et de deux introns a ainsi pu être comparée entre les genres. La structure de ce gène est indiquée dans le Tableau VI.4.

Les tailles indiquées correspondent à l'observation d'une seule séquence pour *Bathymodiolus* et à une moyenne estimée à partir de l'ensemble des populations disponibles pour les *Mytilus* et les *Crassostrea*.

Tableau VI.4. Comparaison de la portion commune du gène du Facteur d'Elongation 1 α entre trois genres de bivalves. Les chiffres correspondent à la taille en nombre de paires de bases.

genre	exon 1	intron 1	exon 2	intron 2	exon 3	Taille locus totale
<i>Crassostrea</i>	140	80	181	109	85	595
<i>Mytilus</i>	140	556	181	235	85	1197
<i>Bathymodiolus</i>	140	624	181	493	85	1523

A cause de sa grande taille chez *Bathymodiolus* (1523 bp), ce locus n'a pas été amplifié sur la totalité de sa longueur pour les populations utilisées lors de cette thèse. En effet des amorces internes, ont été définies afin de permettre le séquençage d'une portion plus courte, mais en série. Il n'est donc pas possible ici de comparer les indices de diversité à partir d'un si faible échantillonnage.

La structure du gène est conservée entre les trois genres et la taille des exons est également fixée. Nous avons observée les mutations fixées entre les genres sur les trois exons. Les résultats sont indiqués dans le Tableau VI.5.

Tableau VI.5. Comparaison des mutations fixées sur les exons du gène EF1 α entre trois genres de bivalves.

	<i>Crassostrea</i> (n= 35)		<i>Mytilus</i> (n= 84)	
	S	NS	S	NS
<i>Bathymodiolus</i> (n=1)	51	15	30	9
<i>Mytilus</i> (n= 84)	43	16		

Il faut considérer le résultat des mutations fixées avec précautions pour les *Bathymodiolus* car nous ne disposons que d'une seule séquence. Cependant le risque d'assimiler du polymorphisme à de la divergence fixée est assez faible car les exons de EF1 présentent une faible diversité nucléotidique.

Nos résultats confirment la plus forte proximité phylogénétique entre les *Bathymodiolus* et les *Mytilus* (9 mutations NS fixées contre 15 et 16 avec *Crassostrea*) et nous indiquent que l'évolution du gène de l'EF1 est similaire entre les trois genres. Il serait intéressant de mener ce type d'analyse comparative sur un plus grand nombre de gènes entre

les trois genres de bivalves afin de détecter si certains gènes présentent des évolutions différentes en fonction des genres.

Contrairement aux exons, la taille des introns est très variable entre les genres (Tableau VI.4). Les introns de *Crassostrea* sont nettement plus petits (165 bp au total) que ceux trouvés pour les deux genres de moules (791 et 1117 au total pour les genres *Mytilus* et *Bathymodiolus* respectivement). Il est difficile de savoir à partir d'un seul gène s'il s'agit d'une tendance générale caractéristique des différents genres. Il faudrait vérifier sur d'autres gènes si la longueur particulièrement important des introns de *Bathymodiolus* est systématique ou non. On pourra alors essayer de déterminer les principaux facteurs permettant la fixation d'intron de si grande taille (par exemple la taille efficace des populations, le taux moyen de recombinaison, ou le taux au taux de mutation...).

Comme nous l'avons dit précédemment, nous ne disposons pas des séquences de *Bathymodiolus* pour la pleine longueur de ce locus. Nous nous sommes donc focalisés sur la partie commune (1^{er} exon et 1^{er} intron) pour réaliser une comparaison des diversités introniques avec *Crassostrea*. L'intron de *Crassostrea* est étudié dans sa totalité (80bp), alors que celui de *Bathymodiolus* est coupé aux alentours de 400 bp car les amorces internes utilisées pour l'amplification ont été définies sur cet intron. Les résultats sont indiqués dans le Tableau VI.6

Tableau VI.6. Comparaison de la diversité du premier intron du locus EF1 α entre les genres *Bathymodiolus* et *Crassostrea*. *N* : nombre de séquences ; *S* : nombre de sites synonymes moyen ; π_s : diversité nucléotidique silencieuse. ; **Nb S** : nombre de mutations synonymes

population	<i>N</i>	<i>S</i>	π_s	nb S	D de Tajima
<i>B. thermo_7S</i>	25	400	0,00020	1	-1,15753 (ns)
<i>B. thermo_21S</i>	16	390	0,00496	11	-1,59368 (*)
<i>B. azo</i>	32	406	0,00108	7	-2,17748 (**)
<i>B. puteo</i>	15	406	0,00282	6	-1,3326 (ns)
<i>C. angul</i>	7	80	0,00360	1	-1,00623 (ns)
<i>C. gigas</i>	26	80	0,03037	16	-1,56798 (ns)

Les analyses ont été réalisées sur les populations du Pacifique (*B. thermophilus*) et de l'Atlantique (*B. azoricus* et *B. puteoserpentis*) ainsi que sur les deux espèces de *Crassostrea* (*C. angulata* et *C. gigas*). Les diversités nucléotidiques introniques observées chez *Crassostrea* sont plus fortes que celles (très faibles) observées chez les *Bathymodiolus*. Même si la

diversité de *C. angul* (0,0036) est du même ordre de grandeur que les valeurs de diversité les plus élevées observées chez *Bathymodiolus* (0,0049 et 0,0028); il faut relativiser car seules 7 séquences ont été analysées. Les valeurs du D de Tajima sont très fortement négatives pour chacune des populations et des genres observés. Il serait nécessaire de réaliser des analyses plus détaillées afin de savoir si les mutations qui apparaissent sur l'intron de *Crassostrea* sont légèrement délétères (comme chez *Bathymodiolus*). Il est cependant important de noter que chez *Crassostrea*, à l'instar de *Bathymodiolus*, l'arbre de coalescence de EF1 α présente un net râteau, sans profondeur de branche et que la majorité des mutations sont les singletons (Figure VI.1).

VI.3.3. Discussion générale sur la comparaison des genres et perspectives

VI.3.3.a. Indices de diversité

Comme nous l'avons précédemment évoqué, nous suspectons que les contraintes démographiques imposées par le milieu hydrothermal (extinction/recolonisation) puissent conduire à une faible diversité génétique des populations de *Bathymodiolus* (effet de goulot d'étranglement). Pourtant la diversité nucléotidique moyenne observée chez les *Bathymodiolus* n'est que 2 ou 2,5 fois inférieures à celle observée chez les espèces du genre *Crassostrea*.

Les bivalves marins sont connus pour avoir une forte diversité (Solé-Cava & Thorpe 1991, Daguin et al. 2001, Bierne et al. 2003). La similarité entre les niveaux de diversité entre *Crassostrea* et *Bathymodiolus* confirme donc que cette forte diversité est bien une caractéristique générale de bivalves marins quelles que soient les contraintes du milieu dans lequel ils vivent.

Un locus présente cependant une très faible diversité quel que soit le genre ou l'espèce étudiée, il s'agit du CO1. Cette faible diversité est étonnante compte tenue des fortes tailles efficaces des populations et surtout de la forte diversité présente aux autres locus. Ce résultat pourrait être lié à des phénomènes sélectifs tels que le balayage, évoqué par exemple par Bazin et al. (2006)) pour expliquer la faible diversité en général...

VI.3.3.b. Nombre de recombinants

La plus grande différence observée lors des comparaisons multi locus entre *Crassostrea* et *Bathymodiolus* concerne le nombre des recombinants intragéniques détectés. Nous avons en effet observé que le nombre de recombinants était nettement plus élevé chez les huîtres. Comme nous n'avons réalisé cette comparaison qu'entre deux genres, il n'est pas possible de déterminer si *Crassostrea* présente un excès de recombinant ou si au contraire *Bathymodiolus* présente des valeurs inférieures à ce qui est généralement rencontré chez les bivalves. Les estimations du nombre minimum de recombinaisons réalisées dans la thèse de M. Faure (Faure 2007) semblent montrer un très faible nombre de recombinant chez *Mytilus*. Le genre *Crassostrea* présenterait donc un excès de recombinants. Il est connu que la recombinaison permet le maintien et la création de diversité dans les génomes (Feldman et al. 1997), qu'elle facilite l'élimination des mutations délétères dans les populations de taille finie (Felsenstein 1974). La présence de recombinaison, chez les espèces asexuée comme chez les espèces sexuées permet également d'accélérer la fixation des mutations avantageuses (Muller 1964) qui seraient apparues sur des haplotypes différents.

Le rapport entre les taux de recombinaison et les mutations délétères est un sujet qui mériterait une étude à part entière chez les bivalves.

VI.3.3.c. Mutation délétères

Le nombre de mutations synonymes observées sur le CO1 semble plus élevé chez les *Bathymodiolus* que chez les autres bivalves. La comparaison avec l'EF1 n'est pas très informative car il semble qu'à ce locus les mutations introniques sont légèrement délétères. Il est très difficile de déterminer si le milieu hydrothermal agit réellement sur les génomes en générant un plus fort taux mutationnel. En effet, il ne suffit pas de comparer le niveau de mutations délétères pour un grand nombre de gènes communs, entre des espèces côtières et une espèce hydrothermale, pour réussir à conclure sur l'influence du milieu hydrothermal. En effet, de nombreux paramètres autres que les facteurs environnementaux (taille efficace, recombinaison, fécondité...) peuvent favoriser les excès de mutations délétères chez certaines espèces. Une méta-analyse prenant en compte de nombreux gènes communs, des répliquats avec des espèces-sœurs et une estimation précise des paramètres de taille efficace, de taux de recombinaison et les paramètres du milieu, serait nécessaire afin de pouvoir enfin mettre en évidence l'influence mutagène du milieu hydrothermal.

VI.3.3.d. Taille des introns

Nous avons mis en évidence que la taille des introns était très variable pour un même locus en fonction des genres étudiés.

Nos observations ne reposent que sur un gène, celui de l'EF1. Une légère variation dans la taille d'un intron avec la présence/ absences de certains indels s'explique facilement par la dérive (Lynch 2002). En revanche dans notre cas, les observations portent sur les deux introns de l'EF1 et les différences de taille sont très importantes et ne sont pas liée à un unique événement évolutif (par exemple un indel).

L'existence d'introns fixés avec des tailles différentes entre les 3 genres de bivalves indique bien que les pressions évolutives en équilibre ne sont pas les mêmes et que cela n'est pas le simple effet du hasard. Les conditions requises pour une évolution neutre de la taille des introns résultent d'un équilibre complexe prenant en compte, par exemple, les taux de recombinaison ou la structure secondaire des gènes... Il a été mis en évidence chez les *Drosophiles* que dans les régions à forte recombinaison, les pressions évolutives étaient plus efficaces et que les tailles d'introns étaient réduites (Hill & Robertson 1966), (Kliman & Hey 1993). Il est intéressant de noter que chez *Crassostrea* aussi la plus forte recombinaison générale qui est suspectée est couplée avec des tailles d'introns inférieures à celles autres bivalves.

Au cours de l'étude approfondie de l'évolution des *Bathymodiolus* nous avons été amenés à proposer un certain nombre d'hypothèses concernant l'évolution des gènes et des bivalves. Ces hypothèses concernent aussi bien l'influence du milieu hydrothermal (taux de mutation, mécanismes de spéciation), que les éléments moteurs de l'évolution des gènes, des génomes, des populations et des espèces (duplication, sélection, recombinaison, introgression entre espèces proches, différenciation primaire, contact secondaire). Certaines de ces hypothèses pourront être validées à partir de l'analyse approfondie des données obtenues sur le genre *Crassostrea* et par la comparaison avec ce qui a déjà été fait sur le genre *Mytilus*

Références bibliographiques

- A -

- Akashi H (1995) Inferring Weak Selection From Patterns of Polymorphism and Divergence at "Silent" Sites in *Drosophila* DNA. *Genetics Society of America* 139:1067-1076
- Alayse-Danet AM, Desbruyères D, Gaill F (1987) The possible nutritional or detoxification role of the epibiotic bacteria of alvinellid polychaetes: review of current data. *Symbiosis* 4:51-62
- Alayse-Danet AM, Gaill F, Desbruyères D (1985) Preliminary studies on the relationship between the pompeii worm, *Alvinella pompejana* (Polychaeta: Ampharetidae), and its epibiotic bacteria. In: Gibbs PE (ed) *Proceedings of the 19th European Marine Biology Symposium*. Cambridge University Press, Cambridge, p 162-172
- Andolfatto P (2005) Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* 437:1149-1152
- Arnold ML (1997) *Natural Hybridization and Evolution*, Vol. Oxford University Press, New York
- Arp AJ, Childress JJ (1981) Blood function in the hydrothermal vent vestimentiferan tube worm. *Science* 213:342-344
- Arp AJ, Childress JJ, Fisher CR (1984) Metabolic and Blood-Gas Transport Characteristics of the Hydrothermal Vent Bivalve *Calyptogena Magnifica*. *Physiol Zool* 57:648-662

- B -

- Bachali S, Jager M, Hassanin A, Schoentgen F, Jolles P, Fiala-Medioni A, Deutsch JS (2002) Phylogenetic analysis of invertebrate lysozymes and the evolution of lysozyme function. *J Mol Evol* 54:652-664
- Bachraty C, Legendre P, Desbruyères D (2008) Biogeographic relationships among deep-sea hydrothermal vent faunas at global scale. *Deep-Sea Research*
- Bailly X, Leroy R, Carney S, Collin O, Zal F, Toulmond A, Jollivet D (2003) The loss of the hemoglobin H₂S-binding function in annelids from sulfide-free habitats reveals molecular adaptation driven by Darwinian positive selection. *Proc Natl Acad Sci U S A* 100:5885-5890
- Baker ET (1995) Characteristics of hydrothermal discharge following a magmatic intrusion. In: Parson LM, Walker CL, Dixon DR (eds) *Hydrothermal vents and processes*. Geological Society, London, p 65-76
- Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290:972-977
- Ballard JWO, Rand DM (2005) The population biology of mitochondrial DNA and its phylogenetic implications. *Annual Review in Ecology and Systematics* 36:621-642
- Banks DA (1985) A fossil hydrothermal worm assemblage from the Tynagh lead-zinc deposit in Ireland. *Nature* 313:128-131
- Barluenga M, Stolting KN, Salzburger W, Muschick M, Meyer A (2006) Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature* 439:719-723
- Barton NH (1986) The effects of linkage and density-dependent regulation on gene flow. *Heredity* 57:415-426
- Barton NH, Hewitt GM (1985) Analysis of Hybrid Zones. *Annu Rev Ecol Syst* 16:113-148
- Bates AE, Tunnicliffe V, Lee RW (2005) Role of thermal conditions in habitat selection by hydrothermal vent gastropods. *Marine Ecology-Progress Series* 305:1-15

- Batista FMM (2007) Assessment of the aquacultural potential of the Portuguese oyster *Crassostrea angulata*. Universidade do Porto
- Bazin E, Glémin S, Galtier N (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science* 312:570-572
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX v. 4.05, logiciel sous WindowsTM pour la génétique des populations. Laboratoire Génome, Populations, Interactions: CNRS UMR 5000, Université de Montpellier II, Montpellier, France.
- Berger MS, Young CM (2006) Physiological response of the cold-seep mussel *Bathymodiulus childressi* to acutely elevated temperature. *Mar Biol* 149:1397-1402
- Bergman CM, Kreitman M (2001) Analysis of conserved noncoding DNA in *Drosophila* reveals similar constraints in intergenic and intronic sequences. *Genome Res* 11:1335-1345
- Bettencourt BR, Feder ME (2001) Hsp70 duplication in the *Drosophila melanogaster* species group: How and when did two become five? *Mol Biol Evol* 18:1272-1282
- Bierne N, Borsa P, Daguin C, Jollivet D, Viard F, Bonhomme F, David P (2003) Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. *Mol Ecol* 12:447-462
- Bierne N, David P, Boudry P, Bonhomme F (2002) Assortative fertilization and selection at larval stage in the mussels *Mytilus edulis* and *M. galloprovincialis*. *Evolution* 56:292-298
- Bierne N, Eyre-Walker A (2006) Variation in synonymous codon use and DNA polymorphism within the *Drosophila* genome. *J Evol Biol* 19:1-11
- Bierne N, Lehnert A, Bedier E, Bonhomme F, Moore SS (2000) Screening for intron-length polymorphisms in penaeid shrimps using exon-primed intron-crossing (EPIC)-PCR. *Mol Ecol* 9:233-235
- Bierne N, Tanguy A, Faure M, Faure B, David E, Boutet I, Boon E, Quere N, Plouviez S, Kempainen P, Jollivet D, Moraga D, Boudry P, David P (2007) Mark-recapture cloning: a straightforward and cost-effective cloning method for population genetics of single-copy nuclear DNA sequences in diploids. *Molecular Ecology Notes* 7:562-566
- Black MB, Lutz RA, Vrijenhoek RC (1994) Gene flow among vestimentiferan tube worm (*Riftia pachyptila*) populations from hydrothermal vents of the eastern Pacific. *Mar Biol* 120:33-39
- Black MB, Trivedi A, Maas PAY, Lutz RA, Vrijenhoek RC (1998) Population genetics and biogeography of vestimentiferan tube worms. *Deep-Sea Research II* 45:365-382
- Blot M, Soyer J, Thiriot-Quievreux C (1987) Preliminary data on the genetic differentiation of *Mytilus desolationis* Lamy 1936 and *Aulacomya ater regia* Powell 1957 (Bivalvia, Mytilidae) in the Kerguelen Islands (Terres Australes et Antarctiques Françaises). *Polar Biol* 7:1-7
- Bolnick DI (2004) Waiting for sympatric speciation. *Evolution* 58:895-899
- Bolnick DI, Fitzpatrick BM (2007) Sympatric speciation: Models and empirical evidence. *Annu Rev Ecol Evol Syst* 38:459-487
- Boudry P, Collet B, Cornette F, Hervouet V, Bonhomme F (2002) High variance in reproductive success of the Pacific oyster (*Crassostrea gigas*, Thunberg) revealed by microsatellite-based parentage analysis of multifactorial crosses. *Aquaculture* 204:283-296
- Boudry P, Heurtebise S, Collet B., Cornette F., Gérard A. (1998) Differentiation between populations of the Portuguese oyster, *Crassostrea angulata* (Lamarck) and the Pacific

- oyster, *Crassostrea gigas* (Thunberg), revealed by *mtDNA* RFLP analysis. *J Exp Mar Biol Ecol* 226:279-291
- Bougault H, Charlou JL, Fouquet Y, Needham HD, Vaslet N, Appriou P, Baptiste PJ, Rona PA, Dmitriev L, Silantiev S (1993) Fast and Slow-Spreading Ridges: Structure and Hydrothermal Activity, Ultramafic Topographic Highs, and CH_4 Output. *J Geophys Res-Solid Earth* 98:9643-9651
- Boulegue J, Benedetti EL, Dron D, Mariotti A, Letolle R (1987) Geochemical and Biogeochemical Observations on the Biological Communities Associated with Fluid Venting in Nankai Trough and Japan Trench Subduction Zones. *Earth Planet Sci Lett* 83:343-355
- Boursot P, Auffray JC, Brittondavidian J, Bonhomme F (1993) The evolution of house Mice. *Annu Rev Ecol Syst* 24:119-152
- Boutet I, Tanguy A, Rousseau S, Auffret M, Moraga D (2003) Molecular identification and expression of heat shock cognate 70 (*hsc70*) and heat shock protein 70 (*hsp70*) genes in the Pacific oyster *Crassostrea gigas*. *Cell Stress Chaperones* 8:76-85
- Bryant D, Moulton V (2004) Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* 21:255-265
- Bucklin A (1988) Allozymic variability of *Riftia pachyptila* populations from the Galapagos Rift and 21°N hydrothermal vents. *Deep-Sea Research* 35:1759-1768
- Buggs RJA (2007) Empirical study of hybrid zone movement. *Heredity* 99:301-312
- Buroker NE, Hershberger WK, Chew KK (1979) Populations genetics of the family Ostreidae. I. Intraspecific studies of *Crassostrea gigas* and *Saccostrea commercialis*. *Mar Biol* 54:157-169
- Butman CA, Carlton JT, Palumbi SR (1995) Whaling Effects on Deep-Sea Biodiversity. *Conserv Biol* 9:462-464

- C -

- Campbell AC, Palmer MR, Klinkhammer GP, Bowers TS, Edmond JM, Lawrence JR, Casey JF, Thompson G, Humphris S, Rona P, Karson JA (1988) Chemistry of hot springs on the Mid-Atlantic Ridge. *Nature* 335:514-519
- Cannat M, Mevel C, Maia M, Deplus C, Durand C, Gente P, Agrinier P, Belarouchi A, Dubuisson G, Humler E, Reynolds J (1995) Thin Crust, Ultramafic Exposures, and Rugged Faulting Patterns at Mid-Atlantic Ridge (22-Degrees 24-Degrees-N). *Geology* 23:49-52
- Cannon GA, Pashinski DJ (1990) Circulation near Axial Seamount. *Journal of Geophysical Research*, B 95:12823-12828
- Cary SC (1994) Vertical transmission of a chemoautotrophic symbiont in the protobranch bivalve, *Solemya reidi*. *Molecular Marine Biology and Biotechnology* 3:121-130
- Cary SC, Shank T, Stein J (1998) Worms bask in extreme temperatures. *Nature* 391:545-546
- Cavanaugh CM (1983) Symbiotic Chemoautotrophic Bacteria in Marine-Invertebrates from Sulfide-Rich Habitats. *Nature* 302:58-61
- Cavanaugh CM, Gardiner SL, Jones ML, Jannasch HW, Waterbury JB (1981) Prokaryotic cells in the hydrothermal vent tube worm *Riftia pachyptila* Jones: possible chemoautotrophic symbionts. *Science* 213:340-342
- Cayré P, Richer de Forges B (2002) Faune mystérieuse des océans profonds. *La Recherche* 355:59-62
- Chabasse C, Bailly X, Sanchez S, Rousselot M, Zal F (2006) Gene structure and molecular phylogeny of the linker chains from the giant annelid hexagonal bilayer hemoglobins. *J Mol Evol* 63:365-374

- Charlesworth B, Morgan MT, Charlesworth D (1993) The Effect of Deleterious Mutations on Neutral Molecular Variation. *Genetics* 134:1289-1303
- Charlou JL, Donval JP, Fouquet Y, Jean-Baptiste P, Holm N (2002) Geochemistry of high H₂ and CH₄ vent fluids issuing from ultramafic rocks at the Rainbow hydrothermal field (36 degrees 14 ' N, MAR). *Chem Geol* 191:345-359
- Cherry R, Desbruyères D, Heyraud M, Nolan C (1992) High levels of natural radioactivity in hydrothermal vent polychaetes. *Comptes Rendus de l'Académie des Sciences de Paris, Série III* 315:21-26
- Chevaldonné P, Desbruyères D, Childress JJ (1992) Some like it hot... and some even hotter. *Nature* 359:593-594
- Chevaldonné P, Jollivet D (1993) Videoscopic study of deep-sea hydrothermal vent alvinellid polychaete populations: biomass estimation and behaviour. *Marine Ecology Progress Series* 95:251-262
- Chevaldonne P, Jollivet D, Desbruyeres D, Lutz RA, Vrijenhoek RC (2002) Sister-species of eastern Pacific hydrothermal vent worms (Ampharetidae, Alvinellidae, Vestimentifera) provide new mitochondrial COI clock calibration. *Cah Biol Mar* 43:367-370
- Chevaldonné P, Jollivet D, Vangriesheim A, Desbruyères D (1997) Hydrothermal-vent alvinellid polychaete dispersal in the eastern Pacific. 1. Influence of vent site distribution, bottom currents and biological patterns. *Limnology and Oceanography* 42:67-80
- Childress JJ, Fisher CR (1992) The biology of hydrothermal vent animals: physiology, biochemistry, and autotrophic symbioses. *Oceanography and Marine Biology Annual Review* 30:337-441
- Childress JJ, Fisher CR, Brooks JM, Kennicutt MC, II, Bidigare R, Anderson AE (1986) A methanotrophic marine molluscan (*Bivalvia*, *Mytilidae*) symbiosis: mussels fueled by gas. *Science* 233:1306-1308
- Comtet T (1998) Structure des populations, reproduction, croissance et phylogéographie des *Mytilidae* des champs hydrothermaux Lucky Strike et Menez Gwen (37°17'N et 37°50'N sur la dorsale médio-atlantique). Thèse de doctorat, Université de Bretagne Occidentale, Brest
- Comtet T, Desbruyeres D (1998) Population structure and recruitment in mytilid bivalves from the Lucky Strike and Menez Gwen hydrothermal vent fields (37 degrees 17 ' N and 37 degrees 50 ' N on the Mid-Atlantic Ridge). *Marine Ecology-Progress Series* 163:165-177
- Comtet T, Desbruyères D (1998) Population structure and recruitment in mytilid bivalves from the Lucky Strike and Menez Gwen hydrothermal vent fields (37°17' N and 37°50' N on the Mid-Atlantic Ridge). *Marine Ecology Progress Series* 163:165-177
- Corliss JB, Dymond J, Gordon LI, Edmond JM, von Herzen RP, Ballard RD, Green K, Williams D, Bainbridge A, Crane K, van Andel TH (1979) Submarine thermal springs on the Galápagos Rift. *Science* 203:1073-1083
- Corpet F (1988) Multiple Sequence Alignment with Hierarchical-Clustering. *Nucleic Acids Res* 16:10881-10890
- Cosson RP, Thiébaud E, Company R, Castrec-Rouelle M, Colaço A, Martins I, Sarradin PM, Bebianno MJ (2008) Spatial variation of metal bioaccumulation in the hydrothermal vent mussel *Bathymodiolus azoricus*. *Marine Environmental Research* 65:405-415
- Craddock C, Hoeh WR, Gustafson RG, Lutz RA, Hashimoto J, Vrijenhoek RJ (1995a) Evolutionary relationships among deep-sea *Mytilids* (*Bivalvia*, *Mytilidae*) from hydrothermal vents and cold-water methane sulfide seeps. *Mar Biol* 121:477-485

- Craddock C, Hoeh WR, Lutz RA, Vrijenhoek RC (1995b) Extensive gene flow among Mytilid (*Bathymodiolus thermophilus*) populations from hydrothermal vents of the Eastern Pacific. *Mar Biol* 124:137-146
- Cronn R, Cedroni M, Haselkorn T, Grover C, Wendel JF (2002) PCR-mediated recombination in amplification products derived from polyploid cotton. *Theor Appl Genet* 104:482-489

- D -

- Daguin C, Bonhomme F, Borsa P (2001) The zone of sympatry and hybridization of *Mytilus edulis* and *M. galloprovincialis*, as described by intron length polymorphism at locus mac-1. *Heredity* 86:342-354
- Dahlhoff EP, Rank NE (2000) Functional and physiological consequences of genetic variation at phosphoglucose isomerase: heat shock protein expression is related to enzyme genotype in a montane beetle. *Proceedings of the National Academy of Science of USA* 97:10056-10061
- Darwin C (1859) *On the Origin of Species by Means of Natural Selection*, Vol. John Murray, London
- Davaine J (1853) *Recherches sur la génération des huîtres*, Vol, Paris
- De la Haba G, Cantoni GL (1959) Enzymatic synthesis of S-adenosyl-L-homocysteine from adenosine and homocysteine. *J Biol Chem* 234:603-608
- de Queiroz K (2005) Ernst Mayr and the modern concept of species. *Proc Natl Acad Sci U S A* 102:6600-6607
- Dell RK (1987) Mollusca of the family Mytilidae (Bivalvia) associated with organic remains from deep water off New Zealand, with revisions of the genera *Adipicola* Dautzenberg, 1927 and *Idasola* Iredale, 1915. *National Museum of New Zealand Records* 3:17-36
- Demets C, Gordon RG, Argus DF, Stein S (1990) Current Plate Motions. *Geophys J Int* 101:425-478
- Deming JW, Reysenbach AL, Macko SA, Smith CR (1997) Evidence for the microbial basis of a chemoautotrophic invertebrate community at a whale fall on the deep seafloor: Bone-colonizing bacteria and invertebrate endosymbionts. *Microsc Res Tech* 37:162-170
- Denis F, Jollivet D, Moraga D (1993) Genetic separation of two allopatric populations of hydrothermal snails *Alviniconcha* spp. (Gastropoda) from two south western Pacific back-arc basins. *Biochemical Systematics and Ecology* 21:431-440
- Desbruyères D, Segonzac M, Bright M (2006) *Handbook of deep-sea hydrothermal vent fauna*, Vol
- Distel DL, Baco AR, Chuang E, Morrill W, Cavanaugh C, Smith CR (2000) Marine ecology - Do mussels take wooden steps to deep-sea vents? *Nature* 403:725-726
- Distel DL, Lee HKW, Cavanaugh CM (1995) Intracellular coexistence of methano- and thioautotrophic bacteria in a hydrothermal vent mussel. *Proceedings of the National Academy of Sciences of the USA* 92:9598-9602
- Dixon DR, Lowe DM, Miller PI, Villemin GR, Colaco A, Serrao-Santos R, Dixon LRJ (2006) Evidence of seasonal reproduction in the Atlantic vent mussel *Bathymodiolus azoricus*, and an apparent link to the timing of photosynthetic primary production. *J Mar Biol Assoc UK* 86:1363-1371
- Donval JP, Charlou JL, Knoery J, Fouquet Y, Costa I, Lourenço N, Ondreas H, Pellé H, Segonzac M, Tivey M (1994) Compared chemistry of hydrothermal fluids collected with the Nautile at Lucky Strike (37°17'N) and the Menez-Gwen (37°50'N) sites on

- the Mid Atlantic Ridge (DIVA 1 cruise - May 1994). Eos, Transactions, American Geophysical Union 75:309
- Douville E, Charlou JL, Oelkers EH, Bienvenu P, Colon CFJ, Donval JP, Fouquet Y, Prieur D, Appriou P (2002) The rainbow vent fluids (36 degrees 14 ' N, MAR): the influence of ultramafic rocks and phase separation on trace metal content in Mid-Atlantic Ridge hydrothermal fluids. Chem Geol 184:37-48
- Dover G (1982) Molecular drive: a cohesive mode of species evolution. Nature 299:111-117
- Duda TF, Palumbi SR (1999) Molecular genetics of ecological diversification: Duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. Proc Natl Acad Sci U S A 96:6820-6823
- Duperron S, Bergin C, Zielinski F, Blazejak A, Pernthaler A, McKiness ZP, DeChaine E, Cavanaugh CM, Dubilier N (2006) A dual symbiosis shared by two mussel species, *Bathymodiolus azoricus* and *Bathymodiolus puteoserpentis* (Bivalvia : Mytilidae), from hydrothermal vents along the northern Mid-Atlantic Ridge. Environ Microbiol 8:1441-1447
- Duperron S, Nadalig T, Caprais JC, Sibuet M, Fiala-Medioni A, Amann R, Dubilier N (2005) Dual symbiosis in a *Bathymodiolus* sp mussel from a methane seep on the gabon continental margin (southeast Atlantic): 16S rRNA phylogeny and distribution of the symbionts in gills. Appl Environ Microbiol 71:1694-1700
- Durand JP, Goudard F, Pieri J, Escoubas JM, Schreiber N, Cadoret JP (2004) *Crassostrea gigas* ferritin: cDNA sequence analysis for two heavy chain type subunits and protein purification. Gene 338:187-195

- E -

- Ejiri S (2002) Moonlighting functions of polypeptide elongation factor 1: From actin bundling to zinc finger protein R1-associated nuclear localization. Bioscience Biotechnology and Biochemistry 66:1-21
- El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L) Skeels] endemic to Morocco. Theor Appl Genet 92:832-839
- Endler JA (1977) Geographic variation, speciation and gene flow, Vol. Princeton Univ. Press, Princeton, NJ
- Engel CR, Daguin C, Serrao EA (2005) Genetic entities and mating system in hermaphroditic *Fucus spiralis* and its close dioecious relative *F. vesiculosus* (Fucaceae, Phaeophyceae). Mol Ecol 14:2033-2046
- Evgen'ev MB, Zatsepina OG, Garbuz D, Lerman DN, Velikodvorskaya V, Zelentsova E, Feder ME (2004) Evolution and arrangement of the hsp70 gene cluster in two closely related species of the virilis group of *Drosophila*. Chromosoma 113:223-232
- Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 47-50
- Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. Trends Ecol Evol 23:347-351
- Eyre-Walker A (2006) The genomics rate of adaptive evolution. Trends in Ecology and Evolution 21:569-575
- Eyre-Walker A, Woolfit M, Phelps T (2006) The Distribution of Fitness Effects of New Deleterious Amino Acid Mutations in Humans. Genetics Society of America 173:891-900

- F -

- Faugères JC, Desbruyères D, Gonthier E, Gribouillard R, Poutiers J, de Resseguier A, Vernet G (1987) Témoins sédimentologiques et biologiques de l'activité tectonique actuelle du prisme d'accrétion de la Barbade. *Comptes Rendus de l'Académie des Sciences de Paris, Série II* 305:115-119
- Faure B, Bierne N, Tanguy A, Bonhomme F, Jollivet D (2007a) Evidence for a slightly deleterious effect of intron polymorphisms at the EF1 alpha gene in the deep-sea hydrothermal vent bivalve *Bathymodiolus*. *Gene* 406:99-107
- Faure B, Chevaldonne P, Pradillon F, Thiebaut E, Jollivet D (2007b) Spatial and temporal dynamics of reproduction and settlement in the Pompeii worm *Advinella pompejana* (Polychaeta : Alvinellidae). *Marine Ecology-Progress Series* 348:197-211
- Faure M (2007) Génétique des barrières semi-perméables au flux génique dans la zone d'hybridation des moules *Mytilus edulis* et *M. galloprovincialis*. Université Montpellier II
- Faure M, David P, Bonhomme F, Bierne N (2008) Genetic hitchhiking in a subdivided population of *Mytilus edulis*. *BMC Evolutionary Biology* 8:164
- Fay JC, Wu C-I (2001) The neutral theory in the genomic era. *Curr Opin Genet Dev* 11:642-646
- Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics* 155:1405-1413
- Felbeck H (1981) Chemoautotrophic potential of the hydrothermal vent tube worm, *Riftia pachyptila* Jones (Vestimentifera). *Science* 213:336-338
- Feldman MW, Otto SP, Christiansen FB (1997) Population genetic perspectives on the evolution of recombination. *Annu Rev Genet* 30:261-295
- Felsenstein J (1974) Evolutionary advantage of recombination. *Genetics* 78:737-756
- Felsenstein J (1984) Distance methods for inferring phylogenies - a justification. *Evolution* 38:16-24
- Ferrini VL, Fornari DJ, Shank TM, Kinsey JC, Tivey MA, Soule SA, Carbotte SM, Whitcomb LL, Yoerger D, Howland J (2007) Submeter bathymetric mapping of volcanic and hydrothermal features on the East Pacific Rise crest at 9 degrees 50 ' N. *Geochem Geophys Geosyst* 8
- Fiala-Médioni A, Le Pennec M (1987) Trophic structural adaptations in relation to the bacterial association of bivalve molluscs from hydrothermal vents and subduction zones. *Symbiosis* 4:63-74
- Fiala-Medioni A, McKiness ZP, Dando P, Boulegue J, Mariotti A, Alayse-Danet AM, Robinson JJ, Cavanaugh CM (2002) Ultrastructural, biochemical, and immunological characterization of two populations of the mytilid mussel *Bathymodiolus azoricus* from the Mid-Atlantic Ridge: evidence for a dual symbiosis. *Mar Biol* 141:1035-1043
- Fiala-Médioni A, Métivier C, Herry A, Le Pennec M (1986) Ultrastructure of the gill of the hydrothermal-vent mytilid *Bathymodiolus* sp. *Mar Biol* 92:65-72
- Fisher CR (1990) Chemoautotrophic and Methanotrophic Symbioses in Marine-Invertebrates. *Reviews in Aquatic Sciences* 2:399-436
- Fisher CR, Childress JJ, Oremland RS, Bidigare RR (1987) The importance of methane and thiosulphate in the metabolism of the bacterial symbionts of two deep-sea mussels. *Mar Biol* 96:59-71
- Fisher RA (1930) *The genetical theory of natural selection*, Vol, Oxford
- Fitch WM, Markowitz E (1970) An improved method for determining codon variability in a gene and its application to rate of fixation of mutations in evolution. *Biochem Genet* 4:579-593

