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)

THESE DE DOCTORAT DE
L'UNIVERSITE PIERRE ET MARIE CURIE

Spécialité

Diversité du Vivant

Présentée par

Mlle Sophie PLOUVIEZ

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE

Phylogéographie comparée des espèces hydrothermales de la dorsale du Pacifique oriental

soutenue le 11 décembre 2009

devant le jury composé de :

M. François BONHOMME,
Directeur de Recherche,CNRS, Montpellier (Rapporteur, Président du Jury)
Mme. Sophie ARNAUD-HAOND, Chargée de recherche, Ifremer (Rapporteur)
Mme. Sarah SAMADI, Chargée de recherche, IRD (Examinateur)
M. Pierre CHEVALDONNE, Chargé de recherche, Endoume (Examinateur)
M. Dominique HIGUET, Professeur, Université Pierre et Marie Curie (Examinateur)
M. Didier JOLLIVET, Chargé de recherche, CNRS, Roscoff (Directeur de thèse)
M. François LALLIER, Professeur, Université Pierre et Marie Curie (Directeur de thèse)

Remerciements

Trois ans déjà et ce doctorat est maintenant terminé. Ce fût pour moi trois années formidables, tant du point de vue scientifique qu'humain. Je tiens donc à remercier un certain nombre de personnes pour ces moments partagés.

Tout d'abord, je tiens à remercier mes directeurs de thèse Didier Jollivet et François Lallier, tous deux scientifiquement brillants et humainement simples, sans qui rien de tout cela n'aurait été possible. Didier, je ne pourrais jamais suffisamment te remercier pour ton implication dans cette thèse. En tant qu'étudiant, c'est vraiment appréciable de pouvoir avoir une discussion scientifique et défendre ses propres idées. J'ai particulièrement apprécié travailler avec toi car tu es quelqu'un qui sait écouter et prendre en compte l'avis de chacun. Merci pour ton soutien en toutes circonstances. Ne change pas !

Merci aux autres membres de mon jury d'avoir accepté de juger mes travaux. Après m'avoir co-encadré en stage de Master 2, ce fût un honneur François (Bonhomme) de t'avoir en tant que rapporteur de thèse. J'espère que cette collaboration sur ces chers Lepeto portera ses fruits et te permettra également de garder un pied dans l'environnement profond. Merci à toi et également à Sophie Arnaud-Haond, Pierre Chevaldonné et Sarah Samadi pour leurs conseils et discussions. Merci aussi à Dominique Higuet d'avoir accepté de faire partie de mon jury, dommage que tu n'aies pas pu assister à ma soutenance.

Et puis une thèse se réalise dans une équipe et pour ma part dans deux équipes. Je tiens donc à remercier les membres des équipes EGPM et GAME. Ce fût un plaisir de travailler avec vous.

Merci à l'ensemble des ex-egpmiens pour ces 2 ans de thèse (précédés par mes stages de Master 1 et 2) passés en votre compagnie. Merci pour votre soutien et vos conseils lors de ma préparation de l'oral d'école doctorale (m'ayant permis de réaliser cette thèse), puis d'oraux de congrès et finalement de ma soutenance de thèse. Merci pour les discussions scientifiques lors des réunions de labo, mais également pour l'ambiance au sein l'équipe, les sorties terrain (me permettant d'observer de temps à autre des espèces non alcoolisées), les chocolats de noël, les

biscuits au retour des congrès, les repas d'équipe… Merci Myriam d'avoir fait suivre mon e-mail de demande de stage à Didier lors de mon Master 1, me permettant ainsi d'avoir un premier pied dans l'équipe. Merci Fred pour tes conseils avisés, notamment lors de mes comités de thèse. Chapeau Claire pour ta parfaite gestion du labo de bio mol, et merci pour tous tes conseils et ta bonne humeur permanente. Je vous souhaite à tous de vous épanouir dans vos nouvelles équipes respectives.

Merci aux gameurs et écophysiens pour leur aide et leur soutien lors de ma dernière année de thèse, les gâteaux d'anniversaire, d'acceptation d'articles scientifiques, de déménagement, de fêtes ou juste sans raison particulière… Merci à Stéphane et Eric pour les discussions hydrothermalo-évolutionnistes et à Dominique pour le génotypage des allozymes. Merci Arnaud pour tes conseils de bio mol et avoir tenté de m'initier à la « philosophie » de Rocky.

Merci également à Eric Thiébaut et Thierry Comtet pour les discussions hydrothermales.

Merci à mes collègues de bureau Baptiste, François, Didier en EGPM puis Arnaud, Matthieu, Joana en GAME, pour l'ambiance amicale qui y régnait et avoir fait progresser mes piètres connaissances musicales et/ou cinématographiques. Merci à Matthieu et Joana pour avoir partagé et participé aux quarts d'heure de folie quotidiens et m'avoir accompagné lors de certaines nuits de rédaction, mais également à Daphné, Mahdi, Alex, Baptiste, François, Didier, Cécile, Manue, Marie, Florence, Jean-Yves… pour les pauses partagées (avec ou sans café).

Je souhaite également remercier les personnes avec qui j'ai pu collaborer pour des articles publiés ou en préparation. Outre mes directeurs de thèse, merci donc à Tim Shank, Baptiste Faure, Claire Daguin-Thiébaut, Frédérique Viard, Dominique Le Guen, Nicolas Bierne, François Bonhomme, Odile Lecompte et Marjolaine Matabos. Merci à Carolyn Engel-Gauthier pour ses éditions et commentaires.

Je remercie également Matthieu Bruneaux pour son aide à l'élaboration du logiciel RabouterSequences (Plouneau pour les intimes), ainsi que Christophe

Caron et Erwan Corre pour leur aide dans le « débuggage » des divers logiciels informatiques. Merci à Chuck Fisher pour m'avoir permis de participer à ma première mission océanographique, awesome !

Merci également à Alain Paoli pour l'organisation, au combien importante, de la salle de conférence le jour J. Ainsi qu'à l'ensemble du personnel du Gulf Stream pour la confection de mon repas de thèse et leur professionnalisme tout au long de ces années.

Pour la partie fun de la thèse, je tiens également à remercier un certain nombre de personnes. Les stationnautes tout d'abord pour leur soutien jour après jour. Merci à Nathalie Oulhen, voisine de dortoir, voisine de bureau (et bientôt quasi voisine de postdoc), pour sa bonne humeur permanente ; mais également pour m'avoir appris à faire un western blot (manipe sans aucun rapport avec ma thèse), et m'avoir aidé dans mon installation à Roscoff (notamment en conduisant une certaine camionnette à vive allure). Merci à Matthieu, Mahdi et Joana pour le repas aux chandelles et le couscous servis dans le bureau en fin de rédaction. Merci à Didier pour m'avoir initié à la pêche aux ormeaux. Merci à Daphné et Nathalie pour les sorties en maillot de bain dans Roscoff. Merci à Baptiste, Manue et Elodie pour cette superbe pêche aux araignées dont les voitures se souviennent encore. Merci à Cécile pour les sorties shopping. Merci à Daphné et Virginie pour les trajets en votre compagnie vers nos régions natales (parfois agrémentés par Zora…). Merci à toutes ces personnes mais également Alex, Julie, François, Marie, Sabrina, Marjo… pour les différentes sorties et soirées organisées ou tout simplement les bons moments partagés. Un grand merci à toutes les personnes ayant participé à l'organisation de ma soirée de thèse, qui fut mémorable !

Merci aussi à mes amis non stationnautes, en particulier Natacha, Perrine et Marianne, et à ma famille pour leur soutien. Merci également à mon professeur de biologie de Terminale, Sylvain Leterme, sans qui mon parcours aurait sans doute été bien différent.

Je dédie cette thèse à ma maman, Christine, qui, je l'espère, aurait été fière de moi.

TABLE DES MATIÈRES

REMERCIEMENTS

CHAPITRE 1 : INTRODUCTION GENERALE A LA PHYLOGEOGRAPHIE 1

1. BIOGEOGRAPHIE ET PHYLOGEOGRAPHIE 1
2. MODELES DE DISPERSION EN POPULATIONS GEOGRAPHIQUEMENT SUBDIVISEES 3
3. APPROCHE DE LA PHYLOGEOGRAPHIE PAR LA THEORIE DE LA COALESCENCE 9
   3.1. Principe de la théorie de la coalescence 9
   3.2. Écart à la neutralité : impact de l'histoire démographique et de la sélection naturelle sur le coalescent 13
   3.3. Structure du coalescent au regard de la géographie 16
   3.4. Coalescence de gènes contre séparation des lignées intra-spécifiques 20
4. RELACHEMENT D'UN ISOLEMENT GEOGRAPHIQUE 22
5. BIOGEOGRAPHIE EN MILIEU HYDROTHERMAL PROFOND 24
6. DORSALE DU PACIFIQUE ORIENTAL (EPR) 30
   6.1. Vitesses d'accrétion et géomorphologie du rift 30
   6.2. Définition et principe d'une faille transformante et d'une microplaque 31
   6.3. Histoire de la formation de la dorsale EPR 32
      6.3.1. La microplaque Bauer: 10°S-18°S 34
      6.3.2. La paléoplaque Mathematician: 12°N-17°N 34
      6.3.3. Capture de la microplaque Bauer et formation des microplaques de Pâques 35
   6.4. Potentielles barrières à la dispersion 36
      6.4.1. Failles transformantes 36
      6.4.2. Courants océaniques profonds 37
7. STRUCTURE GENETIQUE DES ESPECES DE LA DORSALE DU PACIFIQUE ORIENTAL 38
   7.1. Quels modèles de dispersion pour les espèces de l'EPR ? 38
   7.2. Existence de barrières génétiques chez les espèces de l'EPR 40
8. OBJECTIFS DE LA THESE 42

CHAPITRE 2 : METHODES ET BIAIS METHODOLOGIQUES 43

1. STRATEGIE D'ECHANTILLONNAGE 43
2. BANQUES D'ADN COMPLEMENTAIRES 46
3. ACQUISITION DES SEQUENCES 47
   3.1. Séquences du gène mitochondrial Cytochrome Oxydase I 47
   3.2. Séquences des gènes nucléaires : Marquage-Recapture (MR) de séquences 48
4. TRAITEMENT DES SEQUENCES 50
   4.1. Alignement et correction des séquences liées aux erreurs de lecture du séquenceur 50
   4.2. Recombinaison in vitro détectée via la méthode MR 51
   4.3. Correction du jeu de données : suppression des singletons artéfactuels engendrant un multiallélisme 52

5. TRAITEMENT DE LA RECOMBINAISON 54
5.1. Biais liés à la recombinaison dans les analyses de coalescence *54*
5.2. Détection et suppression des séquences recombinantes *55*
6. DETECTION DE GENES CO-AMPLIFIES (POTENTIELLEMENT PARALOGUES) 55
6.1. Bathymodiolus thermophilus *55*
6.1.1. Lysozyme 55
6.1.2. Sulfotransférase 56
6.2. Lepetodrilus elevatus *57*
6.2.1. Lysozyme 58
6.2.2. Veliger Digestive Gland 3 (VDG3) 59
6.2.3. Textilinine 62
7. RECHERCHE D'UNE DIVERGENCE PARTAGEE ENTRE PLUSIEURS ESPECES : MSBAYES 63
8. MIGRATION A TRAVERS UNE BARRIERE AUX FLUX DE GENES : MIGRATE-N 66
9. MODELE D' « ISOLEMENT AVEC MIGRATION » 68
10. ÉCART A LA NEUTRALITE : SELECTION CONTRE DEMOGRAPHIE 69
10.1. Approche multilocus visant à détecter un ou plusieurs locus sous sélection positive : test de Hudson Kreitman Aguadé *69*
10.2. Approche substitutions synonymes / non-synonymes en vu de la détection de sélection positive versus négative : test de MacDonald-Kreitman *70*
10.3. Tests de polymorphisme visant à détecter un écart à la neutralité : tests de Tajima et Fu & Li *71*
10.4. Dater une expansion démographique via le logiciel Fluctuate *72*
10.5. Proposition d'un nouveau test statistique visant à détecter des locus sous sélection diversifiante entre régions géographiques *73*
10.6. Discrimination entre balayage sélectif et goulot d'étranglement via une approche de vraisemblance sur données multilocus : le logiciel Sweep_bott *74*

CHAPITRE 3 : PHYLOGEOGRAPHIE COMPAREE DES ESPECES HYDROTHERMALES DE LA DORSALE DU PACIFIQUE ORIENTAL A L'AIDE DU GENE MITOCHONDRIAL DE LA CYTOCHROME OXYDASE I 77

1. INTRODUCTION A L'APPROCHE DE PHYLOGEOGRAPHIE COMPAREE DES ESPECES 77
2. RESUME DE L'ARTICLE DE PHYLOGEOGRAPHIE COMPAREE ENTRE ESPECES 81
3. ARTICLE : PHYLOGEOGRAPHIE COMPAREE ENTRE ESPECES SUR LE GENE MTCOI 87
4. INTERETS ET LIMITES DE L'APPROCHE DE PHYLOGEOGRAPHIE COMPAREE 102
4.1. Apports de la phylogéographie comparée entre espèces dans l'étude des processus de vicariance *102*
4.2. Apports de la phylogéographie comparée dans l'étude de l'histoire démographique des populations *104*
4.3. Limites de la phylogéographie comparée monogène *105*

CHAPITRE 4 : ETUDE D'UNE BARRIERE SEMI-PERMEABLE A LA DISPERSION PAR UNE APPROCHE MULTI-LOCUS SUR PLUSIEURS ESPECES 107

1. INTRODUCTION A L'ETUDE D'UNE BARRIERE A LA DISPERSION PAR UNE APPROCHE MULTI-LOCUS 107
2. TAXONOMIE ET TRAITS D'HISTOIRE DE VIE DES ESPECES CIBLES 111
2.1. A. pompejana *111*
2.2. B. thermophilus *113*
2.3. L. elevatus *116*
3. MATERIELS ET METHODES DES ESPECES CIBLES 118
4. ANALYSES ET RESULTATS SUR LES ESPECES CIBLES 118
4.1. A.pompejana *118*
4.1.1. Résumé des résultats 118
4.1.2. Article 125
4.2. B. thermophilus *149*
4.2.1. Résumé des résultats 149
4.2.2. Article en préparation 154
4.3. L. elevatus *188*
4.3.1. Introduction 188
4.3.2. Gène mitochondrial COI 189
4.3.3. Analyse des séquences nucléaires : hybridation et barrière(s) au(x) flux de gènes 192
4.3.4. Y a-t-il eu une expansion démographique au Sud de l'EPR ? 207
5. DISCUSSION GENERALE DU CHAPITRE 4 : COMPARAISON DES RESULTATS OBTENUS ENTRE LES ESPECES CIBLES 211

CHAPITRE 5 : CONCLUSION, DISCUSSION GENERALE ET PERSPECTIVES 217

1. L'EPR : UNE OU DEUX PROVINCES BIOGEOGRAPHIQUES ? 218
2. CONNEXION ENTRE EPR ET LES AUTRES PROVINCES BIOGEOGRAPHIQUES 221
3. BARRIERES SEMI-PERMEABLES : ROLE DES FAILLES TRANSFORMANTES ET DES COURANTS DIVERGENTS ? 222
4. PERSPECTIVES DE LA THESE 223
4.1. Évolution des espèces hydrothermales *223*
4.2. Évolution des gènes *225*
4.2.1. Le gène mtCOI 225
4.2.2. Phosphoglucomutase d'*A. pompejana* 227
5. ANNEXE DU CHAPITRE 5 228

REFERENCES BIBLIOGRAPHIQUES 271

ANNEXE 1 : MARQUAGE RECAPTURE DE SEQUENCES 293

ANNEXE 2 : LOGICIEL RABOUTER SEQUENCE 299

ANNEXE 3 : CONDITIONS DE PCR 311

ANNEXE 4 : CODE GENETIQUE MITOCHONDRIAL DES ALVINELLIDAE 315

ANNEXE 5 : ARTICLE RELATION FEMELLES/JUVENILES CHEZ *BRANCHIPOLYNOE SEEPENSIS* 319

Chapitre 1 : Introduction générale à la phylogéographie

1. *Biogéographie et phylogéographie*

Subdiviser le monde vivant en multiples unités taxinomiques a permis aux scientifiques de mieux en appréhender sa diversité. Dès lors, des patrons communs de distribution géographique ont pu être mis en évidence entre taxons, posant ainsi la question des raisons historiques ayant conduit à cette répartition particulière. Repérer ces distributions communes à l'échelle des peuplements et identifier les processus historiques, environnementaux et biologiques, qui les ont façonnées, tel est l'objectif de la biogéographie (Myers & Giller, 1988). L'étude des compositions faunistiques et floristiques a déjà permis l'identification de nombreuses provinces biogéographiques, aussi bien en milieu terrestre (e.g. Cox 2001) qu'en milieu aquatique (e.g. Cox & Moore 2000). Si les barrières biogéographiques terrestres sont, du moins au premier abord, souvent relativement faciles à établir (correspondant à la séparation des continents par les océans ou à l'existence de montagnes ou de déserts), l'interconnexion des océans ainsi que la difficulté d'accessibilité de ce milieu rend plus complexe la visualisation de barrières à la dispersion et donc la définition de provinces biogéographiques marines. L'adaptation des organismes à un environnement donné est également à prendre en considération lors de la définition de provinces biogéographiques, rendant la tâche relativement complexe (e.g. continuum de températures). La profondeur des océans peut ainsi jouer un rôle fondamental dans l'étude de la biogéographie en limitant l'impact des changements climatiques sur la distribution des espèces (Cox & Moore 2000). En effet, en cas de changement de température, par exemple, les espèces peuvent atteindre leur preferendum thermique par migration latitudinale, comme en milieu terrestre, mais également par migration bathymétrique (en milieu terrestre, l'altitude peut également jouer ce rôle mais les points élevés restent relativement localisés).

La subdivision d'une région ancestrale mère en deux provinces biogéographiques filles peut être la conséquence de mécanismes divers. De nombreuses hypothèses ont ainsi été proposées pour expliquer ces subdivisions :

- des changements paléo-environnementaux liés aux mouvements tectoniques des plaques (théorie paléogéographique, Emsley, 1965) ou aux oscillations climatiques (e.g. théorie des refuges permettant aux zones géographiques les plus stables d'abriter certains taxons lors de contextes climatiques défavorables, Haffer, 1969)
- d'autres barrières physiques à la dispersion qui fragmentent l'habitat tels que des courants océaniques, un fleuve, un désert…
- des gradients environnementaux (Endler 1982) qui, par sélection naturelle, peuvent engendrer des gradients de distribution des taxons

Quelle que soit l'hypothèse, les capacités de dispersion et d'adaptation des espèces jouent un rôle fondamental dans leur répartition géographique et leurs permettent de coloniser de nouveaux territoires, de survivre aux conditions locales de nouveaux habitats et s'y maintenir. Ainsi, définir une province biogéographique implique que de multiples espèces ne présentent pas les capacités de dispersion et/ou une adaptabilité suffisantes pour atteindre la province biogéographique voisine et s'y établir. Néanmoins, certaines espèces, de par leur nature plus ubiquiste, sont capables de traverser une barrière biogéographique. Cela ne signifie pas pour autant qu'elles ne sont pas affectées par la barrière biogéographique, certains individus pouvant être plus dispersifs ou mieux adaptés aux contraintes de l'environnement. Pour évaluer l'impact intraspécifique de ces contraintes, il est donc nécessaire d'examiner le polymorphisme de ces espèces et d'apprécier les différences potentielles pouvant exister entre populations situées de part et d'autre de la barrière biogéographique (ségrégation allélique pour un caractère donné). L'utilisation d'outils moléculaires permet de comparer un nombre plus important de caractères considérés neutres (e.g. certains sites nucléotidiques), qu'une approche morphologique, par exemple, et maximise la probabilité de détecter cet impact potentiel. Ainsi, une analyse de la structure génétique de ces populations peut permettre de détecter une concordance entre barrière génétique et barrière biogéographique, et ainsi indiquer que cette dernière a également eu un effet d'isolement sur l'espèce étudiée. Par exemple, la barrière biogéographique de Cap Canaveral, qui sépare les faunes de l'Atlantique et du Golfe du Mexique, a également engendré de manière concomitante de la différenciation génétique

intraspécifique de part et d'autre de cette barrière (Avise 1992). De même, les populations de multiples taxons distribués de part et d'autre de l'archipel indo-australien présentent une divergence génétique accrue entre les océans Indien et Pacifique occidental (e.g. crabe des cocotiers *Birgus latro* : Lavery et al, 1995; crevette tigrée *Penaeus monodon* : Duda & Palumbi, 1999). Néanmoins, les barrières biogéographiques ne correspondent pas nécessairement à des subdivisions génétiques (évolution en allopatrie). La divergence intraspécifique de part et d'autre de la barrière biogéographique de Point Conception (séparant la Californie de l'Oregon) est extrêmement faible, indiquant que cette barrière biogéographique n'aurait pas réellement joué le rôle de barrière intraspécifique en limitant les flux de gènes entre populations (Burton 1998). Lorsqu'elle existe, la relation entre barrière génétique et barrière biogéographique n'est donc pas toujours simple à mettre à en évidence, d'autant plus que la localisation d'une barrière à la dispersion peut varier selon les espèces, engendrant des zones de transition plus ou moins larges entre provinces biogéographiques (e.g. zone de transition le long des côtes chiliennes, Thiel et al 2007).

La phylogéographie a pour objectif de comprendre les processus qui gouvernent la distribution géographique des lignées intraspécifiques (Avise et al 1987) par l'étude de l'histoire évolutive de leurs gènes. Le processus de subdivision phylogéographique le plus couramment admis peut ainsi se décliner en plusieurs étapes : (1) un isolement géographique en allopatrie dépendant notamment du modèle de dispersion de l'espèce considérée et de la force de la barrière rencontrée ; pouvant engendrer, accumulation de mutations de part et d'autre de la barrière et dérive génétique, et/ou maladaptation géographique, (2) un isolement génétique correspondant à cette subdivision géographique.

2. *Modèles de dispersion en populations géographiquement subdivisées*

Si le modèle panmictique correspond à un appariement sans contrainte des individus au sein de la ou des populations considérées (indice de différenciation génétique Fst = 0), à l'inverse, certaines populations subdivisées géographiquement peuvent n'échanger aucun migrant et conduire à une augmentation significative de la différenciation génétique par dérive (Hartl & Clark 2007). De nombreux modèles intermédiaires impliquant un échange de migrants sous contraintes ont alors été décrits,

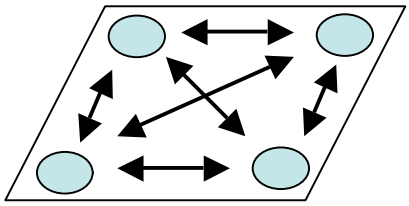
selon un degré plus ou moins variable de connectivité des populations géographiquement subdivisées.

Wright (1931, 1937) intègre la connectivité dans son modèle en îles (Fig. 1.1.A) en y ajoutant un paramètre de migration entre les populations étudiées. Ainsi, toutes les populations sont supposées présenter le même taux de migration m avec l'ensemble des autres populations. La différenciation entre les populations (mesurée par le Fst) peut ainsi permettre d'estimer la fraction d'individus (N x m) échangés entre populations (migrants) par la relation Fst = 1/(1+4Nm) sous l'hypothèse d'un état d'équilibre entre la migration et la dérive à chaque génération (populations démographiquement stables au cours du temps). Néanmoins, supposer que deux populations proches échangent entre elles autant de migrants que deux populations éloignées ne semble pas très réaliste. C'est pourquoi, d'autres modèles intègrent la notion de proximité géographique entre populations.

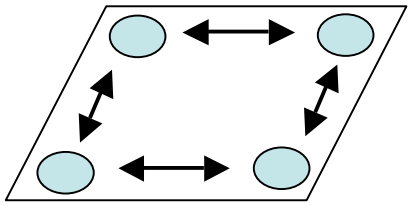
Le modèle en pas japonais (ou stepping stone, Kimura 1953, Fig. 1.1.B) considère que la dispersion ne s'effectue qu'entre populations voisines selon un même taux de migration m (celui-ci étant nul entre populations non adjacentes). Le modèle d'isolement par la distance (Wright 1943, Malécot 1950, Fig. 1.1.C) présente, quant à lui, un taux de migration proportionnel à la distance entre individus considérés. Ainsi, ce modèle permet la connectivité entre individus provenant de populations proches mais également, dans une moindre mesure, entre populations plus distantes (Wright 1943). Il suppose néanmoins que seule la distance est responsable du coefficient d'apparentement entre individus et n'intègre pas la présence potentielle d'une rupture abrupte de la migration par une barrière à la dispersion.

Le modèle avec matrice de migration libre (Bodmer & Cavalli-Sforza 1968, Fig. 1.1.D) permet de fixer un taux de migration différent entre chaque paire de populations, avec des flux pouvant être asymétriques. Ainsi, selon les paramètres de migration choisis, ce modèle peut prendre en compte à la fois l'isolement par la distance mais également de possibles barrières à la dispersion (migration réduite entre deux groupes de populations). Les modèles en îles, en pas japonais et d'isolement par la distance, constituent des cas particuliers du modèle avec matrice de migration libre. Néanmoins, intégrant les avantages de tous les modèles précédents, ce dernier nécessite l'estimation d'un nombre important de

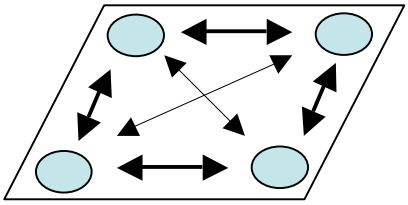
paramètres (n(n-1) paramètres de migration avec n correspondant au nombre de populations). C'est pourquoi il n'est que rarement utilisé.



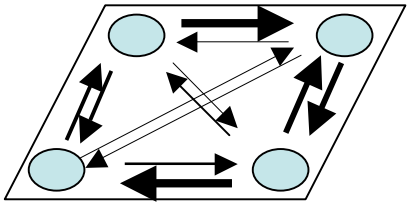
A. Modèle en île



B. Modèle en pas japonais



C. Modèle d'isolement par la distance



D. Modèle avec matrice de migration libre

Fig. 1.1 Modèles de dispersion en populations géographiquement subdivisées.
Les taux de migration sont proportionnels à l'épaisseur des flèches.

Selon les capacités de dispersion d'une espèce, un ajustement à un modèle de populations peut permettre de mieux appréhender son évolution dans le temps et l'espace. Ainsi, une espèce à forte capacité de dispersion sera moins soumise à l'isolement par la distance qu'une espèce à capacité de dispersion réduite et tendra à suivre un modèle en îles. Les espèces à capacité de migration très faible seront, quant à elles, plus proches d'un modèle d'isolement par la distance ou d'un modèle en pas japonais (suivant la fragmentation de l'habitat). L'isolement par la distance sera d'autant plus marqué que les capacités de dispersion des espèces seront faibles (Palumbi 2003, Fig 1.2).

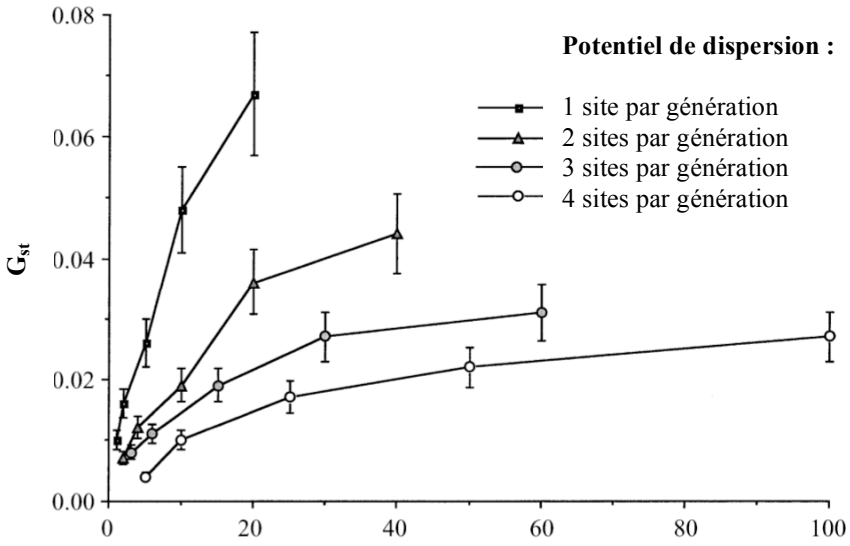


Fig. 1.2 Variations de la force de l'isolement par la distance avec la capacité de dispersion larvaire moyenne (en nombre de sites atteints par génération, la distance entre sites étant constante). Adaptée de Palumbi (2003).
Valeurs du Gst (analogue du Fst et mesurant la proportion de variabilité génétique qui est distribuée entre populations) simulées sur 10 locus bi-alléliques

Si ces modèles permettent de décrire la connectivité entre populations à un temps donné, ils n'intègrent cependant pas la nature dynamique des populations dans le temps. Les modèles de métapopulation (modèle initial proposé par Levin 1970 puis différents modèles dérivés revus dans Harrison & Hastings 1996 ; Fig 1.3) ont été introduits afin de tenir compte de la variabilité dynamique locale des populations en y intégrant la possibilité d'extinction et de recolonisation de ces populations par de nouveaux migrants (notion de colonisateurs).

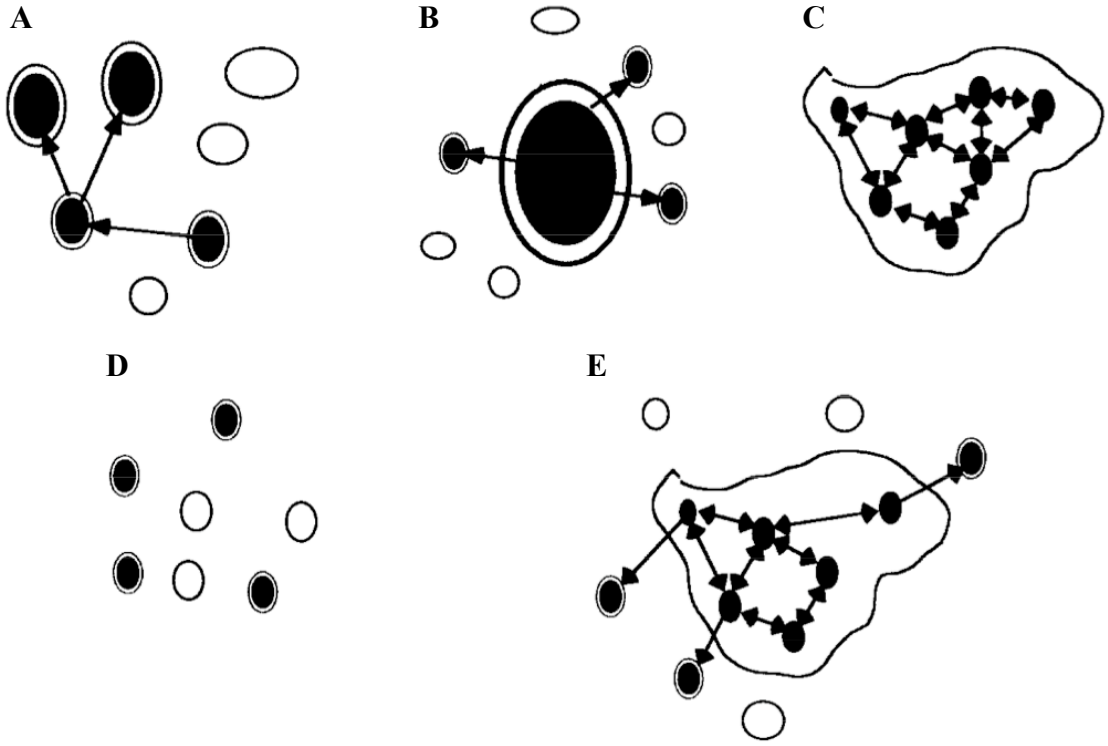


Fig. 1.3 Différents types de métapopulation. (A) type « classique » (Levins) ; (B) type « source-puits » ; (C) type « population fragmentée » ; (D) métapopulation en déséquilibre ; (E) cas intermédiaire combinant (A) et (D).
Symboles remplis : habitats occupés ; symboles vides : habitats vierges ; flèches : dispersion; lignes noires entourant les symboles remplis : limite de distribution des populations locales. D'après Harrison & Hastings (1996).

Deux modèles de recolonisation ont été proposés (Slatkin 1977) : le modèle « migrant » (recolonisation d'un nouveau site à partir de colonisateurs provenant de l'ensemble des autres populations de la métapopulation) et le modèle « propagule » (recolonisation d'un site à partir de colonisateurs provenant uniquement des populations voisines). Il a été montré que le taux d'extinction des populations locales et le modèle de recolonisation influencent fortement la structure génétique de la métapopulation lorsque ces modèles sont comparés aux modèles de référence sans extinction (Wade & McCauley 1988, Whitlock & McCauley 1990, Pannell & Charlesworth 2000).

Dans le modèle « propagule », la recolonisation ne pouvant s'effectuer qu'à partir des populations proches, les populations recolonisées présenteront alors une diversité génétique réduite (notion d'effet fondateur : cf. Pannell & Charlesworth 2000). La

différenciation génétique (Fst) entre les populations augmentera avec le taux d'extinction (Wade & McCauley 1988, Whitlock & McCauley 1990).
Dans le modèle « migrant », la recolonisation s'effectue à partir d'individus provenant de l'ensemble des autres populations. Si le nombre de colonisateurs est suffisamment grand (> 2 fois le nombre de migrants échangés entre populations déjà établies), alors la population recolonisée présentera une diversité génétique supérieure à celle des autres populations et la différenciation génétique entre populations diminuera d'autant plus que le taux d'extinction sera élevé (Wade & McCauley 1988, Whitlock & McCauley 1990). Au contraire, si le nombre de colonisateurs est inférieur à 2 fois le nombre de migrants entre populations établies, la diversité génétique des populations recolonisées sera alors plus faible que dans les autres populations et la différenciation génétique entre populations augmentera avec le taux d'extinction, comme dans le modèle « propagule ».

Extinctions et recolonisations influencent donc la différenciation génétique entre populations. Il est dès lors extrêmement difficile de relier la différenciation génétique au degré d'échange entre populations sans connaître au préalable le taux d'extinction appliqué aux populations. La relation entre Fst et nombre de migrants proposée par Wright (1931 ; Fst = 1/(4Nm+1) doit être adaptée pour tenir compte du caractère dynamique de la métapopulation et inclure 2 nouveaux paramètres : le taux d'extinction et la probabilité que 2 colonisateurs aient la même origine (Whitlock & McCauley 1999).

Ajoutées aux capacités de dispersion d'une espèce (et donc aux potentialités de recolonisation), les caractéristiques de l'habitat peuvent également permettre de choisir le modèle de dispersion le plus approprié. Ainsi, un habitat extrêmement variable dans le temps et dans l'espace induira des extinctions et recolonisations de populations plus fréquentes, mieux décrites par un modèle de métapopulation.

L'isolement génétique provoqué par une barrière à la dispersion sera donc différemment ressenti par les espèces en fonction de leur modèle de population. Selon le modèle considéré, l'impact d'une barrière à la dispersion pourra ainsi être évalué par des approches de génétique des populations basées sur l'écart à l'équilibre Hardy-Weinberg, en utilisant les fréquences alléliques et leur distribution dans l'espace sous l'hypothèse d'un équilibre migration-dérive. Si ces approches en fréquence permettent une vision relativement récente de la différenciation entre populations, la quantification de la

divergence accumulée entre ces populations nécessite, quant à elle, une approche différente basée sur l'accumulation des mutations au cours du temps sous l'hypothèse d'un équilibre mutation-dérive (échelle de temps plus ancienne) et utilisant les arbres phylogénétiques ou les généalogies d'allèles (ou coalescents).

3. *Approche de la phylogéographie par la théorie de la coalescence*

3.1. Principe de la théorie de la coalescence

À chaque génération, des erreurs de réplication de l'ADN se produisent. Lorsqu'elles s'exercent sur les lignées germinales (patrimoine génétique transmis à la descendance), ces erreurs peuvent être transmises à la génération suivante. Les mutations, définies comme changement héréditaire dans le matériel génétique, sont des événements rares (10^{-4} à 10^{-6} nouvelles mutations par gène et par génération) et constituent pourtant la source primordiale de variation génétique entre individus. Les évènements de mutation se produisent aléatoirement dans le temps suivant une loi de Poisson. Au fil du temps, certaines de ces mutations vont se propager au sein d'une population jusqu'à se fixer avec un temps moyen de 4N (où N est la taille de la population diploïde) tandis que d'autres seront éliminées par dérive génétique. À taille de population constante, on atteint rapidement un état d'équilibre entre le nombre de mutations nouvellement apparues et le nombre de mutations qui disparaissent par dérive génétique : c'est l'équilibre mutation-dérive (e.g. Fig. 1.4). De même, un modèle en nombre de sites nucléotidiques infini peut, dans certains cas, être supposé pour simplifier les analyses (chaque nouvelle mutation survient sur un site différent, ce qui a pour conséquence de ne pas considérer l'existence de sites homoplasiques et de mutations pouvant retourner à l'état ancestral). Ces modèles apparaissent néanmoins extrêmement simplificateurs puisqu'il existe des 'hot spots' (ou points chauds) de mutations dans les génomes (Li & Graur 1991).

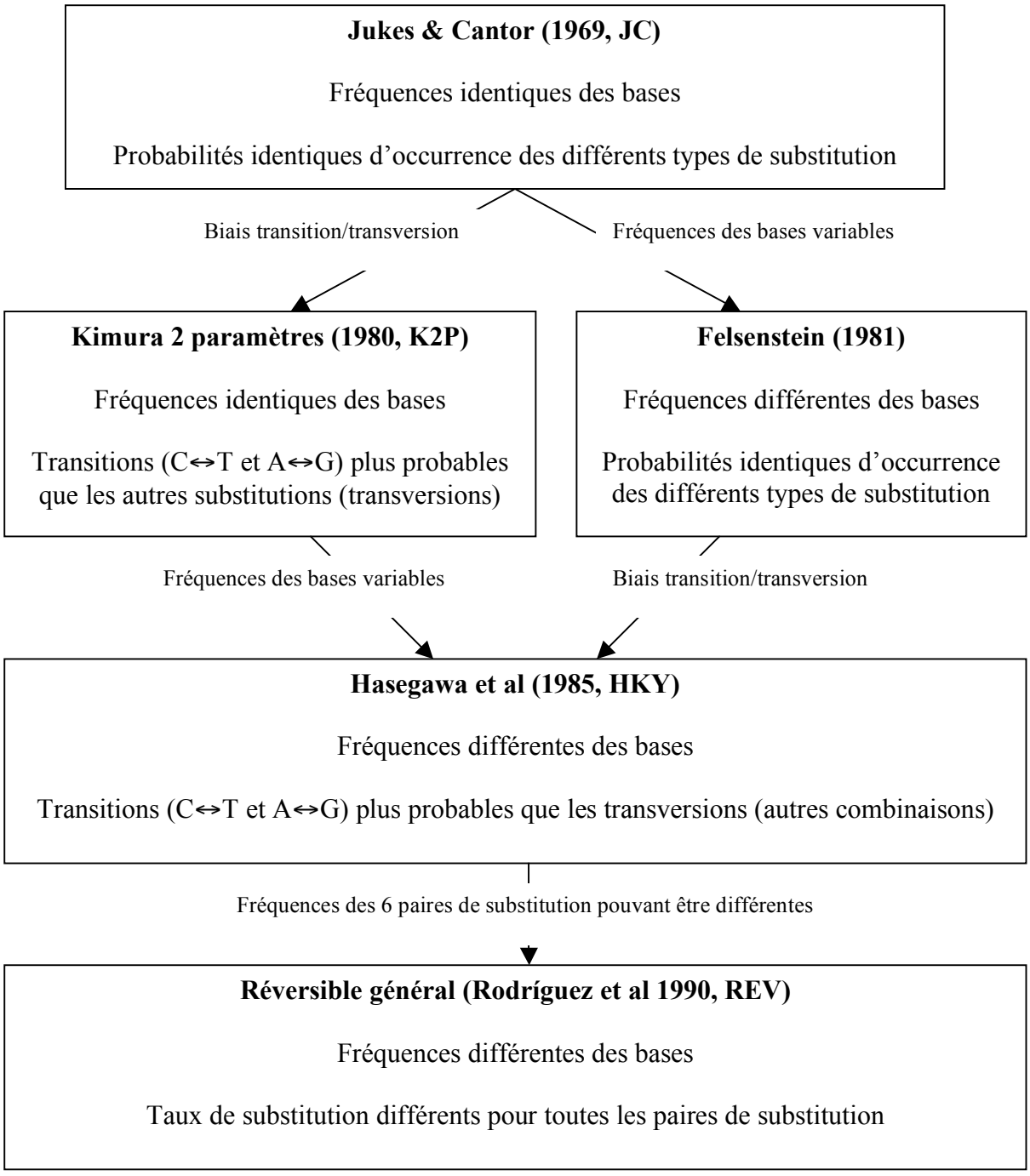


Fig 1.4 Relation entre cinq modèles de substitution. Adaptée de Page & Holmes (1998)

Le devenir d'une mutation va dépendre de l'impact des autres forces évolutives (i.e. sélection, dérive génétique, migration) sur le site muté, la sélection pouvant également affecter les éventuelles mutations qui lui sont liées (notion d'autostop génétique). Ainsi, si le temps d'isolement géographique entre deux populations est suffisamment long, des mutations différentes entre ces populations pourront s'accumuler par sélection différentielle ou simple dérive génétique, pouvant donner lieu à de nouveaux allèles entre ces populations. Le polymorphisme moléculaire (i.e. basé sur les séquences nucléotidiques d'ADN) présent s'inscrit donc comme une empreinte des évènements passés, et notamment des événements d'isolements géographiques. La théorie de la coalescence (Kingman 1982) consiste à utiliser cette empreinte afin de « rembobiner le fil de l'évolution » et de décrypter l'histoire des populations (Fig. 1.4.A). Elle contraste donc avec les modèles existants jusqu'alors, qui comparent les fréquences alléliques des populations de génération en génération (selon le cours du temps) afin d'en observer les changements (Wright 1977).

La théorie de la coalescence consiste à visualiser les relations entre les lignées actuelles en remontant dans le temps, jusqu'à trouver l'ancêtre commun le plus récent à l'ensemble de ces lignées (MRCA : Most Recent Common Ancestor, Fig. 1.5). La fusion de deux lignées au niveau d'un MRCA est appelée événement de coalescence. Pour une espèce diploïde d'effectif constant, la probabilité que deux lignées coalescent à la génération précédente est de 1/2N, où N est la taille de la population. Cette probabilité est la même quelle que soit la génération considérée. Dans un coalescent neutre (absence de sélection et taille de population constante), le temps qui sépare deux lignées qui coalescent augmente de manière exponentielle lorsque l'on remonte dans le temps, certaines lignées ayant été perdues au cours des générations (Kingman 1982, $T4 < T3 < T2 < T1$ sur la Fig. 1.5). Le coalescent représente ainsi la fusion de l'ensemble des lignées échantillonnées. La théorie de la coalescence ne s'intéresse donc qu'aux lignées ayant fournies des descendants actuels (Fig. 1.5.B). Les lignées ayant disparu ne sont pas prises en considération. La théorie de la coalescence peut ainsi permettre de mieux appréhender les processus ayant conduit à l'évolution des populations actuelles en étudiant l'écart du coalescent observé par rapport à un coalescent neutre et la distribution des lignées dans le coalescent au regard de la géographie.

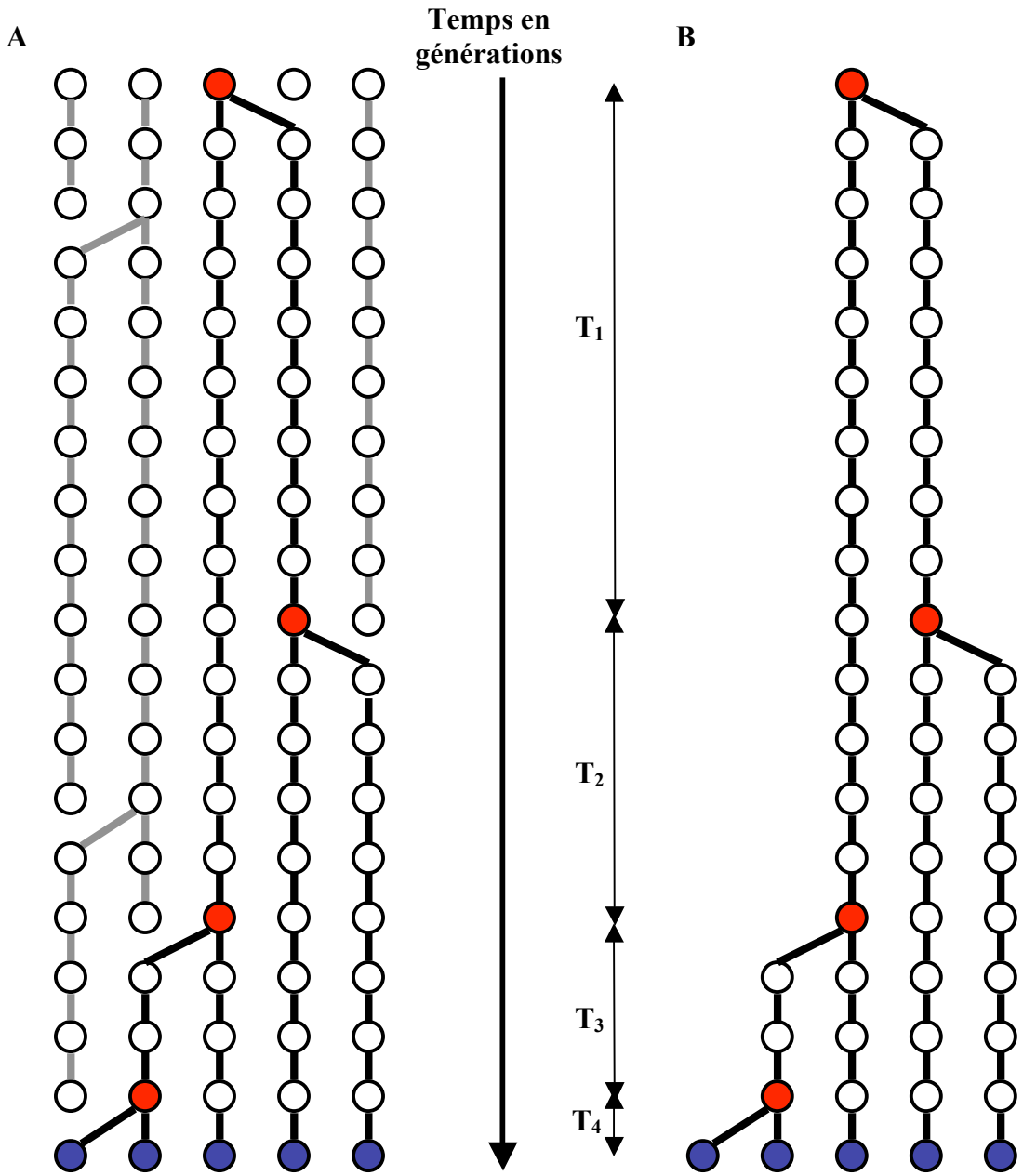


Fig. 1.5 Processus de coalescence sous l'hypothèse neutre.
Chaque cercle représente une lignée. La population est supposée de taille constante. En bleu, lignées actuelles échantillonnées. En rouge, ancêtre commun correspondant à un événement de coalescence. A) Modèle de Wright représentant l'ensemble des lignées au cours du temps. En noir, relation entre lignées ayant données lieu aux lignées actuelles échantillonnées. En gris, relation entre lignées dont la descendance n'a pas été échantillonnée. B) Modèle de coalescence de Kingman ne prenant en compte que les lignées ayant engendrées une descendance échantillonnée.

3.2. Écart à la neutralité : impact de l'histoire démographique et de la sélection naturelle sur le coalescent

L'étude de la distribution des mutations sur un coalescent peut apporter des informations précieuses sur les effets démographiques et/ou sélectifs pouvant intervenir dans les populations. En absence de sélection et d'événement démographique (taille de population constante), les mutations s'accumulent aléatoirement dans le polymorphisme puis dans la divergence d'un gène. La longueur des branches sur un coalescent étant proportionnelle au temps, les branches internes auront donc des longueurs supérieures aux branches terminales (Fig. 1.6.A) puisque ayant accumulé plus de mutations (égal au nombre de générations nécessaires pour arriver au MRCA).

Supposons maintenant qu'un événement démographique ait engendré une augmentation de la taille de la population à un temps t. La population observée actuellement (No) présente alors une taille supérieure à la taille de la population ancestrale au temps t (Nt). Deux lignées ancestrales (plus anciennes que t) présentent alors une probabilité plus grande de coalescer à la génération précédente que deux lignées actuelles ($1/2No < 1/2Nt$). En conséquence, le temps nécessaire pour que deux lignées actuelles coalescent sera plus long qu'entre lignées ancestrales et donc les branches terminales du coalescent ayant subi une expansion démographique seront plus longues que les branches internes (Fig. 1.6.B). Le coalescent aura une forme de râteau ou d'étoile (Hudson & Slatkin 1991). Les mutations auront donc plus de chance de s'accumuler sur les branches terminales. On observera un excès de variants rares (mutations en faible fréquence ou n'étant échantillonnées qu'une seule fois : singletons). À l'inverse, une diminution de taille de population engendrera des branches internes plus courtes que les branches terminales, avec excès de variants en fréquence intermédiaire. Une expansion démographique faisant généralement suite à une réduction de population (effet fondateur), la topologie observée sera alors un râteau avec une branche interne relativement longue (Fig. 1.6.C).

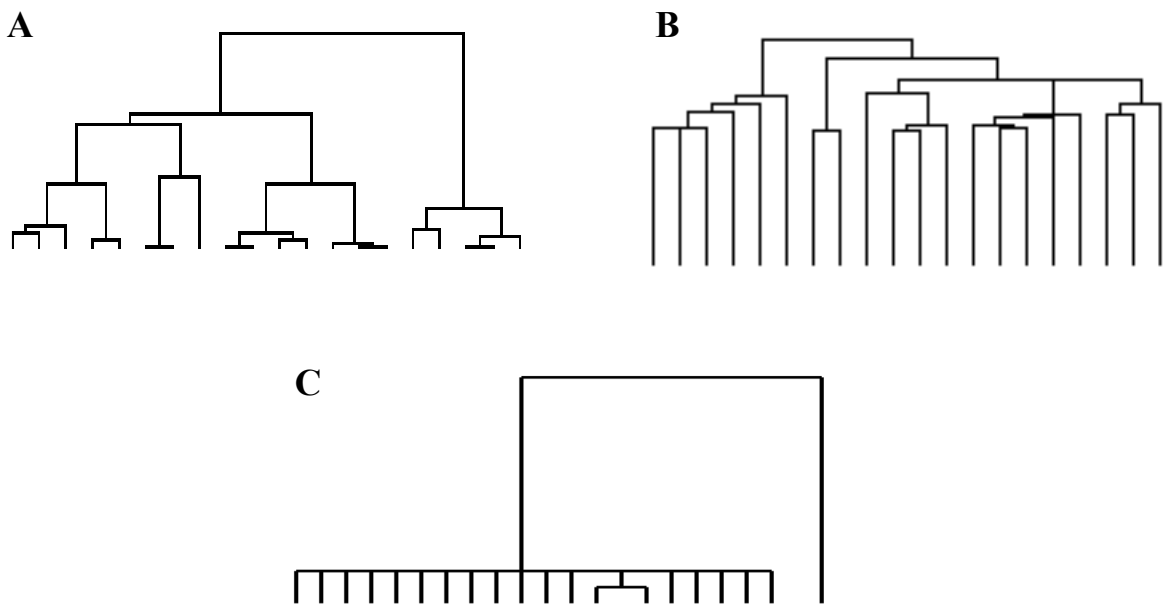


Fig. 1.6 Coalescents simulés selon (A) l'hypothèse neutre, (B) un facteur de croissance démographique de 1000 à l'aide du logiciel TreeToy 1.b, (C) coalescent attendu lors d'un goulot d'étranglement suivi d'une expansion démographique.

Outre les effets démographiques, les gènes sous sélection vont également présenter des topologies de coalescent qui s'écartent du coalescent neutre. Les différents types de sélection peuvent ainsi se classer en deux catégories selon qu'elles engendrent une réduction ou un maintien du polymorphisme.

La sélection positive consiste en la fixation d'un allèle avantageux dans la population diminuant ainsi la diversité génétique de manière drastique lors de l'augmentation en fréquence de la mutation avantageuse dans les générations qui suivent son apparition. De nouvelles mutations pourront alors s'accumuler dans le polymorphisme lorsque l'allèle avantageux aura remplacé toutes les autres lignées ancestrales (notion de balayage sélectif). Le coalescent présentera une topologie en râteau similaire à celle obtenue lors d'un goulot d'étranglement démographique (réduction drastique de la taille de la population et donc de la diversité puis expansion, Fig. 1.6.C). Certains gènes peuvent ne pas être directement sous sélection positive (mutation adaptative apparue ailleurs dans un gène proche) mais être entraînés dans le balayage (par effet autostop ou hitch-hiking) avec le gène qui possède l'allèle avantageux. En absence de recombinaison, ces gènes seront donc en déséquilibre de liaison avec le gène porteur de l'avantage adaptatif.

La sélection purificatrice (négative) conduit également à une réduction de diversité génétique en éliminant les mutations délétères (ou les mutations neutres ou faiblement avantageuses qui lui sont liées par l'effet d'un autostop : sélection d'arrière-plan ou background selection). Par définition, elle affecte tous les gènes ayant une fonction dans l'organisme et pour lesquels la structure de la molécule transcrite et/ou traduite nécessite d'être conservée. De manière générale sur les parties codantes d'un gène, elle aura pour conséquence de réduire la diversité génétique attendue sous l'évolution neutre pour une taille de population donnée (voir Li & Graur 1991). Néanmoins, son action dans le temps est plus diffuse (pas de diminution rapide de diversité génétique) que celle attendue lors d'une propagation d'un allèle avantageux. En conséquence, il est extrêmement difficile de distinguer l'action de la sélection purificatrice (ou sélection d'arrière-plan) d'un taux de mutation faible pour un gène donné ou d'un effet positif (ou balayage sélectif) extrêmement faible de la sélection (Charlesworth et al 1993). Le coalescent aura une topologie très semblable à une topologie neutre mais avec un nombre beaucoup plus réduit d'allèles.

À l'inverse, différents allèles d'un même gène soumis à la sélection balancée tendront à être conservés pendant une période de temps plus longue que des allèles neutres. La sélection balancée engendrera donc un maintien de polymorphisme plus important que sous la neutralité et présentera ainsi un coalescent avec des branches plus longues que dans un coalescent neutre (Fig. 1.7).

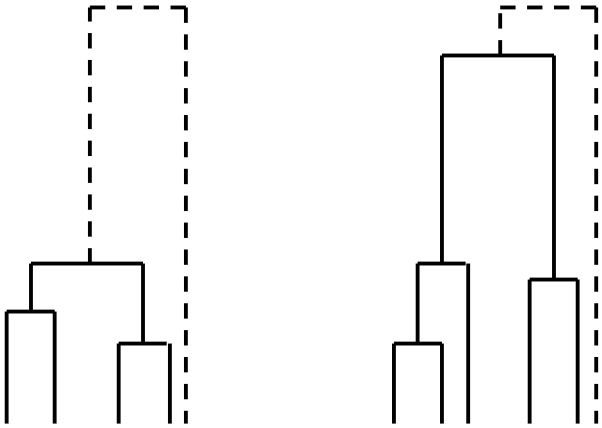


Fig. 1.7 Topologies d'arbres de coalescence (A) d'un gène neutre, (B) d'un gène soumis à sélection balancée.
En pointillés, racine de l'arbre.

Si la topologie d'un coalescent ne permet pas, à elle seule, d'affirmer un écart à une accumulation neutre du polymorphisme, elle constitue un indice précieux permettant de poser des hypothèses pouvant être testées par le biais de tests statistiques. Un effet démographique ou sélectif pouvant avoir le même impact sur la topologie du coalescent d'un gène, il peut être difficile de discriminer ces causes. Pour cela, il est alors nécessaire de procéder par une approche comparative : (1) un gène et plusieurs espèces si l'on soupçonne que l'environnement ait pu affecter la dynamique des populations d'une façon semblable au sein d'un peuplement (e.g. effet vicariant) ou (2) une espèce et plusieurs gènes. En effet, un événement démographique affecte l'ensemble des gènes d'une espèce de la même façon, ce qui n'est pas le cas de la sélection. Au contraire, la sélection n'affectera qu'un nombre limité de gènes (par convergence adaptative) au sein d'une espèce, ou d'un petit nombre d'espèces.

3.3. Structure du coalescent au regard de la géographie

Outre les effets sélectifs et démographiques, la topologie du coalescent peut également fournir de nombreuses informations sur les processus à l'origine de la distribution géographique actuelle de la lignée. Faire figurer sur le coalescent la localisation géographique des lignées échantillonnées permet de visualiser la correspondance possible entre évènements de divergence et région d'échantillonnage, pouvant révéler la présence de barrière aux flux de gènes liée à un isolement géographique (barrière à la dispersion, distance entre sites, voire habitats ou micro-habitats contrastés).

Nombre d'exemples existent dans la littérature ayant mis en évidence des barrières géographiques (voir par exemple les revues de Maggs et al 2008, Nord Est de l'Atlantique ; Soltis et al 2006 Nord Est de l'Amérique). La force de l'isolement génétique dépendra de la durée et de la force de l'isolement géographique.

Une barrière à la dispersion (relativement récente ou faible) pourra ainsi engendrer des différences de fréquences génétiques de part et d'autre de cette barrière. L'utilisation de coalescents et le calcul de fréquences alléliques au sein de chaque site constituent un moyen élégant de détecter ce type de barrière phylogéographique. Outre la représentation de la coalescence utilisant des arbres phylogénétiques, la représentation via des réseaux de séquences permet de visualiser plus facilement les fréquences des différents allèles, chaque cercle étant proportionnel au nombre de séquences identiques échantillonnées. Par exemple, Goldstien et al (2006) mettent en évidence la présence d'une barrière phylogéographique chez trois espèces de patelles du genre *Cellana* qui correspond à l'actuelle séparation des îles nord et sud de la Nouvelle Zélande (Fig 1.8). Ces auteurs attribuent l'isolement géographique au complexe hydrographique dans la région du détroit de Cook. En couplant réseaux d'haplotypes et distribution géographique des haplotypes sur le gène mitochondrial de la Cytochrome b, les auteurs ont ainsi montré que les haplotypes les plus fréquents, pourtant différents entre eux d'une mutation, présentaient des fréquences différentes entre les deux îles, avec des valeurs de différenciation génétique variable selon l'espèce (Φst de 0,829, 0,142 et 0,177 pour *C. ornata*, *C. radians* et *C. flava* respectivement).

Une barrière à la dispersion plus intense ou plus ancienne peut aussi engendrer une divergence sensiblement équivalente entre lignées phylogéographiques au sein d'une même espèce (e.g. Fig. 1.9, de Moritz & Faith 1998), voire de la co-spéciation (e.g. Dooh et al 2006, Jolly et al 2006).

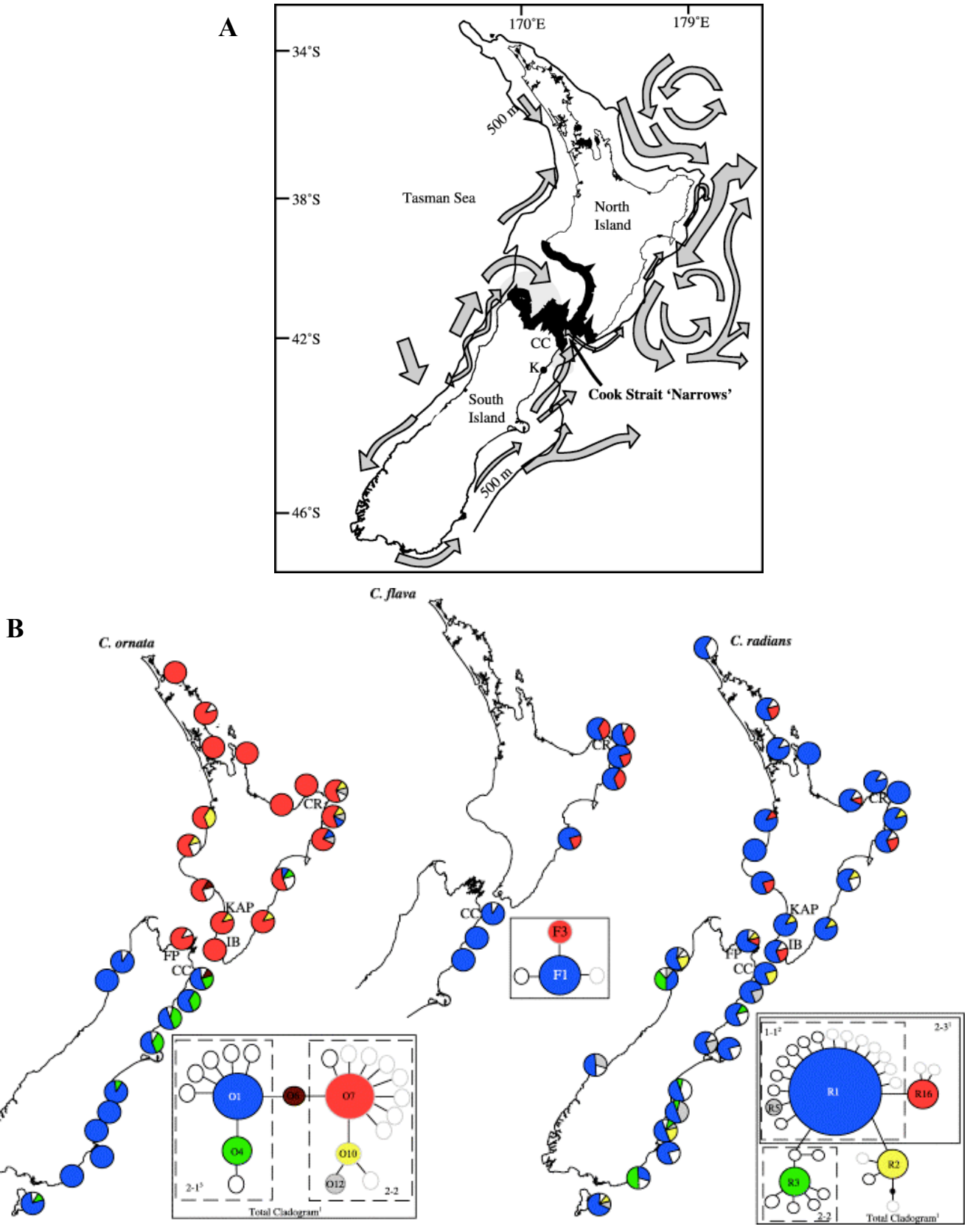


Fig. 1.8 Illustration du couplage entre approches issues de la théorie de la coalescence (réseaux d'haplotypes) et de la génétique des populations « classique » (fréquences haplotypiques) sur 3 espèces du genre *Cellana*. Tirée de Goldstien et al (2006).
(A) Carte géographique soulignant les caractéristiques hydrographiques de la Nouvelle Zélande. (B) réseaux d'haplotypes et distribution de ces haplotypes, les couleurs des haplotypes sur les réseaux correspondent à celles présentent sur les cartes.

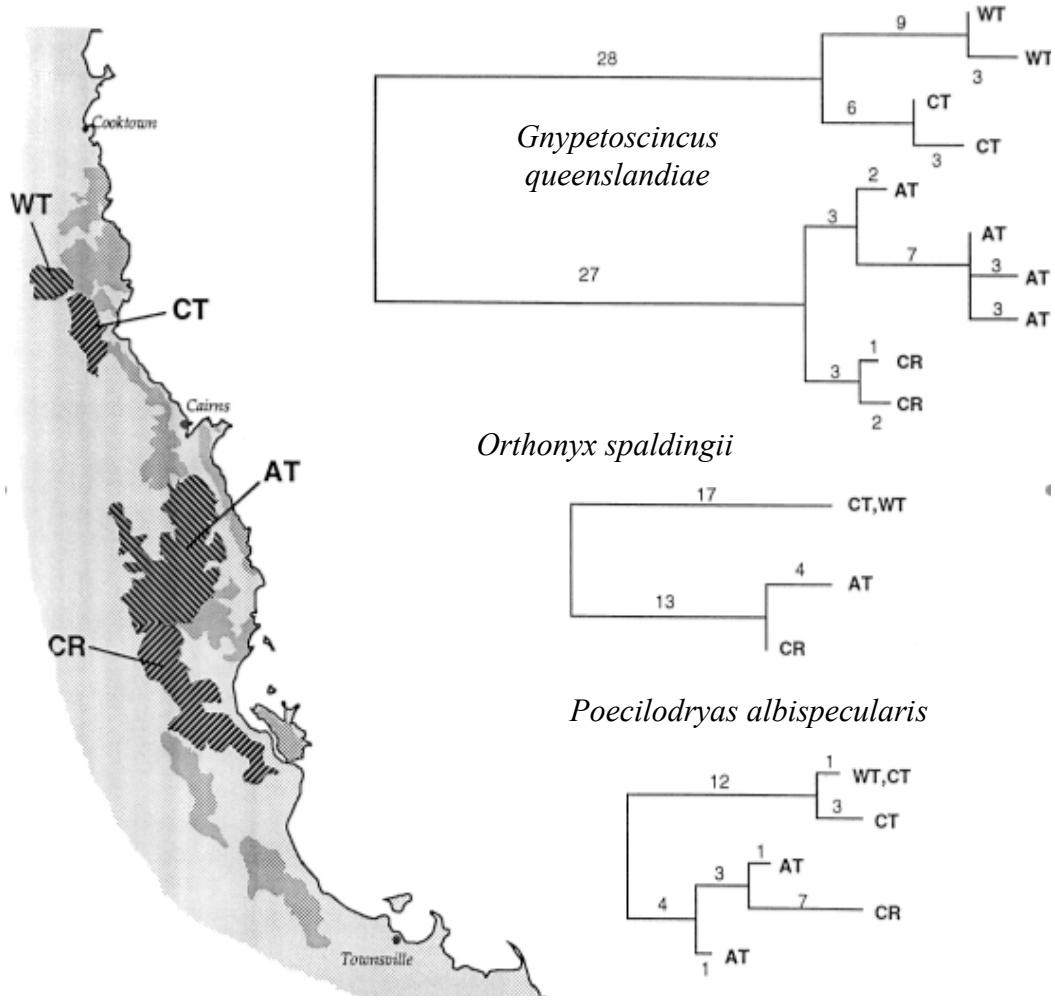


Fig. 1.9 Détection d'une barrière aux flux de gènes via une approche phylogéographique. Tirée de Moritz & Faith (1998)
Arbres phylogénétiques réalisés chez le reptile *Gnypetoscincus queenslandiae* et les oiseaux *Orthonyx spaldingii* et *Poecilodryas albispecularis* au nord est de la forêt tropicale Queensland (Australie). WT, Windsor Tableland ; CT, Carbine Tableland ; AT, Atherton Tableland ; CR Cardwell Range. Les nombres correspondent à l'estimation de la distance de divergence de Kimura-2-P x1000.

Si le temps de divergence séparant deux lignées peut permettre d'avoir une estimation de la date de l'isolement géographique, la persistance d'évènements de migration, même restreints, après l'initiation d'une barrière géographique, peut engendrer une mauvaise estimation du début de l'isolement (Nielsen & Wakeley 2001). Beaucoup de modèles considèrent que les populations géographiquement distinctes : (1) échangent des migrants à un taux constant pendant un temps infini ou (2) ont divergé à partir d'une population ancestrale à un temps t, avec absence totale de flux de gènes entre ces

populations géographiques depuis t. Il est clair que le processus d'isolement est souvent progressif avec des possibilités de retour en arrière (notion de zones de contact secondaire). De fait, la perméabilité de la barrière aux flux de gènes peut être incomplète Une divergence observée faible mais robuste sur un coalescent peut ainsi être due soit à (1) une barrière à la dispersion complète (pas de migration depuis l'isolement) mais récente ou (2) à une barrière à la dispersion plus ancienne mais plus perméable entraînant de la migration sporadique entre les populations isolées et la création d'allèles recombinés entre formes parentales (Nielsen & Wakeley 2001). Si l'isolement génétique des populations s'est effectué en présence de migration, le temps de divergence estimé en considérant l'absence de migration ne correspond alors pas au début du phénomène d'isolement géographique (temps sous estimé). Des modèles d'isolement avec migration ont donc été récemment introduits pour tenir compte des paramètres de migration dans l'estimation des temps de divergence (Nielsen & Wakeley 2001). La position géographique de la barrière et sa largeur peuvent également varier en fonction des capacités de dispersion des espèces, définissant de véritables zones de transition entre ces régions phylogéographiques.

3.4. Coalescence de gènes contre séparation des lignées intra-spécifiques

Si la coalescence est un outil puissant permettant notamment d'appréhender plus aisément l'évolution d'un gène, la topologie d'un coalescent basée sur un gène ne correspond pas nécessairement à celle de l'espèce (ou des lignées intraspécifiques, Hudson 1983, Pamilo & Nei 1988, Maddison 1997). Ainsi le coalescent d'un gène A peut correspondre à la topologie réelle entre lignées intraspécifiques alors que celui d'un gène B présentera une topologie différente (Fig. 1.10). Cela peut être dû à des phénomènes de duplication du gène dans une des lignées, la perte du gène dans une des lignées mais également à la stochasticité du processus de coalescence, notamment si celui-ci est exacerbé par des changements démographiques locaux. Lorsqu'une seule séquence est échantillonnée par lignée, la probabilité qu'un gène ne présente pas une topologie concordante avec la topologie réelle a été évaluée, par Pamilo & Nei (1988), à $(2/3)\ e^{-T}$, où T correspond au temps séparant l'ancêtre commun le plus récent (MRCA) entre les lignées les plus proches et le MRCA à toutes les lignées (Fig. 1.10).

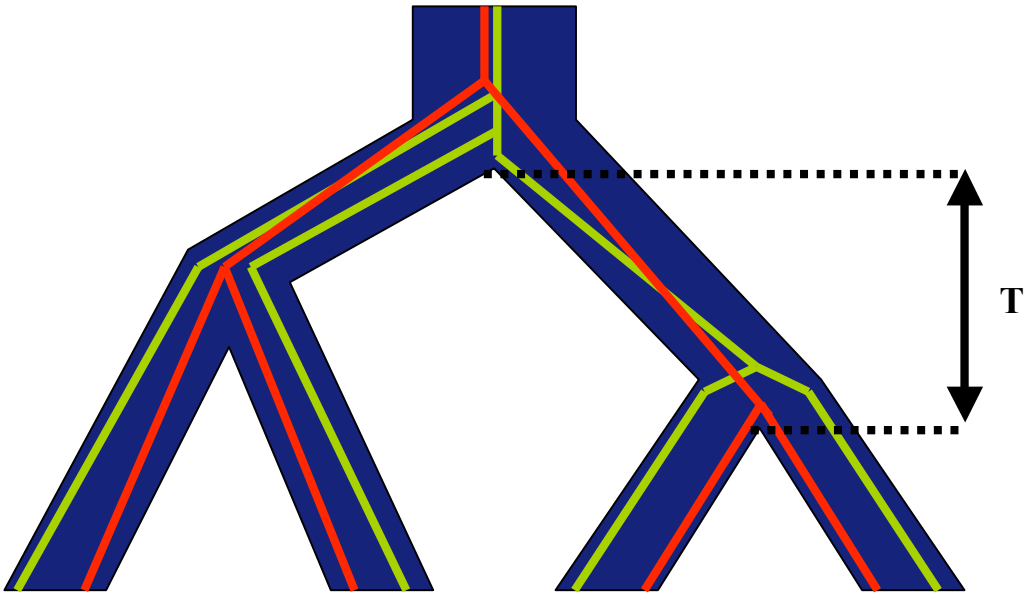


Fig. 1.10 Coalescence de gène contre coalescence de lignées intraspécifiques.
En bleu, représentation de la coalescence réelle des lignées intraspécifiques. En rouge, coalescent d'un gène A en accord avec la topologie réelle de ces lignées. En vert, coalescent d'un gène B ne reflétant pas la coalescence réelle de ces lignées.

Pour illustrer cette différence entre coalescence de gènes et coalescence d'espèces, prenons l'exemple des mandarins diamants du genre *Poephila*. Jennings & Edwards (2005) ont estimé des coalescents de 30 gènes indépendants sur ces oiseaux et ont mis en évidence une divergence génétique entre espèces pour 28 d'entre eux. Si 16 de ces locus montraient un regroupement de *P. acuticauda* (P.a.) et *P. hecki* (P.h.) par rapport à *P. cincta* (P.c.), soit une topologie : ((P.a., P.h.), P.c.) considérée comme représentative de la réalité ; 7 autres locus présentaient une topologie ((P.a., P.c.), P.h.) et les 5 locus restant une topologie ((P.c., P.h.), P.a.). Le calcul de la divergence T à partir des données observées, ne considérant que les 16 premiers locus, est concordant avec le T simulé sur l'ensemble des locus en considérant que ces 16 locus concordent avec la topologie réelle et que les autres locus constituent un 'pool' de gènes s'écartant par dérive du coalescent 'vrai' (Wakeley 2009). Il est donc nécessaire de comparer les topologies de plusieurs gènes et de rester critique quant aux coalescents obtenus sur un faible nombre de gènes pour expliquer l'histoire évolutive d'une espèce.

4. *Relâchement d'un isolement géographique*

Les barrières à la dispersion se mettent en place progressivement (e.g. fermeture de l'Isthme de Panama : Knowlton 2000) et ne sont pas immuables dans le temps. Ainsi, si la mise en place d'une barrière peut engendrer un isolement géographique puis génétique des populations, le relâchement de sa perméabilité peut permettre de nouveaux échanges entre lignées dès lors que l'isolement reproductif n'est pas atteint, et favoriser des zones de contact secondaire entre les entités taxonomiques précédemment séparées (e.g. contact secondaire au niveau du Cap Canaveral chez le crabe du genre *Menippe* : Liu et al 1991, ou le gastéropode du genre *Stramonita* : Bert & Arnold 1995). Si cet isolement des populations a été suffisamment long, ces lignées peuvent ne plus être interfécondes et seront alors considérées comme des espèces à part entière. Il en résulte alors une simple juxtaposition des lignées dans les zones de sympatrie. Dans le cas contraire, la remise en contact secondaire peut être suivie par la formation d'individus hybrides entre lignées génétiques, voire d'une introgression d'allèles d'un génome dans l'autre, si l'individu hybride est capable de se croiser avec les formes parentales. S'ils sont défavorisés, une sélection contre les hybrides peut tendre à augmenter l'isolement reproductif, s'opposant ainsi à la migration (force homogénéisante). L'équilibre entre isolement reproductif et migration formera alors une zone d'hybridation (Barton 1979, Barton & Hewitt 1985), avec la mise en place d'un gradient de fréquences alléliques (cline) autour de la barrière géographique ayant engendré la divergence entre les lignées. L'intensité du cline sera variable en fonction de l'implication du gène étudié dans le processus d'isolement reproductif (ou du degré de liaison des gènes avec ces gènes « verrous », Kruuk et al 1999, Fig. 1.11).

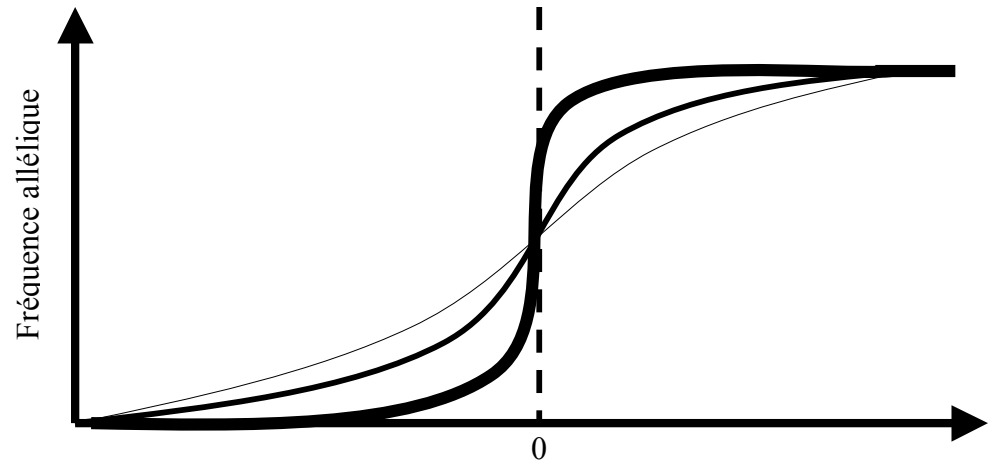


Fig. 1.11 Intensité du cline géographique en fonction de l'implication du gène dans l'isolement reproductif des espèces situées de part et d'autre de la barrière. Adaptée de Kruuk et al (1999)
En pointillés, figuration de la barrière aux flux de gènes. Les courbes représentent le cline de fréquences alléliques pour trois gènes. Plus l'épaisseur de la courbe est importante, plus le gène est impliqué dans l'isolement reproductif (ou lié à un gène impliqué).

5. *Biogéographie en milieu hydrothermal profond*

Au fond des océans, les dorsales océaniques séparant deux plaques tectoniques sont le lieu de l'expansion de la croûte océanique par remontée de magma basaltique. À ce niveau, l'eau de mer froide s'infiltre par les fissures associées aux forces de cisaillement et pénètre jusqu'à la chambre magmatique. Elle se charge alors en métaux lourds, en gaz et en composés réduits, s'échauffe et remonte en surface (Zierenberg et al 2000). Cette eau transformée reste en phase liquide grâce aux pressions hydrauliques importantes présentes à ces profondeurs. Elle constitue le fluide hydrothermal. Le contact entre ce fluide et l'eau froide du fond des océans précipite les éléments métalliques et forme alors les cheminées hydrothermales. Géologues et géochimistes ont longtemps débattu sur la présence de fluide hydrothermal potentiel associé aux dorsales océaniques (zones d'accrétion), et ont ainsi effectué des campagnes océanographiques explorant ces dorsales à la recherche d'infiltration d'eau profonde dans le plancher océanique. Si la vie en absence de lumière et avec des pressions hydrostatiques importantes fût mise en évidence en 1872 lors de l'expédition Challenger (échantillonnage de faune abyssale : Fukushima 2000), les propriétés physico-chimiques de ce fluide (fortes températures, pH acide, anoxie, fortes concentrations en H_2S) laissaient penser à l'impossibilité d'une forme de vie associée au fluide hydrothermal. Pourtant, c'est bien là, associée à ces cheminées actives, que le 15 février 1977, des géochimistes et géologues américains ont découvert la présence d'une faune foisonnante de vers géants et de bivalves sur la ride des Galápagos (Lonsdale 1977).

Bien que les organismes hydrothermaux consomment de l'O_2 (en provenance de la surface des océans) pour vivre, cet écosystème ne dépend pas directement de la lumière et de la production primaire photosynthétique. Il est essentiellement basé sur la chimiosynthèse effectuée par des bactéries vivant libres ou associées aux métazoaires sous forme de symbiose (Childress & Fisher 1992). C'est pourquoi, les espèces hydrothermales se répartissent de manière concentrique autour des émissions de fluide hydrothermal, source des éléments réduits nécessaires à cette chimiosynthèse bactérienne. Cette organisation spatiale particulière dépend des besoins trophiques des animaux, de leurs capacités d'adaptation à la toxicité du fluide et des compétitions inter-spécifiques pour la ressource. Ainsi, les communautés hydrothermales forment de véritables « oasis » (Laubier

1986) séparés parfois par plusieurs dizaines de kilomètres de « désert » abyssal. Elles sont donc entièrement dépendantes des sources hydrothermales actives et, de ce fait, la plupart des espèces formant ces communautés sont spécialistes de ce milieu et disparaissent à l'arrêt des émissions.

Or, la durée de vie de ces sources hydrothermales est limitée dans le temps et dépend en partie de la vitesse d'accrétion de la dorsale sur laquelle elles se forment. Les fissures par lesquelles l'eau ressort sous la forme du fluide hydrothermal peuvent s'obstruer, par colmatage des conduits provoquant ainsi le tarissement de la source. D'autres fissures se reformeront plus ou moins loin de la source éteinte, donnant ainsi naissance à une nouvelle source active. L'instabilité du milieu est donc à la fois spatiale et temporelle. Plus l'accrétion est rapide, plus le milieu est instable. Les populations connaissent ainsi de véritables phases d'extinction par destruction rapide de l'habitat (épanchements de lave sur les sites d'activité) affectant souvent la quasi-totalité d'un champ hydrothermal (Haymon et al 1991, Jollivet 1993, Tunnicliffe et al 1997). Ces phases d'extinction sont très rapidement suivies d'un 'bloom' bactérien et d'une recolonisation massive de cet habitat, d'abord par la faune vagile (crabes, galathées) puis par la faune sessile (vestimentifères) avec des taux de croissance extrêmement élevés (Lutz et al 1994, Shank et al 1998) faisant de ce milieu, un cas exemplaire de fonctionnement en métapopulation (Vrijenhoek 1997, Jollivet et al 1999).

Les sources hydrothermales sont présentes dans tous les océans (Fig. 1.12). Si de nombreux sites hydrothermaux ont d'ores et déjà été découverts, le système de dorsale médio-océanique représente près de 60 000 km de long et reste encore largement inexploré. Bien peu de missions ont été effectuées le long des dorsales Antarctique et Indienne (Fig. 1.12), malgré la présence de sources hydrothermales récemment détectées le long de ces dorsales (Bohrman et al 1999). Il n'est ainsi pas rare de mettre à jour de nouvelles sources actives lors de plongées sur des zones encore non visitées, la dernière découverte en date a ainsi été effectuée en août/septembre 2008 sur la dorsale du Pacifique oriental (ou East Pacific Rise, EPR ; 102,6º Ouest / 2,1º Sud) lors de la mission chinoise DY115-20.

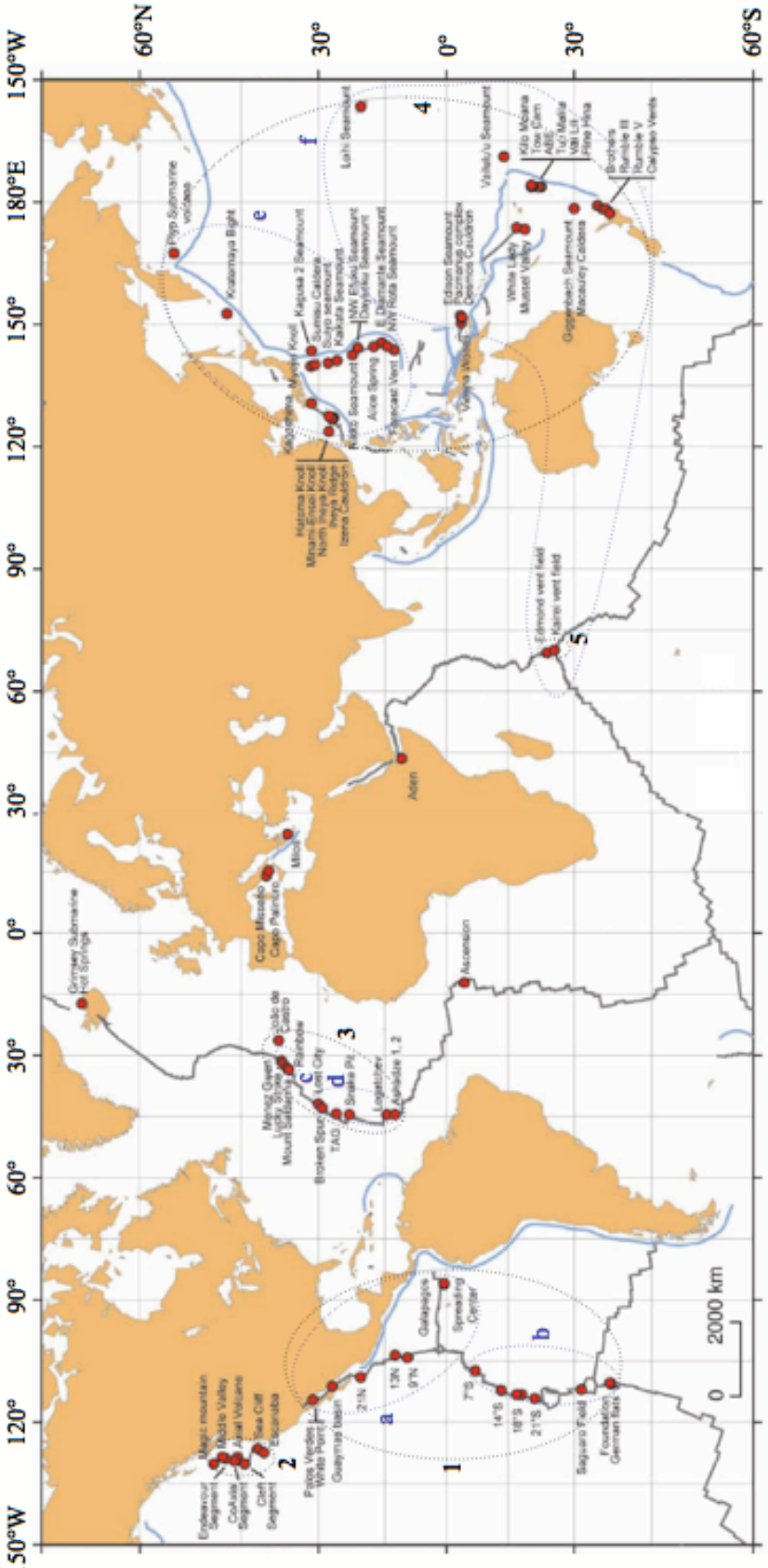


Fig. 1.12 Localisation des sources hydrothermales et des provinces biogéographiques. Adaptée de Desbruyères et al (2006b)
En cercle noir, les provinces biogéographiques généralement admises : 1, dorsale du Pacifique oriental ; 2, dorsale du Pacifique nord ; 3, dorsale médio-atlantique ; 4, bassins arrière arc du Pacifique occidental ; 5, sources hydrothermales de l'océan Indien.
En cercle bleu, les provinces biogéographiques proposées par certains auteurs (voir texte) : a, dorsale du Pacifique oriental nord et Galápagos ; b, dorsale du Pacifique oriental sud ; c, région des Azores ; d, dorsale médio-atlantique au sud des Açores ; e, Pacifique occidental nord ; f, Pacifique occidental sud et sources hydrothermales de l'océan Indien

Depuis la découverte de la faune hydrothermale, le nombre d'espèces décrites n'a cessé d'augmenter (533 recensées dans Desbruyères et al 2006b, mais des dizaines d'autres espèces sont actuellement en cours d'identification). Leur distribution géographique a rapidement suscité l'intérêt des biogéographes. Malgré la forte homogénéité apparente des conditions hydrothermales à l'échelle du globe, la composition faunistique du milieu hydrothermal présente en effet de nombreuses disparités géographiques. L'ensemble des auteurs s'accorde pour définir au moins cinq provinces biogéographiques différentes (Tunnicliffe 1991, Tunnicliffe et al 1998, Van Dover et al 2002, Tyler et al 2003, Fig. 1.12).

L'EPR fut la première province biogéographique identifiée. Sa composition faunistique (Fig. 1.13) est dominée en biomasse par les vers géants *Riftia pachyptila* (polychaetes Siboglinidae), les bivalves *Bathymodiolus thermophilus* et *Calyptogena magnifica* et les espèces de polychètes Alvinellidae (telles que le vers de Pompéi *Alvinella pompejana*, animal résistant à des températures élevées, pouvant atteindre, pendant de courtes durées, 105°C, Chevaldonné et al 1992) et, en nombre, par des gastéropodes de petite taille, notamment les Lepetodrilidae et les Peltospiridae (Hessler et Smithey 1983, Fustec et al 1987).

La seconde province biogéographique décrite fut la dorsale du Pacifique nord. Cette province est composée principalement par une faune similaire à celle de l'EPR mais composée de genres et/ou de familles proches, tels que le ver géant *Ridgeia piscesae* ou les polychètes Alvinellidae *Paralvinella sulfincola* et *P. palmiformis* (Tunnicliffe et al 1985, 1986, Fig. 1.13). Il s'agit de la seule province biogéographique décrite qui ne compte pas de moules (Mytilidae) hydrothermales du genre *Bathymodiolus* (Tyler et al 2003, les moules décrites sur cette province sont des genres *Benthomodiolus* à Gorda et *Idas* et *Adipicola* sur Juan de Fuca). La subduction de la plaque Pacifique sous la plaque Farallon il y a environ 35 millions d'années, aurait divisé la dorsale océanique de part et d'autre de l'Amérique du Nord, engendrant un événement de vicariance majeur ayant conduit à la différenciation de ces deux provinces biogéographiques par la mise en place de faunes jumelles (Tunnicliffe 1988).

La dorsale médio-Atlantique constitue la troisième province biogéographique mise en évidence, probablement séparée des provinces pacifiques lors de la fermeture de l'isthme de Panama (~10 Ma, Van Dover et al 2002). En plus des moulières de *Bathymodiolus azoricus* et *B. puteoserpentis*, la composition faunistique de cette province

contraste fortement avec les deux précédentes. Ainsi, les vestimentifères tubicoles sont absents et de véritables essaims de crevettes Alvinocarididae de l'espèce *Rimicaris exoculata* recouvrent parfois certaines cheminées hydrothermales (Desbruyères et al 2000, Fig. 1.13). On retrouve cependant une dominance numérique de certains petits gastéropodes (*Lepetodrilus atlanticus*, *Protolyra valvatoides*) à certains endroits de la dorsale.

La faune hydrothermale des bassins arrière arc du Pacifique occidental est également dominée par les bivalves du genre *Bathymodiolus (i.e. B. brevior, B. septemdierum, B. elongatus)* mais également par des gastéropodes symbiotiques de grande taille appartenant à la famille des Provannidae tels que *Alviniconcha hessleri* et *Ifremeria nautilei* (Desbruyères et al 1994, Fig. 1.13).

Enfin, découverts récemment, les sites hydrothermaux de l'océan Indien semblerait constituer une cinquième province biogéographique, bien que de nombreuses espèces restent encore à décrire et que les faunes paraissent intermédiaires entre la zone Atlantique et celle du Pacifique occidental (Van Dover et al 2001a). Cette province est dominée par des crevettes Alvinocarididae (proche de celles trouvées en Atlantique), des anémones et des gastéropodes de grande taille dont certains sont proches de ceux trouvés dans le pacifique occidental (Hashimoto et al 2001, Van Dover et al 2001a) alors que d'autres semblent endémique de l'océan Indien (gastéropode possédant un pied avec des écailles, Warén et al 2003, Fig. 1.13). La fermeture de la mer de Tethys (90 Ma) a pu gêner la connectivité entre cette province et la dorsale médio-Atlantique (Hessler & Lonsdale 1991). Il est cependant fort probable que des connections récentes aient pu être établies au travers de l'Atlantique sud et le long de la dorsale Antarctique.

Outre ces cinq provinces biogéographiques généralement admises, des auteurs proposent de diviser certaines de ces provinces en sous provinces biogéographiques différentes sur la base d'analyses multivariées de la composition faunistique en présence/absence (Fig. 1.13). Van Dover et al (2002) proposent ainsi de subdiviser la dorsale médio-atlantique en deux provinces distinctes (plateau des Açores contre region équatoriale de la ride médio-atlantique) en supposant que la formation du plateau des Açores (~20 Ma) ait pu isoler les faunes par vicariance. Une analyse plus récente des compositions faunistiques (Bachraty et al 2009) suggère d'intégrer les sources hydrothermales de l'océan Indien aux bassins arrière arc les plus sud de la province du Pacifique occidental et de séparer la province du Pacifique oriental en deux provinces

(nord et sud). La majorité des espèces sont néanmoins capables de franchir cette barrière biogéographique équatoriale. Pour ces espèces, il est donc intéressant d'étudier l'impact de cette barrière à un niveau phylogéographique. L'histoire de la formation de l'EPR ainsi que ses caractéristiques géotectoniques peuvent également apporter des informations précieuses pouvant permettre d'expliquer l'évolution des espèces dans un cadre géodynamique.

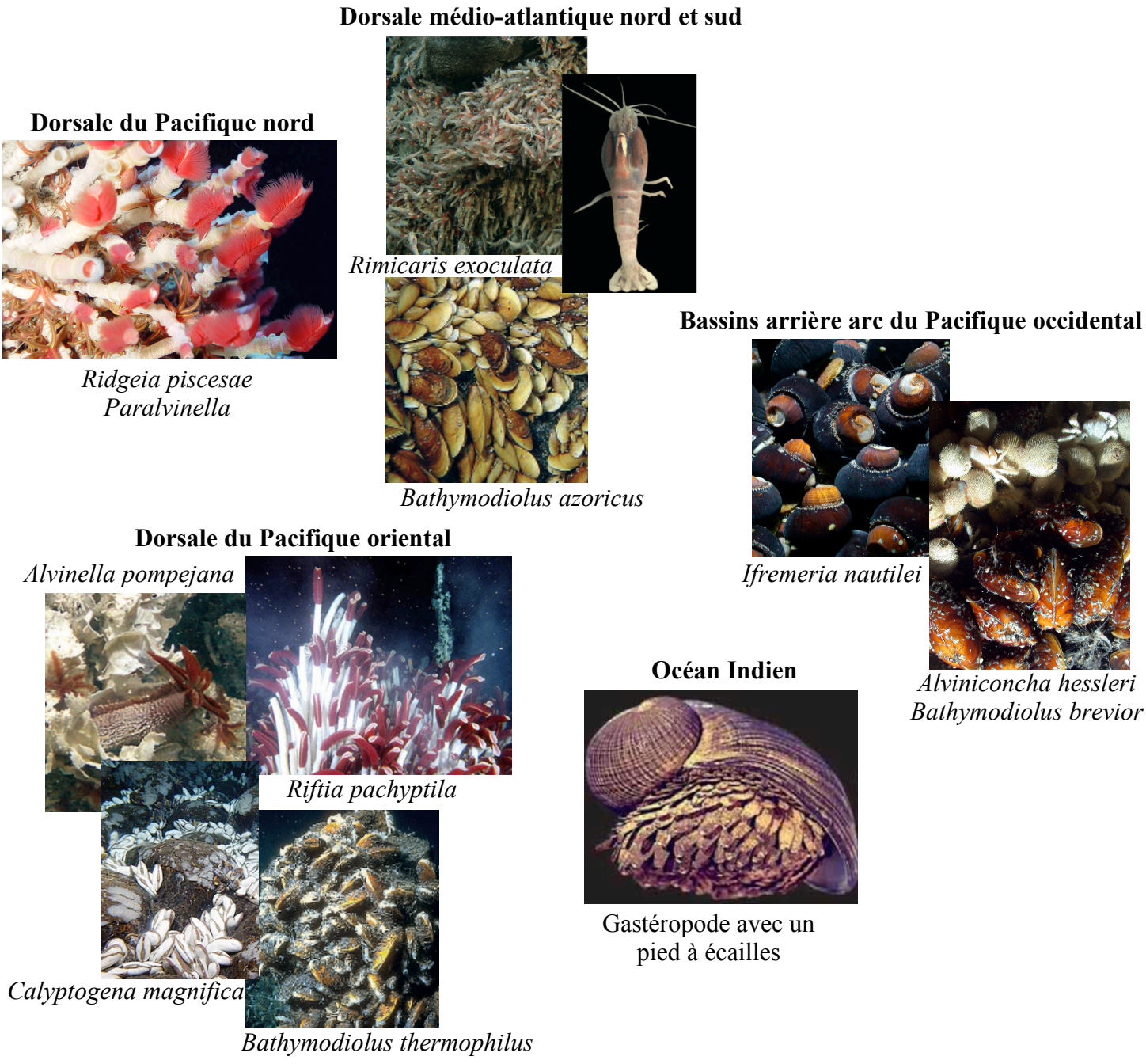


Fig. 1.13 Espèces emblématiques des cinq provinces biogéographiques initialement décrites

6. *Dorsale du Pacifique oriental (EPR)*

6.1. Vitesses d'accrétion et géomorphologie du rift

La morphologie des dorsales varie en fonction (1) de la vitesse d'accrétion et (2) du taux d'approvisionnement en magma (MacDonald et al 1991). En effet, une dorsale à faible vitesse d'accrétion (10 à 40 mm/an) présente généralement une vallée au niveau du rift axial de 1 à 3 km de profondeur (graben axial). C'est le cas de la dorsale médio Atlantique. Une dorsale à vitesse d'accrétion élevée (90 à 170 mm/an) est caractérisée par une élévation du plancher océanique de plusieurs centaines de mètres au niveau de son axe mais pas de vallée axiale (sillon de 50 mètres de large caractérisant une dorsale en dôme). Enfin, une dorsale à vitesse d'accrétion intermédiaire (40 à 90 mm/an) peut présenter un axe élevé avec une vallée généralement peu profonde suivant la quantité de magma fournie. Une forte production de magma peut néanmoins engendrer un axe élevé en dépit d'une vitesse d'accrétion faible.

L'axe des dorsales, séparant deux plaques tectoniques, n'est pas continu. Les ruptures d'axe de la dorsale ont lieu dans les régions présentant une quantité de magma moindre (MacDonald et al 1991). Dans le cas d'accrétion rapide, le magma se concentre en vastes réservoirs formant ainsi, sous la dorsale, un véritable collier de perles. C'est donc entre chaque perle (chaque réservoir, i.e. dans la zone de discontinuité des processus de convection thermique) qu'ont lieu les ruptures de l'axe de la dorsale. Au contraire, lorsque l'accrétion est lente, le magma ne se concentre pas en vaste réservoir mais forme de multiples « gouttelettes » sous la dorsale. Ainsi, les points de rupture de l'axe de la dorsale seront plus nombreux pour une dorsale à vitesse d'accrétion lente, augmentant ainsi par la même, les barrières potentielles à la dispersion. Une dorsale à vitesse d'accrétion élevée présentera, quant à elle, une activité magmatique plus importante, pouvant augmenter la probabilité d'extinction/recolonisation des populations hydrothermales associées.

L'EPR est une dorsale à vitesse d'accrétion variable selon le segment considéré (de 80-125 mm/an et de 125-152 mm/an respectivement le long des frontières Pacifique-Cocos et Pacifique-Nazca : Cormier 1997). Elle présente ainsi un degré de segmentation moindre que la dorsale médio-atlantique (MAR, Magde & Spark 1997). Cependant, on observe une forte diversité structurale entre les différents segments de la dorsale allant des dorsales en

dôme caractérisée par un volcanisme de type effusif (7°25S ou 17°25S avec de nombreux épanchements de lave en coussin) à des dorsales entaillées par des graben profonds et fortement tectonisés (21°25S)(cf Renard et al 1985). On peut ainsi penser que les populations de l'EPR sont moins structurées que celles de MAR. Cependant, la plus grande diversité biologique spécifique de l'EPR (Tunnicliffe 1991) suggère que la faune de la dorsale EPR pourrait être plus ancienne que celle de MAR, laissant plus de temps à la faune pour se différencier. La vitesse d'accrétion de l'EPR sud (entre 14°S et 22°S) est la plus élevée au monde. Les populations de l'EPR sud pourraient donc, potentiellement, avoir un taux de renouvellement plus grand et des goulots d'étranglement (bottleneck) plus fréquents.

L'histoire de la formation de l'EPR ainsi que ses caractéristiques géotectoniques sont donc autant de paramètres à prendre en considération pour expliquer les patrons phylogéographiques des espèces de cette dorsale.

6.2. Définition et principe d'une faille transformante et d'une microplaque

En modifiant l'axe de la vallée axiale de la dorsale, les failles transformantes peuvent engendrer des barrières à la dispersion. Ces failles sont la conséquence des forces de cisaillement portées en différents points de la dorsale selon les vitesses d'accrétion et les mouvements des plaques. Ces failles telles que celles de la zone de fracture Rivera peuvent décaler la dorsale sur plusieurs centaines de kilomètres et constituer des ruptures dans l'écoulement des courants de fonds à l'axe des dorsales. En général, ces failles datent de quelques millions d'années (1 à 4 Ma).

La formation d'une microplaque peut se résumer en trois étapes (Fig. 1.14). La première consiste en une fracturation de la dorsale en deux morceaux (rupture de l'axe par des failles transformantes). La seconde étape est la propagation des deux morceaux de dorsale en sens opposé, constituant ainsi un système de deux morceaux de dorsale parallèles. Dans la troisième étape, les deux morceaux de dorsale se rejoignent pour former une microplaque.

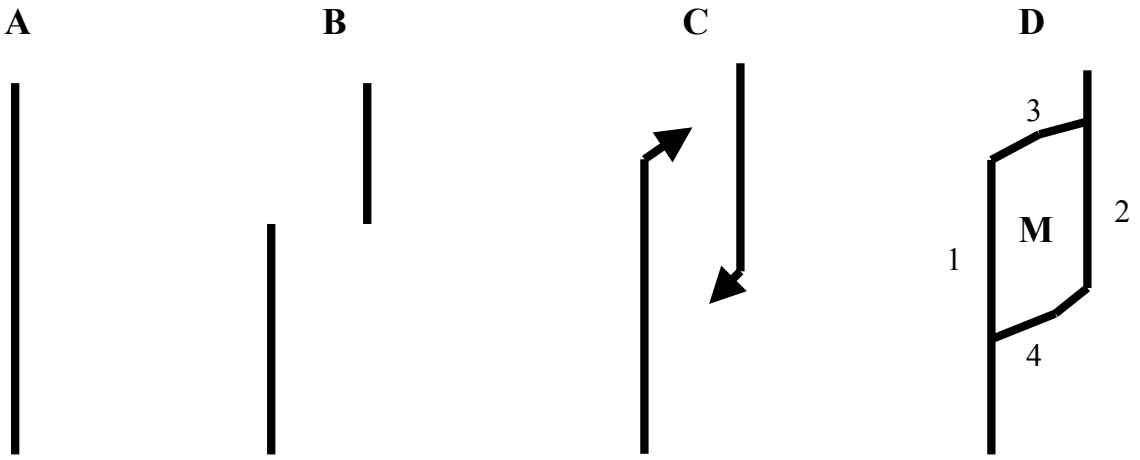


Fig. 1.14 Schématisation de la formation d'une microplaque.
A, dorsale sans fracturation ; B, fracturation de la dorsale ; C, propagation en sens inverse des morceaux de dorsale ; D, isolement de la microplaque (M).
1, Ouest de la microplaque ; 2, Est de la microplaque; 3, jonction nord ; 4, jonction sud.

6.3. Histoire de la formation de la dorsale EPR

La dorsale du Pacifique oriental telle qu'on la considère actuellement (EPR) est située à l'ouest de l'Amérique latine entre les plaques Pacifique et Cocos au Nord et entre les plaques Pacifique et Nazca au Sud. Elle s'étend ainsi entre les latitudes 30°N et 40°S. La dislocation de la plaque Farallon (par subduction sous la plaque Pacifique) a produit un plancher océanique structurellement chaotique entraînant la naissance de la microplaque Mendoza, il y a 24 Ma (Eakins & Lonsdale 2003, Fig. 1.15).

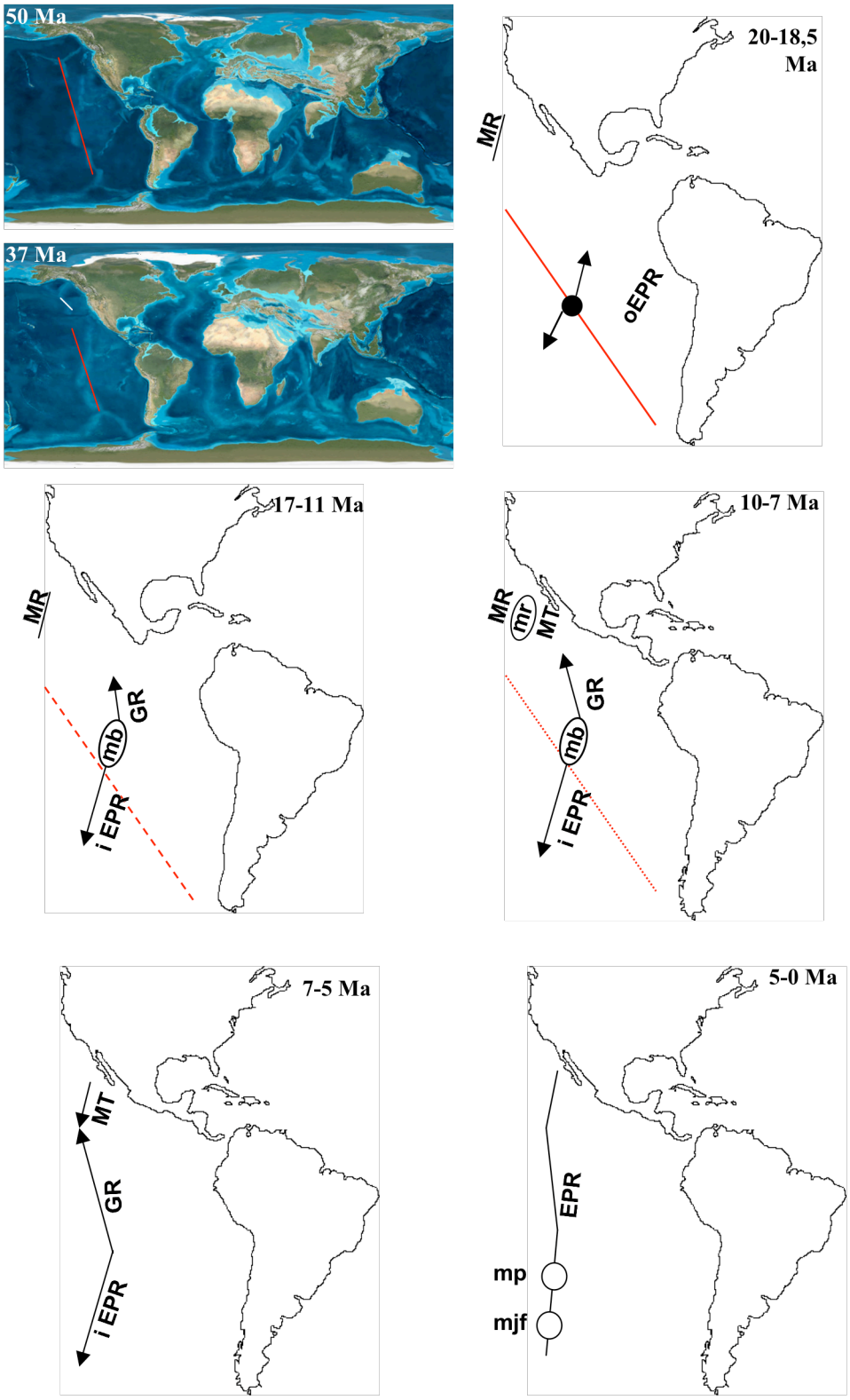


Fig. 1.15 Schématisation de l'histoire évolutive de l'EPR.
En blanc : dorsale nord pacifique
En rouge : dorsale ancestrale qui s'inactive (pointillés) ayant donné lieu à l'EPR actuelle, à partir du point de fracture (cercle noir). Les flèches indiquent la propagation des dorsales.
MR, Mathematician Ridge ; MT, Moctezuma Trough, GR, Galapagos Rise
Cercle vide : microplaque. mr, microplaque mathématicienne ; mb, microplaque Bauer ; mp, microplaque de Pâques ; mjf, microplaque Juan Fernandez

À 20-18,5 Ma, à 13°S, la zone de fracture Marquesas/Mendana donne « naissance » à une autre zone de fracture appelée l'oEPR (EPR d'origine, Fig. 1.15). L'ancienne dorsale s'inactive progressivement et se trouve séparée en plusieurs morceaux fossiles: Selkirk Rise, Roggeveen Rise, Mendoza Rise et Galápagos Rise à l'Est de l'oEPR et Gallego Rise à l'Ouest (Mammerick et al 1980). L'oEPR connaît ensuite différentes modifications avec la formation et/ou capture de diverses microplaques.

6.3.1. La microplaque Bauer (Eakins & Lonsdale 2003) : 10°S-18°S

La microplaque de Bauer a commencé sa formation il y a 17 Ma à partir d'un dédoublement du centre d'accrétion de l'oEPR. Les deux morceaux de dorsale générés par faille transformante sont alors nommés iEPR pour la fraction sud (1, Fig. 1.14 et la Galápagos Rise (GR, ne correspond pas à la dorsale Cocos-Nazca située au large des îles Galápagos) pour la fraction nord (2, Fig. 1.14). De 15 à 11 Ma, les dorsales iEPR et GR se propagent l'une vers l'autre. Les rifts formés à l'extrémité des axes de propagation engendrent un relief escarpé à la marge de ces dorsales. Cette propagation des dorsales s'effectue de manière oblique par rapport à leur pôle de rotation respectif, engendrant ainsi la formation de nouvelles fractures (offsets) perpendiculaires aux axes d'accrétion. À 11 Ma, iEPR et Galapagos Rise se remettent en contact par l'intermédiaire des failles transformantes Garrett (3, Fig. 1.14) et Dana et Woce (4, Fig. 1.14), isolant ainsi la microplaque Bauer.

6.3.2. La paléoplaque Mathematician (Mammerick et al 1980, Mammerick & Klitgord 1982) : 12°N-17°N

À 10 Ma, la Moctezuma Trough (dépression) se forme à l'Est de la Mathematician Ridge entre 12°N et 17°N (Mammerick & Klitgord 1982, Fig. 1.15), parallèlement à celle-ci. Moctozuma Trough et Mathematician Ridge se rejoignent (formation des zones de fracture Riviera et O'Gorman au Nord et au Sud, respectivement), formant la paléoplaque Mathematician. Elles coexistent pendant 3-5 Ma puis la Mathematician Ridge s'inactive (7-5 Ma), engendrant l'annexion de la paléoplaque Mathematician à la plaque Pacifique (Mammerick & Klitgord 1982). La Moctezuma Trough est ainsi rattachée à l'iEPR pour donner l'EPR actuelle (Mammerick et al 1980).

6.3.3. Capture de la microplaque Bauer et formation des microplaques de Pâques (Naar & Hey 1991, Rusby & Searle 1995, ~25°S) et Juan Fernandez (Larson et al 1992, 30°S-36°S)

À ~5,8 Ma, la microplaque Bauer (10°S-18°S) est capturée par subduction, sous la plaque Nazca (Eakins & Lonsdale 2003) et n'existe donc plus actuellement.

À peu près à la même période que la capture de la microplaque Bauer, ~5,89 Ma, l'EPR se fracture à la latitude d'environ 25°S isolant 2 segments de dorsales (Naar & Hey 1991, Rusby & Searle 1995). À 5,25 Ma, le segment de dorsale situé au Sud Est de cette fracture commence sa propagation vers le Nord. À 2,5 Ma, sa vitesse de propagation diminue et s'arc-boute sur le segment de dorsale situé au Nord Ouest, lequel entame une propagation en forme d'arc vers le sud, imprimant un début de rotation de la microplaque de Pâques dans le sens inverse des aiguilles d'une montre (contrairement à la microplaque de Bauer et à la schématisation sur la Fig. 1.14). La dorsale occidentale isole la microplaque au Sud et la dorsale orientale cesse sa propagation. Actuellement, la microplaque de Pâques n'est toujours pas isolée au Nord (Fig. 1.16).

Encore un peu plus au sud, la microplaque Juan Fernandez (Fig. 1.16, Larson et al 1992) est issue de la propagation vers le Sud de ce qui reste du segment sud de la dorsale EPR ayant joué un rôle dans la formation de la microplaque de l'île de Pâques, et sa rencontre, il y a 3,4 Ma, avec la dorsale Pacifique-Antarctique s'étendant vers le nord. À 2,5 Ma, la microplaque est isolée au Sud puis s'isole complètement au Nord (0,7 Ma) via la Deep Endeavour (ou fosse Endeavour).

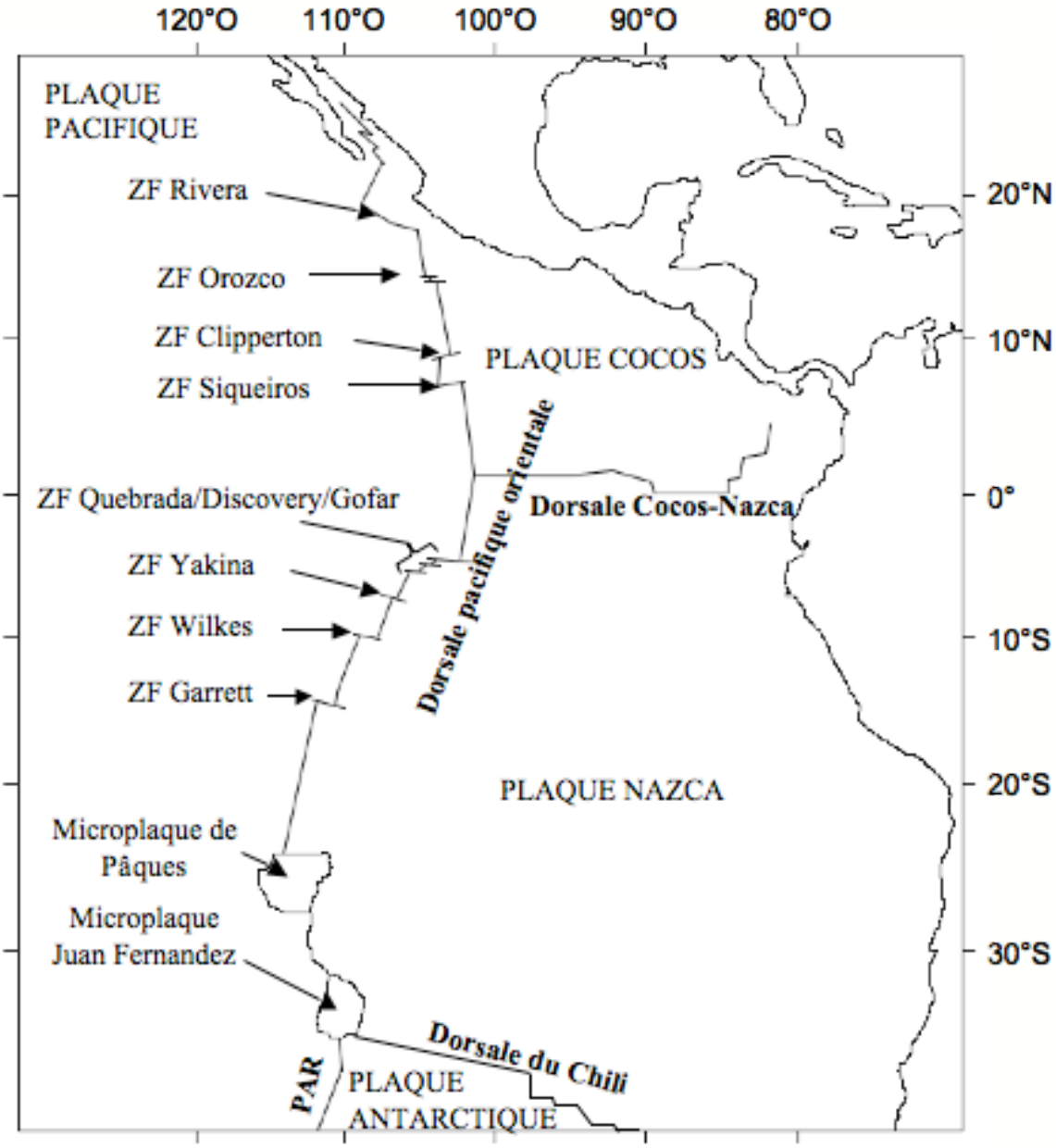


Fig. 1.16 Géomorphologie actuelle de la dorsale pacifique orientale.
PAR, Dorsale Pacifique-Antarctique ; ZF, zone de fracture

6.4. Potentielles barrières à la dispersion

6.4.1. Failles transformantes

Les failles transformantes retrouvées à l'équateur sont datées de 1-2 Ma entre 9°N et 17°S (Kureth & Rea 1981, Naar & Hey 1989, Francheteau et al 1990, Fig. 1.16). L'impact de ces failles sur les espèces en tant que barrière à la dispersion peut différer en fonction

des capacités de dispersion de ces espèces et du comportement de leurs larves potentielles dans la colonne d'eau. Ainsi, Johnson et al (2006) ont montré que la faille transformante Blanco (dorsale pacifique nord) datée de 3,5 Ma serait responsable de la spéciation allopatrique de *Lepetodrilus fucensis* et *L. gordensis* alors que son impact est moindre pour *Ridgeia piscesae* (différenciation génétique mais absence de monophylie réciproque, Young et al 2008).

6.4.2. Courants océaniques profonds

Certains courants océaniques profonds transverses à l'axe de la dorsale peuvent également réduire la dispersion et jouer un rôle de barrière. Ainsi, divers gyres ont été mis en évidence (Reid 1997) le long de l'EPR. Au niveau des zones de contact entre 2 gyres, la dispersion latitudinale des larves pourrait ainsi être limitée. Ces zones de contact sont situées au niveau de l'équateur, aux environs de 13°N, 15°S et de 32°S (Fig. 1.17).

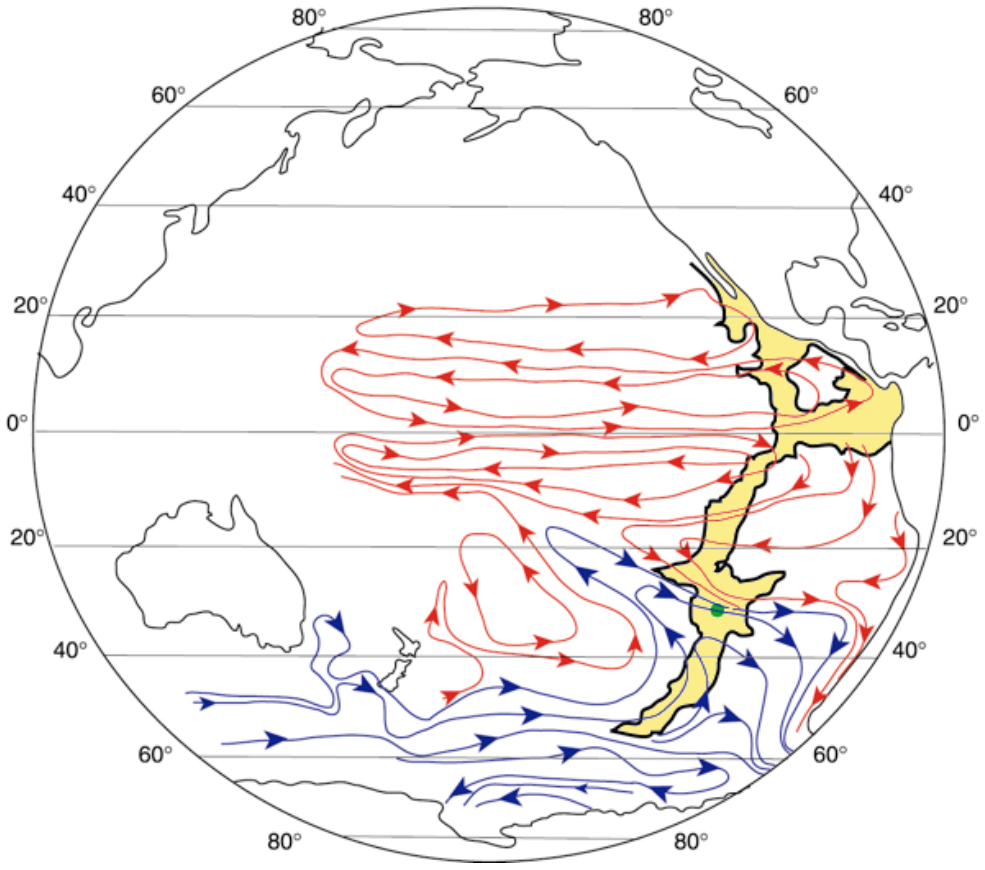


Fig. 1.17 Courants océaniques profonds (en rouge et bleu) proche de l'EPR et de la dorsale Cocos-Nazca (en jaune). Modifiée de Reid 1997.

7. *Structure génétique des espèces de la dorsale du Pacifique oriental*

Les caractéristiques éphémères et fragmentées du milieu hydrothermal ont rapidement suscité de nombreuses interrogations quant aux capacités de dispersion des espèces et au degré de connexion entre les populations. Par exemple, des études menées sur le bivalve *Calyptogena magnifica* ont montré que, bien que présentant de la structure génétique entre populations (Hurtado et al 2003), les larves de cette espèce seraient capables de se disperser sur l'ensemble de leur aire de distribution (i.e. plusieurs milliers de km, Karl et al 1996). De plus, Kojima et al (2004) ont détecté au moins huit événements de migration trans-pacifique, deux évènements entre le Pacifique Nord-Ouest et Sud-Ouest et deux autres entre les océans Pacifique et Atlantique, dans l'histoire évolutive des Vesicomyidae.

L'EPR suscite également un intérêt d'autant plus grand qu'elle présente des vitesses d'accrétion importantes et donc un fort taux d'extinction/recolonisation. Des approches directes visant à comprendre le comportement des larves ont ainsi été réalisées (e.g. recrutement des larves sur des blocs de basalte déployés sur le site 9°50'N, Mullineaux et al 1998 ; processus de développement des embryons d'*A. pompejana* en fonction des conditions environnementales, Pradillon et al 2005) mais restent relativement difficiles à mettre en œuvre du fait de la faible accessibilité du milieu. Des approches indirectes, encore peu nombreuses, basées sur l'analyse de la structure génétique des espèces hydrothermales ont ainsi été initiées afin de mieux comprendre l'impact de la dispersion sur la distribution spatiale des espèces ainsi que l'impact de l'environnement (en termes de fragmentation et d'instabilité) sur le degré de connectivité (migration efficace) et les fluctuations en taille des populations

7.1. Quels modèles de dispersion pour les espèces de l'EPR ?

Sur la base des données génétiques, les espèces hydrothermales étudiées sur l'EPR semblent capables de disperser sur des centaines de kilomètres et répondraient à un modèle classique de métapopulation de Levin (modèle en îles + extinction/recolonisation). Ces données apparaissent en contradiction avec l'hypothèse première d'un isolement linéaire par la distance proposée par Lutz (1988) sur la base d'une distribution fragmentée et uni-

directionnelle de l'habitat hydrothermal et du mode de développement lécithotrophe des larves prépondérant en milieu hydrothermal (cf Tyler & Young 1999). Bien que ne présentant pas les mêmes types de larves (planctotrophes contre lécithotrophes), des études utilisant des marqueurs allozymiques et/ou des séquences du gène mitochondrial Cytochrome Oxydase I (COI) ont ainsi montré que les structures génétiques de *Bathymodiolus thermophilus* (Grassle 1985, Craddock et al 1995) et *Alvinella pompejana* (Jollivet et al 1995) se conformeraient à un modèle en îles (migration importante entre les sites). En modélisant les capacités de dispersion de ces espèces, Jollivet et al (non publiées) proposent d'ajuster la dispersion de *B. thermophilus* à un modèle de métapopulation de type « migrant ». Pareillement, les études allozymiques effectuées sur les gastéropodes *Lepetodrilus pustulosus* et *Eulepetopsis vitrea* (Craddock et al 1997) n'ont pas montré d'isolement par la distance et, suggèrent également une dispersion suivant un modèle en îles. À l'inverse, France et al (1992) furent les premiers à proposer la présence d'un isolement par la distance chez une espèce hydrothermale vagile, l'amphipode *Ventiella sulfuris*, capable de se déplacer en essaims d'un site à un autre. Ces auteurs relient néanmoins cette structure de populations au fait que l'animal a un développement direct, les oeufs fécondés se développant dans un organe appelé le marsupium. Néanmoins, la mise en évidence de ce modèle d'isolement apparaît biaisée par un échantillonnage de deux groupes de sous-populations différenciées fortement distantes respectivement sur l'EPR nord et la dorsale Cocos-Nazca. L'hypothèse d'une telle barrière entre les 2 dorsales avait d'ailleurs été émise par Grassle (1985) et Jollivet et al (1995) pour expliquer l'augmentation de la différenciation génétique observée chez les espèces *Bathymodiolus thermophilus* et *Paralvinella grasslei*. De nouvelles analyses effectuées uniquement avec les sites de l'EPR (Vrijenhoek 1997) ont abouti, sous l'hypothèse du modèle en îles, à de fortes valeurs de migration entre les sites de l'EPR (11,6 individus par génération). Si un isolement par la distance a également été détecté à cette échelle spatiale, il semblerait qu'il s'agisse également d'un second biais résultant de la séparation des populations de part et d'autre de la zone de fracture Riviera (Vrijenhoek 1997). Dans ce cas, la différenciation observée résulte plus de la mise en place de barrières physiques à la dispersion qu'à un effet de la distance géographique. Des échantillonnages et analyses complémentaires s'avèrent donc nécessaires pour infirmer ou confirmer l'existence d'un isolement par la distance chez *V. sulfuris*. Au contraire, la structure génétique du Siboglinidae *Riftia pachyptila* semble, quant à elle, s'ajuster de manière moins critiquable à un modèle avec isolement par la distance (Black et al 1994), et ce malgré une autonomie

dispersive de la larve assez grande (3 semaines à 1 mois : Marsh et al 2001). L'étude, réalisée sur un nombre de champs hydrothermaux plus importants, prend en effet soin de supprimer les sites ayant potentiellement divergé de part et d'autre d'une barrière à la dispersion (Vrijenhoek 1997). À l'heure actuelle, *R. pachyptila* est la seule espèce étudiée de l'EPR pour laquelle un isolement par la distance a été mis en évidence. Cependant, Shank & Halanych (2007) ont toutefois récemment proposé l'existence de migrants sur de longues distances entre 9°50'N et Guyamas chez cette espèce, allant à l'encontre des résultats de Black et al (1994).

7.2. Existence de barrières génétiques chez les espèces de l'EPR

Les premières études génétiques sur l'EPR ne se sont intéressées qu'à une échelle géographique restreinte (EPR nord) sans pouvoir mettre en évidence la présence de barrière à la dispersion à cette échelle (*A. pompejana*, Jollivet et al 1995 ; *B. thermophilus*, Craddock et al 1995) et, ce malgré la présence de la zone de fracture Rivera (la faille la plus longue connue sur l'EPR). Néanmoins, la plupart de ces études démontrent une différenciation des populations entre l'EPR nord et la dorsale Cocos-Nazca (*B. thermophilus* : Grassle 1985 ; *V. sulfuris* : France et al 1992 ; *Paralvinella grasslei* : Jollivet et al 1995 ; *Calyptogena magnifica* : Karl et al 1996).

Les études plus récentes basées sur l'analyse de séquences ADN ont élargi le champ d'investigation en échantillonnant des sites provenant de l'EPR sud, et révélé l'existence de diverses barrières génétiques par des méthodes plus sensibles utilisant la divergence inter-individuelle :

(1) Des clades monophylétiques ont été identifiés chez *A. pompejana* de part et d'autre de l'équateur, suggérant un impact important d'une barrière située à ces latitudes, qu'elle soit de type hydrologique gyre ou de nature tectonique (Hess Deep, failles transformantes Quebrada/Discovery/Gofar). Cette barrière pourrait également, dans une moindre mesure, être responsable de la différenciation des populations de *Tevnia jerichonana*.

(2) La microplaque de Pâques aurait joué un rôle de barrière à la dispersion chez *R. pachyptila* (COI : Hurtado et al 2004) mais également chez *B. thermophilus* (COI et allozymes : Won et al 2003). Won et al (2003) suggèrent ainsi que la

divergence accumulée de part et d'autre de cette microplaque a engendré deux espèces cryptiques chez *B. thermophilus* par spéciation allopatrique. Les polychètes *Branchipolynoe symmytilida* et *A. pompejana* ne semblent cependant pas affectés par cette barrière (Hurtado et al 2004, COI).

(3) Won et al (2003) ont montré l'existence d'une différenciation génétique sur *B. thermophilus* située entre 11°S et 17°S. Ces auteurs ont suggéré que cette différenciation génétique pourrait être due à un gyre profond présent à 15°S.

(4) Une structure génétique a également été observée entre plusieurs populations de l'EPR chez *Riftia pachyptila* via des AFLP (9°N/17°S, Shank & Halanych 2007). De même, deux espèces cryptiques avec potentialités d'hybridation ont été mises en évidence chez *Lepetodrilus elevatus*, la première échantillonnée à 13°N et 9°50'N et la seconde à 9°50'N et 17°S (COI et allozymes, Matabos et al 2008b). Néanmoins, des analyses complémentaires sont nécessaires pour définir plus précisément la localisation des barrières potentielles aux flux de gènes chez ces espèces.

Ces études constituent ainsi les prémices d'une étude de phylogéographie comparée sur la dorsale du Pacifique oriental. Peu d'analyses (cf. Hurtado et al 2004) ont pour le moment été effectuées afin de tester ces hypothèses d'événements de vicariance communs entre espèces, ou d'évaluer le sens de colonisation préférentiel de la dorsale. Si certaines études ont tenté de détecter de l'hybridation entre les lignées divergentes observées sur le COI en couplant ce gène à des données allozymiques (Karl et al 1996, Won et al 2003, Matabos et al 2008b), le pouvoir résolutif des allozymes reste toutefois faible comparé à celui des séquences de gènes nucléaires. L'utilisation de gènes nucléaires pourrait ainsi permettre de mieux évaluer les contours d'une espèce donnée et notamment de savoir si l'existence de lignées mitochondriales distinctes coincide réellement avec l'existence d'espèces cryptiques (notion de monophilies réciproques à tous les gènes) ou, *a contrario* de comprendre ces phénomènes d'hybridation potentiels par l'utilisation de méthodes basées sur la théorie de la coalescence. De plus, la comparaison de plusieurs gènes et/ou plusieurs espèces pourrait permettre de mieux appréhender l'histoire démographique des populations de l'EPR, jusqu'alors largement négligée et notamment son rôle dans la structuration génétique des populations actuelles. Enfin, l'approche multi-locus fournit un moyen particulièrement puissant de détecter des effets sélectifs potentiels sur certains gènes.

8. *Objectifs de la thèse*

L'objectif de ce travail est de réaliser une étude phylogéographique de la dorsale du Pacifique oriental en combinant des approches classiques d'analyses de fréquences alléliques dans les populations et de leurs distributions géographiques avec des méthodes d'analyse plus récentes basées sur la théorie de la coalescence. Ces approches combinées permettent de détecter d'éventuelles barrières à la dispersion, de les localiser spatialement, et de mesurer leur degré de perméabilité. Les études de coalescence permettent de plus d'en évaluer l'âge, de chercher à savoir s'il existe un sens de la colonisation le long de la dorsale et de mieux comprendre l'histoire démographique des espèces en présence. Pour cela, nous avons d'abord opté pour une approche comparative entre sept espèces en regardant plus spécifiquement le gène mitochondrial COI. Nous avons ensuite traité plusieurs cas d'espèces à l'aide de plusieurs types de marqueurs (COI, gènes nucléaires, allozymes) en choisissant des espèces appartenant à 2 embranchements : annélide et mollusque. Les 3 espèces étudiées (le ver de Pompéi *Alvinella pompejana*, le bivalve *Bathymodiolus thermophilus* et le gastéropode patelliforme *Lepetodrilus elevatus*) ont été choisies pour 2 raisons : (1) des stratégies reproductives et des modes de développement larvaire différents et (2) il s'agit d'espèces pour lesquelles des banques de gènes sont disponibles. L'approche multi espèces avait notamment pour but de tester des hypothèses d'événements de vicariance associés à une éventuelle séparation des faunes et chercher à savoir si ces espèces pouvaient avoir eu une histoire démographique commune. L'approche multi gènes nous a permis, quant à elle, d'appréhender les phénomènes d'hybridation entre lignées divergentes, de mieux comprendre les patrons de migration et de détecter des gènes potentiellement sous sélection.

Chapitre 2 : Méthodes et biais méthodologiques

1. *Stratégie d'échantillonnage*

Les échantillons utilisés pour ce travail de thèse sont principalement issus de plusieurs campagnes océanographiques françaises le long de la dorsale EPR. Les premières campagnes (HOT96, HOPE99, PHARE2002) ont été menées au Nord de l'équateur sur les sites situés à 13°N et 9°50'N et avaient notamment pour objectif d'étudier les flux géniques entre les sites hydrothermaux présents au sein d'un ou deux champs hydrothermaux (un champ hydrothermal étant composé de plusieurs sites). La campagne océanographique BIOSPEEDO 2004 visait à élargir l'étude des flux de gènes à des champs hydrothermaux situés au Sud de l'EPR afin de mieux comprendre le rôle des failles transformantes, microplaques et courants océaniques profonds dans la structure génétique des populations. Une collaboration avec Timothy Mark Shank, chercheur au Woods Hole Oceanographic Institute a permis de compléter cet échantillonnage au Nord de la dorsale (campagne Extreme 2003) pour l'espèce *L. ovalis* qui n'avait pas été récoltée lors des campagnes françaises, l'espèce n'ayant pas été trouvée.

Il est nécessaire de prendre conscience de la difficulté d'échantillonner en milieu hydrothermal profond. En effet, les sites étant généralement situés à plus de 2000 m de profondeur, l'accès aux échantillons nécessite des moyens technologiques et humains importants. Les prélèvements sont ainsi effectués par des submersibles hautement résistants aux fortes pressions (sphère titane). Ces submersibles peuvent être habités (pour cette étude, utilisation du sous-marin français Nautile lors des campagnes HOT96, HOPE99 et BIOSPEEDO 2004, et du sous-marin américain Alvin lors de la campagne Extreme 2003, Fig. 2.1.A, 2.1.B, 2.1.C) ou télécommandés depuis la surface à partir d'une laisse avec un câble électroporteur (pour l'alimentation en énergie) et des faisceaux de fibres optiques (pour récupérer les informations) (utilisation du Remotely Operated Vehicule, ROV Victor 6000 pour la campagne PHARE 2002 Fig. 2.1.D). Il convient également de rappeler la difficulté d'observation liée à l'obscurité (l'éclairage des

bathyscaphes permettant un champ de vision de quelques mètres) et le temps d'échantillonnage limité (une dizaine d'heures pour un submersible habité par plongée). Les animaux sont récoltés à l'aide des bras télé-manipulés de ces bathyscaphes (Fig. 2.1.E), placés dans des paniers isothermes (Fig. 2.1.F) et remontés en surface (décompression des organismes). À bord du navire océanographique (N/O L'Atalante pour les campagnes françaises et R/V Atlantis pour la campagne américaine, Fig 2.1.G, 2.1.H), ils sont alors triés par espèce et préservés dans l'éthanol absolu ou congelé à -80°C ou dans l'azote liquide pour les extractions d'ADN/ARN réalisées à terre.

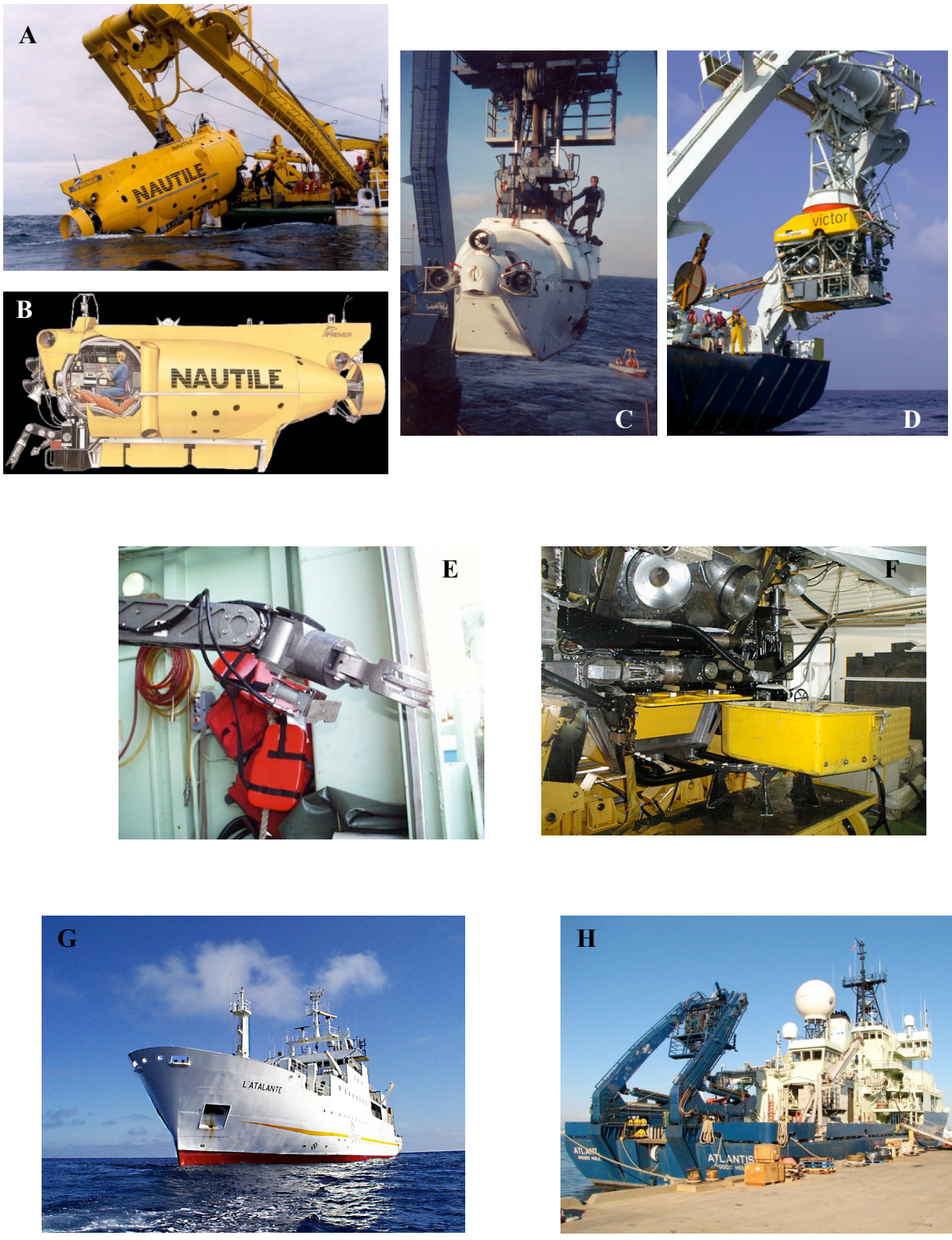


Fig. 2.1 Moyens à la mer et équipements utilisés pour effectuer les échantillonnages nécessaires à la thèse.
A et B, submersible habité Nautile ; C, submersible habité Alvin ; D, ROV Victor 6000
C, pince télé-manipulée (ici de l'Alvin) ; D, panier isotherme du Nautile
G, le navire français N/O L'Atalante ; H, le navire américain R/V Atlantis

Ces différentes campagnes océanographiques ont ainsi permis de récolter des populations d'espèces différentes sur des champs hydrothermaux distribués de 21°N à 21°33'S (Fig. 2.2), permettant ainsi d'effectuer les analyses de phylogéographie comparée réalisées lors de ma thèse.

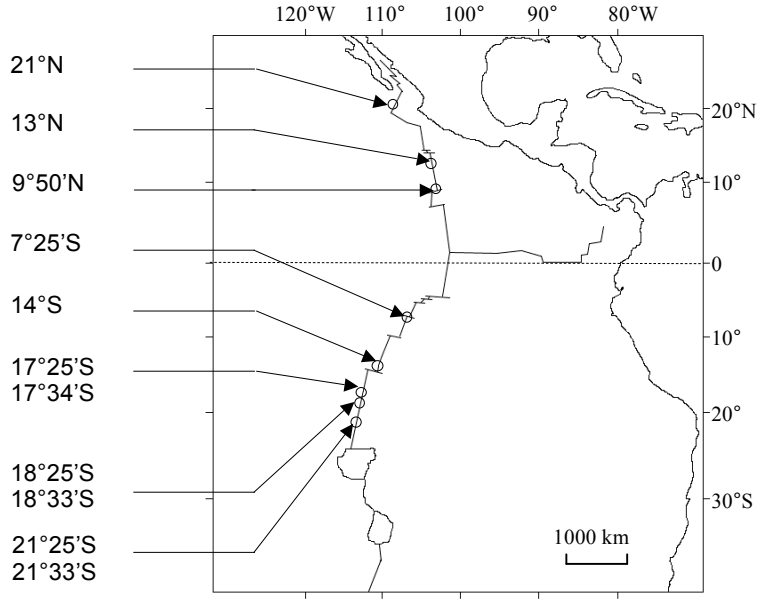


Fig. 2.2 Champs hydrothermaux échantillonnés le long de la dorsale EPR

2. *Banques d'ADN complémentaires*

Les gènes nucléaires utilisés dans le chapitre 3 sont issus de deux banques d'ADNc réalisées par Arnaud Tanguy sur plusieurs individus de *B. azoricus* récoltés sur le site Menez Gwen (dorsale Médio-Atlantique, 37°50N, Tanguy et al 2008) et de *L. elevatus* du site Oasis (EPR, 17°25'S), respectivement. Les 2 banques effectuées sur individus entiers ont été obtenues par utilisation du kit de synthèse d'ADNc SMART^{TM} et du kit de clonage SMART^{TM} (Clontech) après purification des messagers sur résine polyAT tract®mRNA. Les gènes nucléaires obtenus sur *A. pompejana* proviennent de banques ADNc tissulaires (branchie, tissu ventral et pygidium) réalisées par le Génoscope d'Evry dans le cadre du Consortium 'Alvinella' à partir de 2 types de méthodes de clonage : (1) 'Oligocapping' des ARNm par phosphatase alcaline, récupération des messagers sur résine OligodT après élimination de la coiffe 5' et insertion dans un vecteur pME18S-FL3 après RT-PCR et, (2)

la méthode 'CloneMiner' ou les ADNc double brins sont produits par ajout de séquences AttB à chaque bout et clonés dans un vecteur Gateway par recombinaison homologue. Ces banques représentent un total de 100200 séquences assemblées et annotées (Gagnière et al soumis).

L'extraction des ARN totaux a été effectuée à partir d'un protocole utilisant le Tri-Reagent. Cinquante à cent milligrammes de tissu congelé sont placés dans 1 mL de solution Tri-Reagent et broyés à l'aide d'un broyeur à billes. A l'homogénat sont ajoutés 0,2 mL de chloroforme pour provoquer une émulsion afin de séparer la fraction ARN des fractions ADN/protéines/lipides. Après centrifugation (12 000 g, 15 min), les ARN totaux du surnageant sont précipités par ajout de 0,5 mL d'isopropanol. Les ARN totaux sont ensuite culottés par centrifugation (12 000 g, 10 min), lavés dans de l'alcool 70°, séchés et repris dans 50 microlitres d'eau-DEPC (incubation 15 min à 55°C).

Dans le cadre des banques *Bathymodiolus* et *Lepetodrilus*, les ARN messagers (ARNm) sont purifiés sur une résine oligodT à l'aide du kit Promega PolyATtract mRNA Isolation Systems. Les ARN totaux sont hybridés à l'aide d'une sonde Oligo(dT) biotinylé (50 nmol/mL) en présence d'une solution SSC 0,5X. Les ARNm hybridés sont ensuite capturés sur des sphères magnétiques de streptavidine couplées à un anticorps de la biotine et isolés du « pool » des ARNs par aimantation. Les ARNm-sphères sont ensuite lavés plusieurs fois dans des solutions 0,5X SSC et 0,1X SSC et resuspendus dans 1 mL d'eau-DEPC. Les ADNc sont synthétisés par rétrotranscription (8 µl de tampon contenant 1,2 µl de MgCl2 25mM, 4 µl de tampon de transcriptase reverse, 3 µl de dNTP 2mM, 100 U de transcriptase reverse, une heure à 37°C) puis amplifiés, clonés et séquencés.

3. *Acquisition des séquences*

3.1. Séquences du gène mitochondrial Cytochrome Oxydase I

L'ADN mitochondrial étant généralement à hérédité uniquement maternelle (mais voir Skibinsky et al 1994 pour les exceptions), il est théoriquement descendant d'une seule copie (pas de recombinaison) au sein d'un individu. C'est pourquoi le séquençage du gène mtCOI a été effectué en direct, c'est-à-dire sans clonage préalable, à la plateforme Génomer de la Station Biologique de Roscoff. Le fragment du gène COI est amplifié par PCR à l'aide d'amorces universelles (Folmer et al 1994) ou redéfinies à partir de ces amorces, puis purifié sur membrane Millipore et séquencé sur ABI.

3.2. Séquences des gènes nucléaires : Marquage-Recapture (MR) de séquences

Les locus génomiques étudiés sont diploïdes. Les individus séquencés peuvent donc être homozygotes ou hétérozygotes au locus d'intérêt. Afin de ne pas brouiller le signal de la séquence (présence d'indels ou nombreuses substitutions entre les allèles), il est nécessaire, après amplification par PCR et purification du produit, d'isoler les différents allèles par clonage. Le clonage étant un protocole relativement long et coûteux, il n'est pas possible d'effectuer des clonages individuels pour chaque locus. Nous avons donc choisi une solution alternative : combiner les produits d'amplification des individus appartenant à une même espèce, voire une même population, après les avoir préalablement marqués individuellement, et cloner l'ensemble. Pour éviter de biaiser la diversité allélique par clonage, les allèles sont donc « taggés » selon leur appartenance à un individu (Fig. 2.3 technique de marquage / re-capture de séquences : Bierne et al 2007, article présenté en annexe 1). Le marquage de la séquence consiste à ajouter quelques nucléotides permettant une identification postérieure à l'extrémité 5' des amorces utilisées pour l'amplification des individus. Ces nucléotides forment ainsi un « tag » qui, bien que ne s'hybridant pas avec le fragment d'ADN recherché, est amplifié par la Taq polymérase lors de la PCR. Ainsi, les allèles d'un individu amplifié sont identifiés par une combinaison de « tags », lesquels diffèrent d'un individu à l'autre. Ces « tags » sont ensuite retrouvés sur les séquences clonées de part et d'autre de l'insert, ce qui permet, *a posteriori*, d'attribuer les séquences aux individus (re-capture).

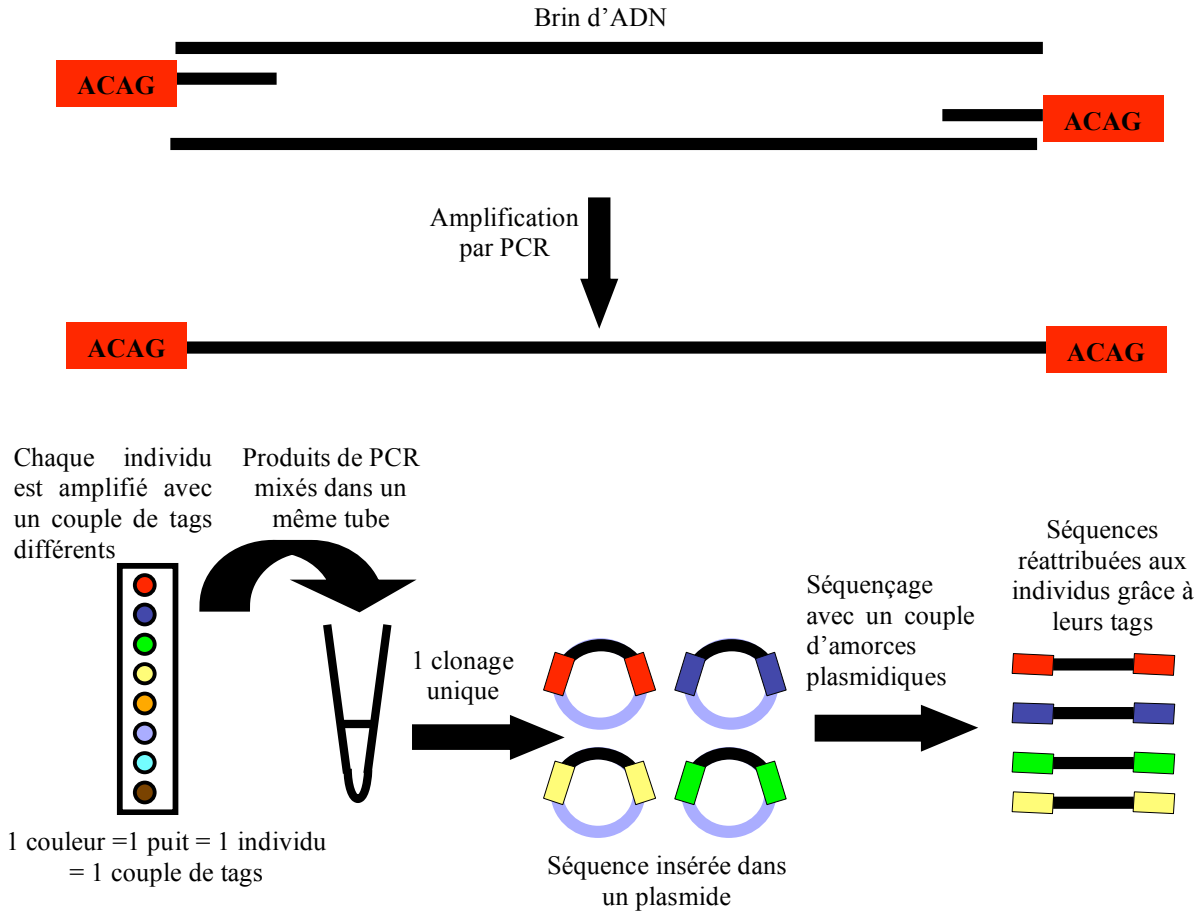


Fig 2.3 Schéma du fonctionnement de la méthode de marquage-recapture de séquences utilisée pour le séquençage des gènes nucléaires.

Des mutations artéfactuelles sur les « tags » ou de la recombinaison *in vitro* pouvant engendrer une mauvaise réassignation des séquences, des précautions particulières ont été prises afin de :

(1) Détecter les mutations artéfactuelles sur les tags. Le choix des tags a été réalisé afin que chaque tag (de 4 nucléotides) diffère d'au moins 2 substitutions permettant ainsi de repérer une éventuelle mutation artéfactuelle sur l'un des tags.

(2) Détecter les recombinaisons *in vitro*. Chaque individu a été amplifié avec un couple d'amorces présentant des « tag » différents (individu 1 : F1R1, individu 2 : F2R2, Fig. 2.3). En conséquence, une séquence recapturée présentant des combinaisons de tags non planifiées (F1R2 par exemple) est considérée comme recombinée *in vitro*.

Pour chaque espèce et chaque gène, nous avons choisi de séquencer 16 individus par champ hydrothermal en effectuant 2 clonages par marquage recapture des séquences issues de 8 individus chacun, permettant ainsi d'augmenter le nombre de séquences recapturées (en comparaison à 1 seul clonage de 16 individus). Trente deux clones par clonage ont ensuite été séquencés au génoscope d'Evry.

4. *Traitement des séquences*

4.1. Alignement et correction des séquences liées aux erreurs de lecture du séquenceur

Une fois le séquençage effectué par les plateformes de Genomer et du Génoscope, les régions plasmidiques (pour les gènes nucléaires) sont éliminées manuellement. Pour chaque gène, les séquences sens (forward) et antisens (reverse) sont alignées via ClustalW (Thompson et al 1994) sous Bioedit version 6.0.6 (Hall 1999) et l'alignement est vérifié visuellement. Les chromatogrammes de chaque séquence sont également alignés entre eux via la version de démonstration de CodonCode aligner 2.0.6 (http://www.codoncode.com/aligner/). L'alignement de ces chromatogrammes permet de détecter une zone de recouvrement commune à l'ensemble des séquences pour laquelle les

chromatogrammes concordent parfaitement entre chaque paire de séquences sens et antisens. La correspondance entre les bases inscrites par le séquenceur et les pics des chromatogrammes est alors vérifiée jusqu'à cette zone de recouvrement incluse (sens : de la première base jusqu'à la fin de la zone de recouvrement, antisens : début de la zone de recouvrement jusqu'à la fin de la séquence). Matthieu Bruneaux et moi-même avons développé un script sous C permettant de « rabouter » les séquences sens et antisens de manière automatique en fournissant un fichier de séquences sens, un fichier antisens (format fasta) et la position du site de coupure (identique pour toutes les séquences) défini par l'utilisateur dans la zone de recouvrement (cf annexe 2). Ce programme n'effectue pas de consensus de séquences, ne modifiant pas les séquences fournies. Plusieurs centaines de séquences d'un même gène peuvent ainsi être raboutées en quelques secondes. Ce programme a donc été utilisé indépendamment pour chaque gène et chaque espèce. Pour les séquences nucléaires, effectuées selon la méthode de marquage-recapture de séquences, nous avons ensuite réattribué chaque séquence à son individu amplifié par l'observation du couple de « tags » (Bierne et al 2007). Pour chaque gène, les bases correspondant aux amorces sont ensuite supprimées pour les analyses afin de ne pas biaiser les estimations de polymorphisme et de divergence (zone conservée du fait de l'amorce).

4.2. Recombinaison *in vitro* détectée via la méthode MR

En utilisant un couple de tags différents pour chaque individu, nous avons pu mettre en évidence 22 séquences présentant une combinaison anormale de tags, indiquant que ces séquences sont issues d'un événement de recombinaison *in vitro* entre individus différents d'un même clonage. Le pourcentage de séquences détectées recombinées *in vitro* entre individus est compris entre 0 et 1,82%, cette dernière valeur ayant été obtenue pour la Globine X d'*A. pompejana* (Tableau 2.1).

Tableau 2.1 Pourcentage de séquences détectées recombinées *in vitro* entre différents individus.
SAHH, S-Adénosyl Homocystéine Hydrolase ; PGM, Phosphoglucomutase ; VDG3, Veliger Digestive Gland 3.

Espèce	Gène	Taille moyenne du gène (pb)	Pourcentage de séquences détectées recombinées *in vitro* entre individus
A. pompejana			
	SAHH	579	0,18
	Globine X	668	1,82
	PGM	1004	0,58
B. thermophilus			
	SAHH	1005	0
	Lysozyme	1302	0,29
	Sulfotransférase	275	0
L. elevatus			
	Lysozyme	782	0,33
	Textilinine	1227	1,64
	VDG3	682	0,18

Aucune corrélation n'a pu être détectée entre l'espèce ou la taille du gène séquencé et le pourcentage de recombinaison *in vitro* détectée. Ces séquences recombinées ont été supprimées des différents jeux de donnée.

4.3. Correction du jeu de données : suppression des singletons artéfactuels engendrant un multiallélisme

Des erreurs d'amplification peuvent engendrer la présence de singletons artéfactuels dans les séquences soit par erreur de réplication de la Taq polymérase ou lors de recombinaison entre allèles si l'ADN est de mauvaise qualité. À cela s'ajoute les mutations somatiques accumulées au cours de la croissance de l'organisme et non-transmises dans les lignées germinales. Si ces erreurs d'amplification ne sont généralement pas visibles lors d'un séquençage direct de produit de PCR (nombre de copies important, masquant ces erreurs), le passage par une étape de clonage, en isolant les copies, les met en évidence. Ainsi, pour un gène de 1000 pb, les séquences diffèrent généralement entre elles d'environ un à deux singletons artéfactuels. C'est pourquoi nous avons choisi d'appliquer une méthode de correction systématique des séquences nucléaires. Dans un premier temps, trois fichiers de « qualité » différente sont réalisés :

(1) jeu qualité 1 (Fig. 2.4). Contient l'ensemble des séquences recapturées, corrigées et réattribuées à un individu.

(2) jeu qualité 2 (Fig. 2.4). Pour chaque individu, une séquence consensus des séquences peu différentes entre elles (quelques singletons) est réalisée. Il est noté entre parenthèses le nombre de séquences ayant permis de réaliser ce consensus de la manière suivante : premier chiffre correspondant au nombre de séquences strictement identiques à la séquence consensus, puis les autres chiffres correspondent aux autres groupes de séquences différentes. Par exemple, cons (3 1 2) signifie : 3 séquences recapturées strictement identiques à la séquence consensus, puis 1 séquence différente recapturée une seule fois, puis 1 autre séquence différente des précédentes recapturée 2 fois. De même, cons (0 1 1 1) signifie qu'aucune séquence ne correspond strictement à la séquence consensus, le consensus a été fait sur la base de 3 séquences proches recapturées une seule fois chacune. Théoriquement (i.e. en absence de co-amplification de gènes), les fichiers de qualité 2 comptent donc ainsi au maximum deux séquences par individu recapturé correspondant aux 2 allèles de cet individu et sont utilisés lors de la réalisation des arbres et réseaux phylogénétiques (permettant de visualiser ainsi de potentiels individus hybrides).

(3) jeu qualité 3 (Fig. 2.4). Une seule séquence par individu est sélectionnée : la séquence consensus issue du plus grand nombre de recaptures. Ce jeu est utilisé pour effectuer les tests de neutralité pour éviter de biaiser la diversité nucléotidique observée, le nombre d'allèles recapturés n'étant pas systématiquement égal à 2 pour tous les individus.

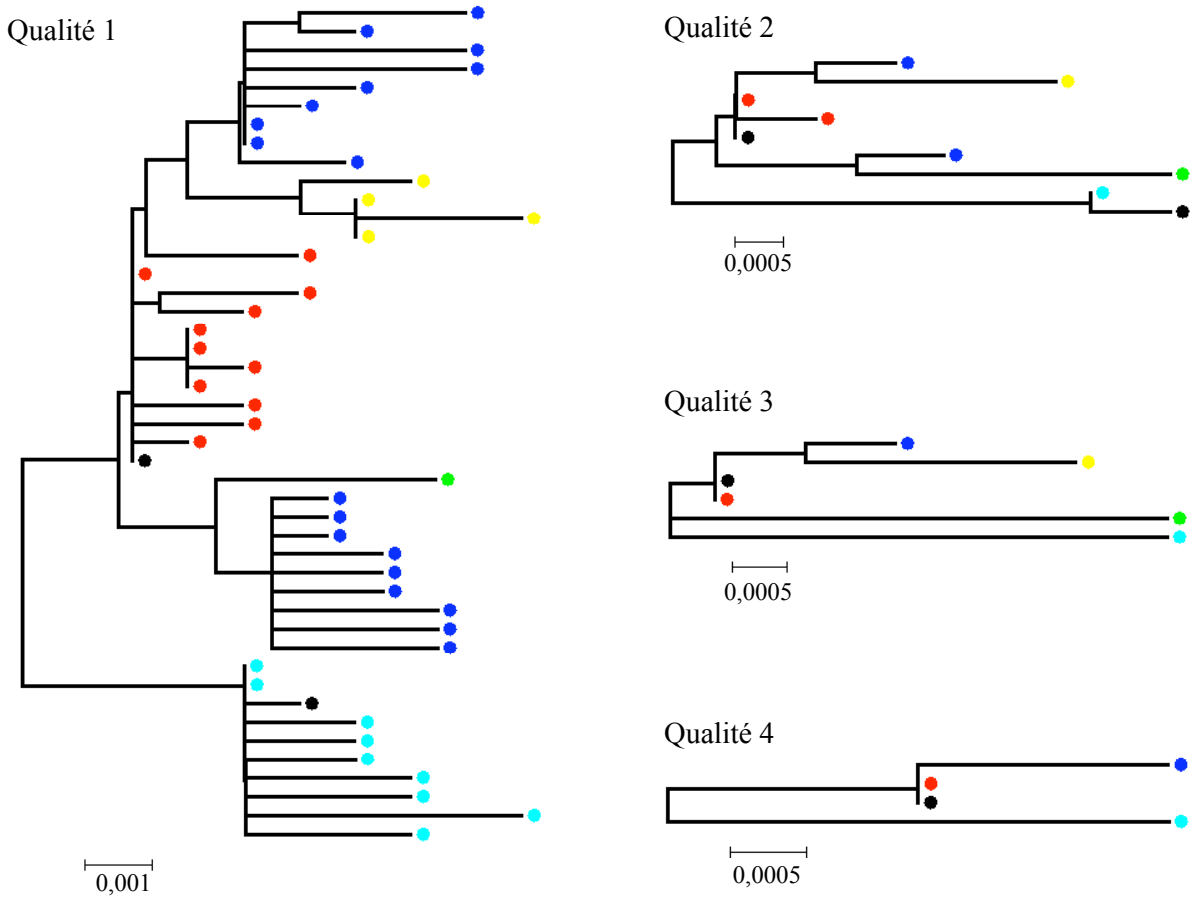


Fig. 2.4 Arbres phylogénétiques réalisés à partir de six individus sur le gène Lysozyme de *B. thermophilus* illustrant les différents jeux de qualité obtenus
Chaque couleur représente un individu différent.

5. *Traitement de la recombinaison*

5.1. Biais liés à la recombinaison dans les analyses de coalescence

Outre les recombinants *in vitro* (artéfactuels), la recombinaison peut avoir une réalité biologique. Cependant, la plupart des analyses basées sur la théorie de la coalescence pose l'hypothèse de l'absence de recombinaison. En effet, la recombinaison augmente la diversité génétique observée et modifie la forme du coalescent en établissant des nœuds intermédiaires entre 2 allèles pour lesquels une divergence existe (rajeunissement de la divergence : Shriner et al 2003). Les sites recombinés présentent une diversité plus élevée que sous l'attendu neutre sans recombinaison, pouvant ainsi biaiser les analyses utilisant

des indices de diversité génétique. Avant d'effectuer les analyses de coalescence, il est donc nécessaire de détecter et de supprimer du jeu de données les séquences potentiellement recombinantes.

5.2. Détection et suppression des séquences recombinantes

Afin de supprimer les séquences présentant de la recombinaison, nous avons tout d'abord utilisé le test des 4 gamètes de Hudson & Kaplan (1985) via le logiciel DNAsp 4.10.3 (Rozas et al 2003) afin de détecter le nombre minimum de points de recombinaison (R_m) au sein du jeu de séquences et les sites nucléotidiques incriminés dans ces évènements de recombinaison. Une mutation reverse (homoplasique) peut également être comptée comme un événement de recombinaison par ce test (un point unique de mutation pouvant être compté comme événement de recombinaison s'il n'est pas en équilibre de liaison avec les sites adjacents comparés, observation personnelle). Ces sites de recombinaison sont comparés deux à deux pour l'ensemble des séquences et les séquences détectées recombinantes (ou présentant des sites homoplasiques) sont éliminées du jeu de données (appelé de qualité 4, Fig. 2.4) jusqu'à atteindre $R_m = 0$. Ce jeu de qualité 4 est alors utilisé pour les analyses de polymorphisme et de divergence posant l'hypothèse de l'absence de recombinaison.

6. *Détection de gènes co-amplifiés (potentiellement paralogues)*

6.1. *Bathymodiolus thermophilus*

6.1.1. Lysozyme

Faure (2008) a mis en évidence la présence de deux gènes paralogues sur le lysozyme, ces gènes étant co-amplifiés avec les amorces utilisées (divergence de 4% entre ces paralogues). Cependant, cet auteur n'ayant recapturé qu'une séquence du deuxième paralogue sur l'EPR, nous n'avons pas redéfini d'amorces spécifiques du premier paralogue. Hormis quatre séquences supplémentaires correspondant au deuxième paralogue, toutes les séquences obtenues sur le Lysozyme correspondaient au paralogue 1. Les analyses ont donc été effectuées uniquement sur le paralogue 1 dans le chapitre 3.

6.1.2. Sulfotransférase

Certains individus recapturés sur le gène de la Sulfotransférase présentaient trois à quatre allèles distincts (substitutions multiples) par individu, laissant penser que nous avions co-amplifié au moins deux gènes avec le couple d'amorces utilisé. Deux clades divergents de 2,6% ont été identifiés (Fig. 2.5). La répartition des séquences entre individus présentant plus de deux allèles distincts nous a permis de poser l'hypothèse d'une paralogie associée à ces deux clades. En effet, certains individus comptaient deux allèles dans le clade 1 et deux autres dans le clade 2, mais aucun individu ne possédait plus de deux allèles recapturés au sein d'un même clade. De plus, le clade 1 présentait une divergence génétique entre les individus échantillonnés au Nord et au Sud de la barrière au flux de gènes préalablement détectée sur d'autres gènes (Fig. 2.5). De même, les séquences du clade 2 se regroupent préférentiellement selon leur géographie (en termes de fréquences). Les clades 1 et 2 ont donc été considérés comme deux gènes distincts, respectivement nommés Sulfotransférase 1 et 2, pour les analyses sur *B. thermophilus* du chapitre 3.