- Fouquet Y, Charlou JL, Ondreas H, RadfordKnoery J, Donval J-P, Douville E, Apprioual R, Cambon P, Pellé H, Landuré JY, Normand A, Ponzevera E, German C, Parson L, Barriga F, Costa I, Relvas J, Ribeiro A (1997) The Rainbow hydrothermal vent field on the MAR (36°14' N). *Eos Trans* 78:832
- Fox IH, Kelley WN (1978) Role of adenosine and 2'-deoxyadenosine in mammalian-cells. *Annu Rev Biochem* 47:655-686
- France SC, Hessler RR, Vrijenhoek RC (1992) Genetic differentiation between spatially-disjunct populations of the deep-sea, hydrothermal vent-endemic amphipod *Ventiella sulfuris*. *Mar Biol* 114:551-559
- Fu Y-X, Li W-H (1993) Statistical Tests of Neutrality of Mutations. *Genetics Society of America* 133:693-709
- Fustec A, Desbruyères D, Juniper SK (1987) Deep-sea hydrothermal vent communities at 13°N on the East Pacific Rise: microdistribution and temporal variations. *Biological Oceanography* 4:121-164

- G -

- Gage JD, Tyler PA (1991) Deep-sea biology: a natural history of organisms at the deep-sea floor, Vol. Cambridge University Press, Cambridge
- Gaill F, Desbruyères D, Laubier L (1988) Relationships between the "Pompeii worms" and their epibiotic bacteria. *Oceanologica Acta N° SP.* 8:147-154
- Gaill F, Desbruyères D, Prieur D (1987) Bacterial communities associated with "Pompeii worms" from the East Pacific Rise hydrothermal vents: SEM, TEM observations. *Microbial Ecology* 13:129-139
- Gaill F, Hunt S (1986) Tubes of deep sea hydrothermal vent worms *Riftia pachyptila* (Vestimentifera) and *Alvinella pompejana* (Annelida). *Marine Ecology Progress Series* 34:267-274
- Galkin SV, Vinogradov GM, team atMs (2004) Russian biological studies using MIR submersibles at north atlantic and east pacific hydrothermal sites. *InterRidge News* 13:27-33
- Galtier N, Depaulis F, Barton NH (2000) Detecting Bottlenecks and Selective Sweeps From DNA Sequence Polymorphism. *Genetics Society of America* 155:981-987
- Gardner JPA, Skibinski DOF (1990) Thermostability Differences of Allozyme Loci in *Mytilus-Edulis*, *Mytilus-Galloprovincialis* and Hybrid Mussels. *Marine Ecology-Progress Series* 64:99-105
- Gebruk AV, Chevaldonne P, Shank T, Lutz RA, Vrijenhoek RC (2000) Deep-sea hydrothermal vent communities of the Logatchev area (14 degrees 45 ' N, Mid-Atlantic Ridge): diverse biotopes and high biomass. *J Mar Biol Assoc UK* 80:383-393
- Genio L, Johnson SB, Vrijenhoek RC, Cunha MR, Tyler PA, Kiel S, Little CTS (2008) New record of "Bathymodiolus" mauritanicus Cosel 2002 from the Gulf of Cadiz (NE Atlantic) mud volcanoes. *J Shellfish Res* 27:53-61
- German CR, Parson LM (1998) Distributions of hydrothermal activity along the Mid-Atlantic Ridge: interplay of magmatic and tectonic controls. *Earth Planet Sci Lett* 160:327-341
- Geyer CJ (1992) Practical Markov Chain Monte Carlo. *Statistical Science* 7:473-511
- Glinka S, Ometto L, Mousset S, Stephan W, DeLorenzo D (2003) Demography and Natural Selection Have Shaped Genetic Variation in *Drosophila melanogaster*: A Multi-locus Approach. *Genetics Society of America* 165:1269-1278
- Goffredi SK, Barry JP (2002) Species-specific variation in sulfide physiology between closely related Vesicomid clams. *Marine Ecology-Progress Series* 225:227-238

- Goffredi SK, Hurtado LA, Hallam S, Vrijenhoek RC (2003) Evolutionary relationships of deep-sea vent and cold seep clams (Mollusca : Vesicomidae) of the "pacifica/lepta" species complex. *Mar Biol* 142:311-320
- Goldman N, Yang ZH (1994) Codon-based model of nucleotide substitution for protein-coding DNA-sequences. *Mol Biol Evol* 11:725-736
- Goulet J (2002) Fstat, a program to estimate and test gene diversities and fixation indices (Version 2.9.3.2).
- Grassle JF (1985a) Hydrothermal vent animals: distribution and biology. *Science* 229:713-717
- Grassle JP (1985b) Genetic differentiation in populations of hydrothermal vent mussels (*Bathymodiolus thermophilus*) from the Galapagos Rift and 13°N on the East Pacific Rise. *Bulletin of the Biological Society of Washington* 6:429-442
- Guinot D, Hurtado LA (2003) Two new species of hydrothermal vent crabs of the genus *Bythograea* from the southern East Pacific Rise and from the Galapagos Rift (Crustacea Decapoda Brachyura Bythograeidae). *C R Biol* 326:423-439
- Gustafson RG, Turner RD, Lutz RA, Vrijenhoek RC (1998) A new genus and five new species of mussels (Bivalvia, Mytilidae) from deep-sea sulfide/hydrocarbon seeps in the Gulf of Mexico. *Malacologia* 40:63-112

- H -

- Hadrill PR, Charlesworth B, Halligan DL, Andolfatto P (2005a) Patterns of intron sequence evolution in *Drosophila* are dependent upon length and GC content. *Genome Biol* 6
- Hadrill PR, Thornton KR, Charlesworth B, Andolfatto P (2005b) Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Res* 15:790-799
- Hadfield MG, Carpizo-Ituarte EJ, DelCarmen K, Nedved B (2001) Metamorphic competence, a major adaptative convergence in marine, invertebrate larvae. *Am Zool* 41:1123-1131
- Halanych KM, Janosik A (2006) A review of molecular markers used for Annelid phylogenetics. *Integrative and comparative biology* 40:533-543
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41:95-98
- Hardison RC (2000) Conserved noncoding sequences are reliable guides to regulatory elements. *Trends Genet* 16:369-372
- Harris DJ, Crandall KA (2000) Intragenomic variation within ITS1 and ITS2 of freshwater crayfishes (Decapoda : Cambaridae): Implications for phylogenetic and microsatellite studies. *Mol Biol Evol* 17:284-291
- Harrison RG (1990) Hybrid zones: windows on evolutionary processes. *Oxf Surv Evol Biol* 7:69-128
- Hartl DL (1988) A primer of population genetics, Vol. Sinauer, Sunderland, MA
- Hasegawa M, Kishino H, Yano TA (1985) Dating of the Human Ape Splitting by a Molecular Clock of Mitochondrial-DNA. *J Mol Evol* 22:160-174
- Hashimoto J (2001) A new species of *Bathymodiolus* (Bivalvia: Mytilidae) from hydrothermal vent communities in the Indian Ocean. *Venus* 60:141-149
- Hashimoto J, Jollivet D, Party KS (1989) The hydrothermal vent communities in the North Fiji Basin: results of Japan-France cooperative research on board KAIYO 88. *La Mer* 27:62-71
- Hashimoto J, Miura T, Fujikura K, Oosaka J (1993) Discovery of vestimentiferan tube-worms in the euphotic zone. *Zoological Science* 10:1063-1067

- Hashimoto J, Okutani T (1994) Four new mytilid mussels associated with deepsea chemosynthetic communities around Japan. *Venus: Japanese Journal of Malacology* 53:61-83
- Hashimoto J, Yamane T (2005) A new species of *Gigantidas* (Bivalvia: Mytilidae) from a vent site on the Kaikata Seamount southwest of the Ogasawara (Bonin) Islands, southern Japan. *Venus* 64:1-10
- Haymon RM, Carbotte S, Wright D, Beedle N, Johnson F, Fornari DJ, Von Damm K, Grebmeier J, Lilley M, McLaughlin E, Olson E, Edmond J, Perfit MR, Smith M, Shanks WC, III, Lutz R, Seewald J, Reudelhuber D (1991a) Active eruption seen on East Pacific Rise. *Eos, Transactions, American Geophysical Union* 72:505-507
- Haymon RM, Fornari DJ, Edwards MH, Carbotte S, Wright D, Macdonald KC (1991b) Hydrothermal vent distribution along the East Pacific Rise crest (9°09'-54'N) and its relationship to magmatic and tectonic processes on fast-spreading mid-ocean ridges. *Earth Planet Sci Lett* 104:513-534
- Haymon RM, Koski RA, Sinclair C (1984) Fossils of hydrothermal vent worms from Cretaceous Sulfide Ores of the Samail Ophiolite, Oman. *Science* 223:1407-1409
- Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proc R Soc Lond Ser B-Biol Sci* 270:313-321
- Hedgecock D, Chow V, Waples RS (1992) Effective population numbers of shellfish broodstocks estimated from temporal variance in allelic frequencies. *Aquaculture* 108:215-232
- Hedgecock D, Sly F (1990) Genetic drift and effective population sizes of hatchery-propagated stocks of the Pacific oyster, *Crassostrea gigas*. *Aquaculture* 88:21-38
- Helfrich KR, Battisti TM (1991) Experiments on Baroclinic Vortex Shedding from Hydrothermal Plumes. *J Geophys Res-Oceans* 96:12511-12518
- Hellberg ME (1994) Relationships between inferred levels of gene flow and geographic distance in a philopatric coral, *Balanophyllia elegans*. *Evolution* 48:1829-1854
- Hessler RR, Lonsdale PF (1991) Biogeography of Mariana Trough hydrothermal vent communities. *Deep-Sea Research* 38:185-199
- Hessler RR, Smithey WM, Jr., Keller CH (1985) Spatial and temporal variation of giant clams, tube worms and mussels at deep-sea hydrothermal vents. *Bulletin of the Biological Society of Washington* 6:411-428
- Hey J (2006a) On the failure of modern species concepts. *Trends Ecol Evol* 21:447-450
- Hey J (2006b) Recent advances in assessing gene flow between diverging populations and species. *Curr Opin Genet Dev* 16:592-596
- Hey J, Machado CA (2003) The study of structured populations - New hope for a difficult and divided science. *Nat Rev Genet* 4:535-543
- Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D-persimilis*. *Genetics* 167:747-760
- Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc Natl Acad Sci U S A* 104:2785-2790
- Hill WG, Robertson A (1966) The effect of linkage on limits to artificial selection. *Genet Res* 8:269-294
- Hilton H, Kliman RM, Hey J (1994) Using Hitchhiking Genes to Study Adaptation and Divergence During Speciation within the *Drosophila-Melanogaster* Species Complex. *Evolution* 48:1900-1913
- Howard DJ (1999) Conspecific sperm and pollen precedence and speciation. *Annu Rev Ecol Syst* 30:109-132

- Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA-sequences. *Genetics* 111:147-164
- Hudson RR, Kreitman M, Aguadé M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153-159
- Hudson RR, Turelli M (2003) Stochasticity overrules the "Three-Times Rules": genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution* 57:182-190
- Hurst GDD, Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proc R Soc B-Biol Sci* 272:1525-1534
- Hurtado LA, Lutz RA, Vrijenhoek RC (2004) Distinct patterns of genetic differentiation among annelids of eastern Pacific hydrothermal vents. *Mol Ecol* 13:2603-2615
- Hurtado LA, Mateos M, Lutz RA, Vrijenhoek RC (2002) Molecular evidence for multiple species of *Oasisia* (Annelida : Siboglinidae) at eastern Pacific hydrothermal vents. *Cah Biol Mar* 43:377-380
- Hurtado LA, Mateos M, Lutz RA, Vrijenhoek RC (2003) Coupling of bacterial endosymbiont and host mitochondrial genomes in the hydrothermal vent clam *Calyptogena magnifica*. *Appl Environ Microbiol* 69:2058-2064
- Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23:254-267
- Huvet A, Balabaud K, Bierne N, Boudry P (2001) Microsatellite analysis of 6-hour-old embryos reveals no preferential intraspecific fertilization between cupped oysters *Crassostrea gigas* and *Crassostrea angulata*. *Mar Biotechnol* 3:448-453
- Huvet A, Boudry P, Ohresser M, Delsert C, Bonhomme F (2000) Variable microsatellites in the Pacific cupped oyster *Crassostrea gigas* and other cupped oyster species. *Anim Genet* 31:71-72
- Huvet A, Gérard A, Ledu C, Phélot P, Heurtebise S, Boudry P (2002) Is fertility of hybrids enough to conclude that the two oysters *Crassostrea gigas* and *Crassostrea angulata* are the same species? *Aquat Living Resour* 15, 42-52

- I -

- Irwin DE (2000) Song variation in an avian ring species. *Evolution* 54:998-1010
- Iwasaki H, Kyuno A, Shintaku M, Fujita Y, Fujiwara Y, Fujikura K, Hashimoto J, Martins LD, Gebruk A, Miyazaki JI (2006) Evolutionary relationships of deep-sea mussels inferred by mitochondrial DNA sequences. *Mar Biol* 149:1111-1122

- J -

- Jackson JBC (1986) Modes of dispersal of clonal benthic invertebrates: consequences for species' distributions and genetic structure of local populations. *Bulletin of Marine Science* 39:588-606
- James RH, Elderfield H, Palmer MR (1995) The chemistry of hydrothermal fluids from the Broken Spur site, 29°N Mid-Atlantic Ridge. *Geochim Cosmochim Acta* 59:651-659
- Jannasch HW (1985) The chemosynthetic support of life and the microbial diversity at deep-sea hydrothermal vents. *Proceedings of the Royal Society of London, B* 225:277-297
- Jannasch HW, Mottl MJ (1985) Geomicrobiology of deep-sea hydrothermal vents. *Science* 229:717-725
- Jaruzelska J, Zietkiewicz E, Labuda D (1999) Is Selection Responsible for the Low Level of Variation in the Last Intron of the *ZFY* Locus? *Mol Biol Evol* 16(11):1633-1640

- Jelmert A, OppenBerntsen DO (1996) Whaling and deep-sea biodiversity. *Conserv Biol* 10:653-654
- Jiggins CD, McMillan WO, King P, Mallet J (1997) The maintenance of species differences across a *Heliconius* hybrid zone. *Heredity* 79:495-505
- Johnson KS, Childress JJ, Hessler RR, Sakamoto-Arnold CM, Beehler CL (1988) Chemical and biological interactions in the Rose Garden hydrothermal vent field, Galapagos spreading center. *Deep-Sea Research* 35:1723-1744
- Johnson SB, Waren A, Vrijenhoek RC (2008) DNA barcoding of *Lepetodrilus* limpets reveals cryptic species. *J Shellfish Res* 27:43-51
- Johnson SB, Young CR, Jones WJ, Waren A, Vrijenhoek RC (2006) Migration, isolation, and speciation of hydrothermal vent limpets (Gastropoda ; Lepetodrilidae) across the Blanco Transform Fault. *Biol Bull* 210:140-157
- Jollivet D (1993) Distribution et évolution de la faune associée aux sources hydrothermales profondes à 13°N sur la dorsale du Pacifique oriental : le cas particulier des polychètes Alvinellidae. Volume 1. Thèses de doctorat, Université de Bretagne Occidentale, Brest, 353pp
- Jollivet D (1996) Specific and genetic diversity at deep-sea hydrothermal vents: an overview. *Biodiversity and Conservation* 5:1619-1653
- Jollivet D, Chevaldonné P, Planque B (1999) Hydrothermal-vent alvinellid polychaete dispersal in the eastern Pacific. 2. A metapopulation model based on habitat shifts. *Evolution* 53:1128-1142
- Jollivet D, Comtet T, Chevaldonne P, Hourdez S, Desbruyeres D, Dixon DR (1998a) Unexpected relationship between dispersal strategies and speciation within the association *Bathymodiolus* (Bivalvia) - *Branchipolynoe* (Polychaeta) inferred from the rDNA neutral ITS2 marker. *Cah Biol Mar* 39:359-362
- Jollivet D, Desbruyères D, Bonhomme F, Moraga D (1995a) Genetic differentiation of deep-sea hydrothermal vent alvinellid populations (Annelida: Polychaeta) along the East Pacific Rise. *Heredity* 74:376-391
- Jollivet D, Desbruyères D, Ladrat C, Laubier L (1995b) Evidence for differences in the allozyme thermostability of deep-sea hydrothermal vent polychaetes (Alvinellidae): a possible selection by habitat. *Marine Ecology Progress Series* 123:125-136
- Jollivet D, Dixon LRJ, Desbruyères D, Dixon DR (1998b) Ribosomal (rDNA) variation in a deep sea hydrothermal vent polychaete, *Alvinella pompejana*, from 13°N on the East Pacific Rise. *Journal of the Marine Biological Association of the U K* 78:113-130
- Jollivet D, Hashimoto J, Auzende JM, Honza E, Ruellan E, Dutt S, Iwabushi Y, Jarvis P, Joshima M, Kawai T, Kawamoto T, Kisimoto K, Lafoy Y, Matsumoto T, Mitsuzawa K, Naganuma T, Naka J, Otsuka K, Otsuki A, Rao B, Tanahashi M, Tanaka T, Temakon JS, Urabe T, Veivau T, Yokokura T (1989) Premières observations de communautés animales associées à l'hydrothermalisme arrière-arc du bassin Nord-Fidjien. *Comptes Rendus de l'Académie des Sciences de Paris, Série III* 309:301-308
- Jollivet D, Lallier FH, Barnay A-S, Bienvenu N, Bonnivard E, Briand P, Cambon-Bonavita M-A, Comtet T, Cosson R, Daguin C, Donva IJ-P, Faure B, Gaillard M, Glippa V, Guillou L, Hourdez S, Le Bris N, Morand S, Pradillon F, Rees J-F, Segonzac M, Shillito B, Thiébaud E, Viard F (2004) The BIOSPEEDO cruise: a new survey of hydrothermal vents along the South East Pacific Rise from 7°24'S to 21°33'S. *InterRidge News* 13:20-26
- Jolly MT (2005) Structures génétiques et histoires évolutives de polychètes inféodés aux sédiments fins envasés dans l'Atlantique Nord Est : les genres *Pectinaria sp.* et *Owenis sp.*, Université Paris VI, 206

- Jones DT, Taylor WR, Thornton JM (1992) The Rapid Generation of Mutation Data Matrices from Protein Sequences. *Comput Appl Biosci* 8:275-282
- Jones WJ, Vrijenhoek RC (2006) Evolutionary relationships within the "Bathymodiolus" childressi group. *Cah Biol Mar* 47:403-407
- Jones WJ, Won YJ, Maas PAY, Smith PJ, Lutz RA, Vrijenhoek RC (2006) Evolution of habitat use by deep-sea mussels. *Mar Biol* 148:841-851
- Joseph N, Krauskopf E, Vera MI, Michot B (1999) Ribosomal internal transcribed spacer 2 (ITS2) exhibits a common core of secondary structure in vertebrates and yeast. *Nucleic Acids Res* 27:4533-4540
- Joyce TM, Cannon GA, Pashinski D, Helfrich KR, Harrington SA (1998) Vertical and temporal vorticity observations at Juan de Fuca Ridge: hydrothermal signatures. *Geophys Res Lett* 25:1741-1744

- K -

- Kadar E, Bettencourt R, Costa V, Santos RS, Lobo-Da-Cunha A, Dando P (2005) Experimentally induced endosymbiont loss and re-acquirement in the hydrothermal vent bivalve *Bathymodiolus azoricus*. *J Exp Mar Biol Ecol* 318:99-110
- Kadko D (1996) Radioisotopic studies of submarine hydrothermal vents. *Rev Geophys* 34:349-366
- Kaplan NL, Hudson RR, Langley CH (1989) The "Hitchhiking Effect" Revisited. *Genetics* 123:887-899
- Karl DM, Wirsén CO, Jannasch HW (1980) Deep-sea primary production at the Galapagos hydrothermal vents. *Science* 207:1345-1347
- Karl SA, Schutz S, Desbruyères D, Lutz R, Vrijenhoek RC (1996) Molecular analysis of gene flow in the hydrothermal vent clam (*Calymene magnifica*). *Molecular Marine Biology and Biotechnology* 5:193-202
- Kazazian HH (2004) Mobile elements: Drivers of genome evolution. *Science* 303:1626-1632
- Kenk VC, Wilson BR (1985) A new mussel (*Bivalvia*, *Mytilidae*) from hydrothermal vents in the Galapagos Rift zone. *Malacologia* 26:253-271
- Kennicutt MC, II, Brooks JM, Bidigare RR, Fay RR, Wade TL, McDonald TJ (1985) Vent-type taxa in a hydrocarbon seep region on the Louisiana slope. *Nature* 317:351-353
- Kiang JG, Tsokos GC (1998) Heat shock protein 70 kDa: Molecular biology, biochemistry, and physiology. *Pharmacol Ther* 80:183-201
- Kiel S, Goedert JL (2006) Deep-sea food bonanzas: early Cenozoic whale-fall communities resemble wood-fall rather than seep communities. *Proc R Soc B-Biol Sci* 273:2625-2631
- Kim SL, Mullineaux LS (1998) Distribution and near-bottom transport of larvae and other plankton at hydrothermal vents. *Deep-Sea Research II* 45:423-440
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111-120
- Kimura M, Weiss GH (1964) Stepping Stone Model of Population Structure + Decrease of Genetic Correlation with Distance. *Genetics* 49:561-&
- Kingman JFC (1982) The coalescent. *Stochastic Process Appl* 13
- Kinlan BP, Gaines SD (2003) Propagule dispersal in marine and terrestrial environments: A community perspective. *Ecology* 84:2007-2020
- Kliman RM, Hey J (1993) Reduced natural-selection associated with low recombination in *Drosophila melanogaster*. *Mol Biol Evol* 10:1239-1258

- Kloor D, Fumic K, Attig S, Tete M, Osswald H, Baric I, Tomiuk J, Kompf J (2006) Studies of S-adenosylhomocysteine-hydrolase polymorphism in a Croatian population. *Journal of Human Genetics* 51:21-24
- Knowlton N (1993) Sibling species in the sea. *Annu Rev Ecol Syst* 24:189-216
- Knowlton N (2000) Molecular genetic analyses of species boundaries in the sea. *Hydrobiologia* 420:73-90
- Kreitman M (1983) Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature* 304:412-417
- Krueger DM, Gustafson RG, Cavanaugh CM (1996) Vertical transmission of chemoautotrophic symbionts in the bivalve *Solemya velum* (Bivalvia: Protobranchia). *Biol Bull* 190:195-202
- Kruuk LEB, Baird SJE, Gale KS, Barton NH (1999) A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. *Genetics* 153:1959-1971
- Kulm LD, Suess E, Moore JC, Carson B, Lewis BT, Ritger SD, Kadko DC, Thornburg TM, Embley RW, Rugh WD, Massoth GJ, Langseth MG, Cochrane GR, Scamman RL (1986) Oregon Subduction Zone - Venting, Fauna, and Carbonates. *Science* 231:561-566
- Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics* 5(2):150-163
- Kwiatowski J, Skarecky D, Hernandez S, Pham D, Quijas F, Ayala FJ (1991) High fidelity of the polymerase chain reaction. *Mol Biol Evol* 8:884-887

- L -

- Lagabrielle Y (2005) La dorsale est-Pacifique entre 10° et 20° S. Alternance du volcanisme et de la tectonique le long de la zone active axiale. *Géomorphologie : relief, processus, environnement* 2:105-120
- Lalou C, Brichet E, Hekinian R (1985) Age Dating of Sulfide Deposits from Axial and Off-Axial Structures on the East Pacific Rise near 12-Degrees-50'n. *Earth Planet Sci Lett* 75:59-71
- Lalou C, Reyss JL, Brichet E (1993a) Actinide-series disequilibrium as a tool to establish the chronology of deep-sea hydrothermal activity. *Geochim Cosmochim Acta* 57:1221-1231
- Lalou C, Reyss JL, Brichet E, Arnold M (1990) Chronology of MAR hydrothermal fields, comparison with fast spreading ridges. *Eos, Transactions, American Geophysical Union* 71:1654
- Lalou C, Reyss JL, Brichet E, Arnold M, Thompson G, Fouquet Y, Rona PA (1993b) New age data for Mid-Atlantic Ridge hydrothermal sites: TAG and Snakepit chronology revisited. *Journal of Geophysical Research, B* 98:9705-9713
- Larson RL, Searle RC, Kleinrock MC, Schouten H, Bird RT, Naar DF, Rusby RI, Hooft EE, Lasthiotakis H (1992) Roller-Bearing Tectonic Evolution of the Juan-Fernandez Microplate. *Nature* 356:571-576
- Latta RG, Mitton JB (1999) Historical separation and present gene flow through a zone of secondary contact in ponderosa pine. *Evolution* 53:769-776
- Launey S, Hedgecock D (2001) High genetic load in the Pacific oyster *Crassostrea gigas*. *Genetics* 159:255-265
- Le NTV, Richardson DR (2002) The role of iron in cell cycle progression and the proliferation of neoplastic cells. *Biochim Biophys Acta-Rev Cancer* 1603:31-46

- Leicht BG, Muse SV, Hanczyc M, Clark AG (1995) Constraints on Intron Evolution in the Gene Encoding the Myosin Alkali Light Chain in *Drosophila*. *Genetics Society of America* 139:299-308
- Levinson W, Oppermann H, Jackson J (1980) Transition series metals and sulfhydryl-reagents induce the synthesis of four proteins in eukaryotic cells. *Biochimica Et Biophysica Acta* 606:170-180
- Lewontin RC (1991) Electrophoresis in the Development of Evolutionary Genetics: Milestone or Millstone? *Genetics* 128:657-662
- Li H, Stephan W (2006) Inferring the Demographic History and Rate of Adaptive Substitution in *Drosophila*. *PLoS genetics* 10:1580-1589
- Li TH, Spearow J, Rubin CM, Schmid CW (1999) Physiological stresses increase mouse short interspersed element (SINE) RNA expression in vivo. *Gene* 239:367-372
- Li W-H, Graur D (1991) *Fundamentals of molecular evolution*, Vol. Sinauer Associates, Sunderland, MA
- Little CTS, Cann JR (1999) Late Cretaceous hydrothermal vent communities from the Troodos ophiolite, Cyprus. *Geology* 27:1027-1030
- Locard A (1889) Révision des espèces françaises appartenant au genre *Mytilus*. *Bull Soc Malac Fr* 6:83-164
- Lonsdale P (1977) Clustering of suspension-feeding macrobenthos near abyssal hydrothermal vents at oceanic spreading centers. *Deep-Sea Research* 24:857-863
- Lowell RP, Rona PA, Von Herzen RP (1995) Seafloor hydrothermal systems. *J Geophys Res* 100:327-352
- Lu G, Basley DJ, Bernatchez L (2001) Contrasting patterns of mitochondrial DNA and microsatellite introgressive hybridization between lineages of lake whitefish (*Coregonus clupeaformis*); relevance for speciation. *Mol Ecol* 10:965-985
- Lundberg KS, Shoemaker DD, Adams MWW, Short JM, Sorge JA, Mathur EJ (1991) High-Fidelity Amplification Using a Thermostable DNA-Polymerase Isolated from *Pyrococcus-Furiosus*. *Gene* 108:1-6
- Lutz RA, Bouchet P, Jablonski D, Turner RD, Warén A (1986) Larval ecology of mollusks at deep-sea hydrothermal vents. *American Malacological Bulletin* 4:49-54
- Lutz RA, Jablonski D, Rhoads DC, Turner RD (1980) Larval dispersal of a deep-sea hydrothermal vent bivalve from the Galapagos Rift. *Mar Biol* 57:127-133
- Lynch M (2002) Intron evolution as a population-genetic process. *Proc Natl Acad Sci U S A* 99:6118-6123
- Lynch M, Blanchard JL (1998) Deleterious mutation accumulation in organelle genomes. *Genetica* 102/103:29-39
- Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290:1151-1155
- Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154:459-473
- Lyngso RB, Song YS, Hein J (2005) Minimum recombination histories by branch and bound. In: *Algorithms in Bioinformatics, Proceedings*, Vol 3692. Springer-Verlag Berlin, Berlin, p 239-250

- M -

- Macdonald KC, Fox PJ (1983) Overlapping Spreading Centers - New Accretion Geometry on the East Pacific Rise. *Nature* 302:55-58
- Macdonald KC, Scheirer DS, Carbotte SM (1991) Midocean Ridges - Discontinuities, Segments and Giant Cracks. *Science* 253:986-994

- Machado CA, Hey J (2003) The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proc R Soc Lond Ser B-Biol Sci* 270:1193-1202
- Machado CA, Kliman RM, Markert JA, Hey J (2002) Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Mol Biol Evol* 19(4):472-488
- Malatesta F, Antonini G, Sarti P, Brunori M (1995) Structure and Function of a Molecular Machine - Cytochrome-C-Oxidase. *Biophys Chem* 54:1-33
- Mammerickx J, Herron E, Dorman L (1980) Evidence for 2 fossil spreading ridges in the southeast Pacific. *Geol Soc Am Bull* 91:263-271
- Mantel N (1967) Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Res* 27:209-&
- Marie B, Genard B, Rees J-F, Zal F (2006) Effect of ambient oxygen concentration on activities of enzymatic antioxidant defences and aerobic metabolism in the hydrothermal vent worm, *Paralvinella grasslei*. *Mar Biol* 150:273-284
- Marsh AG, Leong PKK, Manahan DT (1999) Energy metabolism during embryonic development and larval growth of an Antarctic sea urchin. *J Exp Biol* 202:2041-2050
- Marsh AG, Mullineaux LS, Young CM, Manahan DT (2001) Larval dispersal potential of the tubeworm *Riftia pachyptila* at deep-sea hydrothermal vents. *Nature* 411:77-80
- Marshall E (2005) Taxonomy - Will DNA bar codes breathe life into classification? *Science* 307:1037-1037
- Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16:562-563
- Martineu P (1997) Adaptations physiologiques des polychètes des sources hydrothermales sous-marines à l'hydrogène sulfuré. Thèse de Doctorat, Université du Québec à Montréal
- Maside X, Bartolome C, Charlesworth B (2002) S-element insertions are associated with the evolution of the Hsp70 genes in *Drosophila melanogaster*. *Curr Biol* 12:1686-1691
- Matabos M (2007) Biodiversité des peuplements hydrothermaux de la dorsale du Pacifique oriental : influence des traits d'histoire de vie des organismes et de l'environnement géophysique et physico-chimique. Université Pierre et Marie Curie
- Matabos M, Thiebaut E, Le Guen D, Sadosky F, Jollivet D, Bonhomme F (2008) Geographic clines and stepping-stone patterns detected along the East Pacific Rise in the vetigastropod *Lepetodrilus elevatus* reflect species crypticism. *Mar Biol* 153:545-563
- Maynard Smith J, Haigh J (1974) The hitchhiking effect of a favorable gene. *Genet Res* 23:23-35
- Mayr E (1942) *Systematics and the Origin of Species*, Vol. Columbia University Press, New York
- McDonald JH, Kreitman M (1991) Adaptive Protein Evolution at the Adh Locus in *Drosophila*. *Nature* 351:652-654
- McGoldrick DJ, Hedgecock D (1997) Fixation, segregation and linkage of allozyme loci in inbred families of the Pacific oyster *Crassostrea gigas* (Thunberg): Implications for the causes of inbreeding depression. *Genetics* 146:321-334
- McHugh D (1989) Population structure and reproductive biology of two sympatric hydrothermal vent polychaetes, *Paralvinella pandorae* and *Paralvinella palmiformis*. *Mar Biol* 103:95-106
- McHugh D (1997) Molecular evidence that echiurans and pogonophorans are derived annelids. *Proceedings of the National Academy of Sciences of the USA* 94:8006-8009
- McKiness ZP, McMullin ER, Fisher CR, Cavanaugh CM (2005) A new bathymodioline mussel symbiosis at the Juan de Fuca hydrothermal vents. *Mar Biol* 148:109-116

- McLean JH (1985) Preliminary report on the limpets at hydrothermal vents. *Bulletin of the Biological Society of Washington* 6:159-166
- Métivier B, von Cosel R (1993) *Acharax alinae* n. sp., Solemyidae (Mollusca: Bivalvia) géante du bassin de Lau. *Comptes Rendus de l'Académie des Sciences de Paris, Série III* 316:229-237
- Michot B, Joseph N, Mazan S, Bachellerie JP (1999) Evolutionarily conserved structural features in the ITS2 of mammalian pre-rRNAs and potential interactions with the snoRNA U8 detected by comparative analysis of new mouse sequences. *Nucleic Acids Res* 27:2271-2282
- Miyazaki JI, Shintaku M, Kyuno A, Fujiwara Y, Hashimoto J, Iwasaki H (2004) Phylogenetic relationships of deep-sea mussels of the genus *Bathymodiolus* (Bivalvia : Mytilidae). *Mar Biol* 144:527-535
- Muller HJ (1964) The relation of recombination to mutational advance. *Mutat Res* 1:2-9
- Mullineaux LS, Kim SL, Wiebe PH (1992) Entrainment of larvae and zooplankton into the lateral plume from vents on the East Pacific Rise (9°-10°N). *Eos, Transactions, American Geophysical Union* 73:253
- Mullineaux LS, Mills SW, Sweetman AK, Beaudreau AH, Metaxas A, Hunt HL (2005) Vertical, lateral and temporal structure in larval distributions at hydrothermal vents. *Marine Ecology-Progress Series* 293:1-16
- Mullineaux LS, Speer KG, Thurnherr AM, Maltrud ME, Vangriesheim A (2002) Implications of cross-axis flow for larval dispersal along mid-ocean ridges. *Cah Biol Mar* 43:281-284
- Mullineaux LS, Wiebe PH, Baker ET (1995) Larvae of Benthic Invertebrates in Hydrothermal Vent Plumes over Juan-De-Fuca Ridge. *Mar Biol* 122:585-596
- Muths D, Davoult D, Gentil F, Jollivet D (2006) Incomplete cryptic speciation between intertidal and subtidal morphs of *Acrocnida brachiata* (Echinodermata : Ophiuroidea) in the Northeast Atlantic. *Mol Ecol* 15

- N -

- Naar DF, Hekinian R, Segonzac M, Francheteau J, Pito TT (2004) Vigorous venting and biology at Pito Seamount, Easter Microplate. *Geophysical monograph* 148:305-318
- Naar DF, Hey RN (1989) Recent Pacific-Easter-Nazca plate motion. In: Sinton JM (ed) *Evolution of Mid-Ocean Ridges, Vol 57. I.U.G.G Geophysical Monograph*, p 9-30
- Naar DF, Hey RN (1991) Tectonic evolution of the Easter Microplate. *Journal of Geophysical Research-Solid Earth and Planets* 96:7961-7993
- Nachman MW (1998) Deleterious mutations in animal mitochondrial DNA. *Genetica* 102/103:61-69
- Needham HD (1996) Some features of the North America-Africa plate boundary. *Journal of Continents Abstracts*:1, 834-835
- Nei M (1987) *Molecular Evolutionary Genetics*, Vol. Columbia Univ. Press, New York
- Nei M, Li WH (1979) Mathematical-Model for Studying Genetic-Variation in Terms of Restriction Endonucleases. *Proc Natl Acad Sci U S A* 76:5269-5273
- Nelson K, Fisher CR (2000) Absence of cospeciation in deep-sea vestimentiferan tube worms and their bacterial endosymbionts. *Symbiosis* 28:1-15
- Nielsen R, Wakeley J (2001) Distinguishing Migration From Isolation: A Markov Chain Monte Carlo Approach. *Genetics* 158:885-896
- Niemiller ML, Fitzpatrick BM, Miller BT (2008) Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae : Gyrinophilus) inferred from gene genealogies. *Mol Ecol* 17:2258-2275

Nosil P (2008) Speciation with gene flow could be common. *Mol Ecol* 17:2103-2106

- O -

- O'Mullan GD, Maas PAY, Lutz RA, Vrijenhoek RC (2001) A hybrid zone between hydrothermal vent mussels (*Bivalvia* : *Mytilidae*) from the Mid-Atlantic Ridge. *Mol Ecol* 10:2819-2831
- Oakeshott JG, Gibson JB, Anderson PR, Knibb WR, Anderson DG, Chambers GK (1982) Alcohol-Dehydrogenase and Glycerol-3-Phosphate Dehydrogenase Clines in *Drosophila-Melanogaster* on Different Continents. *Evolution* 36:86-96
- Odonnell K, Cigelnik E (1997) Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. *Mol Phylogenet Evol* 7:103-116
- Ohta S, Laubier L (1987) Deep biological communities in the subduction zone of Japan from bottom photographs taken during "Nautile" dives in the Kaiko Project. *Earth Planet Sci Lett* 83:329-342
- Ohta T (1988) Further simulation studies on evolution by gene duplication. *Evolution* 42:375-386
- Okutani T, Fujiwara Y, Fujikura K, Miyake H, Kawato M (2004) A mass aggregation of the mussel *Adipicola pacifica* (*Bivalvia*: *Mytilidae*) on submerged whale bone. *Venus* 63:61-64
- Olu-Le Roy K, Sibuet M, Fiala-Medioni A, Gofas S, Salas C, Mariotti A, Foucher JP, Woodside J (2004) Cold seep communities in the deep eastern Mediterranean Sea: composition, symbiosis and spatial distribution on mud volcanoes. *Deep-Sea Res Part I-Oceanogr Res Pap* 51:1915-1936
- Olu-Le Roy K, Von Cosel R, Hourdez S, Carney SL, Jollivet D (2007) Amphi-Atlantic cold-seep *Bathymodiolus* species complexes across the equatorial belt. *Deep-Sea Research I* 54: 1890-1911
- Ondreas H, Fouquet Y, Voisset M, RadfordKnoery J (1997) Detailed study of three contiguous segments of the Mid-Atlantic Ridge, South of the Azores (37 degrees N to 38 degrees 30' N), using acoustic imaging coupled with submersible observations. *Mar Geophys Res* 19:231-255
- Ondreas H, Olu K, Fouquet Y, Charlou JL, Gay A, Dennielou B, Donval JP, Fifis A, Nadalig T, Cochonat P, Cauquil E, Bourillet JF, Le Moigne M, Sibuet M (2005) ROV study of a giant pockmark on the Gabon continental margin. *Geo-Mar Lett* 25:281-292
- Orino K, Lehman L, Tsuji Y, Ayaki H, Torti SV, Torti FM (2001) Ferritin and the response to oxidative stress. *Biochem J* 357:241-247
- Orr HA (1995) The Population-Genetics of Speciation - the Evolution of Hybrid Incompatibilities. *Genetics* 139:1805-1813
- Orr HA, Masly JP, Presgraves DC (2004) Speciation genes. *Curr Opin Genet Dev* 14:675-679
- Otto SP, Whitton J (2000) Polyploid incidence and evolution. *Annu Rev Genet* 34:401-437

- P -

- Page RDM, Holmes EC (2000) *Molecular Evolution: A Phylogenetic Approach*, Vol. Blackwell Science
- Pailleret M, Haga T, Petit P, Prive-Gill C, Saedlou N, Gaill F, Zbinden M (2007) Sunken wood from the Vanuatu Islands: identification of wood substrates and preliminary description of associated fauna. *Marine Ecology-an Evolutionary Perspective* 28:233-241

- Palumbi SR (1992) Marine Speciation on a Small Planet. *Trends Ecol Evol* 7:114-118
- Palumbi SR (1994) Genetic-Divergence, Reproductive Isolation, and Marine Speciation. *Annu Rev Ecol Syst* 25:547-572
- Palumbi SR, Baker CS (1994) Contrasting population structure from nuclear intron sequences and mtDNA of Humpback Whales. *Mol Biol Evol* 11:426-435
- Pannell JR, Charlesworth B (1999) Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. *Evolution* 53:664-676
- Parsch J (2003) Selective Constraints on Intron Evolution in *Drosophila*. *Genetics Society of America* 165:1843-1851
- Pascoe PL, Patton SJ, Critcher R, Dixon DR (1996) Robertsonian polymorphism in the marine gastropod, *Nucella lapillus*: Advances in karyology using rDNA loci and NORs. *Chromosoma* 104:455-460
- Pasteur N, Pasteur G, Bonhomme F, Catalan J, Britton-Davidian J (1987) Manuel technique de génétique par électrophorèse des protéines, Vol. Lavoisier, Paris
- Paull CK, Hecker B, Commeau R, Freemanlynde RP, Neumann C, Corso WP, Golubic S, Hook JE, Sikes E, Curray J (1984) Biological Communities at the Florida Escarpment Resemble Hydrothermal Vent Taxa. *Science* 226:965-967
- Pechenik JA (1999) On the advantages and disadvantages of larval stages in benthic marine invertebrate life cycles. *Marine Ecology-Progress Series* 177:269-297
- Peek AS, Feldman RA, Lutz RA, Vrijenhoek RC (1998) Cospeciation of chemoautotrophic bacteria and deep sea clams. *Proc Natl Acad Sci U S A* 95:9962-9966
- Piccino PV, F.; Sarradin, P.-M.; Le Bris, N., Le Guen, D. et Jollivet, D.) (2004) Thermal selection of PGM-allozymes in newly founded populations of the thermotolerant vent polychaete *Alvinella pompejana* (polychaete: alvinellidae). *Proceedings of the Royal Society of London, Serie B*,
- Podos J (2007) Discrimination of geographical song variants by Darwin's finches. *Anim Behav* 73:833-844
- Pond D, Dixon D, Sargent J (1997) Wax-ester reserves facilitate dispersal of hydrothermal vent shrimps. *Marine Ecology Progress Series* 146:289-290
- Powell MA, Somero GN (1983) Blood components prevent sulfide poisoning of respiration of the hydrothermal vent tubeworm *Riftia pachyptila*. *Science* 219:297-299
- Powell MA, Somero GN (1986) Adaptations to Sulfide by Hydrothermal Vent Animals - Sites and Mechanisms of Detoxification and Metabolism. *Biol Bull* 171:274-290
- Pradillon F, Le Bris N, Shillito B, Young CM, Gaill F (2005a) Influence of environmental conditions on early development of the hydrothermal vent polychaete *Alvinella pompejana*. *J Exp Biol* 208:1551-1561
- Pradillon F, Shillito B, Young CM, Gaill F (2001) Developmental Arrest in Vent Worm Embryos. *Nature* 413:698-699
- Presgraves DC (2003) A fine-scale genetic analysis of hybrid Incompatibilities in *drosophila*. *Genetics* 163:955-972
- Prieur D, Benbouzid-Rollet N, Chamroux S, Durand P, Erauso G, Jacq E, Jeanthon C, Mével G, Vincent P (1989) Distribution de divers types métaboliques bactériens sur un site hydrothermal profond (dorsale du Pacifique oriental à 13°N). *Cah Biol Mar* 30:515-530
- Pruski AM, Dixon DR (2003) Toxic vents and DNA damage: first evidence from a naturally contaminated deep-sea environment. *Aquatic Toxicology* 64:1-13

- R -

- Rawson PD, Agrawal V, Hilbish TJ (1999) Hybridization between the blue mussels *Mytilus galloprovincialis* and *M. trossulus* along the Pacific coast of North America: evidence for limited introgression. *Mar Biol* 134:201-211
- Reid JL (1997) On the total geostrophic circulation of the Pacific Ocean: flow patterns, tracers, and transports. *Prog Oceanogr* 39:263-352
- Renard V, Hékinian R, Francheteau J, Ballard RD, Backer H (1985) Submersible observations at the axis of the ultra-fast-spreading East Pacific Rise (17°30' to 21°30'S). *Earth Planet Sci Lett* 75:339-353
- Requena JM, Lopez MC, Jimenezruiz A, Delatorre JC, Alonso C (1988) A Head-to-Tail Tandem Organization of Hsp70 Genes in *Trypanosoma-Cruzi*. *Nucleic Acids Res* 16:1393-1406
- Rhoads DC, Lutz RA, Revelas EC, Cerrato RM (1981) Growth of Bivalves at Deep-Sea Hydrothermal Vents Along the Galapagos Rift. *Science* 214:911-913
- Rice WR (1987) Speciation via habitat specialization: the evolution of reproductive isolation as a correlated character. *Evolutionary Ecology* 1:301-314
- Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends Ecol Evol* 16:351-358
- Roesijadi G, Crecelius EA (1984) Elemental composition of the hydrothermal vent clam *Calyptogena magnifica* from the East Pacific Rise. *Mar Biol* 83:155-161
- Rona PA (1986) Black smokers on the Atlantic seafloor: preliminary report. *Deep-Sea Newsletter* 12:10-11
- Ross MI, Scotese CR (1988) A hierarchical tectonic model of the Gulf of Mexico and Caribbean region. *Tectonophysics* 155:139-168
- Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496-2497
- Ruehl M, Erben U, Schuppan D, Wagner C, Zeller A, Freise C, Al-Hasani H, Loesekann M, Notter M, Wittig BM, Zeitz M, Dieterich W, Somasundaram R (2005) The elongated first fibronectin type III domain of collagen XIV is an inducer of quiescence and differentiation in fibroblasts and preadipocytes. *J Biol Chem* 280:38537-38543

- S -

- Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA-polymerase. *Science* 239:487-491
- Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425
- Salerno JL, Macko SA, Hallam SJ, Bright M, Won YJ, McKiness Z, Van Dover CL (2005) Characterization of symbiont populations in life-history stages of mussels from chemosynthetic environments. *Biol Bull* 208:145-155
- Samadi S, Quemere E, Lorion J, Tillier A, von Cosel R, Lopez P, Cruaud C, Couloux A, Boisselier-Dubayle MC (2007) Molecular phylogeny in mytilids supports the wooden steps to deep-sea vents hypothesis. *C R Biol* 330:446-456
- Sambrook J, Russel DW (2001) *Molecular cloning: a laboratory manual*, Vol. Cold Spring Harbor Laboratory Press
- Sanchez MS, Arnold J, Asmussen MA (2000) Symbiont survival and host-symbiont disequilibria under differential vertical transmission. *Genetics* 154:1347-1365

- Sanchez MS, Hastings A (2003) Uniform vertical transmission and selection in a host-symbiont system. Non-random symbiont distribution generates apparent differential selection. *J Theor Biol* 225:517-530
- Sarver SK, Foltz DW (1993) Genetic Population-Structure of a Species Complex of Blue Mussels (*Mytilus* Spp). *Mar Biol* 117:105-112
- Savolainen V, Anstett MC, Lexer C, Hutton I, Clarkson JJ, Norup MV, Powell MP, Springate D, Salamin N, Baker WJ (2006) Sympatric speciation in palms on an oceanic island. *Nature* 441:210-213
- Schierup MH, Hein J (2000) Consequences of recombination on traditional phylogenetic analysis. *Genetics* 156:879-891
- Schlötterer C, Hauser MT, Vonhaeseler A, Tautz D (1994) Comparative evolutionary analysis of rDNA ITS regions in drosophila. *Mol Biol Evol* 11:513-522
- Segonzac M (1992) The Hydrothermal Vent Communities of Snake Pit Area (Mid-Atlantic Ridge, 23-Degrees-N, 3,480 M) - Megafaunal Composition and Microdistribution. *Comptes Rendus Acad Sci Ser III-Sci Vie-Life Sci* 314:593-600
- Seyfried WE, Janecky DR (1985) Heavy-metal and sulfur transport during subcritical and supercritical hydrothermal Alteration of basalt - influence of fluid pressure and basalt composition and crystallinity. *Geochim Cosmochim Acta* 49:2545-2560
- Shank TM, Halanych KM (2007) Toward a mechanistic understanding of larval dispersal: insights from genomic fingerprinting of the deep-sea hydrothermal vent tubeworm *Riftia pachyptila*. *Marine Ecology-an Evolutionary Perspective* 28:25-35
- Shoemaker DD, Machado CA, Molbo D, Werren JH, Windsor DM, Herre EA (2002) The distribution of *Wolbachia* in fig wasps: correlations with host phylogeny, ecology and population structure. *Proc R Soc Lond Ser B-Biol Sci* 269:2257-2267
- Sibuet M, Olu K (1998) Biogeography, biodiversity and fluid dependence of deep-sea cold-seep communities at active and passive margins. *Deep-Sea Res Part II-Top Stud Oceanogr* 45:517-567
- Sidow A (1996) Gen(om)e duplications in the evolution of early vertebrates. *Curr Opin Genet Dev* 6:715-722
- Siebenaller J, Somero GN (1978) Pressure-adaptive difference in lactate dehydrogenases of congeneric fishes living at different depths. *Science* 201:255-257
- Slatkin M (1977) Gene flow and genetic drift in a species subject to frequent local extinctions. *Theor Popul Biol* 12:253-262
- Smith CR, Baco AR (2003) Ecology of whale falls at the deep-sea floor. *Oceanography and Marine Biology, Vol 41* 41:311-354
- Smith CR, Kukert H, Wheatcroft RA, Jumars PA, Deming JW (1989) Vent fauna on whale remains. *Nature* 341:27-28
- Smith PJ, McVeagh SM, Won Y, Vrijenhoek RC (2004) Genetic heterogeneity among New Zealand species of hydrothermal vent mussels (*Mytilidae* : *Bathymodiolus*). *Mar Biol* 144:537-545
- Solé-Cava AM, Thorpe JP (1991) High levels of genetic variation in natural populations of marine lower invertebrates. *Biol J Linnean Soc* 44:65-80
- Sorensen JG (2003) The evolutionary and ecological role of heat shock proteins. *Ecology Letters* 6:1025-1037
- Speer KG, Marshall J (1995) The growth of convective plumes at seafloor hot springs. *J Mar Res* 53:1025-1057
- Stephan W, Xing L, Kirby DA, Braverman JM (1998) A test of the background selection hypothesis based on nucleotide data from *Drosophila ananassae*. *Proc Natl Acad Sci U S A* 95:5649-5654

Strathmann RR, Kendall LR, Marsh AG (2006) Embryonic and larval development of a cold adapted Antarctic ascidian. *Polar Biol* 29:495-501

- T -

- Tajima F (1989a) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585-595
- Tajima F (1989b) The effect of change in population size on DNA polymorphism. *Genetics* 123:597-601
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596-1599
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial-DNA in Humans and Chimpanzees. *Mol Biol Evol* 10:512-526
- Tang J, Toè L, Back C, Unnasch TR (1996) Intra-specific heterogeneity of the rDNA internal transcribed spacer in the *Simulium damnosum* (Diptera: Simuliidae) complex. *Mol Biol Evol* 13:244-252
- Tanguy A, Bierre N, Saavedra C, Pina B, Bachere E, Kube M, Bazin E, Bonhomme F, Boudry P, Boulo V, Boutet I, Cancela L, Dossat C, Favrel P, Huvet A, Jarque S, Jollivet D, Klages S, Lapegue S, Leite R, Moal J, Moraga D, Reinhardt R, Samain JF, Zouros E, Canario A (2008) Increasing genomic information in bivalves through new EST collections in four species: Development of new genetic markers for environmental studies and genome evolution. *Gene* 408:27-36
- Thiébaud E, Huther X, Shillito B, Jollivet D, Gaill F (2002) Spatial and temporal variations of recruitment in the tube worm *Riftia pachyptila* on the East Pacific Rise (9°50'N and 13°N). *Marine Ecology Progress Series*:234, 147-157
- Thiriou-Quievreux C (2003) Advances in chromosomal studies of gastropod molluscs. *J Molluscan Stud* 69:187-202
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673-4680
- Thorson G (1950) Reproductive and Larval Ecology of Marine Bottom Invertebrates. *Biol Rev Cambridge Philosophic Soc* 25:1-45
- Tissiere A, Mitchell HK, Tracy UM (1974) Protein-synthesis in salivary-glands of *Drosophila melanogaster*: relation to chromosome puffs. *J Mol Biol* 84:389-&
- Toulmond A (1990) Dioxygen binding by the extracellular hemoglobins of two deep-sea hydrothermal vent annelids, *Alvinella pompejana* and *Alvinella caudata*. In: Préaux G, Lontie R (eds) *Invertebrate Dioxygen Carriers*. Leuven University Press, Leuven, p 359-362
- Tunnicliffe V (1991) The biology of hydrothermal vents: ecology and evolution. *Oceanography and Marine Biology Annual Review* 29:319-407
- Tunnicliffe V, Botros M, de Burgh ME, Dinet A, Johnson HP, Juniper SK, McDuff RE (1986) Hydrothermal vents of Explorer Ridge, northeast Pacific. *Deep-Sea Research* 33:401-412
- Tunnicliffe V, Desbruyères D, Jollivet D, Laubier L (1993) Systematic and ecological characteristics of *Paralvinella sulfincola* Desbruyères and Laubier, a new polychaete (family Alvinellidae) from northeast Pacific hydrothermal vents. *Canadian Journal of Zoology* 71:286-297

- Tunnicliffe V, Embley RW, Holden JF, Butterfield DA, Massoth GJ, Juniper SK (1997) Biological colonization of new hydrothermal vents following an eruption on Juan de Fuca Ridge. *Deep-Sea Research I* 44:1627-1654
- Tunnicliffe V, Fowler CMR (1996) Influence of sea-floor spreading on the global hydrothermal fauna. *Nature* 379:531-533
- Tunnicliffe V, Juniper SK, de Burgh ME (1985) The hydrothermal vent community on Axial Seamount, Juan de Fuca Ridge. *Bulletin of the Biological Society of Washington* 6:453-464
- Tyler PA, Young CM (1999) Reproduction and dispersal at vents and cold seeps. *J Mar Biol Assoc UK* 79:193-208

- V -

- Van Dover CL, Desbruyères D, Segonzac M, Comtet T, Saldanha L, Fiala-Médioni A, Langmuir C (1996) Biology of the Lucky Strike hydrothermal field. *Deep-Sea Research I* 43:1509-1529
- Van Dover CL, German CR, Speer KG, Parson LM, Vrijenhoek RC (2002) Marine biology - Evolution and biogeography of deep-sea vent and seep invertebrates. *Science* 295:1253-1257
- Van Dover CL, Humphris SE, Fornari D, Cavanaugh CM, Collier R, Goffredi SK, Hashimoto J, Lilley MD, Reysenbach AL, Shank T, Von Damm KL, Banta A, Gallant RM, Götz D, Green D, Hall J, Harmer TL, Hurtado LA, Johnson P, McKiness ZP, Meredith C, Olson E, Pan IL, Turniseed M, Won Y, Young CR, Vrijenhoek RC (2001) Biogeography and ecological setting of Indian Ocean hydrothermal vents. *Science* 294:810-823
- Vetter RD, Powell MA, Somero GN (1991) Metazoan adaptations to hydrogen sulphide. In: Bryant C (ed) *Metazoan life without oxygen*. Chapman and Hall, London, p 109-128
- Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol Evol* 16:381-390
- Von Cosel R (2002) A new species of bathymodioline mussel (Mollusca, Bivalvia, Mytilidae) from Mauritania (West Africa), with comments on the genus *Bathymodiolus* Kenk & Wilson, 1985. *zoosystema* 24:259-271
- Von Cosel R, Comtet T, Krylova EM (1999) *Bathymodiolus* (Bivalvia : Mytilidae) from hydrothermal vents on the Azores Triple Junction and the Logatchev hydrothermal field, Mid-Atlantic Ridge. *Veliger* 42:218-248
- Von Cosel R, Marshall BA (2003) Two new species of large mussels (Bivalvia : Mytilidae) from active submarine volcanoes and a cold seep off the eastern North Island of New Zealand, with description of a new genus. *Nautilus* 117:31-46
- Von Cosel R, Metivier B, Hashimoto J (1994) Three new species of *Bathymodiolus* (Bivalvia, Mytilidae) from hydrothermal vents in the Lau Basin and the North Fiji Basin, Western Pacific, and the Snake Pit Area, Mid-Atlantic Ridge. *Veliger* 37:374-392
- Von Cosel R, Olu K (1998) Gigantism in Mytilidae. A new *Bathymodiolus* from cold seep areas on the Barbados accretionary Prism. *C R Acad Sci Paris Sci de la Vie* 321:655-663
- Vrijenhoek RC (1997) Gene flow and genetic diversity in naturally fragmented metapopulations of deep-sea hydrothermal vent animals. *J Hered* 88:285-293
- Vrijenhoek RC, Schutz SJ, Gustafson RG, Lutz RA (1994) Cryptic Species of Deep-Sea Clams (Mollusca, Bivalvia, Vesicomidae) from Hydrothermal Vent and Cold-Water Seep Environments. *Deep-Sea Res Part I-Oceanogr Res Pap* 41:1171-1189

- W -

- Wade MJ, McCauley DE (1988) Extinction and recolonization: their effects on the genetic differentiation of local-populations. *Evolution* 42:995-1005
- Wakeley J, Hey J (1997) Estimating ancestral population parameters. *Genetics* 145:847-855
- Watremez P, Kervévan C (1990) Origine des variations de l'activité hydrothermale : premiers éléments de réponse d'un modèle numérique simple. *Comptes Rendus de l'Académie des Sciences de Paris, Série II* 311:153-158
- Watterson GA (1975) Number of Segregating Sites in Genetic Models without Recombination. *Theor Popul Biol* 7:256-276
- Watterson GA (1983) On the Time for Gene Silencing at Duplicate Loci. *Genetics* 105:745-766
- Weinshilboum RM, Otterness DM, Aksoy IA, Wood TC, Her C, Raftogianis RB (1997) Sulfotransferase molecular biology: cDNAs and genes. *Faseb J* 11:3-14
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358-1370
- Whitlock MC (1992) Nonequilibrium population-structure in forked Fungus Beetles - extinction, colonization, and the genetic variance among populations. *Am Nat* 139:952-970
- Whitlock MC, McCauley DE (1990) Some population genetic consequences of colony formation and extinction - genetic correlations within founding groups. *Evolution* 44:1717-1724
- Wiens JJ, Graham CH (2005) Niche conservatism: Integrating evolution, ecology, and conservation biology. *Annu Rev Ecol Evol Syst* 36:519-539
- Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *J Evol Biol* 14:611-619
- Won Y, Hallam SJ, O'Mullan GD, Vrijenhoek RC (2003a) Cytonuclear disequilibrium in a hybrid zone involving deep-sea hydrothermal vent mussels of the genus *Bathymodiolus*. *Mol Ecol* 12:3185-3190
- Won Y, Young CR, Lutz RA, Vrijenhoek RC (2003b) Dispersal barriers and isolation among deep-sea mussel populations (Mytilidae : *Bathymodiolus*) from eastern Pacific hydrothermal vents. *Mol Ecol* 12:169-184
- Won YJ, Hallam SJ, O'Mullan GD, Pan IL, Buck KR, Vrijenhoek RC (2003c) Environmental acquisition of thiotrophic endosymbionts by deep-sea mussels of the genus *Bathymodiolus*. *Appl Environ Microbiol* 69:6785-6792
- Won YJ, Hey J (2005) Divergence population genetics of chimpanzees. *Mol Biol Evol* 22:297-307
- Won YJ, Jones WJ, Vrijenhoek RC (2008) Absence of cospeciation between deep-sea Mytilids and their thiotrophic endosymbionts. *J Shellfish Res* 27:129-138
- Won YJ, Maas PAY, Van Dover CL, Vrijenhoek RC (2002) Habitat reversal in vent and seep mussels: seep species, *Bathymodiolus heckerae*, derived from vent ancestors. *Cah Biol Mar* 43:387-390
- Wright S (1931) Evolution in mendelian populations. *Genetics* 16:97-159
- Wright S (1943) Isolation by distance. *Genetics* 28:114-138
- Wright S (1951) The Genetical Structure of Populations. *Annals of Eugenics* 15:323-354
- Wright S (1965) The interpretation of population structure by F-statistics with special reference to systems of mating. *Evolution* 19:395-420
- Wright S (1977) Evolution and the Genetics of populations. *Experimental Results and Evolutionary Deductions* University of Chicago Press Chicago Vol. 3
- Wu CI, Ting CT (2004) Genes and speciation. *Nat Rev Genet* 5:114-122

- Y -

- Yang Z (2000) Phylogenetic analysis by maximum likelihood (PAML). Version 3.0. University College London, London, England.
- Yang ZH (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555-556
- You CF, Butterfield DA, Spivack AJ, Gieskes JM, Gamo T, Campbell AJ (1994) Boron and Halide Systematics in Submarine Hydrothermal Systems - Effects of Phase-Separation and Sedimentary Contributions. *Earth Planet Sci Lett* 123:227-238
- Young CM, Vazquez E, Metaxas A, Tyler PA (1996) Embryology of vestimentiferan tube worms from deep-sea methane/sulphide seeps. *Nature* 381:514-516
- Young ND, dePamphilis CW (2000) Purifying selection Detected in the plastid Gene *matK* and Flanking Ribozyme Regions Within a Group II Intron of Nonphotosynthetic Plants. *Mol Biol Evol* 17(12):1933-1941

- Z -

- Zhang JZ (2003) Evolution by gene duplication: an update. *Trends Ecol Evol* 18:292-298
- Zhang JZ, Rosenberg HF, Nei M (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci U S A* 95:3708-3713

Annexe 1 : Protocoles d'extraction

Protocole extraction d'ADN au Phénol/Chloroforme

_ (penser à faire chauffer bain à sec à 56°C)

Prélever des tissus de petite taille (moitié taille de l'ongle de l'auriculaire !!)

Réduire les gros morceaux musculueux en les coupant (pieds...)

[Faire un rinçage préalable pour les échantillons conservés dans du BLB :

Mettre les morceaux dans tubes avec 400µl de TpPK

Bien vortexer

Eliminer le surnageant

Répéter l'opération si nécessaire puis commencer digestion]

Ajouter 800µl de Tp PK

Ajouter 20µl de PK 10 mg/ml (ou 10 µl de PK 20 mg/ml)

Vortexer

Bain à sec à 56°C pour digestion des tissus

Ajout 1µl de RNase A (10 mg/ml) par tube et remettre à incuber pendant 1h à 37°C

Ajouter 330 µl de NaCl (5M)

Vortexer

Centrifuger 5000 G pendant 4 min

Récupération du surnageant et mettre dans tube 2ml

_ (Réaliser une solution Phénol / Chloroforme à 1/1)

Sous hotte : Compléter à volume 1/1 les tubes de solution d'ADN avec la solution Phénol/Chlo.
(ou remplir jusqu'à ras bord...)

Mélanger vigoureusement pour former une émulsion (passage protéines dans le Phénol/Chlo)

Centrifuger à 5000G pendant 10 min

Prélever attentivement le surnageant et le transférer dans une nouvelle série de tubes (2 ml) (ou diviser le volume pour le répartir dans 2 séries de tubes)

Précipiter l'ADN par ajout de EtOH (100% à -20°C), compléter jusqu'à 2 vol d'EtOH pour 1 de solution avec ADN. (ou remplir jusqu'à ras bord...)

Laisser passer une nuit au congélateur -20°C.

...

Le lendemain... (Ou quelques heures plus tard quand jolie méduse)

...

_ refroidir les centrifugeuses

_mettre bain à sec à 37°C

Centrifugation des tubes à 12000G pendant 20 min

vider alcool (en faisant attention au culot !)

Ajouter 500µl d'EtOH 70% (pour éliminer les derniers ions)

Vortexer pendant une minute

Centrifuger à 12000t/min pendant 4 min

Vider l'alcool

Bien faire attention à éliminer la dernière goutte (retourner le tube sur papier)

Sécher au SpeedVac (niveau moyen pendant 3 min)

Récupération du culot dans - 100µl TE 1X (jusqu'à 250µl si beaucoup d'ADN) possibilité utilisation H₂O autoclavée

Secouer doucement les tubes et les placer dans bain à sec à 37°C pour favoriser la remise en solution de l'ADN.

Annexe 1 : Protocoles d'extraction

Protocole extraction d'ADN au Chelex

Réaliser une solution de Chelex à 5% (5 mg de billes de Chelex dans 95ml de TE1X)

Penser à faire chauffer le bain marie à sec à 55°C

Mettre 400µl de solution de Chelex 5% par tube (1,5ml)

Ajouter 25µl de protéinase K 10mg/ml

#Ajouter un petit morceau de tissu (1 mm³)

Vortexer

Centrifuger 3 minutes

4 à 12 h au bain mair à 55°C

Chauffer à 95°C pendant 10 à 30 min (dénaturation PK)

Vortexer

Centrifuger brièvement

Conserver à -20°C

Protocole extraction d'ADN au CTAB

Composition du tampon CTAB =

CTAB 2%

PVP 1%

NaCl 1,4M

b-mercaptoethanol 0,2%

EDTA 20 mM

Tris HCl pH 8.0 100 mM

Protéinase K 0,1 mg/mL

H₂O

1) tissu frais = broyage dans l'azote

2) tissu alcoolisé = coupé finement en morceaux et épongé le surplus d'alcool avec KimWipe

Préchauffer la solution de CTAB à 60°C

Ajouter 600 microlitres de solution CTAB

Incuber à 60°C pendant 4 heures (possibilité sur la nuit)

Centrifuger et transférer le surnageant dans un tube 2 mL

Ajouter un volume égal de CIA (Chloroforme + Alcool isoamylique, 96:4) et bien mélanger par renversements pendant 3 minutes

Centrifuger 15 000 g pendant 10 minutes à 4°C

Récupérer surnageant et ajouter 1 microlitre RNase A

Incuber à 37°C pendant 1 heure

Ajouter 2/3 de volume d'isopropanol glacé

Laisser précipiter pendant 1 heure à -20°C (ou la nuit)

Centrifuger 10 minutes à 15 000 g

Laver le culot avec 500 microlitres d'éthanol 75%

Centrifuger 10 minutes à 15 000 g

Sécher le culot (3 minutes au SpeedVac) et reprendre dans 50 à 150 microlitres d'H₂O ou TE

en incubant à 37°C pour bien resuspendre l'ADN

Annexe 2 : Protocoles de clonage

LE protocole du clonage réussi

Ligation Kit pGEM-TVector System I

Mélanger dans un tube :

3.3 µl de 2X Rapid ligation Buffer

2 µl ADN purifié

0.6 µl Vecteur pGEM-T Easy

0.6 µl T4 DNA Ligase 3U/µl (Très fragile, ne sortir que pour prélever le volume désiré)

Centrifuger rapidement

Mettre au réfrigérateur pour la nuit.

Protocole de transformation bactéries MAX efficiency DH5α T1 phage Resistant Competent Cells (Invitrogen) (compétentes maison)

0 : mettre bain marie à 42°C et SOC à température pièce

1 : faire décongeler les bactéries compétentes sur glace. Placer également le nombre de tubes Eppendorff nécessaire sur glace

2 : agiter doucement les bactéries compétentes et en **aliquoter 100 µl** dans chacun des tubes.

3 : recongeler les bactéries non utilisées dans un bain de carboglace/éthanol pendant 5 min avant de les remettre au congélateur à -70°C.

4 : Ajouter **1 µl** de ligation (1 à 4 ng ADN) en agitant la pipette dans les bactéries. Taper doucement les tubes pour mélanger.

5 : laisser incuber les bactéries **sur glace pendant 30 min**

6 : Choc thermique de **45 secondes dans un bain à 42°C** ; ne pas secouer.

7 : placer sur **glace pour 2 minutes**.

8 : Ajouter **0.9 ml de SOC à température de la pièce**.