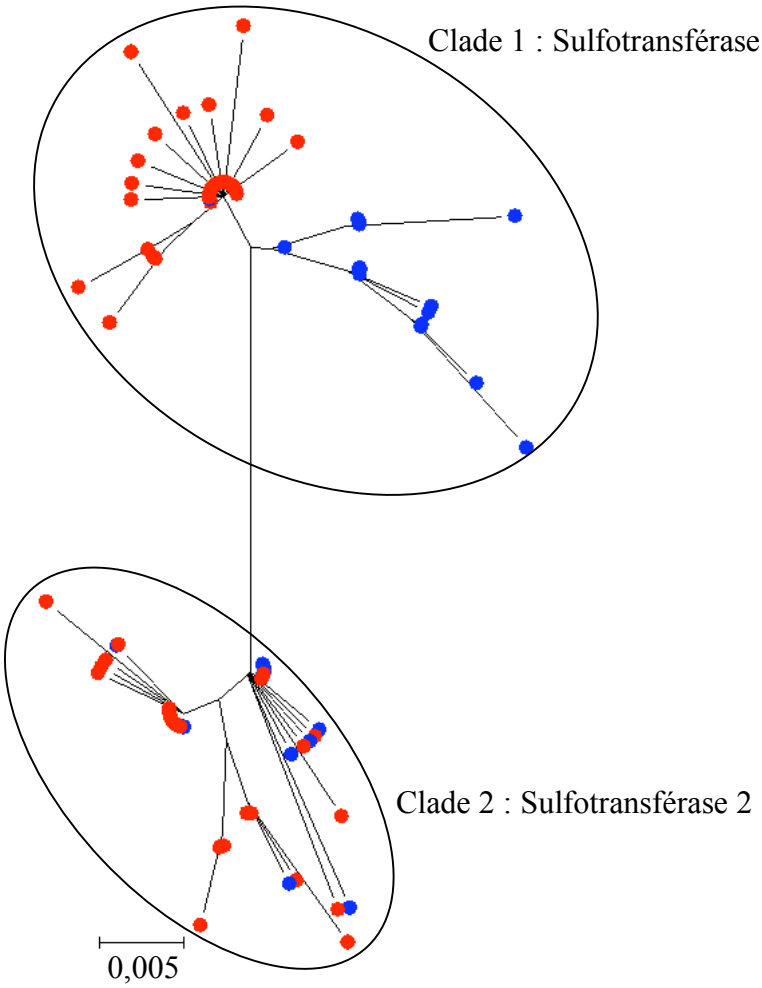


Fig. 2.5 Arbre phylogénétique permettant l'identification de deux gènes co-amplifiés sur la Sulfotransférase.
En bleu, individus échantillonnés à 9°50'N et 7°25'S.
En rouge, individus échantillonnés de 14°S à 21°33'S.

6.2. *Lepetodrilus elevatus*

Matabos et al (2008b) ont décrit l'existence de deux espèces cryptiques de *L. elevatus* divergentes de 6,5% sur le gène mtCOI. Cette étude portait sur trois sites : 13°N, 9°50'N et 17°25'S. Alors que les deux espèces sont présentes en sympatrie à 9°50'N, seule une espèce était échantillonnée à 13°N et la seconde à 17°25'S. Lors de mon stage de Master 2 (Plouviez 2006), nous avons séquencé trois gènes nucléaires pour des individus présents à 13°N et 17°25'S et donc diagnostiques des deux espèces cryptiques. Nous fûmes alors confrontés à un problème de multiallélisme au niveau de ces gènes nucléaires, laissant penser à la co-amplification de plusieurs gènes. Durant la thèse, nous avons donc

tenté de résoudre ces problèmes de multiallélisme afin d'utiliser ces gènes dans le cadre d'une étude phylogéographique.

6.2.1. Lysozyme

Dans un premier temps, une quinzaine de *L. elevatus* ont été clonés et séquencés individuellement (pas d'utilisation de la méthode de MR). Pour chaque individu, nous avons séquencé une dizaine de clones afin de détecter les gènes co-amplifiés. Quatre clades ont ainsi pu être mis en évidence (Fig. 2.6).

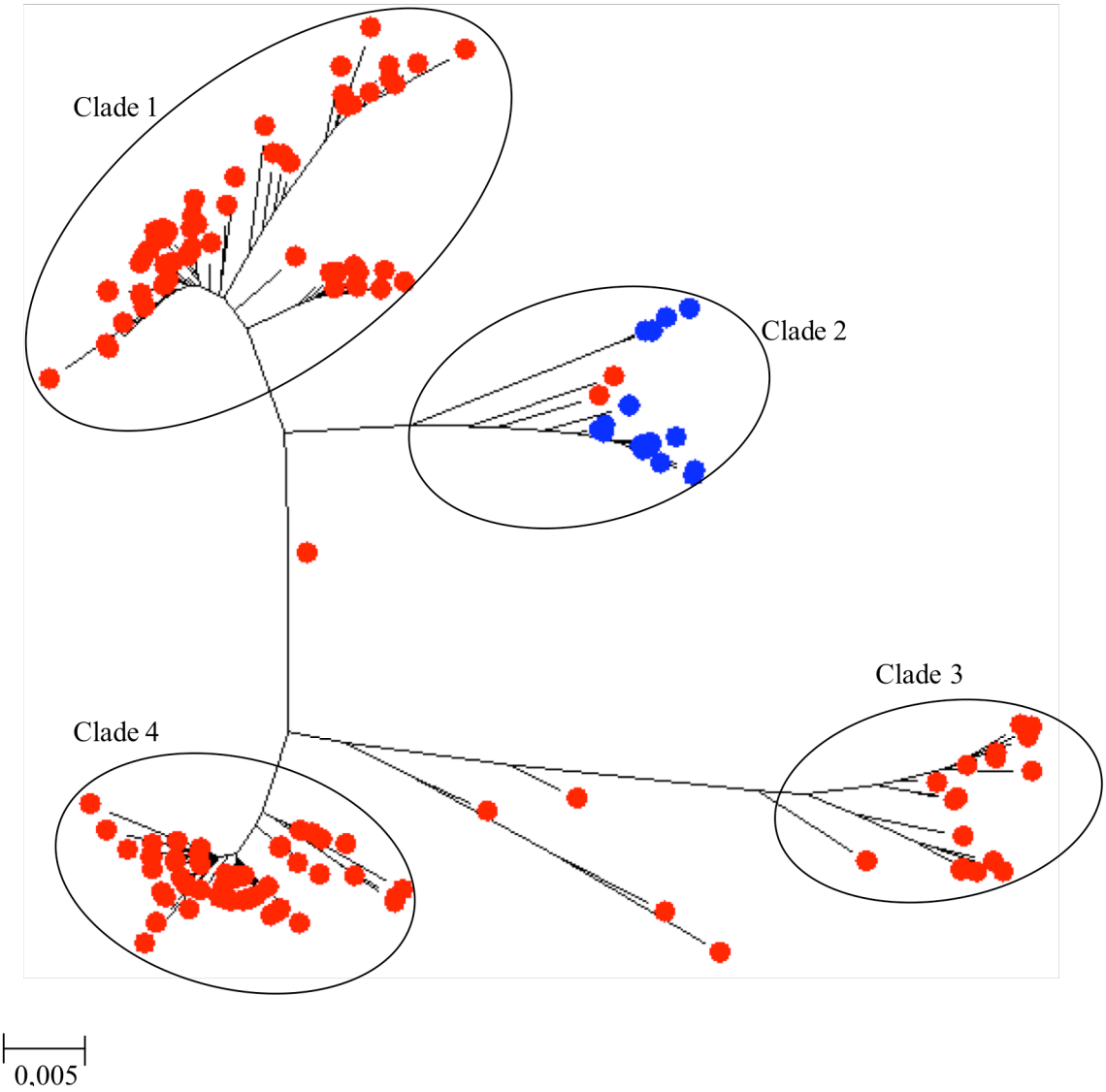


Fig. 2.6 Arbre phylogénétique de type qualité 1 réalisé sur les séquences des individus clonés individuellement sur le Lysozyme.
En bleu et rouge, espèces cryptiques de type 13°N et 17°25'S (Matabos et al 2008), respectivement, sur la base du mtCOI sur la base du mtCOI (Matabos et al 2008). Les allèles présents sur la divergence constituent des recombinants.

Les séquences des clades 1 et 2 présentent le même polymorphisme d'indels (insertions/délétions) et le même type exonique. Elles diffèrent essentiellement par des substitutions situées dans l'intron. Le clade 3 diffère notamment de ces deux clades par un nombre de 13 substitutions non-synonymes fixées sur l'exon 2. Le clade 4 possède le même type d'exon que les clades 1 et 2. Il diffère des autres clades par une délétion d'environ 560 pb au début de l'intron et une insertion de ~260 pb à la fin de celui-ci, ces indels étant également en déséquilibre de liaison à des substitutions introniques multiples. Certains individus comptaient des séquences à la fois dans les clades 1, 3 et 4, indiquant la présence d'au moins deux gènes co-amplifiés. Le clade 2 étant essentiellement constitué de séquences de l'espèce type 13°N, nous avons considéré que la divergence observée entre les clades 1 et 2 (Fig. 2.6) était due à la divergence entre espèces cryptiques et non entre gènes différents. Nous avons alors défini trois couples d'amorces discriminant les clades 1-2, 3 et 4 et testé ces amorces sur une cinquantaine d'individus par espèce. Pour les deux espèces, nous avons ainsi mis en évidence des individus amplifiant pour les trois couples d'amorces, confirmant l'existence d'au moins deux gènes différents. Néanmoins, contrairement aux amorces définies pour les clades 1-2, les clades 3 et 4 n'amplifiaient pas sur l'ensemble des individus, suggérant un mauvais échantillonnage de ces clades 3 et 4. Par conséquent, nous avons considéré que les clades 1 et 2 constituaient un gène unique et séquencé ce gène par la méthode de MR pour l'étude phylogéographique. Les séquences obtenues à l'issue de ce séquençage se sont révélées compatibles avec un gène unique (deux allèles maximum par individus échantillonnés).

6.2.2. Veliger Digestive Gland 3 (VDG3)

Le gène VDG3 présentait également une co-amplification d'allèles appartenant à plusieurs locus (allèles multiples) avec la présence d'au moins deux clades divergents sans rapport avec la présence des 2 espèces cryptiques. Nous avons donc redéfini une amorce sens afin de n'amplifier qu'un seul clade (l'amorce antisens restant inchangée). Nous avons ensuite cloné individuellement les individus de chaque espèce, respectivement de type 13°N et 17°25'S au locus mtCOI (Matabos et al 2008b) et séquencé 6-8 clones par individu. Les séquences obtenues comptaient un maximum de deux allèles par individu, suggérant que les allèles observées proviennent d'un gène unique. Là encore les allèles se

regroupent selon les espèces cryptiques (Fig. 2.7). Nous avons donc utilisé ce couple d'amorces pour le séquençage par la méthode MR.

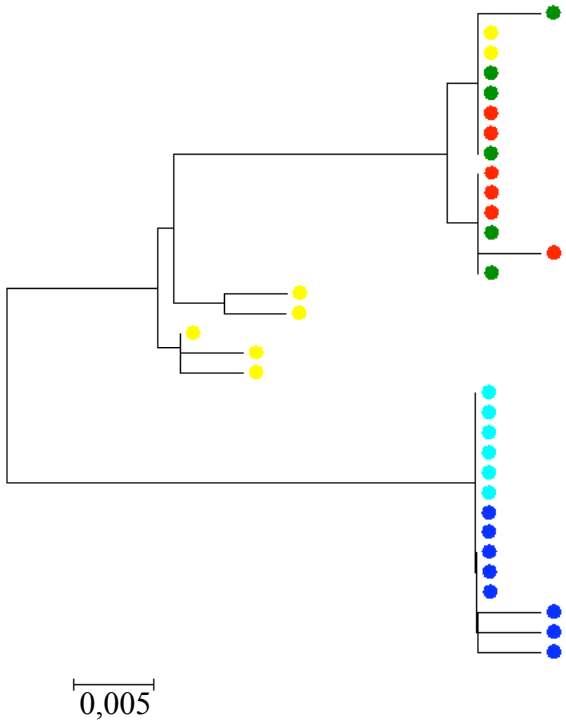


Fig. 2.7 Arbre phylogénétique de type qualité 1 réalisé sur les séquences des individus clonés individuellement sur le VDG3 avec les amorces redéfinies.
Chaque couleur représente un même individu. En bleu, séquences de deux individus correspondant à l'espèce cryptique de type 13°N sur le mtCOI (Matabos et al 2008). En jaune, rouge et vert, les séquences d'individus de l'autre espèce cryptique. Chaque individu présente au maximum deux allèles différents (séquences comptant un polymorphisme de singletons, e.g. individu en jaune, mais correspondant à un même allèle consensus pour le jeu de qualité 2)

Les séquences obtenues par MR se distribuaient en trois clades majoritaires, avec des séquences intermédiaires entre ces clades (Fig. 2.8.A). Si la plupart des individus séquencés présentaient au maximum deux allèles par individu, dix individus comptaient plusieurs allèles bien différents. Nous avons donc examiné, pour chacun de ces individus, s'il existait des allèles recombinés. Pour ces dix individus, nous avons ainsi pu mettre en évidence la présence d'un ou deux allèles appartenant à un des 2 clades préalablement identifiés ainsi que des allèles recombinés intra-clade, des allèles recombinés entre les 2 clades et, de façon encore plus inattendue, des allèles recombinés entre le gène d'intérêt et le gène que nous avions choisi d'exclure par le biais de l'utilisation de la nouvelle amorce sens. Les ADN génomiques de nos individus étant assez dégradés, il est possible que la

Taq polymérase ait associé des morceaux de séquences sens et antisens des différents clades voire des différents paralogues (l'amorce antisens n'ayant pas été redéfinie). Pour ces dix individus, après suppression des séquences recombinées, nous avons obtenu un maximum de deux allèles différents par individu et les clades de l'arbre phylogénétique se sont révélés mieux définis (Fig. 2.8.B). Nous avons donc considéré que ce gène constituait un gène unique et avons donc utilisé ces séquences pour l'analyse phylogéographique de *L. elevatus* dans le chapitre 4.

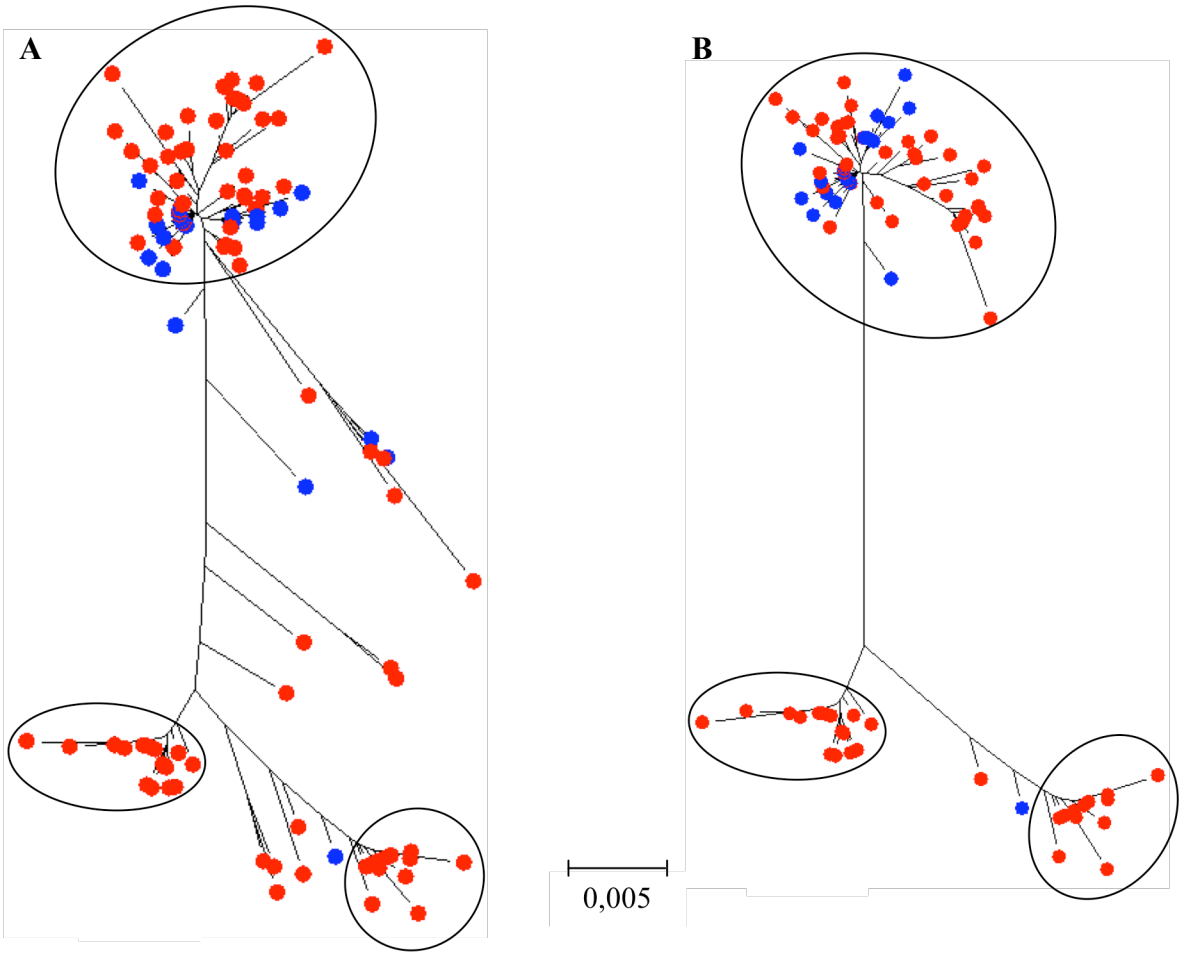


Fig. 2.8 Arbre phylogénétique de type qualité 2 réalisé sur les séquences issues de la méthode MR sur le VDG3 comprenant les séquences recombinantes des dix individus présentant plus de deux séquences bien différentes (A), sans les séquences recombinantes de ces dix individus (B).

6.2.3. Textilinine

Les séquences du gène Textilinine obtenues par la méthode MR se répartissent en quatre clades principaux (Fig. 2.9). Certains individus comptent plus de dix séquences , ce qui a permis la mise en évidence d'allèles multiples (plus de deux allèles par individu), suggérant, là encore, la co-amplification de plusieurs gènes. Un individu recapturé douze fois possède un allèle dans le clade 1, deux dans le clade 2, et un dans le clade 3, indiquant que le clade 2 constituerait un gène différent des deux autres clades. De même, un autre individu recapturé treize fois présente un allèle dans le clade 1, un dans le clade 3 et deux allèles dans le clade 4, suggérant que le clade 4 serait également différent des clades 1 et 2. Nous n'avons pas recapturé d'individus rendant incompatible l'appartenance des clades 1 et 3 à un même gène. De plus, l'ensemble des séquences recapturées provenant de l'espèce cryptique mitochondriale de type 13°N (Matabos et al 2008b) est distribué dans le clade 1, indiquant que la divergence entre les clades 1 et 3 pourraient correspondre à la divergence attendue entre 2 espèces différentes. Les clades 2 et 4 pourraient constituer un deuxième gène co-amplifié, mais aucune séquence de l'espèce cryptique de type 13°N n'a été recapturée dans ces clades (amorces ne permettant pas d'amplifier cette espèce, ou apparition d'une duplication secondaire dans la lignée mitochondriale sud de *L. elevatus*). Par conséquent, bien que l'approche reste critiquable, nous avons considéré que les clades 1 et 3 pouvaient constituer un gène unique utilisable pour une approche de phylogéographie sur *L. elevatus*.

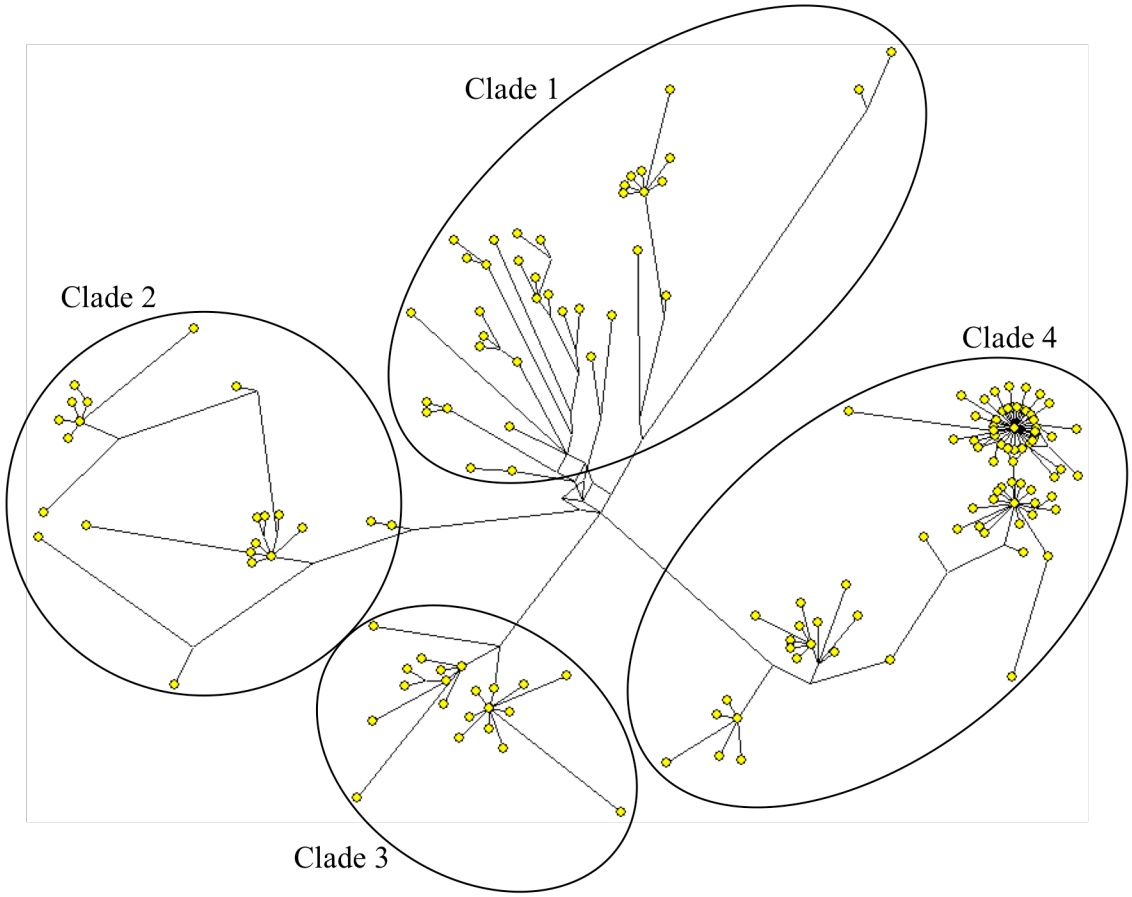


Fig. 2.9 Réseau réalisé en Median Joining des séquences de Textilinine sur le jeu de données de qualité 2 pour lequel des individus présentent des allèles multiples.
Pour faciliter la représentation, le diamètre des cercles n'est pas proportionnel au nombre de séquences recapturées. La longueur des branches est proportionnelle à la divergence.

7. *Recherche d'une divergence partagée entre plusieurs espèces : MsBayes*

Lorsque plusieurs espèces présentent deux clades divergents de part et d'autre d'une potentielle barrière aux flux de gène, il s'avère intéressant de tester si la divergence observée entre ces paires de clades est due à un événement d'isolement commun à toutes les espèces ou à plusieurs évènements différents et ainsi de dater ces évènements. Une nouvelle approche dite 'Hierarchical Approximate Bayesian Computation' (HABC) permet d'estimer, via le logiciel MsBayes (Hickerson et al 2006b), ce nombre

d'évènements de divergence (hypervariable Ψ) et de les dater (temps de divergence moyen E(τ) et Ω = E(τ)/Var(τ)). Cette approche est basée sur des simulations de statistiques résumées (e.g. diversité génétique, θ_W) à partir d'un modèle d'isolement en 2 populations sœurs évoluant sous l'hypothèse d'une variation indépendante de leur taille efficace (Fig. 2.10).

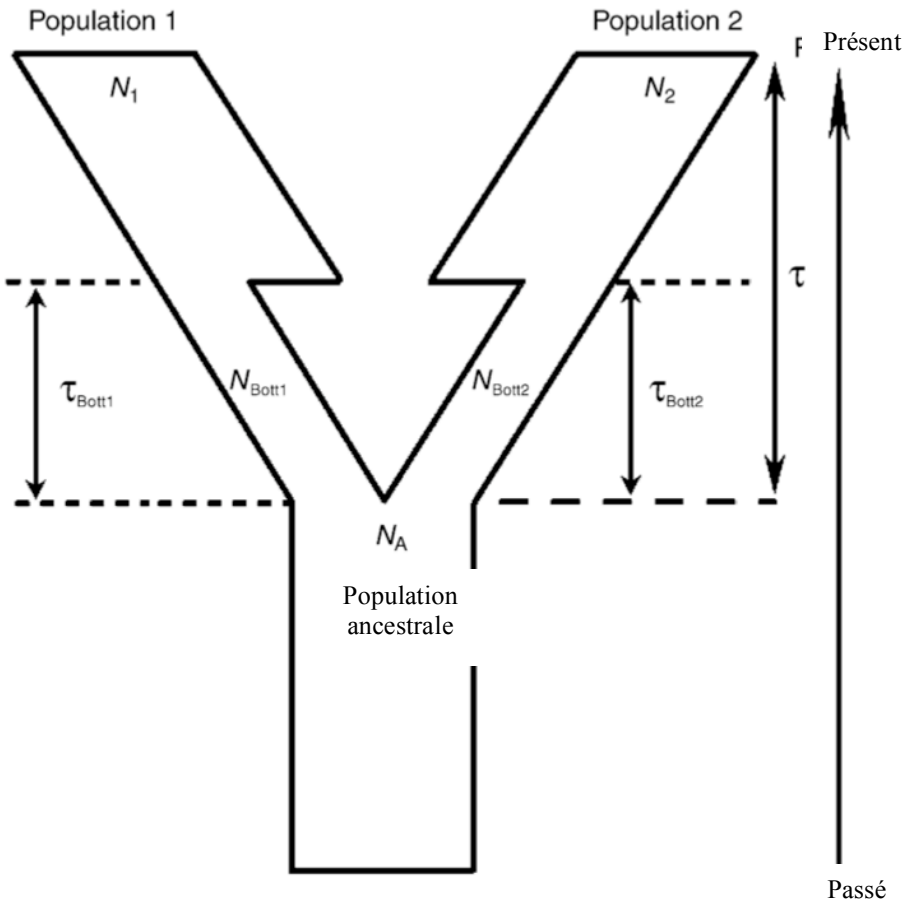


Fig. 2.10 Représentation schématique d'un modèle où une population ancestrale se sépare en deux populations filles 1 et 2 au temps τ. Tirée de Hickerson et al (2006a).
N_A, N_1 et N_2, tailles des populations ancestrales, fille 1 actuelle et fille 2 actuelle, respectivement.
N_{bott1} et N_{bott2}, tailles des populations 1 et 2 après la séparation des populations pour les temps τ_{bott1}, τ_{bott2}
τ, temps d'isolement des populations

Cette approche se divise en trois étapes :

(1) Un fichier contenant les séquences observées des deux clades divergents est fourni pour chaque espèce. À partir de ces jeux de séquences, le logiciel MsBayes estime une série de statistiques résumées pour chaque clade/lignée divergente (Tableau 2.2).

Tableau 2.2 Statistiques résumées estimés à partir des séquences observées ou simulées. Tiré de Hickerson et al (2006a).

Notation	Statistique résumée	Référence
π	Moyenne des différences par paires de séquences	Tajima 1983
π_b	Moyenne des différences par paires de séquences interclades	Takahata & Nei 1985
π_w	Moyenne des différences par paires de séquences intraclades	Takahata & Nei 1985
π_{net}	Moyenne nette des différences par paires de séquences entre clades ($=\pi_w - \pi_b$)	Takahata & Nei 1985
S	Nombre de sites polymorphes	Watterson 1975
θ_w	Nombre de sites polymorphes normalisés par la taille de l'échantillonnage (thêta de Watterson)	Watterson 1975
D	D de Tajima	Tajima 1989

(2) 100 000 hyper-paramètres (inter-espèces : Ψ, E(τ) et Ω) et sous-paramètres (intra-espèce : N_A, N_1, N_2, N_{bott1}, N_{bott2}, τ_{bott1}, τ_{bott2}, τ, calculés pour chaque espèce selon le modèle présenté Fig. 2.10) sont aléatoirement tirés du modèle d'isolement avec goulots d'étranglements dans les lignées produites. Chacune de ces séries de paramètres permet de simuler un jeu différent de données de séquences (100 000 jeux de séquences) pour chaque espèce (paires de clades). Les statistiques résumées sont ensuite calculées pour chaque jeu de séquences simulé

(3) Les statistiques résumées calculées à partir des jeux de séquences simulés sont ensuite comparées à celles des séquences observées. Les distributions des probabilités postérieures des hyper-paramètres (Ψ, E(τ) et Ω) et sous-paramètres (N_A, N_1, N_2, N_{bott1}, N_{bott2}, τ_{bott1}, τ_{bott2}, τ) associés proviennent des 1000 statistiques résumées qui s'approximent le mieux aux statistiques résumées observées.

8. *Migration à travers une barrière aux flux de gènes : Migrate-n*

La migration entre populations peut être analysée via des logiciels tels que Migrate-n (Beerli & Felsenstein 2001), et notamment permettre l'analyse de cette migration à travers une barrière géographique. Migrate-n utilise une approche basée sur la théorie de la coalescence afin d'estimer les valeurs les plus vraisemblables des taux de migration entre populations et celles de leur taille efficace (Beerli & Felsenstein 1999, 2001). La matrice $P = \begin{pmatrix} \Theta_1 & M_{2,1} & M_{3,1} & \dots & M_{n,1} \\ M_{1,2} & \Theta_2 & M_{3,2} & \dots & M_{n,2} \\ \dots & \dots & \dots & \dots & \dots \\ M_{1,n} & M_{2,n} & \dots & M_{n-1,n} & \Theta_n \end{pmatrix}$ contient l'ensemble de ces paramètres estimés (Fig. 2.11), avec $M_{i,j}$ correspondant au taux de migration de la population i vers j divisé par le taux de mutations par génération (μ), et Θ_i correspondant au paramètre démographique de la population i (= 4 Ne_i μ avec Ne_i la taille efficace de la population i). Les valeurs de vraisemblance de cette matrice (*L(P)*) sont calculées selon la relation $L(P) = \sum_G prob(G/P).prob(D/G)$ avec prob(G/P), la probabilité d'observer la généalogie pour les paramètres *P* choisis et prob(D/G) la probabilité d'obtenir le jeu de données observé pour une généalogie simulée.

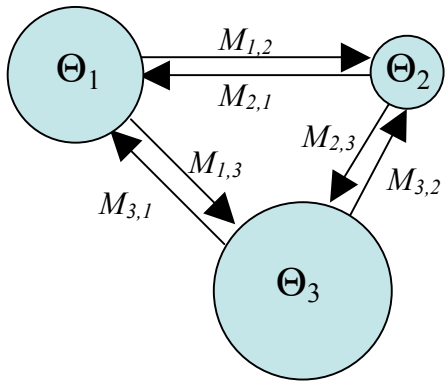


Fig. 2.11 Représentation graphique du modèle utilisée par Migrate-n.
Exemple de trois populations (cercles), les valeurs de $M_{i,j}$ pouvant être nulles

Une approche Metropolis-Hastings de chaînes de Markov Chain selon une procédure Monte Carlo (MCMC, Metropolis et al 1953) est utilisée afin d'échantillonner les valeurs de *P* les plus probables dans l'espace des paramètres possibles (i.e. espace des coalescents possibles). Cette approche MCMC consiste à estimer la vraisemblance de paramètres *L(P)* aléatoirement choisis dans cet espace. Les paramètres de cette matrice de départ sont ensuite légèrement modifiés pour donner une nouvelle matrice de paramètres et sa valeur de vraisemblance associée est comparée à la précédente :

(1) Si $L(P)_{\text{nouvelle matrice}} > L(P)_{\text{matrice de départ}}$ alors le processus d'exploration des paramètres se poursuit en estimant une autre matrice.

(2) Lorsque $L(P)_{\text{matrice n}} < L(P)_{\text{matrice n-1}}$, alors le processus de recherche des paramètres P est stoppé et les paramètres de la matrice n-1 sont considérés comme les plus vraisemblables.

Afin de mieux explorer l'univers des possibles, ce processus est effectué sur plusieurs matrices de départ, permettant ainsi de s'assurer que le choix de la matrice la plus vraisemblable converge entre les simulations et ainsi de maximiser la probabilité que cette matrice reflète les patrons de migration réels entre populations.

En estimant les valeurs les plus probables des taux de migration souvent assymétriques entre populations, le logiciel Migrate-n peut ainsi permettre de mieux appréhender le sens de la migration entre populations et notamment au travers une barrière aux flux de gènes. Néanmoins, ce programme fait l'hypothèse de l'équilibre mutation/dérive (pas de changements démographiques au cours du temps) et d'absence de recombinaison intralocus (et interlocus lorsque plusieurs gènes sont utilisés dans une même simulation : série de génotypes pour un même individu).

Le logiciel Migrate-n permet donc de comparer les compositions génétiques des régions géographiques et peut permettre de définir des taux de migration entre ces régions. Cette approche ne fait pas d'hypothèse sur la spéciation en cours. Ainsi, lorsque deux entités génétiques différentes sont présentes au sein d'une même population, les individus migrants mais aussi les individus potentiellement introgressés entre ces entités seront comptabilisés dans le calcul du taux de migration.

9. *Modèle d' « Isolement avec Migration »*

La séparation d'une population ancestrale en deux populations filles divergentes de part et d'autre d'une barrière physique ou génétique est un processus long qui peut se réaliser en absence totale ou en présence d'un flux de gènes limité. Le modèle d'Isolement avec Migration (logiciel IM et IMa, Nielsen & Wakeley 2001, Hey & Nielsen 2004, 2007) permet de distinguer l'effet relatif de la divergence et des flux de gènes dans le processus ayant conduit à l'isolement génétique de populations. Une approche MCMC est utilisée afin d'estimer les probabilités postérieures des différents paramètres du modèle (Fig. 2.12) soient:

(1) θ_A, θ_1, et θ_2, les paramètres démographiques des populations ancestrales, filles 1 et 2, respectivement (avec $\theta_i = 4N_i u$ pour un gène nucléaire et $2 N_i u$ pour un gène mitochondrial, N_i correspondant à la taille de la population i considérée)

(2) La proportion d'individus de la population 1 qui est remplacée à chaque génération par des migrants de la population 2 (m1) et vice-versa (m2) et

(3) Le temps de début de séparation des lignées (t).

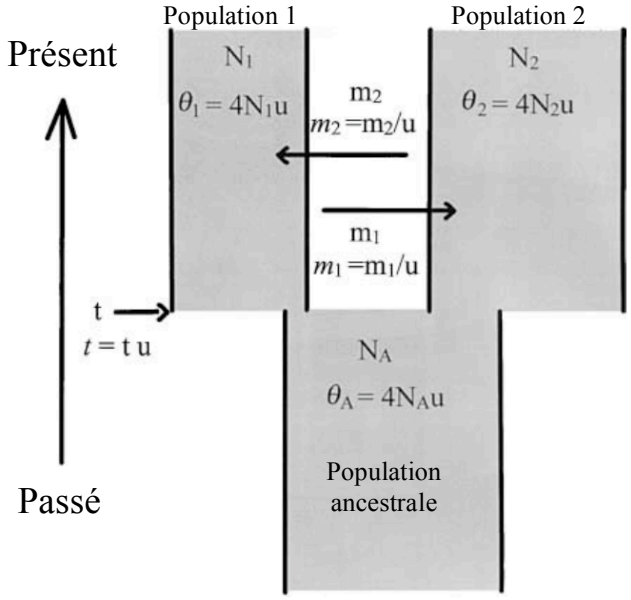


Fig. 2.12 Représentation graphique du modèle d'Isolement avec Migration. Tirée de Nielsen & Wakeley (2001).
u, taux de mutation par génération.

Ce modèle d'Isolement avec Migration peut ainsi permettre d'émettre des hypothèses pour expliquer les flux de gènes potentiellement détectés. Ainsi, la présence de flux de gènes entre les populations pourrait correspondre à un isolement progressif des populations en présence de migration pendant le processus d'isolement (spéciation parapatrique/sympatrique) ou à un tri incomplet des allèles (phase de spéciation toujours en cours). Au contraire, une absence de flux de gènes entre les populations accrédite un processus de type allopatrique dans lequel la barrière est totalement imperméable. Il existe cependant un nombre de cas ou cette barrière cesse d'être imperméable et favorise la mise en place d'une zone de contact secondaire entre les populations préalablement isolées. Dans ce cas, des flux de gènes récents sont en général observés et traduisent des phénomènes d'hybridation avec une possibilité d'introgression d'allèles à certains locus si les croisements des hybrides avec les populations parentales sont encore possibles (e.g. Faure et al 2009).

Contrairement à Migrate-n, le modèle d'Isolement par la distance prend donc en compte la différenciation dans le temps.

10. *Écart à la neutralité : sélection contre démographie*

10.1. Approche multilocus visant à détecter un ou plusieurs locus sous sélection positive : test de Hudson Kreitman Aguadé

Hudson, Kreitman et Aguadé (HKA, 1987) proposent un test statistique multilocus de polymorphisme/divergence basé sur la comparaison de séquences d'ADN intra- et inter-spécifiques. Un nombre *L* de locus sont échantillonnés respectivement chez deux espèces proches. Soient S_i^A et S_i^B le nombre de sites nucléotidiques polymorphes respectivement chez les espèces A et B au locus i. et Di le nombre de différences nucléotidiques entre deux séquences tirées au hasard chez chacune de ces deux espèces au même locus i. Sous l'hypothèse neutraliste, S_i^A, S_i^B et D_i ne sont soumis qu'à la dérive génétique. Si l'ensemble des locus échantillonnés est neutre, alors la statistique

$$X^2 = \sum_{i=1}^{L} \frac{(S_i^A - \hat{E}(S_i^A))^2}{\hat{Var}(S_i^A)} + \sum_{i=1}^{L} \frac{(S_i^B - \hat{E}(S_i^B))^2}{\hat{Var}(S_i^B)} + \sum_{i=1}^{L} \frac{(D_i - \hat{E}(D_i))^2}{\hat{Var}(D_i)}$$ suit une distribution de χ^2 à

2L – 2 degrés de liberté. Au contraire, si cette statistique ne suit pas cette distribution, on peut en déduire qu'au moins un locus s'écarte du modèle neutre. Or, un locus soumis à la sélection darwinienne positive est supposé avoir un rapport polymorphisme sur divergence plus faible qu'un locus neutre. Par conséquent, le test HKA peut permettre de déterminer les locus assujettis à sélection darwinienne positive, sans pour autant donner la raison de cette sélection. Néanmoins, ce test ne permet pas d'affirmer si le locus échantillonné le moins soumis à sélection est neutre ou simplement sous des pressions de sélection plus faible. De plus, ce test fait plusieurs hypothèses pouvant être invalides, en particulier, il suppose l'absence de recombinaison intralocus et de liaison inter-locus mais surtout des tailles efficaces stationnaires depuis la spéciation des deux espèces (Hudson et al 1987). Or, cette dernière est rarement vérifiée (Nei & Kumar 2000).

10.2. Approche substitutions synonymes / non-synonymes en vu de la détection de sélection positive versus négative : test de MacDonald-Kreitman

Le test de McDonald-Kreitman (MK, 1991a) repose également sur la comparaison intra- et inter-spécifique de séquences mais en comparant 2 classes de mutations : les mutations non-synonymes assujetties à la sélection et les mutations synonymes supposées neutres. Il s'agit d'un test utilisant une approche mono-locus. Il confronte le ratio de polymorphisme p_n / p_s au ratio de divergence d_n / d_s (avec p_n et p_s le nombre de différences respectivement non-synonymes et synonymes entre séquences d'une même espèce, d_n et d_s le nombre de différences respectivement non synonymes et synonymes entre séquences d'espèces différentes mais proches). Si le locus étudié a évolué de façon neutre chez les espèces étudiées, alors ces rapports doivent être les mêmes: $p_n / p_s = d_n / d_s$. Si $p_n / p_s < d_n / d_s$, alors des mutations non synonymes se sont fixées dans la divergence entre espèces sous l'effet de la sélection darwinienne positive. Si $p_n / p_s > d_n / d_s$, des mutations non synonymes légèrement délétères ségrègent dans les populations mais ne fixent pas. Un simple test d'homogénéité permet de détecter l'écart à l'hypothèse de neutralité. Le test MK (1991a) permet de tester si un locus particulier est soumis à sélection sans nécessité l'accès à d'autres locus. De plus, ce test a pour avantage de pouvoir comparer plusieurs espèces relativement proches et ne fait pas d'hypothèse sur la taille efficace des populations. Cependant, Graur et Li (1991) et Whittam et Nei (1991) suggèrent qu'un faible nombre de séquences ou d'espèces échantillonnées peut sous estimer le ratio d_n / d_s.

McDonald et Kreitman (1991b) rétorquent que cela n'affecterait pas la validité du test de neutralité puisqu'un faible échantillonnage de séquences et d'espèces entraînerait une sous-estimation égale des deux ratios. De plus, ce test conjecture un taux de mutations synonymes stationnaire durant le processus d'évolution et l'absence de liaison génétique entre sites nucléotidiques (effet autostop négligé).

10.3. Tests de polymorphisme visant à détecter un écart à la neutralité : tests de Tajima et Fu & Li

Tajima (1989) propose un test statistique basé sur des données de polymorphisme intra-population. La statistique $D = \frac{(\hat{\theta}_H - \theta_S)}{\left[Var(\hat{\theta}_H - \theta_S)\right]^{1/2}}$ dépend de la différence entre deux estimateurs du thêta de Watterson (1975) ou diversité nucléotidique : $\hat{\theta}_H$ le nombre moyen de différences entre les paires de séquences échantillonnées et θ_S le nombre total de sites polymorphes observés ; pondérée par l'estimation de la variance de cette différence. Si le polymorphisme génétique s'explique par la seule dérive (polymorphisme neutre), alors ces deux paramètres sont égaux et le D de Tajima est nul. Au contraire, un allèle en fréquence rare entraînera une augmentation du nombre total de sites polymorphes mais aura un impact minime sur le nombre moyen de différences entre paires de séquences (effet sensible à la taille de l'échantillonnage). Un excès de variants rares engendrera ainsi un D de Tajima négatif alors qu'il sera positif pour un excès de variants intermédiaires. L'excès de variants rares indique généralement une sélection positive récente (un allèle avantageux a récemment envahi la population et des variants neutres encore rares sont apparus par mutation), un effet autostop (mutations neutres liées à l'avantage ou de la sélection négative (des allèles faiblement délétères ségrégent en faible fréquence) et correspond à un D négatif. A l'inverse, un excès de variants intermédiaires est généralement dûà de la sélection balancée (qui maintient les allèles à des fréquences intermédiaires) en l'absence de recombinaison et se traduira par un D positif. D suit approximativement une loi de distribution bêta. Cependant, plusieurs conditions doivent être vérifiées : (1) les séquences d'ADN doivent être échantillonnées au hasard dans la population, (2) il n'y a pas de phénomène d'autostop et surtout, (3) la population a atteint l'équilibre mutation / dérive depuis une longue période évolutive. En effet, ce test ne permet pas de faire la différence

entre la sélection darwinienne et l'existence de changements démographiques passés. Ainsi, par exemple, un D de Tajima négatif peut être dû à la sélection mais également à un goulot d'étranglement de la population suivi d'une phase d'expansion. De plus, ce test s'avère peu robuste pour un faible nombre de séquences échantillonnées. Par conséquent, il faut rester prudent lors de l'interprétation des résultats. Fu et Li (1993) proposent une amélioration du test de Tajima. en estimant la diversité nucléotidique uniquement à partir des mutations récemment apparues (singletons présents sur une seule séquence). Ce dernier test apparaît donc beaucoup plus sensible aux mutations artéfactuelles. La comparaison de ces indices sur un gène pour plusieurs espèces et/ou plusieurs gènes d'une même espèce partageant la même zone géographique, telle que l'EPR, peut néanmoins favoriser l'hypothèse sélective ou démographique. En effet, un événement démographique provoqué par une modification de l'habitat (e.g. épanchement de lave détruisant la faune) doit théoriquement affecter l'ensemble des gènes d'une espèce et la majorité des espèces (Hudson et al 1987, Galtier et al 2000). Au contraire, la sélection n'affectera qu'un faible nombre de gènes et d'espèces. En conséquence, un D de Tajima négatif sur une même région géographique pour l'ensemble des espèces et des gènes posera l'hypothèse forte d'une expansion démographique (goulot d'étranglement réduisant la diversité, suivi d'une accumulation de nouvelles mutations : excès de variants rares). Cette hypothèse sera d'autant plus forte que l'expansion supposée est simultanée pour l'ensemble de ces espèces.

10.4. Dater une expansion démographique via le logiciel Fluctuate

Le logiciel Fluctuate (Kuhner et al 1998) permet de dater le début de l'expansion démographique d'une population à partir de la structure généalogique d'un coalescent. En effet, en absence de sélection, une population en expansion doit présenter un coalescent en étoile, avec une séquence majoritaire ancestrale et de multiple séquences dérivées faiblement divergentes de cette dernière. Le logiciel consiste à échantillonner un grand nombre de généalogies neutres dans l'univers des possibles (par la méthode de Metropolis-Hasting utilisant des chaînes de Markov sous procédure Monte Carlo, MCMC, Metropolis et al 1953), à partir des données de séquences fournies, une valeur initiale du paramètre populationnel θ (=2 Ne μ pour un gène mitochondrial et 4 Ne μ pour un gène nucléaire, où Ne correspond à la taille efficace de la population et μ au taux de mutations) et du

paramètre de croissance exponentielle g. Une probabilité postérieure est ensuite attribuée à chacune de ces généalogies (Griffiths & Tavare 1994a, 1994b), la meilleure d'entre elles permettant de réévaluer conjointement les valeurs de θ et g.

Cette approche suppose l'absence de sélection sur le locus considéré, l'absence de recombinaison intra-locus, que la population soit fermée (pas de migration) et une expansion exponentielle de la population au cours du temps (t) à partir d'une population ancestrale de taille efficace Ne_a (Kuhner et al 1998). Elle fait également appel au taux de mutation du locus considéré en faisant l'hypothèse que ce taux de mutation est constant au cours du temps. La relation $\theta_t = \theta_a e^{-g\mu t}$ (où θ_t et θ_a correspondent à θ au temps t et à la population ancestrale, respectivement, Kuhner et al 1998) permet ainsi de corréler la taille de la population au temps et ainsi de déterminer le début de l'expansion démographique. Cette approche peut également être réalisée pour dater le début d'un déclin démographique (g est alors négatif).

10.5. Proposition d'un nouveau test statistique visant à détecter des locus sous sélection diversifiante entre régions géographiques

Nous proposons un nouveau test multi-locus visant à détecter des locus sous sélection diversifiante entre régions géographiques différentes. Ce test est basé sur la différence entre les thêtas de Watterson $\Delta\theta_w$ calculés dans des lignées présentes dans des régions géographiques différentes (e.g. Nord et Sud). Les $\Delta\theta_w$ (= θ_w Nord - θ_w Sud) observées sont ainsi calculées pour chaque gène et doivent être distribuées selon une loi normale sous l'hypothèse d'un modèle neutre d'isolement de ces régions géographiques. Les gènes présentant une valeur de $\Delta\theta_w$ en marge de la distribution normale seraient ainsi de bons candidats pour détecter de la sélection diversifiante entre les régions géographiques considérées. Ce test peut être considéré indépendant des taux de mutation de chaque locus car il est supposé que les gènes considérés présentent le même taux de mutation dans ces deux régions. Le logiciel MsBayes (Hickerson et al 2006b) peut être utilisé afin de simuler les valeurs de θ_w dans chaque paire de lignées sous modèle neutre (nécessité de préciser l'absence de goulot d'étranglement dans MsBayes) et ainsi d'obtenir la distribution des probabilités postérieures des $\Delta\theta_w$ utilisée pour détecter les $\Delta\theta_w$ en marge de la distribution.

Ce test a été utilisé pour les espèces *A. pompejana* et *B. thermophilus* mais n'a pas permis de détecter de locus « outlier », la totalité des locus étudiés étant situés en milieu de distribution. Il serait néanmoins intéressant de tester la puissance statistique de ce test en l'utilisant sur des locus connus comme étant soumis à sélection diversifiante entre régions géographiques différentes.

10.6. Discrimination entre balayage sélectif et goulot d'étranglement via une approche de vraisemblance sur données multilocus : le logiciel Sweep_bott (Galtier et al 2000)

Afin de discriminer l'effet d'une sélection positive (ou, par autostop génétique, de balayage sélectif) d'un goulot d'étranglement de la population, une méthode basée sur des estimations de maximum de vraisemblance de coalescents à plusieurs locus sous l'hypothèse d'un modèle de réduction de la diversité génétique (logiciel Sweep_bott, Galtier et al 2000) a également été utilisée lors des approches multigènes. L'idée est de considérer qu'un goulot d'étranglement se traduira par une réduction de polymorphisme à un temps T et avec une force S, ces paramètres étant identiques entre tous les locus échantillonnés. À l'inverse, le temps et la force d'un balayage sélectif seront différents voire nuls selon le locus étudié. Cette méthode faisant l'hypothèse d'un modèle en nombre de sites infini (une mutation ne pouvant survenir qu'une seule fois sur un site donné), les sites s'écartant de ce modèle (i.e. site à plus de deux états de caractère, les séquences présentant des sites homoplasiques ayant déjà été éliminées pour obtenir un jeu de qualité 4 à R_m nul) sont supprimés pour les analyses utilisant sweep_bott. Nécessitant un codage binaire des séquences avec 0 représentant l'état ancestral du site et 1 l'état dérivé, les états de caractère les plus fréquemment échantillonnés pour chaque site donné sont considérés ancestral (le second état de caractère échantillonné étant ainsi dérivé). Trois modèles sont définis:

(1) Modèle M1 : modèle neutre (absence d'événement démographique et sélectif), comptant p paramètres correspondant aux valeurs de θ pour chacun des p locus

(2) Modèle M2 : modèle « goulot d'étranglement » (bottleneck). Il compte p+2 paramètres : p valeurs de θ, S et T. Pour un événement démographique, S et T sont les mêmes pour l'ensemble des locus.

(3) Modèle M3 : modèle de « balayage sélectif » (selective sweep). Il compte 3p paramètres. En effet, chacun des p locus compte une valeur de θ, un S et un T différents.

Pour chaque modèle, une valeur de maximum de vraisemblance est calculée (selon Griffiths & Tavare 1994a, 1994b). Un test de rapport de vraisemblance (likelihood ratio test, LRT = $2(L_2-L_1)$ distribué selon une loi du χ^2, où L_1 et L_2 correspondent aux maximum de vraisemblance des modèles considérés) effectué entre chaque paire de modèles permet ainsi de définir le modèle le plus probable. Si le modèle M2 est favorisé (significativement différent de M1), alors l'ensemble des locus doit présenter un écart à la neutralité (LRT par locus significatif). Au contraire, si le modèle M3 est favorisé (significativement différent de M1 et M2), les locus sous balayage sélectif doivent montrer un écart à la neutralité, séparément. Dans le cas où, un locus sous balayage sélectif s'additionne à un goulot d'étranglement, on s'attend à ce que : (1) les modèles M2 et M3 soient tous les deux significativement différents du modèle M1, (2) le modèle M3 soit favorisé vis-à-vis du modèle M2 (au moins un locus présentant un T et un S différents des autres locus), (3) tous les locus soient significativement différents du modèle M1 mais, que le locus (ou les locus) sous balayage sélectif présentent un LRT plus fort que le LRT des locus non soumis à balayage.

Chapitre 3 : Phylogéographie comparée des espèces hydrothermales de la dorsale du Pacifique oriental à l'aide du gène mitochondrial de la Cytochrome Oxydase I

1. *Introduction à l'approche de phylogéographie comparée des espèces*

Jusqu'à aujourd'hui, la plupart des études phylogéographiques ont été menées sur une espèce à la fois et visaient à mettre en évidence les processus ayant conduit à une différenciation génétique avec monophylie réciproque (divergence fixée entre lignées soeurs). Néanmoins, si la monophylie réciproque (Fig. 3.1) entre lignées constitue un cas extrême de différenciation génétique, pouvant mener à la spéciation, le processus d'isolement génétique est progressif. Les généalogies de gènes vont ainsi passer par une phase de tri des lignées allant de la polyphylie (absence d'isolement génétique) à de la paraphylie (isolement génétique incomplet) avant de présenter, potentiellement, une monophylie réciproque (Nosil et al 2009, Fig. 3.1).

Pour qu'il y ait monophylie réciproque, le polymorphisme ancestral doit être éliminé par dérive génétique et/ou sélection naturelle. Le temps nécessaire pour atteindre ce stade de différenciation dépendra donc de la taille efficace des populations (Avise 2000). Ainsi, la probabilité d'observer une monophylie réciproque lors d'un isolement de deux populations par une barrière à la dispersion, augmente au cours du temps et s'avère d'autant plus forte que la taille efficace de la population est petite (Fig. 3.2).

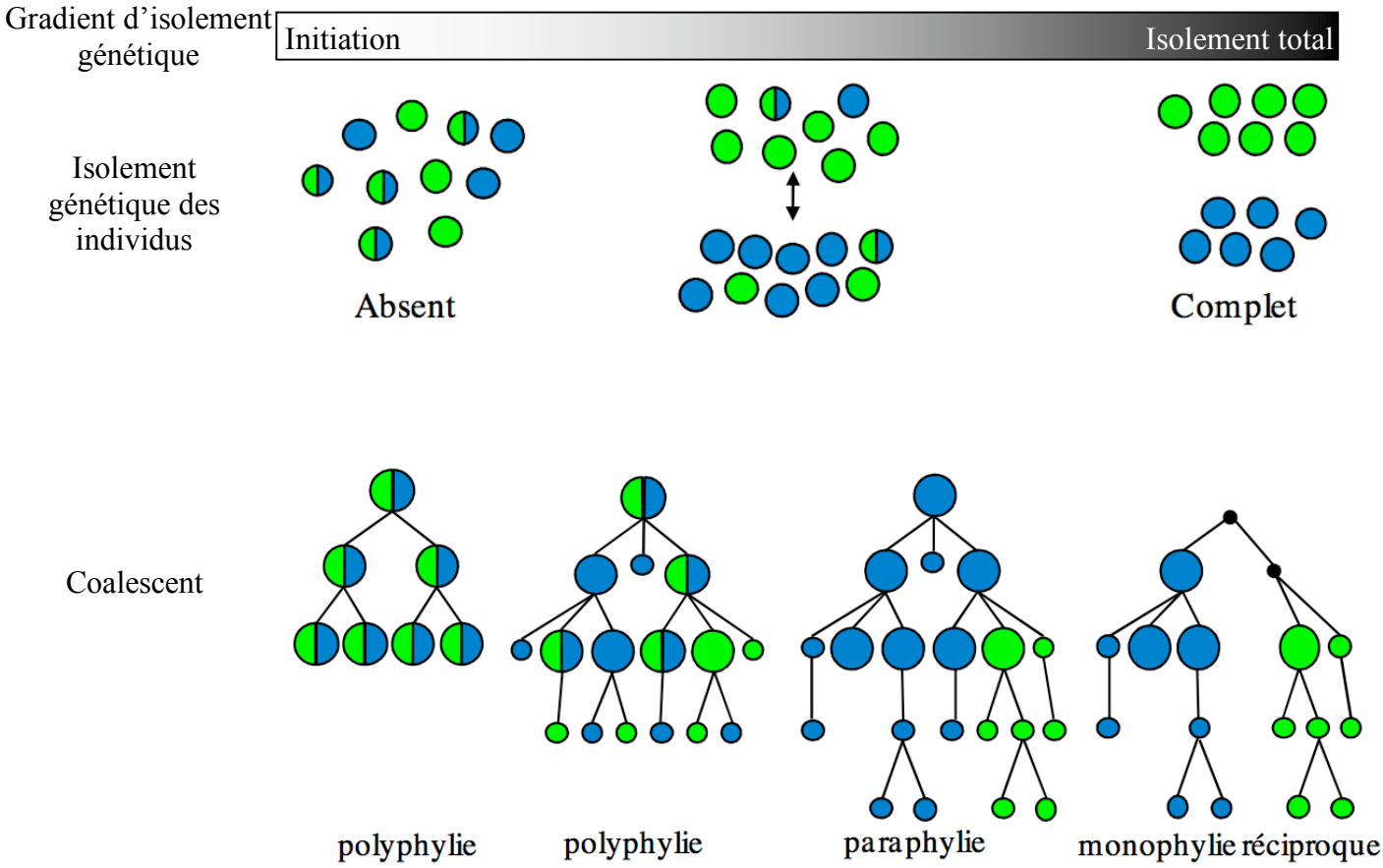


Fig. 3.1 Etapes de mise en place d'un isolement génétique complet et des coalescents théoriques associés à ces étapes. Adaptée de Nosil et al (2009).
Les individus sont représentés par des cercles, la couleur indique le génotype. Les individus bicolores correspondent à des individus hybrides.

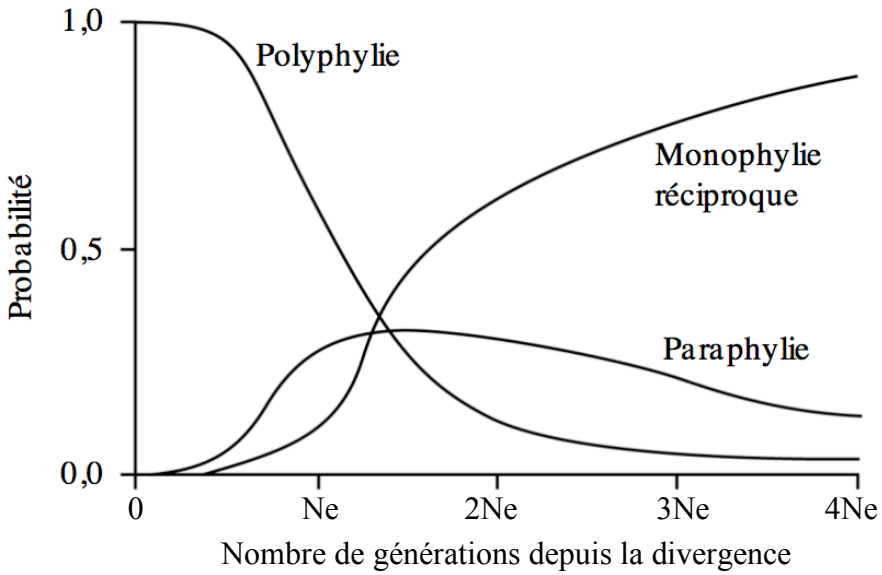


Fig. 3.2 Probabilité d'obtenir un coalescent polyphylétique, paraphylétique ou en monophylie réciproque en fonction du temps écoulé depuis le début de la divergence en allopatrie de deux populations. Adaptée de Neigel & Avise (1986)
La taille efficace *Ne* est supposée identique pour les deux populations et constante au cours du temps

De plus, la monophylie réciproque n'est pas nécessairement immuable. En effet, si l'isolement reproducteur n'est pas complet (spéciation non atteinte), une remise en contact secondaire entre lignées géographiquement isolées, par exemple, peut engendrer de l'hybridation entre ces lignées, donnant ainsi lieu à un coalescent para- voire polyphylétique. Ainsi, un coalescent para/poly-phylétique peut être dû soit à une rétention de polymorphisme ancestrale (Fig. 3.3.A), soit à une remise en contact secondaire après un relâchement de la barrière aux flux de gène (Fig. 3.3.B). Dans les 2 cas, cela se traduit par un mélange d'allèles au niveau des 2 lignées dans les populations échantillonnées mais dans lequel la coalescence de l'allèle « introduit » est généralement profonde s'il est associé à un polymorphisme ancestral non trié, ou terminale si le contact secondaire est récent.

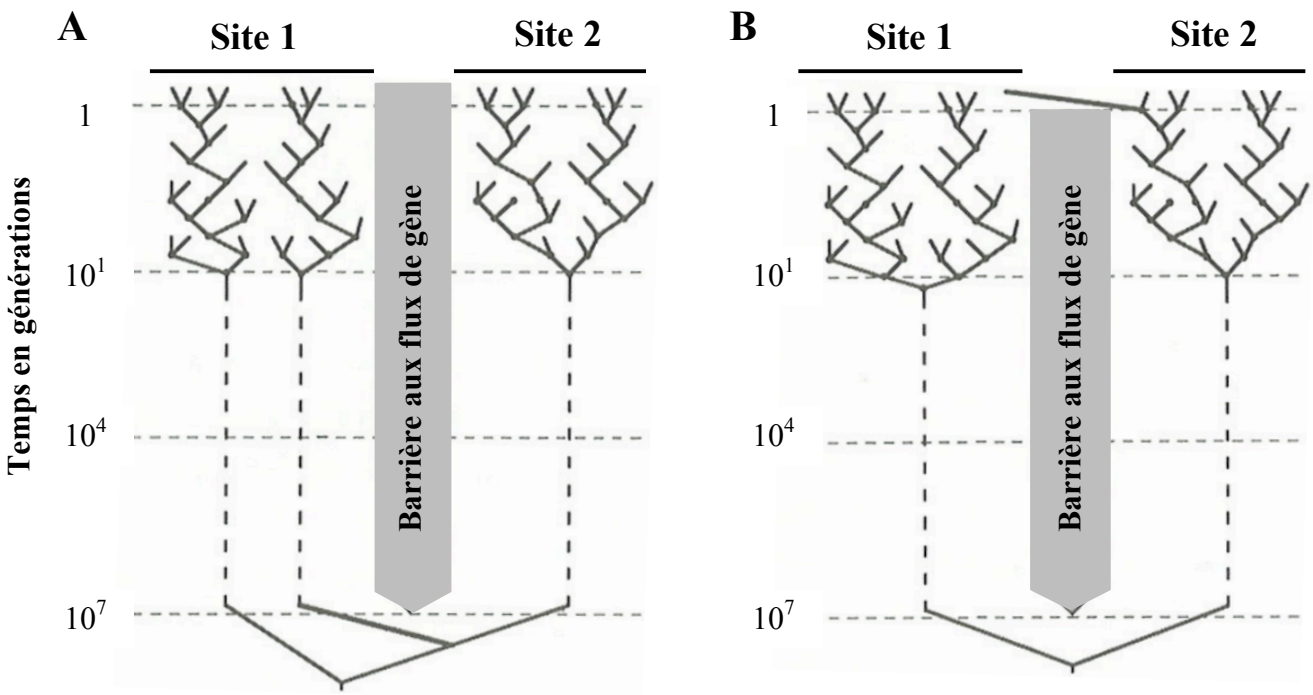


Fig. 3.3 Deux scénarios alternatifs pour expliquer un coalescent para/poly-phylétique entre deux sites géographiquement distincts. Adaptée de Avise (2000)
A, rétention de polymorphisme ancestral ; B, remise en contact secondaire.

Avise (1987) suppose que les capacités de dispersion jouent un rôle majeur dans la structure géographique des lignées. Chez les invertébrés marins, la présence d'une phase larvaire planctonique, sa durée de vie et la possibilité de migration des individus adultes sont autant de traits d'histoire de vie qui favorisent la dispersion. Ils interagissent avec les

contraintes environnementales et peuvent ainsi permettre ou non de franchir des barrières à la dispersion. Généralement, plus les traits d'histoire de vie favorisent la dispersion et plus la structure génétique observée est faible (Palumbi 1994 ; Féral 2002 ; Kinlan & Gaines 2003). L'analyse de séquences du gène mitochondrial Cytochrome Oxydase I sur plusieurs espèces de crépidules (gastéropodes invasifs) le long de la côte est de l'Amérique du Nord, a montré que les espèces à développement direct présentaient une structure génétique plus marquée que celles à développement larvaire planctotrophe (Collin 2001). Wilke & Davis (2000) considèrent que la différenciation génétique observée chez le gastéropode *Hydrobia ventrosa* (développement direct) est plus élevée que chez *H. ulvae* (1-3 jours de stade larvaire planctonique), et pourrait être liée à des différences d'habitat entre les populations d'adultes plutôt qu'au mode de développement lui-même. Rolán-Alvarez et al (1995) suggèrent, quant à eux, que les différences de structure génétique entre deux espèces de littorines (*Littorina mariae* et *L. obtusata*) à développement direct seraient dues à un temps de génération différent entre ces espèces. De nombreux critères (cycle de vie, morphologie, physiologie, comportement, etc) influençant les capacités de dispersion des espèces, mais également leur potentiel d'adaptation à un habitat donné, entrent ainsi en jeu et affectent la topologie des coalescents. Lors d'une étude monospécifique, l'absence de monophylie réciproque entre populations issues de sites géographiquement distincts pourrait alors signifier :

(1) l'absence de barrière à la dispersion entre ces sites

(2) l'existence d'une barrière à la dispersion mais qui ne constituerait pas un obstacle pour celle-ci, présentant potentiellement de meilleures capacités de dispersion que les autres espèces plus structurées

(3) l'existence d'une barrière ayant anciennement affecté cette espèce mais plus récemment perméable aux flux de gènes après la remise en contact secondaire des lignées

Afin de mieux connaître les causes de la distribution géographique des lignées intraspécifiques d'une espèce, il est intéressant de comparer les patrons phylogéographiques de cette espèce à ceux d'autres espèces présentant une distribution géographique similaire. Ainsi, les concordances/discordances phylogéographiques entre espèces informeront sur l'impact d'une barrière aux flux de gènes et donc sur l'intensité ou sur l'âge de la barrière responsable de la structure géographique observée (Zink 2002). Cette approche comparative apporte ainsi des informations précieuses sur les facteurs ayant façonné la distribution géographique des lignées, et peut permettre de déduire la nature de cette barrière (e.g. géologique, courantologie, distance à parcourir…). De plus, la phylogéographie comparée entre espèces peut également, dans certains cas, permettre de déterminer les traits d'histoire de vie spécifiquement responsable des dissemblances de

comportement des espèces face à cette barrière (Bermingham & Avise 1986, Bermingham 1998). De plus, l'analyse conjointe de plusieurs taxons issus de régions ayant des histoires communes telles que celles associées à la fermeture de l'Isthme de Panama ou de l'ouverture du détroit de Béring peut fournir des informations précieuses sur les logiques d'extinction totale ou partielle de certaines lignées voire d'inférer des scénarios sur la recolonisation d'une région/environnement par la lignée ayant survécu dans l'autre région/environnement (Cunningham & Collins 1998).

Dans ce chapitre, nous avons effectué une approche de phylogéographie comparée de sept espèces hydrothermales profondes de la dorsale du Pacifique oriental (EPR) en utilisant le gène mitochondrial de la Cytochrome Oxydase I (mtCOI). Ce chapitre correspond à un article publié dans la revue Molecular Ecology, présenté après un résumé en français. Il a pour objectif de tester l'hypothèse d'une barrière aux flux de gènes au niveau de la zone équatoriale de l'EPR pouvant avoir donné lieu à un événement vicariant majeur ayant séparé les faunes hydrothermales du Pacifique est. Cette étude de ce fait pourrait permettre de valider l'hypothèse de 2 sous-unités biogéographiques proposées par Bachraty et al (2009) et éventuellement de dater cette séparation. Le choix de cette zone a été guidée par les travaux d'Hurtado et al (2004) qui soulignaient la possibilité d'une barrière au niveau du point triple des Galapagos chez les polychètes *Alvinella pompejana* et *Tevnia jerichonana* sur la base de monophilies réciproques de part et d'autre de cette barrière.

2. *Résumé de l'article de phylogéographie comparée entre espèces*

Pour effectuer notre approche de phylogéographie comparée entre espèces, nous avons choisi d'utiliser des espèces appartenant à des familles différentes (gastéropodes : *Lepetodrilus elevatus*, *L. ovalis*, *Eulepetopsis vitrea*, polychètes : *Alvinella pompejana*, *Hesiolyra bergi*, *Branchipolynoe symmytilida*, et un bivalve : *Bathymodiolus thermophilus*, Fig. 3.4) afin d'obtenir une vue d'ensemble des distributions phylogéographiques d'espèces présentant des traits d'histoire de vie contrastés. Ces espèces sont relativement abondantes sur l'EPR, permettant l'échantillonnage nécessaire à une étude de populations, et endémiques de l'EPR. Ces espèces ont été échantillonnées sur neuf champs

hydrothermaux répartis de 21°N à 21°S. Deux à trois cents individus par espèce ont été séquencés sur un fragment du gène mtCOI.

Malgré les disparités biologiques entre ces espèces, les réseaux d'haplotypes du gène mtCOI (Fig. 3.5) révèlent des topologies similaires avec deux clades divergents, à l'exception du polychète commensal des moules hydrothermales, *B. symmytilida*. Cette dernière espèce montre, en effet, un réseau plus complexe suggérant une diversification plus ancienne et l'absence de co-évolution entre ce polychète et son hôte.

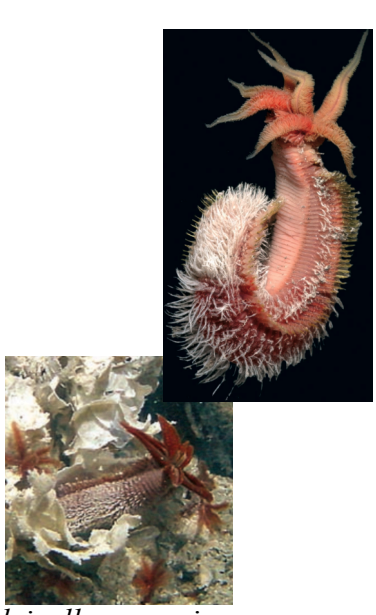
Le polychaete *A. pompejana* et le gastéropode *E. vitrea* présentent exactement la même architecture de réseau d'haplotypes (Fig. 3.5), avec deux clades distints (divergence de 1% pour *A. pompejana* et 0,9% pour E. vitrea) séparés de part et d'autre de l'Equateur.

L. elevatus compte deux clades A et B divergents de 6,5% correspondant aux deux espèces cryptiques définies par Matabos et al (2008b) distribués respectivement de 21°N à 9°50'N (clade A) et de 9°50'N à 21°S (clade B). Quelques individus du clade B ayant été également échantillonnés à 13°N. La population de 21°N est composée uniquement de l'haplotype 1. La population de 9°50'N diffèrent légèrement des autres populations du Sud par la présence de l'haplotype 4 trouvé en plus forte fréquence sur ce site, expliquant la significativité du ϕ_{st} calculé au sein du clade B entre cette population et celles situées plus au Sud.

Les deux clades détectés chez *L. ovalis* (2% de divergence) forment un cline de 14°S à 21°S, alors que les champs hydrothermaux 9°50'N et 7°25'S ne sont composés que d'haplotypes du clade A. Le bivalve *B. thermophilus* montre, à l'inverse, un cline de ces clades divergents au Nord (13°N à 7°25'S) et des séquences du clade B uniquement au Sud (14°S à 21°S).

H. bergi montre un réseau d'haplotype en étoile avec un haplotype central distribué sur l'ensemble de l'EPR et quelques individus du Sud présentant une divergence de 1%.
La distribution géographique des haplotypes met ainsi en exergue la présence d'une zone de transition entre le nord et le sud de la dorsale. Cette zone de transition est confirmée par l'augmentation des valeurs de différenciation génétique (ϕ_{st}) entre 17°S et l'équateur.

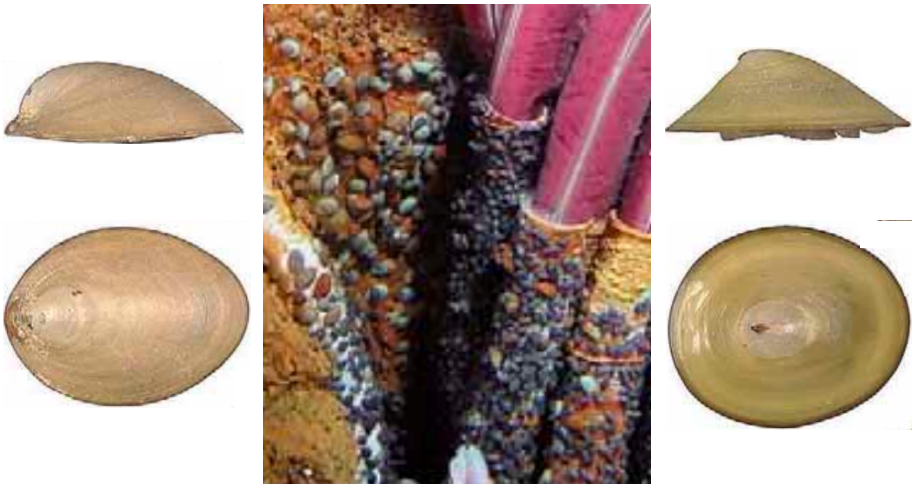
Bathymodiolus thermophilus
(bivalve formant de véritables moulières)



Alvinella pompejana
(polychète tubicole)

Hesiolyra bergi
(polychète libre, proche des tubes d'*A. pompejana*)

Branchipolynoe symmytilida
(polychète commensal des *Bathymodiolus*)



Lepetodrilus elevatus (à gauche) et *L. ovalis* (à droite), gastéropodes abondants sur des roches, *Riftia pachyptila* (au centre) et *Bathymodiolus*

Eulepetopsis vitrea
(gastéropode associé au pôle froid des sources hydrothermales)

Fig. 3.4 Photographies des espèces étudiées par l'approche de phylogéographie comparée sur le mtCOI. Tirées de Desbruyères et al (2006b)

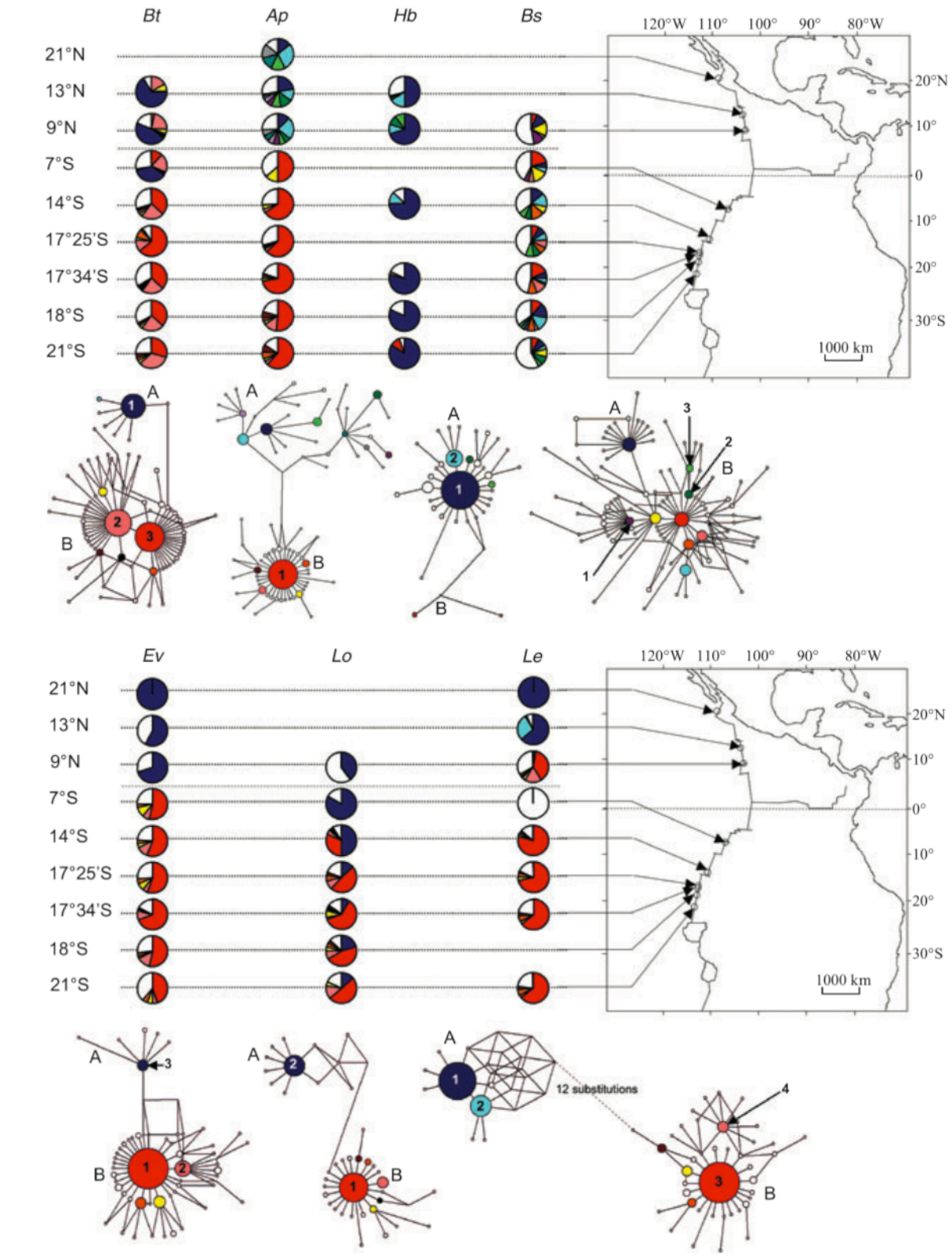


Fig. 3.5 Réseaux d'haplotypes en Median Joining et distributions des fréquences haplotypiques du gène mtCOI pour les sept espèces étudiées.

Sur les réseaux, la taille des cercles d'haplotypes et les connexions entre haplotypes sont proportionnelles au nombre de séquences et de pas mutationnels, respectivement. A et B représentent les deux clades divergents pour chaque espèce.

Sur les distributions d'haplotypes, les haplotypes partagés ayant une fréquence > 2% au sein du clade considéré sont colorés. Les haplotypes privés et haplotypes partagés en fréquence < 2% au sein du clade considéré sont représentés en blanc. *Bt*, *Bathymodiolus thermophilus* ; *Ap*, *Alvinella pompejana* ; *Hb*, *Hesiolyra bergi* ; *Bs*, *Branchipolynoe symmytilida* ; *Ev*, *Eulepetopsis vitrea* ; *Lo*, *Lepetodrilus ovalis* ; *Le, Lepetodrilus elevatus*.

L'hypothèse d'horloge moléculaire ayant été testée et validée sur les jeux de données, nous avons ensuite cherché à déterminer si les divergences observées entre les deux clades de chaque espèce (pouvant correspondre aux lignées ayant divergé de part et d'autre d'une même barrière entre nord et sud EPR) étaient issues d'un même événement de vicariance. Pour cela, nous avons utilisé une nouvelle approche basée sur une méthode d' « Approximate Bayesian Computation » (méthode dite ABC) à l'aide du logiciel MsBayes (Hickerson et al 2007, voir le chapitre 2 pour plus d'explication). Cette analyse a permis de mettre en évidence la présence de deux évènements de divergence datés respectivement à 11,6 Ma pour *L. elevatus* uniquement et, correspondant aux deux espèces cryptiques découvertes par Matabos et al (2008b), et à 1,3 Ma pour les six autres espèces. L'observation de la divergence entre les haplotypes 3 (sud EPR) et 4 (principalement à 9°50'N) chez l'espèce cryptique *L. elevatus* majoritairement trouvée de 9°50'N à 21°33'S, pourrait correspondre également à l'événement de divergence daté à 1,3 Ma. Ainsi, *L. elevatus* aurait à la fois subi un isolement allopatrique ayant donné lieu aux deux espèces cryptiques et à la zone de contact à 9°50'N, mais également un isolement concomitant à celui des autres espèces à 1,3 Ma entre 9°50'N et le sud de l'EPR (haplotype 4 plus fréquent à 9°50'N).

Il y a 12,5-11 Ma, la Mathematician Ridge, située au Nord, a subi d'importantes réorientations magnétiques et bathymétriques, provoquant la formation d'un système de dorsales parallèles (microplate Mathematician) avec la Mathematician Ridge à l'ouest et la Moctezuma Trough à l'est (Mammerickx & Kligtgord 1982, voir chapitre 1). Ce n'est qu'environ 4-5 Ma plus tard (7-5 Ma) que la Moctezuma Trough (survivante des 2 dorsales après fossilisation de la Mathematician Ridge) et l'oEPR aurait fusionné pour former l'EPR actuelle. Cette réorganisation de la dorsale au Nord avec la formation temporaire de la microplaque Mathematician pourrait avoir été à l'origine de la spéciation chez *L. elevatus,* il y a 11,6 Ma. Concernant le deuxième événement de vicariance ayant affecté l'ensemble des espèces aux alentours de l'équateur, il pourrait être lié à la formation de plusieurs failles transformantes entre 9°50'N et 17°25'S (i.e. système de fracture Quebrada/Discovery/Gofar, ou failles transformantes Wilkes et Garrett) qui ont fracturé l'axe de la dorsale en multiples fragments il y a 1-2 Ma (Kureth & Rea 1981, Naar & Hey 1989, Francheteau et al 1990). Cette barrière aurait pu être renforcée par une discontinuité des courants profonds dans cette zone (cf Reid 1997, Hurtado et al 2004).

La formation lente et progressive d'une barrière tectonique associée à des caractéristiques de dispersion et/ou d'adaptation contrastés chez les espèces étudiées, pourraient expliquer que la barrière aux flux de gènes se situe à des positions géographiques légèrement différentes entre espèces (Knowlton & Weigt 1998). Il est extrêmement difficile de détecter un critère biologique et/ou écologique unique pouvant expliquer ces différences. *A. pompejana* et *H. bergi* partagent le même habitat (proche du fluide, *H. bergi* se déplaçant autour voire dans les tubes d'*A. pompejana*) et, présentent néanmoins des distributions d'haplotypes très contrastées. Ces distributions sont marquées par une barrière hermétique à l'équateur pour *A. pompejana*, et par la survivance d'une lignée sud divergente d'*H. bergi* très rarement échantillonnée au-delà de 17°25S. Le microenvironnement seul ne semble donc pas jouer de rôle dans la différence de distribution de ces espèces. Les caractères dispersifs des espèces hydrothermales restent relativement peu connus mais ne semblent pas expliquer non plus ces différences. La taille des œufs, leur contenu en vitellus ou le comportement des larves peuvent influencer la migration verticale des propagules (Mullineaux et al 2005, Arellano & Young 2009). La force et le sens des courants océaniques peuvent varier en fonction de la profondeur. Par conséquent, la migration verticale des propagules entre espèces pourrait jouer un rôle dans les dissemblances observées dans le franchissement de la barrière. Cela pourrait également expliquer les résultats obtenus à l'aide du logiciel Migrate 3.0.3 (Beerli & Felsenstein 1999, 2001), avec *H. bergi* présentant une migration orientée de la lignée mitochondriale la plus fréquente du Nord vers le Sud alors que celles de *L. elevatus*, *L. ovalis*, *B. thermophilus* et *B. symmytilida* auraient plutôt migré du Sud vers le Nord. Des études complémentaires sur les capacités de dispersion des espèces pourraient permettre de mieux comprendre ces dissemblances.