9 : Agiter à **225 rpm (37°C)** pendant une heure

10 : Appliquer sur boîte de culture (LB + ampi) : **20µl de X-Gal 10% (=100mg/ml) + 40µl IPTG (0,2M)** (ou **ajouter 6µl d'IPTG** directement dans tube contenant les bactéries) puis laisser reposer 10 min à 37°C

11 : Appliquer bactéries en répartissant sur 2 boîtes :

-**125µl** de la solution bactérienne

-**250µl** de la solution bactérienne

12 : Laisser à **37°C pendant 10 min**

13 : incuber la nuit à 37°C

Annexe 3 : Récapitulatif des séquences de CO1 utilisées pour la phylogénie des *Bathymodiolinae*

Gène	Numero d'accession / identification	Espèce	localité	Habitat	références
CO1	Perso_950_02	Bathy_thermo_9°50N	EPR_9°50N	Hydro	perso
CO1	perso_950_21	Bathy_thermo_9°50N	EPR_9°50N	Hydro	perso
CO1	Perso_950_28	Bathy_thermo_9°50N	EPR_9°50N	Hydro	perso
CO1	Perso_950_34	Bathy_thermo_9°50N	EPR_9°50N	Hydro	perso
CO1	Perso_7S_03	Bathy_thermo_7S	EPR_7°S	Hydro	perso
CO1	Perso_7S_05	Bathy_thermo_7S	EPR_7°S	Hydro	perso
CO1	Perso_7S_07	Bathy_thermo_7S	EPR_7°S	Hydro	perso
CO1	Perso_7S_08	Bathy_thermo_7S	EPR_7°S	Hydro	perso
CO1	Perso_7S_19	Bathy_thermo_7S	EPR_7°S	Hydro	perso
CO1	Perso_7S_27	Bathy_thermo_7S	EPR_7°S	Hydro	perso
CO1	Perso_7S_36	Bathy_thermo_7S	EPR_7°S	Hydro	perso
CO1	Perso_7S_55	Bathy_thermo_7S	EPR_7°S	Hydro	perso
CO1	Perso_Bt21S_01	Bathy_thermo_21S	EPR_21°S	Hydro	perso
CO1	Perso_Bt21S_15	Bathy_thermo_21S	EPR_21°S	Hydro	perso
CO1	Perso_Bt21S_16	Bathy_thermo_21S	EPR_21°S	Hydro	perso
CO1	Perso_Bt21S_19	Bathy_thermo_21S	EPR_21°S	Hydro	perso
CO1	Perso_Bt21S_20	Bathy_thermo_21S	EPR_21°S	Hydro	perso
CO1	AB257557	Bathy_thermophilus	EPR_9.51 N 104.18 W	Hydro	Fujita unpublished
CO1	AF456308	Bathy_thermophilus	EPR_Bt27	Hydro	Won et al 2003
CO1	AF456307	Bathy_thermophilus	EPR_Bt26	Hydro	Won et al 2003
CO1	AF456306	Bathy_thermophilus	EPR_Bt25	Hydro	Won et al 2003
CO1	AF456305	Bathy_thermophilus	EPR_Bt24	Hydro	Won et al 2003
CO1	AF456304	Bathy_thermophilus	EPR_Bt23	Hydro	Won et al 2003
CO1	AF456303	Bathy_thermophilus	EPR_7S_Bt22	Hydro	Won et al 2003
CO1	AF456302	Bathy_thermophilus	EPR_Bt21	Hydro	Won et al 2003
CO1	AF456301	Bathy_thermophilus	EPR_Bt20	Hydro	Won et al 2003
CO1	AF456300	Bathy_thermophilus	EPR_Bt19	Hydro	Won et al 2003
CO1	AF456299	Bathy_thermophilus	EPR_Bt18	Hydro	Won et al 2003
CO1	AF456298	Bathy_thermophilus	EPR_Bt17	Hydro	Won et al 2003
CO1	AF456297	Bathy_thermophilus	EPR_Bt16	Hydro	Won et al 2003
CO1	AF456296	Bathy_thermophilus	EPR_Bt15	Hydro	Won et al 2003
CO1	AF456295	Bathy_thermophilus	EPR_Bt14	Hydro	Won et al 2003
CO1	AF456294	Bathy_thermophilus	EPR_Bt13	Hydro	Won et al 2003
CO1	AF456293	Bathy_thermophilus	EPR_Bt12	Hydro	Won et al 2003
CO1	AF456292	Bathy_thermophilus	EPR_Bt11	Hydro	Won et al 2003
CO1	AF456291	Bathy_thermophilus	EPR_Bt10	Hydro	Won et al 2003
CO1	AF456290	Bathy_thermophilus	EPR_Bt9	Hydro	Won et al 2003
CO1	AF456289	Bathy_thermophilus	EPR_Bt8	Hydro	Won et al 2003
CO1	AF456288	Bathy_thermophilus	EPR_Bt7	Hydro	Won et al 2003
CO1	AF456287	Bathy_thermophilus	EPR_Bt6	Hydro	Won et al 2003
CO1	AF456286	Bathy_thermophilus	EPR_Bt5	Hydro	Won et al 2003
CO1	AF456285	Bathy_thermophilus	EPR_9°N_Bt4	Hydro	Won et al 2003
CO1	AF456284	Bathy_thermophilus	EPR_Bt3	Hydro	Won et al 2003
CO1	AF456283	Bathy_thermophilus	EPR_Bt2	Hydro	Won et al 2003
CO1	AF456282	Bathy_thermophilus	EPR_18S_animal Farm Bt1	Hydro	Won et al 2003
CO1	AF456322	Bathy_sp._RCV-2003	EPR_Bt41	Hydro	Won et al 2003
CO1	AF456321	Bathy_sp._RCV-2003	EPR_Bt40	Hydro	Won et al 2003
CO1	AF456320	Bathy_sp._RCV-2003	EPR_Bt39	Hydro	Won et al 2003
CO1	AF456319	Bathy_sp._RCV-2003	EPR_Bt38	Hydro	Won et al 2003
CO1	AF456318	Bathy_sp._RCV-2003	EPR_Bt37	Hydro	Won et al 2003

Annexe 3 : Récapitulatif des séquences de CO1 utilisées pour la phylogénie des *Bathymodiolinae*

Gène	Numero d'accession / identification	Espèce	localité	Habitat	références
CO1	AF456317	Bathy_sp._RCV-2003	EPR_32S_Bt36	Hydro	Won et al 2003
CO1	AF456316	Bathy_sp._RCV-2003	EPR_Bt35	Hydro	Won et al 2003
CO1	AF456315	Bathy_sp._RCV-2003	EPR_Bt34	Hydro	Won et al 2003
CO1	AF456314	Bathy_sp._RCV-2003	EPR_Bt33	Hydro	Won et al 2003
CO1	AF456313	Bathy_sp._RCV-2003	EPR_Bt32	Hydro	Won et al 2003
CO1	AF456312	Bathy_sp._RCV-2003	EPR_Bt31	Hydro	Won et al 2003
CO1	AF456311	Bathy_sp._RCV-2003	EPR_Bt30	Hydro	Won et al 2003
CO1	AF456310	Bathy_sp._RCV-2003	EPR_Bt29	Hydro	Won et al 2003
CO1	AF456309	Bathy_sp._RCV-2003	EPR_Bt28	Hydro	Won et al 2003
CO1	perso	Bathy_sp._37S	sEPR_37°S	Hydro	Perso
CO1	perso	Bathy_sp._37S	sEPR_37°S	Hydro	Perso
CO1	perso	Bathy costa rica	seep Costa Rica	Suint.	Perso
CO1	perso	Bathy costa rica	seep Costa Rica	Suint.	Perso
CO1	AB170046	Bathy_septemdierum	Mariana Back-Arc Basin	Hydro	Iwasaki 2006
CO1	AB170041	Bathy_septemdierum	Suiyo Seamount, Izu-Ogasawara Island Arc_haplo SS4	Hydro	Iwasaki 2006
CO1	AB101430	Bathy_septemdierum	Suiyo Seamount, Izu-Ogasawara Island Arc_haplo SS3	Hydro	Iwasaki 2006
CO1	AB101429	Bathy_septemdierum	Suiyo Seamount, Izu-Ogasawara Island Arc_haplo SS2	Hydro	Iwasaki 2006
CO1	AB101428	Bathy_septemdierum	Suiyo Seamount, Izu-Ogasawara Island Arc_haplo SS1	Hydro	Iwasaki 2006
CO1	AB101427	Bathy_septemdierum	Myojin Knoll, Izu-Ogasawara Island Arc_haplo SM4	Hydro	Iwasaki 2006
CO1	AB101426	Bathy_septemdierum	Myojin Knoll, Izu-Ogasawara Island Arc_haplo SM3	Hydro	Iwasaki 2006
CO1	AB101425	Bathy_septemdierum	Myojin Knoll, Izu-Ogasawara Island Arc_haplo SM2	Hydro	Iwasaki 2006
CO1	AB101424	Bathy_septemdierum	Myojin Knoll, Izu-Ogasawara Island Arc_haplo SM1	Hydro	Iwasaki 2006
CO1	AB257535	Bathy_sp._EF5	Japan:Northwest Eifuku Seamount_21.29 N 144.02 E_haplo EF5	Hydro	Fujita unpublished
CO1	AB257534	Bathy_sp._EF3	Japan:Northwest Eifuku Seamount_21.29 N 144.02 E_haplo EF3	Hydro	Fujita unpublished
CO1	AB257542	Bathy_sp._Lau5_sp_B	Lau Back Arc basin_23.32 S 176.43 W_haplo Lau 5	Suint.	Fujita unpublished
CO1	AY649799	Bathy_brevior	Mariana Through	Hydro	Jones et al 2006
CO1	AY275544	Bathy_brevior	Lau Back Arc basin W pac	Hydro	Smith et al 2005
CO1	AY275543	Bathy_aff brevior (<i>marisindicus</i> ??)	Kairei vent_Central Indian Ridge	Hydro	Smith unpublished
CO1	AB170045	Bathy_marisindicus	Kairei Field_ Southern Central Indian Ridge	Hydro	Iwasaki 2006
CO1	AB170044	Bathy_marisindicus	Kairei Field_ Southern Central Indian Ridge	Hydro	Iwasaki 2006
CO1	AB170043	Bathy_marisindicus	Kairei Field_ Southern Central Indian Ridge	Hydro	Iwasaki 2006
CO1	AB170042	Bathy_marisindicus	Kairei Field_ Southern Central Indian Ridge	Hydro	Iwasaki 2006
CO1	AY649798	Bathy_brooksi	West Florida Escarpment	Suint.	Jones et al 2006
CO1	AY649797	Bathy_brooksi	Alaminos Canyon	Suint.	Jones et al 2006
CO1	EF051244	Bathy_brooksi	Alaminos Canyon	Suint.	Cordes 2007
CO1	DQ513449	Bathy_boomerang	Barbados prism Orenoque B	Suint.	Olu et al 2007
CO1	DQ513448	Bathy_boomerang	Barbados prism Orenoque B	Suint.	Olu et al 2007
CO1	DQ513446	Bathy_boomerang	Barbados prism Orenoque B	Suint.	Olu et al 2007
CO1	DQ513444	Bathy_boomerang	Barbados prism Orenoque B	Suint.	Olu et al 2007
CO1	DQ513451	Bathy_aff._boomerang or <i>Bathy sp. A</i>	Congo-Angola margin	Suint.	Olu et al 2007
CO1	DQ513450	Bathy_aff._boomerang or <i>Bathy sp. A</i>	Congo-Angola margin	Suint.	Olu et al 2007

Annexe 3 : Récapitulatif des séquences de CO1 utilisées pour la phylogénie des *Bathymodiolinae*

Gène	Numero d'accession / identification	Espèce	localité	Habitat	références
CO1	DQ513447	Bathy_aff._boomerang or <i>Bathy sp. A</i>	Congo-Angola margin	Suint.	Olu et al 2007
CO1	DQ513445	Bathy_aff._boomerang or <i>Bathy sp. A</i>	Congo-Angola margin	Suint.	Olu et al 2007
CO1	DQ513443	Bathy_aff._boomerang or <i>Bathy sp. A</i>	Congo-Angola margin	Suint.	Olu et al 2007
CO1	DQ513442	Bathy_aff._boomerang or <i>Bathy sp. A</i>	Congo-Angola margin	Suint.	Olu et al 2007
CO1	EF051243	Bathy_sp._A_Nigerian_seep_long_type	Nigerian slope	Suint.	Cordes 2007
CO1	EF051242	Bathy_sp._A_Nigerian_seep_long_type	Nigerian slope	Suint.	Cordes 2007
CO1	DQ513441	Bathy_heckerae	Florida escarpment	Suint.	Olu et al 2007
CO1	AY649794	Bathy_heckerae	Florida escarpment	Suint.	Jones et al 2006
CO1	AY649793	Bathy_heckerae	Blake Ridge	Suint.	Jones et al 2006
CO1	Perso_Put1	Bathy_puteoserpentis	MAR_Logatchev	Hydro	perso
CO1	Perso_Put2	Bathy_puteoserpentis	MAR_Logatchev	Hydro	perso
CO1	AB170062	Bathy_puteoserpentis	MAR_Snakepit	Hydro	Maas 1999
CO1	AY649796	Bathy_puteoserpentis	MAR_Snakepit	Hydro	Jones et al 2006
CO1	Perso_Bputeo_06	Bathy_puteoserpentis	MAR_Logatchev	Hydro	perso
CO1	Perso_Bputeo_12	Bathy_puteoserpentis	MAR_Logatchev	Hydro	perso
CO1	Perso_Bputeo_25	Bathy_puteoserpentis	MAR_Logatchev	Hydro	perso
CO1	Perso_Bputeo_30	Bathy_puteoserpentis	MAR_Logatchev	Hydro	perso
CO1	AB170061	Bathy_azoricus	haplo AZL2	Hydro	Iwasaki 2006
CO1	AB170060	Bathy_azoricus	haplo AZL1	Hydro	Iwasaki 2006
CO1	AY649795	Bathy_azoricus	MAR_Menez Gwen	Hydro	Jones et al 2006
CO1	Perso_Rb_01	Bathy_azoricus	MAR_Rainbow	Hydro	perso
CO1	Perso_Rb_03	Bathy_azoricus	MAR_Rainbow	Hydro	perso
CO1	Perso_Rb_06	Bathy_azoricus	MAR_Rainbow	Hydro	perso
CO1	Perso_Rb_12	Bathy_azoricus	MAR_Rainbow	Hydro	perso
CO1	Perso_Rb_13	Bathy_azoricus	MAR_Rainbow	Hydro	perso
CO1	Perso_Rb_28	Bathy_azoricus	MAR_Rainbow	Hydro	perso
CO1	Perso_MG_12	Bathy_azoricus	MAR_Menez Gwen	Hydro	perso
CO1	Perso_MG_14	Bathy_azoricus	MAR_Menez Gwen	Hydro	perso
CO1	Perso_MG_15	Bathy_azoricus	MAR_Menez Gwen	Hydro	perso
CO1	Perso_MG_16	Bathy_azoricus	MAR_Menez Gwen	Hydro	perso
CO1	Perso_MG_30	Bathy_azoricus	MAR_Menez Gwen	Hydro	perso
CO1	Perso_MG_33	Bathy_azoricus	MAR_Menez Gwen	Hydro	perso
CO1	Perso_MG_35	Bathy_azoricus	MAR_Menez Gwen	Hydro	perso
CO1	Perso_BaMG_MoM_23	Bathy_azoricus	MAR_Menez Gwen	Hydro	perso
CO1	Perso_BaMG_MoM_28	Bathy_azoricus	MAR_Menez Gwen	Hydro	perso
CO1	Perso_BaMG_MoM_40	Bathy_azoricus	MAR_Menez Gwen	Hydro	perso
CO1	AB257543	Bathy_sp._Lau8_sp_A	Lau Back Arc Basin_23.32 S 176.43 W_haplo Lau8	Suint.	Fujita unpublished
CO1	AB257541	Bathy_sp._Lau4_sp_A	Lau Back Arc Basin_22.32 S 176.43 W_haplo Lau4	Suint.	Fujita unpublished
CO1	AB257540	Bathy_sp._Lau3_sp_A	Lau Back Arc Basin_22.32 S 176.43 W_haplo Lau3	Suint.	Fujita unpublished
CO1	AB257539	Bathy_sp._Lau1_sp_A	Lau Back Arc Basin_23.32 S 176.43 W_haplo Lau1	Suint.	Fujita unpublished
CO1	AB255743	Bathy_sp._NewZ	New Zealand:Off New Zealand_Ne5	Suint.	Fujita NZ unpublished
CO1	AB255742	Bathy_sp._NewZ	New Zealand:Off New Zealand_Ne4	Suint.	Fujita NZ unpublished
CO1	AB255741	Bathy_sp._NewZ	New Zealand:Off New Zealand_Ne3	Suint.	Fujita NZ unpublished
CO1	AB255740	Bathy_sp._NewZ	New Zealand:Off New Zealand_Ne2	Suint.	Fujita NZ unpublished
CO1	AB255739	Bathy_sp._NewZ	New Zealand:Off New Zealand_Ne1	Suint.	Fujita NZ unpublished
CO1	AB101434	Bathy_sp._manusensis	Papua New Guinea:Manus Basin, PACKMANUS FIELD_haplo 5	Hydro	Miyazaki 2004

Annexe 3 : Récapitulatif des séquences de CO1 utilisées pour la phylogénie des *Bathymodiolinae*

Gène	Numero d'accession / identification	Espèce	localité	Habitat	références
CO1	AB101433	Bathy_sp._manusensis	Papua New Guinea:Manus Basin, PACKMANUS FIELD_haplo 3-4	Hydro	Miyazaki 2004
CO1	AB101432	Bathy_sp._manusensis	Papua New Guinea:Manus Basin, PACKMANUS FIELD_haplo 2	Hydro	Miyazaki 2004
CO1	AB101431	Bathy_sp._manusensis	Papua New Guinea:Manus Basin, PACKMANUS FIELD_haplo 1	Hydro	Miyazaki 2004
CO1	AB170059	Bathy_aduloides	Japan:Off Kikaijima Island_haplo AK3	Hydro	Iwasaki 2006
CO1	AB170058	Bathy_aduloides	Japan:Off Kikaijima Island_haplo AK5	Hydro	Iwasaki 2006
CO1	AB170057	Bathy_aduloides	Japan:Off Kikaijima Island_haplo AK4	Hydro	Iwasaki 2006
CO1	AB170056	Bathy_aduloides	Japan:Off Kikaijima Island_haplo AK2	Hydro	Iwasaki 2006
CO1	AB170055	Bathy_aduloides	Japan:Off Kikaijima Island_haplo AK1	Hydro	Iwasaki 2006
CO1	AB170054	Bathy_aduloides	Japan:Iheya Ridge, Mid-Okinawa Trough_haploAI1	Hydro	Iwasaki 2006
CO1	AB257523	Adipi_iwaotakii	Japan:Ibaraki, Off Nakaminato_haplo_AIH5	Suint.	Fujita Evo Pro unpub
CO1	AB257522	Adipi_iwaotakii	Japan:Ibaraki, Off Nakaminato_haplo_AIH3	Suint.	Fujita Evo Pro unpub
CO1	AB257521	Adipi_iwaotakii	Japan:Ibaraki, Off Nakaminato_haplo_AIH2	Suint.	Fujita Evo Pro unpub
CO1	AB257520	Adipi_iwaotakii	Japan:Ibaraki, Off Nakaminato_haplo_AIH1	Suint.	Fujita Evo Pro unpub
CO1	AY649804	Idas_macdonaldi	Garden Banks (Atlantic)	Suint.	Jones et al 2006
CO1	AY275546	Idas_washingtonia	Monterey Bay, E. pacific	Bois coulés / os de baleine	Smith unpublished NZ
CO1	AB257519	Adipi_crypta	Japan:Kagoshima, Off Noma Cape_ACN328-1	Os de baleine	Fujita Evo Pro unpub
CO1	AB257518	Adipi_crypta	Japan:Kagoshima, Off Noma Cape_ACN4	Os de baleine	Fujita Evo Pro unpub
CO1	AB257517	Adipi_crypta	Japan:Kagoshima, Off Noma Cape_ACN3	Os de baleine	Fujita Evo Pro unpub
CO1	AB257516	Adipi_crypta	Japan:Kagoshima, Off Noma Cape_ACN2	Os de baleine	Fujita Evo Pro unpub
CO1	AB257515	Adipi_crypta	Japan:Kagoshima, Off Noma Cape_ACN1	Os de baleine	Fujita Evo Pro unpub
CO1	AY649802	Gigant_gladius	Rumble III (W pacific)	Hydro	Jones et al 2006
CO1	AY608438	Gigant_gladius	RVf	Hydro	Smith 2004
CO1	AY608437	Gigant_gladius	RVd	Hydro	Smith 2004
CO1	AY608436	Gigant_gladius	RIIIsb	Hydro	Smith 2004
CO1	AY608435	Gigant_gladius	RIIIb	Hydro	Smith 2004
CO1	AY608434	Gigant_gladius	Mclong	Hydro	Smith 2004
CO1	AY608433	Gigant_gladius	RVg	Hydro	Smith 2004
CO1	AY608432	Gigant_gladius	RVc	Hydro	Smith 2004
CO1	AY608431	Gigant_gladius	RIIIsc	Hydro	Smith 2004
CO1	AY608430	Gigant_gladius	Rve	Hydro	Smith 2004
CO1	AY608429	Gigant_gladius	RVb	Hydro	Smith 2004
CO1	AY608428	Gigant_gladius	Rva	Hydro	Smith 2004
CO1	AY608427	Gigant_gladius	RIIIsa	Hydro	Smith 2004
CO1	AY608426	Gigant_gladius	RIIIa	Hydro	Smith 2004
CO1	AB257525	Gigant_sp._Aitape2	Papua New Guinea_2.55 S 142.14 E_haplo Aitape2	Hydro	Fujita Evo Pro unpub
CO1	AB257524	Gigant_sp._Aitape1	Papua New Guinea_2.55 S 142.14 E_haplo Aitape1	Hydro	Fujita Evo Pro unpub
CO1	AB257529	Gigant_sp._Ashizuri	Japan:Kochi, Off Ashizuri Cape_haplo Ashizuri	?	Fujita Evo Pro unpub

Annexe 3 : Récapitulatif des séquences de CO1 utilisées pour la phylogénie des *Bathymodiolinae*

Gène	Numero d'accession / identification	Espèce	localité	Habitat	références
CO1	AB257555	Gigant_sp._Su4	Japan:Sumisu Caldera_31.28 N 140.04 E_haploSu4	?	Fujita Evo Pro unpub
CO1	AB257554	Gigant_sp. Su2	Japan:Sumisu Caldera_31.28 N 140.04 E_haploSu2	?	Fujita Evo Pro unpub
CO1	AB257553	Gigant_sp. Su1	Japan:Sumisu Caldera_31.28 N 140.04 E_haploSu1	?	Fujita Evo Pro unpub
CO1	AB257546	Gigant_sp._NK5	Japan:Nikko Seamount_23.05 N 142.19 E_NK5	?	Fujita Evo Pro unpub
CO1	AB257545	Gigant_sp._NK2	Japan:Nikko Seamount_23.05 N 142.19 E_NK2	?	Fujita Evo Pro unpub
CO1	AB257544	Gigant_sp._NK1	Japan:Nikko Seamount_23.05 N 142.19 E_NK1	?	Fujita Evo Pro unpub
CO1	AB257538	Gigant_horikoshii	Japan:Kaikata Seamount_26.43 N 141.05 E	Hydro	Fujita Evo Pro unpub
CO1	AB170047	Bathy_hirtus	Japan:Kuroshima Knoll, Off Yaeyama Islands	Suint.	Iwasaki 2006
CO1	AB257552	Bathy_sp._(nouveau_clade_papua)_Sp_B	Papua New Guinea:Sissano_2.49 S 142.14 E_Si3-7	Hydro	Fujita Evo Pro unpub
CO1	AB257550	Bathy_sp._(nouveau_clade_papua)_Sp_B	Papua New Guinea:Sissano_2.53 S 142.16 E_haplo Si2-4	Hydro	Fujita Evo Pro unpub
CO1	AB257549	Bathy_sp._(nouveau_clade_papua)_Sp_B	Papua New Guinea:Sissano_2.53 S 142.16 E_haplo Si2-3	Hydro	Fujita Evo Pro unpub
CO1	AB257548	Bathy_sp._(nouveau_clade_papua)_Sp_B	Papua New Guinea:Sissano_2.53 S 142.16 E_haplo Si2-1	Hydro	Fujita Evo Pro unpub
CO1	AB257531	Bathy_sp._(nouveau_clade_papua)_Sp_B	13.47 N 146 E_haplo C2	Hydro	Fujita Evo Pro unpub
CO1	AB257530	Bathy_sp._(nouveau_clade_papua)_Sp_B	13.47 N 146 E_haplo C1	Hydro	Fujita Evo Pro unpub
CO1	AB257556	Bathy_sp._Kikaijima	Japan:Off Kikaijima Island_28.26 N 130.19 E	?	Fujita Evo Pro unpub
CO1	AB257547	Bathy_sp._(nouveau_clade_papua)_Sp_B	Papua New Guinea:Sissano_2.49 S 142.14 E_haplo Si1-1	Hydro	Fujita Evo Pro unpub
CO1	AB101423	Bathy_japonicus	Japan:Kanagawa, off Hatsushima, sagami bay and Mid-Okinawa though , Minami-Ensei Knoll	Hydro / Suintements	Iwasaki 2006
CO1	AB101422	Bathy_japonicus	Japan:Mid-Okinawa though , Minami-Ensei Knoll	Hydro / Suintements	Iwasaki 2006
CO1	AB170053	Bathy_securiformis	Japan:Dai-ni (No2) Atsumi Knoll, Nankai Trough_LA2	Suint.	Iwasaki 2006
CO1	AB170052	Bathy_securiformis	Japan:Dai-ni (No2) Atsumi Knoll, Nankai Trough_LA1	Suint.	Iwasaki 2006
CO1	AB170051	Bathy_securiformis	Japan:Kuroshima Knoll, Off Yaeyama Islands_haplo_LK4	Suint.	Iwasaki 2006
CO1	AB170050	Bathy_securiformis	Japan:Kuroshima Knoll, Off Yaeyama Islands_haplo_LK3	Suint.	Iwasaki 2006
CO1	AB170049	Bathy_securiformis	Japan:Kuroshima Knoll, Off Yaeyama Islands_haplo_LK2,5	Suint.	Iwasaki 2006
CO1	AB170048	Bathy_securiformis	Japan:Kuroshima Knoll, Off Yaeyama Islands_haplo_LK1	Suint.	Iwasaki 2006
CO1	AY608439	Bathymodiolus_tangaroa	Cape Turnagain (W pacific) Papua New	Suint.	Smith 2004
CO1	AB257551	Bathy_sp._(nouveau_clade_papua)_Sp_A	Guinea:Sissano_2.49 S 142.14 E_haplo Si3-3	Suint.	Fujita Evo Pro unpub

Annexe 3 : Récapitulatif des séquences de CO1 utilisées pour la phylogénie des *Bathymodiolinae*

Gène	Numero d'accession / identification	Espèce	localité	Habitat	références
CO1	AB101421	Bathy_platifrons	Japan: off Hatsushima, Sagami Bay et North Iheya Ridge, Mid Okinawa Trough	Hydro / Suintements	Iwasaki 2006
CO1	AB101420	Bathy_platifrons	Japan: off Hatsushima, Sagami Bay	Hydro / Suintements	Iwasaki 2006
CO1	AB101419	Bathy_platifrons	Japan: off Hatsushima, Sagami Bay	Hydro / Suintements	Iwasaki 2006
CO1	DQ513429	Bathy_aff._childressi	Barbados Prism El Pilar	Suint.	Olu et al 2007
CO1	DQ513425	Bathy_aff._childressi ou Bathymodiolus sp B	Barbados prism Orenoque A	Suint.	Olu et al 2007
CO1	EU288174	Bathy_childressi	isolate 3	Suint.	Genio et al 2008
CO1	EU288173	Bathy_childressi	isolate 1	Suint.	Genio et al 2008
CO1	EU288175	Bathy_childressi	isolate 5	Suint.	Genio et al 2008
CO1	AB257533	Bathy_childressi	27.44 N 91.19 W_CHiG2	Suint.	Fujita Evo Pro unpub
CO1	AB257532	Bathy_childressi	27.44 N 91.19 W_CHiG2	Suint.	Fujita Evo Pro unpub
CO1	EF051246	Bathy_childressi	Alaminos Canyon	Suint.	Cordes 2007
CO1	AY649800	Bathy_childressi	Alaminos Canyon	Suint.	Jones et al 2006
CO1	DQ513438	Bathy_aff._childressi	Barbados prism Orenoque A	Suint.	Olu et al 2007
CO1	EU288170	Bathy_mauritanicus	isolate D41_24	Suint.	Genio et al 2008
CO1	EF051241	Bathy_sp._B_Nigerian_seep__short_type (near mauritan)	Nigerian slope	Suint.	Cordes 2007
CO1	EU288168	Bathy_mauritanicus	isolate D41_22	Suint.	Genio et al 2008
CO1	EU288163	Bathy_mauritanicus	isolate D41_6	Suint.	Genio et al 2008
CO1	EU288162	Bathy_mauritanicus	isolate D41_14	Suint.	Genio et al 2008
CO1	EU288161	Bathy_mauritanicus	isolate D41_10	Suint.	Genio et al 2008
CO1	EU288159	Bathy_mauritanicus	isolate AT608_1	Suint.	Genio et al 2008
CO1	AY649801	Bathy_mauritanicus	West Africa_South Côte Ivoire	Suint.	Jones et al 2006
CO1	DQ513440	Bathy_aff._childressi ou Bathymodiolus sp B	Barbados prism Orenoque A	Suint.	Olu et al 2007
CO1	DQ513439	Bathy_aff._childressi ou Bathymodiolus sp B	Barbados prism Orenoque A	Suint.	Olu et al 2007
CO1	DQ513437	Bathy_aff._childressi ou Bathymodiolus sp B	Barbados prism Orenoque A	Suint.	Olu et al 2007
CO1	DQ513436	Bathy_aff._childressi ou Bathymodiolus sp B	Barbados prism Orenoque A	Suint.	Olu et al 2007
CO1	DQ513435	Bathy_aff._childressi ou Bathymodiolus sp B	Barbados prism Orenoque A	Suint.	Olu et al 2007
CO1	DQ513434	Bathy_aff._childressi ou Bathymodiolus sp B	Barbados prism Orenoque A	Suint.	Olu et al 2007
CO1	DQ513433	Bathy_aff._childressi ou Bathymodiolus sp B	Barbados prism Orenoque A	Suint.	Olu et al 2007
CO1	DQ513432	Bathy_aff._childressi ou Bathymodiolus sp B	Barbados prism Orenoque A	Suint.	Olu et al 2007
CO1	DQ513431	Bathy_aff._childressi ou Bathymodiolus sp B	Barbados prism Orenoque A	Suint.	Olu et al 2007
CO1	DQ513430	Bathy_aff._childressi ou Bathymodiolus sp B	Barbados prism Orenoque A	Suint.	Olu et al 2007
CO1	DQ513428	Bathy_aff._childressi ou Bathymodiolus sp B	Barbados prism Orenoque A	Suint.	Olu et al 2007
CO1	DQ513427	Bathy_aff._childressi ou Bathymodiolus sp B	Barbados prism Orenoque A	Suint.	Olu et al 2007
CO1	DQ513426	Bathy_aff._childressi ou Bathymodiolus sp B	Barbados prism Orenoque A	Suint.	Olu et al 2007
CO1	AY608440	Bathy_sp._3-NZ3	Macauley Cone, W. pacific_isolate Mcsquat	Hydro	Smith 2004
CO1	AB257537	Idasola_japonica	Japan:Kagoshima, Off Noma Cape_Haplo IJN2	Os de baleine	Fujita Evo Pro unpub
CO1	AB257536	Idasola_japonica	Japan:Kagoshima, Off Noma Cape_Haplo IJN1	Os de baleine	Fujita Evo Pro unpub
CO1	AB257528	Adipi_pacifica	Japan:Kagoshima, Off Noma Cape_31.21 N 129.59 E_haplo APN2	Os de baleine	Fujita Evo Pro unpub
CO1	AB257527	Adipi_pacifica	Japan:Kagoshima, Off Noma Cape_31.21 N 129.59 E_haplo APN328-25	Os de baleine	Fujita Evo Pro unpub

Annexe 3 : Récapitulatif des séquences de CO1 utilisées pour la phylogénie des *Bathymodiolinae*

Gène	Numero d'accession / identification	Espèce	localité	Habitat	références
CO1	AB257526	Adipi_pacifica	Japan:Kagoshima, Off Noma Cape_haplo APN3	Os de baleine	Fujita Evo Pro unpub
CO1	AB170040	Adipi_pacifica	Japan:Kagishima, Off Noma Cape	Os de baleine	Iwasaki 2006
CO1	AY649803	Tamu_fisheri	Golfe Mexique, Garden Banks (Atlantic)	Suint.	Jones et al 2006
CO1	AB257514	Bentho_sp.	Japan:Torishima Seamount_30.55 N 141.49 E_haplo 895-2	Hydro	Fujita Evo Pro unpub
CO1	AB257513	Bentho_sp.	Japan:Torishima Seamount_30.55 N 141.49 E_haplo 895-1	Hydro	Fujita Evo Pro unpub
CO1	DQ077892	Bathy_sp._JDF-2005	Juan de Fuca	Hydro	McKiness 2005
CO1	AY275545	Bentho_lignicola	Chatham Rise (New Zealand)	Bois coulés / os de baleine	Smith et al unpublished
CO1	AJ865787	Brachidontes pharaonis_ 93	Mer rouge / Méditerranée	côtier	Shefer et al_2004
CO1	AJ865786	Brachidontes pharaonis_ 92	Mer rouge / Méditerranée	côtier	Shefer et al_2004
CO1	AJ865785	Brachidontes pharaonis_ 91	Mer rouge / Méditerranée	côtier	Shefer et al_2004
CO1	AJ865784	Brachidontes pharaonis_ 90	Mer rouge / Méditerranée	côtier	Shefer et al_2004
CO1	AJ865783	Brachidontes pharaonis_ 89	Mer rouge / Méditerranée	côtier	Shefer et al_2004
CO1	AJ865782	Brachidontes pharaonis_ 88	Mer rouge / Méditerranée	côtier	Shefer et al_2004
CO1	AJ865781	Brachidontes pharaonis_ 87	Mer rouge / Méditerranée	côtier	Shefer et al_2004
CO1	AJ865776	Brachidontes pharaonis_ 82	Mer rouge / Méditerranée	côtier	Shefer et al_2004
CO1	AJ865775	Brachidontes pharaonis_ 81	Mer rouge / Méditerranée	côtier	Shefer et al_2004
CO1	AJ865774	Brachidontes pharaonis_ 80	Mer rouge / Méditerranée	côtier	Shefer et al_2004
CO1	EU018203	Mytilus edulis	?	côtier	Jha et al_2008
CO1	EU018201	Mytilus edulis	?	côtier	Jha et al_2009
CO1	EU018200	Mytilus edulis	?	côtier	Jha et al_2010
CO1	EU018204	Mytilus edulis	?	côtier	Jha et al_2011
CO1	EU018199	Mytilus edulis	?	côtier	Jha et al_2012
CO1	EU018198	Mytilus edulis	?	côtier	Jha et al_2013
CO1	EU018197	Mytilus edulis	?	côtier	Jha et al_2014
CO1	EU018195	Mytilus edulis	?	côtier	Jha et al_2015
CO1	EU018194	Mytilus edulis	?	côtier	Jha et al_2016
CO1	EU018202	Mytilus edulis	?	côtier	Jha et al_2017
CO1	EU018196	Mytilus edulis	?	côtier	Jha et al_2018
CO1	DQ351479	Mytilus galloprovincialis	?	côtier	Zardi et al_2007
CO1	AY363687	Mytilus galloprovincialis	?	côtier	Cao et al_2004
CO1	DQ864430	Mytilus galloprovincialis	?	côtier	Lewis et Jarman_unpublished
CO1	DQ864383	Mytilus galloprovincialis	?	côtier	Lewis et Jarman_unpublished
CO1	DQ399833	Mytilus galloprovincialis	?	côtier	Venetis et al_2007
CO1	AY296816	Trichomya hirsuta	?	côtier	Colgan et al_2003
CO1	DQ198225	Mytilus trossulus	?	côtier	Burzynski et al_2006

TECHNICAL ARTICLE

Mark–recapture cloning: a straightforward and cost-effective cloning method for population genetics of single-copy nuclear DNA sequences in diploids

N. BIERNE,* A. TANGUY,† M. FAURE,* B. FAURE,†‡ E. DAVID,§ I. BOUTET,§ E. BOON,* N. QUERE,† S. PLOUVIEZ,*† P. KEMPPAINEN,¶ D. JOLLIVET,† D. MORAGA,§ P. BOUDRY‡ and P. DAVID**

*Génome Populations Interactions Adaptation, UMR 5171, Université Montpellier II – IFREMER – CNRS, Station Méditerranéenne de l'Environnement Littoral, 34200 Sète, France, †Equipe Evolution & Génétique des Populations Marines, UMR 7144, UPMC – CNRS, Station biologique de Roscoff, BP. 74, Place Georges Teissier, 29682 Roscoff, France, ‡Laboratoire de Génétique et Pathologie, IFREMER, 17390 La Tremblade, France, §Laboratoire des sciences de l'environnement marin, UMR CNRS 6539, Institut Universitaire Européen de la Mer, Université de Bretagne Occidentale 29280 Plouzané, France, ¶Department of Marine Ecology, Tjärnö Marine Biological Laboratory, 45296 Strömstad, Sweden, **Centre d'Ecologie Fonctionnelle et Evolutive – CNRS, 34293 Montpellier cedex 5, France

Abstract

We describe a simple protocol to reduce the number of cloning reactions of nuclear DNA sequences in population genetic studies of diploid organisms. Cloning is a necessary step to obtain correct haplotypes in such organisms, and, while traditional methods are efficient at cloning together many genes of a single individual, population geneticists rather need to clone the same locus in many individuals. Our method consists of marking individual sequences during the polymerase chain reaction (PCR) using 5'-tailed primers with small polynucleotide tags. PCR products are mixed together before the cloning reaction and clones are sequenced with universal plasmid primers. The individual from which a sequence comes from is identified by the tag sequences upstream of each initial primer. We called our protocol mark–recapture (MR) cloning. We present results from 57 experiments of MR cloning conducted in four distinct laboratories using nuclear loci of various lengths in different invertebrate species. Rate of capture (proportion of individuals for which one or more sequences were retrieved) and multiple capture (proportion of individuals for which two or more sequences were retrieved) empirically obtained are described. We estimated that MR cloning allowed reducing costs by up to 70% when compared to conventional individual-based cloning. However, we recommend to adjust the mark:recapture ratio in order to obtain multiple sequences from the same individual and circumvent inherent technical artefacts of PCR, cloning and sequencing. We argue that MR cloning is a valid and reliable high-throughput method, providing the number of sequences exceeds the number of individuals initially amplified.

Keywords: DNA polymorphism, high throughput allele recognition, population genetics, sequence

Received 30 July 2006; revision accepted 4 December 2006

The analysis of gene genealogies by increasingly powerful methods (Balding *et al.* 2001; Slatkin & Veuille 2002; Zhang & Hewitt 2003) and the development of methods to quantify adaptation at the molecular level (Yang &

Bielawski 2000; Fay & Wu 2001) make DNA sequence a major tool in population genetics. Although the literature abounds in studies of mitochondrial DNA (mtDNA) and concerted evolving multiple-copy ribosomal DNA (rDNA) loci, the analysis of single-copy nuclear DNA sequences remains surprisingly infrequent and limited to model organisms (Zhang & Hewitt 2003). The lack of

Correspondence: Nicolas Bierne, Fax: +33 (0)467463399; E-mail: n-bierne@university-montp2.fr

reference sequences in nonmodel organisms does not explain everything (Zhang & Hewitt 2003). Another technical challenge is the difficulty to identify alleles in heterozygous state in outcrossing diploid organisms. Heterozygous individuals at a given locus have two different alleles that should ideally be sequenced independently. Even though alternative methods exist and are continuously explored (Zhang & Hewitt 2003), cloning of PCR products often remains an essential step. While PCR and sequencing have become universally used low-cost techniques, individual cloning still remains time-consuming and expensive. As a consequence, molecular ecologists endeavour to avoid the cloning procedure when possible, restricting the analysis of DNA sequences to mtDNA, rDNA or sex chromosomes in the hemizygous sex when available, or losing the benefit of genealogical information by typing single nucleotide polymorphisms (SNP), even when nucleotide diversity is high. When individual cloning is performed, the cost increases proportionally to sample size, setting a strong limit to the latter.

Here we describe a simple protocol that allows the cloning of PCR products of several individuals from a population sample at once, leading to a less time- and resource-consuming cloning procedure. Our method is based on the observation that cloning can separate single alleles from several individuals as well as it does within a single individual. A simple solution to reduce the number of cloning reactions would therefore be to pool the PCR products of several individuals before cloning and to sequence many clones (e.g. Kronforst *et al.* 2006). However, with such a procedure it is no longer possible to know the individual from which an allele sequence comes from. To solve this problem, PCR products need to be individually marked. The method we found consists of marking individual sequences during the PCR using slightly different primer pairs for each individual. To this aim, every primer is 5'-tailed with a small polynucleotide tag. Tags do not match the matrix DNA sequence in the initial stages of the PCR and does not perturb the reaction. The method is essentially similar to the M13-tailing technique (Oetting *et al.* 1995) although the tail is much smaller. PCR products of similar quantities are mixed together and cloned with standard protocols. Clones are then sequenced with universal plasmid primers flanking the insert. The small polynucleotide tags upstream of primers are therefore sequenced and allow identifying the individual from which the sequence comes from. Using the combination of the forward and reverse primers, it is not necessary to use different primer pairs for each PCR-amplified individual. For instance, we usually used eight different tags for the forward primers and six for the reverse primers, yielding 48 unique combinations by which sequences can be recognized.

PCR products were quantified on agarose gel stained with ethidium bromide then mixed together in such a way

as to equalize concentration of each PCR product. Pools of PCR products were purified with the QIAquick PCR purification kit or the QIAEX II Purification Kit (QIAGEN), and cloned with the pGEM-T Vector System (Promega) according to manufacturer's recommendations. Positive clones were screened for the presence of appropriate-sized inserts by PCR amplifications then sent to the Genoscope platform (www.genoscope.cns.fr/) where plasmid extraction and sequencing with vector-specific primers SP6 (5'-TATTTAGGTGACACTATAG-3') and T7 (5'-TAATACGACTCACTATAGGG-3') were performed.

The method has been tested in four distinct laboratories accounting for 57 experiments of mark-recapture (MR) cloning using various species of marine invertebrates and genes (Table S1, Supplementary material). We present observed rates of capture (i.e. the proportion of individuals for which one or more sequences was obtained), technical artefacts we have encountered and recommendations to accommodate artefacts in the laboratory or during statistical analysis.

Stochastic processes during PCR, ligation, transformation and bacterial growth can sometimes generate an over-representation of a few sequences at the end of the experiment. To circumvent this drift effect, we choose to pool an appreciable number of individuals (usually 48 which corresponds to half a PCR plate). Our aim was not to capture every individual of the initial sample. The average number of sequences obtained and number of individuals captured in each experiment are given in Table S1. The rate of capture (number of individuals captured/initial sample size) increased with the capture effort (number of sequences/initial sample size) but was on average slightly lower than the expectation based on a uniform distribution (Fig. 1). The rate of multiple capture which provides more reliable data (see below) increased linearly with the

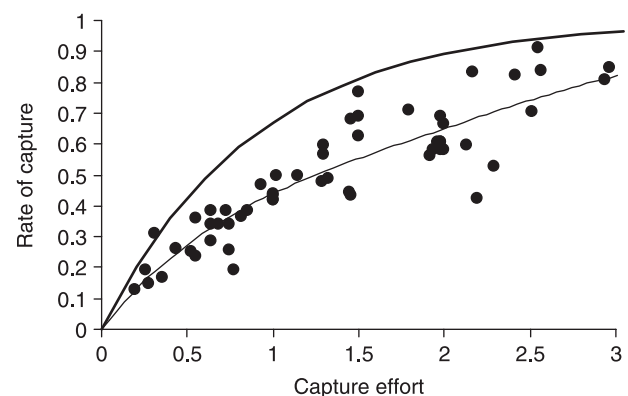


Fig. 1 Rate of capture (number of individuals captured/initial sample size) as a function of the capture effort (number of sequences/initial sample size). The thick line is the expectation based on a uniform distribution and the thin line is a quartic polynomial regression on the data.

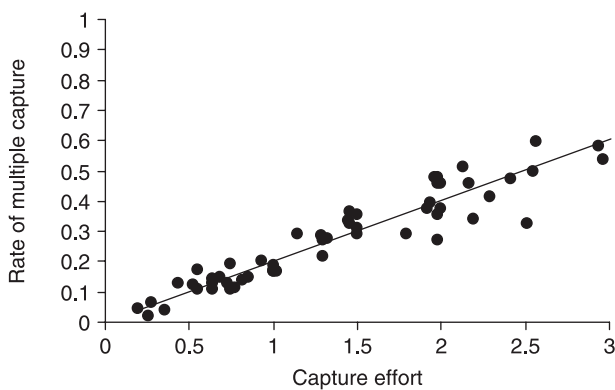


Fig. 2 Rate of multiple captures (number of individuals captured more than once/initial sample size) as a function of the capture effort (number of sequences/initial sample size). The line is a linear regression on the data (slope = 0.2).

capture effort (slope = 0.2) for the range of capture effort investigated in this study (Fig. 2).

In the course of the development of the protocol, we encountered a number of technical artefacts. First, a number of tags were partially or totally deleted during the cloning process. Tag deletion led to an average rate of unassigned sequences of ~7%, but this rate was highly variable depending on the locus studied (Table S1). We suspect that the sequence upstream of the primer in the matrix DNA may have an impact because a high rate of deletion has been observed for a primer immediately designed after a poly T repetition (25%). However, other primers sometimes reached as a high rate of deletion without any visible distinctiveness at the DNA primary structure. Unassigned sequences should not inevitably be removed from the data analysis (see Kronforst *et al.* 2006) but the consequences of their use need to be considered. Second, the impact of classical technical artefacts usually encountered in this kind of protocol – that is mutation during PCR, cloning and sequencing, is not easy to appreciate with our technique. We expect an individual to have a maximum number of two different sequences (i.e. alleles) and when two sequences are observed, the divergence should be in accordance with the global diversity observed. A small proportion of individuals captured several times displayed more than two alleles (~8%). However, in such cases differences were only due to the presence of a single artefactual mutation in one sequence. We also observed individuals with two alleles, of which one was sequenced only once, differing by a single nucleotide, while the average pairwise difference in the whole sample was much greater. Third and most problematically, we observed in a few cases multiple captured individuals for which more than two alleles presented such a divergence that sequence misassignment to this individual was the only valid explanation. Misassignment can occur owing to a mutation

in a tag (during PCR, ligation or bacterial replication) or *in vitro* recombination. Indeed, in some instances one of the sequences retrieved was in good agreement with an event of recombination between divergent alleles present in our sample.

We found no satisfactory solution for tag deletion. Initial experiments were conducted with two-nucleotide tags which was enough to create our 14 primers. Tag length was sometimes increased in successive experiments with no significant impact on this problem. We observed a strong variation in the rate of tag deletion according to the locus analysed (Table S1). We therefore suspect an effect of the primer sequence (hairpin or duplex effect) or the sequence upstream of the primer, although we were unable to find convincing evidence for such an effect.

The problem of artefactual mutations could be circumvented by restricting genetic data analysis to alleles captured several times. However, the rate of artefactual mutations was always low. One can then compare the results obtained with reliable alleles (for which several sequences were captured) and results obtained with the whole data set. Because artefactual mutations should mainly create singletons (mutations observed in a single sequence of the data set) an interesting parameter to evaluate in this respect is the proportion of singleton mutations. One can also choose the data set required depending on the analysis conducted. For instance, any sequences can be used in most analyses of molecular evolution that compare the relative rate of evolution between different categories of mutations within the same sequence (synonymous, nonsynonymous, noncoding, indels). The McDonald–Kreitman test (McDonald & Kreitman 1991a) falls in this category of analysis (McDonald & Kreitman 1991b). In addition, singletons can sometimes be removed from the data in some analyses of molecular evolution (e.g. Bierne & Eyre-Walker 2004; Andolfatto 2005). Here, attention might be called to the fact that such a technical artefact is a ubiquitous problem not restrained to the MR cloning protocol (Zhang & Hewitt 2003).

Misassignment (tag mutation or *in vitro* recombination) could have been a serious problem if the rate was high. When nucleotide diversity is low, misassignment can easily be confounded with standard artefactual mutations. Luckily, marine invertebrates usually exhibit very high nucleotide diversities (π often > 0.01, Table S1). We were able not only to detect misassignment, but also to estimate its rate. The rate of misassignment turned out to be low (< 2%, Table S1). The occurrence of *in vitro* recombination is known to occur at a non-negligible rate during PCR (Meyerhans *et al.* 1990) or cloning (Tang & Unnasch 1995). Such chimeric DNA products are well known in surveys of bacterial 16S rRNA genes (Kopczynski *et al.* 1994). However, this artefact is not easily detected when nucleotide diversity is low. We argue that *in vitro* recombination

is not a more serious bias in MR cloning than in standard protocols but is detected in multiple captures (recombination during PCR) or because of tags rearrangement (recombination during cloning). As for artefactual mutations, the problem can be solved by restricting genetic data analysis to alleles captured several times.

Finally, we would need to estimate the time/money saved with MR cloning over standard protocols for a comparable amount of data collected. The time saved seems obvious to us, as a cloning reaction is far more time-consuming than a sequencing reaction; especially when accounting for the recent technical progress made in the automatization of sequencing. In addition, sequencing platforms have flourished and the sequencing step is increasingly outsourced to these platforms. Estimating the money saved is more difficult because costs and laboratory facilities can vary widely among laboratories and countries. First, we used our estimated costs of primers, PCR, PCR product purification, cloning and sequencing reactions to evaluate the cost of an MR cloning. Then, using our empirical rate of capture (quartic regression in Fig. 1) we estimated the cost of obtaining the same final number of sequences with standard individual-based cloning protocols. However, the estimate we made is an underestimation because we neglected our salaries in the calculation. To take costs of manual work into account, we used in a second estimate prices given by a private company (information one can easily get on the web). The financial gain of an MR cloning protocol primarily depends on the ratio of the cost of a cloning reaction to the cost of a sequencing reaction, which turned out to be five in our case but was estimated to be 15 from the costs provided by private companies. The relative cost of MR cloning to standard protocols of individual cloning is presented in Fig. 3 as a function of the sample size for population genetics analysis. As expected, the bigger is the final sample size, the more is the saving of money provided by MR cloning. MR cloning was estimated reducing costs by up to 70% when compared to conventional individual-based cloning (Fig. 3). We do not claim that MR cloning would be so cost-effective in every laboratory. In addition, one may not plan to obtain a big sample size simply to save money while the genetic information sought could emerge in a small sample size (e.g. Felsenstein 2005). However, we would argue that big sample sizes can often be highly valuable for population genetics inference in nonequilibrium populations for instance when it allows sampling the rare lineage that has survived a bottleneck or a selective sweep or that has introgressed through a barrier to gene flow.

We would conclude that MR cloning is a valid and reliable high-throughput method. From the experience we gained with MR cloning, we would recommend to use an appreciable effort of capture (say two to three) in order to obtain multiple sequences from the same individual (see Fig. 2) and circumvent inherent technical artefacts of PCR,

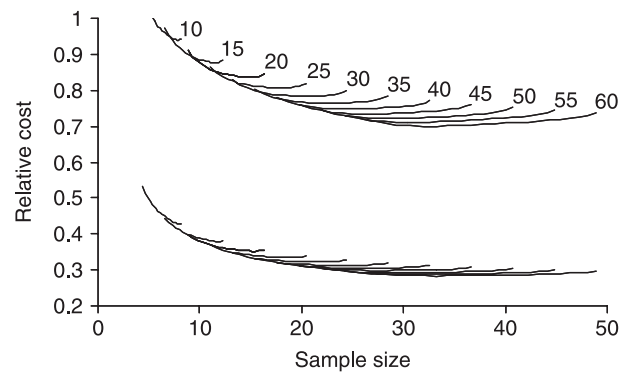


Fig. 3 Estimated cost of MR cloning protocols relative to standard protocols of individual cloning as a function of the sample size for population genetics analysis. The empirically estimated rate of capture of Fig. 1 was used for a gradient of initial sample size (the number of individuals PCR-amplified with tagged primers in the MR cloning), and a gradient of capture effort (number of sequences performed/initial sample size). Numbers closed to curves indicate the initial sample sizes for MR cloning. Each curve is generated with efforts of capture ranging from one to three. The upper series of curves are estimates that neglect salary costs (based on the prices we get for molecular biology kits and products) and the lower series of curves are estimates that include salary costs (based on prices practiced by private companies for a complete outsource of the experiment).

cloning and sequencing. However, the level of precision required depends on the nucleotide diversity observed and the data analysis one wants to conduct. MR cloning offers an opportunity to appreciate the consequences of technical artefacts by comparing more or less stringent data sets (e.g. raw data sets to data sets restricted to sequences obtained more than once).

Acknowledgements

The authors are very indebted to the National Sequencing Centre (Genoscope at Evry) for providing access to their sequencing platform. They thank François Bonhomme, Grant Pogson and Tobias Lindskog for helpful discussions. This research was funded by the national programme GIS 'Génomique Marine', by the Fish & Shellfish node of the NoE 'Marine Genomics Europe' and by La Région Bretagne. This is publication GPIA 2006-144.

Supplementary material

The supplementary material is available from <http://www.blackwellpublishing.com/products/journals/suppmat/MEN/MEN1685/MEN1685sm.htm>

Table S1 Statistics for the 57 experiments of MR cloning

References

Andolfatto P (2005) Adaptive evolution of non-coding DNA in *Drosophila*. *Nature*, **437**, 1149–1152.

- Balding DJ, Bishop M, Cannings C (2001) *Handbook of Statistical Genetics*, p. 847. John Wiley & Sons, Ltd., Chichester, UK.
- Bierne N, Eyre-Walker A (2004) The genomic rate of adaptive amino-acid substitution in *Drosophila*. *Molecular Biology and Evolution*, **21**, 1350–1360.
- Fay JC, Wu CI (2001) The neutral theory in the genomic era. *Current Opinion in Genetics and Development*, **11**, 642–646.
- Felsenstein J (2005) Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? *Molecular Biology and Evolution*, **23**, 691–700.
- Kopczynski ED, Bateson MM, Ward DM (1994) Recognition of chimeric small-subunit ribosomal DNAs composed of genes from uncultivated microorganisms. *Applied and Environmental Microbiology*, **60**, 746–748.
- Kronforst MR, Young LG, Blume LM, Gilbert LE (2006) Multilocus analyses of admixture and introgression among hybridizing *Heliconius* butterflies. *Evolution*, **60**, 1254–1268.
- McDonald JH, Kreitman M (1991a) Adaptive evolution at the Adh locus in *Drosophila*. *Nature*, **351**, 652–654.
- McDonald JH, Kreitman M (1991b) Neutral mutation hypothesis test. *Nature*, **354**, 116.
- Meyerhans A, Vartanian JP, Wain-Hobson S (1990) DNA recombination during PCR. *Nucleic Acids Research*, **18**, 1687–1691.
- Oetting WS, Lee HK, Flanders DJ *et al.* (1995) Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13-tailed primers. *Genomics*, **30**, 450–458.
- Slatkin M, Veuille M (2002) *Modern Developments in Theoretical Population Genetics, the Legacy of Gustave Malecot*. Oxford University Press, Oxford.
- Tang J, Unnasch TR (1995) Discriminating PCR artifacts using directed heteroduplex analysis (DHDA). *BioTechniques*, **19**, 902–905.
- Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution*, **15**, 496–503.
- Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*, **12**, 563–584.

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article was published in an Elsevier journal. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



ELSEVIER

Available online at www.sciencedirect.com

Gene 406 (2007) 99–107

GENE

www.elsevier.com/locate/gene

Evidence for a slightly deleterious effect of intron polymorphisms at the EF1 α gene in the deep-sea hydrothermal vent bivalve *Bathymodiolus*

B. Faure^{a,*}, N. Bierne^b, A. Tanguy^a, F. Bonhomme^b, D. Jollivet^a^a *Evolution et Génétique des Populations Marines, UMR 7144, CNRS-UPMC, Station Biologique de Roscoff, Place Georges Teissier, 29682 Roscoff, France*^b *Département Biologie Intégrative, Institut des Sciences de l'Evolution, UMR5554, UM2-CNRS, Université Montpellier 2, Place Eugène Bataillon, CC 63, 34 095 Montpellier cedex 5, France.*

Received 21 February 2007; received in revised form 26 May 2007; accepted 28 June 2007

Available online 17 July 2007

Abstract

A multilocus analysis was initiated in order to infer the general effect of demography and the indirect effect of positive selection on some chromosome segments in *Bathymodiolus*. Mussels of the genus *Bathymodiolus* inhabit the very hostile, fragmented and variable environment of deep-sea hydrothermal vents which is thought to cause recurrent population bottlenecks via extinction/colonisation processes and adaptation to new environmental conditions. In the course of this work we discovered that the assumption of neutrality of non-coding polymorphisms usually made in genome scan experiments was likely to be violated at one of the loci we analysed. The direct effect of slight purifying selection on non-coding polymorphisms shares many resemblances with the indirect effect of positive selection through genetic hitchhiking. Combining polymorphism with divergence data for several closely related species allowed us to obtain different expectations for the direct effect of negative selection and the indirect effect of positive selection. We observed a strong excess of rare non-coding polymorphisms at the second intron of the EF1 α gene in the two species *Bathymodiolus azoricus* and *Bathymodiolus thermophilus*, while two other loci, the mitochondrial COI gene and an intron of the Lysozyme gene, did not exhibit such a deviation. In addition, the divergence rate of the EF1 α intron was estimated to be unexpectedly low when calibrated using the closure of the Panama Isthmus that interrupted gene flow between the two species. The polymorphism to divergence ratio was similar to the one observed for the other two loci, in accordance to the hypothesis of purifying selection. We conclude that slight purifying selection is likely to act on polymorphic intronic mutations of the EF1 α second intron and discuss the possible relationship with the specific biology of *Bathymodiolus* mussels.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Deleterious mutations; Selective sweep; Gene genealogies; Non-coding DNA; Vent species

1. Introduction

The analysis of DNA sequence polymorphism started two decades ago, with the first report at the alcohol dehydrogenase

Abbreviations: ATP, Adenosine Triphosphate; Ba, *Bathymodiolus azoricus*; Bt, *Bathymodiolus thermophilus*; cDNA; DNA complementary to RNA; Dn, Non-synonymous difference between species; Ds, Synonymous difference between species; EF1 α , Elongation Factor 1 α gene; EPR, East Pacific Rise; HKA, Hudson, Kreitman and Aguadé test; Lyso, Lysozyme gene; MAR, Mid-Atlantic Ridge; MK test, McDonald and Kreitman test; MR-method, Mark-recapture cloning method; mtCOI, Mitochondrial Cytochrome-C oxidase I gene; Ne, Population size; NJ, Neighbor-joining; NoE, Network of Excellence; Pn, Non-synonymous difference within species; Ps, Synonymous difference within species; ROV, Remote Operated Vehicle.

* Corresponding author. Tel.: +33 2 98 29 23 67.

E-mail address: bfaure@sb-roscoff.fr (B. Faure).

locus in *Drosophila melanogaster* (Kreitman, 1983). Ever since, the debate about the 'adaptive' significance of observed nucleotide polymorphisms never really stopped (Lewontin, 1991). It is indeed not trivial to tell the effect of past demographic changes apart from the eventual footprint of indirect selection on linked neutral polymorphisms and the direct effect of selective forces acting on some polymorphisms (Glinka et al., 2003; Haddrill et al., 2005; Li and Stephan, 2006). One idea is that demographic processes affect the entire genome in a similar way, whereas selection acts in a locus specific manner (Hudson et al., 1987; Galtier et al., 2000). The spread of a positively selected allele is known to greatly affect the frequencies of neutral mutations at linked positions and, thus to mimic the effect of a population bottleneck (Fu and Li, 1993). This loss of diversity at linked positions is commonly referred

as the hitchhiking effect (Maynard Smith and Haigh, 1974). The magnitude of the effect depends on the strength of selection on the beneficial mutation, recombination rate between the neutral locus and the beneficial mutation and the effective population size of the species under study (Kaplan et al., 1989). The hitchhiking effect refers to an indirect action of natural selection at the locus under consideration. However, the loss of genetic diversity may also indicate a direct effect of directional selective processes in which most mutations are deleterious (Tajima, 1989b). Within exonic regions for instance, non-synonymous polymorphisms affecting the protein tend to segregate at lower frequencies than synonymous mutations (Fay and Wu, 2001). This observation is taken as evidence that some non-synonymous polymorphisms are slightly deleterious (Eyre-Walker et al., 2006). Non-synonymous mutations contribute to a shift of the allele frequency spectrum towards rare variants which, if assumed neutral, bias the interpretation towards the inference of a selective sweep or a population bottleneck. Slight selection has also been shown to sometimes affect synonymous mutations (Akashi, 1995) and non-coding and/or intronic regions (Leicht et al., 1995; Jaruzelska et al., 1999) whilst these categories of mutations are often assumed to behave neutrally in hitchhiking mapping experiments. Tajima's D is a well-known indicator of the distortion of the allele frequency spectrum from the neutral expectation at mutation-drift equilibrium. Recently, Andolfatto (2005) reported Tajima's D values for different categories of mutations (synonymous, non-synonymous and various non-coding categories) in *D. melanogaster* that suggested a non-negligible fraction of slightly deleterious non-coding polymorphisms. Andolfatto's results challenged the existence of a genuine category of mutations that could be assumed as perfectly neutral in *D. melanogaster* except perhaps synonymous mutations on which selection is thought to be relaxed in this lineage (Akashi, 1996). When neutrality cannot be assumed on any category of mutations comparing polymorphism to divergence data (Kimura, 1983) and/or polymorphism data in closely related species (Moriyama and Powell, 1997) proved useful to tease apart the various forces (drift, migration, selection, recombination) accounting for the evolutionary dynamics of genes within populations and species.

We initiated a multilocus analysis in order to infer the cumulative effect of demography and selection on some portions of the genome in the deep-sea hydrothermal vent mussels of the genus *Bathymodiolus*. Mussels of the genus *Bathymodiolus* (Kenk and Wilson, 1985) are one of the most widespread mollusc dominating deep-sea hydrothermal vents and cold-seep communities throughout the world. These bivalves seem to have diversified a few millions of years ago following an explosive radiation (Jones et al., 2006) more or less concomitant with the closure of the Panama Isthmus. One of the most intriguing features associated with these mussels is that they are living in a very fragmented and variable environment in which extinction/recolonisation events are common place (Jollivet, 1996; Jollivet et al., 1999). Environmental characteristics are prone to cause both recurrent population bottlenecks and adaptation to new environmental conditions making *Bathymodiolus* mussels a good model to investigate the role of selection and drift in

shaping the structure of populations. In the course of this work, we were able to find evidences for the direct action of selection on non-coding polymorphisms at one of the loci we were analysing, the second intron of the Elongation Factor 1 α (EF1 α) gene. Here, we present a comparison of gene genealogies for three portions of genes between two well-separated species of the genus *Bathymodiolus* for which speciation is believed to have occurred prior to the closure of the Panama Isthmus: the mitochondrial Cytochrome-C oxydase I (COI) and two nuclear introns of genes encoding for the Lysozyme (Lys) and the Elongation Factor 1 α (EF1 α).

2. Materials and methods

2.1. Sampling

The two *Bathymodiolus* species were collected during the oceanographic cruises ATOS 2001 and BioSpeedo 2004 on board of the N/O L'Atalante. The first species, *Bathymodiolus azoricus*, was sampled from the site Rainbow (2500 m; 36°13'N–33°54'W) on the Mid-Atlantic Ridge (MAR) using the telemanipulated arm of the ROV Victor6000. The second species, *Bathymodiolus thermophilus*, was sampled at 7°25S vent field (2450 m; 7°25'S–107°47'W) on the East Pacific Rise (EPR) using the telemanipulated arm of the manned submersible Nautilie. Samples were dissected on board, foot tissue and gills were immediately frozen and kept in liquid nitrogen until nucleic acid extraction back in our laboratory. DNA was extracted using the usual phenol-chloroform protocol (Sambrook and Russel, 2001).

2.2. Gene selection

The mitochondrial Cytochrome-C Oxidase subunit I (COI) gene has been already used to describe the phylogeny status of Bathymodiolinae bivalves (Iwasaki et al., 2006) and exhibits reciprocal monophyly between all the well-identified species. This subunit has a critical role during respiration by catalyzing the transfer of electrons between ferrocytochrome *c* and molecular oxygen and contributes to the ATP generation (Malatesta et al., 1995). Previous studies already suggested that it might be under strong purifying selection for organisms living in anaerobic conditions such as vent species (Powell and Somero, 1986) which may explain the very low substitution rate already found for them (Chevaldonné et al., 2002). The second protein, Lysozyme, is an enzyme able to split a (1,4) bond between two amino sugars, *N*-acetylmuramic acid and *N*-acetylglucosamine, of the bacterial peptidoglycan. It was suggested that this enzyme plays a key role in the body defence against bacterial infection (Bachali et al., 2002). The third gene, the elongation factor EF1 α is a subunit of the major factor of the translational machinery, but it also has remarkable functions in limiting cell proliferation and/or apoptosis (Ejiri, 2002).

2.3. Gene amplification, mark-recapture cloning and sequencing

Species-specific primers were designed from conserved regions (mtCOI) or exonic regions flanking introns (nuclear

genes) from *Bathymodiolus* sequences obtained from a collection of ESTs on *B. azoricus* (cDNA library: Marine Genomics Europe NoE, Fish & Shellfish node: Tanguy et al. in prep). A 528 bp long fragment was amplified from the Cytochrome *c* Oxydase subunit I mitochondrial gene using Bat-CO1f: 5'-TGTGGTCTGGAATAATTGGAAC-3'; Bat-CO1r 5'-ATAAAAAGATGTATTRAARTGACG-3' (Olu-Leroy et al., in press). Reactions were performed in 25 μ l containing 1X PCR buffer (supplied with the polymerase enzyme), 2.5 mM MgCl₂, 0.25 μ M of each dNTP, 0.5 μ M of each primer, 0.04 U of Thermoprime Plus *Taq* polymerase (ABgene), 25 ng genomic DNA. Cycling parameters were 94 °C for 2 min, 5 cycles of 94 °C for 35 s, 46 °C for 35 s, 72 °C for 70 s; followed by 30 cycles of 94 °C for 35 s, 50 °C for 35 s, and 72 °C for 70 s and a final elongation at 72 °C for 10 min. PCR products were purified before sequencing with the Multiscreen 96-plate purification kitTM (Millipore) and sequenced using the BigDye terminator chemistry (PerkinElmer) on an ABI 3100 sequencer, following the manufacturer's protocol. Sequences were run in both directions.

Nuclear genes were amplified and cloned following a new method called MR-cloning (Bierne et al., 2007), which allows the cloning of PCR products of several individuals at once. The method consists of marking individual sequences during the PCR using 5'-tailed primers with small polynucleotide tags. Amplifications were conducted with the gene-specific primer pairs (EF1 α gene: BatEF1f 5'-ACGCCTGGGTATTGGACAAACT-3' and BatEF1r 5'-CCAAGAGGGGTCTGACAAATTA-3'; Lysozyme gene: BatLysof 5'-GTTTCCCCAAAATGTATGAGCTGT-3' (as adapted from Bachali et al., 2002) and BatLyso 3'-TAATCTCCGCCTGGACTACCACAATC-3') supplemented with a 4 bp polynucleotide tag at their 5' ends. Eight and six polynucleotide motifs were used for the forward and reverse primers, respectively (Table 1) allowing the amplification of 48 individuals for a single cloning reaction. Reactions were performed in a 20 μ l volume containing 1X PCR buffer (supplied with polymerase enzyme), 2 mM MgCl₂, 0.25 μ M of each dNTPs, 0.4 μ M of each primer, 0.025 U of Thermoprime Plus *Taq* polymerase (ABgene), 20 ng genomic DNA. Cycling parameters were 94 °C for 5 min, 35

cycles of 94 °C for 45 s, 60 °C for 1 min, 72 °C for 2 min; and a final elongation at 72 °C for 10 min for both genes. We then used the mark-recapture cloning protocol described in Bierne et al. (2007). Every individual were PCR-amplified separately with a set of 5'-tagged primers. PCR products of similar quantities were mixed together and cloned into a pGEM-T vector by using a Promega pGEM-T cloning kit (Promega, Madison, WI, USA). Artfactual cross-recombination between alleles of different individuals cannot occur during this protocol as the products cloned are pooled after the PCR reaction. However, recombination of the two alleles of a heterozygous individual remains possible as is nonetheless the case in standard protocols. We used a capture effort of two in the MR-cloning experiment according to Bierne et al. (2007): a total of 96 clones per population for each gene was sequenced at the Génoscope (Evry) (<http://www.genoscope.cns.fr/>) with the universal primers flanking the insert in the plasmid. The small polynucleotide tags upstream of the primers allowed us to identify the individual from which the sequences came from.

2.4. Genetic analyses

For all genes, sequences were checked and edited using the Chromas 2.22 computer program (Technelysium Pty. Ltd., Helensvale, Australia) and aligned using ClustalW (Thompson et al., 1994) in BioEdit Sequence Alignment Editor (Hall, 1999). All sequences were carefully examined for polymorphic mutations. Artfactual mutations and in vitro recombinants were suppressed from the dataset when detected from multiple-recaptures if more than two allelic forms occurred (see Bierne et al., 2007). Neighbour-joining trees were constructed using MEGA 3 (Kumar et al., 2004), with the number of differences and the Kimura-2 parameter methods to delineate gene genealogies within each species.

For each gene, the haplotype diversity (Hd), the number of segregating sites (S) and the nucleotide diversities (π) were obtained from silent, synonymous and non-synonymous sites associated with coding and non-coding regions. The number of fixed mutations was also computed from the three portions of

Table 1
Schematic view of a PCR reaction plate with the 5'-tailed primers and the 8*6 combinations

	CACA-forward primer	CTCT-forward primer	CGCG-forward primer	TATA-forward primer	TGTG-forward primer	AGAG-forward primer	AAAA-forward primer	CCCC-forward primer
CACA-reverse primer	CACA-F/ CACA-R	CTCT-F/ CACA-R	CGCG-F/ CACA-R	TATA-F/ CACA-R	TGTG-F/ CACA-R	AGAG-F/ CACA-R	AAAA-F/ CACA-R	CCCC-F/ CACA-R
CTCT-reverse primer	CACA-F/ CTCT-R	CTCT-F/ CTCT-R	CGCG-F/ CTCT-R	TATA-F/ CTCT-R	TGTG-F/ CTCT-R	AGAG-F/ CTCT-R	AAAA-F/ CTCT-R	CCCC-F/ CTCT-R
CGCG-reverse primer	CACA-F/ CGCG-R	CTCT-F/ CGCG-R	CGCG-F/ CGCG-R	TATA-F/ CGCG-R	TGTG-F/ CGCG-R	AGAG-F/ CGCG-R	AAAA-F/ CGCG-R	CCCC-F/ CGCG-R
TATA-reverse primer	CACA-F/ TATA-R	CTCT-F/ TATA-R	CGCG-F/ TATA-R	TATA-F/ TATA-R	TGTG-F/ TATA-R	AGAG-F/ TATA-R	AAAA-F/ TATA-R	CCCC-F/ TATA-R
TGTG-reverse primer	CACA-F/ TGTG-R	CTCT-F/ TGTG-R	CGCG-F/ TGTG-R	TATA-F/ TGTG-R	TGTG-F/ TGTG-R	AGAG-F/ TGTG-R	AAAA-F/ TGTG-R	CCCC-F/ TGTG-R
AGAG-reverse primer	CACA-F/ AGAG-R	CTCT-F/ AGAG-R	CGCG-F/ AGAG-R	TATA-F/ AGAG-R	TGTG-F/ AGAG-R	AGAG-F/ AGAG-R	AAAA-F/ AGAG-R	CCCC-F/ AGAG-R

genes to estimate the divergence between species based on the Kimura-2-parameters method, using the DnaSP version 4.10 software package (Rozas et al., 2003).

Neutrality tests were performed based on our set of genes using polymorphism and divergence data. Tajima's (1989a) test is only based on extant polymorphism and uses the haplotype sequence information to compare two estimators of Watterson's θ ($4N_e\mu$ for nuclear genes on autosomal chromosomes and $N_e\mu$ for mitochondrial genes) under an infinite-site model without recombination. Significant nonzero negative values of the statistics can result either from a past population contraction (bottleneck) or expansion, from a selective sweep in the chromosomal vicinity of the locus studied or from negative selection on the polymorphic mutations observed. Conversely, positive values are more likely to reflect the maintenance of deep branching lineage resulting from population structure, a recent bottleneck and, direct or indirect balancing selection. Statistical significance of the Tajima's D tests on each locus was assessed by coalescent simulations with the software HKA (J. Hey's web page: <http://lifesci.rutgers.edu/~hey/HeylabSoftware.htm#HKA>). Although the Tajima's D test relies on polymorphism data, coalescence simulations are conditioned on Theta estimates that incorporate the information (e.g. mutation rate) contained in divergence data.

Departure from neutral evolution was also examined using tests that compare the number of mutations accumulated in both divergence and polymorphisms. The MK test (McDonald and Kreitman, 1991) is designed to detect the direct action of selection on non-synonymous mutations. This method compares the number of synonymous and/or non-coding differences (both mutations being referred to as silent) with the number of non-synonymous differences that have been fixed between two species (D_n/D_s ratio) and those that segregate within species (P_n/P_s ratio). If the D_n/D_s ratio is significantly greater than the P_n/P_s ratio, at least one lineage may have evolved under positive Darwinian selection. In the opposite case, significant values of the test indicate the segregation of slightly deleterious non-synonymous polymorphisms. Note that an assumption of the McDonald–Kreitman test is the neutrality of silent mutations which we suspect to be sometimes violated in our data (see below). Another test of departure from neutral evolution that

compares divergence with polymorphism for several loci at a time, the HKA test (Hudson, Kreitman, and Aguadé, 1987) was performed on pairwise combinations of genes in order to test whether differences in tree topologies reflect differing levels of evolution rates or selective processes. In this latter test, a locus under positive selection is supposed to have a polymorphism/divergence ratio smaller than that of a neutral one. We used the average of observed fixed differences per site (D_{XY}) and the number of segregating sites in both species to perform a HKA test in direct mode using the DNAsp 4.10 software.