La forme en étoile des réseaux d'haplotypes ainsi que les D de Tajima significativement négatifs au Sud de l'EPR pour l'ensemble des espèces étudiées, suggèrent une expansion démographique au Sud de la dorsale. Cette expansion démographique serait concomitante entre toutes les espèces (hormis *B. symmytilida*) et datée de moins de 500 000 ans. Cette expansion démographique pourrait être due (1) à une colonisation récente de l'EPR Sud par des effets fondateurs, en provenance, par exemple, de sites situés au Sud de la microplaque de Pâques (23°S), ou (2) à la fréquence plus élevée d'éruptions massives avec ensevelissement de la faune au Sud de l'EPR (fréquentes dans cette région, e.g. Haymon et al 1993, Cowen et al 2007). Ces dernières auraient alors détruit massivement la faune et engendré des goulots d'étranglement plus aigus.

3. *Article : phylogéographie comparée entre espèces sur le gène mtCOI*

MOLECULAR ECOLOGY

Molecular Ecology (2009) doi: 10.1111/j.1365-294X.2009.04325.x

Comparative phylogeography among hydrothermal vent species along the East Pacific Rise reveals vicariant processes and population expansion in the South

S. PLOUVIEZ,* T. M. SHANK,† B. FAURE,‡ C. DAGUIN-THIEBAUT,* F. VIARD,* F. H. LALLIER* and D. JOLLIVET*

Laboratoire Adaptation et Diversité en Milieu Marin, Université Pierre et Marie Curie-Paris 6; and CNRS, UMR7144 Station Biologique de Roscoff, Roscoff, France, †Biology Department MS#33, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA, ‡Department of Biology, Pennsylvania State University, University Park, PA 16802, USA

Abstract

The use of sequence polymorphism from individual mitochondrial genes to infer past demography has recently proved controversial because of the recurrence of selective sweeps acting over genes and the need for unlinked multilocus data sets. However, comparative analyses using several species for one gene and/or multiple genes for one species can serve as a test for potential selective effects and clarify our understanding of historical demographic effects. This study compares nucleotide polymorphisms in mitochondrial cytochrome oxidase I across seven deep-sea hydrothermal vent species that live along the volcanically active East Pacific Rise. Approximate Bayesian Computation (ABC) method, developed to trace shared vicariant events across species pairs, indicates the occurrence of two across species divergence times, and suggests that the present geographical patterns of genetic differentiation may be explained by two periods of significant population isolation. The oldest period dates back 11.6 Ma and is associated with the vent limpet *Lepetodrilus elevatus*, while the most recent period of isolation is 1.3 Ma, which apparently affected all species examined and coincides with a transition zone across the equator. Moreover, significant negative Tajima's *D* and star-like networks were observed for all southern lineages, suggesting that these lineages experienced a concomitant demographic and geographical expansion about 100 000–300 000 generations ago. This expansion may have initiated from a wave of range expansions during the secondary colonization of new sites along the Southern East Pacific Rise (founder effects below the equator) or recurrent bottleneck events because of the increase of eruptive phases associated with the higher spreading rates of the ridge in this region.

Keywords: allopatry, Approximate Bayesian Computation, cytochrome oxidase I, deep-sea, genetic isolation, past demography

Received 11 May 2009; revision received 17 July 2009; accepted 20 July 2009

Introduction

Patterns of genetic variation are powerful tools for elucidating population history over time and space. The extent to which current genetic signals reveal past demography mostly depends on the mutational model of the molecular marker used (e.g. microsatellites, introns or exons) and the analytical methods, either allele frequency (allele size, restriction fragment length polymorphism or single nucleotide polymorphisms) or sequence-based approaches (e.g. Sunnucks 2000). Among them, coalescence-based approaches using nucleotide polymorphisms from one or more species/lineages have arisen as a powerful tool for tracing historical demographic events, genetic exchange and

Correspondence: Sophie Plouviez, Fax: +332 988 92324; E-mail: plouviez@sb-roscoff.fr

© 2009 Blackwell Publishing Ltd

population isolation (e.g. Hey & Nielsen 2004). Particularly powerful are coalescent analyses in a geographical context (i.e. phylogeography), with neutral markers lacking recombination. In these cases, gene genealogies and phylogenetic networks (Posada & Crandall 2001) together with DNA sequence diversity indexes can enable the reconstruction of population histories and identification of the possible reduction and/or expansion of populations (e.g. Emerson *et al.* 2001). Comparative approaches using multiple co-distributed species sharing congruent patterns of genetic structure between reciprocally monophyletic populations can reveal major isolating mechanisms or barriers to dispersal (Bermingham & Avise 1986).

Mitochondrial genes are the most widely used genes for phylogeographical analyses in animals (Avise 1998). The absence of recombination because of the maternal inheritance of the mitochondria (but see Skibinski *et al.* 1994), the supposed neutrality, relatively rapid mutation rates makes it sensitive to population subdivision (Avise 1998), and the ease of data generation of this haploid sequence (Folmer *et al.* 1994) provide major advantages to using this gene. The use of these genes has been recently found to be controversial for the inference of population size and past demographic events because of the recurrence of possible selective sweeps, which mimic the reduction of the effective population size following a bottleneck or a founder event (Bazin *et al.* 2006) and the need for unlinked multiloci data sets. However, even if rarely carried out in the same study, comparison across multiple co-living species could help in disentangling demographic events from selective sweeps. For example, Lessa *et al.* (2003) showed concordant evidence for shared demographic expansion in North American mammals by comparing different species using the mitochondrial cytochrome *b* gene only and Smith & Farrell (2005) revealed concomitant range expansions in *Moneilema gigas* and *Moneilema armatum* beetles using the cytochrome oxidase I (mtCOI) gene.

Deep-sea hydrothermal vents on intermediate- to fast-spreading mid-ocean ridges provide opportunities for testing shared vicariant events and concomitant demographic changes as populations form metapopulations subjected to frequent local extinctions and bottlenecks because of the recurrence of eruptive activity (Haymon *et al.* 1991; Tunnicliffe *et al.* 1997; Jollivet *et al.* 1999). Vent habitats are distributed along oceanic ridges, where sulphide-rich fluid emissions are ephemeral both in time (years to a few decades; Shank *et al.* 1998) and space (Jollivet *et al.* 1999). Distances between sites vary from tens to hundreds of kilometres and can play an important role in population differentiation and allopatric speciation. Previous studies have focussed on the role of plate tectonics in favouring allopatry and subsequent secondary contacts across topographic (transform faults) and/or oceanographic (gyres) barriers (O'Mullan *et al.* 2001; Hurtado *et al.* 2004; Young *et al.* 2008; Faure *et al.* 2009). Moreover, at the scale of geological times, vent field (cluster of sites) displacements imposed by the dynamics of the underlying magmatic chamber are likely to favour a succession of isolation phases and secondary contacts (Jollivet *et al.* 1999). Plate tectonics is thus one of the major forces acting on faunal composition, distinguishing seven hydrothermal vent biogeographical provinces (Tunnicliffe 1991; Van Dover *et al.* 2002; Bachraty *et al.* 2009). Differences in the faunistic composition of vent communities at a global scale are, however, difficult to reconcile with the origin of speciation processes, which in turn need to be assessed at a more restricted spatial scale (i.e. the ridge scale). At this level of spatial organization, well-dated topographic discontinuities that offset the ridge (like transform faults, overlapping spreading centres or microplates) are largely responsible for bottom current disruption, favouring breaks in gene flow that may contribute to vicariance affecting nearly the entire regional species pool. In this geological context, the concomitant isolation of taxa may be followed or not by secondary contacts and population admixtures, the extent of a contact zone and its age mostly depending on both the life-history traits of each species and the level of habitat connectivity. Comparative phylogeographical analyses are thus helpful to identify potential sources of vicariance shared between species. Sequence-based methods based on coalescence theory are then necessary to detect the major ecological/tectonic events that might have influenced most co-evolving species population dynamics and subsequently blurred the phylogenetic information (i.e. divergence time).

Based on faunal composition, the East Pacific Rise (EPR) has been recently subdivided into two different biogeographical provinces one on each side of the equator (Bachraty *et al.* 2009), suggesting the presence of a potential barrier to gene flow at this latitude. The equatorial barrier was, however, established on faunal differences between the two communities based on the presence/absence of species. This study aims at examining whether this hypothesized biogeographical break gave rise to distinct genetic divisions among vent species that co-occur in both Northern and Southern EPR. Indeed, Burton (1998) showed that the abrupt change in the faunal composition of near-shore marine communities at Point Conception (California) did not always correlate with the genetic isolation of its species components. If the biogeographical split coincides with a species genetic break, two additional questions must be addressed: (i) is the position of the break at the same geographical location for all species? and (ii) is the timing

© 2009 Blackwell Publishing Ltd

of separation consistent among taxa and thus represent a 'true' vicariant event between communities? Differences in species' dispersal capabilities may indeed allow them to track spreading vent fields differentially after the time of isolation. Finally, the study also aims to examine hypotheses regarding the possible stepwise population expansion in a given direction along the EPR. Gene flow asymmetries across the barrier would provide useful information about the history of colonization along ridges, particularly if taxa display similar patterns of directional gene flow. Therefore, comparative phylogeographical analysis of mtCOI sequences from seven broadly distributed species from three different taxonomic groups (three gastropod limpets, three polychaetes and one bivalve species) was conducted to test for shared vicariant events and demographic histories among the vent fauna.

Materials and methods

Biological specimens and molecular methods

Specimens from seven morphologically well-described species (*Bathymodiolus thermophilus*, *Alvinella pompejana*, *Hesiolyra bergi*, *Branchipolynoe symmytilida*, *Eulepetopsis vitrea*, *Lepetodrilus ovalis* and *Lepetodrilus elevatus*) were collected along the EPR during either the *Nautile* or DSV *Alvin* submersible expeditions. Nine hydrothermal vent localities were sampled in 1999, 2002, 2003 and 2004 (Table 1). The primary biological and habitat characteristics of the seven sampled species are summarized in Table 2. Animals were collected from sulphide chimneys, mussel beds and vestimentiferan assemblages by grabbing the fauna with the submersible manipulator arm and storing them in an insulated

Table 1 Location, size of faunistic samples, number of haplotypes and admixture between clades at population level along the East Pacific Rise for the seven species

Vent field	Geographical position	Depth (m)		Species						
				Bt	*Ap*	*Hb*	*Bs*	*Ev*	*Lo*	*Le*
21°N	20°49′N	2606	*N*	Abs	7 (7)	0	Abs	1	0	37
	109°06′W		*h* (H_d)		6 (0.952)			1		1 (0.000)
			Adm		7/0			1/0		37/0
13°N	12°43–50′N	2560–2700	*N*	12 (12)	72 (5)	153	0	7	0	150
	103°53–57′W		*h* (H_d)	4 (0.561)	26 (0.911)	23 (0.720)		3 (0.667)		13 (0.523)
			Adm	8/4	72/0	153/0		7/0		147/3
9°N	9°31–51′N	2500–2585	*N*	59 (12)	31 (21)	10	39 (8)	10	15	63
	104°15–18′W		*h* (H_d)	16 (0.729)	16 (0.929)	4 (0.533)	27 (0.968)	4 (0.533)	12 (0.962)	25 (0.828)
			Adm	32/25	30/0	10/0	9/30	10/0	15/0	2/61
7°S	7°25′S	2735–2752	*N*	58 (12)	14	0	20	27	6	1
	107°47–49W		*h* (H_d)	21 (0.837)	7 (0.758)		15 (0.953)	9 (0.687)	2 (0.333)	1
			Adm	22/38	0/14		3/17	0/27	6/0	0/1
14°S	13°59′S	2623–2632	*N*	30	23 (3)	8	17	48	22	44
	112°29′W		*h* (H_d)	14 (0.825)	8 (0.581)	3 (0.464)	13 (0.971)	11 (0.647)	7 (0.714)	8 (0.368)
			Adm	0/30	0/23	8/0	6/11	0/48	11/11	0/44
17°25′S	17°25′S	2578–2590	*N*	33 (12)	51 (21)	0	29	72	52	49
	113°12′W		*h* (H_d)	8 (0.585)	19 (0.610)		22 (0.982)	17 (0.725)	18 (0.749)	10 (0.489)
			Adm	0/33	0/51		4/25	0/72	7/45	0/49
17°35′S	17°35–36′S	2591–2597	*N*	60	36	34	49	51	46	64
	113°15′W		*h* (H_d)	24 (0.817)	10 (0.483)	7 (0.373)	30 (0.957)	15 (0.503)	13 (0.650)	17 (0.588)
			Adm	2/58	0/36	33/1	7/42	0/51	5/41	0/64
18°S	18°26–37′S	2636–2680	*N*	30	29	11	29	47	24	0
	113°23–24′W		*h* (H_d)	14 (0.825)	11 (0.724)	3 (0.345)	20 (0.961)	13 (0.677)	8 (0.757)	
			Adm	0/30	0/29	11/0	6/23	0/47	5/19	
21°S	21°25–33′S	2804–2840	*N*	27	55	19	24	34	22	31
	114°16–18′W		*h* (H_d)	12 (0.818)	14 (0.613)	3 (0.292)	19 (0.982)	15 (0.779)	9 (0.745)	6 (0.353)
			Adm	0/27	0/55	17/2	2/22	0/34	3/19	0/31

N, number of mtCOI sequences used in the study per sample and species (Abs stands for the absence of the species at the studied site. 0 means that no sample could be collected, although the species was observed at the site), including those obtained from GenBank (in brackets); *h*, number of haplotypes; H_d, haplotype diversity; Adm, number of sequences from clade A/clade B; *Bt*, *Bathymodiolus thermophilus*; *Ap*, *Alvinella pompejana*; *Hb*, *Hesiolyra bergi*; *Bs*, *Branchipolynoe symmytilida*; *Ev*, *Eulepetopsis vitrea*; *Lo*, *Lepetodrilus ovalis*; *Le*, *Lepetodrilus elevatus*.

© 2009 Blackwell Publishing Ltd

Table 2 Biological and ecological characteristics of the seven studied vent invertebrates

Species	Fertilization	Fecundity (oocytes/ female)	Egg size (μm)	Dispersal mode	Reproduction	Habitat	Reference
Bt	External	1 000 000	50	Plantoctrophic larva	Discontinuous	Diffuse venting	Le Pennec et al. (1984); Tunnicliffe (1991); Jollivet (1996)
Ap	Internal	200 000	180	Lecithotrophic larva	Nearly continuous	Vent chimney	Jollivet (1996); Chevaldonné et al. (1997); Faure et al. (2007)
Hb						Vent chimney	Jollivet (1996)
Bs	Internal	1000	400	Lecithotrophic larva	Continuous or frequently intermittent	Palleal cavity of mussels	Van Dover et al. (1999); Jollivet et al. (2000); Plouviez et al. (2008)
Ev	Internal	200	230	Lecithotrophic larva	Continuous	Basaltic rocks and mussel shells	Tyler et al. (2008)
Lo	Internal	1000	90	Lecithotrophic larva	Continuous	Mussel shells	Fretter (1988); Mullineaux et al. (1995); Sadosky et al. (2002); Tyler et al. (2008)
Le	Internal	1000	90	Lecithotrophic larva	Continuous	Vestimentiferan tubes and mussel shells	Fretter (1988); Mullineaux et al. (1995); Sadosky et al. (2002); Tyler et al. (2008)

Table 3 Species-specific primers

Species	Primer sequences (5′–3′)	T_a (°C)
Bathymodiolus thermophilus	**F:** TGTGGTCTGGAATAATTGGAAC **R:** ATAAAAAGATGTATTRAARTGACG	50
Alvinella pompejana	**F:** TATTTGGTATTTGGGCAGGTC **R:** GATGGGTCGAAGAATGATGTG	57
Branchipolynoe symmytilida	**F:** CCCTTTACTTTCTATTTGGC **R:** ATTTCGATCTGTTAGGAGTATG	51
Hesiolyra bergi	**F:** CATACAAATAGTGGTACTCGTTC **R:** TTCCTTTTCGGACTATGAG	51
Lepetodrilus elevatus	**F:** TGARCTYGGACAACCRGGAG **R:** RGGGTCAAAGAARGARGTGTT	56

F, forward primer; R, reverse primer; T_a, annealing temperature.

box until the submersible was recovered on board the support ship. Animals were individually preserved in absolute alcohol.

Genomic DNA of the three species of gastropods (*L. elevatus*, *L. ovalis* and *E. vitrea*) was extracted using NucleoSpin Tissue Kit (Macherey-Nagel) following the manufacturer's instructions. Genomic DNA of polychaetes (*A. pompejana*, *B. symmytilida* and *H. bergi*) and the bivalve *B. thermophilus* was extracted using a CTAB extraction procedure (Doyle & Dickson 1987). Tissues were digested in 600 μL of a 2% CTAB buffer solution (1.4 M NaCl, 0.2% 2-mercaptoethanol, 20 mM EDTA, 100 mM Tris–HCl pH 8 and 0.1 mg/mL proteinase K) for 2 h at 60 °C. DNA was then purified by adding Chloroform–Isoamyl alcohol (24:1) and precipitated together with 1 mL of 100% isopropanol at −20 °C for 2 h. Finally, DNA pellets were washed with 70% ethanol and re-suspended in 50 μL of sterile H_2O.

Species-specific primers (Table 3) were developed using the first set of sequences obtained from an initial set of amplification and sequencing using 'universally' applicable mitochondrial cytochrome oxidase I primers (LCOI, HCO1, Folmer *et al.* 1994). This procedure strengthened the amplification of DNA for *B. thermophilus*, *A. pompejana*, *B. symmytilida*, *H. bergi* and *L. elevatus*. Polymerase chain reaction (PCR) amplifications were performed in a 25-μL reaction volume containing 1× reaction buffer (supplied by the manufacturer), 2 mM $MgCl_2$, 0.12 mM of each dNTP, 0.5 μM of each primer, 20 μg/mL Bovine Serum Albumin, 0.75 U Thermoprime Plus DNA polymerase (Thermo Scientific), 2 μL of template DNA and sterile H_2O. Thermal cycling parameters used an initial denaturation at 94 °C for 2 min, followed by 40 cycles at 94 °C for 35 s, appropriate annealing temperature for 35 s (Table 3), and 72 °C for 1 min 20 s, before a final 10-min extension at 72 °C.

© 2009 Blackwell Publishing Ltd

Polymerase chain reaction products were purified with Millipore Montage™ μPCR$_{96}$ Cleanup kit and sequenced on an ABI 3130XL DNA analyser using BigDye® Terminator v3.1 (Applied Biosystems) sequencing chemistry, following the manufacturer's protocol. The sequences (GenBank accession numbers: G0473188–G0475034) were proofread in Chromas version 2.23 (http://www.technelysium.com.au/chromas.html) and aligned in BIOEDIT version 6.0.6 (Hall 1999).

Phylogeographical structure and divergence

For each species, number of haplotypes (h) and haplotype diversities (H_d) within each hydrothermal vent site were determined using DnaSP 4.10.3 (Rozas *et al.* 2003). The genealogical relationships among haplotypes were estimated using the median-joining algorithm (Bandelt *et al.* 1999) of Network software (version 4.5.0.0; http://www.fluxus-engineering.com), allowing for the definition of clades (based on divergence up to 0.5%). The haplotype diversity (H_d), nucleotide diversity (π_n) and Watterson's theta (θ_w) were estimated within each divergent clade using DnaSP 4.10.3.

Geographical distribution of the clades was then assessed by plotting the dominant (ancestral) haplotype-frequency distributions for each locality. To detect putative geographical barriers to gene flow, *F*-statistics (ϕ_{ST}; Hudson *et al.* 1992) were computed via DnaSP 4.10.3, with a sliding window of three populations from 21°S to 21°N to avoid sampling size effect. For example, for *E. vitrea*, the Φ_{ST} calculation grouped populations from 21°S–18°S–17°34′S, the second one from 18°S–17°34′S–17°25′S and so on until reaching the last grouping from 9°N–13°N–21°N. Groups were not always the same among species depending on the locality sampling schedule (sites without sample). For each sliding window, departure of the *F*-statistic value from zero was tested using a permutation test. Clade admixtures (i.e. the presence of divergent clades in the population) were examined for each population depending on the proportion of synthetic haplotypes typifying each mitochondrial lineage to reveal potential geographical clines.

Substitution rates consistent with a molecular clock were tested for each species via the BEAUTI/BEAST 1.4.8 Markov chain Monte Carlo (MCMC) package (Drummond & Rambaut 2007) using a subset of 15–20 informative sequences. A GTR + G + I substitution model was chosen to run an uncorrelated lognormal relaxed molecular clock model using a constant population size coalescent with 10 million steps, a sampling every 500 steps and a burn-in of 100 000 steps to reach convergence. The goodness of fit to a strict clock model was performed by examining the posterior distribution of the standard deviation of the uncorrelated lognormal relaxed clock (Std ulrc). Divergence times (T) were then estimated between clades for each species when possible using the formula $T = D/(2r)$, where D is the average divergence between clades and r the evolutionary rate per site per million years (Kumar *et al.* 1996). Three independent clock calibrations were previously performed on mtCOI, using known historical vicariance events of the ridge system that may have affected vent population demography: (i) the Farallon ridge subduction under the American plate, 28.5 Ma (Chevaldonné *et al.* 2002), the formation of the Cascadia depression in the 450-km-long Blanco transform fault, 5 Ma (Johnson *et al.* 2006), and the formation of the Easter microplate, 5.9 Ma (Faure *et al.* 2009). These calibrations led to the estimation of a mutation rate of 2.2%, 2.8% and 3.8% for several vent annelids, *Lepetodrilus* limpets and *Bathymodiolus* bivalves, respectively, indicating a close evolution rate across vent taxa for the mtCOI gene. Consequently, estimations of divergence times between each taxon pair were performed using a consensual (2.8%) mutation rate for the seven species.

To test for simultaneous divergence across clade pairs and common barriers to gene flow among the seven species, an integrative approach using the MsBayes Approximate Bayesian Computational (ABC) software (Hickerson *et al.* 2006b) was performed. The ABC method simulates sequence data sets for a series of taxon pairs that fit a divergence population model in which one ancestral population of size (N_A) splits into two daughter populations a and b (of size N_a and N_b) that independently endure a bottleneck of varying length τ' (and lead to actual sizes of N_a' and N_b'). Runs consist of three steps, allowing for the estimation of interspecies parameters [i.e. hyperparameters such as the number of possible divergence times (Ψ) across taxon pairs, and the corresponding mean ($E(\tau)$) and variance (Ω) of the divergence time τ]. First, a vector of observed summary statistics is obtained from the observed data set of each taxon pair. Second, through a series of hundreds of thousands of replicates, the hyperparameters and subparameters that typify the coalescent model are randomly drawn from hyper-prior and subprior distributions to simulate a corresponding finite sequence data set for the Y taxon pairs. These simulated data sets are then used to calculate a series of pseudo-summary statistics. These pseudo-summary statistics are compared with the observed ones to produce an approximate sample of simulations (i.e. simulations that yield the summary statistics values nearest to that of the observed ones), which is then used to draw the posterior distributions of the hyper- and subparameters of the taxon pairs, following an arbitrary rejection/acceptance ratio of 0.1–0.2 and a local weighted linear regression.

© 2009 Blackwell Publishing Ltd

Twenty 329-bp-long sequences per clade constituted the input data set. Because of a lack of transversion in the between-clade divergence of some species, the transition–transversion ratio was fixed to two (Jukes & Cantor 1969). Joint prior distributions were used to perform 500 000 simulated draws. Out of these simulated draws, the closest to the observed data set (1000) was used to define the joint posterior distribution. Upper bounds for the prior population mutation parameter for the ancestral population size (θ_A) were chosen to be equal to 5, given the observed pairwise differences within clades (Hickerson *et al.* 2006b). These parameters were unchanged across runs, allowing the cross-taxa comparison of results. ABC analysis was run on the seven species without any constraint on the number of divergence times (Ψ) to estimate the hyperparameters. ABC analysis was also performed fixing $\Psi = 2$ because obtained Ψ values were >1. A final ABC analysis was computed on species after suppressing the pair-clade *L. elevatus* (for which divergence between clades was considered extremely high), without any constraint on Ψ.

Demographic history

Significantly non-null Tajima's *D* and its derived statistics (e.g. the Fu & Li's *F*) have been widely used to trace demographic events (Glinka *et al.* 2003; Akey *et al.* 2004), given that the studied gene region evolved under a lack of selective pressures. To test the hypothesis that vent fauna have undergone demographic changes in response to shared vicariant events, Tajima's (1989) statistic *D* was estimated (using DnaSP 4.10.3) within each species for both the southern and northern clades located from each side of the equatorial biogeographical break following the results of Bachraty *et al.* (2009). The departure of the value of Tajima's *D* from zero was tested using a two-tailed test, assuming that *D* follows the beta distribution (Tajima 1989).

To further examine whether or not species endured the same demographic change, effective population size (N_e) and the exponential population growth parameter (g) were estimated from clades with significantly non-null Tajima's *D*-values using Fluctuate (version 1.4, Kuhner *et al.* 1998). This allowed for the timing of expansion to be generated using the formulation $\theta(t) = \theta_0 e^{-g\mu t}$ in which $\theta = N_e\mu$, where μ is the mutation rate, t is the time elapsed (in generations) and θ_0 is the present state of θ for the clade under scrutiny. For each species, Fluctuate analyses were run with different values of short (i.e. 10, 50, 100, 150) and long (i.e. 2, 10, 20, 30) chains to evaluate convergence between runs. Sampling increment and the number of steps were, respectively, three and fifteen times greater than the number of sequences found within a clade for both short and long chains. To test if expansion signature was because of a wave of colonization (stepwise foundations) or from the recovering of a recent bottleneck, asymmetrical gene flow was checked across the barrier. For each species, the sliding window of Φ_{ST} was used to locate the barrier and populations were then separated into two geographical groups from each part of the barrier, respectively. Watterson's theta (θ_w) was calculated for each group using DnaSP 4.10.3 (Rozas *et al.* 2003) and used as a starting parameter in Migrate 3.0.3 (Beerli & Felsenstein 1999, 2001) to estimate the gene flow parameters (θ**M*) with MCMC runs of 10 short and three long chains.

Results

Phylogeographical structure and divergence across taxon pairs

For each species, the number of haplotypes (*h*) and haplotype diversities (H_d) within each population were assigned (Table 1). Median-joining networks exhibited similar topological patterns across all species, but differed in their structure (Fig. 1). Two divergent clades were detected in all networks except for: (i) *Hesiolyra bergi* in which three divergent but markedly rare haplotypes were found in two of the most southern populations and (ii) for the scaleworm *Branchipolynoe symmytilida* in which connections are more complex, with at least nine nearly equally frequent haplotypes. The clades were geographically structured (with the exception of *B. symmytilida*), in particular with regard to the geographical location of the equator (Fig. 1). In each network, at least the southern clade (B) displayed a 'star'-like topology, with one central and frequent (ancestral) haplotype surrounded by a crown of derived singletons or multiple singletons (however, see *Bathymodiolus thermophilus* for the exception of two closely related, equally frequent haplotypes in the southern lineage). Derived haplotypes were often 'unique' within a given locality. The northern clade was more diversified with more distant (older) COI lineages. Divergence between the northern and the southern clades ranged from 0.9% for *B. thermophilus* to 6.5% for *Lepetodrilus elevatus*. By comparison, divergence for the two most frequent haplotypes of the scaleworm *B. symmytilida* was only 0.4%.

The geographical distribution of haplotypes (Fig. 1) together with Φ_{ST} differentiation tests (Fig. 2) allowed us to discriminate clear patterns of geographical isolation for all species. When grouping localities from south to north, the sliding window Φ_{ST} became significantly different from zero between 17°S and the equator in all

© 2009 Blackwell Publishing Ltd

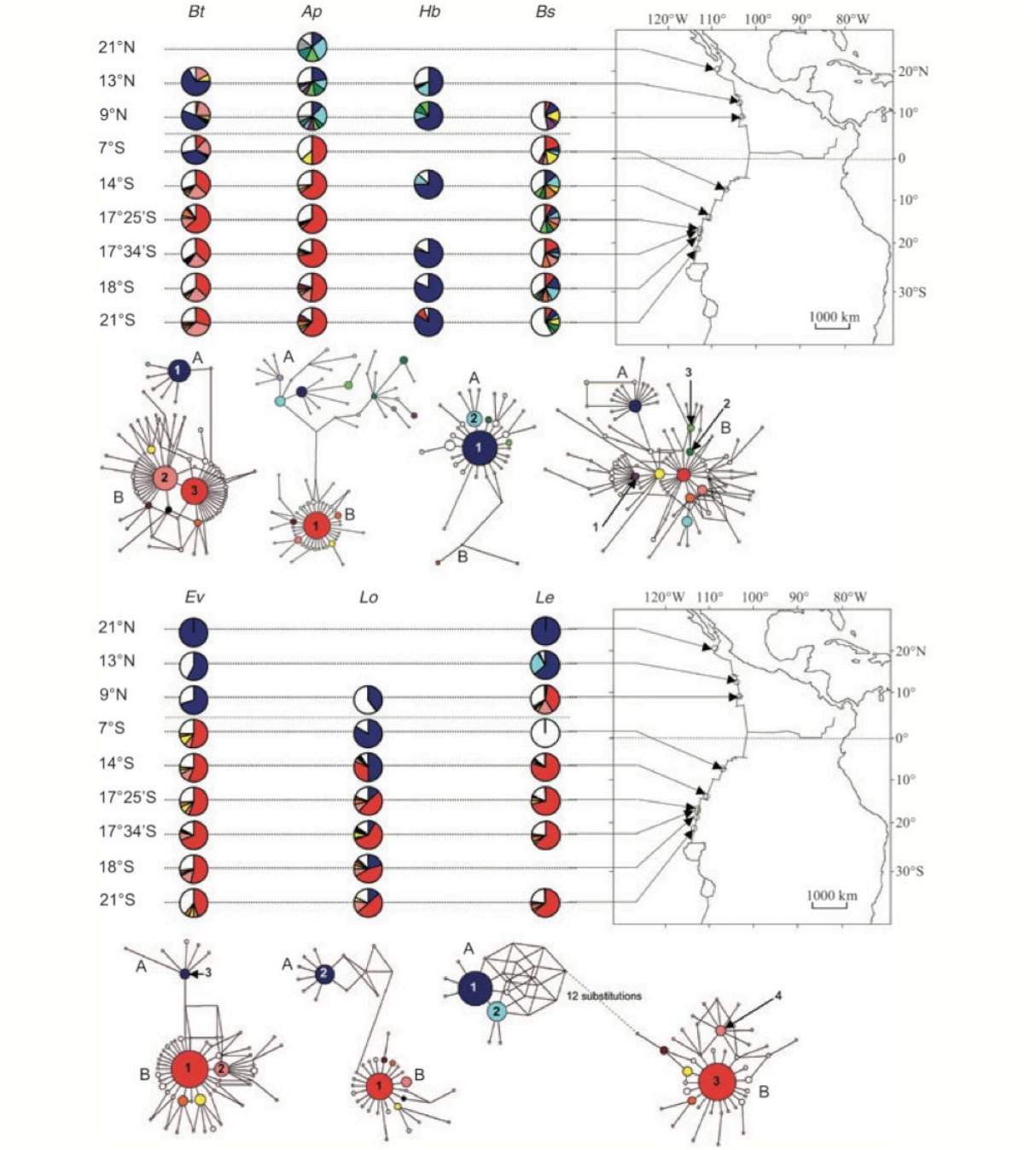


Fig. 1 Median-joining networks and haplotype-frequency distributions of all sampled populations for the seven species. Size of haplotype circles and connections are proportional to number of sequences and mutation step, respectively. A and B represent the two divergent clades for each species. On the haplotype-frequency distributions, shared haplotypes with >2% frequency within its corresponding clade are coloured. Private haplotypes and shared haplotypes with <2% frequency within its clade are in white. *Bt*, *Bathymodiolus thermophilus*; *Ap*, *Alvinella pompejana*; *Hb*, *Hesiolyra bergi*; *Bs*, *Branchipolynoe symmytilida*; *Ev*, *Eulepetopsis vitrea*; *Lo*, *Lepetodrilus ovalis*; *Le*, *Lepetodrilus elevatus*.

species, except *H. bergi*. The occurrence of possible admixtures between divergent haplotypes typifying each clade along the EPR was detected (Table 1). Except for *B. thermophilus* and *Lepetodrilus ovalis*, admixture is often caused by <10% of individuals in the investigated populations.

© 2009 Blackwell Publishing Ltd

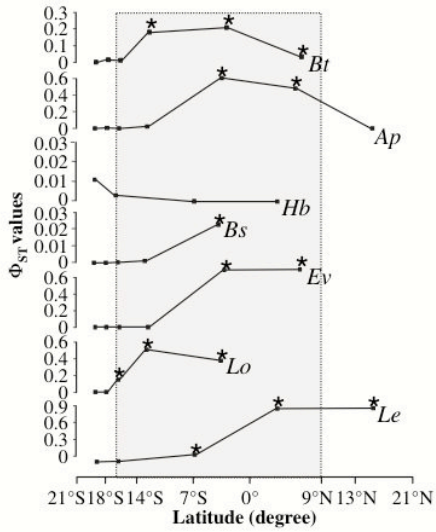


Fig. 2 Distribution of Φ_{ST} values calculating for groups of three populations using a sliding window as a function of distance to 21°S. Each point represents Φ_{ST} values relative to a barycentric position of the three vent field latitudes used in the sliding window. *$P < 0.05$. *Bt, Bathymodiolus thermophilus*; *Ap, Alvinella pompejana*; *Hb, Hesiolyra bergi*; *Bs, Branchipolynoe symmytilida*; *Ev, Eulepetopsis vitrea*; *Lo, Lepetodrilus ovalis*; *Le, Lepetodrilus elevatus*.

The vent polychaete *Alvinella pompejana* (*Ap*) and the gastropod limpet *Eulepetopsis vitrea* (*Ev*) both displayed exactly the same network architecture (Fig. 1). Coalescence trees are separated into two distinct clades (A and B) across the equator with 1% (*Ap*) and 0.9% (*Ev*) mutations accumulated into the divergence, respectively. Only one sequence from the southern clade (B) of *A. pompejana* was sampled in the northern EPR.

The vent limpet *L. elevatus* (*Le*) also displayed a pronounced geographical structure with two highly divergent clades (divergence = 6.5%); however, the clades overlap on the northern part of the EPR at 9°N (Fig. 1). In the northern clade, the population from 21°N displayed only one haplotype (i.e. Hap 1, the most frequent one in other populations), leading to a significant pairwise Φ_{ST} between these two disjunct vent fields with and without considering the presence of clade B at 13°N. In the southern clade, the population from 9°N sharply differed from the SEPR ones by the presence of a unique haplotype (Hap 4) at a high frequency. This haplotype diverged from the most frequent (ancestral) one (Hap 3) by three mutations and possessed a crown of derived haplotypes, indicating that it had time to diversify since the southern vs. northern isolation of lineages.

The bivalve *B. thermophilus* (*Bt*) and the gastropod *L. ovalis* (*Lo*) showed a nearly similar but reverse situation in which haplotypes of one clade were distributed along the entire ridge system, whereas haplotypes from the other clade A were restricted to the northern (*Bt*) or the southern (*Lo*) parts of the EPR.

The two other vent species, *H. bergi* (*Hb*), which lives in sympatry with *A. pompejana*, and *B. symmytilida (Bs)*, which lives commensally with the bivalve *B. thermophilus*, were characterized by a lack of obvious differences across localities throughout the EPR (Fig. 1). However, three individuals of *H. bergi* located in the southern part of the EPR exhibited haplotypes with a clear divergence of nearly 1%, which may represent 'old' surviving mitochondrial lineages, whereas an isolation-by-distance structure was observed in *B. symmytilida*.

Using Migrate 3.0.3, *A. pompejana* and *E. vitrea* displayed a nearly complete absence of mtCOI gene flow across the equatorial region, whereas other species showed high asymmetric gene flow (Table 4) from north to south for *H. bergi* and south to north for the remaining species.

For each species, standard deviation of the uncorrelated lognormal relaxed clock (Std ulrc) was close to zero (mean ranged from 0.416 to 0.608, with maximum probabilities close to zero and upper HPDs under 1), indicating no mutation rate heterogeneity among clades. Thus, these data sets can be utilized for divergence time estimates that assume a constant rate. Clade splitting dates calculated with 0.28% per Myr and the $T = D/(2r)$ formula lead to the following divergence times: T_{Le} = 11.6 Myr, T_{Lo} = 3.6 Myr, T_{Hb} = 1.9 Myr, T_{Ap} = 1.8 Myr, T_{Ev} = 1.6 Myr, T_{Bt} = 1.6 Myr and T_{Bs} = 0.7 Myr.

Table 4 Theta (θ) and migration (θ*M*) parameters in each species

Parameters	*Bt*	*Ap*	*Hb*	*Bs*	*Ev*	*Lo*	*Le*
θ North population	0.014	0.016	0.024	0.050	0.007	0.018	0.040
θ South population	0.034	0.024	0.009	0.065	0.030	0.018	0.013
θ*M* North to South	0.000	0.000	2.338	1.142	0.280	0.377	0.000
θ*M* South to North	21.676	0.572	0.000	90.882	0.000	15.292	28.077

Bt, Bathymodiolus thermophilus; *Ap, Alvinella pompejana*; *Hb, Hesiolyra bergi*; *Bs, Branchipolynoe symmytilida*; *Ev, Eulepetopsis vitrea*; *Lo, Lepetodrilus ovalis*; *Le, Lepetodrilus elevatus*.

© 2009 Blackwell Publishing Ltd

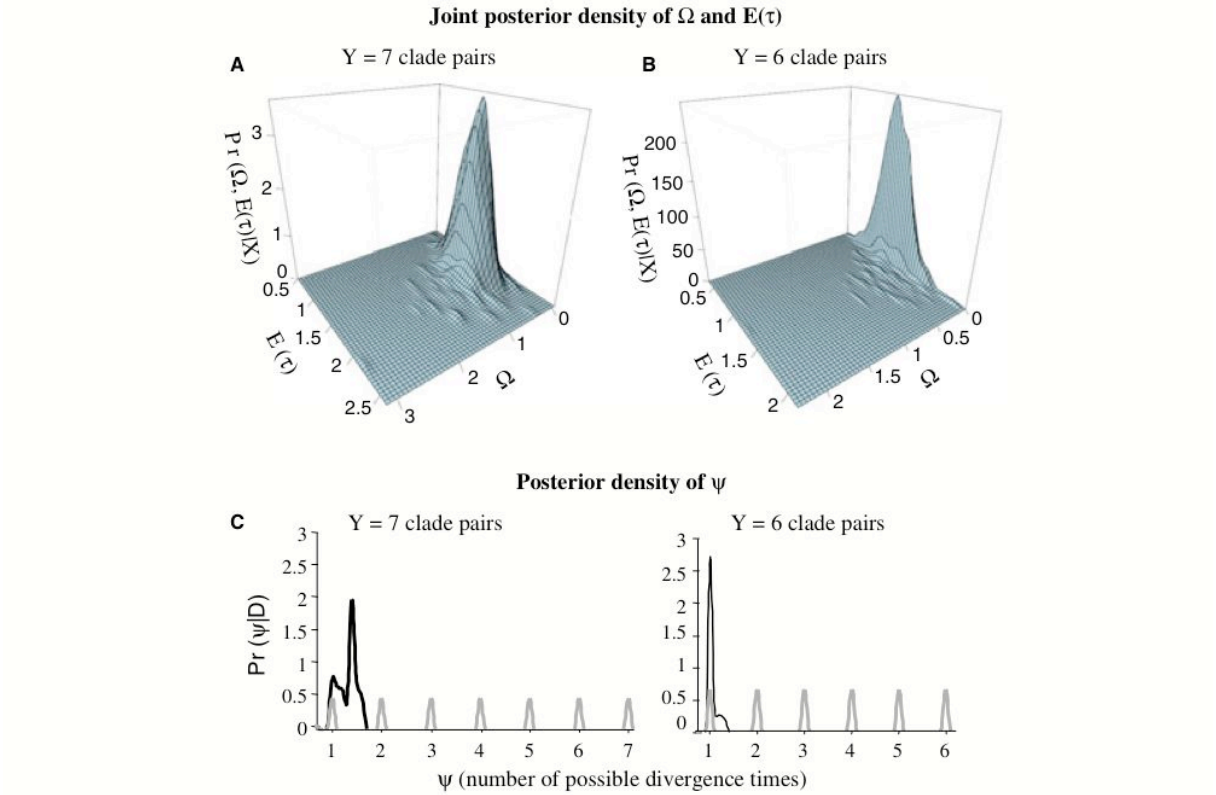


Fig. 3 (A, B) Three-dimensional joint posterior probability densities for $E(\tau)$ and Ω. (C, D) Posterior probability densities for Ψ, the number of divergence times given Y clade pairs. Estimates in (A) and (C) are based on data from all seven species clade pairs, whereas estimates in panels (B) and (D) are based on a data set in which *Lepetodrilus elevatus* was excluded. These estimates use the same uniform prior for θ_A bounded by 0.5 and 5.0 and are based on 500 000 simulated draws from the joint hyper-prior and 1000 draws from the joint posterior using MsBayes Approximate Bayesian Computational software (Hickerson et al. 2006b). In panels (C) and (D), the dotted line is the prior for Ψ and the solid line is the posterior for Ψ.

Estimates of Ω [=Variance(τ)/$E(\tau)$] and the number of shared vicariant events Ψ with MsBayes using our seven taxon pairs did not support a history of simultaneous divergence (Fig. 3A, C). Indeed, the average Ψ value (1.858) was close to 2, indicating the occurrence for two possible isolation times. Moreover, the amount of variance (Ω = 0.384) was high, possibly indicating a biased estimation of the mean divergence time $E(\tau)$ (=1.521), corresponding to 1.65 Myr using the T = [100 $E(\tau)$]/(r*length of gene) equation (Hickerson *et al.* 2006a). Simulating vicariance events across the seven taxon pairs by fixing Ψ = 2 indicated that six species have a simultaneous divergence time [with $E(\tau_1)$ = 1.088 corresponding to 1.2 Myr) and one species had a greater divergence time (with $E(\tau_2)$ = 2.639 corresponding to 2.9 Myr). The most divergent taxon pair was expected to correspond to *L. elevatus* because this species showed the greatest single T-value across species. Simulation excluding *L. elevatus* pairs (Fig. 3B, D) showed a clear-cut estimate of Ψ nearly equal to one (1.316) together with a markedly small variance (Ω = 0.047 < 0.1). This latter estimate seems to be robust if the ancestral coalescent variance has a larger effect on the total genetic divergence when divergence times are recent: a plausible explanation for EPR vent fauna in light of the allopatric distribution across the equator. In this study, the high variance of coalescent estimates across taxon pairs was greatly compensated by the choice of the prior for θ_A (0.5–5), highly stringent boundaries for this parameter being the consequence of the exceptionally small ranges of observed pairwise differences within populations within each species pair. Simultaneous divergence time $E(\tau)$ was estimated to be 1.187, corresponding to a vicariant event 1.3 Ma. Divergence time (1.1 Ma using the $T = D/2r$ equation) between Hap 3 and Hap 4 for *L. elevatus* was congruent with this simultaneous divergence time estimated for the six other species, suggesting that all species were subjected to this vicariant event 1.3 Ma. Results using BEAST were consistent with a common and recent

© 2009 Blackwell Publishing Ltd

vicariant event for all taxa (including the southern lineage of *L. elevatus*), with average times of the most recent common ancestor (T_{MRCA}) ranging from 9.2×10^{-4} to 2.3×10^{-3} with a strong overlapping of HPDs.

Shared demographic histories in the South

To examine how clades evolved after being putatively isolated, Tajima's *D* was calculated for each clade separately over the seven species pairs. Results (Table 5) indicated a difference between the southern and northern clades. Tajima's *D* was always significantly negative in the southern clade, regardless of the species analysed. A significant excess of rare variants was also detected in the polychaetes *H. bergi* and *B. symmytilida* for which only one clade was detected over the entire range. Moreover, when considering only populations without admixture (Table 1), 25%, 7% and 42% of populations had pairwise distributions differing significantly from the neutral hypothesis in 21°N–13°N, 9°N–14°S and 17°25′S–21°S regions, respectively. Fluctuate analyses of the southern lineages (Fig. 4; using 100 short and 20 long chains allowing convergence between runs for all species) estimated the beginning of a global population expansion between 100 000 and 300 000 generations for all species except *B. symmytilida* (~1 million generations), with exponential population growth parameter *g* ranging from 3640 (*L. elevatus*) to 10 000 (*H. bergi*) when discarding *B. symmytilida* (g = 2160).

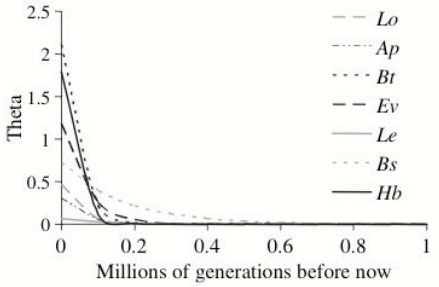


Fig. 4 Graph of theta parameter (where $\theta = 2N_e\mu$, N_e = effective population size and μ = the mutation rate) over number of generations based on Metropolis-Hastings Monte Carlo coalescent analysis using Fluctuate version 1.4. Pattern of growth is based on estimates of g (the exponential growth rate of the population) generated jointly with θ. *Bt*, *Bathymodiolus thermophilus*; *Ap*, *Alvinella pompejana*; *Hb*, *Hesiolyra bergi*; *Bs*, *Branchipolynoe symmytilida*; *Ev*, *Eulepetopsis vitrea*; *Lo*, *Lepetodrilus ovalis*; *Le*, *Lepetodrilus elevatus*.

Discussion

The seven deep-sea hydrothermal vent species studied in this work possess different life-history traits (Table 2), which are likely to influence their ability to expand their range and colonize new localities. Despite biological disparities across species, mtCOI network topologies yielded similar topologies—two divergent clades located in Northern and Southern EPR, respec-

Table 5 Haplotype diversity (H_d), nucleotide diversity (π_n), Watterson's theta per site from number of segregating sites (θ_w) and Tajima's *D* of clades from studied species and overall Φ_{ST} values for each species

Species/clade (Φ_{ST})	*n*	H_d (SD)	π_n (SD)	θ_w (SD)	*D*
Bathymodiolus thermophilus (0.255***)					
Northern clade	64	0.208 (0.068)	0.000 (0.000)	0.002 (0.001)	−2.011*
Southern clade	245	0.812 (0.019)	0.003 (0.000)	0.016 (0.004)	−2.452***
Alvinella pompejana (0.533***)					
Northern clade	108	0.911 (0.014)	0.009 (0.000)	0.014 (0.004)	-0.983^{NS}
Southern clade	210	0.618 (0.040)	0.002 (0.000)	0.015 (0.004)	−2.564***
Hesiolyra bergi (0.018**)					
Major clade	232	0.614 (0.036)	0.002 (0.000)	0.015 (0.004)	−2.379**
Southern minor clade	3	—	—	—	—
Branchipolynoe symmytilida (0.031***)	208	0.970 (0.005)	0.007 (0.000)	0.022 (0.005)	−2.137**
Eulepetopsis vitrea (0.578***)					
Northern clade	18	0.562 (0.134)	0.002 (0.001)	0.005 (0.003)	−1.849*
Southern clade	279	0.646 (0.032)	0.003 (0.000)	0.016 (0.004)	−2.369**
Lepetodrilus ovalis (0.484***)					
Northern clade	51	0.722 (0.047)	0.003 (0.000)	0.004 (0.002)	-1.143^{NS}
Southern clade	136	0.628 (0.048)	0.002 (0.000)	0.012 (0.003)	−2.477**
Lepetodrilus elevatus (0.900***)					
Northern species	186	0.432 (0.036)	0.001 (0.000)	0.005 (0.002)	-1.599^{NS}
Southern species	253	0.590 (0.037)	0.003 (0.000)	0.016 (0.004)	−2.326**

n, number of sequences; SD, standard deviation; —Too small sample size to estimate indices. $^{NS}P > 0.05$; $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

© 2009 Blackwell Publishing Ltd

tively, with only one exception, the commensal scaleworm *Branchipolynoe symmytilida*. Similar topologies raise the question of whether the equatorial barrier, previously described by Bachraty *et al*. (2009), may have played a role in promoting a genetic break(s) across vent species and subsequent gene flow limitations after the isolation of populations. To test the 'vicariance' hypothesis, two main questions need to be answered: (i) did ancestral species split at the same time and (ii) was the barrier impermeable enough to impose the same exact geographical patterns across the northern and southern vent lineages?

Number of splitting events in vent populations along the East Pacific Rise

To date vicariant events, the most currently used approaches consist of: (i) estimating divergence times for each taxon pair; (ii) comparing these dates across taxa, accounting for stochastic processes associated with the coalescence process; and (iii) estimating the demographic evolution of populations. This technique has been widely used to propose vicariant events when speciation has occurred in the distant past (several millions years) or when the target taxa represent conspecific species displaying the same life-history traits across the same habitat. For example, simultaneous vicariant events have been documented for coastal invertebrates inhabiting sediments along the European coasts of the North Atlantic (Jolly *et al.* 2006) or the intertidal limpets separated across the Greater Cook Strait in New Zealand (Goldstien *et al.* 2006). In our case, the vicariance date of 11.6 Myr estimated for *Lepetodrilus elevatus* coincides with two cryptic, possibly hybridizing, species detected by Matabos *et al*. (2008) at 9°50'N/EPR using allozymes and mtCOI sequences. However, divergence times of the other remaining species are more recent and markedly close to each other (from 3.6 Ma for *Lepetodrilus ovalis* to 1.6 Ma for *Bathymodiolus thermophilus* and 0.7 Ma for *B. symmytilida*), suggesting that they may be the result of a same vicariant event.

However, testing whether co-distributed taxa share a common history of simultaneous vicariance just by superimposing divergence times may lead to erroneous conclusions if the vent community mixes species groups of different origins, possibly via arriving into the ridge system at different times. According to Hickerson *et al.* (2006b), variation between estimated divergence times is largely explained by the mutational and coalescent variance. A recent approach (Hickerson *et al.* 2006b) allows testing simultaneous divergence using ABC by incorporating differences in the demographic history of each sister population during the isolation process. Simulations using our seven taxon pairs indicated two possible splitting events (simulation without fixing Ψ), with the first, most recent event affecting six species and the second one with a much older divergence time (simulation with $\Psi = 2$) affecting only one species, *L. elevatus*, according to the previous species-by-species estimation of divergence times. Although the most ancient divergence time estimated for one species (*L. elevatus*) seems to be underestimated (2.9 Myr), the simultaneous divergence time (1.2 Myr) estimated with $\Psi = 2$ is consistent with the 1.3 Myr detected for the most recent event when using $\Psi = 1$ and the six remaining taxon pairs. Moreover, *L. elevatus* divergence between Hap 3 and 4 (1.1 Ma) inside clade B is congruent with this simultaneous divergent time, indicating that *L. elevatus* was submitted to this more recent isolation event as well. Consequently, the results are congruent with the hypothesis of two independent isolation events: 11.6 Ma for *L. elevatus* and 1.3 Ma for the other taxon pairs. As allopatric speciation is usually caused by gene-flow disruption due to physical/tectonic barriers that modify hydrothermalism along ridge axes (Jollivet 1996), the two possible splitting dates subdividing these vent species into two distinct phylogeographical clades are probably explained by the formation of geological discontinuities that progressively offset the ridge crest.

Geological barriers to gene flow along the East Pacific Rise

Plate tectonics most probably influence genetic structure and speciation of the hydrothermal vent fauna, but the exact events are tough to pinpoint. Modelling the formation of the East Pacific ridge system based on fossil records and magnetic/gravimetric anomalies suggested that the subduction of the Farallon plate under the American plate provoked multiple reorganizations of the now-extinct Pacific/Farallon ridges, while forming the present EPR (Mammerickx *et al.* 1980). Two major co-occurring tectonic events could coincide with the divergence date of the *L. elevatus* split 11.6 Ma: (i) the Bauer microplate rotation and (ii) the Mathematician Ridge reorientation. The formation of the Bauer Overlapping Spreading Centre initiated about 17 Ma (at latitudes located between 10° and 15°S) evolved into a microplate between 15 and 11 Ma (Eakins & Lonsdale 2003). Such reorganization provoked ridge offset and the formation of two parallel active ridges that could have been responsible for modifying the bottom current patterns in this region and subsequently the *L. elevatus* divergence. During the same period (12.5–11 Ma, Mammerickx & Klitgord 1982), the Mathematician Ridge situated further North (12°–17°N) was subjected to abrupt changes in magnetic and bathymetric orientations, provoking the fossilization of transform faults

© 2009 Blackwell Publishing Ltd

and the formation of a parallel ridge system, the Moctezuma trough, in this northern region (Mammerickx & Klitgord 1982). Because the Mathematician Ridge reorganization event appears to be much closer to latitudes at which the two present divergent lineages overlap (between 9° and 13°N), this latter scenario seems to be more consistent with the *L. elevatus* split.

The seven vent species studied seem to share a simultaneous vicariant event ≈1.3 Ma. Moreover, this vicariant event coincides with a significant north/south differentiation of vent populations for all species, with the exception of the polychaete *Hesiolyra bergi*. However, this polychaete species displays three divergent haplotypes sampled at the most southern sites, suggesting that the second clade was either under-sampled or had become nearly extinct. Even if *B. symmytilida* and *L. elevatus* seem to display a more complex population history, both species conform to the hypothesis of a recent North/South isolation. This is particularly clear when considering the genetic differentiation observed between populations located at 9°N and 14°S for *B. symmytilida* and the southern lineage (clade B) of *L. elevatus* at the same geographical sites. By superimposing geographical distribution of haplotypes and the position at which genetic differentiation increases significantly across species, this barrier is likely to be positioned between the Equator and 17°S, and defines a clear transition zone between the Northern and Southern EPR. Most transform faults located between 9°N and 17°S (i.e. Quebrada/Discovery/Gofar fracture zone system, or Wilkes and Garrett transform faults) deeply offset the ridge axis into separate segments around 1–2 Ma (Kureth & Rea 1981; Naar & Hey 1989; Francheteau *et al.* 1990), a scenario consistent with the simultaneous 1.3 Ma divergence time. For example, the 450-km-long Blanco Transform fault is known to be responsible for the speciation of *Lepetodrilus fucensis* and *Lepetodrilus gordensis* (Johnson *et al.* 2006) with a divergence of 7.3% on the mtCOI gene and allele frequency inversion at the phosphoglucomutase gene. However, the impact of this fault as a barrier to gene flow was less intense for the tubeworm *Ridgeia piscesae*, for which isolation did not lead to reciprocal monophyletic clades but only abrupt haplotype-frequency differentiation in populations located at each part of the ridge offset (Young *et al.* 2008).

The slow and progressive formation of a tectonic barrier (rate at which a ridge offset is typically 5–10 cm per year) could explain why slight discrepancies in population differentiation still hold across species, with some species being able to cross the barrier when others are not (Knowlton & Weigt 1998). Therefore, the slight differences observed in the barrier positioning may be attributable to differing dispersal capabilities. Indeed, egg size, egg vitellogenin content and larval developmental mode (see Table 2) are good apparent indicators of dispersal capabilities and vary greatly across species. These characteristics affect the buoyancy of propagules and are likely to change the vertical dispersal of a larva in the water column (e.g. Mullineaux *et al.* 2005). Thus, larvae could be subjected to different water currents and subsequent divergent trajectories. This may be particularly the case of *Bathymodiolus* bivalves, which have planktotrophic larvae capable of reaching the upper water column (Arellano & Young 2009) as opposed to *Lepetodrilus* gastropods, the larvae of which are mainly found beneath hydrothermal plume layers, <200 m above the seafloor (Mullineaux *et al.* 2005). The extremely great size of mature oocytes (400 μm) of the scaleworm *Branchipolynoe* spp. may also explain why this species is so weakly affected by the equatorial barrier, as large yolky eggs may delay larval metamorphosis for months (Jollivet *et al.* 2000). The frequency of available vent habitat may also affect the ability of vent species to cross the barrier, as diffuse venting systems are more prevalent along the ridge compared with vent chimneys. However, the habitat distribution does not seem to play an important role to this extent as lineages from the two chimney-living species (*Ap* and *Hb*) have a markedly contrasted geographical range.

Regardless of the species, the precise spatial positioning of the barrier remains difficult to establish; first, because some of these barriers may have disappeared (i.e. the Bauer microplate) and second, because migration events and subsequent secondary contacts probably mask the exact position of the barriers. Secondary contact zones detected in this study often display clines of clade-specific haplotype frequencies (responsible for north/south EPR differentiation). Such clines have been detected particularly in *L. ovalis* and *B. thermophilus* and might be attributed to preferential migration routes along the EPR, as suggested by the strong South to North asymmetric gene flow revealed by Migrate results. However, differential lineage extinctions between northern and southern EPR (regional extinctions of *L. ovalis* clade B in northern EPR and/or *B. thermophilus* clade A in the southern EPR) could also explain such haplotype distributions.

Evidence for simultaneous population expansions in the south

Comparing haplotype coalescence trees and overall gene diversities between reciprocal monophyletic clades of the examined vent taxa (with exception of *H. bergi* and *B. symmytilida* for which a single clade is preponderant) indicated large discrepancies between the northern and southern lineages. Star-like topologies

© 2009 Blackwell Publishing Ltd

together with reduced gene diversities were indeed observed for southern lineages (generally not for the northern ones) and are strongly suggestive of non-equilibrium dynamics. These results were supported by significant negative Tajima's *D* for the southern clades. Furthermore, a greater proportion of populations showed departure to equilibrium between 17°25′S and 21°S when compared with the most northern populations. These departures to equilibrium may be explained by either a recent demographic expansion along the Southern EPR (*sensu* Harpending *et al.* 1998) or alternatively recurrent selective sweeps at the mitochondrial locus (Bazin *et al.* 2006). However, the positive selection hypothesis here imposes that a series of nearly simultaneous fixations of advantageous alleles co-occurred in different species, and only for the most southern cryptic lineage: a situation in which mitochondria would have been sensitive to positive fixation and able to sweep at an extremely high rate in the south but not in the north. In addition, Fluctuate analyses performed on our seven vent taxa revealed a nearly simultaneous expansion of southern populations for all species, except the commensal scaleworm *B. symmytilida*. Altogether, the selective hypothesis is highly unlikely because simultaneous selective sweeps across different vent taxa are highly improbable.

If we accept the idea that populations expanded on one side of the barrier after having been separated about 1.3 Ma, the nearly simultaneous expansion would date back to 100 000 to 300 000 generations for all species except *B. symmytilida* (around 800 000 generations). Assuming that one to two generations occur per year for nearly all vent species, the date of expansion would fall in the last 500 000 years and coincide well with an expansion subsequent to the splitting of the two vicariant lineages. Two hypotheses are proposed: (i) this demographic expansion coincides with a geographical expansion of most lineages to the south by subsequent founder events until discouraged by the Eastern Microplate around 23°S and (ii) the southern EPR endured a large catastrophic/eruptive event that destroyed most of the vent fauna throughout the region, thus causing a large bottleneck, about 0.5 Ma. The second hypothesis appears to be more probable, as eruptive events can be frequent occurrences along the EPR, with the observation of two eruptions in the last 20 years near 9°50N (Haymon *et al.* 1993; Cowen *et al.* 2007). Such an assumption is also in agreement with the lack of concordant asymmetric flow across species from north to south. Indeed, vent species did not display similar patterns of orientated gene flow across the barrier, whereby gene flow was mainly orientated in the opposite direction from south to north. Such eruptive events may also be responsible for a possible bottleneck at 21°N (only one haplotype has been sampled for *L. elevatus*). Multiple catastrophic events along the southern EPR could, therefore, explain this cross-species demographic pattern.

Recent theoretical studies argue that mitochondrial genes may not be appropriate to perform phylogeographical analyses despite some advantages, such as the nonrecombination of the gene (Bazin *et al.* 2006). This study illustrates the usefulness of mitochondrial markers when used across a set of species sharing the same environment, regional distribution and more or less the same history. Comparing multiple phylogeographical patterns can help in discriminating demographic vs. selective effects, and thus yielding a better understanding of the microevolutionary processes that shape the geographical structure of populations.

Acknowledgements

We thank the chief scientists and 'Nautile' and 'Alvin' crews for their technical support and efforts during our oceanographic expeditions: HOPE99, PHARE2002 (F. Gaill and N. Lebris), Extreme 2003 (AT11-4 cruise) and BIOSPEEDO2004. We are very grateful to Stéphane Hourdez and Eric Thiébaut for collecting and sorting polychaetes and gastropods, and to Marjolaine Matabos for her help in the diagnosis of Lepetodrilid gastropods. We are also truly indebted to the sequencing genomic platform (GENOMER, Station Biologique de Roscoff, France) for DNA direct sequencing. This work was supported by the GDR Ecchis, ANR-06-BDV-005 (Deep Oases) and ANR-05-BLAN-0407 (Alvi_Stress_Adapt). S. Plouviez was supported by a PhD grant from the Université Pierre et Marie Curie. T.M. Shank was supported by the National Science Foundation (OCE-01-29394 and OCE-03-16348) and a Fellowship from the Deep Ocean Exploration Institute, Woods Hole Oceanographic Institution.

References

Akey JM, Eberle MA, Rieder MJ *et al.* (2004) Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biology*, **2**, 1591–1599.

Arellano SM, Young CM (2009) Spawning, development and the duration of larval life in a deep-sea cold-seep mussel. *Biological Bulletin*, **216**, 149–162.

Avise JC (1998) The history and purview of phylogeography: a personal reflexion. *Molecular Ecology*, **7**, 371–379.

Bachraty C, Legendre P, Desbruyères D (2009) Biogeographic relationships among deep-sea hydrothermal vent faunas at global scale. *Deep Sea Research Part I: Oceanographic Research Papers*, **56**, 1371–1378.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Bazin E, Glemin S, Galtier N (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science*, **312**, 570–572.

© 2009 Blackwell Publishing Ltd

Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*, **152**, 763–773.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences, USA*, **98**, 4563–4568.

Bermingham E, Avise JC (1986) Molecular zoogeography of fresh-water fishes in the Southeastern United-States. *Genetics*, **113**, 939–965.

Burton RS (1998) Intraspecific phylogeography across the point conception biogeographic boundary. *Evolution*, **52**, 734–745.

Chevaldonné P, Jollivet D, Vangriesheim A, Desbruyeres 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.

Chevaldonné 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. *Cahiers De Biologie Marine*, **43**, 367–370.

Cowen JP, Fornari DJ, Shank TM *et al.* (2007) Volcanic Eruptions at East Pacific Rise near 9°50′N. *EOS, Transactions of the American Geophysical Union*, **88**, 81–83.

Doyle JJ, Dickson E (1987) Preservation of plant samples for DNA restriction endonuclease analysis. *Taxon*, **36**, 715–722.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**(214), doi:10.1186/1471-2148-7-214.

Eakins B, Lonsdale P (2003) Structural patterns and tectonic history of the Bauer microplate, Eastern Tropical Pacific. *Marine Geophysical Researches*, **24**, 171–205.

Emerson BC, Paradis E, Thebaud C (2001) Revealing the demographic histories of species using DNA sequences. *Trends in Ecology & Evolution*, **16**, 707–716.

Faure B, Chevaldonné P, Pradillon F, Thiébaut E, Jollivet D (2007) Spatial and temporal dynamics of reproduction and settlement in the Pompeii worm *Alvinella pompejana* (Polychaeta: Alvinellidae). *Marine Ecology Progress Series*, **348**, 197–211.

Faure B, Jollivet D, Tanguy A, Bonhomme F, Bierne N (2009) Speciation in the deep-sea: multilocus analysis of divergence and gene flow between two hybridizing species of hydrothermal vent mussels. *PLoS ONE*, **4**(8), doi:10.1371/journal.pone.0006485.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Francheteau J, Armijo R, Cheminee JL *et al.* (1990) 1 Ma East Pacific Rise oceanic-crust and uppermost mantle exposed by rifting in Hess Deep (Equatorial Pacific-Ocean). *Earth and Planetary Science Letters*, **101**, 281–295.

Fretter V (1988) New archaeogastropod limpets from hydrothermal vents; superfamily *Lepetodrilacea*. II. Anatomy. *Philosophical Transactions of the Royal Society of London. Series B*, **318**, 33–82.

Glinka S, Ometto L, Mousset S, Stephan W, De Lorenzo D (2003) Demography and natural selection have shaped genetic variation in Drosophila melanogaster: a multi-locus approach. *Genetics*, **165**, 1269–1278.

Goldstien SJ, Schiel DR, Gemmell NJ (2006) Comparative phylogeography of coastal limpets across a marine disjunction in New Zealand. *Molecular Ecology*, **15**, 3259–3268.

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.

Harpending HC, Batzer MA, Gurven M *et al.* (1998) Genetic traces of ancient demography. *Proceedings of the National Academy of Sciences, USA*, **95**, 1961–1967.

Haymon RM, Fornari DJ, Edwards MH *et al.* (1991) Hydrothermal vent distribution along the East Pacific Rise crest (9-degrees-09′-54′n) and its relationship to magmatic and tectonic processes on fast-spreading midocean ridges. *Earth and Planetary Science Letters*, **104**, 513–534.

Haymon RM, Fornari DJ, Vondamm KL *et al.* (1993) Volcanic-eruption of the midocean ridge along the East Pacific Rise crest at 9-degrees-45-52′n—direct submersible observations of sea-floor phenomena associated with an Eruption event in April, 1991. *Earth and Planetary Science Letters*, **119**, 85–101.

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.

Hickerson MJ, Dolman G, Moritz C (2006a) Comparative phylogeographic summary statistics for testing simultaneous vicariance. *Molecular Ecology*, **15**, 209–223.

Hickerson MJ, Stahl EA, Lessios HA (2006b) Test for simultaneous divergence using approximate Bayesian computation. *Evolution*, **60**, 2435–2453.

Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA-sequence data. *Genetics*, **132**, 583–589.

Hurtado L, Lutz R, Vrijenhoek R (2004) Distinct patterns of genetic differentiation among annelids of eastern Pacific hydrothermal vents. *Molecular Ecology*, **13**, 2603–2615.

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. *Biological Bulletin*, **210**, 140–157.

Jollivet D (1996) Specific and genetic diversity at deep-sea hydrothermal vents: an overview. *Biodiversity and Conservation*, **5**, 1619–1653.

Jollivet D, Chevaldonne P, Planque B (1999) Hydrothermal-vent alvinellid polychaete dispersal in the eastern Pacific. 2. A metapopulation model based on habitat shifts. *Evolution*, **53**, 1128–1142.

Jollivet D, Empis A, Baker MC *et al.* (2000) Reproductive biology, sexual dimorphism, and population structure of the deep sea hydrothermal vent scale-worm, *Branchipolynoe seepensis* (Polychaeta: Polynoidae). *Journal of Marine Biological Association of the United Kingdom*, **80**, 55–68.

Jolly MT, Viard F, Gentil F, Thiebaut E, Jollivet D (2006) Comparative phylogeography of two coastal polychaete tubeworms in the Northeast Atlantic supports shared history and vicariant events. *Molecular Ecology*, **15**, 1841–1855.

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: *Mammalian Protein Metabolis* (ed. Munro HN), pp. 21–123. Academic Press, New York.

© 2009 Blackwell Publishing Ltd

Knowlton N, Weigt LA (1998) New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **265**, 2257–2263.

Kuhner MK, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics*, **149**, 429–434.

Kumar S, Balczarek KA, Lai Z-C (1996) Evolution of the *hedgehog* gene family. *Genetics*, **142**, 965–972.

Kureth CL, Rea DK (1981) Large-scale oblique features in an active transform-fault, the Wilkes fracture-zone near 9-degrees-S on the East Pacific Rise. *Marine Geophysical Researches*, **5**, 119–137.

Le Pennec M, Hily A, Lucas A (1984) Structures gonadiques particulières d'un Mytilidae profond des sources hydrothermales du Pacifique oriental. *Comptes Rendus de l'Académie des Sciences de Paris, Série III*, **299**, 725–730.

Lessa EP, Cook JA, Patton JL (2003) Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary. *Proceedings of the National Academy of Sciences, USA*, **100**, 10331–10334.

Mammerickx J, Klitgord KD (1982) Northern East Pacific Rise—evolution from 25 My Bp to the present. *Journal of Geophysical Research*, **87**, 6751–6759.

Mammerickx J, Herron E, Dorman L (1980) Evidence for two fossil spreading ridges in the southeast Pacific. *Geological Society of America Bulletin*, **91**, 263–271.

Matabos M, Thiebaut E, Le Guen D *et al.* (2008) Geographic clines and stepping-stone patterns detected along the EPR in the vetigastropod *Lepetodrilus elevatus* reflects species crypticism. *Marine Biology*, **153**, 545–563.

Mullineaux LS, Wiebe PH, Baker ET (1995) Larvae of benthic invertebrates in hydrothermal vent plumes over Juan-De-Fuca Ridge. *Marine Biology*, **122**, 585–596.

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.

Naar DF, Hey RN (1989) Speed limit for oceanic transform faults. *Geology*, **17**, 420–422.

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

Plouviez S, Daguin C, Hourdez S, Jollivet D (2008) Juvenile and adult scale-worms, *Branchipolynoe seepensis*, in Lucky Strike hydrothermal vent mussels are genetically unrelated. *Aquatic Biology*, **3**, 79–87.

Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology & Evolution*, **16**, 37–45.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.

Sadosky F, Thiebaut E, Jollivet D, Shillito B (2002) Recruitment and population structure of the vetigastropod Lepetodrilus elevatus at 13 degrees N hydrothermal vent sites on East Pacific Rise. *Cahiers de Biologie Marine*, **43**, 399–402.

Shank TM, Fornari DJ, Von Damm KL *et al.* (1998) Temporal and spatial patterns of biological community development at nascent deep-sea hydrothermal vents (9°50′N, East Pacific Rise). *Deep Sea Research Part II: Topical Studies in Oceanography*, **45**, 465–515.

Skibinski DOF, Gallagher C, Beynon CM (1994) Mitochondrial-DNA Inheritance. *Nature*, **368**, 817–818.

Smith CI, Farrell BD (2005) Range expansions in the flightless longhorn cactus beetles, *Moneilema gigas* and *Moneilema armatum*, in response to Pleistocene climate changes. *Molecular Ecology*, **14**, 1025–1044.

Sunnucks P (2000) Efficient genetic markers for population biology. *Trends in Ecology & Evolution*, **15**, 199–203.

Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Tunnicliffe V (1991) The biology of hydrothermal vents: ecology and evolution. *Oceanography and Marine Biology. An Annual Review*, **29**, 319–407.

Tunnicliffe V, Embley RW, Holden JF *et al.* (1997) Biological colonization of new hydrothermal vents following an eruption on Juan de Fuca Ridge. *Deep-Sea Research Part I-Oceanographic Research Papers*, **44**, 1627.

Tyler PA, Pendlebury S, Mills SW *et al.* (2008) Reproduction of gastropods from vents on the East Pacific Rise and the Mid-Atlantic Ridge. *Journal of Shellfish Research*, **27**, 107–118.

Van Dover CL, Trask JL, Gross J, Knowlton A (1999) Reproductive biology of free-living and commensal polynoid polychaetes at the Lucky Strike hydrothermal vent field (Mid-Atlantic Ridge). *Marine Ecology-Progress Series*, **181**, 201–214.

Van Dover CL, German CR, Speer KG, Parson LM, Vrijenhoek RC (2002) Evolution and biogeography of deep-sea vent and seep invertebrates. *Science*, **295**, 1253–1257.

Young CR, Fujio S, Vrijenhoek RC (2008) Directional dispersal between mid-ocean ridges: deep-ocean circulation and gene flow in *Ridgeia piscesae*. *Molecular Ecology*, **17**, 1718–1731.

This study is a component of S.P.'s PhD project, which investigates comparative phylogeography of deep-sea hydrothermal vent species. Her research uses multiple species and markers to examine demographic processes and species dispersion along the East Pacific Rise. The PhD is co-supervised by D.J. and F.H.L. T.M.S. B.F., C.D. and F.V. made significant contributions to this study through the design of sampling strategies, field collections and editing that improved the manuscript, initially written by the first author.

© 2009 Blackwell Publishing Ltd

4. *Intérêts et limites de l'approche de phylogéographie comparée*

Les approches de phylogéographie monospécifique permettent de révéler des barrières aux flux de gènes ayant impacté l'espèce étudiée et de poser des hypothèses quant à l'histoire évolutive de cette espèce. Les apports de la phylogéographie comparée ajoutent une dimension supplémentaire en détectant des changements environnementaux majeurs (cataclysmes tels qu'une éruption volcanique majeure, changements climatiques, fermeture ou ouverture d'un passage entre 2 régions) ayant impacté l'ensemble de la faune et la flore associé, tant du point de vue de la vicariance que de l'histoire démographique.

4.1. Apports de la phylogéographie comparée entre espèces dans l'étude des processus de vicariance

L'estimation du temps de divergence T entre deux lignées répond à une équation, en apparence, simple : $T = D/2r$ où D correspond à la divergence génétique observée entre ces lignées et r, au taux de substitution du gène considéré. Il semble alors aisé, pour une espèce donnée, d'estimer la date d'isolement de deux lignées monophylétiques et d'en rechercher la cause dans l'histoire de l'espèce (e.g. causes environnementales associées à la mise en place de la barrière). Néanmoins, cette corrélation entre divergence génétique et mise en place d'une barrière géographique pose en réalité de nombreux problèmes (e.g. Rambaut & Bromham 1998, Welch et al 2005).

Tout d'abord, la corrélation temps de divergence/divergence génétique nécessite l'accès au taux de substitution du gène considéré. L'estimation d'un taux de substitution s'effectue généralement en comparant la divergence entre un ADN ancien issu de fossiles datés et un ADN actuel d'un même taxon (Shapiro et al 2004 : Rise and fall of the Beringian steppe bison, Science) ou par le biais de la divergence génétique entre lignées sœurs situées de part et d'autre d'une barrière géographique bien datée géologiquement (Knowlton & Weigt 1998, Chevaldonné et al 2002). Il peut être également obtenu de façon expérimentale en suivant l'accumulation des nouvelles mutations entre les lignées généalogiques d'une espèce donnée sur plusieurs générations (Holwell et al 2003). Néanmoins, cette technique est sujette à caution du fait que les mutations délétères sont purgées plus rapidement dans la divergence que dans le polymorphisme, donnant lieu à des

taux de mutation nettement plus élevés lorsqu'estimée par cette approche (Emerson 2007). Une fois estimé par l'intermédiaire d'une datation géologique précise, ce taux de substitution pourrait alors être utilisé pour estimer un temps de divergence entre lignées pour une autre zone géographique. Néanmoins, le taux de substitution est souvent variable, non seulement selon l'espèce étudiée, mais aussi au cours du temps de manière indépendante dans chaque lignée évolutive (deux lignées actuelles pouvant avoir évolué selon des taux de substitution différents à partir d'une même lignée mère : e.g. Ho 2009). Afin de corréler la divergence génétique observée au temps de divergence, il est donc nécessaire de vérifier l'hypothèse d'horloge moléculaire (Zuckerkandl & Pauling 1965) qui suppose un taux de substitution constant au cours du temps et entre lignées évolutives.

Un écart à la neutralité du gène considéré (balayage sélectif) peut également engendrer une mauvaise estimation du temps de divergence. En effet, la réduction du polymorphisme liée à de la sélection purifiante sur le gène considéré, engendrera ainsi une sous-estimation de la divergence entre clades par rapport à un coalescent neutre (voir l'introduction générale partie 3.2). La divergence estimée au gène sous sélection pourrait alors ne pas correspondre à l'événement ayant donné lieu à l'isolement des espèces ou des lignées intra-spécifiques.

Sous les hypothèses de l'horloge moléculaire et de l'accumulation neutre des mutations, la divergence génétique entre clades peut alors permettre de dater l'événement d'isolement de ces clades. Si ces clades sont sub-divisés géographiquement, alors cette datation peut être reliée à l'existence d'une potentielle barrière géographique aux flux de gènes. Une approche phylogéographique espèce par espèce permet, en comparant les différentes dates obtenues, d'appréhender si ces espèces présentent des dates de divergence proches, potentiellement issues d'un même événement d'isolement géographique. Dans le cas de la barrière équatoriale de l'EPR, les dates de divergence coïncident relativement bien même s'il existe une forte variance associée due à la stochasticité du processus de tri des allèles sur une période de temps aussi courte (1-2 millions d'années). Il est en effet évident que plus l'événement de spéciation est ancien et plus la proportion de gènes ayant atteint la monophylie réciproque est élevée, ce qui fournit une meilleure estimation inter-espèces (Neigel & Avise 1997). Cette estimation peut être néanmoins rapidement biaisée si le temps de divergence devient trop long et génère une saturation du signal phylogénétique au niveau du gène étudié (Avise 1994). L'approche comparative d'Hickerson et al (2006) avec le logiciel MsBayes, apparaît beaucoup plus informative. Sous l'hypothèse d'un taux de mutation constant au cours du temps (horloge moléculaire), elle permet de tester le

nombre d'évènements d'isolement génétique et de déterminer les dates les plus probables auxquelles ses évènements de vicariance ont eu lieu en imposant une histoire démographique propre à chaque lignée sœur.

De plus, la phylogéographie comparée permet de préciser le positionnement de la barrière aux flux de gènes lorsque les espèces étudiées possèdent des capacités de dispersion légèrement différentes (e.g. Maggs et al 2008, l'étude d'une seule espèce pouvant engendrer une mauvaise interprétation de la position géographique historique de cette barrière). L'approche comparative d'Hickerson et al (2006b) avec le logiciel MsBayes, est beaucoup plus informative. Elle permet de tester le nombre d'évènements d'isolement génétique et de déterminer les dates les plus probables auxquelles ces évènements de vicariance ont eu lieu.

4.2. Apports de la phylogéographie comparée dans l'étude de l'histoire démographique des populations

Outre les informations de vicariance, la phylogéographie comparée peut également s'avérer un outil puissant pour détecter des évènements démographiques et notamment de distinguer l'effet sélectif de l'effet démographique si ce dernier a été causé par une perturbation environnementale « globale » ayant affectée l'ensemble de la faune (e.g. glaciations, éruption volcanique, etc…). En effet, la topologie d'un coalescent à un gène donné pour une espèce donnée ne permet pas de dissocier les causes ayant engendré un écart à l'accumulation neutre des mutations dans le polymorphisme. Si les tests d'écarts à la neutralité (tests de Tajima, Fu & Li ou Fay & Wu) sont congruents entre plusieurs espèces pour une zone géographique donnée ou une lignée intra-spécifique donnée), il est alors peu probable que cet écart soit dû à des évènements sélectifs concomitants sur le gène mais plus à une expansion démographique ou un goulot d'étranglement récent(e) affectant toutes les espèces (Zink 2002). Dans notre cas (espèces hydrothermales de l'EPR), l'hypothèse la plus parcimonieuse est alors une modification de l'habitat ayant engendré un événement démographique particulier sur l'ensemble des espèces. Cette réduction de la taille des populations (goulot d'étranglement ou bottleneck) suivie d'une expansion démographique à partir des individus restants semble ici liée à une réduction de l'activité hydrothermale ou à une série d'épanchements de magma (volcanisme effusif) ayant

engendré une destruction importante de la faune associée. L'autre hypothèse d'une colonisation récente de l'EPR sud à partir d'une succession d'effets fondateurs du nord vers le sud est beaucoup plus difficile à considérer eu égard à la disparité des modes de dispersion observés chez les espèces étudiées et à l'absence d'une réelle migration asymétrique orientée du nord vers le sud au travers de la barrière équatoriale.

4.3. Limites de la phylogéographie comparée monogène

Si la comparaison de patrons phylogéographiques concordants entre espèces peut permettre de mieux évaluer la date de l'événement d'isolement, les études intégratives, telles que celles d'Hickerson et al (2006), nécessitent de comparer le même gène pour un ensemble d'espèces ayant la même histoire et reposent sur un modèle de spéciation avec changements démographiques indépendants. Or, l'obtention de séquences à un même gène sur plusieurs espèces peut s'avérer techniquement difficile en absence d'amorces dites « universelles ». Pour cette raison, les approches intégratives de phylogéographie comparée ont, pour le moment, été menées presque exclusivement sur le gène mitochondrial de la Cytochrome Oxydase I (Hickerson et al 2006a, Hickerson & Meyer 2008, Plouviez et al 2009 ; exception : Hickerson et al 2006b, étude menée sur des introns peu variables). Or, les gènes mitochondriaux présentent généralement une héritabilité uniquement maternelle (mais voir par exemple, Skibinski et al 1994 pour la double héritabilité mitochondriale). Ceci implique que les patrons phylogéographiques observés par cette approche intégrative ne reflètent que la distribution des lignées femelles. L'utilisation de gènes nucléaires (hérédité biparentale) pourrait donc complémenter ces approches. De même, le développement de nouvelles méthodes d'analyses permettant d'inclure plusieurs gènes dans les approches intégratives sur plusieurs espèces, pourrait minimiser les risques d'observer des patrons de distribution propres au gène étudié et non représentatif de l'espèce.

Chapitre 4 : Étude d'une barrière semi-permeable à la dispersion par une approche multi-locus sur plusieurs espèces

1. *Introduction à l'étude d'une barrière à la dispersion par une approche multi-locus*

L'utilisation d'un seul locus dans les études de phylogéographie pouvant conduire à des conclusions erronées, en inférant l'histoire du gène plutôt que l'histoire de l'espèce (Hudson 1983, Pamilo & Nei 1988, Maddison 1997), Avise & Ball (1990) soulignent la nécessité de tester la concordance du signal phylogéographique entre plusieurs locus indépendants afin de reconstruire l'histoire évolutive des populations. Ainsi, par exemple, si plusieurs locus regroupent de la même manière des populations différenciées génétiquement de part et d'autre d'une barrière à la dispersion préalablement supposée, cette concordance entre locus apporte du crédit à l'existence d'une telle barrière et permet éventuellement de mieux situer sa position géographique. Néanmoins, la stochasticité du processus de coalescence implique que, pour que plusieurs gènes présentent une différenciation génétique concordante à travers une barrière (accumulation de divergence sur des gènes différents), la cessation des flux de gènes doit être relativement longue (Avise & Ball 1990). Il n'en demeure pas moins vrai qu'une convergence dans la disparité des distributions alléliques (analyse des fréquences alléliques) peut être également observée à une échelle de temps moins grande et conduire aux mêmes conclusions si les effets dynamiques liés au fonctionnement des populations sont négligeables (effet « métapopulation »).

Le nombre optimal de locus pour estimer des paramètres de taille efficace (Pluzhnikov & Donnelly 1996, Felsenstein 2006, Carling & Brumfield 2007), de flux de gènes (Beerli & Felsenstein 1999) ou de taux d'expansion démographique (Kuhner et al 1998), dépend de la complexité de la dynamique de ces populations, de l'écart potentiel des gènes étudiés au modèle d'évolution neutre et, de l'histoire évolutive de l'espèce

considérée (Edwards & Beerli 2000, Hickerson et al 2007, Peters et al 2007). Felsenstein (2006) suggère néanmoins, qu'il est généralement plus informatif d'augmenter le nombre de locus indépendants considérés plutôt que le nombre d'individus ou la longueur des séquences utilisées.

Du fait de leur taille efficace plus faible liée à leur héritabilité généralement uniquement maternelle, les marqueurs cytoplasmiques (mitochondriaux mais également chloroplastiques chez les végétaux) ont une plus grande probabilité de montrer une structure phylogéographique prononcée que les marqueurs nucléaires (Moore 1995). Ainsi, quand l'ADN mitochondrial présente une monophylie réciproque entre deux populations, les gènes nucléaires peuvent n'arborer qu'une para- ou polyphylie (Fig. 4.1, Zink & Barrowclough 2008). L'accumulation et la fixation des mutations dans le polymorphisme ne dépend pas du taux de mutation du gène considéré, mais presque exclusivement de sa taille efficace (Zink & Barrowclough 2008).

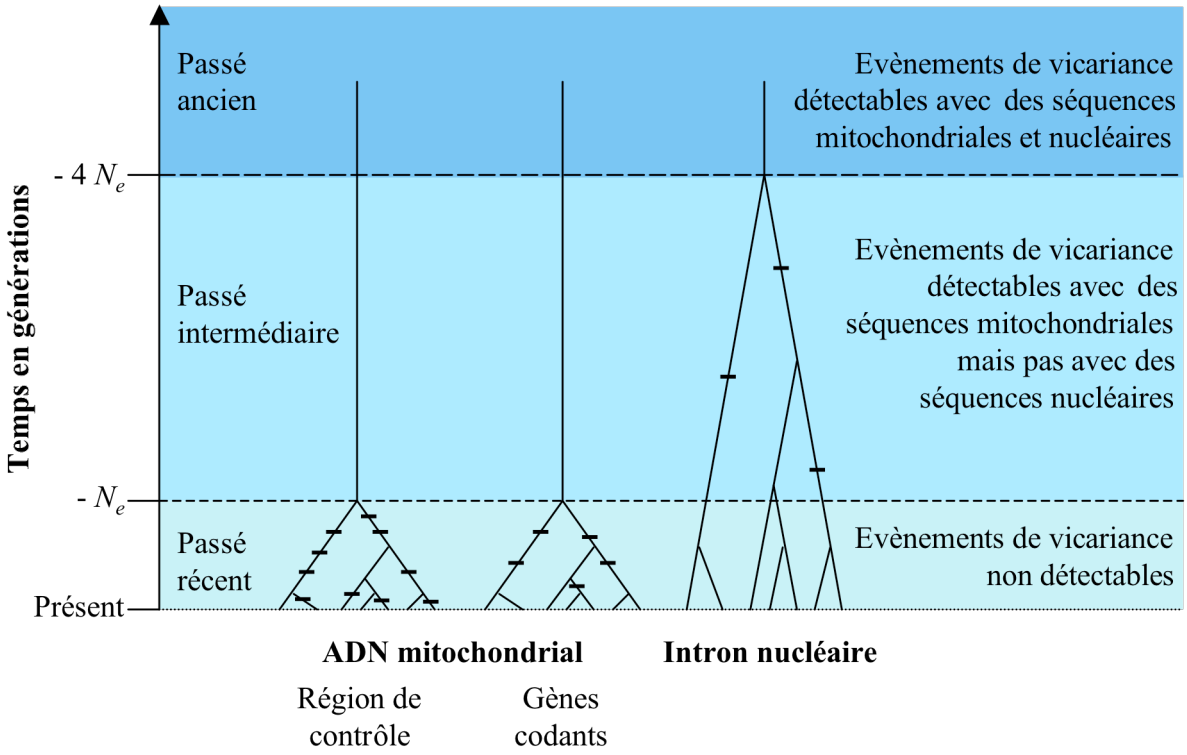


Fig. 4.1 Détectabilité des évènements de vicariance et coalescents de gènes mitochondriaux et nucléaires associés sous l'hypothèse neutre. Adaptée de Zink & Barrowclough (2008).
Les mutations sont représentées sur les coalescents, permettant de comprendre que le taux de mutations n'affecte pas le temps de coalescence.

Cependant, la structure phylogéographique observée par des marqueurs cytoplasmiques a également plus de chance d'avoir une origine purement stochastique, sous l'effet de la dérive génétique (les séquences échantillonnées dans les populations ont une chance plus grande d'apparaître sous une monophylie réciproque car la monophylie réciproque est plus rapidement atteinte pour ce type de marqueur à taille efficace Ne, Kuo et Avise 2005, Irwin 2002, Zink & Barrowclough 2008) et nécessite donc d'être confirmée par des locus nucléaires indépendants afin d'inférer l'histoire phylogéographique de l'espèce. De plus, les marqueurs à hérédité uniparentale ne tracent l'histoire que d'un sexe. Or, mâle et femelle peuvent parfois avoir subi des histoires évolutives différentes (e.g. migration différentielle entre sexes, philopatrie, Hoelzer 1997) et ainsi extrapoler cette histoire aux deux sexes peut engendrer des conclusions erronées.

Parce qu'ils reflètent à la fois l'évolution des mâles et des femelles, les gènes nucléaires peuvent donc améliorer la compréhension de l'histoire phylogéographique de

l'espèce étudiée, mais également permettre d'évaluer la perméabilité d'une barrière génétique via l'étude de l'hybridation potentielle entre lignées divergentes. Si un individu hybride est capable de donner une descendance (introgression), ces individus introgressés peuvent influencer l'histoire des lignées de différentes manières (cf. revue de Seehausen 2004) : (1) engendrer l'homogénéisation génétique de ces lignées, (2) le transfert de matériel génétique entre lignées, potentiellement favorisé par de la sélection positive, (3) un renforcement de l'isolement reproducteur entre lignées. Cependant, ces individus peuvent également être retenus dans une zone de contact secondaire entre les lignées divergentes (zone d'hybridation) suivant un équilibre entre la dispersion et la sélection contre les hybrides (Barton & Hewitt 1985). L'étendue géographique d'une zone d'hybridation dépend de cet équilibre et peut ainsi permettre de mieux comprendre l'impact d'une barrière génétique dans la structure des populations.

Dans le chapitre précédent, nous avons mis en évidence la présence d'un événement de vicariance partagé entre sept espèces (1,3 Ma) le long de la dorsale du Pacifique oriental sud laissant supposer l'existence d'une spéciation de type allopatrique en cours et, d'un second événement plus ancien (11,6 Ma) ayant conduit à une spéciation chez *L. elevatus* au niveau de la dorsale EPR nord (décalage géographique de la zone de contact secondaire). Néanmoins, l'utilisation du seul gène mtCOI ne nous a pas permis de déterminer avec précision l'emplacement de la barrière aux flux de gènes pour chacune de ces espèces. En effet, si *A. pompejana* et *E. vitrea* présentent une monophylie réciproque de part et d'autre de l'équateur, les lignées divergentes de *B. thermophilus*, par exemple, se distribuent selon un cline géographique le long de l'EPR, la lignée majoritairement distribuée au Nord n'étant pratiquement pas échantillonnée de 14°S à 21°33'S (2 individus à 17°25'S). La question de savoir si la présence de lignées mitochondriales divergentes traduit une spéciation en cours avec ou non la formation de zones de contact secondaire mérite également toute notre attention. De la même façon, sous l'hypothèse d'un contact secondaire (i.e. *L. elevatus, L. ovalis* ou *B. thermophilus*), il semble également important de déterminer si ces lignées sont capables de s'hybrider, de préciser la nature de l'hybridation et d'estimer le niveau d'introgression entre celles-ci. De plus, la détection préalable d'hybrides via des marqueurs allozymiques (Matabos et al 2008b) entre les espèces cryptiques de *L. elevatus* dans une zone de contact située à 9°50'N pose la question de l'étendue de cette zone d'hybridation (seules les populations de 13°N, 9°50'N et 17°25'S avaient été génotypées par ces auteurs). Dans ce présent chapitre, nous avons

donc choisi de comparer les patrons phylogéographiques contrastés révélés par le mtCOI à ceux obtenus sur des marqueurs nucléaires chez *A. pompejana*, *B. thermophilus* et *L. elevatus*, appartenant à des classes distinctes et pour lesquelles des banques d'ADN complémentaires nous permettaient d'accéder à de tels marqueurs (cf. Chapitre 2). Ces analyses phylogéographiques multi-locus seront présentées après une brève présentation de la taxonomie et des principaux traits d'histoire de vie des espèces cibles.

2. *Taxonomie et traits d'histoire de vie des espèces cibles*

La phase larvaire constitue le vecteur principal de colonisation des espèces à cycle bentho-pélagique (Pechenik 1999). Néanmoins, la capacité natatoire des larves étant extrêmement limitée, cette dispersion s'effectue essentiellement de manière passive par l'intermédiaire des courants de fond orientés dans l'axe du graben (Chevaldonné et al 1997). Ainsi, toute modification topographique de la dorsale (segmentation de la vallée axiale par des failles transformantes, microplaque) ou altération de la circulation océanique (inflations bathymétriques) peuvent modifier l'écoulement des eaux de fond et, engendrer des barrières à la dispersion. Ces barrières peuvent être responsables de véritables discontinuités génétiques, faisant du milieu hydrothermal un environnement propice à la spéciation par vicariance à l'échelle de la dorsale océanique.

2.1. *A. pompejana*

La famille des Alvinellidae est endémique des sources hydrothermales de l'océan Pacifique (Desbruyères & Laubier 1991) et se compose d'une douzaine d'espèces dont deux appartenant au genre *Alvinella* (Desbruyères et al 2006b) : *A. pompejana* et *A. caudata* (les autres étant du genre *Paralvinella*), trouvées uniquement sur l'EPR. La séparation phylogénétique des espèces sur la base de leurs isoformes enzymatiques a montré que les regroupements d'espèces ne se faisaient pas en fonction de leur proximité géographique mais en plutôt en fonction de l'habitat (notamment le preferendum thermique : Jollivet et al 1995) laissant supposer qu'une spéciation de type 'écologique' ait pu prédater une spéciation de type allopatrique (existence d'espèces jumelles entre EPR et dorsale nord Pacifique : Tunnicliffe 1991).

Le polychète tubicole *Alvinella pompejana* a été observé à des températures pouvant atteindre 105°C sur de courtes périodes (enroulement d'un individu sur une sonde indiquant 105°C), avec des températures à l'ouverture du tube de 20-45°C (Chevaldonné et al 2002). Cette espèce a ainsi été décrite comme l'un des métazoaires les plus tolérants aux températures élevées. *A. pompejana* est une espèce pionnière dans la colonisation de nouvelles sources hydrothermales (Jollivet 1993, Desbruyères et al 1998). Cette espèce présente des oocytes de 200 μm riche en vitellus. La fécondation des ovocytes est interne et fait suite à un appariement des mâles et des femelles au cours duquel le sperme du mâle est transféré à la femelle tête à tête. Les spermatozoïdes sont transférés dans une spermathèque et fécondent les ovocytes matures au moment de leur émission dans la colonne d'eau pour donner lieu à des larves lécithotrophes (Chevaldonné & Jollivet 1993, Chevaldonné et al 1996, Jollivet 1996, Pradillon & Gaill 2003, Fig. 4.2). Le ralentissement du développement embryonnaire en eau froide, détecté expérimentalement chez cette espèce, retarderait la métamorphose des larves et pourrait ainsi, en augmentant la durée de dispersion dans la colonne d'eau, permettre une dispersion sur de plus longues distances (dépendant alors des courants profonds circulant à l'axe de la dorsale : Pradillon et al 2001).

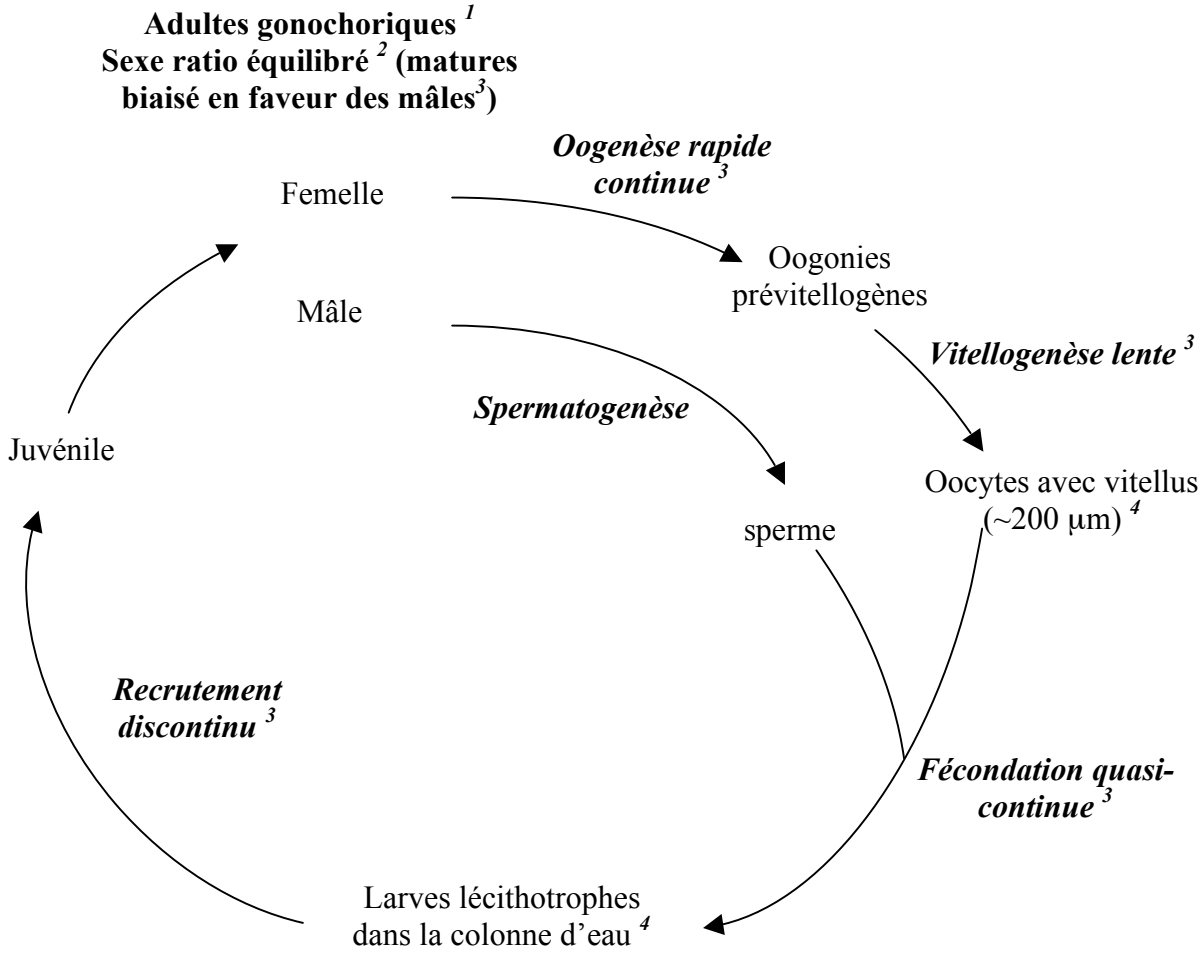


Fig. 4.2 Cycle de vie d'*Alvinella*.
La longueur des flèches n'est pas proportionnelle au temps.
1 Zal et al (1994), 2 Pradillon et al (2005), 3 Faure et al (2007), 4 Chevaldonné et al (1997)

2.2. *B. thermophilus*

Les bivalves du genre *Bathymodiolus* constituent l'un des taxons les plus cosmopolites et présentant la plus forte biomasse des sources hydrothermales (importantes moulières recensées en de nombreux points du globe sur quasiment tous les sites hydrothermaux connus exception faite de la dorsale Nord Pacifique (Desbruyères et al 2006b). Composés d'une vingtaine d'espèces (Won et al 2008), ce genre présente également une forte diversité spécifique et est également distribué au niveau des zones de

suintements froids. Les espèces du genre *Bathymodiolus* apparaissent clairement polyphylétiques et se trouvent mélangées à d'autres genres provenant d'habitats variés (carcasses de baleine, suintements et bois coulés) : *Gigantidas, Idas, Adipicola* (Samadi et al 2007, Lorion et al 2009). La branche évolutive dans laquelle se trouve *B. thermophilus* est, quant à elle, composée uniquement d'espèces appartenant au genre *Bathymodiolus* et pourrait avoir un ancêtre commun ayant vécu dans les zones de suintements froids du Golfe du Mexique (Faure 2008). Cette espèce peut être considérée comme une espèce-architecte et abritent de nombreuses espèces de macrofaune de petite taille (e.g. gastéropodes, polychètes) (Jollivet 1996).

Le cycle de vie des *Bathymodiolus* est présenté en figure 4.3. Les *Bathymodiolus* sont généralement gonochoriques ou hermaphrodites protandres, (e.g. *B. elongatus*, Le Pennec & Benniger 1997, *B. azoricus* Comtet et al 1999). La fécondation des gamètes s'effectue dans la colonne d'eau pour former des œufs (50 μm) puis des larves planctotrophes (Lutz et al 1980, Le Pennec 1988). Ces larves sont capables de remonter dans la colonne d'eau jusqu'à atteindre les couches éclairées de l'océan (Arellano & Young 2009), plus riches en nourriture. Il a été ainsi montrer que les larves pouvaient passer environ 4 mois dans la colonne d'eau entre leur émission (prodissochonque I : 80 μm) et leur recrutement (prodissochonque II : 400 μm) (Dixon et al 2007). Le recrutement a été suggéré discontinu chez *B. azoricus* par Comtet & Desbruyères (1998) mais continu chez *B. thermophilus* par Van Dover et al (2001b). Par cette remontée verticale, les larves pourront alors être soumises à des courants océaniques différents des courants profonds, par leur vitesse, mais également leur direction (Thomson et al 1990, Cannon & Pashinski 1997). L'accès à la dispersion via ces courants moins profonds pourrait ainsi favoriser la migration à travers une barrière physique et permettre un contact secondaire entre lignées divergentes.

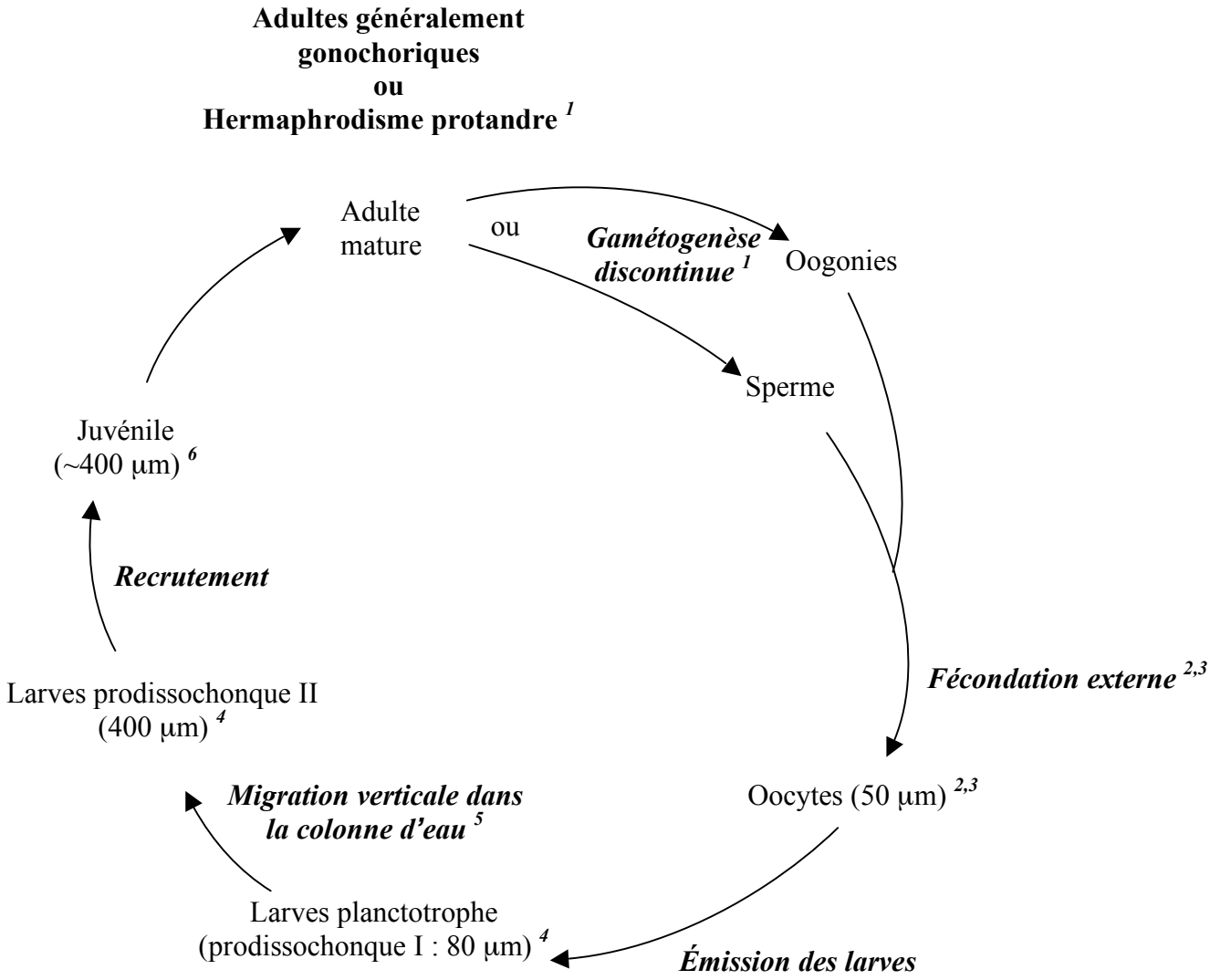


Fig. 4.3 Cycle de vie de *Bathymodiolus*
La longueur des flèches n'est pas proportionnelle au temps.
1 Le Pennec & Beninger (1997), 2 Lutz et al (1980), 3 Le Pennec (1988), 4 Dixon et al (2006), 5 Arellano & Young (2009), 6 Mullineaux et al (1998)

2.3. *L. elevatus*

La famille des Lepetodrilidae représente une des familles de gastéropodes patelliformes les plus abondantes des sources hydrothermales profondes. Certaines espèces ont également été décrites sur les bois coulés (*Lepetodrilus elevatus* McLean 22°N, EPR) et carcasses de baleines (*L. pustulosus* McLean, Monterey Bay, 36°N, Johnson et al 2008). De plus, cette famille compte également une forte diversité spécifique avec actuellement plus de 20 espèces morphologiquement distinctes (Warén & Boucher 2001, Desbruyères et al 2006b) dont la plupart appartiennent au genre *Lepetodrilus*. Venant s'ajouter à ce décompte, de nombreux complexes d'espèces (e.g. complexe « *fucensis* » : 2 espèces, Johnson et al 2006 ; complexe « *elevatus* » : 4 espèces, Matabos et al 2008b, Johnson et al 2008 ; complexe « *schrolli* » : 3 espèces, Johnson et al 2008) ont été mis en évidence au sein de ce genre. Originaires du Mésozoïque (McLean 1988), les Lepetodrilidae, et plus particulièrement le genre *Lepetodrilus*, ont donc connu depuis lors une véritable radiation évolutive, suscitant un grand intérêt pour l'étude des processus de spéciation ayant conduit à une telle diversification.

Les *Lepetodrilus* adultes se déplacent relativement lentement (< 0,1 cm/s, observation personnelle sur boîte de pétri sur *L. schrolli*) et ont une activité plutôt grégaire et sédentaire (empreintes des coquilles sur les tubes de vestimentifères dues au broutage des bactéries sur le tube (cf. *L. elevatus* et *L. pustulosus*). Leur transport via la faune vagile (phorésie), notamment par les crabes hydrothermaux (cf. Johnson et al 2008), constitue un vecteur de dispersion efficace mais limité à l'échelle du champ hydrothermal (dispersion de proche en proche souvent liée à la migration massive des crabes lors de l'extinction d'un site d'activité : Jollivet 1993). Au contraire, l'aptitude des larves à entrer en dormance ou à présenter d'importantes réserves vitellines leur permet de pallier au manque de nourriture dans l'eau de fond oligotrophe et donc de disperser sur de longues distances (Jollivet et al 2000 ; Pradillon et al 2001 ; Metaxas et al 2004). Le cycle de vie de *L. elevatus* est présenté en figure 4.4.

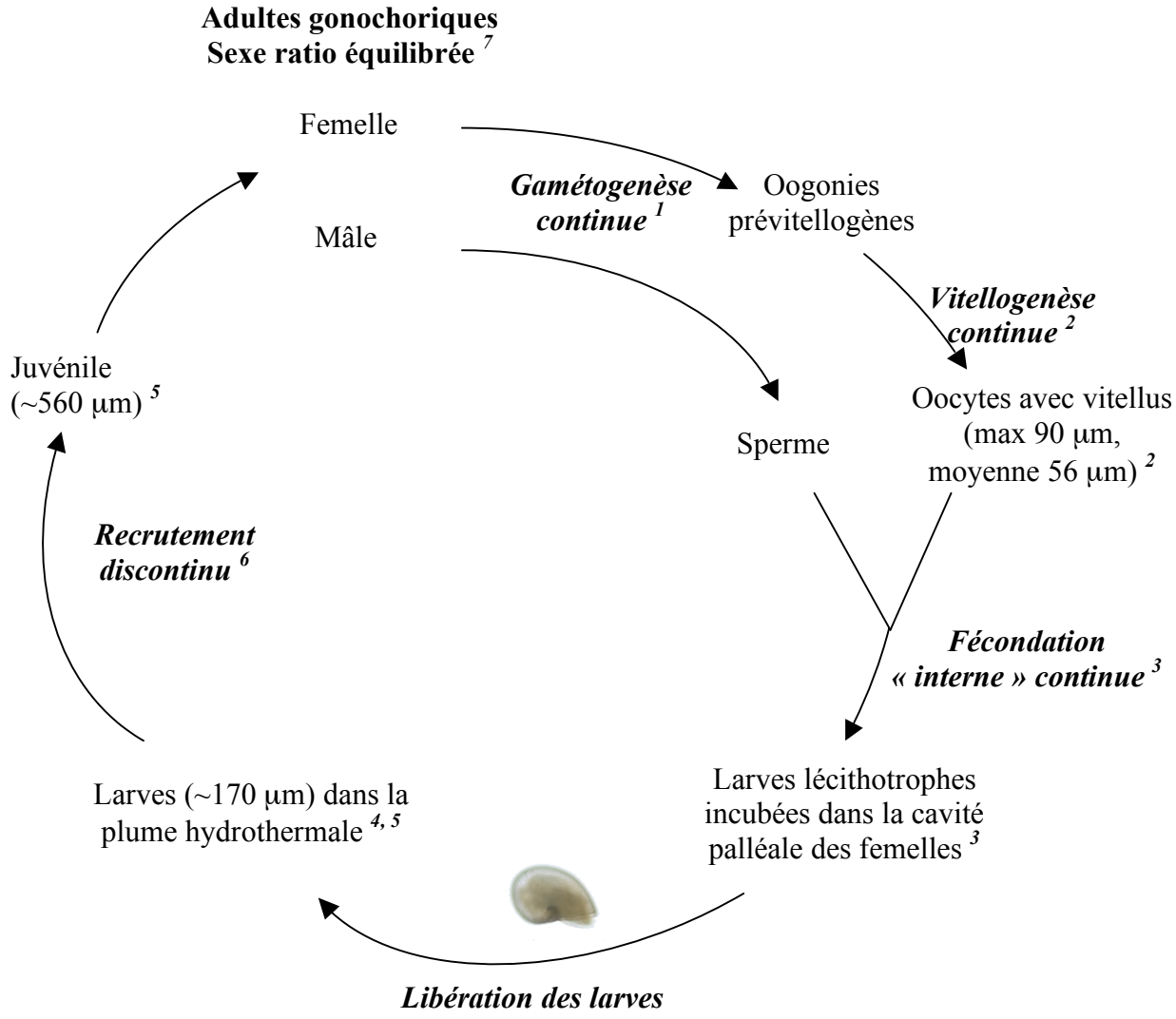


Fig. 4.4 Cycle de vie de *L. elevatus*.
La longueur des flèches n'est pas proportionnelle au temps.
1 Kelly & Metaxas (2007), 2 Tyler et al (2008), 3 Fretter (1988), 4 Mullineaux et al (1995), 5 Mullineaux et al (1996), 6 Mullineaux et al (1998), 7 Matabos et al (2008a).

Une espèce de *Lepetodrilus* donnée est généralement présente sur une seule dorsale ou un segment de dorsale et un habitat particulier, pouvant suggérer une diversification liée à des processus de vicariance. Néanmoins, certaines espèces présentent des aires de distribution superposées (e.g. *L. elevatus*, *L. pustulosus*, *L. ovalis*, *L. cristatus* sur l'EPR) au sein d'un segment de dorsale mais ayant un habitat clairement défini (tubes de vestimentifères, cheminée, basaltes ou coquilles de bivales). Dans certains cas, les espèces se répartissent différemment à l'échelle d'un site, selon leur preferendum thermique (Bates et al 2005, Mills et al 2007) mais aussi potentiellement selon la composition

physico-chimique du fluide et/ou des relations interspécifiques entre compétiteurs proches (Matabos et al 2008a). Ces différences de distribution liées à l'environnement local pourraient ainsi favoriser la spéciation écologique ou la spéciation parapatrique.

3. *Matériels et méthodes des espèces cibles*

L'extraction d'ADN, l'amplification et le séquençage du gène mtCOI ont été réalisés selon le protocole décrit dans Plouviez et al (2009). Les gènes nucléaires des différentes espèces cibles ont été amplifiés selon les conditions décrites dans l'annexe 3. Le clonage de ces gènes nucléaires a été réalisé par Marquage-Recapture de séquences (Bierne et al 2007 présenté en annexe 1 et chapitre 2) selon le même protocole de clonage pour *A. pompejana*, *B. thermophilus* et *L. elevatus*. Généralement, deux clonages comprenant chacun 8 individus ont été réalisés pour chaque population et 32 clones par clonage ont été séquencés. Concernant *B. thermophilus*, les séquences des individus échantillonnés à 7°25'S et 21°33'S ont été réalisées par Dr Baptiste Faure sur un clonage unique par population de 48 individus avec un effort de recapture de X1 (i.e. 96 clones séquencés par clonage). Pour *L. elevatus*, le nombre d'individus clonés a été augmenté dans la zone de contact (9°50'N), ainsi qu'à 21° 33'S (où une troisième espèce cryptique de *L. elevatus* a été observée sur la base d'une divergence mitochondriale) : soient 32 individus à 9°50'N et 24 à 21°33'S. L'effort de capture, quel que soit le nombre d'individus, est de X2 (e.g. 32 séquences par clonage comprenant 8 individus). Les séquences ont ensuite été traitées selon le protocole décrit dans la partie 'Matériels et Méthodes' du chapitre 2 afin d'éliminer la recombinaison *in vitro* et de ne garder qu'un gène supposé unique et diploïde.

4. *Analyses et résultats sur les espèces cibles*

4.1. *A.pompejana*

4.1.1. Résumé des résultats

Malgré une microstructure génétique significative de type chaotique au sein d'un champ hydrothermal donné (i.e. 13°N/EPR), des études allozymiques menées sur le polychète hydrothermal *A. pompejana* au nord de l'EPR (de 21°N à 9°50'N) n'ont pas

permis de détecter de réelle structure géographique entre ces champs hydrothermaux (Jollivet et al 1995, Jollivet 1996), suggérant de fortes capacités de dispersion chez cette espèce ou l'action de processus sélectifs homogénéisant. Cependant, un modèle flux-propagule basé sur la biologie de cette espèce suggère que, si les larves migrent dans la colonne d'eau durant 15 à 30 jours, la renverse périodique des courants de marée ne permet pas une migration efficace inter-champs de celles-ci (Chevaldonné et al 1997), impliquant une différenciation génétique des populations entre ces champs par simple dérive. L'absence de structure génétique à l'échelle de l'EPR Nord pourrait alors s'expliquer par

(1) un manque de résolution des marqueurs enzymatiques,

(2) une sélection balancée sur ces marqueurs maintenant une homogénéité des fréquences allozymiques directement liée à la nature de l'habitat hydrothermal, comme le proposent Piccino et al (2004) pour le locus phosphoglucomutase (PGM), celui-ci présentant un allèle 90 thermorésistant plus avantageux dans les populations nouvellement fondées ou les conditions thermiques sont plus élevées mais en équi-fréquence avec l'allèle 100 dans les populations plus « agées »,

(3) un déplacement de l'activité hydrothermale le long de l'axe (10 mètres par année : Watremez et al 1990, Kervévan 1991) pouvant, de manière transitoire, permettre la connectivité des champs hydrothermaux par rapprochement et ainsi empêcher la différenciation génétique de s'établir dans le temps (modélisé par Jollivet et al 1999).

Le séquençage du gène mtCOI sur des individus issus de populations réparties sur l'ensemble de la dorsale EPR a permis de révéler la présence d'une barrière aux flux de gènes à l'Equateur (Hurtado et al 2004, Plouviez et al 2009), donnant lieu à deux clades divergents quasi-monophylétiquement réciproques entre les individus du Nord et du Sud de l'EPR. Nous avons donc cherché à déterminer l'impact de cette barrière datée à 1,3 Ma sur les lignées mitochondriales divergentes en utilisant à la fois les variations en fréquence des marqueurs enzymatiques précédemment utilisés uniquement dans la partie Nord de l'EPR et l'histoire évolutive des allèles (séquences) de trois gènes nucléaires (dont le gène PGM pour lequel une sélection des formes 90 et 100 est suspectée).

Si les valeurs de différenciation génétique obtenues sur les marqueurs allozymiques sont proches de celles obtenues à micro-échelle (au sein d'un champ hydrothermal : 13°N/EPR), six de ces marqueurs permettent néanmoins de confirmer la présence d'une barrière aux flux de gènes à l'Equateur.

D'une manière plus frappante, l'analyse des parentés alléliques par des approches de coalescence (Fig. 4.5) montre de manière très claire l'existence d'une divergence entre les lignées alléliques du gène de la PGM (réseau d'allèles révélant 4 clades divergents : deux échantillonnés majoritairement au Nord et les deux autres échantillonnés majoritairement au Sud) et, à une échelle moindre entre 2 lignées d'allèles du gène codant pour la GlobineX (2 lignées d'allèles plus faiblement divergentes présentant des fréquences alléliques différentes dans les deux régions géographiques). Pour ce dernier gène, ces différences de fréquences se traduisent par une différenciation génétique significative entre les populations de part et d'autre de cette barrière. À l'inverse, aucune divergence n'a été mise en évidence dans la phylogénie d'allèles associée au gène S-Adenosyl Homocystéine Hydrolase (SAHH) et aucune différenciation génétique Nord/Sud n'a été détectée sur la base des fréquences alléliques observées.

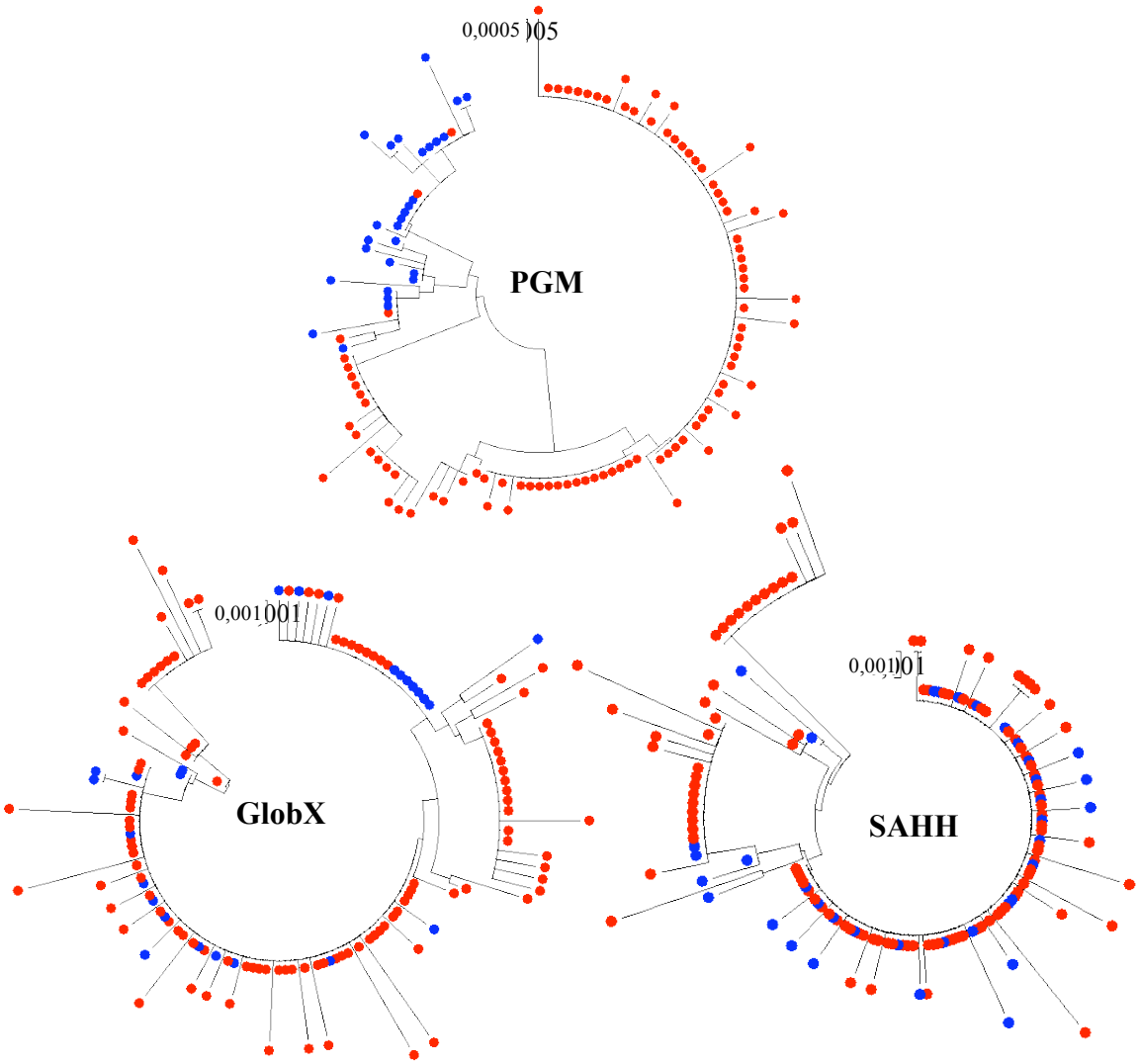


Fig. 4.5 Arbres phylogénétiques réalisés selon la méthodes des plus proches voisins sur les gènes nucléaires d'*A. pompejana* à partir des jeux de séquences de qualité 2 (2 allèles maximum par individu).
En bleu et rouge : séquences d'individus échantillonnés respectivement au Nord et au Sud de l'EPR.

L'utilisation d'un modèle d' « Isolement avec Migration » de Hey & Nielsen (2007) a révélé la présence d'une migration asymétrique à travers l'Equateur, du Nord vers le Sud de la dorsale avec un temps de divergence moyen estimé sur les gènes mtCOI et nucléaires qui concorde avec celui proposé par l'approche multispécifique au gène mtCOI (1,3 Ma, Plouviez et al 2009, Fig. 4.6). Ce sens de migration pourrait expliquer la différenciation génétique observée des champs hydrothermaux 7°25'S-14°S versus 17°25'S-21°33'S sur les allozymes, détectée via le logiciel Structure (mise en évidence de 3 groupes génétiquement différenciés : Prichard et al 2000), le groupe 2 composé essentiellement d'individus de 7°25'S et 14°S pourrait indiquer que ces latitudes correspondent à une zone de transition entre les populations du Nord de la dorsale (majoritaire dans le groupe 1) et les populations du Sud de la dorsale (majoritaire dans le groupe 3). L'existence de Fis significatifs au niveau des locus enzymatiques sur les sites 7°25S, 14°S et 17°25S renforce cette hypothèse de zone de contact secondaire. De plus, une analyse RFLP/typage d'un indel basée sur les sites fixés des 4 lignées alléliques du gène PGM ont permis de détecter des individus hybrides/introgressés majoritairement aux sites 7°25'S à 14°S, indiquant que ces sites pourraient correspondre à une zone d'hybridation entre les lignées mitochondriales.

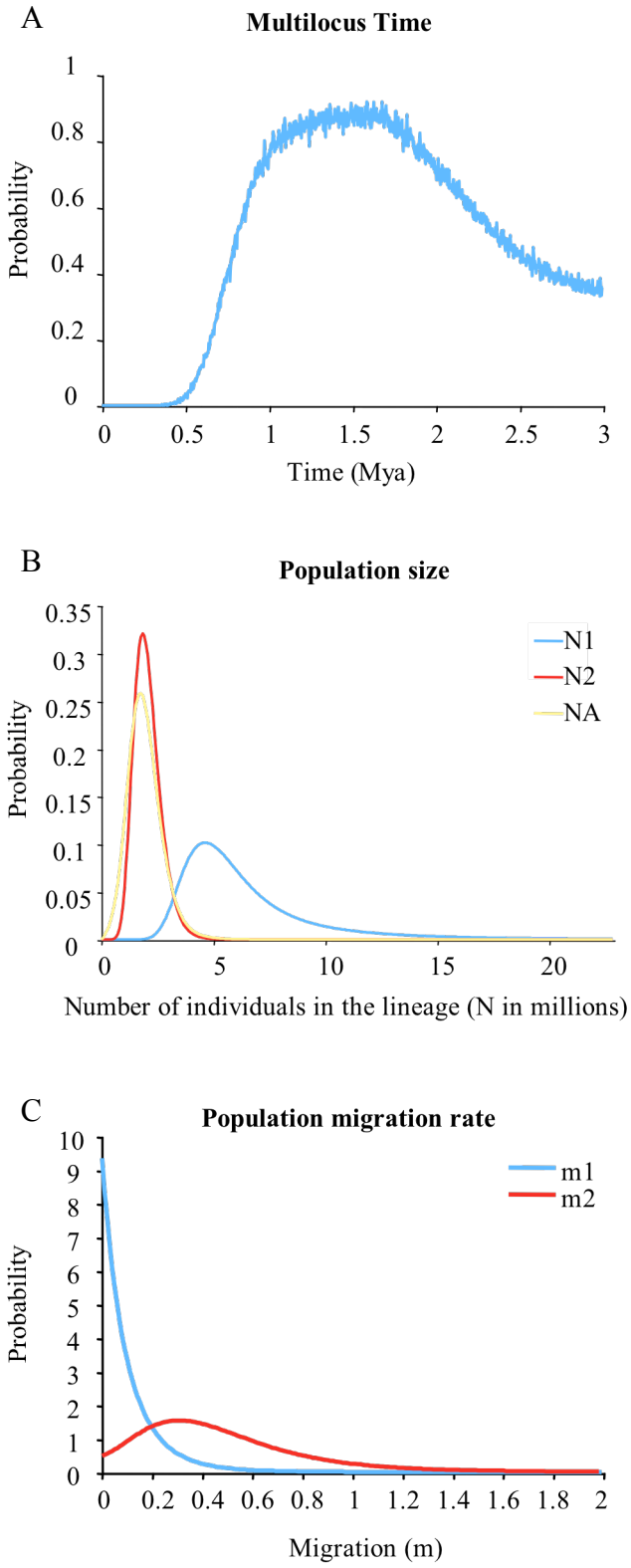


Fig. 4.6 Distributions des probabilités postérieures marginales du temps de divergence entre populations Nord et Sud (A), de la taille efficace de la population ancestrale (NA) et de ces populations filles (Nord N1, Sud N2), (B), et des taux de migration directionnels (dans le sens du temps: m1 Sud vers Nord, m2 Nord vers Sud) (C), estimées par un modèle d'Isolement avec Migration via IMa.

Enfin, les D de Tajima et les F* de Fu & Li, significatifs au Sud de la dorsale pour l'ensemble des gènes, ainsi qu'une analyse visant à discriminer l'effet d'une expansion démographique d'un balayage sélectif, ont permis de confirmer que l'accumulation non neutre des mutations dans le polymorphisme de la lignée sud est bien due à une expansion démographique telle que précédemment suggérée par Plouviez et al (2009). Cette dernière analyse a également mis en évidence un balayage sélectif sur le gène SAHH, pouvant expliquer l'absence de différenciation génétique entre les populations Nord et le Sud de l'EPR lorsque l'on considère ce gène uniquement. Ce dernier résultat valide l'hypothèse du passage facilité d'un allèle avantageux à travers une barrière physique à la dispersion lorsque les espèces ont encore peu divergé : une situation assez similaire à celle trouvée par Faure et al (2008) entre *Mytilus edulis* et *M. galloprovincialis*. Bien que n'étant pas détecté comme tel par le logiciel Sweep_bott (Galtier et al 2000), la topologie du coalescent du gène GlobX pourrait également être apparentée à un balayage sélectif plus récent (et donc moins détectable) sur ce gène.

La comparaison des génotypes RFLP des 4 allèles du gène PGM avec les génotypes allozymiques a permis de corréler les clades 3 et 4 aux allèles 90 et 100 trouvés au Sud de l'EPR et ainsi de montrer qu'une mutation Q/E non-synonyme engendrant une différence de charge nette entre les protéines sur l'exon 4 expliquait la différence de mobilité électrophorétique observée entre les allèles 90 et 100 au Sud. Cependant, cette corrélation n'étant pas trouvée au Nord, il semble que l'allélisme 90/100 observé dans les populations nord soit associé à une mutation affectant un autre site exonique du gène. Cette homoplasie de charge souligne ainsi une limite des locus enzymatiques dans la détection de structure génétique et l'intérêt de coupler ces marqueurs à d'autres types de marqueurs, en particulier basés sur le polymorphisme de séquences. L'approche multilocus effectuée a ainsi permis d'apporter du crédit aux hypothèses d'un effet vicariant allopatrique et d'une expansion démographique dans la lignée Sud proposées par l'approche multispécifique au gène mtCOI, mais également de révéler le caractère semi-perméable, au moins transitoirement, de cette barrière à l'Equateur pour *A. pompejana*, malgré la monophylie réciproque observée sur le mtCOI.

4.1.2. Article

Plouviez *et al. BMC Evolutionary Biology* 2010, **10**:220
http://www.biomedcentral.com/1471-2148/10/220

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Determining gene flow and the influence of selection across the equatorial barrier of the East Pacific Rise in the tube-dwelling polychaete *Alvinella pompejana*

Sophie Plouviez[1,2,3], Dominique Le Guen[1,2], Odile Lecompte[4], François H Lallier[1,2], Didier Jollivet[1,2*]

Abstract

Background: Comparative phylogeography recently performed on the mitochondrial *cytochrome oxidase I* (*mtCOI*) gene from seven deep-sea vent species suggested that the East Pacific Rise fauna has undergone a vicariant event with the emergence of a north/south physical barrier at the Equator 1-2 Mya. Within this specialised fauna, the tube-dwelling polychaete *Alvinella pompejana* showed reciprocal monophyly at *mtCOI* on each side of the Equator (9°50'N/7°25'S), suggesting potential, ongoing allopatric speciation. However, the development of a barrier to gene flow is a long and complex process. Secondary contact between previously isolated populations can occur when physical isolation has not persisted long enough to result in reproductive isolation between genetically divergent lineages, potentially leading to hybridisation and subsequent allelic introgression. The present study evaluates the strength of the equatorial barrier to gene flow and tests for potential secondary contact zones between *A. pompejana* populations by comparing the *mtCOI* gene with nuclear genes.

Results: Allozyme frequencies and the analysis of nucleotide polymorphisms at three nuclear loci confirmed the north/south genetic differentiation of *Alvinella pompejana* populations along the East Pacific Rise. Migration was oriented north-to-south with a moderate allelic introgression between the two geographic groups over a narrow geographic range just south of the barrier. Multilocus analysis also indicated that southern populations have undergone demographic expansion as previously suggested by a multispecies approach. A strong shift in allozyme frequencies together with a high level of divergence between alleles and a low number of 'hybrid' individuals were observed between the northern and southern groups using the *phosphoglucomutase* gene. In contrast, the *S-adenosylhomocysteine hydrolase* gene exhibited reduced diversity and a lack of population differentiation possibly due to a selective sweep or hitch-hiking.

Conclusions: The equatorial barrier leading to the separation of East Pacific Rise vent fauna into two distinct geographic groups is still permeable to migration, with a probable north-to-south migration route for *A. pompejana*. This separation also coincides with demographic expansion in the southern East Pacific Rise. Our results suggest that allopatry resulting from ridge offsetting is a common mechanism of speciation for deep-sea hydrothermal vent organisms.

* Correspondence: jollivet@sb-roscoff.fr
[1]Université Pierre et Marie Curie-Paris 6, Laboratoire Adaptation et Diversité en Milieu Marin, Roscoff, France
Full list of author information is available at the end of the article

Background

Many population genetic models have been proposed to describe how individuals disperse between geographically separated localities including n-island models [1,2], stepping-stone [3] or isolation-by-distance [4,5] models, and have subsequently been adapted to a metapopulation context to integrate population dynamics through time (i.e. local extinction and recolonisation rates; reviewed in [6,7]). Nevertheless, a physical barrier to dispersal can disrupt the relationship between genetic differentiation and migration rate: all models show that the establishment of a long-term physical barrier to dispersal abruptly reduces or stops gene flow between populations and leads to a rapid change in allele frequencies on either side of the barrier that may be dampened or accelerated depending on the type of selection at specific loci. This often leads to the fixation of mutations (divergence) and reciprocal monophyly for many loci, provided that no lineage extinction or population admixture occurs [8]. Secondary contact events can produce genetic incompatibilities and clinal allelic distributions whose shapes depend primarily on the strength of selection against hybrids [9].

Deep-sea hydrothermal vents represent a patchily distributed habitat (sometimes separated by thousands of kilometres) suggestive of a one-dimensional stepping-stone model over 60 000 kilometres of globe-encircling ridge crests [10]. Because of their dependence on 'hot' sulphidic fluids from vent chimneys, deep-sea hydrothermal vent species display an island-like distribution along ridges. This habitat distribution raises questions about vent species' dispersal among sites and the putative occurrence of physical barriers to dispersal [11,12]. It has been hypothesised that hydrothermal vent species disperse primarily in their local neighbourhood [13], as in the Kimura and Weiss [3] one-dimensional stepping-stone model, and thus follow an isolation-by-distance model [11]. However, in most cases, allozyme studies have failed to detect isolation-by-distance among populations of hydrothermal vent species (e.g. [14,15]), suggesting that an n-island model [1] is a better description of gene flow in vent systems.

Amongst vent species, the deep-sea hydrothermal vent polychaete *Alvinella pompejana*, which only lives at the top of hydrothermal-vent chimneys of the East Pacific Rise (EPR) between 38°S and 27°N (Guaymas Basin), has been described as a pioneer species because it is the first species to colonise new hot-sulphide edifices [16]. The larval development of this tube-dwelling worm is still largely unknown, but large yolk-rich eggs (150-200 µm), suggest that larval development is lecithotrophic and could be delayed in cold (1.5-2°C) water away from vent fields [12]. This worm species reproduces via internal fertilization and has nearly continuous gametogenesis, maximising its reproductive effort [17]. During mating, spermatozoids are transferred to the female in a spermatecae, and multiple paternity is possible [18]. This kind of reproductive mode is known to strongly favour larval retention in the immediate surrounding of the source population, which may be advantageous in unstable and patchy environments [19]. Given this set of reproductive characteristics and the recurrent extinction and recolonisation of vent edifices, Jollivet *et al.* [20,21] have proposed that this species follows a propagule-flux model of dispersal (i.e. (re)colonisation of a previously extinct or new vent site from the closest populations, [22]). A hydrodynamic model predicts that effective migration rates - accounting for *A. pompejana*'s larval biology - cannot be maintained across vent fields if larvae travel more than 15-30 days in the water column [12]. Nevertheless, using allozymes this species lacks any clear genetic structure from latitudes 9°50'N to 21°N despite a microgeographical population differentiation at the scale of vent fields (i.e. tens of kilometres) [23]. These results suggest that the potential for larvae to disperse may be higher than previously thought provided that no physical barrier to gene flow exists along the ridge system [24]. Accordingly, Pradillon *et al.* [25] reported an experiment where *A. pompejana* embryos can arrest their development in cold water, potentially delaying their metamorphosis during dispersal from vents. To better explain the absence of isolation-by-distance in the northern EPR, three hypotheses have been proposed: (1) allozymes lack the power to detect genetic differentiation between populations; (2) balancing selection - due to the homogeneity of vent conditions along the whole ridge system - maintains homogeneous allozyme frequencies; or (3) displacement of habitat patches over geological times promote metapopulation dynamics, thereby maintaining gene flow [24]. The first hypothesis was tested by Audzijonyte and Vrijenhoek [26] who suggested that undersampling biased some of the previous analyses. The second hypothesis was supported by Piccino *et al.* [27] who observed balancing selection on the phosphoglucomutase (PGM) locus with one particularly thermostable allele being more frequent in young populations. Interestingly, this study also showed allele frequencies were homogenous at the metapopulation scale, proving support for the third hypothesis. The third hypothesis was tested by a model which showed that bursts of colonisation events following the emergence of new vent sites along with population reconnections were sufficient to prevent genetic differentiation between distant sites given limited dispersal of *A. pompejana* larvae [21]. The model predictions were however, based on the possibility that new vent

sites could occur across physical barriers such as transform faults.

Recently, genetic divergence of nearly 1% was observed at the mitochondrial *cytochrome c oxidase subunit I* (*mtCOI*) gene between the northern and southern EPR populations of *A. pompejana* [28,29], suggesting the occurrence of an impermeable equatorial barrier to gene flow. Furthermore, Plouviez *et al.* [29] demonstrated that this barrier affected most EPR vent fauna 1 to 2 Mya, giving credence to the role of transform faults in provoking vicariant events between vent communities and allopatric speciation. In the present study, an extended analysis of allozyme and *mtCOI* haplotype distributions over the geographic range of the species, including the southern EPR, was performed to better address the ability of allozymes to detect genetic differentiation and to discriminate between the three hypotheses described above. Three additional nuclear genes, including the *PGM* locus, were used to evaluate the strength of this equatorial barrier to gene flow and to test for potential secondary contact. We also attempted to quantify migration and/or introgression rates between the previously isolated northern and southern EPR populations to disentangle the effects of selective and demographic processes contributing to the genetic break.

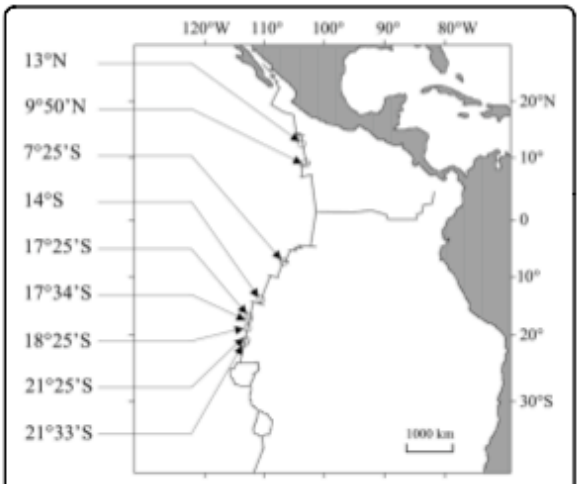


Figure 1 Map of the south-east Pacific and *Alvinella pompejana* sampling locations along the East Pacific Rise. Dots represent sampled hydrothermal vent fields. The bracket between 9°50'N and 7°25'S indicates the Quebrada, Discovery and Gofar transform faults from north to south, respectively.

Methods

Collection

Alvinella pompejana polychaetes were collected with the manned submersible *Nautile* during two oceanographic cruises in 1999 (HOPE) and 2002 (PHARE) at northern EPR sites (Figure 1) and one cruise in 2004 (BIOSPEEDO) at southern EPR sites from latitudes 7°25'S to 21°33'S. Onboard (R/V *L'Atalante*), samples were measured at the fourth setigerous segment, sexed, and cut in two: the cephalic portion was frozen immediately at -80°C for allozyme analyses and the body was preserved in 95% ethanol for DNA analyses.

Allozyme genotyping

Allozymes, encoded by nine enzyme systems, were examined from seven hydrothermal vent field samples (Table 1), following the protocols of Pasteur *et al.* [30]. Proteins from each frozen gill tissue were extracted according to the method in Piccino *et al.* [27]. Allozyme electrophoreses were conducted on 12% starch gels using three different buffer systems: (1) Tris-citrate pH 8.0 (TC 8.0) for phosphoglucomutase (PGM, E.C. 5.4.2.2), mannose phosphate isomerase (MPI, E.C. 5.3.1.8), 6-phosphogluconate dehydrogenase (6PGD, E.C. 1.1.1.44) and glucose phosphate isomerase (GPI, E.C. 5.3.1.9); (2) Tris-HCl pH 8.5/8.2 (THCl 8.5) for triose phosphate isomerase (TPI, E.C. 5.3.1.1), leucine amino peptidase (LAP, 3.4.11.1) and hexokinase-1 (HK, 2.7.1.1); and (3) Tris-citrate pH 6.7/6.3 (TC 6.7) for acid phosphatase (ACP, E.C. 3.1.3.2) and isocitrate dehydrogenase (IDH, E.C. 1.1.1.42).

DNA sequencing

Genomic DNA was extracted using a CTAB extraction procedure following the protocol described in Plouviez *et al.* [29]. *A. pompejana mtCOI* sequences from this previous study and the sequences of the outgroup *A. caudata* were obtained with the 'universal' LCOI and HCOI primers [31] using the previously described sequencing protocol. Specific primers (additional file 1) were designed to amplify partial exon-intron sequences of three nuclear genes, either identified during transcriptome sequencing (the *S-adenosylhomocysteine hydrolase* (*SAHH*) and *globin X* (*GlobX*) genes [32]), or from the complete *phosphoglucomutase* gene sequence in *A. pompejana* previously obtained using degenerate primers (D.J., unpublished data). A non-synonymous change was discovered in exon 4 of *PGM* and specifically targeted in subsequent analyses. Sizes of the DNA fragments and their exon-intron structures (boundaries) are given in additional file 2.

Polymerase chain reactions (PCR) and cloning were performed using the mark-recapture (MR) method developed by Bierne *et al.* [33] with 4-nucleotide tails added to the 5' end of each primer. Each individual was amplified by PCR separately with a different combination of 5'-tailed primers. PCR were conducted in a 25 µl volume that included 1X buffer (supplied by manufacturer), 2 mM $MgCl_2$, 0.05 mM (0.25 mM for *PGM*) of each dNTP, 0.48 µM (0.4 µM for *PGM*) of each primer

Table 1 Location of individuals and sample size for allozymes (n_{allo}), *mtCOI* (n_{mtCOI}), and nuclear genes.

Vent field	Geographic position	Depth (m)	n_{allo}	n_{mtCOI}	n_{GlobX}	n_{SAHH}	n_{PGM}	n_{RFLP}
13°N	12°43-50'N	2560-2700	40	67	19	17	20	74
	103°53-57'W				**10**	**11**	**12**	
9°50'N	9°31-51'N	2500-2585	0	10	7	20	8	0
	104°15-18'W				**4**	**15**	**6**	
7°25'S	7°25'S	2735-2752	17	14	14	13	15	12
	107°47-49W				**9**	**8**	**11**	
14°S	13°59'S	2623-2632	15	20	16	16	15	16
	112°29'W				**9**	**14**	**11**	
17°25'S	17°25'S	2578-2590	34	30	15	16	10	15
	113°12'W				**10**	**9**	**8**	
17°34'S	17°35-36'S	2591-2597	19	36	18	15	12	17
	113°15'W				**12**	**10**	**10**	
18°25'S	18°26-37'S	2636-2680	52	29	17	17	14	14
	113°23-24'W				**11**	**14**	**10**	
21°25'S	21°25'S	2804-2831	12	30	16	16	13	22
	114°16-17'W				**12**	**11**	**10**	
21°33'S	21°33'S	2771-2846	23	25	14	20	17	11
	114°16-18'W				**10**	**14**	**13**	

n_{allo}: number of individuals genotyped for allozymes, n_{mtCOI}: number of individuals sequenced for the *mtCOI* gene, n_{GlobX}, n_{SAHH}, n_{PGM}: number of recaptured sequences for each nuclear gene using a maximum of two consensus sequences (alleles) per individual with the number of recaptured individuals given in bold, n_{RFLP}: number of genotyped individuals for *PGM* RFLPs.

(additional file 1), 0.5 U of Taq polymerase (Thermoprime plus), 5 µl of template DNA and sterile H_2O. Thermal cycling parameters used an initial denaturation step at 94°C for 3 min, followed by 40 cycles at 94°C for 30 s, T_a for 20 s and 72°C for 2 min 30 s, before a final 10 min extension at 72°C. As per the MR protocol, eight PCR-products were pooled together per cloning experiment with two cloning experiments per vent locality (16 individuals per sampled site). PCR products were isostochiometrically mixed, purified using QIAQuick™ columns, and ligated into a pGEM-T vector (pGEM-T cloning kit, Promega, Madison, WI, USA). For each cloning experiment, 32 positive clones were sequenced on both strands at Genoscope (Evry) with universal plasmid primers (SP6 and T7), leading to a total number of approximately 3500 nuclear sequences. Sequences from all genes were proof-read using CodonCode Aligner 2.0.6 http://www.codoncode.com/aligner/. Sequence alignments were initially performed with ClustalW [34] in BioEdit version 6.0.6 [35] and adjusted manually. For nuclear genes, *in vitro* recombinants between individuals from the same cloning set were detected by their abnormal combination of 5'-tails and removed from the dataset. Multiple recaptures also allowed us to discard intra-individual *in vitro* recombinants (PCR recombination of heterozygous alleles) and artefactual/somatic mutations.

Detecting a putative barrier to gene flow

Linkage disequilibrium between allozyme loci was examined for each locality and for the complete dataset using Genetix version 4.05.2 [36] to detect redundant information. For each locality, allele frequencies were then calculated for each locus and departures from Hardy-Weinberg equilibrium were estimated using the Weir and Cockerham [37]'s estimator *f* and tested with 1000 permutations. To detect putative barriers to gene flow, genetic differentiation was estimated along the EPR from a multilocus Fst as calculated using Weir and Cockerham's θ [37], adding populations one at a time from the most southern locality (21°33'S) to 13°N. At each step, the F index was tested with 1000 permutations of the dataset. Exact G test was performed using Genepop version 4.0.10 [38] for all populations.

Isolation-by-distance was tested along the whole EPR using a Mantel Spearman test [39] with 5000 permutations using Genepop version 4.0.10 [38]. Because previous studies [28,29] detected a barrier to gene flow across the Equator for the *mtCOI* marker, this test was also performed without the 13°N population. To better understand population structure, the number of genetically-differentiated groups (*K*) was evaluated using individual assignments to these *a priori* groups based on their multilocus allozyme genotypes from an admixture model with a 150 000 burn-in length period followed by 50 000 MCMC repetitions. The most appropriate number was obtained by testing *K* values from 1 to 15 using the Structure program (version 2.2) [40]. Convergence of the Markov chain simulations was checked for stationarity by monitoring ten iterations.

Average recapture rates were 0.60, 0.74 and 0.63 for *GlobX*, *SAHH* and *PGM* genes, respectively. For all

statistical analyses based on recaptured sequences, one consensus allele was randomly selected per individual to avoid any putative bias due to the cloning method. Allele diversity *Hd*, nucleotide diversity π_N, and Watterson's theta θ_W were estimated for northern and southern EPR sites separately with DNAsp 4.10.3 [41]. For each gene, allele networks were constructed using Network 4.5.1.0 [42] to search for divergence across the Equator (all recaptured alleles were used). Genetic differentiation indexes (ϕ_{st}) were computed using Arlequin 3.1 [43] from nucleotide sites by grouping populations from either side of the Equator and departures from zero (no differentiation) were tested using 1000 permutations.

Because most coalescent-based approaches impose strong assumptions on the absence of recombination, recombinants were detected by the Hudson and Kaplan [44]'s four-gamete test, using DNAsp 4.10.3, and discarded from the dataset for the IMa and Sweep_bott analyses.

Detection of gene flow between the two geographic groups of *A. pompejana* using an isolation-with-migration model

Population sizes, migration rates, and the time of population splitting were estimated using the isolation-with-migration model (IMa software, [45]) by grouping populations from each part of the geographical barrier (i.e. 9° 50'N and 13°N on one side and the remaining southern populations on the other side). This model allowed us to calculate posterior density probabilities of these three parameters assuming that migration could still occur after the splitting event of an ancestral population into two descendant populations. To check convergence of the Markov chain Monte Carlo (MCMC) with the true stationary distribution, multiple runs were performed using different starting points and autocorrelation between parameter values was assessed over the course of the runs. Upper bounds of the uniform prior distributions used for the final run were: q = 30 (prior population size parameters), m = 1 (migration rates) and t = 2 (divergence time). A swapping procedure was also used to enhance the mixing of chains using a geometric heating scheme with 15 parallel chains. Inheritance scalars were assigned to 1 for nuclear genes (autosomal) and 0.25 for *mtCOI* to adjust for their expected population sizes. The Hasegawa-Kishino-Yano model [46] was chosen to allow for multiple substitutions at the same site.

To better estimate the date since the splitting event, IMa parameters of divergence between populations (t_{IMa}) were converted into years. Because mutation rates have not yet been estimated for alvinellid polychaetes, with the exception of the *mtCOI* gene [47] which is known to have a higher mutation rate compared to nuclear genes, a geometric mean was derived from *mtCOI*, as well as *PGM* and *SAHH* for which a mutation rate has been obtained from the hydrothermal vent gastropod *Lepetodrilus fucensis* across the Blanco transform fault [48] and the hydrothermal vent mussels *Bathymodiolus thermophilus* across the Easter microplate [49], respectively. The *GlobX* gene was not used because there is no adequate calibration for its mutation rate. Before calculating the geometric mean, an average mutation rate was assessed per locus per generation accounting for different lengths of the three genes (*mtCOI* 507 bp, *PGM* 940 bp, *SAHH* 531 bp), assuming one generation per year. The geometric mean $U = 0.64 \times 10^{-6}$, close to the value previously found using eight genes (*mtCOI* and 7 nuclear genes, $U = 0.63 \times 10^{-6}$) on *B. thermophilus* [49], was then used to recalibrate the divergence time from IMa to time in years ($T = t_{IMa} / U$).

Detecting putative 'hybrids' between the north and south EPR using the PGM gene

Because substantial divergence has been observed between the northern and southern EPR clades at the nuclear *PGM* gene, fixed mutations were used to detect and locate 'hybrids' between these two geographic regions (heterozygous individuals with one allele from the northern clade and one allele from the southern clade or individuals sampled in a geographic region with alleles typical of the clade of the other region). A restriction fragment length polymorphism (RFLP) analysis combined with genotyping insertion-deletions (indels) was performed to create a composite *PGM* genotype for each individual. Among the four clades observed, clade 3 displayed two restriction sites for the *Rsa*I enzyme whereas the three remaining clades only had one. Clade 4 displayed two restrictions sites for *Ssp*I whereas the three remaining clades have three sites. Clades 1 and 2 differed by several fixed substitutions for which no restriction site was observed and the presence of a 9 bp indel. *Rsa*I and *Ssp*I digestions were performed on *PGM* PCR products following manufacturer's protocols and the presence/absence of the indel was genotyped using new nested primers defined from each side of the indel. Individuals were amplified in 15 µl solutions that included 1X buffer (supplied by manufacturer), 1.25 mM $MgCl_2$, 0.2 mM each dNTP, 0.1 µM of fluorescent-labelled (IRD700™) forward primer, 0.27 µM of unlabelled forward primer, 0.33 µM of unlabelled reverse primer (additional file 1), 0.5 U of Taq polymerase (Thermoprime plus), 4 µl of template DNA and sterile H_2O. Thermal cycling parameters used an initial denaturation at 94°C for 2 min, followed by 30 cycles at 94°C for 30 s, annealing temperature (T_a, see additional file 1) for 30 s and 72°C for 25 s, before a final 10 min extension at 72°C. PCR products were then electrophoresed on a denaturing acrylamide 41 cm gel in a Li-Cor NEN Global IR2 DNA analyzer.

Detecting demographic and/or selective events from coalescence trees

Because gene genealogies contain information on demographic and/or selective events, neighbour-joining trees (distances obtained by the maximum composite likelihood) were performed using Mega 4.0.2 [50], separately on northern and southern EPR populations and rooted with an *A. caudata* sequence for non-ambiguous portions of each gene. Departures from neutral expectations were assessed by estimating Tajima (*D*, [51]), Fu and Li (*F**, [52]) and Fay and Wu (*H*, [53]) indexes in each group of populations to test for putative demographic and/or selective events. A multi-locus HKA test [54] was computed for each group with one sequence of *A. caudata* used as an outgroup using the HKA software (J. Hey's web page: http://lifesci.rutgers.edu/~heylab/HeylabSoftware.htm#HKA). This test compares divergence and polymorphism at several loci to determine if any loci display a departure from neutral evolution. McDonald Kreitman tests (MK, [55]) were also performed to test for selection. We tested for the fit of either a bottleneck or a selective sweep model for all four loci (including *mtCOI*) in northern and southern EPR populations (excluding introgressed alleles) using Sweep_bott [56] with 100 000 first-step iterations, 1 000 000 second-step iterations and 20 optimisation processes with a theta range from 1 to 30. This test compares three distinct coalescent models (including a null constant-size model: M1) using likelihood ratio tests (LRTs) and estimates whether the strength and the time since bottleneck are sufficiently similar between loci to suggest a historical bottleneck (M2 model). If there is a departure from the M1 model but no detected bottleneck, then the selective sweep model is considered as more probable. A strong selective sweep at a single locus can significantly reduce the fit to a bottleneck model resulting in false rejection of a bottleneck hypothesis. To better discriminate the bottleneck and the sweep models, departure from the M1 model was tested individually for each locus (called M3 model). If all loci showed a departure from M1 model, then a bottleneck would occur even if the sweep model was found more probable using the multilocus approach. If only a few loci gave a departure from this model, then the bottleneck was ruled out and the selective sweep model selected.

Results

Equatorial barrier to gene flow

Linkage disequilibria were not significant for any pair of loci in any population ($p > 0.05$ using the Garnier-Géré and Dillman [57] test), indicating that each locus behaves independently. The number of alleles ranged from 3 (GPI, IDH, ACP, LAP, HK) to 5 (TPI, see: additional file 3), however most loci displayed only one to 3 alleles at moderate frequency (> 5% in the total population) and were shared among all populations. In contrast, rare alleles were often private (additional file 3). Common alleles reversed in frequency at the TPI locus across the Equator, with the 13°N site showing opposite trends in allele frequencies compared to the 7°25'S site (Figure 2). A similar result was observed at the PGM locus but with a nearly complete replacement of allele 100 by allele 78 and the frequency of allele 90 slightly decreasing from north to south. Significant heterozygote deficiencies (Fis > 0, see additional file 3) were not linked to any particular locus and were not systematically observed at any location. However, a more detailed examination of Fis values at loci with two or three common alleles indicated that the highest values were found at 14°S and 17°25'S.

The multilocus estimate of Weir and Cockerham's θ [37] over all populations was not significantly different from zero ($\theta = 0.097$, $p > 0.05$, 1000 permutations) but Fst significantly differed from zero with the Robertson and Hill's correction ([58], $RH = 0.060$, $p < 0.05$, 1000 permutations). The exact G test for genic differentiation was significant for the overall populations ($p < 0.05$). Tests for isolation-by-distance showed a significant correlation between the genetic distance of allozymes $\theta/(1-\theta)$ and geographic distance across the entire EPR region (Mantel Spearman test: $p = 0.014$) but was not significant when the 13°N site was removed from the dataset ($p = 0.115$). Partitioning genetic structure using the Bayesian clustering method of Pritchard *et al.* [40] identified three genetically distinct groups present in different proportions within each vent field. The first group was primarily made up of individuals from 13°N, whereas the second group clustered

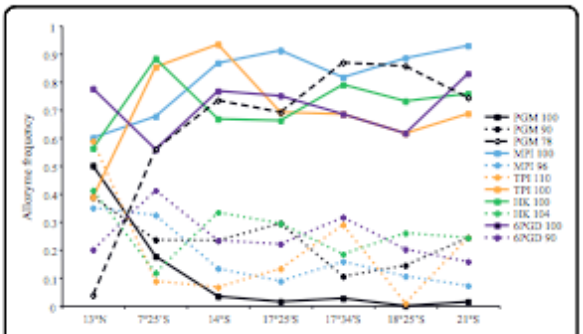


Figure 2 Allozyme frequency clines along the East Pacific Rise. Frequencies of the most frequent alleles found in *A. pompejana* populations for PGM, MPI, TPI, HK and 6PGD allozyme loci are plotted according to latitude. Each colour represents a distinct allozyme locus. Individuals from 9°50'N were not used in the analysis simply because they have been only preserved in alcohol.

Remarque: le code génétique mitochondrial de *Drosophila* a été utilise pour le mtCOI: cf annexe 4

individuals mostly from 7°25'S-14°S, and the third group clustered 17°25'S to 21°33'S individuals.

DNA sequence variation among the three nuclear genes (*SAHH*, *GlobX* and *PGM*) and *mtCOI* were analysed in the northern and southern populations of *A. pompejana* across the Equator (Table 2, Genbank accession numbers: HM183084-HM183493). As shown by Plouviez *et al.* [29], the network of *mtCOI* haplotypes displayed two reciprocally monophyletic clades (~1% divergence) in the northern and southern EPR, respectively (Figure 3). Similarly, the *PGM* locus possessed two sets of two divergent clades that were geographically structured (average northern/southern EPR divergence = 0.76%): clades 1 and 2 were mainly found in the northern EPR, whereas clades 3 and 4 were distributed in the southern EPR (Figure 3). Populations from the northern and southern EPR were significantly different from each other at this locus ($ø_{st}$ = 0.18, $p < 0.01$). Although populations appeared to be genetically differentiated across the Equator at the *GlobX* locus ($ø_{st}$ = 0.04, $p < 0.05$), these populations were genetically homogeneous across regions at the *SAHH* locus ($ø_{st}$ value = 0.00, $p > 0.05$).

Testing an isolation model with migration between the two geographic groups of *A. pompejana*

The isolation-with-migration (IM) model of Hey and Nielsen [45] was used to test the occurrence of historical gene flow between northern and southern populations of *A. pompejana* or recent hybridisation and introgression following secondary contact. The estimated IM parameters are presented in additional file 4. Using all four genes, the split between the two groups of populations showed the highest likelihood values at divergence times of around 1-2 Mya (Figure 4A). IMa calculates migration rate based on coalescent analysis by looking from the present backward in time reversing the direction of migration when looking forward in time. The magnitude of migration from the north to the south (m_n) and from the south to the north (m_s) shows that gene flow occurred predominantly from the north to the south with no migration being detected in the opposite direction (Figure 4B).

Single-locus estimates of migration also exhibited asymmetry from the north to the south, although the parameters for *PGM*, *SAHH*, and *GlobX* were not significantly different from zero on the basis of their 90% highest posterior density (HPD) interval. This was clearly suggested by the allele networks for *PGM* and *GlobX* (Figure 3) for which a few alleles from northern clades were sampled in the southern EPR region. The mean time of migration events from the north to the south ranged from 25 000 to 300 000 years ago (Table 3), indicating that introgression occurred well after the initial isolation (~1.6 Mya) but not recently. The mean time of migration events (mtime) at *PGM* was, however, ten times lower than the two other loci suggesting more recent introgression (~23 000 years) at this locus.

Detection of hybrids at PGM between northern and southern populations

Based on the *PGM* network (Figure 3), we detected three individuals sampled on the southern EPR with at least one allele from a northern clade (two from 14°S and one from 7°25'S), but with the southern mitochondrial type. Owing to fixed substitutions and indels between the four *PGM* clades, all individuals collected at these sites were genotyped by combining RFLP and intron length polymorphisms. By comparing RFLP/indel genotypes and mitochondrial haplotypes of each individual, four 'hybrids' (33.3%) between the northern and southern groups were detected at 7°25'S, two (16.7%) at 14°S and only one (1.7%) at 13°N. No 'hybrids' were found at the other latitudes.

Table 2 Nucleotide sequence variation in *Alvinella pompejana*

Gene	EPR locality	*n*	*Hd* (SD)	π_N (SD)	θ_W (SD)	*D*	*F**
mtCOI							
	northern	108	0.911 (0.014)	0.009 (0.000)	0.014 (0.004)	-0.983^{NS}	-2.164^{NS}
	southern	210	0.618 (0.040)	0.002 (0.000)	0.015 (0.004)	-2.564***	-4.330**
GlobX							
	northern	14	0.890 (0.060)	0.003 (0.000)	0.0034 (0.0017)	-0.872^{NS}	-1.504^{NS}
	southern	73	0.860 (0.031)	0.0054 (0.001)	0.012 (0.004)	-1.721^{+}	-4.107**
SAHH							
	northern	26	0.471 (0.119)	0.002 (0.001)	0.004 (0.002)	-1.989*	-2.7696*
	southern	80	0.604 (0.064)	0.0031 (0.001)	0.011 (0.003)	-2.202**	-4.484**
PGM							
	northern	18	0.921 (0.030)	0.0056 (0.001)	0.007 (0.003)	-0.693^{NS}	-1.437^{NS}
	southern	74	0.725 (0.055)	0.0042 (0.001)	0.009 (0.003)	-1.836*	-3.950**

n, number of sequences; *Hd*, Haplotype diversity, π_n nucleotide diversity; θ_w, Watterson's theta per site from the number of segregating sites, *D*, Tajima's *D* index; *F**, the Fu and Li index; *SD*, standard deviation; ^{NS}: $P > 0.10$, $^{+}$: $P < 0.10$, *: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$.

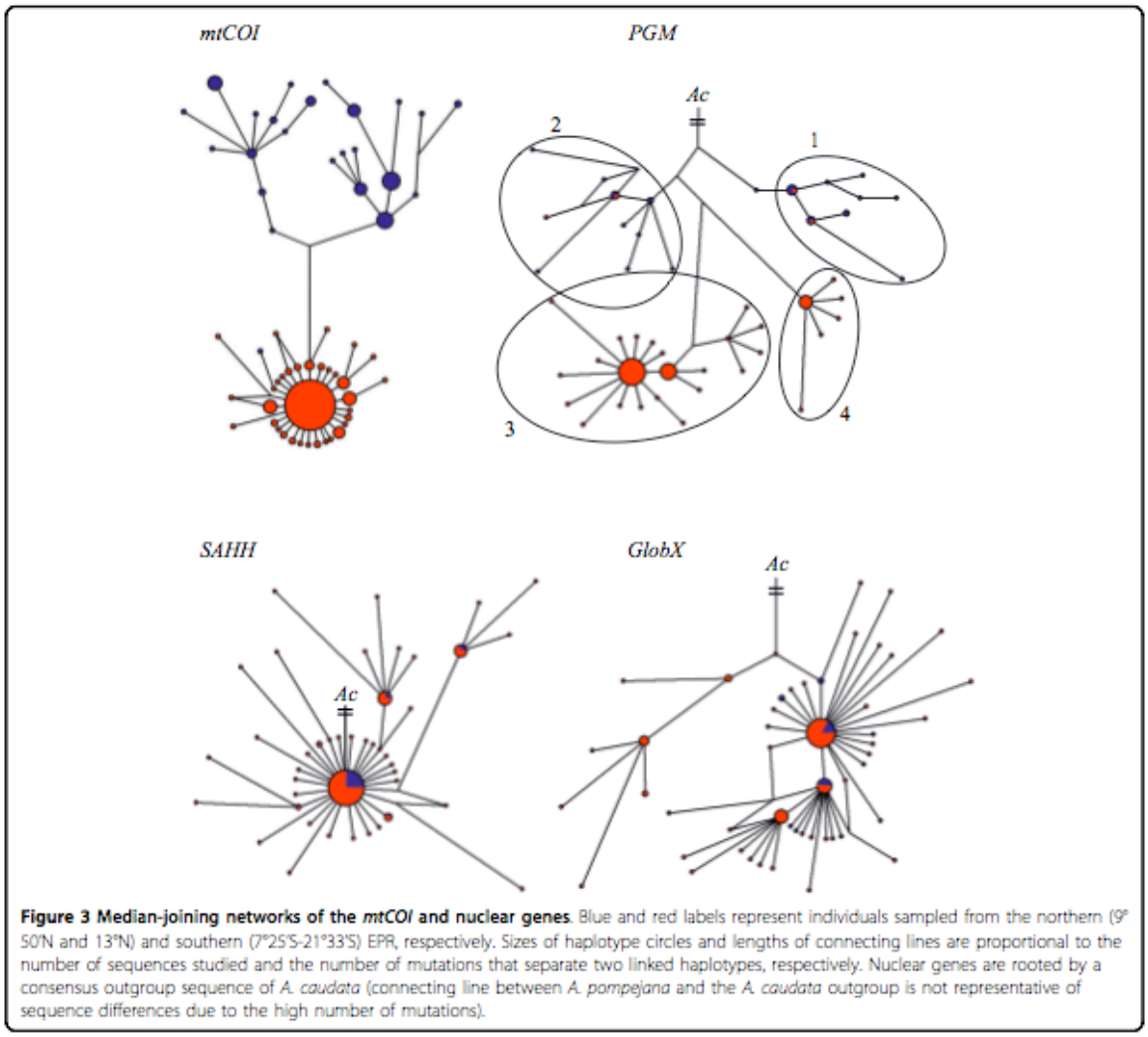


Figure 3 Median-joining networks of the *mtCOI* and nuclear genes. Blue and red labels represent individuals sampled from the northern (9° 50'N and 13°N) and southern (7°25'S-21°33'S) EPR, respectively. Sizes of haplotype circles and lengths of connecting lines are proportional to the number of sequences studied and the number of mutations that separate two linked haplotypes, respectively. Nuclear genes are rooted by a consensus outgroup sequence of *A. caudata* (connecting line between *A. pompejana* and the *A. caudata* outgroup is not representative of sequence differences due to the high number of mutations).