3. Results

3.1. Size and composition of the PCR fragments

The proportion of coding sequences greatly varied between the three portions of genes. While the mitochondrial Cytochrome-*C* oxydase I fragment was only made of 176 codons, the two other nuclear genes corresponded to a single complete intron flanked by two exonic truncated regions of varying sizes. The amplified portion of the Lysozyme gene corresponded to the third large intron of 1198 bp flanked by two small truncated exonic regions of c.a. 9 bp and 90 bp (limited by the primer positions). The intron exhibited a clear diagnostic signature for the two species due to the length of a poly-T region and three indels of one base pair. The portion of the EF1 α used in this study also corresponded to the first part of the second intron of the gene (from base 128 to base 539) which overall size was about 627 bp. The remaining part of the amplified fragment (position 1 to 127) corresponded to exon 2. Species were clearly identified from each other by a diagnostic indel that varies in length from 11 bp to 5 bp for *B. azoricus* and *B. thermophilus*, respectively. The length of the locus corresponds to the PCR product without the primers sequences.

3.2. Correction for artifactual singletons

Inherent artifactual mutations associated with molecular methods can generate singleton mutations that can potentially bias the analyses and neutrality tests we performed. For example, the number of polymerase errors has been previously

Table 2
Summary statistics of nucleotidic polymorphism for mitochondrial and nuclear sequences in two bathymodioline species

Locus	Species	<i>n</i>	No. sites	S	Fixed mutations	Ka	Ks	Coding region		Non-coding π int	<i>D</i>	Hd	Tajimas's <i>D</i>
								π Nsyn	π syn				
COI	Bt	48	528	20	47	0.022	0.391	0.00082	0.02516		0.1198	0.857	-0.4883
	Ba	44	528	26				0.00070	0.02145			0.904	-1.6176
Lyo	Bt	39	1297	69	44	0.023‡	0.093‡	0.00698	0.00456	0.00714	0.0489	0.989	-1.7163
	Ba	44	1297	117				0.01001	0.01148	0.01466		1.000	-1.4453
EF α	Bt	70	528	40	9	0.008‡	0.050‡	0.00178	0.00106	0.00258	0.0198	0.643	-2.7759***
	Ba	51	534	34				0.00163	0.00432	0.00307		0.704	-2.6793***

Bathymodiolus thermophilus (Bt), and *B. azoricus* (Ba). Locus abbreviations are given in the Materials and methods. *n*, number of alleles sampled; S, segregating sites; Fixed mutations, mutations fixed between the two populations; π Nsyn, π syn and π silent represent nucleotide diversities on exonic and intronic regions, respectively; *D*, divergence derived from the Kimura-2 parameters model of substitutions; Hd, haplotype diversity. Tajima's *D* values (***) $P < 0.001$. ‡: Ka and Ks values obtained from pairwise alignments of longer partial cDNA between *B. azoricus* and *B. thermophilus* (115 codons for EF1 and 163 codons for Lyso).

estimated by Palumbi and Baker (1994) to be 1–2 polymerase errors per 1000 bases of cloned PCR products. It was therefore necessary to watch carefully to identify and eliminate such artifacts. With the MR-cloning protocol (Bierne et al. 2007), we had the possibility to capture a high number of sequences for each individual. Artifactual singletons could usually be removed by restricting the data analysis to multiple captured sequences (Bierne et al. 2007). However, this procedure would result in too few sequences kept for the analysis. We therefore decided to use the following correction: we suppressed singleton mutations when more than two alleles were obtained from one individual, assuming that these variants were due to amplification errors, and did not reflect true genetic variation. This manipulation also potentially removes true polymorphisms but it should be conservative in the present study where we observed a trend for an excess of rare variants. For example Tajima's *D* values should be slightly biased upward.

3.3. Gene polymorphisms

Gene diversities and numbers of segregating sites for coding and non-coding sequences are presented for the two species in Table 2. High nucleotide diversity seemed to be a general trend for the two *Bathymodiolus* species. However, nucleotide diversity greatly varied between genes. The overall nucleotide diversity was found to be three to five times lower in the EF1 α as compared to the two other genes. In addition, π values at this locus were very similar in the intronic (412 bp) and exonic (127 bp) regions for both species.

3.4. Species divergence

The divergence rate also greatly varied among genes, from 0.02 for EF1 α to 0.12 for COI (Table 2). Most of the divergence occurred at silent sites in both nuclear genes and on synonymous

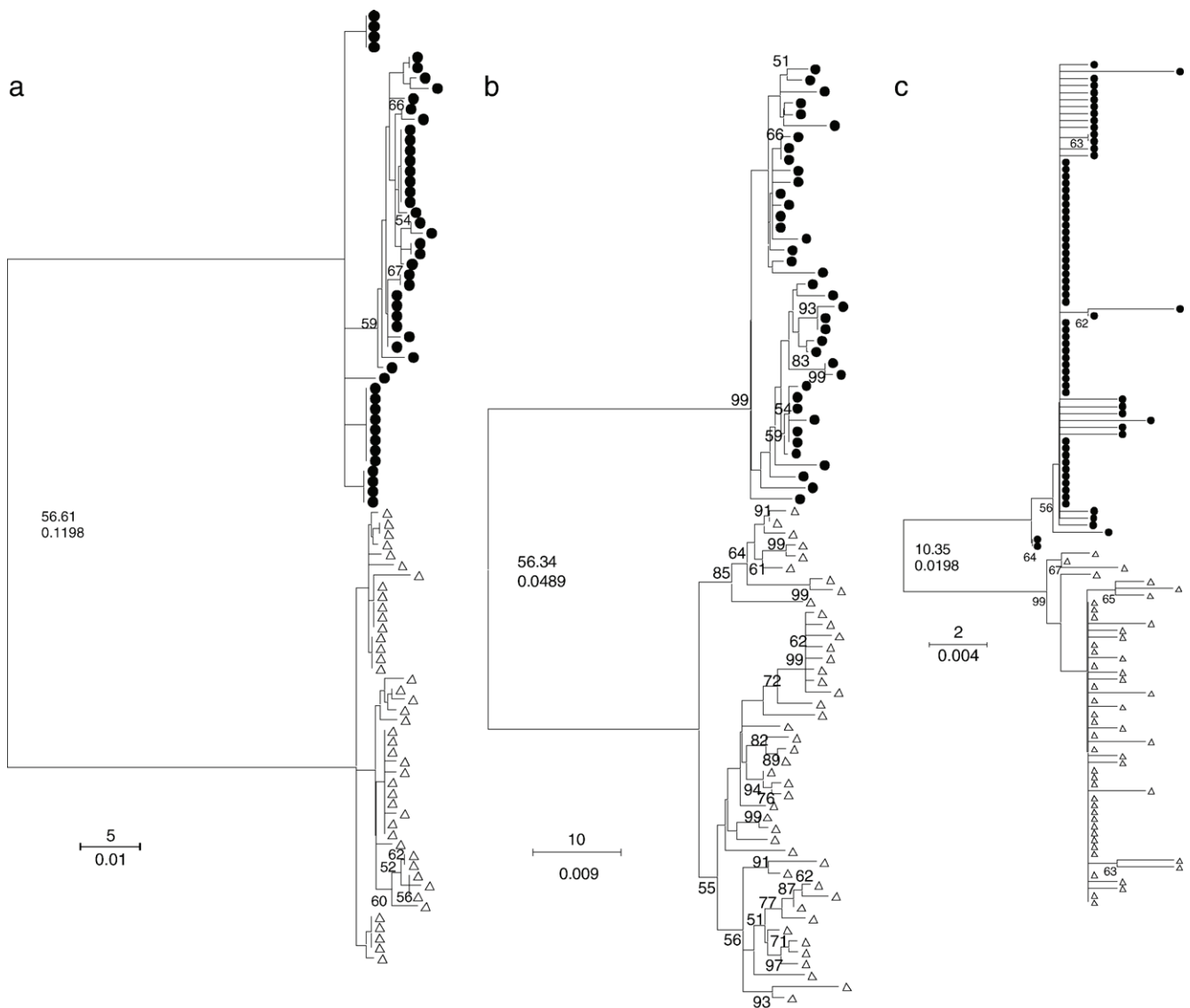


Fig. 1. Neighbour-joining trees, based on the number of differences, between *Bathymodiolus thermophilus* (black circle), and *Bathymodiolus azoricus* (white triangle) for mtCO1 (a), Lysozyme (b), and EF1 (c). Top values are the number of differences; the bottom values the K2P equivalent distances. Bootstrap proportions upper than 50% are presented (percentage of 1000 replicates).

Table 3
Results of the Hudson–Kreitman–Aguade test, across the three loci

		Locus	
		COI	Lyso
Locus 2	EF1	0.1595 NS	0.6660 NS
	Lyso	0.2504 NS	

NS, Non-significant deviation from the neutral expectations.

sites (more than 80%) in the mitochondrial gene. The divergence observed at the EF1 α locus was very low compared to Lysozyme. This suggests an unexpectedly slow rate of evolution. The estimated divergence was about three times lower than that obtained for the Lysozyme and 6 times lower than that of the mitochondrial gene. The difference remained unchanged when restricting the comparison to intronic regions which rules out the effect of a strong selective constraint on the protein.

3.5. Gene genealogies

Gene genealogies within each species were traced back based neighbour-joining trees using the observed number of nucleotide differences (Fig. 1). The trees indicated that the two species are well-separated and represents good outgroups for each other. The shapes of the Lysozyme and COI trees were quite similar with 2–3 clades that were not supported by bootstrap values. The mitochondrial gene, which is more sensitive to demographic changes because of its lower effective population size (N_e) did not seem strongly affected by any departure from demographic equilibrium. EF1 α however displayed very distinct genealogies with short branch lengths that coalesce rapidly near the most recent common ancestor. In contrast, gene genealogies observed at the COI and Lysozyme locus were more structured with several allelic lineages. Divergence between allelic clades explained most of the nucleotide diversity that can be found at this locus for both species (Table 2) and could not exclude the trace of a population admixture.

3.6. Departure from neutrality

Great negative and significant D values associated with Tajima's D tests provided evidence for a departure from neutrality in the distribution of mutations at the EF1 locus in both species. These values remained highly significant when the test was applied solely on the intronic part of the sequences. In contrast, D values were not significant for the other two genes. The stochastic variance of D can be high. Although Tajima's D values proved to be significant at the EF1 α locus, we did not prove that they were significantly different from the values observed at the other two loci, be they non-significantly different from zero at the 5% threshold. However, genealogies produced in two different species are expected to be independent replicates. Therefore, the lowest Tajima's D is not expected to be necessarily observed at the same locus in both species. Here, Tajima's D values observed at the locus EF1 α were strongly negative for both species with values nearly similar in the two independent genealogies. To strengthen our findings, we used the software HKA that incorporates the

divergence process in coalescent simulations. Simulations confirmed that negative values obtained at the EF1 α locus were highly significant (100% of simulated values greater than the observed value for both species). We therefore conclude to be very unlikely that demographic effects could have produced a departure of a similar magnitude at the exact same locus in both species.

The MK test was also performed, assuming silent (synonymous + non-coding) mutations as neutral references to infer the possible effect of selection on non-synonymous mutations. None of the three loci did depart from the neutral expectation using Fisher's exact tests ($1 < P < 0.351$, NS). There was however a lack of power in the application of the test owing to the very few mutations fixed between species in the coding regions.

Results of the HKA test from pairwise combination of genes are shown in Table 3. None of the sequenced loci deviated from neutral expectations.

4. Discussion

We analysed *B. azoricus* and *B. thermophilus* genomes at three loci. The use of marine bivalves as a target to investigate molecular polymorphism is of crucial importance because they may be considered as some of the most highly polymorphic organisms on Earth. They should therefore be very useful to detect putative adaptive or slightly deleterious polymorphism from large datasets of nuclear sequences. This preliminary multilocus analysis allows us to discuss the following points: (i) a marked effect of demography on the pattern of DNA diversity can be discarded in both species; (ii) the indirect effect of recurrent hitchhiking events does not easily explain the departure from neutrality observed at one intron of the EF1 α gene; (iii) the direct effect of negative selection on intronic polymorphisms may be a more appropriate explanation; and (iv) combining polymorphism data of several closely related species is recommendable in genome scan experiments.

4.1. No evidence for demographic changes in the recent past

Star-shaped genealogies are generally produced if a population has gone through a bottleneck. However, in contrast to selective sweeps, demography is expected to affect genetic diversity at every locus (Page and Holmes, 2000). In the present study, only the EF1 α locus departed highly significantly from the neutral expectation at mutation-drift (demographic) equilibrium. Signature of a strong demographic event (bottleneck or recent expansion) is therefore unlikely but may be partially masked by population admixture if any. Furthermore, the effective population size of mitochondrial genes is expected to be smaller on average than the one of nuclear genes because mtDNA is haploid, often inherited uniparentally and is more affected by selection on one of its genes because of the absence of recombination (Ballard and Rand, 2005; Bazin et al., 2006). As a consequence mtDNA is expected to be more sensitive to a departure from demographic equilibrium (Hudson and Turelli, 2003). The fact that the mtCOI locus did not depart from the

standard coalescent could be taken as a solid argument to refute a past demographic change. The cause for the lack of diversity at the EF1 α locus should therefore imply selection, either indirectly because of selective sweeps or directly as a consequence of slight purifying selection on observed polymorphisms.

4.2. The indirect selection hypothesis requires two equally old selective sweeps in the two distinct species

We observed low levels of sequence polymorphism at the EF1 α locus within each of the two species. This finding greatly differs from what we observed in the two other genes, and especially the Lysozyme intron. The NJ tree at locus EF1 α shows a star-shaped gene genealogy in the two *Bathymodiolinae* species together with significant negative Tajima's D . In addition, the length of the tree leaves – i.e. the number of singletons – was similar in the two species. These observations could suggest that an advantageous mutation swept to fixation nearly simultaneously in both species after speciation. The age of a sweep in units of mutational time, τ , can be estimated from the mismatch distribution using the method of Rogers (1995). We estimated τ to be ~ 1 in *B. azoricus* and ~ 0.5 in *B. thermophilus*, two very similar values. Selective sweeps have been found in a variety of eukaryote genomes and may affect introns from the same gene differently (Nachman and Crowell, 2000). However, selective sweeps may be expected to have a weak probability to occur simultaneously within two distinct species and would also have had a profound impact on the HKA test. Indeed, the basis of the HKA test is that selective sweeps decrease diversity from the expectation at mutation-drift equilibrium; an expectation that can be assessed with divergence data that essentially depends on the mutation rate. The HKA test did not reveal any departure from neutrality for the three genes. It means that the EF1 α locus did not exhibit a larger polymorphism to divergence ratio than the other two genes — the diversity observed is in accordance with a low effective mutation rate at the locus either because mutation is lower at this locus or because of selective constraints. Although the hypothesis of direct selection on intron polymorphisms is usually discarded, the arguments against the selective sweep hypothesis together with the observation that the intron of the EF1 α gene is evolving at an abnormally slow rate, led us to seriously consider this hypothesis.

4.3. The second intron of the EF1 α gene is under slight purifying selection

Introns generally exhibit higher levels of polymorphism and divergence than exonic regions because they are subjected to weaker selective constraints (Li and Graur, 1991). However, in some specific cases, introns may show to a low level of variability due to selective constraints associated with functional and/or structural specificities (Jaruzelska et al. 1999; Young and dePamphilis, 2000; Parsch, 2003). We argue that polymorphic mutations observed at the second intron of the EF1 α are slightly deleterious in the deep-sea hydrothermal vent bivalve *Bathymodiolus*. The locus EF1 α greatly differs from the locus Lysozyme in the number of fixed mutation in the non-

coding region. The number of fixed mutations in the intron of the EF1 α locus is more than three times higher than for the lysozyme intron (1 fixation for 100 bp at EF1 α as opposed to 3.7 at the Lysozyme). Similarly, the non-synonymous/synonymous divergence ratio (Ka/Ks) obtained from cDNA was twice lower for the EF1 α coding sequence (0.12) than for the Lysozyme one (0.24) in accordance with a stronger selective constraint on the protein. At the same time, the within species nucleotide diversity of the intronic region is 3 to 5 times lower at the locus EF1 α as compared to that of the Lysozyme. Purifying selection is known to greatly affect both intraspecific polymorphism and interspecific divergence rates. Tajima (1989a,b) demonstrated that when purifying selection is sufficiently weak to allow deleterious mutations to reach appreciable frequencies (say $-10 < N_e s < -1$), it greatly affects the D statistic because although such deleterious mutations are observed in the polymorphism they segregate at low frequency. This may thus explain why we found significant negative D values at the EF1 α locus but no departure from neutrality using the HKA test — the number of substitutions accumulated in the divergence is low because slightly deleterious mutations do not reach fixation. Furthermore, similarly strong negative D values were independently observed in both species which is unlikely to be explained simply by stochasticity. Reduced levels of intraspecific polymorphism and interspecific divergence of introns have been already described in *D. melanogaster* (Leicht et al., 1995) and in plastid genes (Young and dePamphilis, 2000). The former authors interpreted this lack of genetic variation as a result of direct selective effect in which deleterious mutations are purged from the intronic sequence because of its alternative pre-mRNA splicing function in the gene. Regulation and splicing functions have been widely reported for intronic regions but occur in general on the first or the last intron of the gene (Green, 1986; Wakabayashi-Ito and Shigekazu, 1994). Regulation may thus represent an interesting hypothesis to interpret our finding mostly because the 5' flanking region and intron 1 of the elongation factor 1 α are known to be under strong selective constraints due to their promoting function in the gene translation (Wakabayashi-Ito and Shigekazu, 1994). Selective constraints on introns may also be associated with their sizes (Castillo-Davis et al., 2002; Lynch, 2002) or the stability of their secondary structure like the preservation of stem structures (Stephan and Kirby, 1993). In this context, it is noteworthy that the second intron of EF1 α in *Bathymodiolus* is much larger (627 bp) than its orthologs in the blue mussel *Mytilus edulis* (272 to 295 bp), *Mytilus trossulus* (230 bp), and the oyster *Crassostrea gigas* (89 bp) and seems to evolve at a much lower rate (N. Bierne, unpublished data).

4.4. Teasing apart the direct effect of negative selection and the indirect effect of positive selection from independent genealogies of closely related species

We obtained polymorphic and divergence information by contrasting two sets of sequences from two distinct species and not by using a single sequence for the outgroup, as is often the case in genome scan studies. This allowed us to get a better

picture of how genes are evolving and especially to discriminate between the direct (negative) and indirect (positive) action of selection in shaping polymorphism of non-coding and coding portions in nuclear genes. Indeed, juxtaposing polymorphisms of two closely related species represented a powerful tool to refute the hypothesis of a selective sweep at the EF1 α intron. With only polymorphism data for a single species we would probably have concluded that a selective sweep occurred in the chromosomal vicinity of the gene without challenging the usual assumption of neutrality on non-coding polymorphisms. This observation therefore poses the question of the bias generated by direct negative selection in genome scan experiments, especially when applied to non-model species like ours.

5. Conclusions

To date, very little information on nuclear polymorphisms has been collected on bivalves and, especially deep-sea hydrothermal vent organisms. The concomitant analysis of polymorphism and divergence data at three loci in two *Bathymodiolus* species produced the first evaluation of nucleotide diversity in bivalves inhabiting “extreme” and ephemeral habitats. Levels of polymorphism (and mutation rates) were high and not very different from what has previously been reported for other invertebrates (Li and Graur, 1991; Leicht et al., 1995; France et al., 1999; Glinka et al., 2003). The hypothesis that vent organisms should have small effective population sizes in response to the tremendous level of habitat instability at all spatial scales does not seem to hold in the case of *Bathymodiolus*. Although bottlenecks may occur locally due to the dynamics of the vent system, they are not sufficient to decrease enough the effective population size of both *Bathymodiolus* species to impact their nucleotide diversities. However, the second intron of the EF1 α locus exhibited low nucleotide diversities in both species which could either result from recurrent selective sweeps in the chromosomal vicinity of the gene or direct purifying selection acting on polymorphic intronic mutations themselves. We nonetheless provided argument in favour of the second hypothesis — slight purifying selection on the observed polymorphism. Our conclusions are however only based on three genes. More nuclear loci are now needed to explore how nucleotide diversity and divergence have evolved in these peculiar bivalve species and to test whether selective sweeps and purifying selection occur more widely in such an extreme and unpredictable environment.

Acknowledgments

We gratefully acknowledge the captains and crew of the NO ‘L’Atalante’, and the pilots and co-pilots of the ROV Victor 6000 and ‘Nautile’ groups for collecting material. We are particularly grateful to P-M Sarradin, chief scientist of the ATOS’01 cruise, who allowed us to conduct this work. We also want to thank an anonymous referee for his useful criticisms on an earlier version of the manuscript. This work was funded in part by the Marine Genomics Europe NoE (Fish & Shellfish node) and the French national programme GIS “Génomique Marine” (The Bivalvomix Project, coord. N. Bierne).

References

- Akashi, H., 1995. Inferring weak selection from patterns of polymorphism and divergence at “silent” sites in *Drosophila* DNA. *Genetics* 139, 1067–1076.
- Akashi, H., 1996. Molecular evolution between *Drosophila melanogaster* and *D. simulans*: reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. *Genetics* 144, 1297–1307.
- Andolfatto, P., 2005. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* 437, 1149–1152.
- Bachali, S., et al., 2002. Phylogenetic analysis of invertebrate lysozymes and the evolution of lysozyme function. *J. Mol. Evol.* 54, 652–664.
- Ballard, J.W.O., Rand, D.M., 2005. The population biology of mitochondrial DNA and its phylogenetic implications. *Ann. Rev. Ecol. Syst.* 36, 621–642.
- Bazin, E., Glémin, S., Galtier, N., 2006. Population size does not influence mitochondrial genetic diversity in animals. *Science* 312, 570–572.
- Bierne, N., et al., 2007. Mark-recapture cloning: a straightforward and cost-effective cloning method for population genetics of single copy nuclear DNA sequences in diploids. *Mol. Ecol. Notes* 7, 562–566.
- Castillo-Davis, C.I., Mekhedov, S.L., Hartl, D.L., Koonin, E.V., Kondrashov, 2002. Selection for short introns in highly expressed genes. *Nat. Genet.* 31, 415–418.
- Chevaldonné, P., Jollivet, D., Desbruyeres, D., Lutz, R.A., Vrijenhoek, R.C., 2002. Sister-species of eastern Pacific hydrothermal vent worms (Ampharetidae, Alvinellidae, Vestimentifera) provide new mitochondrial *COI* clock calibration. *Cah. Biol. Mar.* 43, 367–370.
- Ejiri, S.-I., 2002. Moonlighting functions of polypeptide elongation factor 1: from actin bundling to zinc finger protein R1-associated nuclear localisation. *Biosci. Biotechnol. Biochem.* 66, 1–21.
- Eyre-Walker, A., Woolfit, M., Phelps, T., 2006. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* 173, 891–900.
- Fay, J.C., Wu, C.-I., 2001. The neutral theory in the genomic era. *Curr. Opin. Genet. Dev.* 11, 642–646.
- France, S.C., Tachino, N., Duda, T.F.J., Shleser, R.A., Palumbi, S.R., 1999. Intraspecific genetic diversity in the marine shrimp *Penaeus vannamei*: multiple polymorphic elongation factor-1 α loci revealed by intron sequencing. *Mar. Biotechnol.* 1, 261–268.
- Fu, Y.-X., Li, W.-H., 1993. Statistical tests of neutrality of mutations. *Genetics* 133, 693–709.
- Galtier, N., Depaulis, F., Barton, N.H., 2000. Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. *Genetics* 155, 981–987.
- Glinka, S., Ometto, L., Mousset, S., Stephan, W., DeLorenzo, D., 2003. Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. *Genetics* 165, 1269–1278.
- Green, M.R., 1986. PRE-mRNA splicing. *Ann. Rev. Gene.* 20, 671–708.
- Haddrill, P.R., Thornton, K.R., Charlesworth, B., Andolfatto, P., 2005. Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Res.* 15, 790–799.
- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.
- Hudson, R.R., Kreitman, M., Aguadé, M., 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116, 153–159.
- Hudson, R.R., Turelli, M., 2003. Stochasticity overrules the “Three-Times Rules”: genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution* 57, 182–190.
- Iwasaki, H., et al., 2006. Evolutionary relationships of deep-sea mussels inferred by mitochondrial DNA sequences. *Mar. Biol.* 149, 1111–1122.
- Jaruzelska, J., Zietkiewicz, E., Labuda, D., 1999. Is selection responsible for the low level of variation in the last intron of the *ZFY* locus? *Mol. Biol. Evol.* 16 (11), 1633–1640.
- Jollivet, D., 1996. Specific and genetic diversity at deep-sea hydrothermal vents: an overview. *Biodiver. Conserv.* 5, 1619–1653.
- Jollivet, D., Chevaldonné, P., Planque, B., 1999. Hydrothermal-vent alvinellid polychaete dispersal in the eastern Pacific. 2. A metapopulation model based on habitat shifts. *Evolution* 53, 1128–1142.
- Jones, W.J., Won, Y.-J., Mass, P.A.Y., P.J.S., Lutz, R.A., Vrijenhoek, R.C., 2006. Evolution of habitat use by deep-sea mussels. *Mar. Biol.* 148, 841–851.

- Kaplan, N.L., Hudson, R.R., Langley, C.H., 1989. The “Hitchhiking Effect” revisited. *Genetics* 123, 887–899.
- Kenk, V.C., Wilson, B.R., 1985. A new mussel (*Bivalvia*, *Mytilidae*) from hydrothermal vents in the Galapagos Rift zone. *Malacologia* 26, 253–271.
- Kimura, M., 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, London.
- Kreitman, M., 1983. Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature* 304, 412–417.
- Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* 5 (2), 150–163.
- Leicht, B.G., Muse, S.V., Hanczyc, M., Clark, A.G., 1995. Constraints on intron evolution in the gene encoding the myosin alkali light chain in *Drosophila*. *Genetics* 139, 299–308.
- Lewontin, R.C., 1991. Electrophoresis in the development of evolutionary genetics: milestone or millstone? *Genetics* 128, 657–662.
- Li, W.-H., Graur, D., 1991. *Fundamentals of Molecular Evolution*. Sinauer Associates, Sunderland, MA.
- Li, H., Stephan, W., 2006. Inferring the demographic history and rate of adaptive substitution in *Drosophila*. *PLoS Genet.* 10, 1580–1589.
- lynch, M., 2002. Intron evolution as a population-genetic process. *Proc. Natl. Acad. Sci. U. S. A.* 99, 6118–6123.
- Malatesta, F., Antonini, G., Sarti, P., Brunori, M., 1995. Structure and function of a molecular machine: cytochrome *c* oxidase. *Biophys. Chem.* 54, 1–33.
- Maynard Smith, J., Haigh, J., 1974. The hitchhiking effect of a favorable gene. *Genet. Res.* 23, 23–35.
- McDonald, J.H., Kreitman, M., 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351, 652–654.
- Moriyama, E.N., Powell, J.R., 1997. Synonymous substitution rates in *Drosophila*: mitochondrial versus nuclear genes. *J. Mol. Evol.* 45, 378–391.
- Nachman, M.W., Crowell, S.L., 2000. Contrasting evolutionary histories of two introns of the Duchenne muscular dystrophy gene, *Dmd*, in humans. *Genetics* 155, 1855–1864.
- Olu-Le roy, K., Von Cosel, R., Hourdez, S., Carney, S.L., Jollivet, D., in press. Amphi-Atlantic cold-seep *Bathymodiolus* complexes of species across the Equatorial belt. *Deep-Sea Res.* accepted, manuscript number: DSR1-D-06-00062R1.
- Page, R.D.M., Holmes, E.C., 2000. *Molecular Evolution: a Phylogenetic Approach*. Blackwell Science.
- Palumbi, S.R., Baker, C.S., 1994. Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Mol. Biol. Evol.* 11, 426–435.
- Parsch, J., 2003. Selective constraints on intron evolution in *Drosophila*. *Genetics* 165, 1843–1851.
- Powell, M.A., Somero, G.N., 1986. Adaptations to sulfide by hydrothermal vent animals: sites and mechanisms of detoxification and metabolism. *Biol. Bull.* 171, 274–290.
- Rogers, A.R., 1995. Genetic evidence for a Pleistocene population explosion. *Evolution* 49 (4), 608–615.
- Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X., Rozas, R., 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19, 2496–2497.
- Sambrook, J., Russel, D.W., 2001. *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory Press.
- Stephan, W., Kirby, D., 1993. RNA folding in *Drosophila* shows a distance effect for compensatory fitness interactions. *Genetics* 135, 97–103.
- Tajima, F., 1989a. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595.
- Tajima, F., 1989b. The effect of change in population size on DNA polymorphism. *Genetics* 123, 597–601.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- Wakabayashi-Ito, N., Shigekazu, S., 1994. Characterization of the regulatory elements in the promoter of the human elongation factor-1 α gene. *J. Biol. Chem.* 269 (47), 29831–29837.
- Young, N.D., dePamphilis, C.W., 2000. Purifying selection detected in the plastid gene *matK* and flanking ribozyme regions within a group II intron of nonphotosynthetic plants. *Mol. Biol. Evol.* 17 (12), 1933–1941.

Spatial and temporal dynamics of reproduction and settlement in the Pompeii worm *Alvinella pompejana* (Polychaeta: Alvinellidae)

Baptiste Faure^{1,*}, Pierre Chevaldonné², Florence Pradillon³, Eric Thiébaud¹,
Didier Jollivet¹

¹Station Biologique de Roscoff, UMR 714CNRS—Université Pierre & Marie Curie-Paris 6, Adaptation et diversité en milieu marin, B.P. 74, Place Georges Teissier, 29682 Roscoff Cedex, France

²Station Marine d'Endoume, Centre d'Océanologie de Marseille, UMR 6540 CNRS—Université Méditerranée, Diversité, Evolution et Ecologie Fonctionnelle Marine, Rue de la Batterie des Lions, 13007 Marseille, France

³Université Pierre et Marie Curie-Paris 6, UMR 7138 CNRS-UPMC-MNHN-IRD, Systématique, Adaptation et Evolution, 7 quai Saint-Bernard, 75252 Paris Cedex 05, France

ABSTRACT: The tubicolous polychaete *Alvinella pompejana* (Desbruyères & Laubier) inhabits deep-sea hydrothermal vents. To assess its reproductive and recruitment strategies in response to such extreme conditions, populations of *A. pompejana* were sampled at various spatial scales: among hydrothermal vents within the 13°N/East Pacific Rise vent field and among vent sectors along the southern East Pacific Rise. Additional samples were collected at 1 vent 31 d apart to determine short-term temporal variations. Such a sampling strategy allowed us to conduct an integrated study of both the spatial variation of the reproductive dynamics (i.e. fecundity, oocyte distribution) and the population structure of this polychaete. Despite a 1:1 sex ratio for each population, a male-biased sex ratio was found at sexual maturity, which could maximise reproductive success in this spermathecae-bearing species. Oocyte size distributions in mature females distinguished 5 maturity stages related to the presence of 1 to 5 distinct oocyte size groups. Coelomic fecundity, which is among the highest known in polychaetes (max. 978 000 oocytes), depends on the maturity stage, but not on the size of the female. The analysis of oocyte distributions strongly highlighted that reproduction is semi-continuous. Mature oocytes are however discretely recruited in the gonoducts from a massive stock of coelomic oogonia and spawned in pulses that may be synchronised between reproductive females. Size-frequency distributions of organisms displayed polymodal structure, suggesting discontinuous recruitment that may be locally affected by the vent dynamics or variations in larval supply. The discrepancy between semi-continuous reproduction and a polymodal size-frequency distribution may be explained by: (1) a limited amount of recruits within populations, (2) rapid growth during the early benthic phase, and/or (3) an accumulation of individuals of different ages in the adult cohort. Nearly continuous production of gametes and storage of mature oocytes and/or sperm seem to be a general trend selected in hydrothermal-vent polychaete species.

KEY WORDS: Hydrothermal vents · Sex ratio · Oogenesis · Semi-continuous reproduction · Population structure · East Pacific Rise

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

Reproduction mode is a crucial element of population dynamics on which recruitment efficiency, population growth rate, or extinction rate directly depend.

Generally speaking, continuous recruitment implies continuous reproduction, while discontinuous recruitment often implies discontinuous but synchronised reproduction. Polychaete annelids display a large variety of reproduction modes due to the simplicity and the

great phenotypic plasticity of their reproductive system and also to their adaptive ability to live in contrasting marine habitats (Wilson 1991, Giangrande 1997). According to the interaction between environmental factors and reproduction, polychaetes can be divided into 4 main reproductive patterns (Olive 1984): (1) semelparous species that breed only once during their life, with a high degree of intra-population synchrony and reproduction highly controlled by endogenous factors; (2) iteroparous species that reproduce more than once during their life time, but at an annual interval, with a high degree of within-population synchrony; in these forms, reproduction is governed by a coupling between internal and external factors; (3) iteroparous species with no within-population synchrony and reproduction controlled by endogenous factors; and (4) semi-continuous species with continuous reproduction where short-term intra-population synchrony may be found.

Polychaete reproduction is even more intricate as gametogenesis may be more or less rapid according to species (Eckelbarger 1983). While vitellogenesis is often slow and takes a few months for most semelparous or iteroparous species, it can be fast and achieved in <2 wk for semi-continuous or continuous breeders. Furthermore, gamete size and the way gametes are produced and emitted are also very diverse. Although polychaetes usually exhibit external fertilisation, some use internal fertilisation with or without sperm storage (Zal et al. 1994, Jollivet et al. 2000, Pradillon & Gaill 2003). Egg size and gamete mating appear to be under great phylogenetic, morphological and ecological constraints. As observed for numerous polychaete families, species of small size are not able to invest as much energy in egg production in a single breeding episode as a species of larger size, and exhibit lower fecundity, larger oocytes and generally a continuous reproduction pattern (Schroeder & Hermans 1975, McHugh 1993). While life-history traits may be the direct evolutionary consequence of selective pressures, total fecundity or egg size are also influenced by ecological factors such as 'patch' size and feeding resources associated with a given habitat. This is particularly relevant for harsh marine environments like deep-sea hydrothermal vents where environmental conditions are 'extreme' and highly variable.

Rapid changes in the faunal composition of hydrothermal-vent communities are the reflection of the high instability of the vent emissions and the great level of habitat patchiness (Copley et al. 1997, Shank et al. 1998, Sarrazin et al. 1999). Hydrothermal vents are often extremely dynamic, which can lead to population extinctions associated with recurrent colonisations of the new emergent emissions (Jollivet 1996). Such spatio-temporal variability may thus have a profound

impact on the organism life-history traits and over time could have promoted specific adaptive modes of reproduction and dispersal in vent animals. Only few studies have been conducted on the reproductive and recruitment strategies of hydrothermal-vent polychaetes. They have focused either on families known from other marine environments (e.g. Ampharetidae, Polynoidae) (McHugh 1989, McHugh & Tunnicliffe 1994, Jollivet et al. 2000) or on endemic families such as Alvinellidae or Siboglinidae (Zal et al. 1995, Copley et al. 1999, Thiébaud et al. 2002, Pradillon et al. 2005). They all showed peculiar adaptive convergence on reproductive mode and larval development. Indeed, most of them display continuous oogenesis (McHugh 1989, Copley et al. 2003, Pradillon et al. 2005), organs for oocyte storage with spermathecae and thus internal fertilisation of oocytes (Zal et al. 1994, Jollivet et al. 2000, Hilario et al. 2005). These species are also characterised by a high fecundity relative to their body size and produce mature oocytes of great size consistent with the hypothesis of lecithotrophic or direct development (McHugh & Tunnicliffe 1994, Zal et al. 1995, Jollivet et al. 2000).