Detecting demographic changes and/or selective effects

Tajima (*D*) and Fu and Li (*F**) indexes were significantly negative for all loci in the south, but only *SAHH* displayed significant values in the north. All genes exhibited departures from neutral equilibrium between drift and mutation in the southern EPR, but not in the northern EPR. Similarly, for each nuclear gene, Fay and Wu tests were not significantly different from the neutral expectation in the northern populations, but were significantly negative in the southern populations.

The *PGM* neighbour-joining tree displayed a topology similar to a gene experiencing balancing selection (i.e., the maintenance of 'old' lineages), whereas trees from the two other nuclear genes (*SAHH* and *GlobX*) had more star-like topologies typical of a selective sweep (Figure 5). The HKA multilocus test was significant using all genes (probability from χ^2 distribution < 0.00), indicating that at least one gene has experienced positive selection (or gene hitchhiking). After excluding *mtCOI*, the HKA test became nonsignificant ($p\ \chi^2 > 0.05$), suggesting this gene could be experiencing strong purifying selection or a sweep. However, McDonald-Kreitman tests failed to reject neutral expectations at any locus ($p > 0.05$) when comparing *A. pompejana* polymorphism to the divergence of the *A. caudata* outgroup.

The Sweep_bott analyses performed on northern and southern groups of populations using the four loci are summarised in Table 4. When the analysis was performed over all genes, LRTs indicated that the bottleneck (M2) and selective sweep (M3) models both showed a significantly better fit with the dataset than

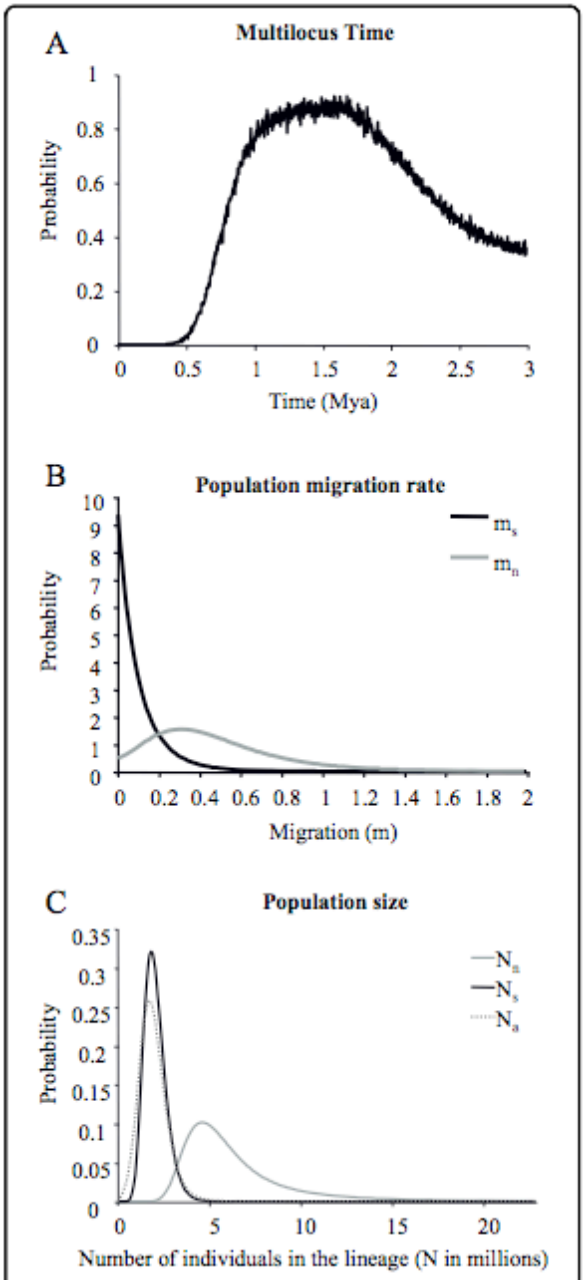


Figure 4 Marginal posterior probability distribution for parameters estimated by the isolation-with-migration model. Marginal posterior probabilities are estimated for (A) divergence time between the northern and southern EPR, (B) the forward-in-time migration rates between the northern and southern EPR, where m_s represents south-to-north migration and m_n north-to-south and (C) effective population sizes of the ancestral population (N_a) and the two sister populations (i.e. N_n and N_s for the northern EPR population and the southern EPR population, respectively). Values estimated using IMa and 90% highest posterior density intervals of these parameters are presented in additional file 4.

Table 3 Number ($m_{\#}$) and mean time (mtime) of migration events, obtained from IM analyses

Locus	$m_{\#s}$	$m_{\#n}$	$mtime_s$ (My)	$mtime_n$ (My)
mtCOI	0	0	-	-
SAHH	0	0-7	-	0.347
GlobX	0	0-7	-	0.388
PGM	0	1	-	0.023

Forward in time, labels (s) and (n) correspond to the south-to-north and north-to-south orientation of migration, respectively. For *SAHH* and *GlobX* genes, $m_{\#}$ values correspond to the range of the posterior probability distribution without considering m = 0.

the constant-size (M1) model. However, it was not possible to discriminate between the M2 and M3 hypotheses. When the analysis was performed on each gene individually, LRTs for M3 were all significantly better when compared with the M1 model for the southern populations, suggesting that a bottleneck might have occurred in this region. When the same analysis was done on the northern populations, LRTs for M3 were only significant for the *SAHH* gene, indicating that in addition to a bottleneck this gene may have undergone a selective sweep, at least in the northern populations. In effect, the M2 and M3 models are not mutually exclusive if a gene has experienced a moderate sweep in a bottlenecked population.

Discussion

Emergence and strength of the equatorial barrier to gene flow

A genetic break at the *mtCOI* gene between the northern and southern EPR populations has been identified in previous studies by using a large number of hydrothermal vent species [28,29]. For some species, such as the gastropod *Lepetodrilus ovalis* and the bivalve *Bathymodiolus thermophilus*, a clinal distribution of the northern and southern *mtCOI* haplotypes was observed and attributed to the occurrence of a secondary contact zone between the northern and southern EPR [29]. In contrast, *A. pompejana* exhibited a clear equatorial break at this locus, raising the question of the permeability of this barrier to gene flow. The strength of the break between the northern and southern populations of *A. pompejana* appears to have significantly affected both allozyme and nuclear loci despite the recent formation of the barrier. Although the *SAHH* gene exhibited very little differentiation across this barrier, both the *GlobX* and *PGM* loci displayed a clear geographic separation of alleles. Divergence between northern and southern clades was estimated at ~1-2 Mya using the IMa multilocus estimation with very low migration rates. This estimate is in agreement with the shared vicariant event estimated at ~1.3 Mya from a series of vent species using approximate Bayesian computation

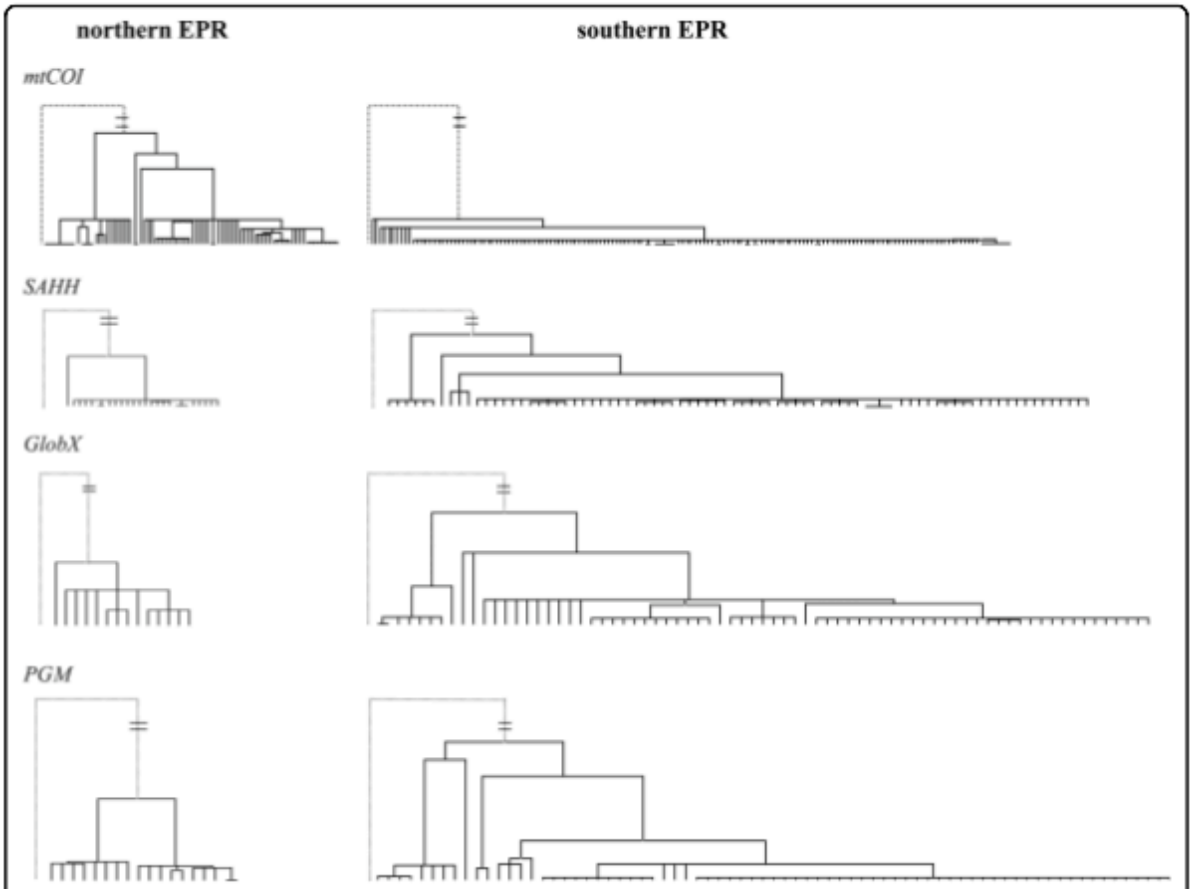


Figure 5 Ultrametric neighbour-joining topologies of four genes in *A. pompejana* in the northern and southern EPR. The outgroup *A. caudata* (dotted lines) was used to root the phylogenetic trees that were constructed from the most well-aligned parts of the four studied genes.

on the *mtCOI* gene [29]. This isolation may be the consequence of the formation of multiple transform faults ~1-2 Mya in the region (Figure 1, [59-61]) reinforced by a strong transverse deep-sea current at the Equator [62].

Slight, but significant differentiation at six allozyme loci across the equatorial barrier indicates that 1-2 My may not be sufficient to affect the net charge state of allozymes to create divergence, but likely reduced migration rates enough to modify allozyme frequencies by random genetic drift. This is particularly obvious in the distributions of allozymes and RFLP variants at the *PGM* locus (see Figure 6). Comparing patterns of genetic differentiation between PGM allozymes and DNA sequences indicated that allozymes are less sensitive at detecting the break. Moreover, the nearly complete replacement of the most frequent allele (100) in the northern populations by allele 78 in the south suggested that genetic drift is the main force acting to separate these populations at this locus. This equatorial break contrasts with the complete lack of genetic differentiation previously observed along thousands kilometres of vent fields on both sides of the EPR ([18], along the southern EPR in the present study) and confirms that isolation-by-distance by itself is not a valid argument to explain the observed decrease of gene flow for *A. pompejana*. In the absence of physical barriers, the species is able to disperse farther than expected based on moderate inter-field distances and the high bottom-current velocities [12]. This may be explained by either a developmental arrest of embryos in cold abyssal water masses and subsequent delay in metamorphosis [25], or the displacement of vent sites due to the movement of hydrothermal activity along the ridge axis, allowing bursts of colonisation by reconnecting previously isolated populations [21].

Questions addressing potential migration during an ongoing allopatric isolation process may be explored in two ways, depending on the primary cause for separation

Table 4 MLEs and likelihood ratio tests computed using Sweep_bott software for bottleneck and selective models.

	Multilocus	Single-locus			
		mtCOI	*GlobX*	*SAHH*	*PGM*
northern EPR					
M1 MLE	-99.718	-37.728	-11.171	-14.341	-36.475
M2 MLE	-92.660				
M3 MLE	-90.460	-35.171	-9.712	-11.170	-34.407
LRT	M2-M1: 14.116				
	$p < 0.005$				
	M3-M1: 18.516	5.114	2.918	6.342	4.136
	$p < 0.025$	(NS)	(NS)	$p < 0.025$	(NS)
	M3-M2: 4.400				
	(NS)				
southern EPR					
M1 MLE	-159.367	-4.149	-57.112	-38.064	-60.042
M2 MLE	-153.211				
M3 MLE	-139.745	-0.290	-52.196	-30.557	-56.704
LRT	M2-M1: 12.312				
	$p < 0.005$				
	M3-M1: 39.243	7.718	9.830	15.014	6.677
	$p < 0.005$	$p < 0.025$	$p < 0.01$	$p < 0.005$	$p < 0.005$
	M3-M2: 23.282				
	$p < 0.005$				

M1, neutral model; M2, bottleneck model; M3, selective sweep model; NS, non-significant. Significances of the likelihood ratio tests (LRT) were tested according to a χ^2 table with a df equal to the difference of the number of parameters estimated for each model (df = 2, 8 and 6 for M2-M1, M3-M1 and M3-M2 multilocus LRTs, respectively, and df = 1 for the M3-M1 single-locus LRT).

of populations (transform faults or bottom water circulation): (1) is there a progressive isolation of the north and south populations of *A. pompejana* by the offsetting of ridge segments at the Equator? or (2) is there emergence of a secondary contact between previously isolated populations due to the relaxation (i.e. attenuation of the transverse circulation) of the barrier?

Permeability of the barrier

Variation in the magnitude of the genetic differentiation among loci has been widely reported during the process of allopatric speciation [63] and is commonly detected across barriers to gene flow [64-66]. This is mainly attributable to: (1) the sensitivity of the molecular method used to detect polymorphism, which in turn depends on the time scale at which the isolation is observed (i.e. the accumulation of non-synonymous mutations that change the net charge of a protein is a long process) [67], (2) the stochasticity of the coalescence process [68], and (3) the formation of a genetic barrier when secondary contacts occur [9,49,69]. Fitting an isolation-with-migration model [45] between the northern and southern EPR populations help to estimate both historical and contemporary migration and better understand the role of the equatorial barrier on the isolation process. Even if the estimated migration rates were not significantly different from zero, the shape of the posterior probability distribution of m_n (north to south) was shifted in comparison to m_s (see Figure 4B) and suggests that the barrier is porous. While migration was weak or absent at *mtCOI*, all nuclear genes displayed an asymmetric number of migration events across the Equator from north to south (Table 3), suggesting that allele migration can occur between the northern and southern groups. This was confirmed by detecting a few intermediate individuals at 7°25'S and 14°S between the northern and southern *PGM* groups.

The fact that gene flow is virtually absent for at least two loci (*mtCOI* and *PGM*) raises the question of whether the barrier is becoming increasingly impermeable to dispersal with time, or whether it represents a semi-permeable genetic barrier following recent secondary contact. In the first case, one can expect that most "migrating" alleles (i.e. alleles typifying one lineage found in the other one) would coalesce deeply in the evolutionary tree as they will represent incomplete lineage sorting. In the other case (emergence of a tension zone), all "migrating" alleles should coalesce shortly in the coalescence tree, leading to very low mtimes (times at which gene flow was maximum).

Mean times of migration at the three nuclear genes indicated that gene exchanges are probably not recent and may have been maximal ~20 000 to 300 000 years ago and support the hypothesis of incomplete lineage sorting and ancestral variation [see [49,70]], favouring an increasingly stronger barrier scenario. However, the high level of gene flow heterogeneity across both nuclear and mitochondrial loci favours the notion of a semi-permeable barrier to gene flow as previously defined by Harrison [71]. Natural selection is expected to prevent gene flow in regions of the genome linked to genetic incompatibilities and to generate interloci differences in migration rates [72]. To this extent, the restriction of 'hybrid' individuals as estimated from the *PGM* RFLP screening to a narrow zone located between 7°25'S-14°S could be indicative of the occurrence of a genetic barrier with differential selection against hybrids [69,73,74]. However, it may also simply reflect the fact that very few individuals are able to cross the geographic barrier by chance maintaining low frequencies according to the migration/drift equilibrium. Currently, the present number of nuclear loci is too small to strongly indicate the

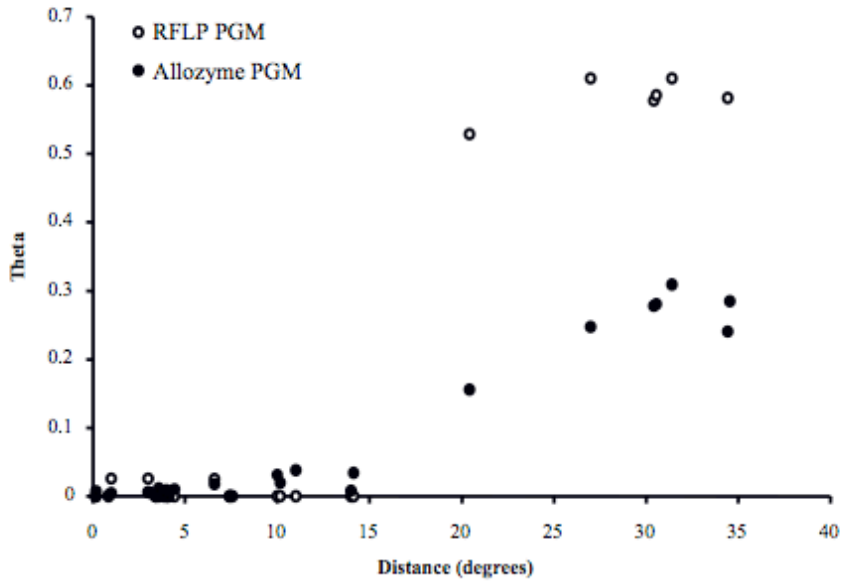


Figure 6 Sensitivity of RFLP and allozyme PGM markers for detecting the equatorial genetic break. The distance θ is plotted against the geographic distance between hydrothermal vent localities. The RFLP dataset shows a clearer departure from the isolation-by-distance model's expectation, indicating that markers to better detect the equatorial genetic break.

presence of a genetic barrier. Given that the vicariant event was quite recent (1-2 Mya), genetic incompatibilities and effective selection against hybrids may not have had time to arise supporting a model of progressive isolation of *A. pompejana* populations.

Demographic expansion in the southern EPR hypothesis and evidence for a possible selective sweep at the *SAHH* gene

A comparative phylogeographic approach involving seven vent species using the *mtCOI* gene (including *A. pompejana*) revealed the occurrence of a concomitant demographic expansion in the southern EPR within the last 0.5 My [29]. In the present study, southern populations (but not necessarily northern populations) displayed significantly negative Tajima's D and Fu and Li statistics for all genes. Moreover, the bottleneck model (M2) was significantly better than the constant-size model (M1) in describing the genealogies associated with the southern populations. Assuming a similar distribution of nucleotide diversity across the studied genes (as expected for a demographic event such as expansion [54,56]), these results are in agreement with an expansion of southern populations, as hypothesised by Plouviez *et al.* [29]. Multiple extinction and recolonisation events possibly caused by a higher rate of tectonic rearrangements and eruptive phases in the southern EPR [75,76], would have favoured the recurrence of bottlenecks through time as previously seen from the long-term monitoring of targeted vent fields since 1991 [77,78]. This also fits well with the drastic change of allozyme frequencies at the PGM locus on each side of the barrier as genetic drift may be greatly enhanced by recurrent bottlenecks and/or founder effects [79,80]. However, IMa results provided a complete reverse scenario (Figure 4C) suggesting that only the northern populations increased in size since the population splitting. Although unexpected, this finding may be due to several causes: (1) the very high proportion of divergent allelic forms found in the northern populations, a situation typifying an 'old' diversified lineage, (2) the nearly equal sample sizes used in the analysis by suppressing both obvious recombinant alleles and randomly-selected alleles in the south, and (3) the possible non-neutral behaviour of at least one loci (i.e. the *SAHH*). The two former explanations seem unlikely. In the first case, the expectation would be an ancestral population size nearly similar to that of the northern one. In the second case, the Sweep_bott analysis (based on the same sequence dataset) would have validated the bottleneck scenario in the northern populations. The detection of a bottleneck in the southern populations does not preclude the possible additional effect of a selective sweep at any specific locus. The *SAHH* locus was the only locus that conformed to a selective sweep model on both sides of the EPR with virtually no genetic differentiation across the expected barrier, if an initial isolation step in allopatry is assumed. An advantageous allele (or a neutral allele linked to a gene under selection by hitch-hiking) is more

likely to cross a genetic barrier and to spread within populations of the recipient lineage if the allele is advantageous for both lineages (e.g. [81,82]). Analyses of nucleotide polymorphism at this locus for northern and southern populations of *Alvinella caudata*, the most closely-related and syntopic species to *A. pompejana*, indicated a clear fixed divergence of alleles across the equatorial barrier analogous to divergences observed at both the *mtCOI* and the *PGM* loci in *A. pompejana* (S.P., unpublished data). This suggests that the lack of lineage sorting found in *A. pompejana* at the *SAHH* locus is probably not the result of a slow evolutionary rate at this locus but rather a case of high migration rate across the barrier. The spread of an allele across a permeable barrier is expected to be related to the magnitude of the selective advantage [83], suggesting that the sweep may be recent or very strong, even if the Fay and Wu test is only significantly negative in the southern EPR. Moreover, Faure *et al.* [49] proposed that diversifying selection helped to accelerate the evolution toward reciprocal monophyly at this specific gene in the Atlantic deep-sea vent mussels, suggesting that the *SAHH* gene may be indeed very sensitive to selection or linked to a potential positively selected gene.

Conclusion

In summary, our multilocus study of *A. pompejana* populations confirmed the presence of an equatorial barrier, which plays an important role in structuring populations of the species *A. pompejana*, as proposed by Plouviez *et al.* [29]. This geographic barrier appears to be still permeable to migration. However, low migration rates that would occur primarily from north to south may also indicate that the barrier is becoming more and more impermeable suggesting that transform faults are a likely mean to promote allopatric speciation along oceanic ridges. In addition, discriminating a bottleneck model from a selective sweep model through the comparison of *mtCOI* and nuclear genes strongly supports the southern demographic expansion hypothesis previously suggested by a multispecies analysis using *mtCOI* only. Combining multilocus and multispecies analyses therefore allows us to propose that allopatry is probably one of the major forces promoting speciation and biodiversity within the vent fauna along the one-dimensional framework of oceanic ridges.

Additional material

Additional file 1: Primers designed for DNA amplification on *Alvinella pompejana*. Table of sequences for forward and reverse primers used in the study.

Additional file 2: Size and exon-intron structure of the four sequenced genes. Schematic drawings of the gene portion used in which exonic regions are represented by boxes.

Additional file 3: Allele frequencies, expected and observed heterozygosities and Fis estimated from enzyme loci across vent fields. Table of allozyme frequencies and heterozygosities for each sampled population.

Additional file 4: Estimates and 90% Highest Posterior Density (HPD) intervals of demographic parameters from IMa multilocus analysis. Table of 90% HPD intervals for population sizes, migration rates and time since population splitting estimated from the IMa Bayesian computations.

Acknowledgements

We thank the chief scientists and 'Nautile' crews for their technical support and effort during our oceanographic expeditions: PHARE2002, HOPE99 and BIOSPEEDO2004. We are very grateful to Stéphane Hourdez and Baptiste Faure for collecting and sorting alvinellid polychaetes and to Grant Pogson, Carolyn Engel-Gauthier, Thomas Schultz and the two anonymous referees for their valuable comments and editorial suggestions on the manuscript. We would like to thank Christophe Caron and Erwan Corre for their help in bioinformatics. We are also truly indebted to the sequencing genomic plateforms GENOMER (Station Biologique de Roscoff, France) and GENOSCOPE (Evry, France) for the sequencing of the *mtCOI* fragment and our clone collection, respectively. This work was supported by the GDR Ecchis, the ANR-06-BDV-005 (Deep Oases: coord. Daniel Desbruyères) and ANR-05-BLAN-0407 (Alvi_Stress_Adapt: coord. Didier Jollivet). S. Plouviez was supported by a PhD grant from the Université Pierre et Marie Curie.

Author details

[1]Université Pierre et Marie Curie-Paris 6, Laboratoire Adaptation et Diversité en Milieu Marin, Roscoff, France. [2]CNRS, UMR 7144 Station Biologique de Roscoff, BP 74, Place Georges Teissier, 29682, Roscoff, France. [3]Duke University Marine Laboratory, 135 Marine Lab Road, 28516, Beaufort NC, USA. [4]IGBMC, Laboratoire de Bioinformatique et de Génomique Intégratives, Parc de l'Innovation, Campus Illkirch, Strasbourg, France.

Authors' contributions

SP carried out sequence acquisition and RFLP/indel genotyping as well as the molecular genetic analyses and drafted the manuscript. DLG genotyped individuals using allozymes. OL annotated the cDNA library of *A. pompejana*. FHL organised two of the three oceanographic cruises, helped sample specimens and provided helpful elements of discussion. DJ designed the study, carried out sampling, performed analyses with SP and edited the manuscript. All authors have read and approved the final manuscript.

Authors' information

This study is part of SP's PhD thesis on the comparative phylogeography of deep-sea hydrothermal vent species along the East Pacific Rise (EPR). Her research involves multiple species and markers to examine speciation and demographic processes along the EPR. DLG is lab technician in charge of genotyping and biochemical analyses. OL works on comparative genomics and bioinformatics. FHL has broad knowledge of vent ecology and has spent most of his career working on the physiology of deep-sea hydrothermal vent species. DJ is a population geneticist working on the evolution and dispersal of deep-sea hydrothermal vent species. He has recently developed a genomic approach to better understand the role of gene adaptation in the evolution of vent species.

Received: 2 March 2010 Accepted: 22 July 2010 Published: 22 July 2010

References

1. Wright S: **Evolution in mendelian populations.** *Genetics* 1931, **16**:97-159.
2. Wright S: **The distribution of gene frequencies in populations.** *Proc Nat Acad Sci USA* 1937, **23**:307-320.

3. Kimura M, Weiss GH: **Stepping stone model of population structure and the decrease of genetic correlation with distance.** *Genetics* 1964, **49**:561-576.
4. Wright S: **Isolation by distance.** *Genetics* 1943, **28**:114-138.
5. Malécot G: **Quelques schémas probabilistes sur la variabilité des populations naturelles.** *Ann Univ Lyon Sci* 1950, **13**:37-60.
6. Harrison S, Hastings A: **Genetic and evolutionary consequences of metapopulation structure.** *Trends Ecol Evol* 1996, **11**:180-183.
7. Pannell JR, Charlesworth B: **Effects of metapopulation process on measures of genetic diversity.** *Philos Trans R Soc Lond* 2000, **355**:1851-1864.
8. Cunningham CW, Collins TM: **Beyond area relationships: extinction and recolonization in molecular marine biogeography.** *Molecular Approaches to Ecology and Evolution* Birkhäser Verlag BaselDeSalle R, Schierwater B 1998, 297-321.
9. Barton NH, Hewitt GM: **Analysis of hybrid zones.** *Annu Rev Ecol Syst* 1985, **16**:113-148.
10. Schlager W: **Earth's layers, their cycles and Earth system science.** *Austrian J Earth Sci* 2008, **101**:4-16.
11. Vrijenhoek RC: **Gene flow and genetic diversity in naturally fragmented metapopulations of deep-sea hydrothermal vent animals.** *J Hered* 1997, **88**:285-293.
12. Chevaldonné P, Jollivet D, Vangriesheim A, Desbruyères D: **Hydrothermal-vent alvinellid polychaete dispersal in the eastern Pacific .1. Influence of vent site distribution, bottom currents, and biological patterns.** *Limnol Oceanogr* 1997, **42**:67-80.
13. Lutz RA: **Dispersal of organisms at deep-sea hydrothermal vents: a review.** *Oceanol Acta* 1988, **8(Suppl)**:23-29.
14. Black MB, Lutz RA, Vrijenhoek RC: **Gene flow among Vestimentiferan tube worm (*Riftia pachyptila*) populations from hydrothermal vents of the Eastern Pacific.** *Mar Biol* 1994, **120**:33-39.
15. Craddock C, Hoeh WR, Lutz RA, Vrijenhoek RC: **Extensive gene flow among Mytilid (*Bathymodiolus thermophilus*) populations from hydrothermal vents of the Eastern Pacific.** *Mar Biol* 1995, **124**:137-146.
16. Jollivet D: **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.** *PhD thesis* Université de Bretagne occidentale, IFREMER centre de Brest 1993.
17. Faure B, Chevaldonné P, Pradillon F, Thiébaut E, Jollivet D: **Spatial and temporal dynamics of reproduction and settlement in the Pompeii worm *Alvinella pompejana* (Polychaeta: Alvinellidae).** *Mar Ecol Prog Ser* 2007, **348**:197-211.
18. Pradillon F, Gaill F: **Oogenesis characteristics in the hydrothermal vent polychaete *Alvinella pompejana*.** *Inv Reprod Dev* 2003, **43**:223-235.
19. Byers JE, Pringle JM: **Going against the flow: retention, range limits and invasions in advective environments.** *Mar Ecol Prog Ser* 2006, **313**:27-41.
20. Jollivet D, Dixon LRJ, Desbruyères D, Dixon DR: **Ribosomal (rDNA) variation in a deep sea hydrothermal vent polychaete, *Alvinella pompejana*, from 13 degrees N on the East Pacific Rise.** *J Mar Biol Ass UK* 1998, **78**:113-130.
21. Jollivet D, Chevaldonné P, Planque B: **Hydrothermal-vent alvinellid polychaete dispersal in the eastern Pacific. 2. A metapopulation model based on habitat shifts.** *Evolution* 1999, **53**:1128-1142.
22. Slatkin M: **Gene flow and genetic drift in a species subject to frequent local extinction.** *Theor Pop Biol* 1977, **12**:253-262.
23. Jollivet D, Desbruyères D, Bonhomme F, Moraga D: **Genetic differentiation of deep-sea hydrothermal vent Alvinellid populations (Annelida, Polychaeta) along the East Pacific Rise.** *Heredity* 1995, **74**:376-391.
24. Jollivet D: **Specific and genetic diversity at deep-sea hydrothermal vents: An overview.** *Biodiv Conserv* 1996, **5(12)**:1619-1653.
25. Pradillon F, Shillito B, Young CM, Gaill F: **Deep-sea ecology - Developmental arrest in vent worm embryos.** *Nature* 2001, **413**:698-699.
26. Audzijonyte A, Vrijenhoek RC: **When gaps realy are gaps: statistical phylogeography of hydrothermal vent invertebrates.** *Evolution* .
27. Piccino P, Viard F, Sarradin PM, Le Bris N, Le Guen D, Jollivet D: **Thermal selection of PGM allozymes in newly founded populations of the thermotolerant vent polychaete *Alvinella pompejana*.** *Proc R Soc Lond* 2004, **271**:2351-2359.
28. Hurtado L, Lutz R, Vrijenhoek R: **Distinct patterns of genetic differentiation among annelids of eastern Pacific hydrothermal vents.** *Mol Ecol* 2004, **13**:2603-2615.
29. Plouviez S, Shank TM, Faure B, Daguin-Thiébaut C, Viard F, Lallier FH, Jollivet D: **Comparative phylogeography among hydrothermal vent species along the East Pacific Rise reveals vicariant processes and population expansion in the south.** *Mol Ecol* 2009, **18**:3[illegible]
30. Pasteur N, Pasteur G, Bonhomme F, Catalan J, Britton-Dav[illegible] *technique de génétique par électrophorèse de protéines* Paris[illegible]
31. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R: **DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates.** *Mol Mar Biol Biotech* 1994, **3**:294-299.
32. Gagnière N: **Développement d'une suite logicielle pour l'analyse et l'annotation intégrative automatiques de transcrits et de protéines. Application aux banques d'ADNc de l'annélide polychète *Alvinella pompejana*.** *PhD thesis* Université de Strasbourg, IGBMC Laboratoire de Bioinformatique et de Génomique Intégratives 2009.
33. Bierne N, Tanguy A, Faure M, Faure B, David E, Boutet I, Boon E, Quere N, Plouviez S, Kemppainen P, Jollivet D, Boudry P, David P: **Mark-recapture cloning: a straightforward and cost-effective cloning method for population genetics of single copy nuclear DNA sequences in diploids.** *Mol Ecol Notes* 2007, **7**:562-566.
34. Thompson JD, Higgins DG, Gibson TJ: **Clustal-W - improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucl Acids Res* 1994, **22**:4673-4680.
35. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucleic Acids Symposium Series* 1999, **41**:95-98.
36. Belkir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F: **Genetix 4.05, logiciel sous Windows TM pour la génétique des populations.** Laboratoire Génome, Populations, Interactions, Adaptations, UMR 5000, Université Montpellier 2 2004 [http://www.genetix.univ-montp2.fr/genetix/genetix.htm].
37. Weir BS, Cockerham CC: **Estimating F-statistics for the analysis of population-structure.** *Evolution* 1984, **38**:1358-1370.
38. Raymond M, Rousset F: **GENEPOP, version 1.2: population genetics software for exact tests and ecumenicism.** *J Hered* 1995, **86**:248-249.
39. Mantel N: **The detection of disease clustering and a generalized regression approach.** *Cancer Res* 1967, **27**:209-220.
40. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data.** *Genetics* 2000, **155**:945-959.
41. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R: **DnaSP, DNA polymorphism analyses by the coalescent and other methods.** *Bioinformatics* 2003, **19**:2496-2497.
42. Bandelt HJ, Forster P, Rohl A: **Median-joining networks for inferring intraspecific phylogenies.** *Mol Biol Evol* 1999, **16**:37-48[http://www.fluxus-engineering.com].
43. Excoffier L, Laval G, Schneider S: **Arlequin (version 3.0): An integrated software package for population genetics data analysis.** *Evolutionary Bioinformatics* 2005, **1**:47-50.
44. Hudson RR, Kaplan NL: **Statistical properties of the number of recombination events in the history of a sample of DNA-sequences.** *Genetics* 1985, **111**:147-164.
45. Hey J, Nielsen R: **Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics.** *Proc Nat Acad Sci USA* 2007, **104**:2785-2790.
46. Hasegawa M, Kishino H, Yano TA: **Dating of the human ape splitting by a molecular clock of mitochondrial-DNA.** *J Mol Evol* 1985, **22**:160-174.
47. Chevaldonné P, Jollivet D, Desbruyères D, Lutz RA, Vrijenhoek RC: **Sister-species of eastern Pacific hydrothermal vent worms (Ampharetidae, Alvinellidae, Vestimentifera) provide new mitochondrial COI clock calibration.** *Cah Biol Mar* 2002, **43**:367-370.
48. Johnson SB, Young CR, Jones WJ, Warén A, Vrijenhoek RC: **Migration, isolation, and speciation of hydrothermal vent limpets (Gastropoda; Lepetodrilidae) across the Blanco transform fault.** *Biol Bull* 2006, **210**:140-157.
49. Faure B, Jollivet D, Tanguy A, Bonhomme F, Bierne N: **Speciation in the deep sea: multi-locus analysis of divergence and gene flow between two hybridizing species of hydrothermal vent mussels.** *PLoS One* 2009, **4(8)**:e6485.
50. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24**:1596-1599.

51. Tajima F: **Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism.** *Genetics* 1989, **123**:585-595.
52. Fu XY, Li WH: **Statistical test of neutrality of mutations.** *Genetics* 1993, **133**:693-709.
53. Fay JC, Wu CI: **Hitchhiking under positive Darwinian selection.** *Genetics* 2000, **155**:1405-1413.
54. Hudson RR, Kreitman M, Aguade M: **A test of neutral molecular evolution based on nucleotide data.** *Genetics* 1987, **116**:153-159.
55. McDonald JH, Kreitman M: **Adaptive protein evolution at the Adh locus in *Drosophila*.** *Nature* 1991, **351**:652-654.
56. Galtier N, Depaulis F, Barton NH: **Detecting bottlenecks and selective sweeps from DNA sequence polymorphism.** *Genetics* 2000, **155**:981-987.
57. Garnier-Géré P, Dillmann C: **A computer-program for testing pairwise linkage disequilibria in subdivided populations.** *J Hered* 1992, **83**:239-239.
58. Robertson A, Hill WG: **Deviations from Hardy-Weinberg proportions: sampling variances and use in estimation of inbreeding coefficients.** *Genetics* 1984, **107**:703-718.
59. Kureth CL, Rea DK: **Large-scale oblique features in an active transform-fault, the Wilkes fracture-zone near 9-degrees-S on the East Pacific Rise.** *Mar Geophys Res* 1981, **5**:119-137.
60. Naar DF, Hey RN: **Speed limit for oceanic transform faults.** *Geology* 1989, **17**:420-422.
61. Francheteau J, Armijo R, Cheminee JL, Hekinian R, Lonsdale P, Blum N: **1 Ma East Pacific Rise oceanic-crust and uppermost mantle exposed by rifting in Hess Deep (Equatorial Pacific-Ocean).** *Earth Planet Sci Lett* 1990, **101**:281-295.
62. Reid JL: **On the total geostrophic circulation of the South-Pacific ocean: flow patterns, tracers, and transports.** *Progr Oceanogr* 1997, **39**:263-352.
63. Lewontin RC, Krakauer J: **Distribution of gene frequency as a test of theory of selective neutrality of polymorphisms.** *Genetics* 1973, **74**:175-195.
64. Pogson GH, Mesa KA, Boutilier RG: **Genetic population-structure and gene flow in the Atlantic cod *Gadus morhua* - a comparison of allozyme and nuclear rflp loci.** *Genetics* 1995, **139**:375-385.
65. Ting CT, Tsaur SC, Wu CI: **The phylogeny of closely related species as revealed by the genealogy of a speciation gene, Odysseus.** *Proc Nat Acad Sci USA* 2000, **97**:5313-5316.
66. Broughton RE, Harrison RG: **Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets.** *Genetics* 2003, **163**:1389-1401.
67. Sunnucks P: **Efficient genetic markers for population biology.** *Trends Ecol Evol* 2000, **15**:199-203.
68. Hudson RR, Turelli M: **Stochasticity overrules the "three-times rule": Genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA.** *Evolution* 2003, **57**:182-190.
69. Barton N, Bengtsson BO: **The barrier to genetic exchange between hybridizing populations.** *Heredity* 1986, **57**:357-376.
70. Nosil P, Harmon LJ, Seehausen O: **Ecological explanations for (incomplete) speciation.** *Trends Ecol Evol* 2009, **24**:145-156.
71. Harrison RG: **Hybrid zones: windows on evolutionary processes.** *Oxf Surv Evol Biol* 1990, **7**:69-128.
72. Wang RL, Wakeley J, Hey J: **Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives.** *Genetics* 1997, **147**:1091-1106.
73. Barton NH, Hewitt GM: **The genetic-basis of hybrid inviability in the grasshopper *Podisma pedestris*.** *Heredity* 1981, **47**:367-383.
74. Kruuk LEB, Baird SJE, Gale KS, Barton NH: **A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids.** *Genetics* 1999, **153**:1959-1971.
75. Cormier M-H: **The ultrafast East Pacific Rise: instability of the plate boundary and implications for accretionary processes.** *Philos Trans R Soc Lond* 1997, **355**:341-367.
76. Fouquet Y, Auzende JM, Ballu V, Batiza R, Bideau D, Cormier MH, Geistdoerfer P, Lagabrielle Y, Sinton J, Spadea P: **Variability of the present-day hydrothermal activity along an ultra-fast spreading ridge - East Pacific Rise between 17- degrees-C and 19-degrees-S (Naudur Cruise).** *C R Acad Sci Ser A* 1994, **319**:1399-1406.
77. Haymon RM, Fornari DJ, Edwards MH, Carbotte S, Wright D, Macdonald KC: **Hydrothermal vent distribution along the East Pacific Rise crest (9-degrees-09'- 54'N) and its relationship to magmatic and tectonic processes on fast-spreading midocean ridges.** *Earth Planet Sci Lett* 1991, **104**:513-534.
78. Tunnicliffe V, Embley RW, Holden JF, Butterfield DA, Massoth GJ, Juniper SK: **Biological colonization of new hydrothermal vents following an eruption on Juan de Fuca ridge.** *Deep-Sea Res Part I* 1997, **44**:1627.
79. Gilpin M: **The genetic effective size of a metapopulation.** *Biol J Linn Soc* 1991, **42**:165-175.
80. Leberg PL: **Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis.** *Evolution* 1992, **46**:477-494.
81. Pialek J, Barton NH: **The spread of an advantageous allele across a barrier: The effects of random drift and selection against heterozygotes.** *Genetics* 1997, **145**:493-504.
82. Faure MF, David P, Bonhomme F, Bierne N: **Genetic hitchhiking in a subdivided population of *Mytilus edulis*.** *BMC Evol Biol* 2008, **8**:164-178.
83. Morjan CL, Rieseberg LH: **How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles.** *Mol Ecol* 2004, **13**:1341-1356.

doi:10.1186/1471-2148-10-220
Cite this article as: Plouviez *et al.*: Determining gene flow and the influence of selection across the equatorial barrier of the East Pacific Rise in the tube-dwelling polychaete *Alvinella pompejana*. *BMC Evolutionary Biology* 2010 **10**:220.

Additional file 1

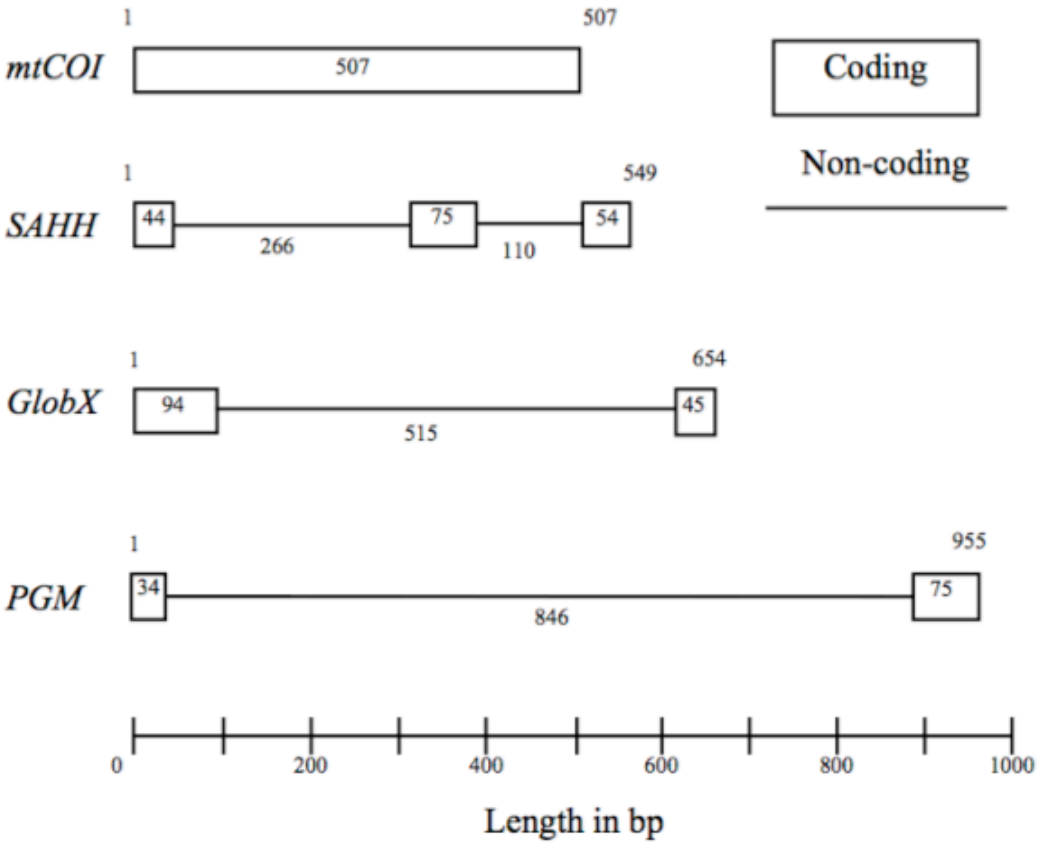
Primers designed for DNA amplification on *Alvinella pompejana*.

Gene	Primer sequences (5'-3')	T_a (°C)
mtCOI	**F:** TATTTGGTATTTGGGCAGGTC	57
	R: GATGGGTCGAAGAATGATGTG	
GlobX	**F:** ----AGTTGACTGAAGAACGCAGAGAAGCGGT	55
	R: ----CGCGAAGATCCTTGAAGTACTCTTGGTA	
SAHH	**F:** ----CTTGTCAACCTCGGCTGTG	57
	R: ----CTGCCTGCTCGTCTGTTAG	
PGM	**F:** ----CCCAGGTGGTCCAAATGC	57
	R: ----GTACAGATGTCAGCCTTCAGGTC	
Indel polymorphism	**F, F*:** GCCATGTTTTTAATTGATTCTGTC	58
in *PGM* exon4	**R:** TCACTACAAAATGAGCTGATGTGC	

----: corresponds to the nucleotide '-bp tag' of primers for the Mark-Recapture sequence method. F, forward primer; F*, forward labelled primer IRD700™; R, reverse primer; T_a, annealing temperature

Additional file 2

Size and exon-intron structure of the four sequenced genes.



Additional file 3

Allele frequencies, expected and observed heterozygosities and Fis estimated from enzyme loci across vent fields.

Locus	Vent field						
	13°N	7°25'S	14°S	17°25'S	17°34'S	18°33'S	21°25'-33'S
PGM							
(N)	40	17	15	34	19	52	35
78	0.0349	0.3889	0.4000	0.4286	0.5000	0.4904	0.4286
90	0.3837	0.4167	0.5667	0.5571	0.4737	0.5096	0.5571
100	0.5116	0.1667	0.0333	0.0143	0.0263	0.0000	0.0143
112	0.0698	0.0278	0.0000	0.0000	0.0000	0.0000	0.0000
H exp.	0.5849	0.6466	0.5178	0.5057	0.5249	0.4998	0.5057
H obs.	0.5814	0.8889	0.8667	0.8857	1.0000	0.9808	0.8857
Fis	-0.128	0.022	0.104*	-0.030	-0.231	-0.281	0.139
MPI							
(N)	40	17	15	34	19	52	35
70	0.0125	0	0	0	0	0	0
92	0.0375	0	0	0	0.0263	0.0096	0
96	0.35	0.3235	0.1333	0.0882	0.1579	0.1058	0.0714
100	0.6	0.6765	0.8667	0.9118	0.8158	0.8846	0.9286
H exp.	0.5159	0.4377	0.2311	0.1609	0.3089	0.2062	0.1327
H obs.	0.675	0.2941	0.1333	0.1765	0.3684	0.1923	0.0857
Fis	-0.308	0.370*	0.451*	-0.097	-0.167	0.077	0.366

Locus	**Vent field**						
	13°N	7°25'S	14°S	17°25'S	17°34'S	18°33'S	21°25'-33'S
GPI							
(N)	40	17	15	34	19	52	35
70	0.0125	0	0	0	0	0	0
100	0.9875	1	1	0.9118	0.9737	0.9904	0.9429
120	0	0	0	0.0882	0.0263	0.0096	0.0571
H exp.	0.0247	0	0	0.1609	0.0512	0.019	0.1078
H obs.	0.025	0	0	0.0588	0.0526	0.0192	0.0571
Fis	-0.012	-	-	0.644*	0.000	0.000	0.481
6PGD							
(N)	40	17	15	34	19	52	35
80	0	0	0	0	0	0.0096	0
90	0.2	0.4118	0.2333	0.2206	0.3158	0.2019	0.1571
100	0.775	0.5588	0.7667	0.75	0.6842	0.6154	0.8286
110	0.025	0.0294	0	0.0294	0	0.1731	0.0143
H exp.	0.3588	0.5173	0.3578	0.388	0.4321	0.5505	0.2886
H obs.	0.3	0.4118	0.2	0.3235	0.5263	0.4808	0.2286
Fis	0.190	0.261	0.468	0.187	-0.192	0.136	0.222

Locus	**Vent field**						
	13°N	7°25'S	14°S	17°25'S	17°34'S	18°33'S	21°25'-33'S
IDH							
(N)	40	17	15	34	19	52	35
100	0.9	0.8824	0.8333	0.9853	0.8947	0.9712	0.9429
120	0.0625	0.1176	0.1333	0	0.1053	0.0096	0.0571
140	0.0375	0	0.0333	0.0147	0	0.0192	0
H exp.	0.1847	0.2076	0.2867	0.029	0.1884	0.0564	0.1078
H obs.	0.15	0.2353	0.3333	0.0294	0.2105	0.0577	0.1143
Fis	0.204	-0.097	-0.129	0.000	-0.091	-0.013	-0.046
ACP							
(N)	40	17	15	34	19	52	35
90	0.0125	0	0	0	0.0263	0.0096	0
100	0.9875	0.6765	0.9667	1	0.9737	0.9808	0.8143
120	0	0.3235	0.0333	0	0	0.0096	0.1857
H exp.	0.0247	0.4377	0.0644	0	0.0512	0.0379	0.3024
H obs.	0.025	0.5294	0.0667	0	0.0526	0.0385	0.3714
Fis	0.000	-0.150	0.000	-	0.000	-0.005	-0.214

Locus	**Vent field**						
	13°N	7°25'S	14°S	17°25'S	17°34'S	18°33'S	21°25'-33'S
LAP							
(N)	40	17	15	34	19	52	35
90	0	0	0	0.0294	0	0	0
100	1	1	0.9333	0.9706	1	1	1
110	0	0	0.0667	0	0	0	0
H exp.	0	0	0.1244	0.0571	0	0	0
H obs.	0	0	0.1333	0.0588	0	0	0
Fis	-	-	-0.037	-0.015	-	-	-
TPI							
(N)	40	17	15	34	19	52	35
82	0	0.0294	0	0.0294	0	0.0481	0.0143
90	0	0.0294	0	0.1471	0.0263	0.3269	0.0571
100	0.3875	0.8529	0.9333	0.6912	0.6842	0.6154	0.6857
110	0.5875	0.0882	0.0667	0.1324	0.2895	0.0096	0.2429
120	0.025	0	0	0	0	0	0
H exp.	0.5041	0.263	0.1244	0.4823	0.4474	0.512	0.4673
H obs.	0.6	0.1765	0	0.6176	0.4211	0.75	0.4
Fis	-0.232	0.358	1.000*	-0.257	0.086	-0.457	0.158

Locus	**Vent field**						
	13°N	7°25'S	14°S	17°25'S	17°34'S	18°33'S	21°25'-33'S
HK							
(N)	40	17	15	34	19	52	35
96	0.025	0	0	0.0441	0.0263	0.0096	0
100	0.5625	0.8824	0.6667	0.6618	0.7895	0.7308	0.7571
104	0.4125	0.1176	0.3333	0.2941	0.1842	0.2596	0.2429
H exp.	0.5128	0.2076	0.4444	0.4736	0.3421	0.3985	0.3678
H obs.	0.725	0.2353	0.2667	0.2059	0.3158	0.4231	0.4857
Fis	-0.388	-0.097	0.429	0.580***	0.104	-0.052	-0.308

N: number of individuals; H exp.: expected heterozygosity, H obs.: observed heterozygosity, *: $P < 0.05$, ***: $P < 0.001$.

Additional file 4

Estimates and 90% Highest Posterior Density (HPD) intervals of demographic parameters from IMa multilocus analysis.

Parameters	θ_n	θ_s	θ_a	t	m_s	m_n	N_n (x1000)	N_s (x1000)	N_a (x1000)	M_s	M_n	t (My)
Estimate (HiPt)	10.498	4.215	4.050	1.021	0.005	0.315	4085	1640	1576	0.026	0.038	1.589
Lower 90% HPD	5.703	2.397	1.571	0.477	0.005	0.005	2219	933	611	0.001	0.000	0.743
Higher 90% HPD	24.053	7.191	7.191	1.809	0.255	0.905	9360	2798	2798	19.385	3.727	2.816

For θ and N parameters, labels (n) and (s) correspond to the northern and southern groups of populations, respectively and label (a) corresponds to the parameters for the ancestral population. Forward in time, m_s and m_n correspond to south to north and north to south genes migration parameters, respectively.

$N_n = \theta_n /(4UG)$, $N_s = \theta_s /(4UG)$, $N_a = \theta_a /(4UG)$, $M_s = 2N1m_s$, $M_n = 2N2m_n$, t (years) = t/U, with U the geometric mean of the mutation rate per year across the loci and G the number of years per generation.

4.2. *B. thermophilus*

4.2.1. Résumé des résultats

Contrairement à l'espèce *A. pompejana* qui associe divergence sur le mtCOI et différenciation géographique de part et d'autre de la barrière équatoriale (constituée en partie par le complexe de failles transformantes Discovery/Gofar), la moule hydrothermale *B. thermophilus* présente deux clades divergents sur le gène mitochondrial sans situation d'allopatrie stricte de 13°N à 7°25'S, ce qui rend le positionnement de la barrière entre populations Nord et Sud EPR difficile sur la base d'un seul gène. De plus, les 2 clades divergents étant mélangées sur les sites 13°N, 9°50N et 7°25S, plusieurs scenarii de colonisation pourraient expliquer la distribution géographique de ces clades:

(1) un isolement physique puis génétique de part et d'autre de la barrière par vicariance, suivi d'une remise en contact secondaire au Nord de l'EPR liée aux caractéristiques particulières de dispersion chez cette espèce,
(2) un isolement sympatrique (parapatrique) des deux clades mitochondriaux au Nord suivi de la migration d'un seul de ces clades vers le Sud à travers la barrière,
(3) l'extinction du clade nord au Sud de l'EPR due au taux de remaniement tectonique des sites plus élevé dans cette région.

Une comparaison des flux de gènes entre le gène mtCOI, des marqueurs allozymiques et le polymorphisme de séquences de quatre gènes nucléaires a ainsi été réalisée afin de :

(1) préciser la position géographique de la barrière aux flux de gènes
(2) déterminer le degré de perméabilité de cette barrière, en évaluant la migration à travers celle-ci (intensité, asymétrie)
(3) vérifier par l'approche multi-locus l'hypothèse d'expansion démographique au Sud de la dorsale proposée par Plouviez et al (2009).

L'analyse des allozymes effectuée de 7°25'S à 21°33'S n'a pas permis de mettre en évidence de structure génétique sur cette portion de dorsale et confirme l'absence d'isolement par la distance chez *B. thermophilus* suggérée par Craddock et al (1995) sur l'EPR Nord et les sites Galápagos. Les quatre gènes nucléaires présentent, quant à eux, une divergence marquée en 2 clades d'allèles (Fig. 4.6) associée à une différenciation

génétique entre les champs hydrothermaux 7°25'S et 14°S, et confirment donc la localisation géographique supposée d'une barrière dans cette zone géographique. Les gènes S-Adenosyl Homocystéine Hydrolase (SAHH) et Sulfotransférase 1 (Sulfo 1) présentent ainsi deux clades divergents répartis respectivement de part et d'autre de cette barrière géographique. De plus, si les clades divergents observés sur le gène Lysozyme (Lyso) se distribuent sur l'ensemble de la dorsale, un indel de 1 pb présent dans les deux clades (et donc impliquant une recombinaison entre ces clades) est structuré géographiquement. En effet, la « délétion » est trouvée uniquement au Sud de la barrière, indiquant que celle-ci serait récemment imperméable au passage des nouvelles mutations, au moins pour ce gène.

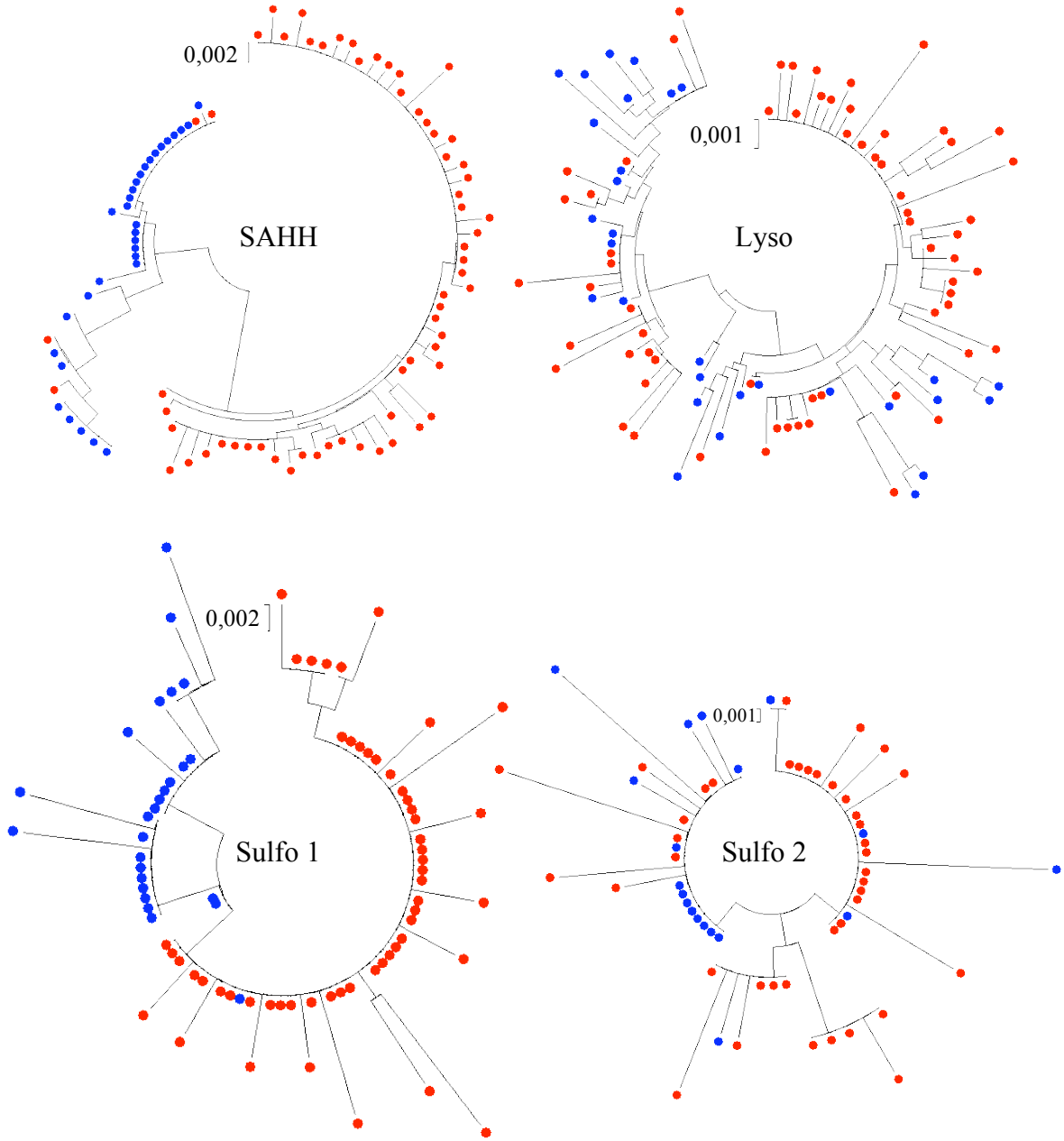


Fig. 4.6 Arbres phylogénétiques réalisés selon la méthodes des plus proches voisins sur les gènes nucléaires de *B. thermophilus* à partir des jeux de séquences de qualité 2 (2 allèles maximum par individu).
En bleu et rouge : séquences d'individus échantillonnées respectivement au Nord (9°50'N-7°25'S) et au Sud (14°S-21°33'S) de la barrière.

L'utilisation d'un modèle d' « Isolement avec Migration » (IMa, Hey & Nielsen 2007) estime le temps de divergence entre les populations de part et d'autre de la barrière de 1 à 2 Ma. Ce résultat est en accord avec la date de mise en place de la barrière (environ 1,3 Ma, Fig. 4.7) proposée par l'approche multispécifique. La migration à travers la barrière, bien que relativement faible, à la fois du Nord vers le Sud et du Sud vers le Nord (Fig. 4.7), avec le sens de migration variant selon les gènes, indiquant une semi-perméabilité de la barrière génétique aux flux de gènes. Un balayage sélectif (ou effet autostop) a été détecté sur le gène Lyso via le logiciel Sweep_bott et conforté par les D de Tajima et F* de Fu & Li significativement négatifs au nord et au sud. Si les gènes Sulfo 1 et 2 contribuent exclusivement à la significativité du test HKA, ces 2 locus ne présentent pas de coalescents s'écartant de l'attendu neutre (Sweep_bott) et l'absence de mutation fixée entre *B. thermophilus* et l'espèce Atlantique *B. azoricus* suggère que ces gènes sont plutôt soumis à une forte sélection purifiante.

Les indices de Tajima et Fu & Li significatifs au Sud pour l'ensemble des gènes ne le sont que pour le gène Lyso au nord. Ces résultats complémentés par les analyses d'adéquation à un modèle de goulot d'étranglement (bottleneck) effectuées par le logiciel Sweep_bott et la mise en évidence par le logiciel IMa d'une taille efficace de la lignée mitochondriale légèrement Sud supérieure à celles des lignée Nord et la lignée ancestrale (Fig. 4.7), confirment l'hypothèse d'expansion démographique au Sud émise par l'approche multi-espèces sur le gène mtCOI (Plouviez et al 2009). Contrairement à l'espèce *A. pompejana* pour qui l'expansion démographique n'a été détectée qu'au Sud de la dorsale, l'analyse de IMa suggère également la possibilité d'une expansion démographique au Nord de la barrière chez *B. thermophilus*.

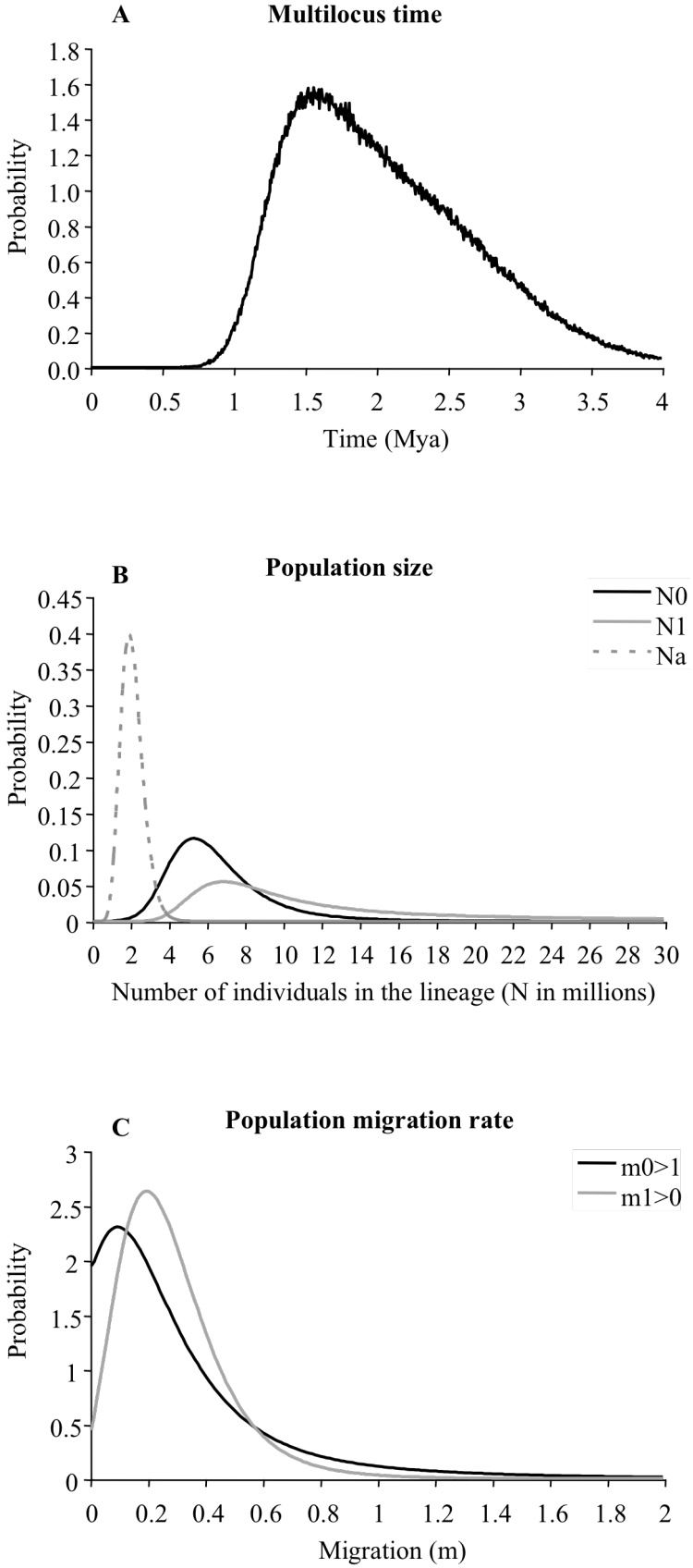


Fig 4.7 Distribution des probabilités posterieures marginales du temps de divergences entre populations (A), de la taille efficace de la population ancestrale (Na) et et de ces populations filles (N0 au Nord et N1 au Sud) (B) et des taux de migration entre populations (dans le sens du temps, m0>1 du Sud vers le Nord et m1>0 du Nord vers le Sud) (C).

4.2.2. Article en préparation

Divergence population genetics reveal the secondary and semi-permeable nature of a barrier to gene flow between hydrothermal vent mussel populations of the East Pacific Rise

Plouviez S. [1,2], Faure B. [1,2,3,4,5], Le Guen D. [1,2], Lallier F.H. [1,2], Bierne N. [3,4], Jollivet D. [1,2]

[1] Université Pierre et Marie Curie-Paris 6, Laboratoire Adaptation et Diversité en Milieu Marin, Roscoff, France
[2] CNRS, UMR 7144 Station Biologique de Roscoff, BP 74, Place Georges Teissier, 29682, Roscoff, France
[3] Université Montpellier 2, Place Eugène Bataillon, Montpellier, France
[4] CNRS, Institut des Sciences de l'Evolution, UMR 5554, Station Méditerranéenne de l'Environnement Littoral, Sète, France
[5] Current address: Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA

Key words: allozymes, nuclear genes, divergence, coalescence, barrier, secondary contact zone

Abstract

Comparative phylogeography of deep-sea hydrothermal vent species has showed the co-occurrence of several genetic breaks between the northern and southern parts of the East Pacific Rise (EPR). However, the width and position of transition zones were variable among species. We studied in more details the position and strength of the barrier between the deep-sea vent mussel *Bathymodiolus thermophilus*, which proved difficult to establish with mtDNA alone. Eight allozymes and DNA sequence polymorphisms at four nuclear genes were added to improve the fine-scale geographic structure of the mussels previously obtained from the mtCOI data. The barrier was positioned between 7°25'S and 14°S, which slightly differs from the one clearly established in other species and possibly reflects the planktonotrophic larval life-style of the species. An analysis of gene genealogies with a coalescent approach revealed contrasted levels and directions of secondary introgression across the barrier. Although intermediate hybrid populations were not observed, the position, the secondary nature and especially the semi-permeability of the barrier suggest that reproductive isolation evolved between populations and is responsible for the genetic structure observed. Our results also confirmed the scenario of a historical demographic expansion in the South EPR as previously proposed from a multi-species approach.

Introduction

The origin and causes of genetic differentiation between populations are often difficult to establish. In addition to the demographic characteristics of the species studied (e.g. population sizes, dispersal rates) and the relation to its habitat distribution (e.g. natural barriers), genetic structure also relies on the history of populations (Avise 1994, Hewitt 2001) and to the evolution of reproductive isolation (Barton and Hewitt 1985). The co-occurrence of a genetic break with a barrier to dispersal does not necessarily means that the barrier itself is the primary cause of the observed genetic structure. The physical barrier can also be a suture zone that delays, rather than prevents, genetic homogenisation of previously isolated populations (Hewitt 2001) or it can have trapped a genetic barrier produced by partial reproductive isolation (Barton 1979, Barton & Hewitt 1985). Testing for the existence of reproductive isolation needs experimental work which is not always achievable. Hydrothermal vent species well illustrate this difficulty because experiments are uneasy to conduct on board as well as in the field. Comparing phylogeographic patterns across several co-distributed species represents a solution to detect shared vicariant events, and to locate more precisely contact zones that also act as barrier to gene flow (Maggs et al 2008). When conducted on a single locus however, such an approach does not inform much about the existence of reproductive isolation. Multilocus comparisons can provide patterns able to discriminate between a genetic barrier and a dispersal barrier: contrary to a natural barrier to dispersal that affect the whole genome, a genetic barrier produced by partial reproductive isolation is expected to prevent gene flow in regions of the genome where isolation genes operate while gene flow can still be substantial in regions lacking such genes (Barton & Hewitt 1985), resulting in the so-called semi-permeable barrier to gene flow (Harrison 1990). Polymorphism to divergence analyses on unlinked loci from different genomes and/or genome compartments (i.e. maternally-inherited mitochondrial versus nuclear loci) could thus provide useful additional information about the strength of a barrier, its geographic range and directionality.

Comparative phylogeographic analyses of deep-sea hydrothermal vent species has previously showed the occurrence of a genetic break between the northern and southern parts of the East Pacific Rise (EPR, see Hurtado et al 2004, Won et al 2003, Plouviez et al 2009). These studies established a shared vicariant event among the species studied and suggested the emergence of a concomitant barrier to dispersal about 1.5 to 2 Mya at the

Equator –*i.e.* a potential suture zone. However, the width and position of transition zones did not perfectly match among species: some species such as the tube-dwelling polychaete *Alvinella pompejana* displaying an abrupt separation of the populations across the Equator (Plouviez et al 2010) whereas others showed genetic patterns closer to an isolation-by-distance model (Won et al 2003). Interestingly, the deep-sea mussel *Bathymodiolus thermophilus* exhibited a smooth clinal distribution of mitochondrial lineages along the northern EPR, the centre of the cline being positioned southward (Plouviez et al 2009). The mitochondrial lineages overlap also coincides with the northern range limit (i.e. 13°N/EPR) of the species, suggesting a possible directional migration of the southern lineage across the barrier and thus questioning about the role of dispersal versus reproductive isolation on the population distribution of *B. thermophilus*. *Bathymodiolus* mussels still possess a dispersive planktotrophic larval stage that allows great dispersal capabilities in the water masses. According to closely-related species, the planktonic larval duration may be up to 6 months (Dixon et al 2006, Arellano & Young 2009) with planktotrophic larvae able to reach upper-lighted layers of the ocean for feeding to finally reach a size of about 400 μm prior to metamorphosis and settlement (Mullineaux et al 2005, Arellano & Young 2009). Such dispersal characteristics could thus favour population connectivity along previously isolated ridges. One might thus suspect that partial reproductive isolation derived from an initial isolation period may now contribute to the maintenance of this geographic structure over the Equator.

In the present study, comparing the spatial distribution of mtCOI haplotypes with those of allozymes and samples of sequences at 4 nuclear genes will address several questions about the evolution of *B. thermophilus* populations located around the north/south barrier of the East Pacific Rise which are:

(1) are allele distributions at other loci consistent with the 7°25'S-14°S barrier to gene flow proposed using mtCOI (Plouviez et al 2009)?

(2) how great is the permeability of the genetic barrier?

(3) could gene flow be really asymmetric between the north and south regions?

(4) does the hypothesis of population expansion in the south EPR still hold when using a series of nuclear genes from *Bathymodiolus*?

To answer these questions, alleles of four nuclear loci were cloned and sequenced in 7 populations located along the whole EPR. Sequence patterns were then compared

against mitochondrial haplotype and allozyme distributions and several evolutionary hypotheses tested using a coalescent approach.

Material and Methods

Collection

Bathymodiolus thermophilus specimens were sampled from seven deep-sea hydrothermal vent fields along the East Pacific Rise (Table 4.1) using the tele-manipulated arm of the manned submersible Nautile operated from the oceanographic vessels Le Nadir and L'Atalante during two oceanographic cruises: at 9°50N during HOT 1996 and from latitudes of 7°25S to 21°33S during BIOSPEEDO 2004. During both cruises, all fresh specimens were measured and dissected on board and tissues (mantle and muscles) were preserved in 80% alcohol. In addition, the anterior muscle of each individual was also frozen in liquid nitrogen for allozyme analyses during BIOSPEEDO 2004.

Table 4.1 Location and sample size of populations sampled for allozymes (n_{allo}), mtCOI (n_{mtCOI}), and nuclear genes.
n_{SAHH}, n_{Lyso}, n_{Sulfo1}, n_{Sulfo2}, total number of recaptured alleles for each nuclear gene. In bold number of recaptured individuals.

Vent field	Geographical position	Depth (m)	n_{allo}	n_{mtCOI}	n_{SAHH}	n_{Lyso}	n_{Sufo1}	n_{Sulfo2}
9°50N	9°50'N	2530	0	45	10	16	14	10
	104°17'W				**6**	**11**	**9**	**8**
7°25S	7°25'S	2735	10	48	33	29	22	12
	107°47'W				**25**	**19**	**17**	**11**
14°00S	13°59'S	2623	38	30	19	24	15	12
	112°29'W				**12**	**14**	**12**	**10**
17°25'S	17°25'S	2575	68	21	16	15	14	12
	113°12'W				**11**	**10**	**12**	**8**
17°34'S	17°35-36'S	2595	23	60	15	24	14	14
	113°15'W				**10**	**15**	**12**	**10**
18°33S	18°33'S	2636	33	30	17	19	15	9
	113°24'W				**11**	**12**	**12**	**7**
21°33S	21°33'S	2800	103	27	28	24	13	9
	114°18'W				**19**	**14**	**8**	**6**

Allozyme genotyping

Eight enzyme loci were genotyped for each individual of *B. thermophilus* collected from 7°25'S to 21°33'S (Table 4.1) following the protocols of Pasteur et al (1987). Proteins from each frozen adductor muscle were extracted according to the Piccino et al (2004)'s method. Enzyme electrophoreses were conducted on 12% starch gel using two different buffer systems: (1) Tris-Citrate pH 8.0 (TC 8) for Phosphoglucomutase (*Pgm*, E.C. 5.4.2.2), Mannose phosphate isomerase (*Mpi*, E.C. 5.3.1.8), Octopine deshydrogenase (*Odh*, 1.5.1.11) and Leucine amino peptidase *(Lap*, 3.4.11.1); (2) Tris-Citrate pH 6.7/6.3 (TC 6.7) for Glucose phosphate isomerase (*Gpi,* 5.3.1.9), Malate deshydrogenase-1 and -2 (*Mdh 1* and *2*, 1.1.1.37), and Hexokinase-1 (*Hk*, 2.7.1.1). Loci were numbered according to the decreasing anodal electromorph mobility in multi-loci systems. Alleles were numbered according to their relative mobility from the most frequent allele (labelled as 100) previously obtained for the Atlantic species *B. azoricus*, this species being used as a reference (see Boutet et al 2009).

DNA sequencing

Genomic DNA was extracted using a CTAB-PVP extraction procedure following Jolly et al (2006)'s protocol. Mitochondrial lineages of *B. thermophilus* were identified from cytochrome oxidase I gene (mtCOI) sequences previously obtained by Plouviez et al (2009: Table 4.1). Sequences from 4 nuclear genes were obtained from the same individuals using the mark-recapture (MR) cloning technique developed by Bierne et al (2007) with a capture effort of X2, using primers developed by Faure et al (2009) for the intron-exons gene portions of S-Adenosyl Homocystein Hydrolase (SAHH) and Lysozyme (Lyso) and specific primers TCTTTAAAGTCAGGATCACATTGG (5'-3' Forward) and TAAGGCAAAGTGGAACAACGAGACCGC (5'-3' Reverse) developed for the two other exonic paralogous genes of the Sulfotransferase (called Sulfo1 and Sulfo2, respectively). PCR-amplification were conducted in a 20-µl solution that included 1x buffer (supplied by manufacturer), 2 mM $MgCl_2$, 0.25 mM (0.3 mM for Sulfotransferase) of each dNTPs, 0.4 µM of each primers, 0.5 U of Taq polymerase (Thermoprime plus), 0.5µl of a 2% solution of BSA, 2 µl of template DNA and sterile H_2O. Thermal cycling parameters used an initial denaturation at 94°C for 5 min, followed by 35 cycles at 94°C for 35s, 60°C for 60s and 72°C for 2min, before a final 10 min extension at 72°C. Each individual was PCR-amplified separately with a different set of 5' tailed primers (4-nucleotides tail). PCR products of similar quantities were mixed together and cloned into a

pGEM-T vector by using the Promega pGEM-T cloning kit. Clones were sequenced with universal plasmid primers SP6 and T7. The individual sequences were identified with the help of the 5'tag sequence characterizing each primers set.

Statistical Methods

The program Genetix 4.05.2 (Belkir et al 2004) was used to perform genetic population analyses on allozymes. For each locus, allele frequencies, heterozygosities and Weir & Cockerham (1984)'s *f* estimator (departure from Hardy-Weinberg equilibrium tested by a 1000-permutations test) were estimated for each population along the EPR. The overall genetic differentiation across populations was estimated at all loci using the Weir & Cockerham (1984)'s q estimator and tested from being significantly different from zero by 1000 permutations. A Principal Component Analysis was performed using PCAGEN (Goudet 1999) to discriminate populations according to their allozyme-frequency distribution. The isolation-by-distance model was tested with a Mantel Spearman test with 5000 permutations using Genepop 4.0.10 (Raymond & Rousset 1995).

DNA sequences obtained from the MR-cloning method were proofread using CodonCode Aligner 2.0.6 (http://www.codoncode.com/aligner/). Sequence alignments were initially performed with ClustalW (Thompson et al 1994) using BioEdit version 6.0.6 (Hall 1999) and improved manually. The number of recaptured alleles varied from one population to another and allele sample size of each gene and population is indicated in Table 4.1. Because of the random nature of the recapture, it was not possible to distinguish 'true' homozygotes from heterozygotes with a few recaptured alleles. Consequently, only the most recaptured allele was retained from each individual to avoid sample bias when performing demographic analyses and genetic diversity estimations (see: Table 4.1). Multiple recaptures allowed us to discard intra-individual *in vitro* recombinants and putative artefactual/somatic mutations. Recombinants between different individuals (1-2% of the dataset for each population) from the same PCR set were detected thanks to abnormal combination of the 5'-tails and removed. This corresponded to 1-2% of the dataset for each population.

For each locus, phylogenetic relationships among alleles were estimated using the median joining algorithm of the Network software (version 4.5.0.0; www.fluxus-engineering.com: Bandelt et al 1999) to detect potential divergent clades. Geographic distribution of divergent clades was then examined to locate potential barrier to gene flow, by plotting synthetic clade-specific allele-frequency distributions for each locality. These

clades were thus defined as north or south according to the allele-frequency distributions. Localities from each side of the putative barrier were grouped together to test for genetic differentiation across this barrier using F-statistics (f_{st}; Hudson et al 1992) computed via DNAsp 4.10.3 and tested from 1000-permutations.

With the same software, nucleotide diversity (p_n) and Watterson's theta (q_w) were estimated for each gene within (1) each geographic locality and (2) each clade. A demographic event, should impact the whole genome in the same way. Conversely, selective effects could act only on few genes. To disentangle the effects of selective vs demographic effects, departure to neutral evolution (i.e. no selection and no demographic event) was thus tested for each gene among localities by Tajima's D tests (1989) and Fu & Li's F tests (1993). Secondary introgression between divergent lineages is an important source of departure from the mutation/drift equilibrium and is a process likely to affect the D and F statistics. Secondary introgression between populations at equilibrium is expected to generate positive D and F values, while the effect of a combination of demographic disequilibrium within populations with secondary intergradations is difficult to predict. In an attempt to identify the possible effect of demography before the contact, D and F were also estimated within each genetic clade for loci that were not differentially fixed at extreme populations.

Because coalescent model assumes no recombination, a four-gamete test (Hudson & Kaplan 1985) was performed on each sequence set and putative recombinants were removed from the dataset by either shortening the length of the studied gene or discarding the most obvious recombinants. A multi-locus Hudson-Kreitman-Aguadé test (HKA, 1987) was performed separately for the Northern and Southern *B. thermophilus* clades with one sequence of *B. azoricus* as outgroup via the software HKA (J. Hey's web page: http://lifesci.rutgers.edu/~heylab/HeylabSoftware.htm#HKA). This test compares divergence and polymorphism at several loci to detect if some of these loci display departure to neutral evolution. Discrimination between bottleneck versus selective sweep was also tested separately using sweep_bott (Galtier & Depaulis 2000) with 100 000 iterations at the first step and 1 000 000 iterations at the second step with 20 optimization processes using a theta range from 1 to 30. The test was done for each clade using all genes, except Sulfo2, for which no divergent clade has been detected. Because this software assumes an Infinite Site mutation model and no reverse substitution, sites that did not follow this assumption were discarded from the dataset.

Fitting an Isolation with Migration model between the two geographic groups

Population sizes, migration rates across the barrier, and time of population splitting were estimated by fitting an isolation with migration model (IMa2 program). The HKY mutation model (Hasegawa et al 1985) was chosen because of the detection of sites with more than two character states in the dataset. With two populations, the method derives analytically posterior probability distributions from sampled gene genealogies for effective sizes of both the ancestral and sister clades (θa, θ0, θ1 respectively), the directional gene-flow rates between the derived lineages (m0>1 and m1>0) and the time elapsed since the populations splitting (T). A gene-by-gene estimation of parameters was also performed using the IM software to search for (1) a preferred introgression orientation of each gene and (2) whether hybridization events are recent (introgressed alleles due to a recent secondary contact) or old (ongoing isolation process). To check convergence of the Markov Chain Monte Carlo (MCMC) toward the true stationary distribution, multiple runs were performed using different starting points and autocorrelation between parameter values was assessed over the course of the runs. A swapping procedure was also used to enhance the mixing of chains using a geometric heating scheme with 6 Metropolis coupling chains. Inheritance scalar was assigned to 1 for nuclear genes (autosomal) and 0.25 for mtCOI to adjust for their expected population sizes. To convert IMa2 parameters into divergence time, effective population sizes and migration rates, the geometric mean of mutation rates per year over loci (U) was estimated from a previous calibration published by Faure et al (2009), which was also consistent with values obtained from the divergence time across the Isthmus of Panama when comparing *B. thermophilus* and *B. azoricus*. These parameters were converted into number of individuals within lineage (N) and time in years according to the relationship: $N = \theta / (4UG)$ and $T\ (\text{years}) = T_{IMa} / U$, with θ and T_{IMa} parameters estimated by IMa2 and G being the number of years per generation.