Living on hydrothermal vent chimneys throughout the East Pacific Rise (EPR) from 32°S to 21°N, the Pompeii worm *Alvinella pompejana* has certainly been one of the most studied species over the last few years in terms of reproductive biology and population structure. It is a gonochoric species with extra-ovarian gametogenesis (Pradillon & Gaill 2003). During pairing, sperm is transferred and stored inside females within spermathecae after pseudocopulation. This mode of sperm transfer is thought to increase the mating probability of gametes, and thus reproductive success (Zal et al. 1994). According to egg size, development is supposed to be lecithotrophic or direct, although larval stages have never been collected. Thus, its reproductive system, egg size and sperm morphology are more consistent with restricted dispersal capabilities (Chevaldonné et al. 1997, Jollivet et al. 1999). However, its reproduction and recruitment dynamics remain poorly documented because of small sample sizes and sampling coverage, lack of temporal series of collections, and inability to keep animals alive after recovery on board. Recently, only 1 study showed different stages of female maturity within single local samples, therefore suggesting continuous gametogenesis at the scale of a local population (Pradillon et al. 2005). Conversely, *A. pompejana* shows signs of the polymodal population structure associated with discontinuous recruitment that suggest the synchrony of individuals for at least 1 stage of the life cycle (Jollivet 1993).

To better understand how reproduction and recruitment dynamics affect population persistence of polychaetes associated with highly instable and fragmented

hydrothermal habitats, the aims of the present study are: (1) to describe the successive stages of *Alvinella pompejana* reproduction dynamics, (2) to assess the relationships between reproductive and recruitment patterns over different spatial scales (local vs. regional scale) and (3) to appreciate short-term variations in oocyte and individual sizes. This work uses statistical analyses performed on demographic and reproductive structures of *A. pompejana* coming from a large series of samples collected over both the northern East Pacific Rise (NEPR) and the southern East Pacific Rise (SEPR).

MATERIALS AND METHODS

Study area and sampling. Samples of the deep-sea hydrothermal vent polychaete *Alvinella pompejana* were collected during 3 oceanographic cruises, Hot'96 (February to March 1996), Phare'02 (April to June 2002) and Biospeedo'04 (April to May 2004), on board the RV 'Nadir' and RV 'Atalante' with the hydraulic arm of either the manned French submersible 'Nautile' (Hot'96 and Biospeedo'04) or the French ROV 'VICTOR6000'

(Phare'02). After collection, all worms were placed into an insulated basket until they reached the surface. The number of individuals collected and the vent locality's coordinates for of each sample are presented in Table 1.

During the Phare'02 cruise, *Alvinella pompejana* samples were collected at a local scale over several vent chimneys from 6 vent sites 0.1 to 4 km distant from each other at 13°N/EPR (Fig. 1). During the Biospeedo'04 cruise, sampling was performed at a regional scale from 8 disjunctive hydrothermal vent sectors that were 2 to 1650 km distant from each other over a large portion of the SEPR (Fig. 1). Two successive collections of *A. pompejana* were conducted on a single hydrothermal vent chimney at the Totem vent site (13°N/EPR) (Fig. 1) during the Hot'96 cruise. The worms were sampled in the same assemblage, at the same exact location, 31 d apart, to estimate changes in individual size and reproductive stages over a short time scale.

Habitat characteristics of *Alvinella pompejana* have recently been described by Desbruyères et al. (1998).

Biometry and sex ratio. After recovery on board, the gender of specimens was identified by observation of the genital pore, which is wider for females than for

Table 1. *Alvinella pompejana*. Date of collection, location and size of samples collected from vent chimneys at the northern and southern East Pacific Rise

Oceanographic survey	Name	Date	Location	Sample size
Northern East Pacific Rise				
Hot'96	Totem	16 Feb 1996	12° 48.80' N, 103° 56.500' W	48
		18 Mar 1996	12° 48.80' N, 103° 56.500' W	64
Total				112
Phare'02	Elsa	21 May 2002	12° 48.143' N, 103° 56.307' W	5
		26 May 2002	12° 48.143' N, 103° 56.307' W	39
		29 May 2002	12° 48.143' N, 103° 56.307' W	38
	Genesis	25 May 2002	12° 48.641' N, 103° 56.442' W	63
		Julie	25 May 2002	12° 49.058' N, 103° 56.571' W
	28 May 2002		12° 49.058' N, 103° 56.571' W	5
	Jumeaux	30 May 2002	12° 48.646' N, 103° 56.436' W	103
	Pulsar	27 May 2002	12° 49.842' N, 103° 56.802' W	40
	Parigo	19 May 2002	12° 48.589' N, 103° 56.423' W	36
		20 May 2002	12° 48.589' N, 103° 56.423' W	9
		22 May 2002	12° 48.589' N, 103° 56.423' W	6
		23 May 2002	12° 48.589' N, 103° 56.423' W	1
		25 May 2002	12° 48.589' N, 103° 56.423' W	4
	26 May 2002	12° 48.589' N, 103° 56.423' W	5	
	29 May 2002	12° 48.589' N, 103° 56.423' W	16	
	Total			
Southern East Pacific Rise				
Biospeedo'04	Last Hope	10 Apr 2004	7° 25.451' S, 107° 47.702' W	32
	Lucky Eric	4 May 2004	13° 59.209' S, 112° 28.956' W	57
	Rehu Marka	24 Apr 2004	17° 24.964' S, 113° 12.128' W	93
	Nadir	30 Apr 2004	17° 25.891' S, 113° 12.365' W	17
	Calvin	28 Apr 2004	17° 35.806' S, 113° 14.888' W	86
	Fromveur	21 Apr 2004	18° 25.964' S, 113° 23.349' W	132
	Near Jasmine	15 Apr 2004	21° 25.295' S, 114° 16.465' W	46
	Krasnov	14 Apr 2004	21° 33.604' S, 114° 17.884' W	32
Total				495

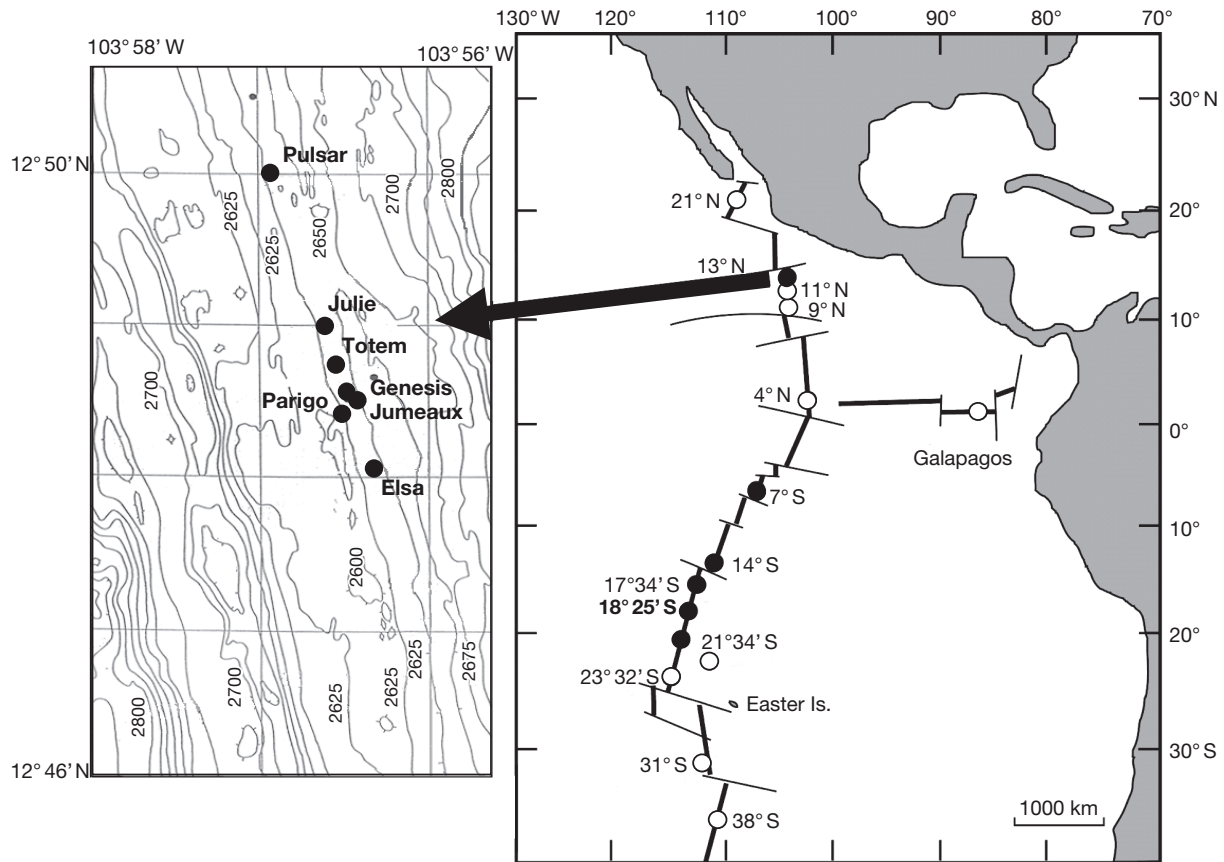


Fig. 1. Sampling sites for the deep-sea *Alvinella pompejana* worms used in our study (isobaths in m). ○: hydrothermal vent; ●: sampling sites. The 13°N region is magnified

males (Pradillon & Gaill 2003). To analyse demographic structure, the dorsal width of the 4th setigerous segment (S4) was used as an index of the individual size as previously shown by Jollivet (1993): $\log(\text{width}) = 2.45 + 0.32 \log(\text{dry weight})$ ($n = 30$, $r = 0.96$). The S4 segment is easily recognisable when sizing individuals, and is conveniently very rigid due to the presence of a large pair of hooks. All measurements were made to the nearest 0.05 mm using a calliper under a binocular microscope. The smallest individuals (<3 mm S4 width) were not easily and definitively assigned to a gender and have therefore been considered juveniles.

Sexual maturity, female fecundity and oocyte measurements. For each female, coelomic fluid was carefully removed by making a ventral incision to the body wall of the trunk to release the fluid following the recommendations of Olive (1970). Specimens were then cut open along the anterior–posterior axis, from the end of the body to the first setigerous segment. The body cavity was extensively washed out in order to collect all coelomic oocytes. The coelomic fluid with oocytes of each mature female was resuspended in

50 ml of a solution of 3% borate-buffered formalin in seawater. Oocytes (oogonia to mature oocyte) were counted from 2 oocyte suspension aliquots of 0.25 ml. In addition, spermiducts were also dissected on males collected during Biospedo'04 and checked for sperm in order to discriminate mature males and assess their size of first sexual maturity. In this study, the term 'mature female' was used for females undergoing oocyte maturation in the coelom but not yet able to reproduce, whereas the term 'reproductive female' was restricted to females displaying mature oocytes inside gonoducts. 'Immature females' correspond to females without any genital product.

Due to the discoid shape of oocytes, the maximum oocyte diameter crossing the micropyle was the only measurement taken into account. For females collected during the Hot'96 and Phare'02 cruises, the maximum diameter was measured using a microscope with a Spot video camera (Diagnostic Instrument Inc., Version 3 for Windows) from a subsample of 200 oocytes. This technique enabled us to digitise oocytes and then to measure their diameter after calibration (with an error of 2 μm). For Biospedo'04

samples, maximum oocyte diameters were measured to the nearest 5 μm using an eye-piece micrometer under a Leitz inverted microscope from a subsample of 200 oocytes.

Size-frequency histograms. S4 size-frequency histograms have been produced for each sample and each gender, while oocyte size-frequency histograms have been drawn for each mature female. The class interval (1 mm for S4 size and 10 μm for oocytes) was chosen in order to minimise the number of null classes and to take into account the measurement error (Jollivet 1993). All size-frequency histograms were then smoothed using a moving average of 3 classes to rule out non-significant peaks (Frontier & Pichod-Viale 1998). Modal decomposition of size-frequency histograms was done according to the method of Bhattacharya (1967). Mean sizes, SDs and proportions of each identified Gaussian component were then refined using the MIX 3.1 program package (Ichthus Data System; MacDonald & Green 1988). This method provides the best fit between theoretical admixtures of normal distributions and the observed one, by the maximum likelihood method.

We chose to use modal decomposition for the oocyte distribution, which, in most cases, resembles a log-normal distribution, in order to have an objective element to describe the stage of oocyte maturation. Because preliminary analyses showed nearly identical standard deviations between oocyte size groups within and between distributions of the 98 mature females, a second series of MIX decompositions was performed on the same histograms assuming equal modal variances under the assumption of constant oocyte maturation rate, which has been widely reported for polychaetes (Zal et al. 1995).

Statistical analyses. For each collected sample, several population descriptors have been calculated as follows: (1) sex ratio, (2) the proportion of mature males and females in the sample, (3) the proportion of oocyte maturation stages in the female population and (4) the size at first sexual maturity in males and females.

Departure from a sex ratio of 1:1 was tested for each sample and for each cruise collection using a standard χ^2 goodness-of-fit-test.

Size-frequency distributions between males and females from NEPR and SEPR, or between NEPR and SEPR for all individuals were compared using a standard Kolmogorov-Smirnov 2-sample test. Recruitment synchronisation within the 13°N/EPR vent field was evaluated using a Kruskal-Wallis test, because of the small sample sizes. Results were evaluated by comparing sample distributions. For the SEPR samples of larger size, mean sizes among vent sectors were compared using 1-way ANOVA with S4 size as the variable and vent sector as the fixed factor for each size

group. For each modal component, the variance within groups was calculated as the average of the variances within each sector, weighted by its df, whereas the variance among groups was estimated from the variance of means among sectors (Sokal & Rohlf 1995).

Synchronisation of oocyte maturation across vent sites (13°N/EPR) or across vent sectors (SEPR) was tested by performing 1-way ANOVAs for the first 2 oocyte modal components using the site or sector as a fixed factor and modal values of the spatial samples as the variable. Only the first 2 cohorts were chosen for these analyses, because they represented the cohorts with the highest proportions of oocytes and because most of the females presented at least these 2 cohorts. Within a vent or a sector, synchronisation of oocyte maturation was analysed using a 1-way ANOVA, with females as a random factor and oocyte size as a variable. A multiple linear regression was performed to test for size, latitudinal and time effects on the mode of the first oocyte cohorts. The significance of partial regression coefficients was tested by permutations according to Anderson & Legendre (1999).

To assess short-term changes in population structure and gametogenesis at the Totem site in 1996, modal sizes of individuals and modal diameter of oocytes were compared across a short time interval using a Student's *t*-test.

RESULTS

Population structure in *Alvinella pompejana*

Sex ratio, male and female distributions, and size at first sexual maturity

During Phare'02, 491 individuals were sampled on several vent chimneys at 13°N/EPR. The relative proportions of juveniles and the proportion of males and females, together with their average size, are presented in Table 2. In spite of the environmental instability of vent sites and the fragmentation of the habitat, there is no significant departure from a 1:1 sex ratio for either a given sample or the whole worm collection (χ^2 test; $p > 0.05$). Only 7.5% of the whole collection corresponds to individuals too small to be sexed (S4 width <3 mm) and referred to as juveniles. The minimum size at which females reached their first sexual maturity corresponds to an S4 width of 6 mm.

During the Biospedo'04 cruise, 495 ind. were sampled, sexed and measured (Table 2). The observed sex ratio in this collection was not found to be different from the expected value of 1:1 (χ^2 test; $p > 0.05$), although 2 samples departed significantly from this expectation at the 17°24' and 21°33'S vent sectors.

Table 2. *Alvinella pompejana*. Sex ratio and average size of males (m), females (f) and juveniles (j) in samples collected at the northern (13°N) and southern East Pacific Rise (EPR) during the Phare'02 and Biospedo'04 cruises. A χ^2 test was used to compare observed and expected numbers of males and females under the null hypothesis of gender equality (1:1). *p < 0.05 (sex ratios with no asterisk are not significant); S4: 4th setigerous segment

Vent site	Number of individuals				Number of mature individuals			Proportion of mature individuals		S4 width mean (mm)			S4 width mean of mature individuals (mm)	
	m	f	j	Sex ratio φ/σ	m	f	Sex ratio φ/σ	m	f	m	f	ind. mean	m	f
Northern EPR														
Elsa (pooled)	41	34	7	1.21	8			0.24		5.67	5.85		7.57	
Genesis	31	32	0	0.97	12			0.37		5.40	5.92		7.56	
Julie	59	56	11	1.05	13			0.23		4.39	4.59		7.01	
Jumeaux	45	44	14	1.02	6			0.14		5.63	6.31		8.73	
Pulsar	20	15	5	1.33	2			0.13		5.78	6.37		8.40	
Parigo	37	40	0	0.90	10			0.25		6.57	6.86		6.55	
Total Phare'02	233	221	37	1.05	51			0.23		5.57	5.98		8.36	
Southern EPR														
7° 25' S	16	16	0	1.00	9	5	1.80*	0.56	0.31	6.25	5.96	6.13	6.79	7.71
13° 58' S	22	27	8	0.81	5	4	1.25*	0.23	0.15	4.67	5.23	4.78	8.13	8.80
17° 24' S	52	41	0	1.27*	13	16	0.81*	0.25	0.39	6.21	6.69	6.42	7.80	7.48
17° 25' S	9	8	0	1.13	6	3	2.00*	0.67	0.38	5.76	7.23	6.40	6.89	9.93
17° 34' S	41	42	3	0.98	19	14	1.36*	0.46	0.33	5.95	6.32	5.97	7.83	8.47
18° 25' S	68	62	2	1.10	17	8	2.13*	0.25	0.13	4.99	4.84	4.92	7.11	7.92
21° 25' S	23	21	2	1.10	7	4	1.75*	0.30	0.19	6.00	5.86	5.91	9.11	9.65
21° 33' S	13	19	0	0.68*	9	5	1.80*	0.69	0.26	7.44	7.23	7.31	8.38	10.22
Total Biospedo'04	244	236	15	1.03	85	59	1.44*	0.35	0.25	5.63	5.90	5.75	7.69	8.36

The proportion of juveniles (3%) and mature individuals (around 25% of the total number of females) were close to those observed at 13°N/EPR 2 yr apart. Nevertheless, the proportion of mature females varied between 13 and 39% across latitudinal samples (Table 2). The size of females at first sexual maturity corresponds to 5.80 mm, and is not different from the value found at 13°N/EPR. The size of males at first sexual maturity is quite a bit smaller (2.90 mm) than that of females. With the exception of 17°24'S, mature females were always bigger than mature males at each site, and a significant difference in the proportion of mature males and females was found in favour of males (χ^2 test; p > 0.05).

Male and female size distributions obtained from the whole 13°N/EPR and SEPR collections are shown in Fig. 2. Distributions are similar in shape between genders and between collections concerning the number and the position of the modal components, the third containing only mature individuals. Kolmogorov-Smirnov 2-sample tests showed no significant statistical difference between male and female distributions of NEPR (p > 0.10)

and SEPR (p > 0.10), but no significant difference between cruises either (p > 0.10). However, discrepancies can be found among size-frequency histograms associated with local populations. Indeed, as shown in Fig. 3, populations from the vent chimneys Parigo, Rehu Marka

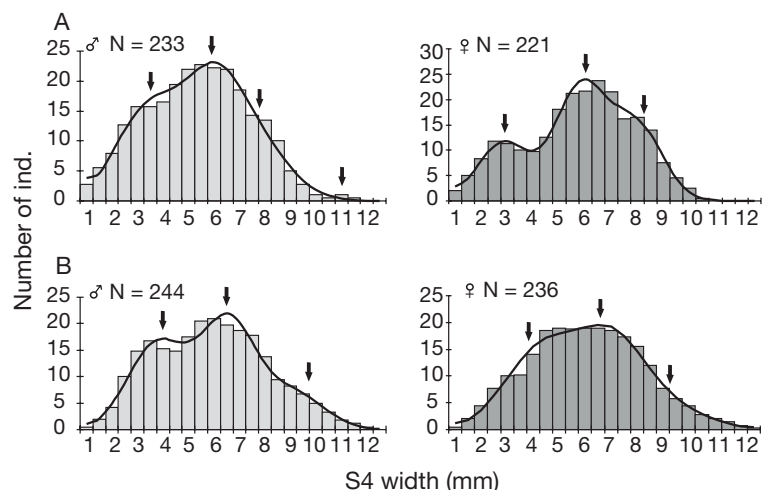


Fig. 2. *Alvinella pompejana*. Size-frequency histograms of males (light grey) and females (dark grey) over the whole *A. pompejana* collection sampled: (A) at 13°N on the East Pacific Rise (EPR) during the Phare'02 cruise and (B) over the southern EPR during the Biospedo'04 cruise. Curve: theoretical admixtures calculated by MIX 3.13; arrows: modal position of cohort; N: number of individuals; S4: 4th setigerous segment

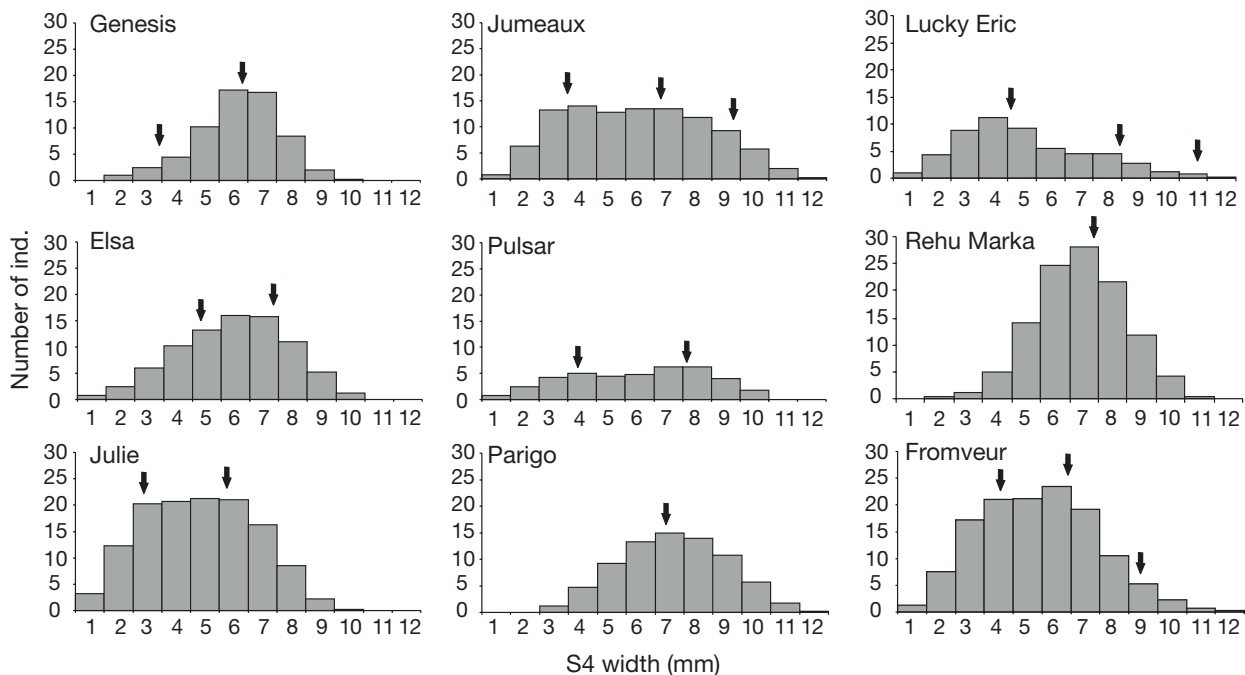


Fig. 3. *Alvinella pompejana*. Subsample of 9 size-frequency distributions for which individuals were the most numerous (6 vent sites at 13°N on the East Pacific Rise [EPR]: Genesis, Elsa, Julie, Jumeaux, Pulsar and Parigo; 3 vent sites on the southern EPR: at 13°58' S [Lucky Eric]; at 17°24' S [Rehu Marka]; at 18°25' S [Fromveur]). Discrepancies between histograms may be an indication of the recruitment heterogeneity according to peculiar dynamics conditions occurring locally (currents or vent activity). Arrows: modal position of cohort

and Genesis lacked at least 1 size group. This supports the hypothesis of a spatial heterogeneity in larval recruitment, survival, or individual growth rate.

Synchronisation of recruitment at a local scale (13°N/EPR)

Size-frequency distributions and modal decompositions have been performed for samples containing at least 50 ind. (i.e. Elsa, Genesis, Julie, Jumeaux, Pulsar and Parigo) (Table 3). Note that temporal samples collected (max. 10 d apart) at the vent sites Elsa and Parigo (see Table 1) were pooled together after a Mann-Whitney *U*-test indicating no difference between the temporal distributions ($U = 458.5$, $p = 0.847$ and $U = 565.5$, $p = 0.0738$, respectively). The number of modal components varies according to locality: only 1 size group was found at Parigo, 2 at Elsa, Genesis and Julie and 3 at Jumeaux. A Kruskal-Wallis test indicated that localities differ significantly in distribution ($H = 49.677$, $df = 7$, $p < 0.0001$). An *a posteriori* test, however, clearly showed that samples are not grouped according to the sampling date. Such spatial heterogeneity in distributions contradicts the assumption of recruitment synchrony at a field scale. However, synchronisation may be masked by differences in individual growth rates

and mortality between sites or by differences in larval settlement over the 13°N/EPR field. The similarity of modal values within each component however indicates that differences in distribution are more likely due to a lack of some components at some locations and, thus, discrepancies in recruitment (Table 3, Fig. 3).

Synchronisation of recruitment at a regional scale (SEPR)

Size-frequency distributions and modal decompositions have been performed for the whole fields and over each vent sector for 13°58', 17°24', 17°34' and 18°25' S, and by pooling worms coming from 21°25' and 21°33' S (Table 3). As previously observed within the vent sector 13°N/EPR, the number of modal components varies according to vent sectors: at least only 1 or 2 size groups (out of 3) occur at 17°24' and 17°34' S, respectively, whereas 3 modes can be identified for all other vent sectors. The unique modal value of the sample from 17°24' S is 6.40 mm, which corresponds to Size group 2 of the other samples and suggests either an absence of recent recruitment in this sector or a spatial discrimination between adults and juveniles at this particular site. A 1-way ANOVA was performed to compare modal positions for Size groups 1, 2 and 3

Table 3. *Alvinella pompejana*. Modal decomposition of size-frequency distributions of the 4th setigerous segment width within samples collected at northern (13°N) and southern East Pacific Rise (EPR). The modal decomposition was performed using MIX 3.1 software (McDonald & Green 1988).
n: number of individuals collected

Vent site	Proportion	Mean \pm SD (mm)
Northern EPR		
Elsa (p = 0.9990; $\chi^2 = 0.380$; df = 6; n = 82)		
Component 1	0.58	4.40 \pm 1.6
2	0.42	6.83 \pm 1.18
3	–	–
Genesis (p = 0.9997; $\chi^2 = 0.258$; df = 13; n = 63)		
Component 1	0.08	2.70 \pm 0.75
2	0.92	5.90 \pm 1.21
2	–	–
Julie (p = 0.9995; $\chi^2 = 0.297$; df = 6; n = 126)		
Component 1	0.42	2.62 \pm 1.07
2	0.58	5.54 \pm 1.36
3	–	–
Jumeaux (p = 0.9862; $\chi^2 = 0.143$; df = 3; n = 103)		
Component 1	0.26	2.70 \pm 0.89
2	0.58	5.75 \pm 1.69
3	0.16	8.70 \pm 1.08
Pulsar (p = 0.9999; $\chi^2 = 0.138$; df = 6; n = 40)		
Component 1	0.44	3.28 \pm 1.37
2	0.56	7.12 \pm 1.31
3	–	–
Parigo (p = 0.9951; $\chi^2 = 1.724$; df = 9; n = 77)		
Component 1	–	–
2	1	6.64 \pm 1.80
3	–	–
Southern EPR		
13° 58' S (p = 1.0000; $\chi^2 = 1.705$; df = 17; n = 57)		
Component 1	0.70	3.52 \pm 1.18
2	0.28	7.31 \pm 1.18
3	0.02	11.00 \pm 1.18
17° 24' S (p = 1.0000; $\chi^2 = 1.712$; df = 17; n = 93)		
Component 1	–	–
2	1.00	6.40 \pm 1.33
3	–	–
17° 34' S (p = 1.0000; $\chi^2 = 3.195$; df = 18; n = 86)		
Component 1	0.39	3.76 \pm 1.44
2	0.61	7.36 \pm 1.55
3	–	–
18° 25' S (p = 0.9907; $\chi^2 = 5.735$; df = 16; n = 132)		
Component 1	0.45	3.31 \pm 1.10
2	0.46	5.80 \pm 1.10
3	0.08	8.40 \pm 1.10
21°25–33' S (p = 1.0000; $\chi^2 = 2.206$; df = 18; n = 78)		
Component 1	0.45	4.17 \pm 1.11
2	0.30	7.01 \pm 1.11
3	0.25	9.79 \pm 1.11

between vent fields of the SEPR. This test showed no significant differences among modal values for Size Group 1, 2, or 3 respectively, $F = 0.09$, $p > 0.05$; $F_c = 0.32$, $p > 0.05$; $F = 1.98$, $p > 0.05$), suggesting that local recruitments may be more or less synchronised along the ridge axis over 15° of latitude when they occurred.

Reproduction dynamics in *Alvinella pompejana*

Reproductive stages

The high number of modal decompositions that have been performed (98 females: 20 000 measured oocytes) and the constancy of the modal standard deviation across the oocyte cohorts (12 μ m) allowed us to discriminate 5 oocyte maturation stages in *Alvinella pompejana* populations according to the number of oocyte cohorts (1 to 5), with decreasing numbers of gametes (Fig. 4).

Fecundity range

Female fecundity varies greatly among individuals and mainly depends on the maturation stage of the female. Females with 1 oocyte cohort have an average number of 5810 oocytes, 2-cohort females have 28900 oocytes, 3-cohort females have 35330 oocytes and 4-cohort females have 135112 oocytes. Reproductive females with 5 oocyte cohorts display 200319 oocytes on average, which is higher than the counts by Desbruyères et al. (1998). However, reproductive females with 5 cohorts have high but very unequal fecundities that range from 67000 to about 978000 oocytes. All females having empty oviducts present <5 cohorts and very low fecundities. The number of oocytes, as expected, slightly increases with body size, but this trend is not significant ($R^2 = 0.149$, $p > 0.05$). Furthermore, there is no statistic correlation between body size and mean oocyte diameter when analyzing the 2 collections (NEPR, SEPR) of mature females (see Fig. 5A, B). A significant positive relationship ($R^2 = 0.442$, $p < 0.01$) was reported between mean oocyte diameter and number of oocyte 'cohorts' found in the coelom (Fig. 6).

Asynchrony of oocyte maturation between females

Oocyte maturation stages are not spatially or temporally synchronised. The proportions of these stages do not seem to be correlated with the increasing size of the worm and do appear to be heterogeneous among the collected samples (Table 4). Within a given size group, some females appear to begin oogenesis when others are active spawners. The occurrence of distinct oocyte maturity stages among females throughout samples may be a good indicator of a continuous breeding system in this peculiar polychaete. However, reproductive females having 5 oocyte 'cohorts' represent the most frequent reproductive stage found in maturing females, suggesting that oogenesis is probably very fast. Simultaneous observations of oviducts

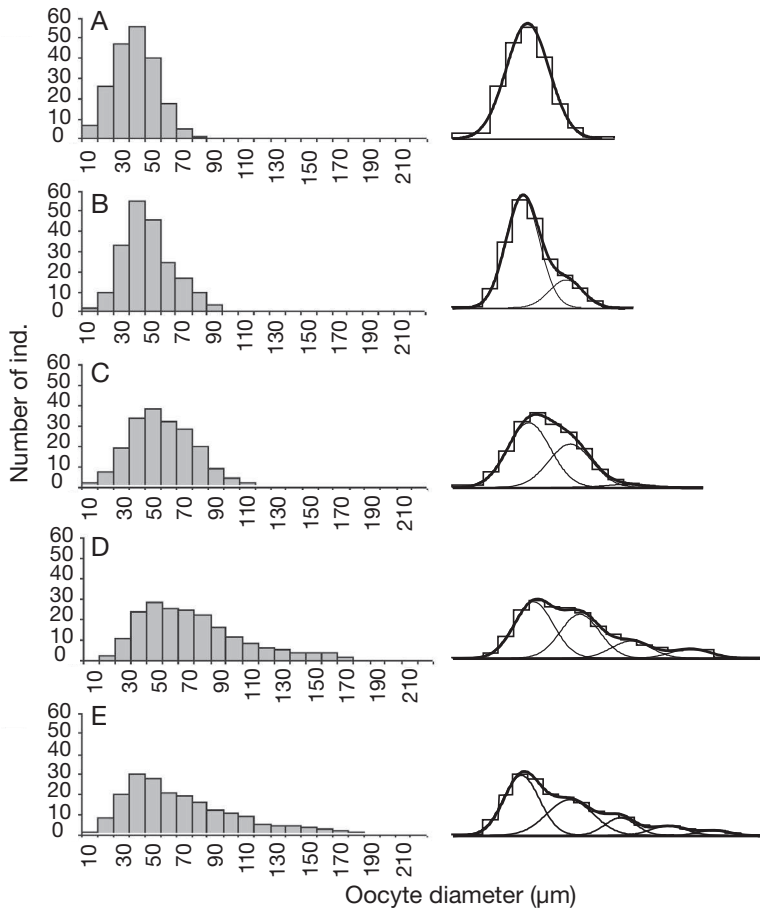


Fig. 4. *Alvinella pompejana*. Presentation of the 5 female maturity stages (mature to reproductive), with oocyte size-frequency distribution (left) and their modal decomposition (right) in Cohorts 1 to 5 (A to E, respectively)

and coelomic fluid showed that some large females initiated oogenesis (none to 1 single cohort of 20 to 50 µm oogonia) while they still had 200 µm mature oocytes in their oviducts.

Variation of oocyte distributions within vent sectors

Although mature females exhibit some asynchrony in their reproductive state, discontinuous recruitment as expected from demography could still be explained by the synchronisation of spawning events. According to this hypothesis, oocyte cohorts are expected to be synchronised among all females over the whole process of oogenesis. We therefore tested this hypothesis under the assumption of a constant maturation rate of oocytes. Modal positions were compared between females within the vent sectors 13°N/EPR and 17°S/EPR only for the first 2 oocyte 'cohorts' for which a sufficient number of individuals was available. A 1-way ANOVA shows that there is no significant difference

between modal positions of these 2 subsequent 'cohorts' within the 13°N/EPR vent field (first cohort: $F_c = 0.08$, $p > 0.05$; second cohort: $F_c = 0.22$, $p > 0.05$). A similar test performed on the oocyte distributions of females that came from 2 hydrothermal vents <50 km away from each other (17°24' and 17°34' S) did not show a significant difference either (first cohort: $F_c = 0.08$, $p > 0.05$; second cohort: $F_c = 0.16$, $p > 0.05$), suggesting that spawning could be synchronised at a vent sector scale (Fig. 7). On the other hand, this may also be an indication of continuous gametogenesis, in which dynamics are insensitive to locality or sampling date.

Variation of oocyte distributions at the regional scale (SEPR)

The extensive sampling of *Alvinella pompejana* from 7°24' to 21°33'S along the SEPR allowed us to test the hypothesis of a latitudinal shift of reproduction in *A. pompejana*. In order to discriminate between a latitudinal effect and a time effect, the chronology of the vent field sampling did not correlate with latitude (see Table 1). Therefore, a 1-way ANOVA test was performed using modal positions of oocyte 'cohorts' as variables and females as replicates within each vent field. Modal means and related standard deviations are presented in Fig. 8 as a scatterbox graph. Although populations were separated by 100s of km, the only significant difference was detected for Cohort 1 ($F = 4.59$, $p < 0.01$). For Cohorts 2, 3, 4 and 5, differences were not significant ($p > 0.05$), indicating a great reproductive homogeneity across mature females over the whole EPR.