'Hybrid' individuals detection using SAHH RFLP analysis

Because the SAHH nuclear gene displays a strong divergence between northern and southern populations, occurrence of putative introgressed/hybrid individuals between mitochondrial lineages were checked by looking for individuals that possessed one allele typical from each clade (called N/S individuals). A RFLP approach was thus done using a fixed site previously found in the divergence between the two sets of SAHH alleles to detect N/S individuals by screening all individuals from the *Bathymodiolus* collection onto

a 2% agarose gel. The Hinf I restriction enzyme was applied on the SAHH PCR-product fragments to discriminate individuals with and without this restriction site. Incubation of PCR-products was done at 37°C during 1h30' within a 20µl total volume containing 17µl of PCR product, 1X buffer (supplied by the manufacturer) and 10U of Hinf I (Ozyme^{TM}).

Results

Geographic distribution of allozymes

A series of differentiation tests were performed on *B. thermophilus* populations located along the SEPR from 7°25S to 21°33S but unfortunately not on the northern populations for which no frozen material was available. Amongst the eight allozyme loci, four (*Mpi*, *Odh*, *Mdh 1* and *Hk)* were considered as monomorphic with the most frequent allele occurring at a frequency greater than 95% in all populations. The remaining 4 loci (*Pgm*, *Lap*, *Gpi* and *Mdh 2*) were found polymorphic enough to investigate their allelic distribution over the range of the EPR (see in supplementary material: the Table of allele frequencies). The global test of genetic differentiation did not show any significant heterogeneity in allele frequencies among populations (Fst = 0.007, p value > 0.05), and thus rejected the hypothesis of any genetic structure along this portion of the EPR. The Mantel Spearman test failed (p value > 0.17) to reveal any trend toward isolation-by-distance with no obvious relationship between the genetic distance (θ(1-θ)) and the geographic distance between vent fields (Fig 4.8). A Principal Component Analysis of the allozyme dataset (Fig 4.9) also indicated that most of the genetic variance was not explained by geography (see Fig. 4.9: first axis extracted 62.93% of the total sampling variance, Fst: 0.011), although the second axis (29.07% of inertia, Fst: 0.005) separated the 7°25'S population from the most southern populations with a slight shift of allele frequencies at loci *Pgm* and *Lap-1*.

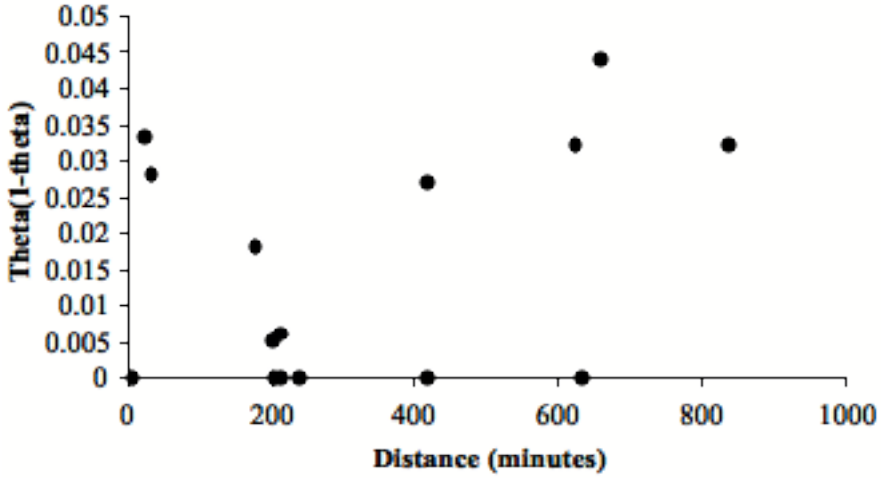


Fig. 4.8 Distribution of allozyme differentiation (measured by using the distance theta(1-theta)) against the geographic distance between sites.

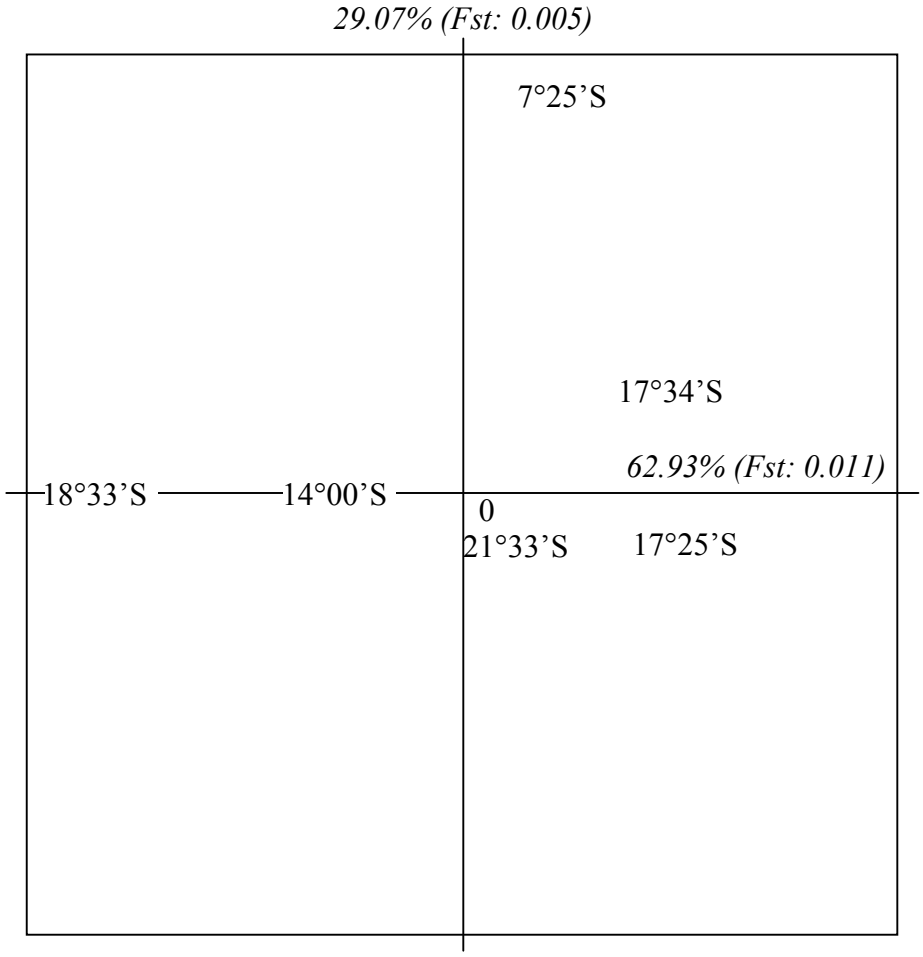


Fig. 4.9 Principal Component Analysis of allozyme frequencies between the southern populations of *Bathymodiolus thermophilus*.
Percentage of inertia for X and Y axis are 62.93% (Fst: 0.011) and 29.07% (Fst: 0.005), respectively.

Polymorphism and divergence from DNA sequences

Networks of alleles revealed a pronounced geographic structure of the mussel *Bathymodiolus thermophilus* between 9°50'N-7°25'S and the most southern latitudes along the EPR. All gene trees, including the mtCOI tree, were shaped into two clades separated by a more or less pronounced divergence (see Fig. 4.10). The SAHH gene revealed the presence of two 2%-divergent clades well established from each part of the previously suggested 7°25'S-14°S mitochondrial break (Plouviez et al 2009, Fig 4.10). SAHH clade 1 contains most of sequences from individuals sampled in 9°50'N and 7°25'S whereas clade 2 corresponds to individuals only sampled from the others southern EPR sites. Two 0.3%-divergent clades were also found in Lysozyme gene but are more difficult to attribute to a particular geographic area. However, a 1bp indel was only found in sequences sampled in the southern EPR sites below the latitude of 7°25S (Fig 4.10), indicating that the 7°25'S-14°S barrier is playing a role in impeding the spread of new mutations. The paralogous gene Sulfo 1 also displayed a clear geographic structure between the north and the south of the barrier with a 0.5% divergence between the two clades whereas the less-sampled paralogous Sulfo 2 gene did not reveal any clear structure (Fig 4.10). This geographic break leads to a great genetic differentiation between the two geographic groups of populations (i.e. below and above the 7°25'S-14°S latitude) with ϕ_{st} values significantly different from zero for all sampled genes (Table 4.2). This north/south differentiation is reinforced by the very steep gradient of the clade-specific allele-frequency for at least 3 loci between 7°25'S and 14°S: a situation usually typifying a secondary contact zone (Fig 4.11).

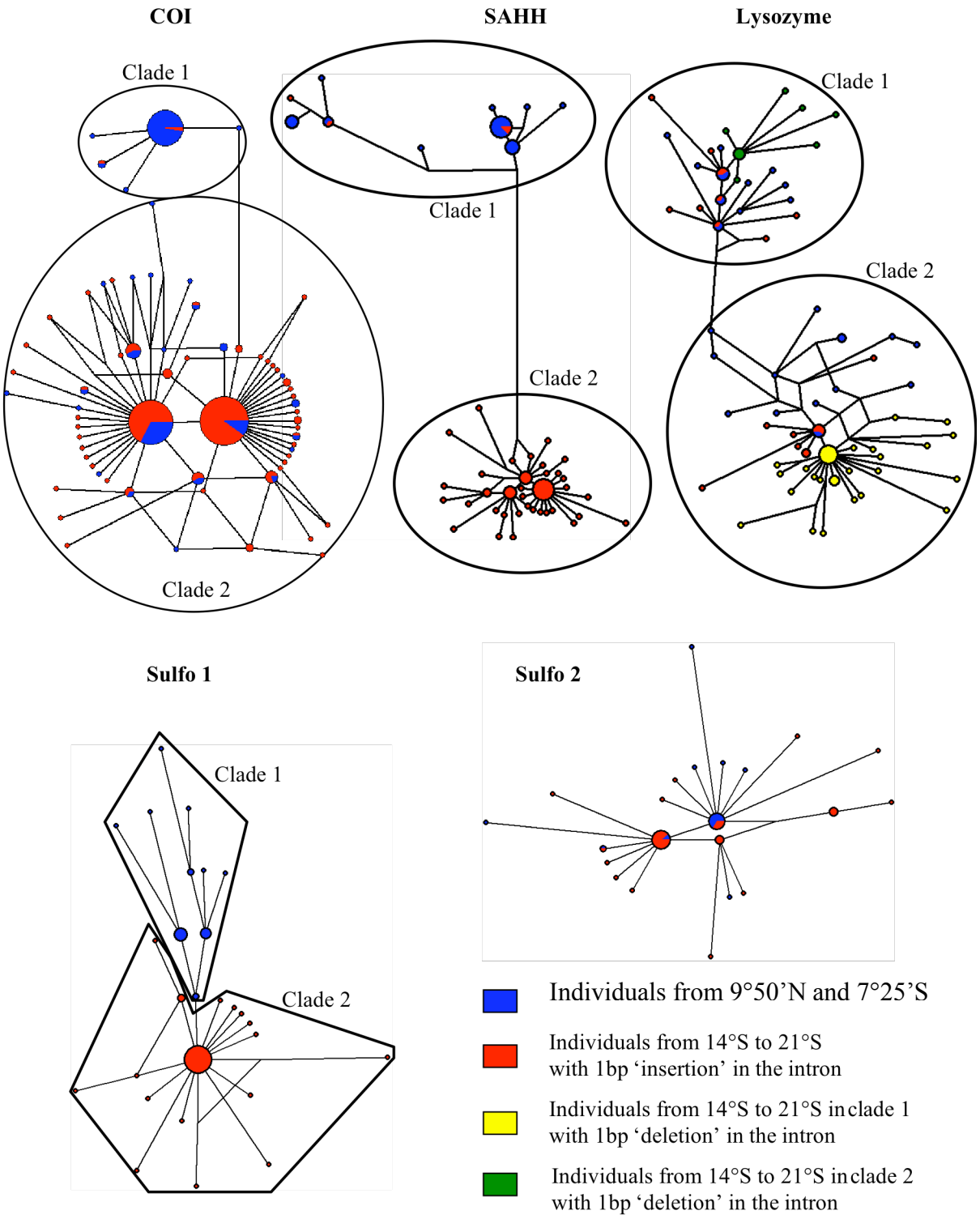


Fig. 4.10 Median Joining Networks on nuclear genes.
SAHH, S-Adenosyl Homocystein Hydrolase; Lyso, Lysozyme; Sulfo 1 and 2, Sulfotransferase paralogues 1 and 2 together with the Cytochrome Oxidase I (mtCOI: adapted from Plouviez et al 2009).

Table 4.2 Summary statistics of nucleotide polymorphism when separated into two clades of divergent alleles corresponding to the separation of the two mussel forms.
N, number of sequences; *S* number of alleles; θ_W, Waterson's theta; π_n, nucleotide diversity; *D*, Tajima's index; *F**, Fu & Li index. ϕ_{st} values correspond to levels of differentiation between the Northern and Southern populations located from each part of the barrier.
*: P< 0.05, **: p<0.01; ***: P<0.001

Locus	Clade	*N*	*S*	θ_W	π_n	*D*	*F**	ϕ_{st}
COI								
	1	43	4	0.0018	0.0004	-1.7624^{NS}	-2.4199^{NS}	0.2500^{***}
	2	218	49	0.0158	0.0029	-2.4330^{**}	-5.2179^{**}	-
SAHH								
	1	35	22	0.0063	0.0074	0.4331^{NS}	-0.3548^{NS}	0.7174^{***}
	2	59	55	0.0151	0.0039	-2.5339^{***}	-5.7349^{**}	-
Lyso								
	1	34	43	0.0085	0.0035	-2.1297^{*}	-4.3284^{**}	0.1548^{*}
	2	61	82	0.0142	0.0044	-2.3880^{**}	-4.9828^{**}	-
Sulfo1								
	1	26	13	0.0165	0.0100	-1.3420^{NS}	-2.6859^{*}	0.4726^{***}
	2	56	19	0.0192	0.0041	-2.5475^{***}	-5.1387^{**}	-
Sulfo2								
		60	30	0.0298	0.0089	-2.2756^{**}	-5.1879^{**}	-

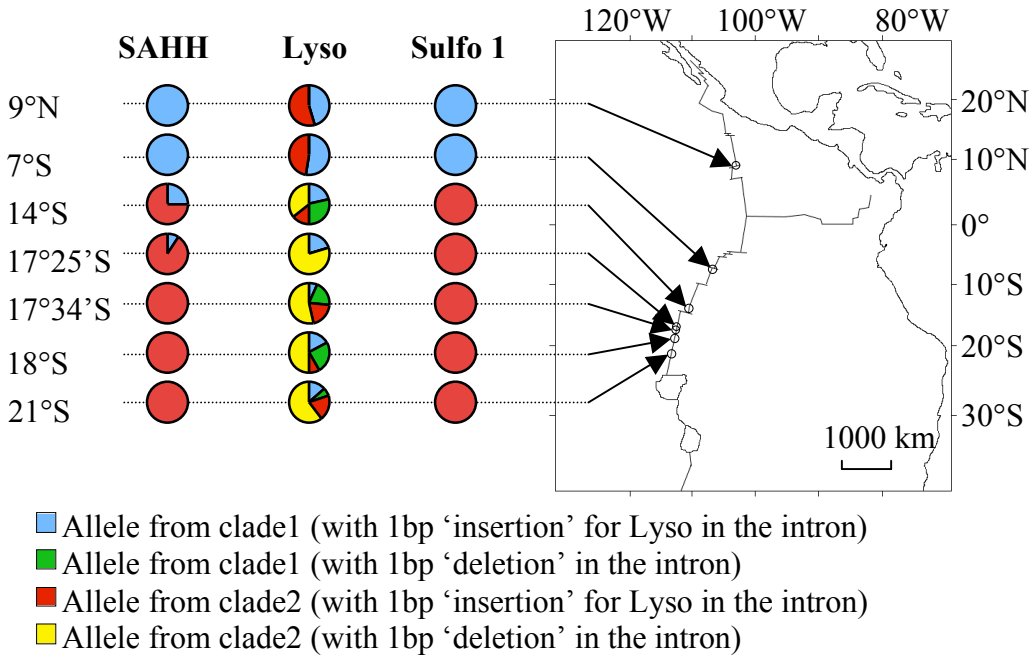


Fig. 4.11 Distribution of divergent alleles along the East Pacific Rise for three nuclear genes.
SAHH, S-Adenosyl Homocystein Hydrolase; Lyso, Lysozyme; Sulfo 1, Sulfotransferase paralogues 1

Tajima's *D* and Fu & Li's *F** indexes estimated from geographic localities (Table 4.3) did not reveal any clear demographic patterns among vent populations (gene diversity was not significantly different from that expected from the neutral accumulation of mutations within each sampled gene). By contrast, these two statistics estimated for both geographic clades were all highly negative and significantly different from neutral expectations in the southern clade (Table 4.2). Departures to neutral expectations were only detected at two loci (Lyso and Sulfo2) in the northern clade. Moreover, the Fu & Li's tests, which are more sensitive to demographic events, provided p-values closest to the 5% threshold when performed on the northern clade. The multi-locus HKA test performed using all genes except Sulfo2, showed a significant departure to neutral evolution (P = 0.0004) in both the northern and southern clades, indicating that at least one gene could be under a strong negative selection or positive selection. Sulfo 1 displayed the highest divergence to polymorphism deviation when compared to other genes. When Sulfo 1 was deleted from the dataset, the HKA test conformed to the model of neutral evolution (P =

0.2812). This gene displayed a polymorphism/divergence ratio of 8.66: a value that is more than eight times greater than the expected one (1.05) with a very low divergence between *B. thermophilus* and *B. azoricus*, a characteristic of balancing or purifying selection. In the northern clade, likelihood ratio tests (LRT) performed by the Sweep_bott analysis using all genes (Table 4.4) also revealed departures from the neutral hypothesis (the bottleneck M2 and/or the selective sweep model M3 were both significantly better than the constant-size model M1). The multiloci approach however did not allow to choose between M2 and M3. When looking at genes separately, the Lysozyme was the only gene that displayed a significant LRT between M1 and M3, indicating that M3 should be more probable than M2 with a moderate selective sweep on the Lysozyme gene. In the southern clade, LRTs also showed that M2 and M3 were more appropriate than M1 (Table 4.4). When looking at genes more specifically, nearly all gene genealogies (with the exception of Sulfo 2) were better explained by the alternative models M2 or M3 suggesting that the bottleneck (M2) model may be valid for, at least the southern clade.

Table 4.3 Summary statistics of nucleotide polymorphism according to locality for mitochondrial and nuclear genes.
Parameters are the same than those in Table 4.2.

Locus	Locality	S	θ_W	π_n	D	F^*
COI						
	9°50N	15	0.007	0.007	0.158	-1.164
	7°25S	20	0.009	0.007	-0.488	-1.165
	14°00S	11	0.005	0.003	-1.446	-2.196
	17°25'S	6	0.003	0.002	-1.365	-1.771
	17°34'S	27	0.011	0.003	-2.259**	-3.924**
	18°33S	13	0.006	0.003	-1.803	-3.008*
	21°33S	10	0.005	0.003	-1.513	-1.939
SAHH						
	9°50N	18	0.009	0.009	-0.490	-0.406
	7°25S	14	0.004	0.007	1.749	1.841**
	14°00S	24	0.011	0.011	-0.097	0.741
	17°25'S	41	0.016	0.010	-1.974*	-2.580**
	17°34'S	11	0.005	0.003	-1.299	-1.507
	18°33S	8	0.003	0.003	-0.887	-1.428
	21°33S	28	0.011	0.006	-1.929*	-2.756*
Lyso						
	9°50N	14	0.005	0.005	-0.065	0.054
	7°25S	15	0.004	0.004	-0.310	-0.399
	14°00S	23	0.006	0.005	-0.852	-1.525
	17°25'S	25	0.008	0.005	-1.970**	-2.446**
	17°34'S	14	0.004	0.003	-0.948	-0.515
	18°33S	26	0.007	0.006	-0.926	-1.394
	21°33S	34	0.009	0.005	-1.633	-2.214
Sulfo1						
	9°50N	3	0.005	0.006	0.794	0.405
	7°25S	10	0.014	0.011	-0.836	-1.593
	14°00S	4	0.006	0.003	-1.747	-2.285
	17°25'S	6	0.009	0.005	-1.894*	-2.505*
	17°34'S	5	0.008	0.004	-1.831*	-2.410*
	18°33S	4	0.006	0.004	-1.865*	-2.419*
	21°33S	2	0.004	0.004	-0.275	-0.115

Table 4.3 suite

Locus	Locality	*S*	θ_W	π_n	*D*	*F**
Sulfo2						
	9°50N	5	0.009	0.006	-1.595	-1.876
	7°25S	9	0.014	0.009	-1.517	-2.074
	14°00S	5	0.008	0.008	0.023	-0.014
	17°25'S	6	0.011	0.008	-1.064	-1.336
	17°34'S	5	0.008	0.007	-0.632	-0.754
	18°33S	9	0.017	0.014	-0.906	-1.047
	21°33S	5	0.010	0.009	-0.826	-0.851

Table 4.4 Likelihoods and likelihood ratio test (LRT) computed via sweep_bott to discriminate selective sweep to bottleneck in northern and southern EPR.
M1, neutral model; M2, bottleneck model; M3, selective sweep model; NS, non-significant. Significances of the LRT were tested according to a $\chi 2$ table.

	Likelihood	SAHH	Lyso	Sulfo1	Sulfo2
North					
M1	-138.255	-40.107	-52.684	-27.066	-18.398
M2	-132.459				
M3	-129.096	-39.301	-48.558	-25.554	-15.683
LRT (significance)	M2-M1: 11.518 (0.5%) M3-M1: 18.312 (2.5%) M3-M2: 6.726 (NS)	1.612 (NS)	8.252 (2.5%)	3.024 (NS)	5.430 (NS)
South					
M1	-135.190	-56.247	-60.755	-	-18.187
M2	-117.380				
M3	-103.950	-46.785	-40.955	-	-16.210
LRT (significance)	M2-M1: 35.619 (0.5%) M3-M1: 62.480 (0.5%) M3-M2: 26.861 (0.5%)	18.924 (0.5%)	39.600 (0.5%)	-	3.954 (NS)

Detection of 'hybrid' individuals using the SAHH gene

Because of its strong divergence across the 7°25'S-14°S barrier, the nuclear gene SAHH was chosen to estimate the number of putative 'hybrids' within populations over the whole EPR. The RFLP analysis of this marker revealed that only 13 individuals displayed at least one nuclear allele corresponding to the northern clade in the southern EPR. Except one individual showing two northern type alleles in the southern populations, all the remaining individuals possessed one allele of each clade. These N/S individuals were mainly located at 14°S and decreased abruptly in frequency further south (Fig 4.12). No 'hybrid' individual was detected at 7°25'S and further north. N/S individuals on SAHH were not associated with a specific mtCOI clade: the SAHH northern-type allele being found both in association with haplotypes from the mtCOI clades 1 and 2 whereas the SAHH southern-type allele was only restricted to the mtCOI clade 2.

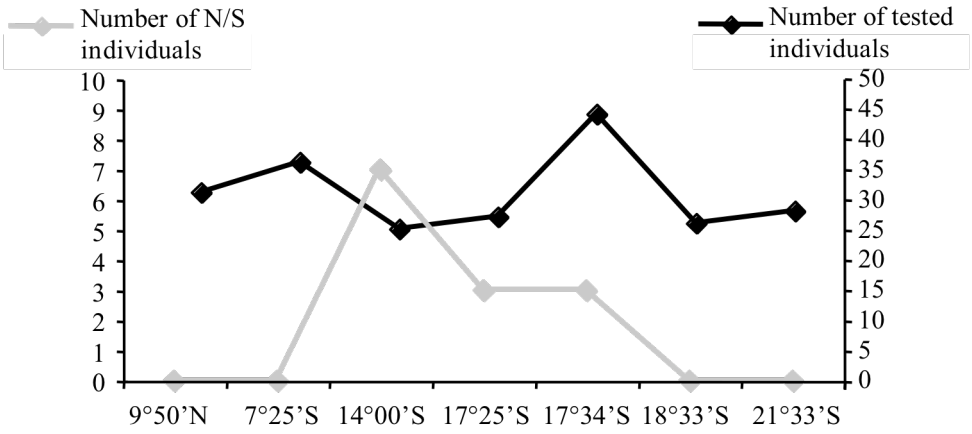


Fig 4.12 Detection of individuals with both North and South type alleles in SAHH or allele type that did not correspond in geographic position across the barrier using RFLP analyses.

Testing the hypothesis of Isolation with migration across the barrier

The Isolation with Migration (IM) model of Hey and Nielsen (2007) was used to test the occurrence of historical gene flow between populations across the barrier and the possibility for a recent gene introgression following a secondary contact. The estimated IMa2 parameters are presented in Table 4.5. Divergence time between the two geographic groups of populations displayed the highest likelihood values around 1-2 Mya (Fig 4.13A). Effective population size of derived populations (N0 for the northern group and N1 for the southern group) were both greater when compared with the ancestral (Na) one (Table 4.5, Fig 4.13B), indicating that actual populations of the vent mussel were both expanding, but at a highest rate for southern populations (N1 being slightly higher than N0).

According to the sofware, m0>1 and m1>0 correspond to north to south and south to north movements of genes when considering migration backward in time. Considering the conventional sense of migration from one population to another (time forward), m0>1 and m1>0 indicated that migration occurred in both directions but is slightly orientated from north to south (Fig. 4.13C, Table 4.5). The gene-to-gene approach confirmed the occurrence of gene exchanges across the barrier but without any clear directionality as expected when a genetic barrier is expected (Table 4.6). Mean times at which gene flow events occurred are very close from zero therefore suggesting that the northern and southern populations were recently in contact when considering the 1.3 Mya value estimated as a starting point for the population splitting (Table 4.6).

Table 4.5 Estimates and the 95% Highest Posterior Density (HPD) intervals of demographic parameters from IMa2 multiloci analysis.
For all parameters, labels 0 and 1 correspond to the northern and southern groups of populations, respectively (see details in text) and label 'a' corresponds to parameters for the ancestral lineage. The effective size (θ), the time since lineages splitting (t) and migration (m) in which, forward in time, m0>1 represents migration from south to north and m1>0, migration from north to south are parameters directly estimated from IM. These parameters were converted in number of individuals (N) in each lineage and time in years (t).

Parameters	θ_0	θ_1	θ_a	t	m0>1	m1>0
Estimate (HiPt)	9.450	12.150	3.450	0.691	0.095	0.195
Lower 95% HPD	3.750	3.750	1.650	0.447	0.000	0.000
Higher 95% HPD	21.45	164.2	7.350	1.479	1.085	0.645

Parameters	N0 (million)	N1 (million)	Na (million)	t (years million)
Estimate (HiPt)	1.258	41.775	1.963	1.368
Lower 95% HPD	0.554	14.596	1.057	0.991
Higher 95% HPD	2.668	846.091	3.171	1.862

Table 4.6 Number (m#) and time (mtime) of migration events between northern and southern groups of populations.
In forward time, 0>1 and 1>0 correspond to south to north and north to south migration, respectively. Between coma: lower and higher values of 95% Highest Posterior Density

Locus	m0>1#	m1>0#	mtime 0>1 (million)	mtime 1>0 (million)
All loci	7 (0-32)	13 (1-32)	0.001 (0-0.895)	0.001 (0-0.937)
COI	3 (0-13)	3 (0-13)	0.001 (0-0.707)	0.001 (0-0.743)
SAHH	0 (0-11)	2 (0-12)	0.005 (0-0.949)	0.003 (0-0.913)
Lyso	0 (0-12)	3 (0-13)	0.013 (0-0.933)	0.001 (0-0.955)
Sulfo1	0 (0-11)	0 (0-11)	0.005 (0-0.969)	0.433 (0-1.041)
Sulfo2	1 (0-13)	2 (0-13)	0.009 (0-0.913)	0.001 (0-0.995)

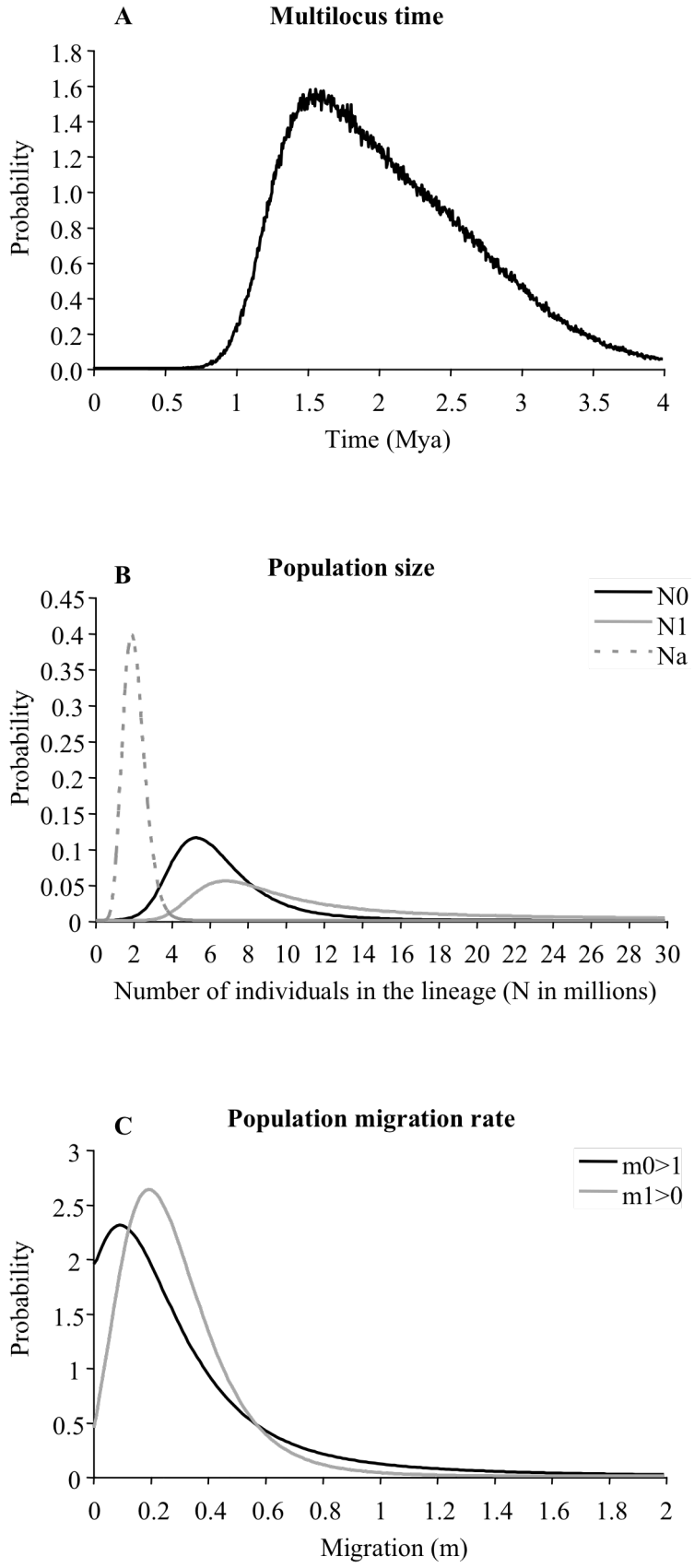


Fig 4.13 Marginal posterior probability distribution for divergence time between lineages (A) effective population sizes of the ancestral population and its two derived ones (B) and directional gene-flow rates between the two mt lineages (C). Estimated by the isolation with migration model IMa. Parameters correspond in those from the Table 4.

Discussion

Addressing more clearly the exact geographic position of a barrier to gene flow in deep-sea vent mussels

One-dimensional but highly fragmented habitats such as deep-sea hydrothermal vents represent an ideal environment to test for stepping-stone or isolation-by-distance models of populations (Vrijenhoek 1997). However, hydrothermal activity is also ephemeral and moves along the ridge axis depending on the dynamics of the heat convection associated with movements of the underneath magmatic chamber (Watremez et al 1990, Kervévan 1991). Moreover, eruptive phases lead to massive fauna extinction at the scale of a given vent field (Haymon et al 1991, Tunnicliffe et al 1997, Marcus et al 2009) able to seriously alter expectations of such population models (Jollivet et al 1999). Because of its unpredictable nature and the level of fragmentation, one would expect that vent species should develop high dispersal capabilities in order to promote long-distance (re)colonisation of the new venting sites. However, most vent species exhibit a lecithotrophic mode of larval development known to typify short-term dispersal ability in coastal and intertidal environments (Chevaldonné et al 1997, Tyler & Young 1999, Jollivet et al 2000, Marsh et al 2001). Paradoxically, genetic structures that fit the isolation-by-distance model were very scarce and often related to the presence of, at least, one geographic barrier (cf the Hess Deep triple junction: see France et al 1992, Black et al 1994, Jollivet et al 1995, Vrijenhoek 1997). On the contrary, many species, such as the polychaete *Alvinella pompejana* (Jollivet et al 1995), the bivalve *Bathymodiolus thermophilus* (Craddock et al 1995) or the gastropods *Lepetodrilus pustulosus* and *Eulepetopsis vitrea* (Craddock et al 1997) displayed a lack of geographic structure and were supposed to follow an island-like model of dispersal, at least over the range of the North EPR.

Species from the deep-sea *Bathymodiolus* genus has thus been previously described as a challenger for long-distance dispersal because of its planktotrophic mode of larval development. For example, occurrence of amphi-Atlantic species with virtually no genetic divergence, suggested that gene flow may still occur across the Atlantic (Olu-LeRoy et al 2007). Absence of genetic differentiation was also reported at the scale of the Gulf of Mexico between populations from the Mississippi Canyon and Alaminos Canyon, 550 km apart, by Carney et al (2006) for the seep species *B. childressi*. In the oriental Pacific, Craddock et al (1995) also failed to show any isolation-by-distance for *B. thermophilus*

when comparing allozyme distributions between the Northern EPR and the Galapagos sites, and subsequently concluded about extended dispersal capabilities of the mussel.

Recently, Won et al (2003) detected an abrupt change in the frequencies of mitochondrial haplotypes between both the Southern and the Northern parts of the EPR and the southern EPR and the Pacific-Antarctic ridge located further South (32-38°S) and subsequently suggested that *Bathymodiolus* may also be sensitive to physical barriers such as microplate and or transform faults. In the present study, the isolation-with-migration model using the multiloci approach revealed a 1-2 Mya divergence time between geographic groups from each side of the Equator. These results are consistent with the dating of the ~1.3 Mya splitting event occurring between Northern and Southern EPR estimated from the multi-species approach (Plouviez et al 2009), and hypothesised to be due to the formation of the Gofar/Discovery multiple transform fault complex about 1-2 Mya near 4°S (Kureth & Rea 1981; Naar & Hey 1989; Francheteau et al 1990). Surveys of He-3 plumes produced by the venting activity along the East Pacific Rise also indicated the occurrence of a westward flow centered at 15°S (Lupton & Craig 1981, Lupton 1998). This flow, possibly linked to the anticyclonic circulation of water masses in the eastern Pacific (Fujio & Isamato 1991), creates strong cross-axis currents able to produce a hydrodynamic barrier at these latitudes. If the deep transverse water-masse circulation could have primarily reinforced isolation between Northern and Southern EPR, putative fluctuations in the strength and orientation of the water movements over time may have also played a crucial role in allowing transient genetic exchanges across the barrier, particularly for species capable of reaching the upper water column, such as *Bathymodiolus* (Arellano & Young 2009). This may explain why *Bathymodiolus* displays a wider clinal distribution of mtCOI haplotypes when compared to other vent species such as the archaeogastropod limpets or alvinellid species for which populations are abruptly separated between 9°50N and 7°25S (Plouviez et al 2009), these species being channelized in the axial valley of the ridge with a great retention around source populations (Chevaldonné et al 1997, Adams & Mullineaux 2008).

In *B. thermophilus*, the precise location of the barrier was difficult to estimate because of the clinal distribution of the three main mtCOI haplotypes from 9°50N to 17°25S (Plouviez et al 2009). In the present study, even if allozyme loci failed to display any significant geographic patterns along the EPR, results from the PCA indicated the

relative isolation of mussels located at 7°25'S when compared to the most southern populations and thus were in agreement with the hypothesis of 7°25'S-14°S barrier to gene flow. In addition, the distribution of two sets of divergent alleles at four nuclear genes confirmed the 7°25'S-14°S position of a barrier with significant genetic differentiation (ϕ_{st}) values for all genes between these two regions. This geographic break also corresponded to a strong divergence between two clades of alleles at the SAHH locus and a similar north/south distribution of clades starting from 7°25'S was also depicted in the gene Sulfo1. For these genes and the mtCOI, the distribution of specific-clade alleles reveals an abrupt cline of frequencies between 7°25S and 14°S. The gene Lyso was the only gene exhibiting two sets of divergent alleles that were well mixed within each population over the whole EPR. A 1bp-indel was however only observed at latitudes greater than 7°25S for both allelic clades. Based on the allele network, the deletion appears to have been recently derived and is present at a high frequency in the most southern populations, at least within clade 2. Even if this mutation is homoplasic between clades and not produced by recombination, this finding reinforces the hypothesis that the barrier locates between 7°25S and 14°S but also indicates that the barrier is nearly impermeable for newly appeared mutations. Consequently, the use of multiple nuclear loci allowed here to locate more precisely this barrier between 7°25'S and 14°S: a situation which is best explained by the strong anticyclonic water-masse circulation in this region.

Variations of the tension-zone position across genes reveal the existence of a semi-permeable genetic barrier

Evolutionary forces at play around the barrier, however pose questions about the evolution of the allopatric separation of the two groups of mussel populations (9°50'N-7°25'S and 14°S-21°33'S) and the presence of a putative genetic barrier that would also involve partial, pre- and/or post-zygotic reproductive isolation between the two geographic forms of *B. thermophilus*. Migration across the equatorial barrier has been already proposed by Plouviez et al (2009) because of the lack of reciprocal monophily between northern and southern populations of several taxa when using mtCOI. In *B. thermophilus*, the two divergent mtCOI lineages were only mixed in the northern populations suggesting that migration may be orientated from south to north. Such directionality was not detected at the multilocus scale and even found in the reverse direction for some nuclear genes when using the software IMa2. Moreover, the very few N/S individuals (13) detected so

far from the RFLP analysis of the SAHH gene suggested that the tension zone may be positioned further south between 14°S and 17°34S. Discrepancies between nuclear and mitochondrial loci when estimating migration rates across a barrier have been highly documented during the process of allopatric isolation. Most of the time, mitochondrial haplotypes become rapidly fixed from both sides of the barrier whereas nuclear alleles are still able to introgress from one genome to another depending on selective forces acting on the gene itself. This is mostly due to both the high evolutionary rate of mitochondrial genes and their 4-times smaller effective population size. As an example, an extensive allele introgression between nuclear genomes has been found between the geographically-fixed mitochondrial lineages of *Bathymodiolus puteoserpentis* and *Bathymodiolus azoricus* that inhabit deep-sea hydrothermal vents from the Mid Atlantic Ridge (Faure et al 2009). However, the reverse situation is also documented (but see Rawson & Hilbish 1998, Arnaud-Haond et al 2003) and mostly relies on an initial low mitochondrial haplotype diversity, a sex-biased migration or a selective sweep affecting the whole mitochondrial genome. In the present study, the mitochondrial haplotype admixture was only found on the northern side of the barrier whereas putative 'hybrids' were only found on the southern side of the barrier for nuclear genes. The non-overlapping of tension zones when positioned using either mitochondrial or nuclear genes is a good indication that the barrier is semi-permeable since the strength and direction of migration greatly varied across genes. Such a finding is in agreement with the occurrence of a genetic barrier that can act as a filter for allele introgression in both directions. If right, the size of the contact zone between previously-isolated mussel populations largely exceeds the ridge portion found between 7°25'S and 14°S and could only be explained by either very great dispersal capabilities of the mussel larvae or a poorly selective genetic barrier. The fact that mitochondrial haplotypes only admix in the north part of the EPR is a convincing argument of repeated asymmetric introgression of mitochondria from *south* to *north*, and suggests that reproductive isolation between these species is incomplete and asymmetric (i.e., weaker for the cross between a *north* female and a *south* male). However, asymmetric introgression does not explain why SAHH N/S 'hybrid' and COI admixture zones are not superimposed. Such a discrepancy may therefore suggest that gametic incompatibilities also occur on the other side of the EPR, at least for the SAHH gene. The fact that the position of the tension zone is mainly dependent on the gene under scrutiny is therefore a convincing argument toward the existence of a semi-permeable genetic barrier acting on

the deep-sea mussels over a great portion of the ridge and not only a physical point barrier to dispersal.

Testing the hypothesis 'Population expansion in the south' from a multilocus dataset

Expansion of the southern populations after the isolation process was already hypothesized from the multi-species approach because significant negative Tajima's D were only found in the southern clade in all species (Plouviez et al 2009). Such an assumption however requires to be also tested using a multigene approach. The Isolation-with-Migration model fitted on our dataset was consistent with a demographic population expansion of *B. thermophilus* in the southern EPR but also suggested that populations from the north part of the EPR also expanded slightly since both derived effective population sizes appear to be greater than the ancestral one: a situation also encountered in the mid-Atlantic Ridge for the *Bathymodiolus* mussels (Faure et al 2009). Tajima's D and Fu & Li F* statistics were tested both within each locality (where a clade admixture may be present) and each clade of alleles (supposed to reflect each derived taxa after the isolation process). Indeed, local hybridization (clade admixture) would locally enhance D and F* to positive values whereas population expansion should produce significant negative values for all loci. Tests applied at the population scale (i.e. within each vent field) were generally not significantly different from the neutral expectations. Allele admixtures were however detected between the northern and southern clades from 9°50'N to 14°S when looking at gene networks and genetic exchanges across the barrier have been widely observed until 17°34'S. Migration after isolation therefore may have masked the expansion of populations. On the contrary, Tajima's D and Fu & Li F* were all significantly negative in the southern clade and, thus in accordance with the expansion hypothesis. However, while smallest, some genes also displayed significant negative values in the northern clade. A more sophisticated likelihood method was thus used to discriminate between a selective sweep model and a bottleneck model (see: Galtier and Depaulis 2000). The Sweep_bott test gave interesting information about sweep versus bottleneck hypotheses. Using the multiloci approach, departures to the mutation/drift equilibrium under neutrality were detected both in Northern and Southern clades and found compatible with either a bottleneck or a selective sweep model. However, the selective sweep model better fitted gene genealogies within the southern clade. When looking at genes separately, only one gene (Lyso) displayed a significant departure to the constant-size model for both clades,

indicating that at least one positive selective effect may have impacted the genetic structure of *Bathymodiolus thermophilus* during the isolation process. A moderate selective sweep may have therefore affected the lyzozyme gene on one side of the EPR and further spread throughout the barrier. Such a scenario could explain why the two divergent clades are presently well mixed over the whole EPR. Monolocus likelihood ratio tests were also significantly different from neutral expectations at COI and SAHH for the southern clade. Together with the lack of polymorphism found at Sulfo1 for which very few singletons have been detected in the same clade, these departures from the constant-size model are in good agreement with Tajima's and Fu & Li's tests and fully support the hypothesis of a population expansion in the south during the isolation period.

Present results using nuclear genes thus confirmed the hypothesis of a southern demographic expansion in the EPR as previously proposed by Plouviez et al (2009). The greatest tectonic instability of this region and its subsequent higher extinction rate when compared with the northern province could explain why such a bottleneck was only detected in the south. The supposedly more probable migration route from north to south obtained from the multi-loci analysis and the relative strength of the 7°25'S-14°S barrier reinforces this scenario as it seems to prevent the progressive and stepwise colonisation of the southern EPR by populations from the north: a scenario that would have also led to an expanding phase.

Acknowledgments

We thank the chief scientists and 'Nautile' and 'Alvin' crews for their technical support and efforts during the oceanographic expeditions: HOT96 and BIOSPEEDO2004. We are very grateful to Claire Daguin-Thiébaut and Frédérique Viard for collecting/preserving mussels during the BIOSPEEDO cruise. We are also truly indebted to the sequencing genomic plateforms: GENOMER (Station Biologique de Roscoff, France) and GENOSCOPE (Evry, France) for the sequencing of the mtCOI fragment and our clone collections, respectively. This work was supported by the GDR Ecchis and the ANR-06-BDV-005 (Deep Oases: coord. D. Desbruyères). We also want to thank A. Tanguy for his contribution to the *B. azoricus* cDNA library construction, which provided exonic sequences for the nuclear genes and the 'Bivalvomix' GIS programme (coord. N. Bierne) that partly supported the sequencing costs of the library together with the NoE "Marine Genomics Europe". S. Plouviez and B. Faure were supported by a PhD grant from the

Université Pierre et Marie Curie and the NoE "Marine Genomics Europe" (Fish & Shellfish node), respectively.

References

Adams DK, Mullineaux LS (2008) Supply of gastropod larvae to hydrothermal vents reflects transport from local larval sources. *Limnology and Oceanography* **53**: 1945-1955.

Arellano SM, Young CM (2009) Spawning, development and the duration of larval life in a deep-sea cold-seep mussel. *Biological Bulletin* **216**: 149-162.

Arnaud-Haond S, Bonhomme F, Blanc F (2003) Large discrepancies in differentiation of allozymes, nuclear and mitochondrial DNA loci in recently founded Pacific populations of the pearl oyster *Pinctada margaritifera. Journal of Evolutionary Biology* **16**: 388-398.

Avise JC (1994) *Molecular markers, Natural History and Evolution*. New York: Chapman & Hall.

Barton NH (1979) Gene Flow Past a Cline. *Heredity* **43**: 333-339.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16**: 113-148.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37-48.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) Genetix 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, Adaptations, UMR 5000, Université de Montpellier 2. Available at: www.genetix.univ-montp2.fr/ genetix/genetix.htm

Bierne N, Tanguy A, Faure M, Faure B, David E, et al (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. 689

Black MB, Lutz RA, Vrijenhoek RC (1994) Gene Flow among Vestimentiferan Tube Worm (*Riftia pachyptila*) Populations from Hydrothermal Vents of the Eastern Pacific. *Marine Biology* **120**, 33-39.

Boutet I, Tanguy A, Le Guen D, Piccino P, Hourdez S, Legendre P, Jollivet D (2009) Global depression in gene expression as a response to rapid thermal changes in vent mussels. *Proceedings of the Royal Society of London Series B*. doi:10.1098/rspb.2009.0503

Carney SL, Formica MI, Divatia H, Nelson K, Fisher CR, Schaeffer SW (2006) Population structure of the mussel "*Bathymodiolus*" *childressi* from Gulf of Mexico hydrocarbon seeps. *Deep-Sea Research* I **53**: 1061-1072.

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.

Craddock C, Hoeh WR, Lutz RA, Vrijenhoek RC (1995) Extensive Gene Flow among Mytilid (*Bathymodiolus thermophilus*) Populations from Hydrothermal Vents of the Eastern Pacific. *Marine Biology* **124**, 137-146.

Craddock C, Lutz RA, Vrijenhoek RC (1997) Patterns of dispersal and larval development of archaeogastropod limpets at hydrothermal vents in the eastern Pacific. *Journal of Experimental Marine Biology and Ecology* **210**, 37-51.

Dixon DR, Lowe DM, Miller PI*, et al.* (2006) Evidence of seasonal reproduction in the Atlantic vent mussel Bathymodiolus azoricus, and an apparent link with the timing of photosynthetic primary production. *Journal of the Marine Biological Association of the United Kingdom* **86**, 1363-1371.

Faure B, Jollivet D, Tanguy A, Bonhomme F, Bierne N (2009) Speciation in the deep-sea: multilocus analysis of divergence and gene flow between two hybridizing species of hydrothermal vent mussels. *PlosOne*.

France SC, Hessler RR, Vrijenhoek RC (1992) Genetic differentiation between spatially-disjunct populations of the deep-sea, hydrothermal vent-endemic amphipod *Ventiella sulfuris*. *Marine Biology* **114**, 551-559.

Francheteau J, Armijo R, Cheminee JL*, et al* (1990) 1 Ma East Pacific Rise Oceanic-Crust and Uppermost Mantle Exposed by Rifting in Hess Deep (Equatorial Pacific-Ocean). *Earth and Planetary Science Letters* **101**: 281-295.

Fu YX, Li WH (1993) Statistical test of neutrality of mutations. *Genetics* **133**: 693-709.

Fujio SZ, Imasato N (1991) Diagnostic Calculation for circulation and water mass movement in the deep Pacific. *Journal of Geophysical Research-Oceans* **96**, 759-774.

Galtier N, Depaulis F, Barton NH (2000) Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. *Genetics* **155**, 981-987.

Goudet (1999) **PCA**-GEN, a program to perform principal component analysis (**PCA**) on genetic data, Version 1. 2.

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.

Harrison RG (1990) Hybrid zones: windows on evolutionary processes. *Oxf. Surv. Evol. Biol.* **7**: 69-128.

Hasegawa M, Kishino H, Yano T (1985) Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**: 160-174.

Haymon RM, Fornari DJ, Edwards MH*, et al.* (1991) Hydrothermal Vent Distribution Along the East Pacific Rise Crest (9-Degrees-09'-54'n) and Its Relationship to Magmatic and Tectonic Processes on Fast-Spreading Midocean Ridges. *Earth and Planetary Science Letters* **104**, 513-534.

Hewitt GM (2001) Speciation, hybrid zones and phylogeography – or seeing genes in space and time. *Molecular Ecology* **10**: 537-549.

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.

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, Slatkin M, Maddison WP (1992) Estimation of Levels of Gene Flow from DNA-Sequence Data. *Genetics* **132**, 583-589.

Hurtado L, Lutz R, Vrijenhoek R (2004) Distinct patterns of genetic differentiation among annelids of eastern Pacific hydrothermal vents. *Molecular Ecology* **13**, 2603-2615.

Jollivet D, Chevaldonne P, Planque B (1999) Hydrothermal-vent alvinellid polychaete dispersal in the eastern Pacific. 2. A metapopulation model based on habitat shifts. *Evolution* **53**, 1128-1142.

Jollivet D, Desbruyères D, Bonhomme F, Moraga D (1995) Genetic Differentiation of Deep-Sea Hydrothermal Vent Alvinellid Populations (Annelida, Polychaeta) Along the East Pacific Rise. *Heredity* **74**, 376-391.

Jollivet D, Empis A, Baker MC*, et al.* (2000) Reproductive biology, sexual dimorphism, and population structure of the deep sea hydrothermal vent scale-worm, *Branchipolynoe seepensis* (Polychaeta : Polynoidae). *Journal of the Marine Biological Association of the United Kingdom* **80**, 55-68.

Jolly MT, Viard F, Gentil F, Thiebaut E, Jollivet D (2006) Comparative phylogeography of two coastal polychaete tubeworms in the Northeast Atlantic supports shared history and vicariant events. *Molecular Ecology* **15**, 1841-1855.

Kervévan (1991) Modélisation tridimentionnelle de la convection hydrothermale en zone de rift océanique. Ph.D. dissertation, Université de Bretagne Occidentale, Brest, France.

Kureth CL, Rea DK (1981) Large-Scale Oblique Features in an Active Transform-Fault, the Wilkes Fracture-Zone near 9-Degrees-S on the East Pacific Rise. *Marine Geophysical Researches* **5**: 119-137.

Lupton J (1998) Hydrothermal helium plumes in the Pacific Ocean. *Journal of Geophysical Research-Oceans* **103**, 15853-15868.

Lupton JE, Craig H (1981) A Major He-3 Source at 15-Degrees-S on the East Pacific Rise. *Science* **214**, 13-18.

Maggs CA, Castilho R, Foltz D, Henzler C, Jolly MT, Kelly J, Olsen J, Perez KE, Stam W, Väinölä R, Viard F, Wares J (2008) Evaluating signatures of glacial refugia for north atlantic benthic marine taxa. *Ecology* 89: 108-122.

Marcus J, Tunnicliffe V, Butterfield DA (2009) Post-eruption succession of macrofaunal communities at diffuse flow hydrothermal vents on Axial Volcano, Juan de Fuca Ridge, Northeast Pacific. *Deep-Sea Research Part Ii-Topical Studies in Oceanography* **56**, 1586-1598.

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.

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.

Naar DF, Hey RN (1989) Speed Limit for Oceanic Transform Faults. *Geology* **17**: 420-422.

Olu-Le Roy L, 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.

Pasteur N, Pasteur G, Bonhomme F, Catalan J, Britton-Davidian J (1987) *Manuel technique de génétique par électrophorèse de protéines*. Paris: Lavoisier.

Piccino P, Viard F, Sarradin PM*, et al* (2004) Thermal selection of PGM allozymes in newly founded populations of the thermotolerant vent polychaete Alvinella pompejana. *Proceedings of the Royal Society of London Series B-Biological Sciences* **271**, 2351-2359.

Plouviez S, Shank TM, Faure B, Daguin-Thiébaut C, Viard F, Lallier FH, Jollivet D (2009) Comparative phylogeography among hydrothermal vent species along the East Pacific Rise reveals vicariant processes and population expansion in the South. *Molecular Ecology* **18**: 3903-3917.

Plouviez S, Le Guen D, Lecompte O, Lallier FH, Jollivet D (2010) Determining gene flow and influence of selection across the equatorial barrier of the East Pacific Rise in the tube-dwelling polychaete *Alvinella pompejana*. *BMC Evolutionary Biology* **10**: 220.

Rawson PD, Hilbish TJ (1998) Assymetric introgression of female and male lineage mitochondrial DNA (mtDNA) haplotypes within the European hybrid zone between *Mytilus edulis* and *Mytilus galloprovincialis*. *Evolution* **52**, 100-108.

Raymond M, Rousset F (1995) GENEPOP, version 1.2: population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**: 248-249.

Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**: 4673-4680.

Tunnicliffe V, Embley RW, Holden JF*, et al.* (1997) Biological colonization of new hydrothermal vents following an eruption on Juan de Fuca Ridge. *Deep-Sea Research Part I-Oceanographic Research Papers* **44**, 1627-1644.

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

Vrijenhoek RC (1997) Gene flow and genetic diversity in naturally fragmented metapopulations of deep-sea hydrothermal vent animals. *Journal of Heredity* **88**, 285-293.

Watremez P, Kervévan C (1990) Origin of Variations in Hydrothermal Activity - 1st Results of a Simple Numerical-Model. *Comptes Rendus De L Academie Des Sciences Serie Ii* **311**, 153-158.

Weir B, Cockerham C (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.

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

Supplementary material.

S1: Allele frequencies of enzyme loci for *B. thermophilus* populations of each vent field of the SEPR.N, population size; H exp, expected heterozygosity; H obs, observed heterozygosity.

	7°25'S	14°S	17°25'S	17°35'S	18°33'S	21°33'S
Pgm						
(N)	10	38	137	23	33	103
Allele						
64	0	0	0.0036	0	0	0
74	0.05	0.0132	0.0109	0	0.0455	0.0194
81	0.7	0.5789	0.5219	0.5652	0.5758	0.5194
85	0	0	0.0036	0	0	0.0146
89	0.25	0.3816	0.4307	0.3696	0.3636	0.4223
95	0	0.0263	0.0292	0.0652	0.0152	0.0243
H exp.	0.445	0.5184	0.5412	0.5397	0.534	0.5507
H obs.	0.5	0.5789	0.5255	0.6087	0.3636	0.4272
Mpi						
(N)	10	38	137	23	33	103
Allele						
100	0.05	0.0395	0.0182	0.0435	0	0
120	0.95	0.9474	0.9781	0.9348	1	0.9903
140	0	0.0132	0	0.0217	0	0.0097
160	0	0	0.0036	0	0	0
H exp.	0.095	0.1008	0.043	0.1238	0	0.0192
H obs.	0.1	0.1053	0.0438	0.087	0	0.0194
Odh						
(N)	10	38	137	23	33	103
Allele						
60	0	0.0658	0.0073	0	0.0303	0.0097
80	1	0.9342	0.9526	1	0.9545	0.9612
100	0	0	0.0401	0	0.0152	0.0291
H exp.	0	0.1229	0.091	0	0.0877	0.0752
H obs.	0	0.1316	0.0949	0	0.0909	0.0777
Lap						
(N)	10	38	137	23	33	103
Allele						
75	0	0.0132	0	0	0	0.0049
82	0.9	0.7632	0.719	0.7609	0.7879	0.7233
90	0	0.0132	0.0073	0	0.0303	0.0097
100	0.1	0.2105	0.2482	0.2174	0.1364	0.233
110	0	0	0.0255	0.0217	0.0455	0.0291
H exp.	0.18	0.3729	0.4208	0.3733	0.3577	0.4216
H obs.	0.2	0.3684	0.4088	0.3043	0.303	0.4078

	7°25'S	14°S	17°25'S	17°35'S	18°33'S	21°33'S
Gpi						
(N)	10	38	137	23	33	103
Allele						
88	0	0.0132	0.0219	0.0435	0	0
100	0	0	0	0	0	0.0049
112	0.85	0.9605	0.938	0.8913	0.9545	0.9466
116	0.05	0	0	0	0	0
120	0.1	0.0263	0.0401	0.0652	0.0455	0.0485
H exp.	0.265	0.0765	0.1181	0.1994	0.0868	0.1016
H obs.	0.3	0.0789	0.1095	0.2174	0.0909	0.1068
Mdh low						
(N)	10	38	137	23	33	103
Allele						
74	1	1	0.9927	1	1	1
100	0	0	0.0073	0	0	0
H exp.	0	0	0.0145	0	0	0
H obs.	0	0	0.0146	0	0	0
Mdh high						
(N)	10	38	137	23	33	103
Allele						
100	0	0.1711	0.073	0.0217	0.2727	0.1019
120	1	0.8289	0.927	0.9783	0.7273	0.8981
H exp.	0	0.2836	0.1353	0.0425	0.3967	0.1831
H obs.	0	0.3421	0.146	0.0435	0.5455	0.2039
Hk						
(N)	10	38	137	23	33	103
Allele						
70	0.05	0	0.0036	0.0217	0	0
100	0.9	1	0.9964	0.9783	1	1
130	0.05	0	0	0	0	0
H exp.	0.185	0	0.0073	0.0425	0	0
H obs.	0.2	0	0.0073	0.0435	0	0

4.3. *L. elevatus*

4.3.1. Introduction

Le complexe d'espèces *L. elevatus* situé sur l'EPR, compte deux espèces majoritairement abondantes divergentes de 6,5% sur le gène mtCOI (Matabos et al 2008b) : *L. elevatus* McLean (appelée espèce Nord dans notre étude) distribuée de 21°N à 9°50'N et une seconde espèce (que nous appellerons espèce Sud) de 9°50'N à 23°S (Johnson et al 2008, Plouviez et al 2009). Le site 9°50'N constitue donc une zone de sympatrie entre ces deux espèces cryptiques, mais l'espèce Sud y est majoritaire. Cette distribution géographique suggère une spéciation allopatrique de ces deux espèces, datée à environ 11,6 Ma et une barrière géographique probablement située entre 13°N et 9°50'N, coïncidant temporellement et spatialement avec la réorientation de la ride Mathématicienne (Mammerickx & Klitgord 1982, Plouviez et al 2009). Le processus de spéciation a pu se renforcer par une spéciation écologique, l'espèce Nord étant trouvée principalement en association avec les siboglinidae *Riftia pachyptila* et l'espèce Sud avec les moules hydrothermales *B. thermophilus*. Néanmoins, la séparation des 2 espèces n'est pas complète, des individus hybrides ayant été détectées à 9°50'N à partir d'un génotypage allozymique (Matabos et al 2008b). Le statut d'hybride entre ces espèces cryptiques n'a pour l'instant pas été évalué de façon certaine car les individus séquencés sur l'ADN mitochondrial ne correspondaient pas aux individus génotypés en allozymes. Une telle approche mériterait d'être effectuée en élargissant le champ d'investigation géographique à l'ensemble de l'EPR. De plus, le second événement de vicariance plus récent (1,3 Ma) observé pour toutes les espèces entre les populations sud et nord de l'EPR semble avoir également impacté l'espèce Sud de *L. elevatus* à l'équateur, en séparant la population de 9°50'N des autres populations situées plus au sud. Cet effet est observable en comparant les séquences du gène mtCOI entre individus de la lignée sud. La mise en place d'une barrière secondairement à la séparation des espèces de *L. elevatus* pourrait jouer un rôle fondamental dans la dynamique d'hybridation des espèces, notamment en termes de longueur de la zone d'introgression et nécessite une attention particulière.

Nous avons ainsi utilisé, en plus du gène mtCOI, trois gènes nucléaires séquencés sur des individus répartis de 21°N à 21°S, afin de mieux comprendre le rôle des barrières

(physique et génétique) 13°N-9°50N et 9°50'N-7°25S dans la structuration du complexe d'espèces et la dynamique d'hybridation.

4.3.2. Gène mitochondrial COI

4.3.2.1. *Doubles pics sur les chromatogrammes mtCOI de certains individus*

Sur l'ensemble des individus séquencés le long de l'EPR, six individus appartenant à la zone de contact entre les 2 espèces cryptiques (9°50'N) présentaient des profils de séquences COI atypiques. En effet, si les autres individus présentaient des chromatogrammes clairs correspondant à l'une ou l'autre des lignées mtCOI, les chromatogrammes de ces 6 individus (séquencés en direct) comptaient des doubles pics suggérant l'amplification de 2 haplotypes (Fig. 4.14). Bien que des exceptions existent (e.g. Skibinski et al 1994 ; Zouros et al 1994), le génome mitochondrial est théoriquement hérité uniquement par la voie maternelle et ne devrait donc pas présenter de tels profils. De plus, les doubles pics observés correspondent systématiquement aux pics attendus pour les mutations fixées de chacune des 2 espèces. Ainsi, ces individus présenteraient à la fois l'haplotype de l'espèce nord et l'haplotype de l'espèce sud.

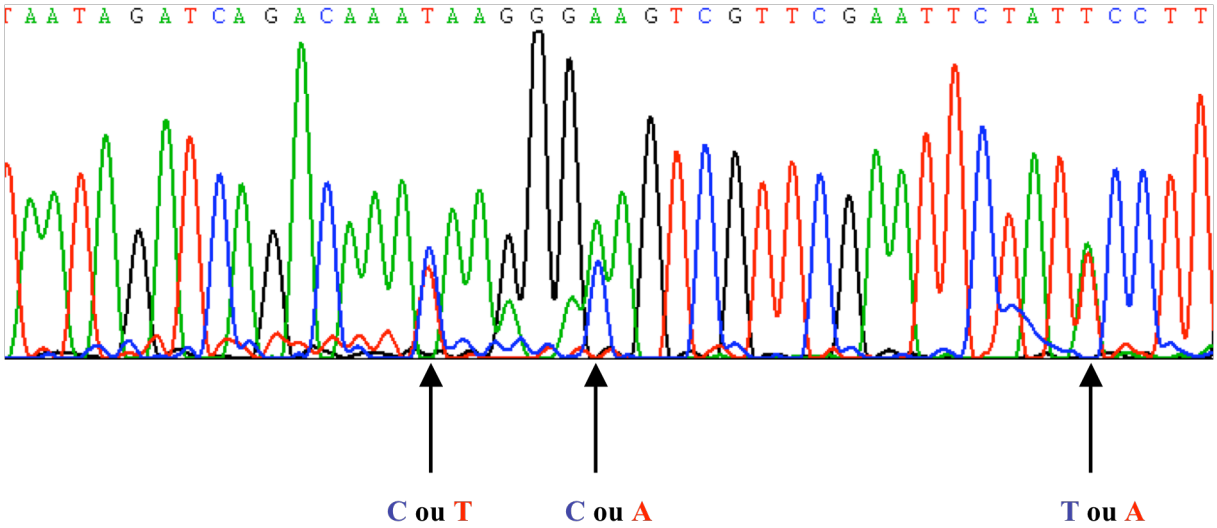


Fig. 4.14 Portion de chromatogramme présentant des doubles pics correspondant respectivement à une base fixée chez l'espèce Nord (indexée en bleu sous le chromatogramme) et une base fixée chez l'espèce Sud (indexée en rouge).

Pour nous assurer de la présence de ces 2 types de mitochondrie et valider l'hypothèse d'hétéroplasmie chez ces 6 individus 'hybrides', nous avons cloné leur produit de PCR (clonage individuel). Les clones repiqués (8 par individus lorsque les conditions le permettaient) sont ensuite amplifiés via des amorces plasmidiques selon le protocole décrit en annexe 3. Une digestion enzymatique de l'ADN (1 µl d'enzyme de restriction Sac1 à une concentration de 20U/µl, 2,5 µl de tampon 10X associé à cette enzyme, 5 µl de produit de PCR, réalisée sur les clones repiqués, et 12,5 µl d'H_20, 3h à 37°C) permettant de générer des patrons RFLP discriminant les mitochondries de type nord et sud, a ensuite été réalisée sur ces clones repiqués. Ces individus présentaient bien les 2 types de profils de RFLP (Fig. 4.15).

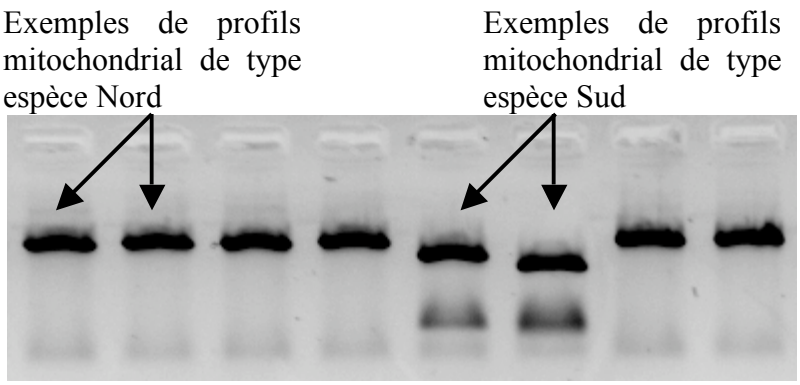


Fig. 4.15 Visualisation sur gel d'agarose des deux types de profil RFLP sur les 8 clones d'un individu présentant les deux types de mitochondrie.
Chaque puit correspond un profil RFLP réalisé sur un clone repiqué.

Nous avons séquencé alors les clones correspondant à ces 2 profils RFLP, vérifiant ainsi qu'ils correspondaient bien aux profils mitochondriaux des 2 espèces cryptiques. Les 2 types de mitochondries sont donc bien amplifiés pour ces 6 individus échantillonnés à 9°50'N aux chromatogrammes ambigus. Bien que les témoins négatifs de PCR soient sans ambiguïté, une seconde PCR et un second séquençage en direct de ces individus ont été réalisés afin d'éliminer de possibles contaminations de produits d'amplification (lors de la purification). Les mêmes résultats ont été obtenus (remarque : ces individus n'ont pas été utilisés dans Plouviez et al 2009).

4.3.2.2. *Hypothèse de double hérédité mitochondriale*

En général, l'ADN mitochondrial des animaux est exclusivement hérité maternellement (Birky 2001) et toutes les cellules d'un même organisme présentent la même copie de cet ADN, hormis quelques rares mutations (homoplasmie). Cependant, certaines espèces comptent parfois plusieurs ADN mitochondriaux au sein d'une même cellule : c'est l'hétéroplasmie. L'ensemble des ADN mitochondriaux au sein d'une cellule peut alors être considéré comme une population. Un ADN mitochondrial mutant pourra ainsi, par dérive, transmettre son ADN à une cellule fille (issue de la réplication de la cellule initiale) et donc, si cette mutation est délétère, diminuer la valeur sélective de cette cellule (moins de mitochondries fonctionnelles, Breton et al 2007). L'ADN mutant entrera ainsi en compétition, lors de sa réplication, avec les ADN non-mutants. Des processus de sélection contre les mutants mitochondriaux incapables d'effectuer la phosphorylation oxydative ont ainsi été mis en évidence chez la levure (Reid 1980, Birky 1995). L'hérédité parentale unique est un mécanisme qui est supposé avoir été favorisé car elle permettrait d'éviter ces conflits intracellulaires (Rand 2001). Cependant, certaines espèces présentent une double hérédité mitochondriale, notamment les moules marines de l'ordre des Mytiloida, les moules d'eau douces de la superfamille des Unionoidea ou les palourdes de l'ordre des Veneroida (e.g. Skibinski et al 1994, Liu et al 1996, Passamonti & Scali 2001). La présence de deux mtCOI appartenant chacune à une espèce cryptique différente semble donc indiquer que *L. elevatus* peut également, au moins chez certains hybrides, probablement la génération F1, présenter une double hérédité parentale.

L'hybridation entre espèces différentes peut avoir des conséquences sur l'hérédité parentale (Wood et al 2003). En effet, chez *Mytilis edulis* et *M. galloprovincialis*, la double hérédité maternelle est généralement présente chez les mâles adultes alors que la mitochondrie mâle est éliminée chez les femelles (même si elles peuvent subsister dans certains tissus des femelles, Dalziel et al 2002, Breton et al 2007). Cependant, les mâles hybrides entre ces deux espèces, avec une mitochondrie mâle provenant de *M. edulis* et une mitochondrie femelle de *M. galloprovincialis*, élimineront plus aisément la mitochondrie mâle qu'un croisement issu de lignées pures (Wood et al 2003). Au contraire, les mâles hybrides ayant une mitochondrie mâle issue de *M. galloprovincialis* et une mitochondrie femelle de *M. edulis* conserveront plus facilement la mitochondrie mâle que chez les *M. galloprovincialis* non hybrides (Wood et al 2003). Les chromatogrammes des *L. elevatus* non hybrides ne présentant pas de double pic apparent, nous pouvons ainsi supposer que seuls les hybrides F1 possèderaient la double hérédité parentale, la mitochondrie mâle étant très probablement éliminée chez les individus non-hybrides lors de la ségrégation méiotique (à l'instar des femelles Mytilidae, Cao et al 2004) à la fois chez les femelles et chez les mâles. Bien que nécessitant de plus amples études, l'hybridation entre les deux espèces cryptiques pourrait avoir favorisé le maintien de la lignée mitochondriale mâle chez certains hybrides, comme chez les *M. galloprovincialis* mâles hybrides (Wood et al 2003). Ceci reste néanmoins une hypothèse à tester.

4.3.3. Analyse des séquences nucléaires : hybridation et barrière(s) au(x) flux de gènes

4.3.3.1. *Recapture de séquences nucléaires*

Le nombre d'individus recapturés pour chaque gène nucléaire est présenté dans le tableau 4.7 (chaque individu ayant été séquencé sur le mtCOI). Le faible nombre de séquences recapturées pour le gène supposé unique de la Textilinine résulte de la suppression de nombreuses séquences correspondant au(x) gène(s) paralogues co-amplifié(s) avec nos amorces (cf. Chapitre 2). De plus, nous n'avons pas réussi à amplifier le fragment de ce gène chez les individus de la population 21°N. Il en résulte que seuls 10 individus ont été recapturés sur les trois gènes nucléaires, mais 60 ont été recapturés sur deux de ces gènes.

Tableau 4.7 Nombre d'individus recapturés aux différents sites hydrothermaux. $N_{3gènes}$ et $N_{2gènes}$ représentent le nombre d'individus recapturés sur les 3 gènes nucléaires et sur 2 des gènes, respectivement.

Localité	Lysozyme	VDG3	Textilinine	$N_{3gènes}$	$N_{2gènes}$
21°N	11	8	0	0	6
13°N	18	16	9	1	12
9°50'N	20	23	10	6	14
14°S	11	13	1	1	8
17°25'S	11	11	3	0	8
17°35'S	12	6	0	0	4
21°33'S	19	10	5	2	8

4.3.3.2. *Polymorphisme et divergence de séquences pour les 3 gènes nucléaires étudiés*

Le clonage ne permettant pas de recapturer les deux allèles de chaque individu, pour limiter les biais dans les fréquences alléliques et simplifier la représentation des arbres phylogénétiques, seule une séquence par individu a été utilisée pour l'ensemble des analyses (jeux de données de qualité 3). Pour chaque gène, un arbre phylogénétique a été construit selon la méthode d'agglomération des plus proches voisins (Neighbour Joining) via le logiciel Mega 4.0 (Tamura et al 2007). Pour les gènes Lysozyme et VDG3, la distribution des allèles a été réalisée au sein de chaque population selon leur clade d'appartenance. La distribution allélique selon le clade d'appartenance n'a pas pu être réalisée sur le gène Textilinine en raison du faible nombre d'individus recapturés au sein de chaque site (généralement inférieur à 10 individus par site, Tableau 4.7).

Les allèles du gène Lysozyme (Fig. 4.16.A) peuvent être divisés en trois clades : le clade 1 comptant l'ensemble des séquences des sites 21°N et 13°N (espèce Nord) tandis que les séquences des individus provenant des sites 14°S à 21°33'S, sont distribuées dans les clades 2 et 3. Les individus en provenance de la zone de contact 9°50'N, se distribuent

dans les 3 clades. La position du clade 3 dans l'arbre phylogénétique montre que ce clade dérive directement du clade 2. Les clades 1-2, 1-3 et 2-3, présentent respectivement une divergence de 0,5%, 0,9% et 0,2%. La distribution des clades au sein de chaque site (Fig. 4.16.B) révèle que :

(1) les populations des champs 21°N et 13°N sont composées uniquement du clade 1,

(2) la population du champ 9°50'N compte des séquences des 3 clades avec une majorité de séquences du clade 2,

(3) les populations des champs situés entre 14°S et 17°34'S présentent un nombre à peu près égal de séquences des clades 2 et 3

(4) la population du champ 21°33'S est constituée presque exclusivement de séquences du clade 3 (une seule séquence recapturée dans le clade 2).

Hormis les individus échantillonnés à 9°50'N, aucun individu ayant ses deux allèles recapturés n'a été détecté hybride (i.e. 1 allèle dans le clade 1 et le deuxième dans le clade 2 ou 3). Les individus présentant deux allèles recapturés confirment la distribution géographique observée sur le jeu de qualité 3. Quatre des six individus (9°50'N) présentant les deux types de mitochondrie ont été recapturés. Parmi ces 4 individus, les deux allèles n'ont été recapturés que pour 1 individu et sont distribués dans les clades 1 et 2, respectivement. Pour les 3 autres individus n'ayant qu'un allèle recapturé, ces allèles se distribuent dans le clade 1 pour deux individus et dans le clade 2 pour le troisième individu.

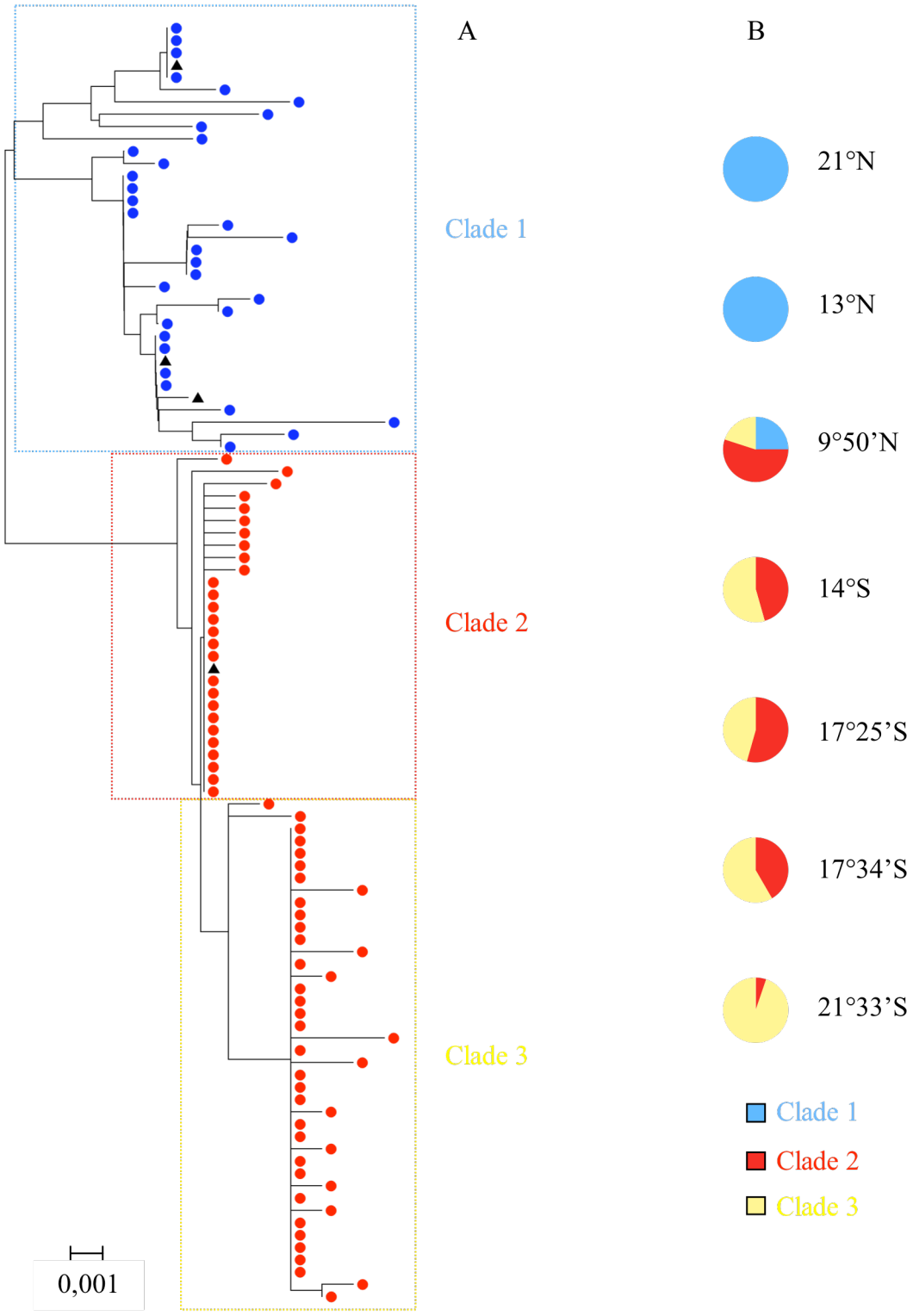


Fig. 4.16 Arbre phylogénétique de qualité 3 (A) et distribution géographique des clades (B) du gène Lysozyme.
Sur l'arbre phylogénétique :
En bleu, individus à mitochondrie de type 'espèce Nord'. En rouge, individus à mitochondrie de type 'espèce Sud'. En noirs, individus à deux types de mitochondries.
Sur la distribution des clades :
Chaque couleur (bleue, rouge, jaune) correspond aux séquences échantillonnées dans les clades (1, 2 et 3, respectivement) de l'arbre phylogénétique.

L'arbre phylogénétique des allèles du gène VDG3 (Fig. 4.17.A) révèle également 3 clades. Les divergences entre clades 1-2, 1-3 et 2-3 sont respectivement de 2,8%, 4% et 1,9%. Le clade 1 compte en majorité des séquences de la lignée mitochondriale Nord mais également quelques allèles appartenant à la lignée mitochondriale Sud ou aux individus ayant les deux types de mitochondrie. Au contraire, les clades 2 et 3 contiennent uniquement des séquences recapturées de la lignée mitochondriale Sud (mais une séquence de 13°N éliminée lors de la sélection aléatoire des séquences présentait une séquence de type clade 3 et une mitochondrie de type Nord). La distribution des clades dans les différentes populations (Fig. 4.17.B) révèlent que, si les 3 clades sont représentés en fréquence variable dans la majorité des populations, seules des séquences du clade 1 ont été échantillonnées à 21°N, 13°N et 17°25'S (hormis la séquence de 13°N échantillonnée dans le clade 3 mais ayant été aléatoirement éliminée dans le jeu de données de qualité 3).