Table 4. *Alvinella pompejana*. Distribution of female reproductive stages in latitudinal samples collected along the southern East Pacific Rise

	Number of oocyte cohorts					Total
	1	2	3	4	5	
13° N	1	0	1	6	17	25
7° S	0	0	0	3	2	5
13° S	0	0	0	1	3	4
17° S	1	2	1	4	20	28
18° S	0	1	2	1	4	8
21° S	2	0	0	3	4	9
Total	4	3	4	18	50	79

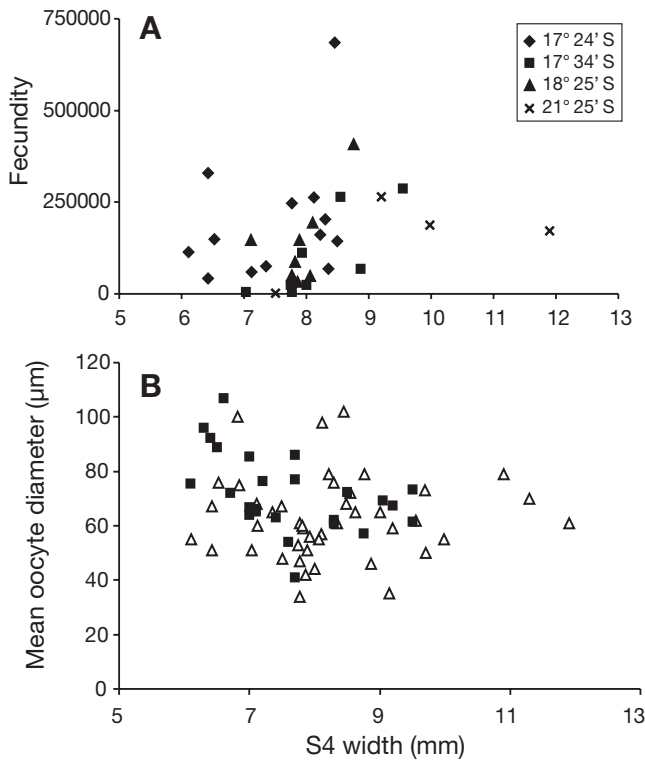


Fig. 5. *Alvinella pompejana*. (A) Positive but not significant relationship between female fecundity and female S4 width. (B) Negative but non-significant relationship between mean oocyte diameter and size of female (S4 width) (N = 54 females). Filled symbols: females from Phare'02; open symbols: females from Biospeedo'04. Note different x-axis scales

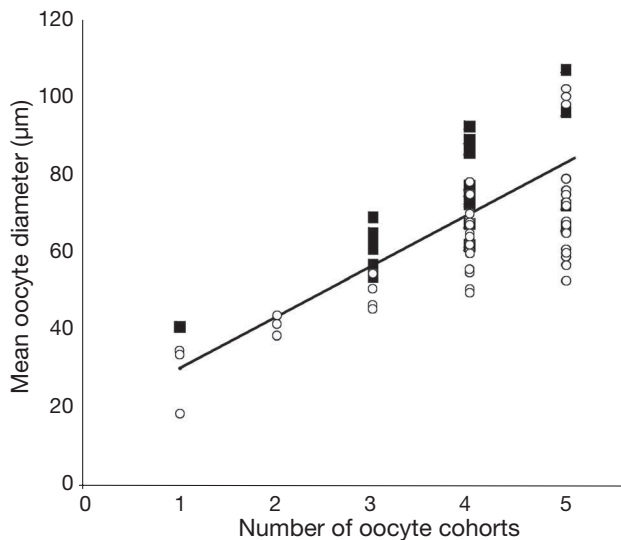


Fig. 6. *Alvinella pompejana*. Positive relationship between mean oocyte diameter (µm) and female 'reproductive' stage (represented by the number of oocyte cohorts) (N = 54 mature females, $y = 9.703x + 26.916$, $R^2 = 0.442$; $p < 0.01$). Filled symbols: females from Phare'02; open symbols: females from Biospeedo'04

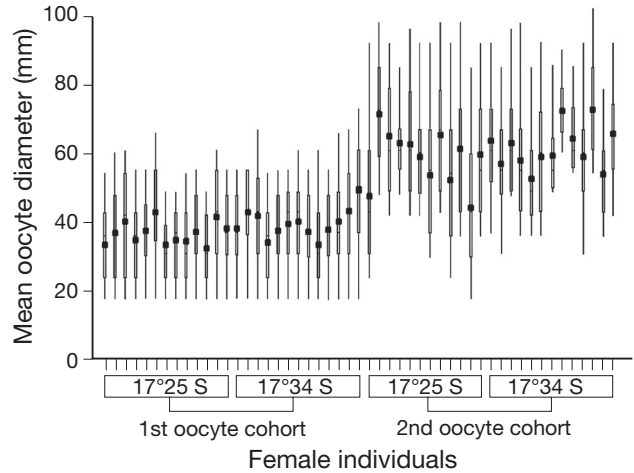


Fig. 7. *Alvinella pompejana*. Scatterbox showing the modal oocyte diameter per ind. together with minima, maxima and SD for the first 2 oocyte cohorts of all the mature females of *A. pompejana* collected at the 2 vent fields 17° 24' and 17° 34' S on the East Pacific Rise

A multiple linear regression with latitude, the female S4 width and the sampling date as explicative variables was conducted on the first oocyte modal position. This analysis showed only a significant decrease in oocyte diameter with the sampling date (Table 5). A graphic representation of the oocyte diameter against the sampling date suggests that the decrease in oocyte diameter (oogonia renewal) may be cyclic (Fig. 9).

Short-term evolution of the Totem population

Despite an extensive effort to sample the Totem population, our analyses could only be conducted on small sample sizes: 48 worms (16 February 1996) and 6 worms (18 March 1996). From modal decompositions of size-frequency histograms, 2 distinct size groups were found at each date (Table 6). A Student's *t*-test

Table 5. Equation parameters (*b*, *t*) and associated statistics (P-perm) of the multiple linear regression of the first oocyte cohort across females of *Alvinella pompejana* collected along the southern East Pacific Rise. Explicative variables correspond to latitude, sampling date and S4 width

Variable	<i>b</i>	<i>t</i>	P-perm
Intercept	42.572	9.093	
Latitude	0.218	0.437	0.340
Date	-0.312	-3.213	0.001*
S4 width	-0.059	-0.349	0.379
$R^2 = 0.19038$	$F = 3.84069$		0.0150*
Number of permutations: 999			

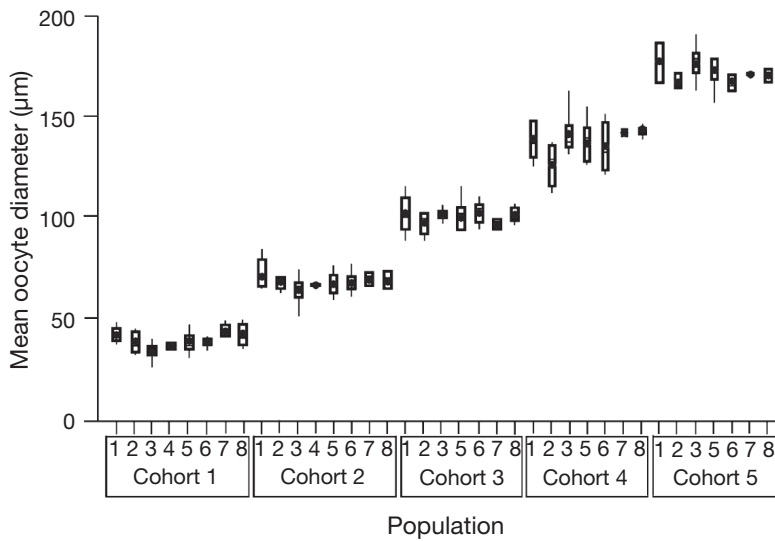


Fig. 8. *Alvinella pompejana*. Scatterbox showing the mean oocyte diameter across females of *A. pompejana* from each population together with minima, maxima and SD for the 5 oocyte cohorts. Numbers (vent fields)—1: 7° 25' S; 2: 13° 58' S; 3: 17° 24' S; 4: 17° 25' S; 5: 17° 34' S; 6: 18° 25' S; 7: 21° 25' S; 8: 21° 33' S

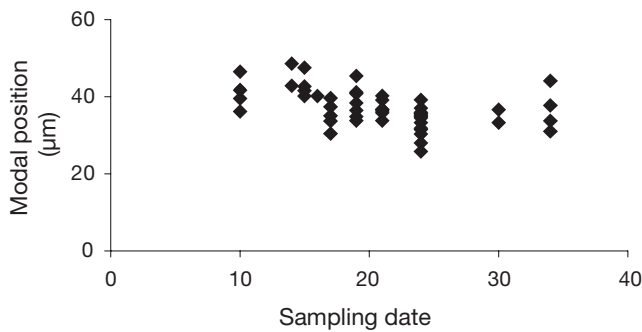


Fig. 9. *Alvinella pompejana*. Relationship between the modal position of the first oocyte cohort and the sampling date (in days since the beginning of the Biospeedo'04 cruise)

comparison of the modal positions between the 2 samples was found to be statistically significant ($t = 2.423$, $df = 43$, $p < 0.005$) and highly significant ($t = 4.26$, $df = 64$, $p < 0.01$) for Modal components 1 and 2, respectively. Modal differences between the 2 temporal samples could be attributable to the ageing of the cohorts, but also to migration, sampling impact, or small-scale environmental heterogeneity.

The temporal evolution of oocyte maturation was followed by examining 10 females and 9 females at each date, respectively. MFodal values averaged over females resulting from the decomposition of size-frequency histograms suggested a tiny shift in oocyte maturation (Table 6). A comparison of these modal values using a t -test reveals, however, no significant difference.

DISCUSSION

Sex ratio, effective population size and mature male excess

Alvinella pompejana is a tube-dwelling worm which lives in a highly patchy and unstable environment. The life span of alvinellid chimneys is likely within the order of a decade, and the cessation of activity causes the death of most associated fauna (Tunncliffe 1991). To endure local extinction, *A. pompejana* appears to have adapted to these vent conditions through a strategy that maximises reproductive success and subsequent effective dispersal. An equilibrate sex ratio of 1:1 is the most effective way to maximise the reproductive effort, providing that partners have an equal chance to mate, which is often associated with an internal mode of fertilisation in annelids (Giangrande 1997).

In *A. pompejana*, although balanced when considering the overall population, the sex ratio appears to be significantly biased towards males in the breeding population (i.e. mature individuals). Increasing the number of reproductive males against reproductive females would decrease the effective population size (Hartl 1988), and thus is likely to become detrimental for the

Table 6. *Alvinella pompejana*. Modal value means of the oocyte diameter across females for each oocyte 'cohort' between temporal samples collected subsequently, 31 d apart at the vent site Totem (13° N/EPR)

	Modal position	SD
Dive 106 (16 Feb 1996)		
Width (mm)		
1	3.60	1.57
2	8.57	1.70
Oocyte diameter (µm)		
1	40.09	6.38
2	70.25	10.48
3	119.86	14.61
4	156.69	6.47
5	180.82	3.87
Dive 1088 (18 Mar 1996)		
Width (mm)		
1	2.72	1.67
2	7.32	1.67
Oocyte diameter (µm)		
1	38.46	6.03
2	59.82	5.07
3	97.03	11.68
4	144.70	12.00
5	Absent	Absent

species over evolutionary time. To understand whether an excess of reproductive males could be beneficial for the species, one needs to consider that *A. pompejana* is a gonochoric species that uses internal fertilisation to reproduce (Chevaldonné & Jollivet 1993). Females possess a pair of oviducts that are both connected to a spermathecae in which spermatozooids are stored and which opens as a unique genital pore (Pradillon & Gaill 2003). The presence of spermathecae in females allows them to be fertilised by several males and, even more, to be inseminated while they are still immature. A specific case of immature insemination of females has already been reported for the coastal polychaete *Dinophilus gyrocellatus* (Westheide 1990). During the Biospedo'04 cruise, this hypothesis was confirmed by the discovery of a young immature female that was already inseminated (i.e. presence of spermatozooids inside the spermatheca). As a consequence, a sex-ratio bias in favour of reproductive males represents an obvious adaptive advantage, as it increases the probability that females become fertilised as early as possible and by >1 male. Such a reproductive strategy has already been described in species with very small population sizes, which mate more often in order to increase the effective population's size (Wilson 1963). The maintenance of genetic diversity is favoured and can be highly beneficial for a species that endures recurrent population extinctions and possible founder effects. Previous *in situ* observations indeed showed that newly formed chimneys are usually recolonised by very few young individuals (Chevaldonné 1996). In such patches all individuals are immature and non-reproductive. A similar situation has been described for the early successional stages of the Northeast Pacific alvinellid *Paralvinella palmiformis* (Copley et al. 2003). Reproductive mechanisms whereby global genetic diversity is maintained despite recurrent population bottlenecks that decrease the effective population size may have been an adaptive response to the vent conditions over time. Because this kind of reproductive mode has also been reported for other hydrothermal vent polychaetes (*Paralvinella grasslei*: Zal et al. 1994, *Branchiopolynoe seepensis*: McHugh & Tunnicliffe 1994, Jollivet et al. 2000) in which spermathecae are present and pseudo-copulation occurs, one can propose that this kind of reproductive mode has been selected by polychaetes in unstable and ephemeral habitats such as hydrothermal vents, as it increases the reproductive success of the species and promotes genetic diversity in small populations.

Population structure and recruitment dynamics

Size-frequency distributions of the worms combined across localities for each cruise are quite similar in

shape and dominated by the medium-sized individuals. A refined analysis of the observed global distributions together with those obtained within each locality, however, clearly indicated that they significantly depart from normal distribution with a regular polymodal structure where modal values are equally spaced. Following modal decompositions, nearly all populations are characterised by 3 size groups, which correspond to immature individuals, individuals that possess fully developed genital tracts and reproductive individuals. Reproductive females always have the largest sizes, whereas reproductive males can also be found in the second size group. Such observations favour the hypothesis of regular arrivals of recruits inside populations. Conversely, the fact that very few immature individuals are usually found in the samples, coupled with the observation that size histograms are the same independent of the latitude and sampling date, is a strong argument in favour of a low but semi-continuous settlement of individuals. From our dataset we can therefore propose that *Alvinella pompejana* exhibits a semi-continuous mode of recruitment, where frequent pulses of settlers are likely to provide the observed distributions. This assumption however relies upon the hypothesis that young *A. pompejana* have a very fast growth rate and rapidly reach Cohort 2, where they accumulate. The dominance of Cohort 2 reflects a decrease in growth rate, a high mortality rate, a short life duration of the worm, or a bad sampling technique. Although numerous factors could be involved to explain the short-term variations of demographic structures at Totem (e.g. migration, small-scale environmental heterogeneity), the shift between modal peaks of the subsequent temporal distribution could reflect a high growth rate of individuals.

Oocyte distributions and reproduction dynamics

The modal decomposition of oocyte-size histograms showed the presence of 1 to 5 discrete oocyte cohorts according to the reproductive state of the animal, displaying a log-normal decrease in abundance from oogonia to mature oocytes. The smallest mature females were characterised by a single cohort of small-sized oocytes (20 to 40 μm), whereas the largest mature females displayed 5 distinct and equally spaced cohorts of oocytes. These latter females were the only ones to present mature oocytes (180 μm) inside their oviducts and, thus, the only ones able to spawn. The examination of numerous oocyte distributions and their cross-comparison provided no doubt about the quality of the modal decomposition, with modal positions of the oocyte cohorts equidistant and with constant standard deviations. Consequently, this kind of

oocyte distribution first suggests periodic short-term pulses in the oocyte production with a linear growth rate of oocyte maturation until recruitment in the oviducts is completed. This assumption is strengthened by the fact that reproductive stages are positively correlated with oocyte size. Together with the fact that no more than 4000 mature oocytes can accumulate inside both oviducts before each spawning, regularly spaced oocyte cohorts could be indicative of a reproduction pattern with discrete and regular spawning events. However, the fact that reproductive stages are not synchronised in females among populations is a strong argument in favour of semi-continuous reproduction in *Alvinella pompejana*. Most populations indeed display a variable proportion of each reproductive stage even if the highest proportion of reproductive females corresponds to the 5-cohort stage (i.e. females able to spawn). In polychaetes, most continuous reproducers exhibit non-synchronised gametogenic stages across individuals within populations (Giangrande 1997).

The log-normal shape of the oocyte size distributions is, however, very surprising and poses an intriguing question about the decreasing number of oocytes in each 'cohort' during maturation. Similar observations have been obtained on the deep-sea ice worm *Hesio-caeca methanicola* and have been explained as an artefact due to organism damage or partial spawning during fixation (Eckelbarger et al. 2001). This, however, was not the case for *A. pompejana*, as females were stressed but alive and still intact prior to dissection and oviducts prevented animals from expelling coelomic oocytes outside.

The oocyte size-frequency distributions with a lot of small oocytes (>100000) and only a few observed in the oviducts (4000) suggest that most oogonia are not able to reach their final stage of maturation in the coelom. Because no degenerating oocyte has been found in >10000 oocyte measurements, the death of a large number of oogonia during the maturation process is unlikely. We propose that the decrease in oocyte numbers from the beginning of synthesis to oviduct recruitment within each female could reflect: (1) a very fast growth rate of individuals, (2) different rates of oogenesis and vitellogenesis, or (3) endogenous hormones that control the maturation process in order to fuel the oviducts on demand (i.e. transfer of x mature oocytes in oviducts implies the maturation of x oogonia in the coelom). The second idea relies on the fact that oogenesis is rapid and continuous while vitellogenesis is a slow process. Oocyte production is then a function of the animal size, so that the number of oocytes produced increased continuously over time. Therefore, smaller oocytes should be found in much higher numbers than larger oocytes. Such a hypothesis

fits well with the occurrence of a yolk autosynthesis mode of production, which alternatively indicates a low oocyte maturation rate (Eckelbarger 1994). The third idea relies on the fact that the massive stock of oogonia in the coelom can be viewed as a vitelline reserve that would secure the full development of a small subset of oocytes via nurse or 'helper' cells (Eckelbarger 1986). However, Pradillon & Gaill (2003) previously showed that vitellogenesis is achieved inside the coelomic cavity without 'helper' cells via autosynthetic mechanisms of yolk production.

Spawning and recruitment synchrony over large spatial scales

The polymodal nature of size histograms for both individual worms and oocytes suggests that spawning and recruitment are non-continuous and possibly periodic, at least over short periods of time. Discontinuous recruitment has been reported for nearly all hydrothermal vent species examined so far (Tyler & Young 1999), especially vent polychaetes (McHugh 1989, Zal et al. 1995, Copley et al. 2003), and has led to the hypothesis that external factors such as tidal cycles or food source seasonality must synchronise the recruitment by pulses. By contrast, most reproductive characteristics of *Alvinella pompejana*, like the co-occurrence of all reproductive stages in various proportions within each population, support the hypothesis of semi-continuous reproduction. Nevertheless, the lack of significant differences in the modal position of oocyte sizes suggests that females may be highly synchronised within and between hydrothermal vent fields at the ridge scale, despite their heterogeneity in reproductive stages. In this case, discrete spawning events may occur within a short time interval. Although we did not detect great variation in oocyte modal diameters between females from SEPR vent fields, multiple regression analysis of the first modal cohort between females indicated a significant decrease in oocyte diameter with time. These results suggest that a periodic signal in oocyte maturation may occur in females, possibly in response to regular pulses of spawning. This observation is in agreement with the occurrence of periodic settlement peaks in populations and the shift of oocyte modes observed between temporal samples collected during Hot'96. Although nearly continuous, the reproductive cycle can be partly influenced by environmental factors, such as the lunar cycle, as previously stated by Chevalloné et al. (1991), Zal et al. (1995) and Copley et al. (1999).

Based on this present dataset, we propose that *Alvinella pompejana* has 2 contrasting life phases: a rapid growth phase during the early benthic phase and an

intense reproductive phase with frequent spawning events later. In the second phase, most of the energy may be allocated to gamete production. The largest specimens could correspond to reproductive individuals of different age classes and reproductive stages as described for the deep-sea ophiuroids (Gage & Tyler 1991). Such a hypothesis may explain why no positive relationship between the size of the worm and its reproductive stage or oocyte diameter was found. Consequently, oocyte diameter or female maturity stage may possibly reflect the age of the worm in a better way than the individual size.

Size-frequency histograms displayed a great heterogeneity in the number of modal components within and between vent fields. This is mainly due to the lack of, at least, one size group within specific samples. As an example, the sample at 17°34'S/EPR was only composed of small and large individuals. By contrast, the sample from the nearby vent field at 17°24'S only displayed medium-sized individuals. The presence of a large number of immature individuals at the former site clearly indicated the arrival of new recruits at the chimney, although they were totally absent at the latter site. Such an observation suggests that: (1) local vent dynamics may occasionally prevent the settlement and recruitment at a given site or (2) larval supply may only be episodic and occasionally prevent local recruitment according to hydrodynamic constraints. Furthermore, size heterogeneity of modal components among localities can be explained by variation in the growth rate of females in habitats.

CONCLUSIONS

Based on our study, *Alvinella pompejana* displays several population and reproductive features: (1) a 1:1 sex ratio but a shift in sexual maturity between genders, (2) a very high fecundity reaching 200 000 oocytes on average, (3) a heterogeneity in maturation state of individuals within populations and (4) reproduction dynamics that follow a semi-continuous breeding pattern with short-term synchrony. Like other hydrothermal vent polychaetes, polymodal demographic structures are characteristic of *A. pompejana*. While this kind of structure is generally associated with discontinuous and synchronised reproduction, the limited number of juveniles in populations and the temporal and spatial similarities of size histograms between cruises and localities suggested the frequent arrival of a limited number of recruits within populations, followed by a fast growth rate among recruits and accumulation of individuals in the adult cohort.

Morphological adaptations of the reproductive apparatus, such as oviducts and spermathecae, and semi-

continuous oogenesis reinforce the idea that these reproductive peculiarities could have recently evolved in response to the unstable nature of the vent habitat, in order to maximise reproductive success and maintain genetic diversity. Because some of these characteristics are shared between most hydrothermal polychaetes (Zal et al. 1994, Copley et al. 2003, Hilario et al. 2005), we propose that they represent a recent adaptive convergence to vent patchiness and instability, in order to minimise the extinction risk for the species.

Acknowledgements. We gratefully acknowledge the captains and crews of the RV 'Nadir' and RV 'Atalante', and the pilots and co-pilots of the 'Nautile' team. We are particularly grateful to F. Gaill and N. LeBris, chief scientists of the Hot'96 and Phare'02 cruises, who allowed us to conduct this work. We thank the 3 anonymous referees for their very valuable comments on the first draft of the manuscript. This work was partially funded by the Région Bretagne (contrat Accueil & Emergence).

LITERATURE CITED

- Anderson MJ, Legendre P (1999) An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *J Stat Comput Simulation* 62: 271–303
- Bhattacharya CG (1967) A simple method of resolution of a distribution into Gaussian component. *Biometric* 23:115–135
- Chevaldonné P (1996) Ecologie des cheminées hydrothermales actives. Thèse de doctorat, Université de la Méditerranée Aix-Marseille
- Chevaldonné P, Jollivet D (1993) Videoscopic study of deep-sea hydrothermal vent alvinellid polychaete populations: biomass estimation and behaviour. *Mar Ecol Prog Ser* 95: 251–262
- Chevaldonné P, Desbruyères D, Le Haître M (1991) Time-series of temperature from three deep-sea hydrothermal vent sites. *Deep-Sea Res* 38:1417–1430
- Chevaldonné P, Jollivet D, Vangriesheim A, Desbruyères D (1997) Hydrothermal-vent alvinellid polychaete dispersal in the eastern Pacific. I. Influence of vent site distribution, bottom currents and biological patterns. *Limnol Oceanogr* 42:67–80
- Copley JTP, Tyler PA, Murton BJ, Van Dover CL (1997) Spatial and interannual variation in the faunal distribution at Broken Spur vent field (29°N, Mid-Atlantic Ridge). *Mar Biol* 129:723–733
- Copley JTP, Tyler PA, Van Dover CL, Schultz A, Dickson P, Singh SM, Sulanowska M (1999) Subannual temporal variation in faunal distribution at the TAG hydrothermal mound (26°N, Mid-Atlantic Ridge). *PSZNI Mar Ecol* 20: 291–306
- Copley JTP, Tyler PA, Van Dover CL, Philip SJ (2003) Spatial variation in the reproductive biology of *Paralvinella palmiformis* (Polychaeta: Alvinellidae) from a vent field on the Juan de Fuca Ridge. *Mar Ecol Prog Ser* 255:171–181
- Desbruyères D, Chevaldonné P, Alayse AM, Jollivet D and 14 others (1998) Biology and ecology of the 'Pompeii worm' (*Alvinella pompejana* Desbruyères and Laubier), a normal dweller of an extreme deep-sea environment: a synthesis of current knowledge and recent developments. *Deep-Sea Res II* 45:383–422

- Eckelbarger KJ (1983) Evolutionary radiation in polychaete ovaries and vitellogenic mechanisms: their possible role in life history patterns. *Can J Zool* 61:487–504
- Eckelbarger KJ (1986) Vitellogenic mechanisms and the allocation of energy to offspring in polychaetes. *Bull Mar Sci* 39:426–443
- Eckelbarger KJ (1994) Diversity of metazoan ovaries and vitellogenic mechanisms: implications for life history theory. *Proc Biol Soc Wash* 107:193–218
- Eckelbarger KJ, Young CM, Ramirez Llodra E, Brooke S, Tyler P (2001) Gametogenesis, spawning behavior, and early development in the 'iceworm' *Hesiocaeca methanicola* (Polychaeta: Hesioniidae) from methane hydrates in the Gulf of Mexico. *Mar Biol* 138: 761–775
- Frontier S, Pichod-Viale D (1998) Ecosystèmes: structure, fonctionnement, évolution. Masson, Paris
- Gage JD, Tyler PA (1991) Deep sea biology: a natural history of organisms at the deep-sea floor. Cambridge University Press, Cambridge
- Giangrande A (1997) Polychaete reproductive patterns, life cycles and life histories: an overview. *Oceanogr Mar Biol Annu Rev* 35:323–386
- Hartl DL (1988) A primer of population genetics, 2nd edn. Sinauer, Sunderland, MA
- Hilario A, Young CM, Tyler PA (2005) Sperm storage, internal fertilisation, and embryonic dispersal in vent and seep tubeworms (Polychaeta: Siboglinidae: Vestimentifera). *Biol Bull (Woods Hole)* 208: 20–28
- Jollivet D (1993) Distribution et évolution de la faune associée aux sources hydrothermales profondes à 13°N sur la dorsale du Pacifique oriental: le cas particulier des polychètes Alvinellidae, Vol 1. Thèses de doctorat, Université de Bretagne Occidentale, Brest
- Jollivet D (1996) Specific and genetic diversity at deep-sea hydrothermal vents: an overview. *Biodiversity Conserv* 5: 1619–1653
- Jollivet D, Chevaldonné P, Planque B (1999) Hydrothermal-vent alvinellid polychaete dispersal in the eastern Pacific. II. A metapopulation model based on habitat shifts. *Evolution* 53:1128–1142
- Jollivet D, Empis A, Baker MC, Hourdez S, Comtet T, Jouin-Toulmond C, Desbruyères D, Tyler PA (2000) Reproductive biology, sexual dimorphism, and population structure of the deep sea hydrothermal vent scale-worm, *Branchiopolynoe seepensis* (Polychaeta: Polynoidae). *J Mar Biol Assoc UK* 80:55–68
- MacDonald PDM, Green PEJ (1988) Users guide to program MIX 2.3: an interactive program for fitting mixtures of distributions. Technical notes on Release 2.3. Ichthus Data Systems, Hamilton
- McHugh D (1989) Population structure and reproductive biology of two sympatric hydrothermal vent polychaetes, *Paralvinella pandorae* and *Paralvinella palmiformis*. *Mar Biol* 103:95–106
- McHugh D (1993) A comparative study of reproduction and development in the polychaete family Terebellidae. *Biol Bull (Woods Hole)* 185:153–167
- McHugh D, Tunnicliffe V (1994) Ecology and reproductive biology of the hydrothermal vent polychaete *Amphisamytha galapagensis* (Ampharetidae). *Mar Ecol Prog Ser* 106:111–120
- Olive PJW (1970) Reproduction of a Northumberland population of the polychaete *Cirratulus cirratus*. *Mar Biol* 5: 259–273
- Olive PJW (1984) Environmental control of reproduction in Polychaeta. *Fortschr Zool* 29:17–38
- Pradillon F, Gaill F (2003) Oogenesis characteristics in the hydrothermal vent polychaete *Alvinella pompejana*. *Invertebr Reprod Dev* 43(3):223–235
- Pradillon F, Zbinden M, Mullineaux LS, Gaill F (2005) Colonisation of newly-opened habitat by a pioneer species, *Alvinella pompejana* (Polychaeta: Alvinellidae), at East Pacific Rise vent sites. *Mar Ecol Prog Ser* 302:147–157
- Sarrazin J, Juniper SK, Massoth G, Legendre P (1999) Physical and chemical factors influencing species distributions on a hydrothermal sulfide edifice of the Juan de Fuca Ridge, northeast Pacific. *Mar Ecol Prog Ser* 190: 189–112
- Schroeder KC, Hermans CO (1975) Annelide: Polychaeta. In: Giese AC, Pearse JS (eds) Reproduction of marine invertebrates. Academic Press, New York
- Shank TM, Fornari DJ, Von Damm KL, Lilley MD, Haymon RM, Lutz RA (1998) Temporal and spatial patterns of biological community development at nascent deep-sea hydrothermal vents (9°50'N, East Pacific Rise). *Deep-Sea Res II* 45:465–515
- Sokal RR, Rohlf FJ (1995) Biometry. The principles and practice of statistics in biological research, 3rd edn. Freeman, New York
- Thiébaud E, Huther X, Shillito B, Jollivet D, Gaill F (2002) Spatial and temporal variations of recruitment in the tube worm *Riftia pachyptila* on the East Pacific Rise (9°50'N and 13°N). *Mar Ecol Prog Ser* 234:147–157
- Tunnicliffe V (1991) The biology of hydrothermal vents: ecology and evolution. *Oceanogr Mar Biol Annu Rev* 29: 319–407
- Tyler PA, Young CM (1999) Reproduction and dispersal at vent and cold seeps. *J Mar Biol Assoc UK* 79:193–208
- Westheide W (1990) Polychaetes: interstitial families. In: Kermarck DM, Barnes RSK (eds) Synopses of the British fauna, new series. Universal Book Services, Oegstgeest
- Wilson EO (1963) Social modifications related to rareness in ant species. *Evolution* 17:249–253
- Wilson WH (1991) Sexual reproductive modes in polychaetes: classification and diversity. *Bull Mar Sci* 48:500–516
- Zal F, Desbruyères D, Jouin-Toulmond C (1994) Sexual dimorphism in *Paralvinella grasslei*, a polychaete annelid from deep-sea hydrothermal vents. *C R Acad Sci Paris* 317:42–48
- Zal F, Jollivet D, Chevaldonné P, Desbruyères D (1995) Reproductive biology and population structure of the deep-sea hydrothermal vent worm *Paralvinella grasslei* (Polychaeta: Alvinellidae) at 13°N on the East Pacific Rise. *Mar Biol* 122:637–648

Résumé: Ce travail vise, à travers une approche de coalescence multi-locus et multi-spécifique, à comprendre les modes de spéciation et les mécanismes d'évolution agissant chez des Mytilidae (genre *Bathymodiolus*) soumis à des pressions environnementales extrêmes : celles des milieux hydrothermaux profonds.

Dans un premier temps, la phylogénie de l'ensemble des espèces appartenant à la famille des Bathymodiolinae, a été étudiée par une utilisation complémentaire d'un marqueur mitochondrial et d'un marqueur nucléaire. Dans une seconde partie, une approche multi locus a été menée sur une zone hybride décrite sur la ride médio-atlantique afin d'étudier la structure génétique et le niveau d'introgession entre les deux espèces locales (*Bathymodiolus azoricus* et *B. puteoserpentis*). Une approche bayésienne (utilisation du logiciel Isolation with Migration IMA) a permis de déterminer l'intensité et l'orientation des flux géniques entre les deux espèces. Ceux-ci sont variables d'un locus à l'autre, indiquant un contact secondaire après une période d'isolement plutôt qu'une différenciation parapatrique progressive. Une étude menée sur les espèces de la dorsale est-pacifique nous a permis de décrire une nouvelle zone hybride entre l'espèce locale déjà connue (*B. thermophilus*) et une nouvelle espèce localisée à 37°S. L'étude de certains gènes ayant montré des particularités dans leurs histoires évolutives a également été abordée, mettant en évidence de la sélection purifiante sur un intron du gène du Facteur d'Elongation 1 α et un relâchement des pressions de sélection sur certaines copies du gène de la protéine de choc thermique HSP70.

Notre étude met en évidence que ce sont principalement les conditions géologiques (volcanisme, tectonisme et failles transformantes) qui structurent les espèces hydrothermales et que cela induit sur le long terme de la spéciation allopatrique. Nous confirmons aussi que les bivalves marins sont des champions dans le domaine des fortes diversités nucléotidiques et que le polymorphisme qui en résulte est constitué de nombreuses mutations légèrement délétères.

Multi locus study of selective pressures acting on a marine bivalve in hybridization conditions: example of the hydrothermal vent mussel *Bathymodiolus* from the MAR and EPR

Abstract: This work aimed, through a multi-locus multi-species study of gene coalescence, at understanding the modes of speciation and evolutionary mechanisms acting on Mytilid bivalves (genus *Bathymodiolus*) submitted to the extreme environmental conditions that prevail in the deep sea.

Firstly, the phylogeny of all species belonging to the family Bathymodiolinae was studied by the dual use of a nuclear and a mitochondrial marker.

In a second part, a multilocus approach was carried out on a hybrid zone existing on the Mid Atlantic Ridge to unravel the genetic structure and measure the level of reciprocal introgression between *Bathymodiolus azoricus* and *B. puteoserpentis*. A Bayesian approach (through the implementation of the Isolation with Migration IMA software) allowed us to measure the intensity and the directionality of residual gene flow between the two species, this varying among genes and pointing toward a secondary contact after a period of isolation rather than a continuous differentiation.. Another study on the East-Pacific Rise has allowed us to describe a new hybrid zone between the local species *B. thermophilus* and a new species localised at 37°S. Certain genes having demonstrated peculiarities in their evolutionary histories have been subjected to specific scrutiny, evidencing purifying selection acting on an intron of the gene Elongation Factor EF1 α , or a relaxation of selective pressures on certain copies of the Heat Shock Protein HSP70.

Our study underlines the fact that the genetic structuring of hydrothermal species such as deep-sea mussels is ascribable primarily to geological conditions (volcanism, tectonics and transforming faults) and that this results on the long term in allopatric speciation. We also confirm that marine bivalves, including the deep-sea ones, are among the more polymorphic species on earth and that this polymorphism contains many slightly deleterious mutations.

DISCIPLINE : Biologie des populations et Evolution

MOTS-CLES

***Bathymodiolus*, Hybridation, Introgression, Sélection, Polymorphisme, Environnement Hydrothermal**

Institut des Sciences de l'Evolution – Montpellier (I.S.E.-M.)
Unité Mixte de Recherche UMII – CNRS (UMR 5554)
Université de Montpellier II CC 065
34095 MONTPELLIER cedex 05