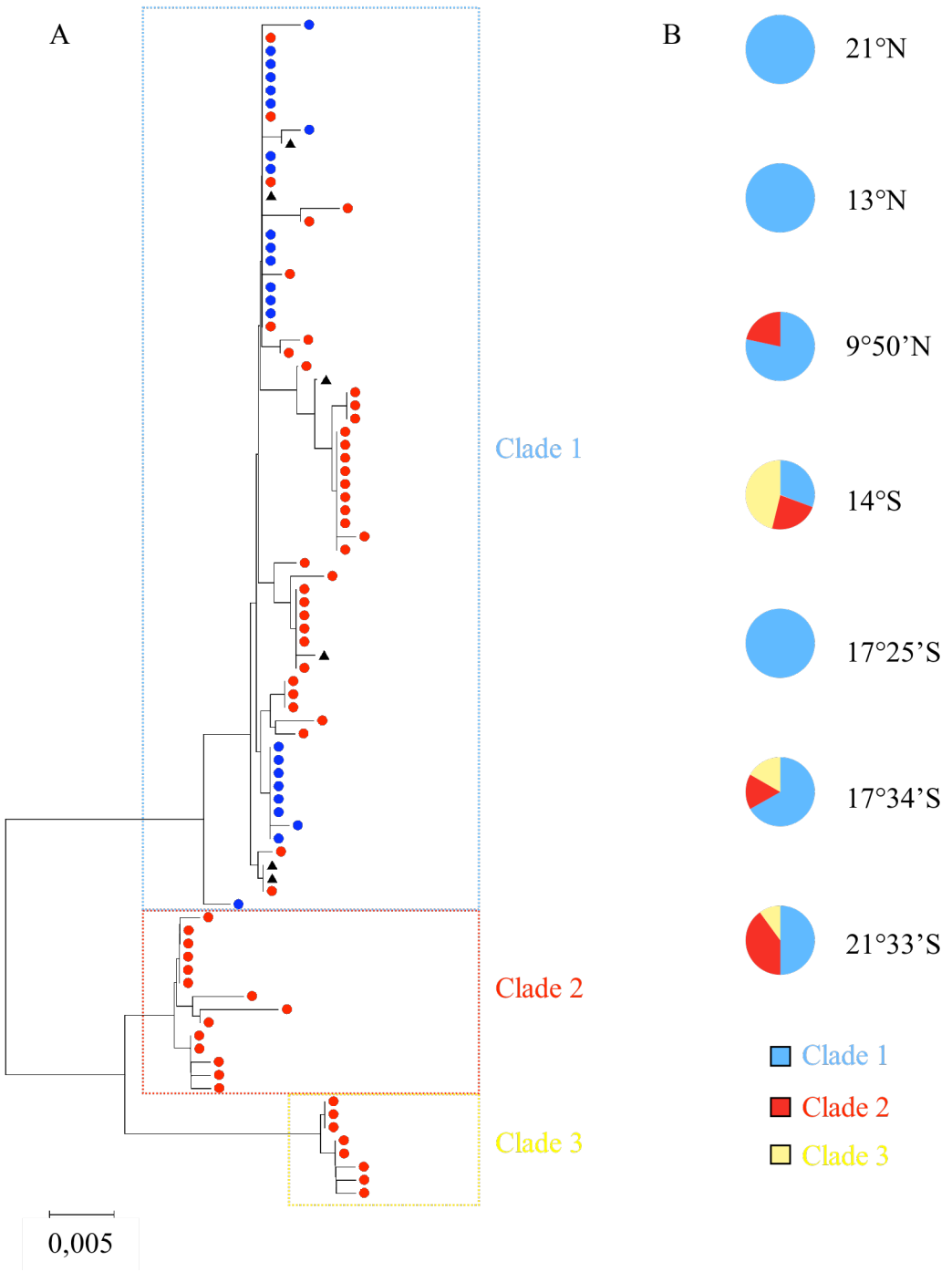
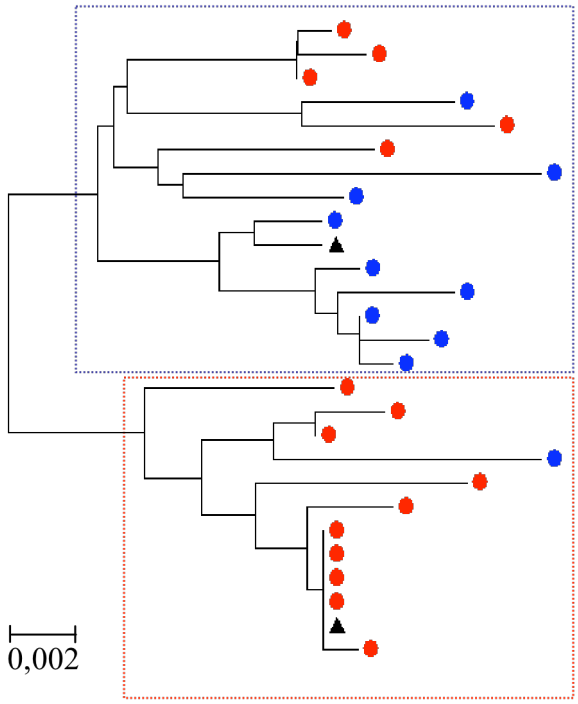


Fig. 4.17 Arbre phylogénétique (A) et distribution géographique des clades (B) du gène VDG3.

Sur l'arbre phylogénétique :
En bleu, individus à mitochondrie de type 'espèce Nord'. En rouge, individus à mitochondrie de type 'espèce Sud'. En noir, individus à deux types de mitochondries.

Sur la distribution des clades ;
Chaque couleur (bleue, rouge, jaune) correspond aux séquences échantillonnées dans les clades (1, 2 et 3, respectivement) de l'arbre phylogénétique.

L'arbre phylogénétique des allèles du gène Textilinine peut être divisé en deux clades qui divergent de 0,6% (Fig. 4.18). Le clade 1 contient l'ensemble des séquences des individus échantillonnés à 21°N et 13°N, ces individus correspondant à la lignée mitochondriale nord, plusieurs séquences d'individus de 9°50'N (appartenant à la lignée mitochondriale sud ou ayant les deux types de mitochondrie) et une séquence de 17°25'S appartenant à la lignée mitochondriale sud. Le clade 2 présente quant à lui uniquement des séquences appartenant à la lignée mitochondriale Sud échantillonnées à 9°50'N et 21°33'S et un allèle issu d'une lignée mitochondriale nord échantillonnée à 9°50N.



Localité	N_{introg}	N
13°N	0	9
9°50'N	5	10
14°S	0	1
17°25'S	1	3
21°33'S	0	4

Fig. 4.18 Arbre phylogénétique du gène Textilinine et nombre d'individus introgressés (tableau).
En bleu, séquences d'individus de mitochondrie de type espèce Nord.
En rouge, séquences d'individus de mitochondrie de type espèce Sud.
En noir, séquences d'individus présentant deux types de mitochondries (Nord et Sud).
N_{introg} : nombre d'individus introgressés, N : nombre total d'individus

Les valeurs de différenciation génétique (ϕ_{st}) ont été calculées pour chaque gène sur la totalité des populations puis entre paires de populations voisines pour les gènes Lysozyme et VDG3 via le logiciel Arlequin 3.1 (Excoffier et al 2005). Pour le gène Textilinine, le nombre de séquences recapturées pour le gène unique étant faible pour certaines populations, les ϕ_{st} ont été calculés en regroupant les populations suivant les barrières aux flux de gènes détectées sur le gène mtCOI (13°N-9°50'N et à l'Equateur), c'est-à-dire : (1) 13°N (les individus du site 21°N n'ayant pas pu être amplifiés), (2) 9°50'N et (3) 14°S à 21°33'S. La significativité de l'écart de chaque valeur de ϕ_{st} vis-à-vis de zéro est évaluée par un test de 1000-permutations par ce même logiciel.

Les ϕ_{st} calculés pour chaque gène sur l'ensemble des populations sont significatifs ($\phi_{st\ Lysozyme}$ = 0,518***, $\phi_{st\ VDG3}$ = 0,235***, $\phi_{st\ Textilinine}$ = 0,267***). Les ϕ_{st} entre chaque paire de populations voisines (Fig. 4.19) montrent, aux locus Lysozyme et VDG3 , 2 zones de différenciation génétique significative, l'une située entre 21°33'S et 17°34'S, et l'autre entre 9°50'N et 13°N. Le gène VDG3 présente également un ϕ_{st} significatif entre les populations de part et d'autre de l'Equateur (i.e. 9°50'N et 14°S), mais pas le gène Lysozyme. Pour le gène Textilinine, les ϕ_{st} calculés entre 13°N et 9°50'N (ϕ_{st} = 0,007) ainsi qu'entre 9°50'N et le Sud de l'EPR (ϕ_{st} = 0,014) ne sont pas significatifs. Cependant, le ϕ_{st} entre 13°N et le Sud de l'EPR est significatif (p = 0,046*).

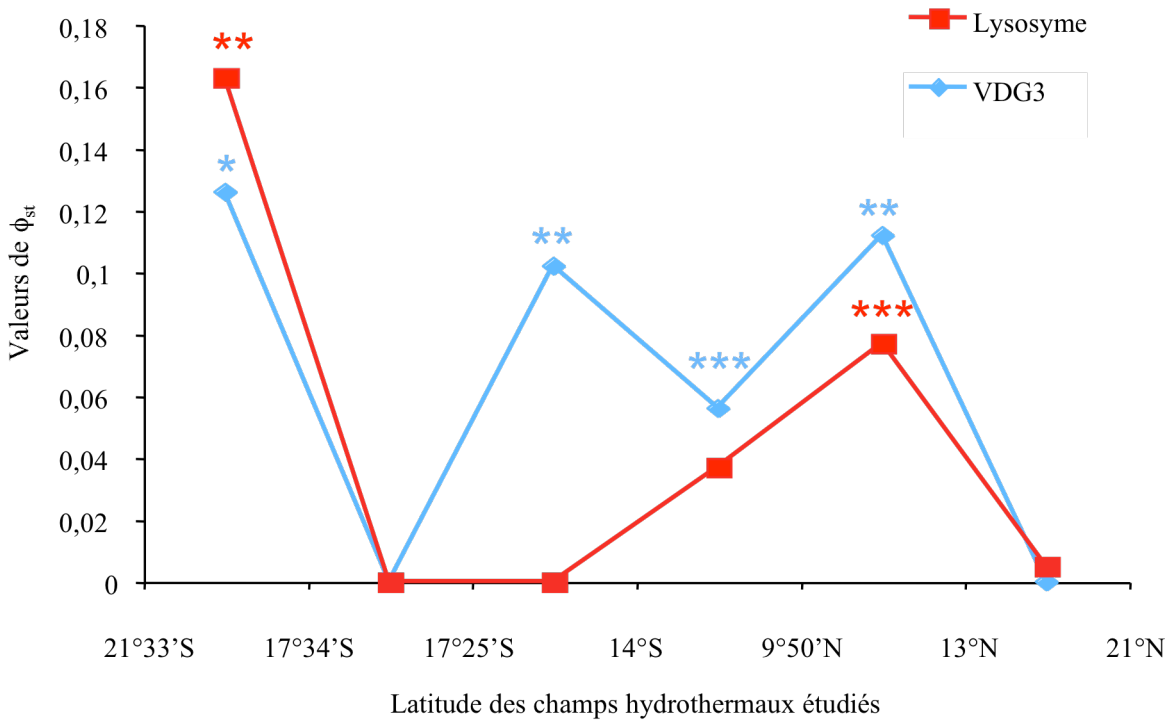


Fig. 4.19 Valeurs de ϕ_{st} entre paires de populations voisines sur les gènes Lysozyme (bleu) et VDG3 (rouge).
* p-value < 0,05 ; ** p < 0,01 ; *** p < 0,001

Les gènes nucléaires Lysozyme et Textiline présentant une forte ségrégation de leurs allèles selon les deux types mitochondriaux, une étude des mauvaises assignations d'un type mitochondrial à son 'background' nucléaire a été effectuée. Bien évidemment, il s'agit d'une estimation biaisée par la recapture des allèles, les individus n'ayant pas été vraiment génotypés aux locus nucléaires (allèles manquants). Un individu est alors considéré introgressé si au moins un de ses allèles nucléaires est échantillonné dans un clade ne correspondant pas à l'assignation de la lignée mitochondriale (e.g. mtCOI Sud et un allèle dans le clade 1 pour le gène Lysozyme et/ou Textilinine). Le gène VDG3 ne présentant pas une correspondance claire entre les clades et les lignées mitochondriales, il n'a pas été utilisé pour cette estimation. Le nombre d'individus détectés hybrides ou introgressés entre les deux espèces cryptiques est faible (7 individus soit 6%, Tableau 4.8). Ces individus proviennent en majorité (6/7) de la zone de contact 9°50'N (seul 1 individu détecté introgressé à 17°25'S).

Tableau 4.8 Nombre d'individus considérés introgressés, en comparant les clades mtCOI aux clades des gènes Lysozyme et Textilinine, selon leur type mitochondrial et leur latitude.

Nombre d'individus détectés introgressés/nombre d'individus considérés.
- : absence du type mitochondrial considéré (ou non échantillonné)

Latitude	mtCOI Nord	mtCOI Sud	mtCOI Nord et Sud
20°N	0/11	-	-
13°N	0/22	-	-
9°50'N	1/2	3/14	2/5
14°S	-	0/11	-
17°25'S	-	1/10	-
17°34'S	-	0/12	-
21°33'S	-	0/19	-

4.3.3.3. *Rôle des barrières aux flux de gènes dans la dynamique de l'introgression*

Le contact secondaire entre deux lignées divergentes géographiquement séparées peut, lorsque la barrière reproductive n'est pas totale, donner lieu à de l'hybridation entre ces lignées. Si la dispersion est suffisamment efficace et que les hybrides ne sont pas contre-sélectionnés, cette remise en contact secondaire tendra vers une homogénéisation génétique rapide de ces lignées. En revanche, un équilibre entre dispersion et sélection contre les hybrides peut maintenir une zone d'hybridation de largeur variable (Barton & Hewitt 1985). Ainsi, alors que la zone hybride entre les moules côtières *Mytilus edulis* et *M. galloprovincialis* se distribue en mosaïque sur des centaines de kilomètres en formant une succession de zones de transition entre le Nord et le Sud de l'Atlantique (Bierne et al 2003), la zone « hybride » de la sauge buissonnante *Artemisia tridentata* se limite à un front d'une quinzaine de mètres entre les aires géographiques des 2 lignées (Graham et al 1995). La détection d'individus hybrides entre les deux espèces cryptiques de *L. elevatus* définies par Matabos et al (2008b) à 9°50'N (via des marqueurs enzymatiques), associée à la distribution géographique de ces espèces (précisée par Johnson et al (2008) et Plouviez et al (2009) sur le mtCOI), suggérait que le champ hydrothermal 9°50'N pouvait constituer une zone de tension dont l'étendue, le mode de fonctionnement, et l'impact sur l'isolement génétique des espèces restaient à établir.

La distribution géographique (21°N-9°50'N et 9°50'N-21°S, respectivement) des allèles appartenant aux deux clades divergents sur chacun des gènes Lysozyme et Textilinine concorde avec la distribution des clades mtCOI des deux espèces et confirme la présence d'une zone de contact secondaire à 9°50'N. De plus, la majorité des individus possédant une signature 'hybride' avec un allèle 'nord' et un allèle 'sud' découlant de chacun des clades (hybrides ou introgressés) a été échantillonnée sur ce champ hydrothermal. Le gène VDG3 présente également des clades divergents. Cependant, la distribution géographique de ces clades reste difficile à établir et ne correspond pas aux patrons de distribution observés aux 3 autres gènes. En effet, si les clades 2 et 3 ne comptent que des séquences échantillonnées au sein des individus de l'espèce sud, le clade 1 correspond à un mélange équiprobable de séquences échantillonnées dans les 2 lignées

mitochondriales trouvées au Nord et au Sud de l'EPR. De ce fait, il est difficile de savoir si (1) le clade 1 correspond à la lignée Nord, une partie des allèles 'nord' ayant introgressé la lignée sud dans un passé relativement ancien (certains allèles 'introgressés' ayant eu le temps de se diversifier dans la lignée sud) ou (2) si cette distribution des allèles n'est pas uniquement le reflet d'un polymorphisme 'ancestral' non trié (i.e. séparation des lignées n'ayant pas permis de fixer suffisamment de mutations dans la divergence pour atteindre une monophylie réciproque sur ce gène). La distinction entre une introgression ancienne d'allèles et un polymorphisme ancestral n'est pas chose aisée, lorsque l'arbre phylogénétique présente une para- ou une polyphylie (Nosil et al 2009), d'autant plus que les gènes nucléaires nécessitent un nombre plus important de générations pour atteindre cette monophylie réciproque (Zink & Barrowclough 2008), en comparaison des locus mitochondriaux. En tout état de cause, il semble néanmoins que, si le mélange d'allèles 'sud' et 'nord' dans le clade 1 reflète une introgression, alors celle-ci doit être relativement ancienne, car certaines branches terminales échantillonnées dans la lignée sud coalescent relativement profondément dans le clade 'nord' de l'arbre phylogénétique, indiquant une accumulation de polymorphisme depuis l'événement d'introgression potentiel (Fig. 4.17.A). Dans ce cas, l'ancienneté de l'événement pourrait expliquer pour ce gène, l'importance de la zone géographique balayée par l'introgression au sud de l'EPR (allèles 'introgressés' retrouvés jusqu'à 21°33S), si tant est que l'on considère le gène VDG3 comme étant un marqueur neutre. Une telle hypothèse apparaît relativement réaliste même pour une espèce à faible capacité dispersive si l'on considère que l'événement de spéciation entre les lignées sud et nord date d'au moins 10 millions d'années (Plouviez et al 2009) et que la remise en contact entre les 2 lignées pourraient largement pré-dater la mise en place de la seconde barrière équatoriale, il y a 1,3 millions d'années. À l'inverse, le faible nombre d'individus ayant une signature 'hybride' sur les gènes Lysozyme et Textilinine et leur position géographique relativement restreinte (essentiellement à 9°50'N) pourraient être la conséquence d'une hybridation récente des 2 lignées (durée d'introgression trop courte pour être détectée au niveau des clades) ou la conséquence d'une trop forte contre-sélection des individus hybrides (barrière génétique efficace de part et d'autre de 9°50N) qui masquerait, à ces locus, l'hybridation ancienne de ces 2 lignées. Il convient cependant de noter que la plupart de ces individus ne sont pas hybrides de première génération (introgression), indiquant que la contre-sélection des hybrides n'est pas totale.

Si la sélection contre les hybrides peut avoir jouer un rôle important dans la limitation géographique de l'hybridation pour certains gènes, la barrière à la dispersion récemment mise en place entre 9°50'N et 14°S (1,3 Ma, Plouviez et al 2009), mise en évidence au locus mitochondrial et confirmée sur le gène Lysozyme, pourrait avoir récemment réduit le flux migratoire de l'espèce Sud entre 9°50'N et le Sud EPR, limitant par la même, le contact secondaire entre les espèces à la zone située à 9°50'N et donc, l'étendue de la zone d'introgression aux nouveaux allèles apparus depuis la mise en place de la seconde barrière (la zone d'introgression beaucoup plus étendue observée au gène VDG3 étant probablement le reflet d'une histoire plus ancienne que la mise en place de cette barrière). Ainsi, la population de la lignée sud échantillonnée sur le champ hydrothermal 9°50'N serait partiellement isolée de celles des autres champs hydrothermaux situés plus au Sud, pouvant ainsi expliquer l'étroitesse de la zone d'hybridation vers le Nord et le Sud pour les gènes Lysozyme et Textilinine malgré une histoire évolutive plus ancienne des 2 lignées.

La mise en place récente (1,3 Ma) de la barrière à l'équateur pourrait également être à l'origine de la présence des 2 clades divergents retrouvés uniquement au Sud de l'EPR (i.e. clades 2 et 3) pour les gènes Lysozyme et VDG3. En effet, les clades 2 des 2 gènes pourraient correspondre à l'isolement de la population 9°50'N au Nord de la barrière alors que les clades 3 correspondraient aux populations isolées au sud de la barrière de 14°S à 21°33'S. Cette hypothèse s'appuie sur le fait que la divergence entre les clades 1 et 2 est plus faible que celle trouvée entre les clades 1 et 3 pour les 2 gènes, et que le clade 3 est majoritairement retrouvé au sud de la barrière 9°50'N-14°S. Une remise en contact des populations de ces clades 2 et 3 auraient ensuite pu brouiller le signal d'isolement géographique entre ces populations (pas de monophylie réciproque) laissant supposer que cette seconde barrière n'est pas très étanche pour ce gastéropode hydrothermal.

4.3.3.4. *Isolement génétique des populations de 21°N et 21°33'S et sens de colonisation*

Outre l'isolement génétique des populations de part et d'autre des barrières génétique 13°N-9°50'N et physique 9°50'N-14°S observées, les populations situées en limite d'aire géographique de l'espèce (21°N et 21°33S) présentent un particularisme génétique sur au moins l'un des gènes échantillonnés : la population 21°N se distingue par une réduction de polymorphisme au locus mtCOI (différence significative avec la population 13°N, Plouviez et al 2009) et la population 21°33'S présente une série d'allèles relativement divergents en forte fréquence aux gènes nucléaires Lysozyme et Textilinine (différence significative avec 17°34'S). Outre un décalage géographique des isolats génétiques associés à la mise en place de la seconde barrière, ces observations pourraient également traduire des événements démographiques locaux.

Ne comptant qu'un unique haplotype mitochondrial, indiquant une réduction drastique de la diversité génétique de la population 21°N, la différenciation génétique entre 13°N et 21°N pourrait être due à un goulot d'étranglement de la population de 21°N suite à une extinction massive des émissions qui aurait pu avoir lieu au début des années 80 (Hessler et al 1985). Ce résultat suggère ainsi que la recolonisation des nouvelles zones d'émission ne s'effectuerait que très partiellement à partir des sites situées plus au sud sur l'EPR, très probablement en raison de la présence de la faille transformante Rivera qui est l'une des plus longues du système EPR (décalage de la dorsale sur plus de 500 kilomètres). Il convient néanmoins de noter que Shank & Halanych (2007) ont suggéré que des immigrations récentes de *Riftia pachyptila* en provenance de 9°50N étaient détectables dans les populations du bassin de Guaymas par utilisation de marqueurs AFLP. Cependant, aucune réduction de diversité n'a pu être notée sur les gènes Lysozyme et VDG3 au niveau de la population 21°N, comme on l'attendrait pour un événement démographique.

Aucune faille transformante majeure n'est présente entre 17°34'S et 21°33'S, indiquant que la différenciation entre ces populations ne serait pas liée à l'existence d'une barrière tectonique surnuméraire. La dynamique des courants profonds (de l'ouest vers l'est à ces latitudes et quasi-perpendiculaire à l'axe du graben, Reid 1997) pourrait être responsable de l'isolement de ces populations. Cependant, une telle différenciation entre la population de 21°33S et les autres populations de l'EPR sud n'a actuellement pas été mise

en évidence chez les autres espèces. Si une migration du Sud vers le Nord à travers la barrière équatoriale a été détectée sur la lignée Sud de l'espèce *L. elevatus* pour le gène COI (Plouviez et al 2009), l'arbre phylogénétique du gène Lysozyme suggère que la population 21°33'S pourrait être dérivée d'un effet de fondation à partir des populations situées plus au Nord (e.g. Animal Farm : 18°33S). En effet, le clade 3 de ce gène, majoritaire à 21°33'S, est dérivé du clade 2 échantillonné en forte fréquence sur les autres populations de la lignée sud (9°50'N à 17°34'S).

Avec la détection des deux barrières aux flux de gènes (13°N-9°50'N et 9°50'N-14°S) et la différenciation génétique des populations 21°N et 21°33'S, *L. elevatus* semble présenter une structure génétique plus marquée que les 2 autres espèces étudiées dans ce chapitre (une seule barrière détectée au sein des espèces *A. pompejana* et *B. thermophilus*). Cette structure génétique plus forte pourrait indiquer que l'espèce a des capacités de dispersion plus restreintes et/ou une sensibilité plus forte aux modifications de l'environnement local (extinction). Il est néanmoins important de considérer ici que l'histoire évolutive plus complexe de *L. elevatus* (2 barrières et 2 périodes d'isolement) peut également fournir une explication convaincante à cette structure génétique plus marquée. L'hypothèse d'une dispersion relativement limitée (type proche en proche) de l'espèce fournit cependant un élément de compréhension supplémentaire pour expliquer le nombre relativement important d'espèces décrites chez les Lepetodrilidae (famille pouvant franchir plus difficilement les barrières à la dispersion et ainsi diverger plus rapidement de part et d'autre d'une barrière par spéciation allopatrique). En effet, la diversification d'espèces en allopatrie partiellement due aux capacités de dispersion réduites des larves (non-planctotrophes), a été mise en évidence chez des gastéropodes marins du genre *Conus* (Cuhna et al 2005). La radiation des Lepetodrilidae pourrait ainsi être le reflet d'une dispersion larvaire limitée. Des études plus approfondies sur les capacités de dispersion de ces espèces pourraient ainsi compléter ces informations génétiques et permettre de mieux comprendre la radiation de cette famille. Néanmoins, il est fort probable qu'une part de cette radiation évolutive soit également issue de spéciations adaptatives (spécialisation écologique).

4.3.4. Y a-t-il eu une expansion démographique au Sud de l'EPR ?

4.3.4.1. *Indices de diversité et tests d'écart à la neutralité*

La diversité nucléotidique (π_n) et le thêta de Watterson (θ_W) ont été calculés via le logiciel DNAsp 4.10.3 (Rozas et al 2003) au sein de chaque clade pour chaque gène et sont présentés dans le tableau 4.9. Pour ces gènes, si la recombinaison *in vitro* détectée entre individus a pu être éliminée, et que le nombre de recombinants inter-allèles au sein du même individu ait été minimisé par suppression des allèles surnuméraires, la recombinaison naturelle non détectée peut néanmoins persister dans nos jeux de séquences et est souvent visible lorsque l'on regarde la longueur des branches terminales de certains allèles. Le 'four-gamete test' de Hudson & Kaplan (1985), visant à détecter le nombre minimum d'événements de recombinaison (R_m), a donc été effectué pour chacun des gènes. Les R_m calculées pour chaque gène sur la totalité des séquences montrent des valeurs non nulles, allant de 5 pour le gène VDG3 à 9 pour les deux autres gènes nucléaires, indiquant la persistance d'allèles recombinés dans nos jeux de données. Néanmoins, il n'était pas possible de supprimer ces séquences potentiellement recombinantes afin d'obtenir un R_m global nul car cela impliquait une suppression d'un nombre trop important de séquences. Les logiciels basés sur la coalescence faisant l'hypothèse de l'absence de recombinaison (e.g. Migrate-n, IM, Sweep_bott), ils n'ont donc pas été utilisés dans cette étude. Afin de détecter d'éventuels écarts à la neutralité, les indices de Tajima (*D*, 1989) et de Fu & Li (*F**, 1993) ont été réalisés pour chacun des clades, leur valeur moyenne n'étant théoriquement pas affectée par la recombinaison (seule la variance est impactée, Schierup & Hein 2000). Cependant, la recombinaison, en augmentant le nombre de mutations en fréquence intermédiaire, influence la distribution des différences par paire de séquences (courbes de mismatch, Fig. 4.20).

Tableau 4.9 Diversité nucléotidique (π_n), thêta de Watterson (θ_W), indices de Tajima (D) et de Fu & Li (F^*), et nombre minimum d'événement de recombinaison (R_m) au sein de chaque clade pour les gènes nucléaires.

N, nombre de séquences (correspondant également au nombre d'individus)

^{NS} p > 0,10 ; $^{+}$ p < 0,10 ; * p < 0,05 ; ** p < 0,01 ; *** p < 0,001

Locus	Clade	N	π_n	θ_W	D	F^*	R_m
Lysozyme	1	33	0,011	0,017	-1,387^{NS}	-2,263^{+}	8
	2	28	0,001	0,005	-2,445**	-4,263**	0
	3	40	0,002	0,008	-2,554***	-4,644**	0
VDG3	1	68	0,005	0,010	-1,675^{+}	-3,324**	1
	2	14	0,003	0,004	-1,573^{NS}	-2,135^{+}	1
	3	8	0,002	0,002	-1,064^{NS}	-1,336^{NS}	0
Textilinine	1	17	0,014	0,020	-1,305^{NS}	-1,747^{NS}	4
	2	13	0,008	0,012	-1,420^{NS}	-2,197^{+}	2

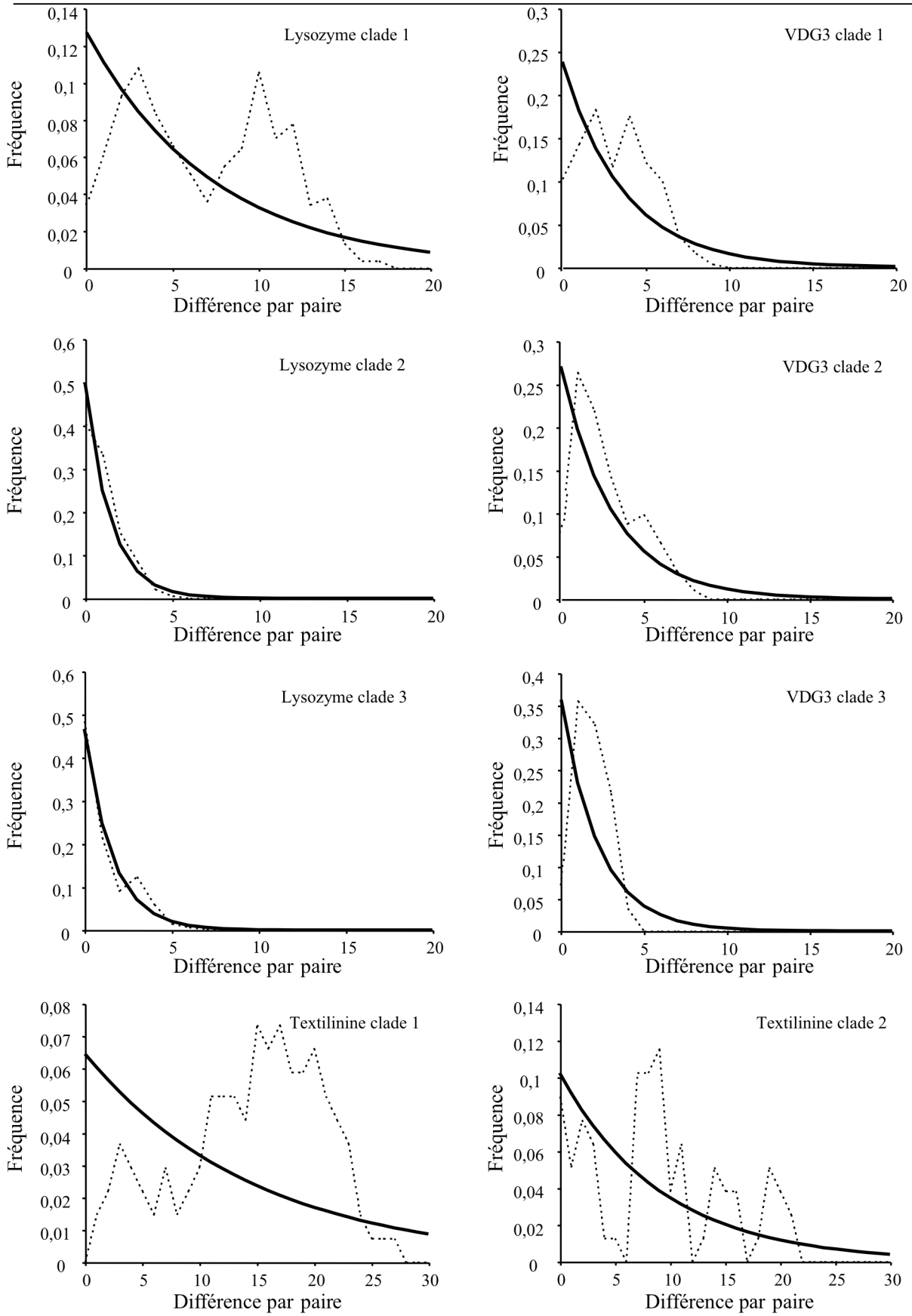


Fig. 4.20 Histogrammes de fréquence du nombre de différences par paire de séquences au sein de chaque clade pour les 3 gènes nucléaires.
Traits pointillés, courbes observées. Traits pleins, courbes attendues sous l'hypothèse d'une homogénéité du clade considéré et de la neutralité.

4.3.4.2. *Expansion démographique au Sud*

Matabos et al (2008b) ont suggéré l'existence d'un goulot d'étranglement récent dans la lignée sud de *L. elevatus* à 17°25'S en comparant, sur des marqueurs allozymiques et des marqueurs DALP, l'hétérozygotie attendue observée à celle attendue sous l'hypothèse mutation/dérive (software Bottleneck, Cornuet & Luitkart 1996). Une expansion démographique plus globale a été ensuite mise en évidence sur 7 espèces le long de l'EPR Sud, sur le gène mtCOI, et datée à moins de 500 000 ans (Plouviez et al 2009), les valeurs de D et F* obtenues chez la lignée sud de *L. elevatus* étant en accord avec cette hypothèse. Les gènes nucléaires utilisés sur les espèces *A. pompejana* et *B. thermophilus* (cf. parties 1 et 2 de ce chapitre) ont confirmé ce résultat. Chez *L. elevatus*, les valeurs significativement négatives du D de Tajima et du F* de Fu & Li recensées uniquement dans la lignée sud ainsi que les courbes de mésappariement 'mismatch', qui présentent un excès de variants rares (avec un *Rm* = 0) au Sud sur le gène Lysozyme, sont également en accord avec cette hypothèse. Pour le gène VDG3, la distribution géographique des clades ne permet pas de relier les valeurs de ces indices au Sud de la dorsale. Le gène Textilinine, qui discrimine bien le Nord et le Sud de l'EPR par 2 clades divergents, présente des indices de Tajima et Fu & Li qui ne sont pas significatifs au Nord et en limite de significativité au Sud (significatifs à 10%). Un échantillonnage plus conséquent de ces clades pourrait permettre de mieux évaluer la significativité de ces tests. Si la valeur moyenne de ces indices n'est théoriquement pas affectée par la recombinaison (Schierup & Hein 2000), les courbes de mésappariement révèlent quant à elles plusieurs pics de variants en fréquence intermédiaires, reflétant probablement un fort effet de la recombinaison pour le gène Textilinine mais beaucoup moins marqués pour les gènes VDG3 et Lyso pour lesquels un effet sélectif ou démographique est plus probable. En conséquence, même si les gènes VDG3 et Textilinine ne permettent pas d'inférer réellement un changement démographique au sud de l'EPR, les autres locus (mtCOI, Lysozyme et allozymes) s'accordent en faveur d'une expansion démographique au Sud de la dorsale, pouvant être due à de multiple extinctions/recolonisations dans cette région soumise à des vitesses d'accrétion élevées (125-152 mm/an, Cormier 1997).

5. *Discussion générale du chapitre 4 : comparaison des résultats obtenus entre les espèces cibles*

L'approche multi-locus utilisée dans ce chapitre sur trois espèces (*A. pompejana*, B. *thermophilus* et *L. elevatus*) appartenant à des classes distinctes d'organismes Lophotrochozoaires, permet de dégager un certain nombre de points communs quant à leur histoire phylogéographique et le rôle de barrières physiques à la dispersion le long de l'EPR. Quatre points communs majeurs peuvent ainsi être soulignés :

(1) La présence d'un patron commun de vicariance entre ces espèces, géographiquement ciblé entre le nord et le sud de la dorsale. Alors que l'approche multispécifique sur le seul gène mtCOI ne permettait pas de définir précisément la position géographique de la barrière aux flux de gènes responsable de cette vicariance, l'approche multi-locus a permis de préciser la localisation de cette barrière chez *B. thermophilus* et de confirmer la position préalablement établie sur le mtCOI pour les deux autres espèces. De plus, la datation multi-locus estimée chez *A. pompejana* et *B. thermophilus* confirme celle évaluée via le logiciel MsBayes par l'approche multispécifique à 1,3 Ma.

(2) Le caractère semi-perméable de la barrière aux flux de gènes même si celle ci apparaît relativement étanche pour les 3 espèces

(3) La détection d'individus introgressés

(4) La confirmation via l'approche multilocus de l'hypothèse d'expansion démographique au Sud de l'EPR proposée au préalable par l'approche multispécifique. Ainsi, l'effet d'un goulot d'étranglement suivi d'une forte expansion des populations observé sur l'EPR Sud semble relativement clair et lié à un facteur extrinsèque fortement structurant à l'échelle de la communauté et non de l'espèce (effet tectonique ou volcanique majeur). Pourtant, révéler un changement démographique passé à partir d'un déséquilibre mutation/dérive dans le polymorphisme d'une espèce n'est pas chose aisée et a donné lieu à de véritables débats entre partisans de l'existence d'un goulot d'étranglement (e.g. Przeworski et al 2001, Wall et al 2002), ou de multiple balayages sélectifs (e.g. Harr et al 2002, Glinka et al 2003, Baines et al 2002) dans les populations africaines et non-africaines de *Drosophila melanogaster*. En compilant les locus

de l'ensemble des études précédentes (soient 105 locus) et en y ajoutant 10 locus non codants liés au chromosome X, Haddrill et al (2005) ont alors conclu à l'existence d'un goulot d'étranglement (certains locus étant également soumis à diverses formes de sélection). Comparer plusieurs locus et plusieurs espèces co-distribuées peut donc s'avérer extrêmement efficace (et peut être même plus efficace que d'utiliser des dizaines de locus sur une seule espèce) pour détecter des évènements démographiques passés, si ceux-ci sont le fruit d'événements écologiques ou climatiques majeurs affectant la majorité des espèces d'une communauté (explosion d'un volcan, changements rapides du climat (glaciations), catastrophes de type tempête/ouragan ou chute de météores).

Mais outre ces points communs entre espèces, l'approche multi-locus a également permis de détecter certaines caractéristiques propres aux traits d'histoire de vie de chaque espèce et ainsi de révéler des dissimilitudes entre espèces notamment en termes de positionnement de la barrière et/ou de la taille des zones de contact. Ces variations, dues à l'histoire évolutive particulière de chaque taxon, notamment au niveau de leurs stratégies de reproduction et de dispersion, constituent un fort bruitage au signal de séparation des faunes que constitue la barrière. Ainsi, *L. elevatus* présente une structuration génétique plus marquée que les deux autres espèces (pour lesquelles une seule barrière est détectée), avec deux événements d'isolement, l'un génétique (création d'une zone de tension à 9°50N suite à une séparation très ancienne des faunes : 11,6 Ma) où le mécanisme principal est la contre-sélection des hybrides et l'autre physique et plus récente (création d'une barrière physique à la dispersion entre 9°50'N et 14°S, il y a 1,3 Ma). Ces barrières perméables aux flux de gènes contribuent à une différenciation génétique complexe à laquelle pourraient s'ajouter des effets de différenciation localisés au niveau des sites 21°N (extinction locale du site, il y a quelques dizaines d'années) et 21°33'S (point de rencontre avec une troisième espèce cryptique de *L. elevatus*). Cette structuration génétique plus forte pourrait s'expliquer par (1) des capacités de dispersion plus faible chez cette espèce et/ou (2) une plus forte sensibilité à l'extinction des émissions hydrothermales en étant plus spécifiquement associé au vestimentifère *Riftia pachyptila*.

Si *L. elevatus* présente, tout comme les deux autres espèces, une phase de dispersion essentiellement larvaire, les larves sont incubées dans la cavité palléale de la femelle jusqu'au stage veliger avant d'être émises dans la plume hydrothermale (Fretter et al 1988, Mullineaux et al 1995). Les larves de cette espèce sont lécitotrophes et n'ont pas été

décrites dans des couches d'eau éclairées à fortes concentrations en plancton (contrairement à *Bathymodiolus*, Arellano & Young 2009). La survie des larves dans une colonne d'eau oligotrophe telle que celle trouvée au niveau de la vallée axiale des dorsales océaniques implique nécessairement d'importantes réserves vitellines. Or, les œufs de *L. elevatus* présentent une taille inférieure à ceux d'*A. pompejana* (qui possède également une larve lécitotrophe ne migrant, *a priori*, pas dans la colonne d'eau : Tyler et al 2008). Du fait d'une incubation des œufs dans la cavité palléale des femelles, d'une part, et de la faible importance des réserves vitellines, d'autre part, on peut donc suggérer que la durée de dispersion dans la colonne d'eau des larves de *L. elevatus* doit être plus faible que celle supposée chez *B. thermophilus* ou *A. pompejana*.

L'hypothèse d'une plus forte sensibilité à l'extinction locale est plus difficile à appréhender car peu de données sont disponibles quant aux conditions de survie de cette espèce. L'espèce *L. elevatus* présente un preferendum thermique d'environ 4,5-11°C (Mills et al 2007) et est essentiellement associée aux *B. thermophilus* et *Riftia pachyptila*. Lorsqu'une source perd de son activité et se tarît, la température du milieu diminue jusqu'à revenir à la température de l'eau de fond (1-5°C), les vestimentifères associés au fluide meurent très rapidement (quelques dizaines de jours) alors que les moulières peuvent rester en place quelques années supplémentaires (évolution du site Parigo 84-90, Jollivet 1993) même si les modioles perdent rapidement leurs symbiontes, leurs réserves en glycogène et leur capacité de reproduction (Fiala-Médioni et al 1986, Raulfs et al 2004). Ces changements faunistiques peuvent ainsi modifier la distribution locale de *L. elevatus*. Pour atteindre les conditions 'idéales' de leur preferendum, les individus devraient donc se déplacer au plus proche du fluide. *L. elevatus* pourrait entrer en compétition avec d'autres espèces, et notamment *L. ovalis*, qui est associé plus spécifiquement aux moules hydrothermales, avec un preferendum thermique plus faible (environ 1,8-6°C, Mills et al 2007). Une telle compétition entre *L. elevatus* et *L. ovalis* sur la gamme de température préférentielle commune, concorde avec les observations de distribution de ces espèces. En effet, alors que *L. ovalis* a été échantillonné en grand nombre à 7°25'S et 18°33'S (sites en voie d'extinction avec populations nécrophages), l'espèce *L. elevatus* y est relativement rare (cf. article de Matabos et al en préparation, annexe du chapitre 5). Cette compétition sur les gammes de températures basses et le nombre limité de vestimentifères 'survivants' lors du tarissement de la source sur les gammes de températures hautes du preferendum, pourraient ainsi induire une sensibilité forte aux changements de température et une

extinction locale supérieure pour cette espèce. Ainsi, Bengtsson (1989) a montré que la compétition interspécifique entre trois espèces de zooplancton augmentait la probabilité d'extinction locale dans un système de métapopulation. En supposant qu'un tel changement de température n'engendre pas de compétition aussi forte avec d'autres espèces pour *A. pompejana* et *B. thermophilus*, *L. elevatus* pourrait présenter une plus forte sensibilité que ces espèces à des changements de température du fluide et ainsi présenter des extinctions plus fréquentes et une structure génétique plus marquée. Cette hypothèse reste néanmoins hautement spéculative et demande à être testée expérimentalement (e.g. exclusion compétitive sur différentes gammes de températures). De plus, d'autres facteurs, tels que la composition chimique du fluide ou la disponibilité en nourriture, pourraient également influencer la compétition interspécifique et potentiellement modifier le régime d'extinction locale (cf Matabos et al 2008a).

L'hypothèse d'une plus forte sensibilité à l'extinction est également à relativiser au vu des tailles efficaces de populations. En effet, une espèce qui présente une grande taille efficace de populations est théoriquement moins assujettie à la dérive génétique et aurait ainsi une plus faible probabilité d'extinction qu'une espèce ayant une taille efficace plus faible (Li & Graur 1991). L'espèce *L. elevatus* présentent généralement des tailles de populations bien supérieures à celles de *A. pompejana* ou même à celle de *B. thermophilus* (Jollivet, communication personnelle), pouvant laisser penser que *L. elevatus* présente une taille efficace supérieure aux deux autres espèces étudiées et donc, au contraire, devrait mieux résister aux extinctions locales. Néanmoins, une grande taille de population ne signifie pas forcément que la taille efficace l'est également. Ainsi, la taille efficace estimée sur *L. fucensis* (Juan de Fuca, dorsale du Pacifique Nord, Johnson et al 2006) à l'aide d'un modèle d'Isolement avec Migration (Hey & Nielsen 2004, 2007), est du même ordre de grandeur que celle estimée par ce même modèle dans ce chapitre sur *A. pompejana*, mais 10X plus faible que celle estimée sur *B. thermophilus*. Cette similitude de taille efficace entre *L. elevatus* et *A. pompejana* pourrait s'expliquer par le mode de reproduction de ces 2 espèces. Ces 2 espèces utilisent en effet l'appariement mâle/femelle avec transmission du sperme et fécondation interne des ovocytes pour se reproduire (Fretter 1988, Pradillon & Gaill 2003), ce qui nécessite une recherche de partenaire dans son proche entourage (reproduction entre individus proches).

Une deuxième différence observée entre les 3 espèces cibles réside dans la position géographique de la barrière la plus récente. La position géographique de la barrière ayant

engendré une vicariance concomitante entre les espèces il y a 1,3 Ma, varie en effet selon l'espèce considérée. Cette barrière est située à l'Equateur (entre 9°50'N et 7°25'S) pour *A. pompejana* mais elle se positionne entre 7°25'S et 14°S pour *B. thermophilus* (1 seul individu échantillonné à 7°25'S pour *L. elevatus*). De la même façon, la zone de contact entre lignées mitochondriales est plus ou moins étendue géographiquement en fonction de l'espèce. Ainsi, les lignées mtCOI d'*A. pompejana* sont bien séparées géographiquement. Au contraire, *B. thermophilus* présente un mélange de ces deux lignées divergentes au Nord de sa barrière et les espèces Nord et Sud de *L. elevatus* sont essentiellement en contact à 9°50'N. De plus, alors que *L. elevatus* et *A. pompejana* présente une zone d'hybridation relativement restreinte (avec des hybrides trouvés essentiellement à 9°50'N et 7°25'S-14°S, respectivement pour ces deux espèces) qui s'apparente plus à la notion de front d'hybridation, la zone d'hybridation de *B. thermophilus* s'étend à la fois au Nord et au Sud de la barrière présumée (au Nord : 2 lignées mitochondriales alors que certains gènes tels que Sulfo 1 et SAHH présentent un unique clade, au Sud : détection d'hybrides sur le gène SAHH de 7°25'S à 17°34'S) dans une logique plus proche de la zone d'hybridation en mosaïque décrite par Bierne et al (2003) chez le genre *Mytilus*. Ces différences de distribution géographique (des lignées mitochondriales et des hybrides) sont en accord avec les capacités de dispersion supposées plus importantes chez *B. thermophilus* que chez les autres espèces et le vecteur de dispersion (courants de fond canalisés par la vallée axiale versus courants de surface). En effet, par analogie aux larves de *Bathymodiolus childressi* (Arellano & Young 2009), les larves planctotrophes de *B. thermophilus* pourraient être capables d'atteindre des couches d'eau soumises à la lumière du soleil, leur permettant ainsi d'accéder à une plus grande quantité de nourriture et passer rapidement d'une larve pedivéligère de 90 μm à une post-larve dont le diamètre atteint 400 μm (Le Pennec 1988, Dixon et al 2007). Elles sont donc susceptibles de rester plus longtemps dans la colonne d'eau que les larves d'*A. pompejana* et *L. elevatus*, mais également de disperser en suivant des courants de surface (potentiellement plus forts que les courants profonds et de direction généralement différente, Thomson et al 1990, Cannon & Pashinski 1997), ces derniers étant peu ou pas du tout affectés par la mise en place de barrières tectoniques telles que le décalage de la dorsale par faille transformante.

Chapitre 5 : Conclusion, discussion générale et perspectives

La distribution mais également la variabilité de l'activité des sources hydrothermales profondes font de cet environnement un exemple clé du concept de métapopulation (Vrijenhoek 1997, Jollivet et al 1999), avec de nombreux événements d'extinctions de populations et une (re)colonisation rapide des sites actifs parfois distants de plusieurs centaines de kilomètres. À cela s'opposent des conditions environnementales 'extrêmes' (H_2S, CO_2, CH_4, métaux Fe, Mn, Cu Zn, et As) relativement stables dans l'espace et perdurant depuis la création des océans (Hannigton & Jonasson 1995), générant ainsi de fortes pressions de sélection homogénéisante. De plus, suivant le modèle de dispersion adopté, les espèces pourront présenter une homogénéisation (modèle migrant) ou de fortes différenciations génétiques entre les champs hydrothermaux géographiquement distants (modèle propagule). Outre la distance entre sites, des barrières à la dispersion, éventuellement couplées à des origines de colonisation différentes entre provinces géographiques (pré-compétences d'espèces phylogénétiquement indépendantes à coloniser des environnements réduits), ont également façonné la distribution de la faune hydrothermale. Certaines barrières majeures ayant profondément impacté la composition de la faune, plusieurs provinces biogéographiques ont alors été définies (Van Dover et al 2002, Tyler & Young 2003, Bachraty et al 2009). L'objectif de cette thèse visait à mieux comprendre l'histoire de la colonisation et le rôle des failles transformantes et des courants profonds dans la distribution actuelle des espèces présentes le long de la dorsale du Pacifique oriental. Pour cela, deux approches complémentaires de phylogéographie ont été réalisées : (1) une approche multi-spécifique menée sur 7 espèces en utilisant le gène mitochondrial Cytochrome Oxydase I et (2) une approche multi-locus effectuée sur 3 espèces cibles *A. pompejana*, *B. thermophilus* et *L. elevatus*. Ces approches ont ainsi permis de dégager des hypothèses fortes sur l'histoire évolutive de ces espèces, telles que (1) la présence d'un événement de vicariance commun entre espèces, daté de 1,3 Ma, (2) une expansion démographique récente (< 500 000 ans) de ces espèces au Sud de l'EPR et (3) la détection de zones de contact secondaire plus ou moins larges entre le Nord et le Sud de la dorsale chez les 3 espèces étudiées via l'approche multi-locus. L'approche multi-locus nous a également permis de détecter des locus soumis à divers types de sélection (i.e. balayage sélectif, sélection balancée, sélection purifiante).

1. *L'EPR : une ou deux provinces biogéographiques ?*

La dorsale du Pacifique oriental (EPR) fût la première dorsale considérée comme une province biogéographique, supposée séparée de la dorsale du Pacifique Nord suite à la subduction de la plaque Américaine sous la plaque Farallon (Tunnicliffe 1991), il y a environ 35 millions d'années. Cette considération en tant que province biogéographique à part entière a été étudiée sur la base de sa composition faunistique. Plusieurs études ont alors défini d'autres provinces biogéographiques sur cette même base (e.g. Tunnicliffe et al 1998, Van Dover et al 2002, Tyler et al 2003, Bachraty et al 2009). Si ces auteurs s'accordent généralement sur la définition de cinq provinces biogéographiques principales (EPR, Dorsale du Pacifique Nord, Bassins arrière-arc du Pacifique ouest, dorsale Médio Atlantique, Océan Indien), certains proposent de rassembler deux provinces en une seule (e.g. Bassins arrière-arc du Pacifique ouest et Océan Indien : Tyler et al 2003, Bachraty et al 2009) ou de séparer une province en deux distinctes (e.g. dorsale Médio Atlantique, Van Dover et al 2002). L'EPR a ainsi été récemment considérée comme pouvant correspondre à deux provinces biogéographiques différentes, définies de part et d'autre de l'équateur (Bachraty et al 2009). Si en milieu terrestre les barrières à la dispersion permettent de définir les provinces biogéographiques, elles correspondent généralement aux frontières entre continents. A l'inverse, les provinces biogéographiques marines sont plus difficiles à appréhender (Cox & Moore 2001), notamment du fait de l'interconnexion des océans, l'existence d'une dimension supplémentaire : la profondeur et de l'accessibilité moindre au milieu.

Dans un travail où j'ai pris une part importante d'activité, Matabos et al (en préparation, cf. résultats présentés en annexe de ce chapitre) ont étendu le jeu de données de Bachraty et al (2009) en y ajoutant la composition des faunes échantillonnées en 2004 durant la campagne océanographique Biospeedo. L'analyse en présence/absence des espèces le long de cette dorsale confirme ainsi l'existence d'une différence de composition de ces faunes entre l'EPR Nord et Sud et renforce l'hypothèse de deux provinces biogéographiques de part et d'autre de l'Equateur. Ces nouvelles analyses placent la zone de séparation des faunes en 2 points de rupture, l'équateur et la zone autour de 14°S. Ce deuxième clivage a été affiné par l'ajout des collections récoltées pendant la campagne BioSpeedo par comparaison des travaux de Bachraty et al (2009) qui considéraient cette

séparation à partir de 17°25'S. Les résultats obtenus en phylogéographie dans les chapitres 3 et 4 montrent que cette position géographique de la séparation des faunes correspond également au point de rupture phylogéographique le plus récent retrouvé chez la plupart des espèces étudiées. De plus, l'analyse en 'barcode' du marqueur mitochondrial COI de certaines espèces rares de gastéropodes révèle l'existence de lignées mitochondriales divergentes, géographiquement structurées autour du point 14°S qui pourraient correspondre à des espèces cryptiques, de part et d'autre de l'Equateur, renforçant également cette séparation en deux provinces biogéographiques. Enfin, les niveaux de divergence observés sur plusieurs espèces (5,5-9,5% : *L. pustulosus, L. tevnianus, P. laevis et P. operculata*) s'apparentent à celui trouvé entre les 2 espèces cryptiques de *L. elevatus*, et renforcent l'hypothèse d'un premier effet vicariant plus ancien, probablement daté à 11-12 millions d'années.

S'il peut sembler évident qu'une barrière biogéographique, engendrant des différences de composition faunistique, puisse engendrer une forte structuration génétique des espèces qui la traversent, une concordance parfaite entre les patrons phylogéographiques et la composition faunistique ne peut pas toujours être établie. En effet, la composition de la faune marine intertidale nord-américaine (Engle et al 1999) et la structure phylogéographique observée sur plusieurs espèces (e.g. Soltis et al 2006) semblent s'accorder sur la côte est américaine afin de définir une barrière biogéographique entre le golfe du Mexique et la côte Atlantique de la Floride, avec la détection d'une zone d'hybridation au niveau du Cap Canaveral pour un grand nombre d'espèces sœurs (e.g. crabes : Bert & Harrison 1988, gastéropodes : Liu et al 1991, ou bivalves : Bert & Arnold 1995). Au contraire, Burton (1998) a montré que, bien qu'engendrant un changement net des communautés entre les provinces biogéographiques de l'Oregon et de la Californie (Doyle 1985), Point Conception ne correspondrait pas à une barrière phylogéographique intraspécifique.

La concordance entre une barrière biogéographique en termes de composition faunistique et une barrière phylogéographique doit donc être considérée avec précaution. En effet, cette concordance géographique ne signifie pas forcément une concordance temporelle de l'événement de vicariance. Ainsi, Cunningham et Collins (1994) distinguent la « congruence » des patrons de distribution, qui nécessite une concordance à la fois géographique et temporelle des barrières, de la « pseudo-congruence » des patrons phylogéographiques, qui correspond à une concordance géographique de ces barrières mais sur une période longue de colonisation du milieu, cette colonisation ayant eu lieu à des

périodes de temps différentes, et les barrières pouvant être issues de processus historiques différents. Donoghue & Moore (2003) proposent ainsi une classification du niveau de concordance entre quatre groupes dépendant de la concordance géographique et temporelle des événements de vicariance (Fig. 5.1). Congruence et pseudo-congruence se définissent ainsi par une correspondance géographique entre barrière et séparation des faunes alors que congruence et incongruence correspondent à une concordance temporelle de l'isolement des taxons (la pseudo-incongruence étant donc une absence de correspondance visible à la fois géographique et temporelle de l'isolement observé).

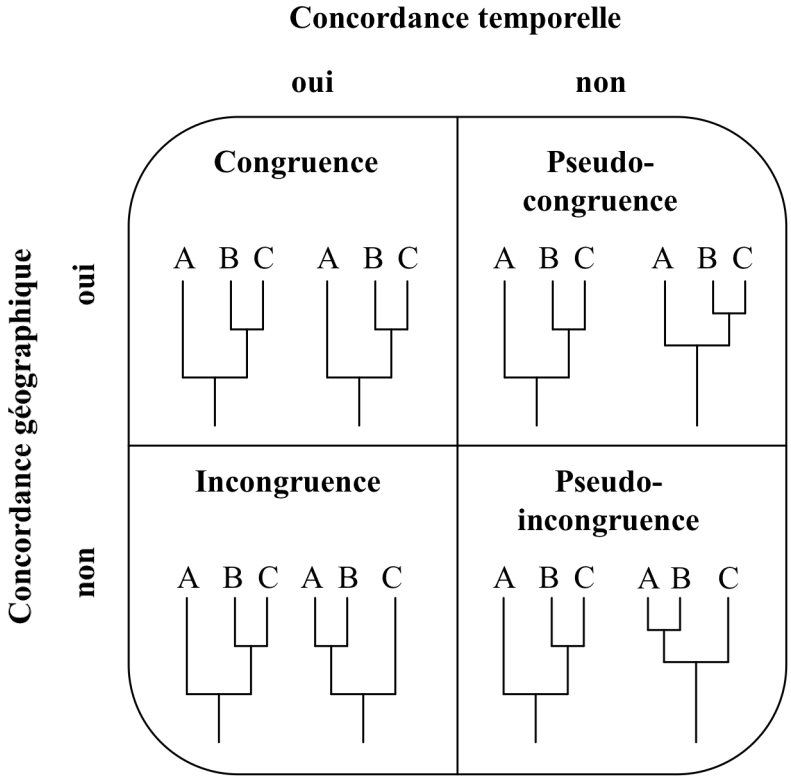


Fig. 5.1 Classification de la concordance géographique et/ou temporelle entre barrières biogéographiques.

Concernant l'EPR, nous avons mis en évidence un événement d'isolement génétique entre le Nord et le Sud de la dorsale à la fois géographiquement (si l'on admet que les différences observées entre espèces sont le reflet de capacités de dispersion différentes, de réponses différentes à l'extinction locale, et/ou de barrières génétiques plus ou moins fortes entre lignées (sélection contre les hybrides)) et temporellement concordants (1,3 Ma). Nous pouvons donc considérer que cet événement de vicariance récent a conduit à un

isolement allopatrique pour chaque espèce et reflète une barrière congruente entre ces espèces.

Au contraire, l'échelle de temps ayant conduit à la différenciation des faunes de part et d'autre de l'Equateur (concordance géographique avec la barrière intraspécifique) serait issue d'au moins un événement plus ancien de vicariance, l'estimation de la divergence entre espèces cryptiques allant de 6,5% pour *L. elevatus* (Matabos et al 2008b) à 12,5% pour le gastéropode Neomphalidae *Pachydermia laevis* (Matabos et al en préparation, cf. annexe de ce chapitre). En conséquence, l'isolement intraspécifique et celui détecté sur les communautés correspondent, *a priori*, géographiquement mais pas temporellement, et peuvent ainsi être considérées comme pseudo-congruentes. Néanmoins, l'isolement intraspécifique à 1,3 Ma et celui des communautés sont probablement issus d'évènements tectoniques distincts, le premier étant probablement lié à la présence de nombreuses failles transformantes récentes (i.e. complexe Quebrada / Discovery / Gofar, Kureth & Rea 1981, Naar & Hey 1989, Francheteau et al 1990) alors que le second pourrait être dû à la réorientation de la ride Mathématicienne (Mammerickx & Kligtgord 1982) ou la rotation de la microplaque Bauer (Eakins & Lonsdale 2003). Cette correspondance géographique apparente pourrait ainsi n'être qu'une convergence géographique. Néanmoins, l'EPR constituerait deux provinces biogéographiques différentes définies de part et d'autre de l'Equateur.

2. *Connexion entre EPR et les autres provinces biogéographiques*

Bachraty et al (2009), en effectuant un codage présence/absence des espèces hydrothermales échantillonnées sur 63 champs hydrothermaux (Desbruyères et al 2006), proposent que l'EPR Nord pourrait avoir constitué un rôle central dans la dispersion des espèces hydrothermales. Ces auteurs suggèrent ainsi qu'une proportion significative de la dispersion en provenance de cette province biogéographique vers l'ensemble des autres provinces pourrait être à l'origine de la diversité actuelle recensée sur les sources hydrothermales. Ils mettent également en évidence la présence d'une dispersion des sources hydrothermales du Sud Ouest Pacifique vers le Nord Ouest Pacifique. Néanmoins, l'EPR Nord est également la province biogéographique la plus étudiée et pourrait, de ce fait, n'être pas aussi central que cela en raison du biais d'échantillonnage. Ces résultats de dispersion sont donc à prendre avec précaution. Si les assemblages des communautés de l'EPR Sud sont relativement similaires à ceux de l'EPR Nord, l'observation d'espèces

typiques des bassins arrière arc du Sud Ouest Pacifique ou de la dorsale Médio Atlantique au niveau de l'EPR Sud (et notamment 21°33S i.e. *Eosipho*, *Shinkailepas*, *Chorocaris*), suggérait que l'EPR Sud puisse constituer une zone de transition biogéographique entre l'EPR Nord et ces deux autres provinces (Jollivet et al 2004). Au sein de l'EPR Sud, les champs hydrothermaux les plus au Sud (i.e. 17°25'S à 21°33'S) présentent des compositions faunistiques différentes des sites situés plus au Nord (i.e. 7°25'S et 14°S) et une diversité taxonomique supérieure (Matabos et al en préparation, annexe de ce chapitre), suggérant que les champs 17°25'S à 21°33'S pourraient effectivement constituer une zone de transition des faunes entre l'EPR Nord et d'autres provinces biogéographiques telles que le Sud Ouest Pacifique. Il semblerait donc que la dispersion des faunes de l'EPR Nord (si on accepte l'hypothèse d'un rôle central de l'EPR Nord dans la colonisation des sources par la faune moderne actuelle) pourrait également avoir transité via l'EPR Sud pour atteindre les autres provinces biogéographiques.

3. *Barrières semi-perméables : rôle des failles transformantes et des courants divergents ?*

Si ce travail de thèse nous a permis de mettre en évidence l'importance des failles transfomantes et des courants profonds en tant que barrière aux flux de gènes, l'approche multi-locus sur trois espèces cibles a montré le caractère semi-perméable de ces barrières, avec la détection d'individus hybrides ou introgressés. Cependant, il est difficile de savoir de quand date cette semi-perméabilité et donc de conclure quant à un isolement allopatrique en cours avec migration (modèle d'Isolement avec Migration, Nielsen & Wakeley 2001) ou une remise en contact secondaire de lignées divergentes (Nosil et al 2009). En effet, le modèle d'Isolement avec Migration suggère une migration entre lignées relativement anciennes en situation parapatrique. Dans notre cas, à la vue du mécanisme très progressif de séparation des segments d'une dorsale (formation d'un décalage lent et continu par un effet de coulissage donnant lieu à la faille transformante), il serait sans doute intéressant de considérer que le mode de spéciation prévalent correspond à une situation d'allopatrie, en admettant une cessation très progressive des flux de gènes semblable au modèle décrit par Hey et Wakeley au moins pendant les 2 à 3 premiers millions d'années de l'isolement. Chez *A. pompejana* et *B. thermophilus*, tout concorde ainsi à montrer que l'isolement est actuellement en cours. Ce scénario est également

soutenu par le non-échantillonnage de l'indel de 1 pb au Nord de l'EPR sur le gène Lysozyme pouvant signifier que cet indel récemment apparu au Sud est désormais bloqué par la barrière aux flux de gènes (mais l'absence de cet indel au Nord serait à vérifier par génotypage). Au contraire, la détection d'individus introgressés (entre lignées divergentes) détectés en comparant directement la lignée mitochondriale et le 'background' nucléaire des lignées ou via des génotypages RFLP sur les gènes PGM (*A. pompejana*) et SAHH (*B. thermophilus*), ainsi que la présence d'individus introgressés à double mitochondries chez *L. elevatus* suggèrent plutôt une remise en contact secondaire de ces lignées divergentes : une situation relativement logique chez *L. elevatus* ou le processus de spéciation est engagé depuis beaucoup plus longtemps et pour lequel une zone de tension est manifeste à 9°50N.

4. *Perspectives de la thèse*

Ce travail de thèse nous a permis de dégager différentes perspectives, tant d'un point de vu de l'évolution des espèces hydrothermales et de la compréhension du milieu hydrothermal, que de l'évolution des gènes étudiés.

4.1. Évolution des espèces hydrothermales

La plupart des logiciels utilisant les modèles de coalescence ne prenant pas en compte la recombinaison comme force de diversité allélique, les analyses effectuées durant cette thèse ont, pour la plupart, été réalisées en éliminant les séquences recombinantes des jeux de données. Or, certains logiciels récents basés sur des méthodes Bayésiennes permettent de prendre en compte la recombinaison si celle-ci n'est pas trop élevée (e.g. PopABC, MIMAR). Ils supposent néanmoins de fournir un certain nombre de paramètres initiaux (prior) pas toujours aisés à définir lorsque les caractéristiques biologiques des espèces sont peu connues (comme c'est le cas ici, e.g. estimation *a priori* de la taille efficace des populations). Néanmoins, ces logiciels n'en demeurent pas moins intéressants et peuvent permettre de complémenter nos analyses, notamment pour les espèces cryptiques de *L. elevatus* pour lesquelles la recombinaison entre allèles était trop importante pour être totalement supprimée des jeux de données. Pour ces espèces, certaines manipulations complémentaires basées sur le génotypage par RFLP pourraient également apporter des

informations précieuses en permettant de quantifier plus finement l'hybridation le long de l'EPR, à l'instar des gènes PGM d'*A. pompejana* et SAHH de *B. thermophilus*. Ainsi, un site de restriction pour l'enzyme Mse1 a été détecté sur le gène Lysozyme chez *L. elevatus* en liaison avec une mutation fixée discriminant les deux clades et pourrait ainsi permettre de mieux évaluer l'étendue de la zone d'hybridation entre ces espèces.

La mise en évidence d'une différenciation génétique entre populations variant selon la latitude et les espèces étudiées pose la question de la raison de ces disparités. Nous avons proposé dans les chapitres 3 et 4, des hypothèses qui pourraient expliquer ces différences. Néanmoins, les connaissances biologiques (e.g. cycle de vie, capacités de dispersion,…) mais également écologiques (e.g. compétition intra et interspécifique, niveau de fragmentation de l'habitat associé à l'espèce, adaptation à l'environnement local,…) de ces espèces restent toutefois limitées. Afin de mieux comprendre l'évolution des espèces, il semble nécessaire de mieux appréhender leur capacité actuelle à coloniser de nouveaux sites et leur sensibilité vis-à-vis des modifications biotiques (relations entre espèces sympatriques) et abiotiques (e.g. température, composition du fluide hydrothermal). Certaines espèces n'ont été que très peu étudiées et leur cycle de vie reste extrêmement peu connu, c'est le cas notamment du polychaete *Hesiolyra bergi* ou du gastéropode *Eulepetopsis vitrae*. Des observations biologiques directes pourraient ainsi permettre d'apporter des informations quant à son mode de reproduction, par exemple. Outre les observations directes, des approches indirectes peuvent permettre d'évaluer les capacités de dispersion de ces espèces, via l'étude de la connectivité entre populations. De telles études ont déjà été engagées au niveau de l'EPR nord en modélisant la circulation des masses d'eau au niveau de la dorsale à l'aide de mouillage de courantomètres, de piégeage de larves dans la colonne d'eau et la plume hydrothermale ou la mise en place de modules de colonisation sur des séries à long terme (cf. Mullineaux et al 1998). L'utilisation de marqueurs génétiques suffisamment polymorphes (e.g. microsatellites, AFLP) pourrait également être extrêmement informatif et permettre, par exemple, d'assigner des individus migrants à leur population d'origine, si une différenciation suffisamment forte entre populations est révélée. Une autre manière d'observer de manière indirecte la connectivité entre populations serait une étude de la géochimie des coquilles chez les bivalves ou les gastéropodes. En effet, le fluide hydrothermal peut présenter des compositions chimiques légèrement différentes entre champs hydrothermaux. Les coquilles des mollusques étant élaborées en utilisant les composés chimiques présents dans le milieu lors de leur

formation, la composition chimique des coquilles larvaires (encore présentes chez l'adulte) peut ainsi renseigner sur la provenance de chaque individu, si tant est que cette coquille larvaire se soit formée au contact du fluide hydrothermal et pas dans la colonne d'eau entre ces sites (un cas envisageable chez les gastéropodes Lepetodrilidae mais peu vraisemblable chez les Bathymodiolinae ou les larves sont directement émises dans la colonne d'eau en périphérie des points d'émission). Ce genre d'approche géochimique a déjà été réalisée avec succès sur des coquilles de bivalves côtiers (*Mytilus californianus*, *M. edulis*, Becker et al 2005, 2007), de gastéropodes intertidaux (*Concholepas concholepas* : Zacherl et al 2003, *Kelletia kelletii* : Zacherl et al 2005) ou sur des otholithes de poissons (Thorrold et al 2007). Les gastéropodes qui incubent leurs larves dans leur cavité palléale (larves formées sur le champ hydrothermal d'origine), tels que les *Lepetodrilus*, pourraient ainsi constituer de bons candidats à une approche géochimique de connectivité des populations.

Une meilleure compréhension du milieu profond lui-même, par une investigation des courants profonds par exemple, pourrait également permettre de mieux expliquer les patrons de distribution des espèces. De plus, la découverte récente d'un champ hydrothermal actif à 4°S sur la dorsale EPR avec une faune hydrothermale associée, pourrait permettre de préciser la position de la barrière géographique entre l'EPR Nord et l'EPR Sud.

4.2. Évolution des gènes

4.2.1. Le gène mtCOI

L'étude du gène mitochondrial de la Cytochrome Oxydase I a permis de mettre en évidence un code génétique particulier chez le polychète *A. pompejana* avec la méthionine qui serait codée par deux codons différents ATG et ATA (ce dernier codant habituellement pour l'isoleucine). Il serait extrêmement intéressant de poursuivre cette étude du code génétique mitochondrial des Alvinellidae en séquençant d'autres gènes mitochondriaux afin de vérifier si ce code génétique est partagée par d'autres espèces d'Alvinellidae, voire d'autres annélides. D'ores et déjà, une étude des génomes mitochondriaux de plusieurs annélides dont *Riftia pachyptila, Alvinella pompejana* et *Paralvinella grasslei* est possible. Si des séquences de ce gène ont été obtenues chez *A. caudata*, le nombre de séquences

effectuées restaient cependant trop faibles pour pouvoir avoir un accès réel au polymorphisme. La raison de ce faible nombre de séquences obtenues sur *A. caudata* est également très intéressante. En effet, sur 48 individus séquencés en direct, seule une dizaine de séquences présentaient des chromatogrammes interprétables. Or, une fois ces séquences alignées, nous avons détecté trois types de séquences différentes. Les deux premiers types étaient différents par des substitutions synonymes, suggérant la présence de deux lignées mitochondriales (voire espèces cryptiques). Le troisième type de séquences était beaucoup plus intrigant puisqu'il différait des deux autres par la présence d'un indel affectant le cadre de lecture de ce gène. Ainsi, la présence de cet indel sur le gène mtCOI laisse penser à une possible duplication du gène qui pourrait ou non être passé dans le génome nucléaire. Ce genre de transfert à déjà été décrit dans la littérature sous le nom de numts en analysant les génomes de certaines espèces modèles (e.g. coléoptère *Tribolium* ou l'abeille *Apis*, Pamilo et al 2007). Le nombre important de chromatogrammes illisibles sur le mtCOI de cette espèce pourrait être dû à la co-amplification de ces gènes dupliqués ou d'individus présentant une hétéroplasmie mitochondriale (cf. hybrides de *L. elevatus*). De plus amples études pourraient ainsi permettre de mieux comprendre l'évolution de ce gène chez *A. caudata*. Tout d'abord, le clonage des produits de PCR des individus présentant des chromatogrammes illisibles pourrait permettre de savoir si la co-amplification de plusieurs lignées paralogues est bien à l'origine de ces chromatogrammes illisibles. Ensuite, une extraction séparée de l'ADN mitochondrial et l'ADN génomique suivie d'une amplification par PCR sur ces deux types d'ADN pourrait permettre de savoir si une duplication de ce gène est passée dans le génome nucléaire ou si cette duplication potentielle s'est effectuée dans le génome mitochondrial.

La détection d'individus présentant une double hérédité parentale et étant hybrides entre les deux espèces cryptiques de *L. elevatus*, est également un point intéressant à creuser. Cependant, elle nécessite d'effectuer un échantillonnage adapté de *L. elevatus*, notamment à 9°50'N, d'en déterminer le sexe puis d'examiner la double hérédité en fonction du sexe des individus. Si un nombre suffisant d'individus à double hérédité parentale est détecté, il serait alors possible de regarder si ces individus présentent un sexe particulier (e.g. si seuls les mâles ont deux mitochondries comme chez les *Mytilus* : Zouros et al 1994). Examiner le caractère de double hérédité des mitochondries chez les larves ou même les œufs fécondés, pourrait également apporter une information importante, à savoir : si la mitochondrie du père (transmise par le sperme) est transmise à la descendance

puis perdue au cours du développement ou de la ségrégation méiotique. L'existence d'individus probablement F1 à double mitochondries laisse supposer en effet que les mécanismes pré-zygotiques n'empêchent pas la double hérédité parentale. Néanmoins, il va sans dire que ces questions dépendent de l'échantillonnage et présentent actuellement un véritable défi.

4.2.2. Phosphoglucomutase d'*A. pompejana*

La comparaison allozymes/RFLP sur le gène PGM d'*A. pompejana* a permis de détecter la mutation non synonyme probablement responsable de la différence de charge entre allèles 90 et 100 sur les individus du Sud de l'EPR. Néanmoins, cette mutation n'explique pas cette différence sur les individus du Nord de l'EPR, indiquant la présence d'une homoplasie de charge. Le séquençage de la totalité du gène de la PGM montre qu'un seul remplacement non-synonyme Q/E en fréquence intermédiaire explique le polymorphisme de la lignée nord, le site étant différent de celui trouvé dans la lignée sud (D. Jollivet, données non publiées). Il serait donc extrêmement intéressant de cibler une étude sur ce gène au voisinage de la (ou des) mutation(s) responsable(s) du changement de charge et dans les parties exoniques les plus éloignées de cette mutation, dans les 2 lignées d'*Alvinella pompejana*. Ce gène étant supposé sous sélection balancée, tout du moins au Nord de la dorsale, cette mutation devrait être accompagnée d'un excès de diversité nucléotidique à son voisinage dans l'hypothèse d'une sélection balancée (cf. Filatov & Charlesworth 1999). Une telle étude par fenêtre coulissante sur la diversité du gène n'a malheureusement pas été menée sur la lignée sud d'*Alvinella pompejana* car la mutation Q/E incriminée se trouvait à l'une des extrémités du fragment d'ADN amplifié.

5. *Annexe du chapitre 5*

Faunal changes and geographic crypticism indicate the occurrence of a biogeographic transition zone along the Southern East-Pacific Rise (SEPR)

Matabos M.[1,2*], Plouviez S.[2,3], Hourdez S.[2,3], Desbruyères D.[4], Legendre P.[5], Warén A.[6], Jollivet D.[2,3] and Thiébaut E.[2,3],

[1] ***Muséum National d'Histoire Naturelle, Département Milieux et Peuplements Aquatiques, UMR BOREA 7208 (MNHN, UPMC, CNRS), CP 53, 61 rue Buffon, F-75231 Paris cedex 05, France***

[2] ***CNRS, UMR 7144, BP 74, F-29682 Roscoff cedex, France***

[3] ***Université Pierre et Marie Curie-Paris 6, UMR 7144, Station Biologique de Roscoff, Laboratoire Adaptation et Diversité en Milieu Marin, BP74, F-29682 Roscoff cedex, France***

[4] ***Ifremer, Département Etude des Ecosystèmes Profonds, BP70, F-29280 Plouzané, France***

[5] ***Département de sciences biologiques, Université de Montréal, CP 6128, Succursale Centre-ville, Montréal, Québec, Canada H3C 3J7***

[6] ***Swedish Museum of Natural History, Box 5007, SE-10405, Stockholm, Sweden***

****corresponding author: mmatabos@uvic.ca***

Present address: School of Earth and Ocean Sciences, University of Victoria, PO Box 3065, STN CSC, Victoria, BC V8W 3V6, Canada

ABSTRACT.

Aim. Following their discovery in 1977 during the exploration of the Galapagos Spreading Center, deep-sea hyrothermal vents have been now reported along all active mid-ocean ridges and back-arc basins spreading centers. Although 5 to 7 distinct biogeographic provinces have been then described, delimitations between biogeographic entities remain questionable due to methodological issues. Here we examine biogeographic patterns of the hydrothermal vent fauna along the East Pacific Rise (EPR) and determine the relative role of regional and local factors on the distribution of biodiversity associated with mussel beds along a poorly explored zone, the southern EPR.

Location. East Pacific Rise

Methods. Species list along the EPR was compiled from the literature and completed with data recovered during the last French cruise BIOSPEEDO, held in 2004 along the southern Eastern Pacific. Biogeographic patterns were assessed by combining identification of morphological species and a molecular barcoding approach. The identification of a geographic break was performed using a multivariate regression tree. The spatial variation of the community structure in the mussel beds of SEPR in relation to environmental factors was analysed through the measure of different diversity indices, a cluster analysis and a redundancy canonical analysis.

Results. Sequencing of the Cytochrome Oxidase I gene revealed the occurrence of several cryptic species complexes along the EPR separating southern and northern clades on both part of the Equator. Furthermore, the Biospeedo cruise led to the discovery of at least 10 new species between 7°S and 21°S. The resulting shift in community structure highlighted by the MRT analysis was located at 14°S and 17°25S suggesting that the southern part of the SEPR (17°-21°S) constitutes a biogeographic transition zone in the vent fauna along the EPR. At regional scale, the latitude combined with the type of venting sampled had a significant impact on community structure associated with mussel beds along the SEPR. However those factors were not sufficient to explain variation in species composition suggesting a strong influence of other local factors such as small-scale habitat heterogeneity and/or biotic interactions.

Keywords East Pacific Rise, mussel beds, hydrothermal vents, species richness, biodiversity, biogeographic patterns, community structure, barcoding.

Running title. Biogeographic transition zone along the south EPR

INTRODUCTION

The discovery of dense communities of endemic species associated with hydrothermal vents on the Galapagos Rift in 1977 offered a new perception of deep-sea benthic fauna, far from what was commonly accepted in the vast poorly-productive abyssal plain (Lonsdale, 1977). Hydrothermal vents are discontinuously distributed in oases along oceanic ridges and characterized by an ephemeral lifespan vent fauna related to tectonic events and volcanic activities occurring along the ridge crest. The associated marine communities harbour high densities of macrobenthic fauna along with low species richness and low diversity resulting from the overwhelming numerical dominance of a few symbiotic species. They are clustered around vents where they thrive on microbial chemoautotrophic primary production (Van Dover, 2000). These oases are characterized by a low number of rare species which are heterogeneously distributed in time and space depending on the vent conditions and faunal succession occurring during the early colonisation of new vent emissions (Jollivet, 1996; Shank *et al.* 1998). The vent fauna has adapted to the great instability of the habitat where strong variations of the geological setting (in relation with the magmatic and/or tectonic activity) and the physical-chemical conditions (related to the temporal dynamic of the fluid discharge) naturally occur on a broad range of spatial (centimetres to thousands kilometres) and temporal (seconds to years) scales (Juniper & Tunnicliffe, 1997; Sarrazin *et al.*, 1997).

The endemism at the specific and generic levels has raised questions about vent species origin and evolution: it was suggested that they originated and then independently evolved from late Cretaceous and early Tertiary, around 65 Ma (Little & Vrijenhoek, 2003). But the colonization pathways and the links between the hydrothermal systems in the global ocean are still poorly understood despite the growing literature addressing biogeographic issues during the last decade (e.g. Tunnicliffe & Fowler, 1996; Tunnicliffe *et al*., 1998; Tyler *et al.*, 2003; Desbruyères *et al*. 2006a; Bachraty *et al.*, 2009). The present distribution of vent fauna mainly reflects historical events (i.e. plate tectonics) and species abilities to colonize distant territories (i.e. dispersion and migration) (review in Tunnicliffe *et al.*, 1998). While dispersal and migration are major processes linking neighbouring vent faunas (Tunnicliffe & Fowler, 1996), the interval between vents depends on the geodynamic nature of the ridge (i.e. volcanism or tectonism).

Based on faunal composition, the global ocean ridge system consists in a nearly continuous ca. 60000 km volcanic chain representing five main biogeographic provinces: the Northeast Pacific, the East Pacific Rise (EPR), the Northwestern Pacific back arc

basins, the Mid-Atlantic ridge (MAR), and the Southwest Pacific and Indian Ocean. However, depending on the authors, five to seven biogeographic entities can be erected from the literature often differing in delimitations within an ocean (Tunnicliffe, 1991; Tunnicliffe *et al.*, 1998; Van Dover *et al.*, 2002; Tyler *et al.*, 2003; Bachraty *et al.*, 2009). As an example, while the Mid-Atlantic ridge is considered as a single biogeographic province by most authors (e.g. Tunnicliffe *et al.*, 1998; Desbruyères *et al.,* 2006a; Tyler *et al.*, 2003; Bachraty *et al.*, 2009), Van Dover *et al.*, (2002) defined the Azores plateau and the MAR as two distinct provinces. Similarly, depending on the authors, the Western Pacific back-arc basins (BAB) comprise one or two (north *vs* south) biogeographic provinces (e.g. Tyler *et al.*, 2003; Bachraty *et al.*, 2009). The differences among authors depend on the methodology and criteria used to define the biogeographic units, as no clear comprehensive classification exists for the deep-sea (Spalding *et al.*, 2007), and the global database of species inventory is far from being complete and still growing with years.

The specialized nature of vents and their geographical isolation on active spreading centres are interesting features to study vicariance, species dispersal and colonisation pathways (Tunnicliffe, 1988). Two complementary approaches are commonly used in biogeography: historical and ecological biogeography. While ecological biogeography investigates the current mechanisms of communication and dispersion at the regional scale, historical biogeography studies tectonic/climatic events that have lead to the creation or dislocation of geographic barriers. When considering oceanic ridges at a global scale, tectonic events, creating physical barriers such as transform faults or microplates, can lead to the isolation of species in allopatry. Geographic isolation can yield pairs of vicariant species evolving independently, but filling the same ecological niche. This process has been suggested to be an important driving force of vent fauna evolution (Tunnicliffe & Fowler, 1996; Tunnicliffe *et al.*, 1998) and appears to be more and more accepted for species in most marine taxa (Heads, 2005). Ecological biogeography studies the factors that define the spatial distributions of species in the present time (Monge-Najera, 2008), thus processes acting at regional to local scales. At a regional oceanic scale, topographic features, geodynamic environment and hydrodynamism highly influence the extinction/recolonisation rate of the fauna, which, in turn, govern metapopulation processes by creating gaps in species distributions (e.g. Jollivet *et al.* 1999, Won *et al.* 2003). Those processes are likely to affect the regional pool of species. At a local scale (i.e. vent fields), community attributes can be highly variable within a site, along the fluid mixing gradient (Sarrazin *et al.,* 1997; Sarrazin & Juniper, 1999; Sarrazin *et al.*, 1999), and

between sites according to the successional stage or age of the site, which is related to the temporal dynamics of vents resulting from the waning and waxing of hydrothermal fluid discharge (Shank *et al.*, 1998; Sarrazin *et al.*, 1999; Mullineaux *et al*., 2000). Biotic factors, such as facilitation, predation, competition for food or space, by controlling recruitment and abundances, may play a significant role in structuring local benthic communities (Mullineaux *et al.*, 1998; Micheli *et al.*, 2002; Mullineaux *et al.*, 2003; Sancho *et al.*, 2005). Finally, spatial habitat heterogeneity, related to spatial variations in fluid fluxes, results in the occurrence of multiple microhabitats at small scale (Sarrazin *et al.*, 1997; Mills *et al.* 2007) and could promote sympatric speciation. This process, also called ecological speciation, has been suggested to fix divergence c.a. 20 times faster than species formation in allopatry but is rarely observed in natura (Via, 2001; Briggs, 2006). Thus, local factors do not only act at the community level but can also have broader significance. Ecology and biogeography are thus two complementary approaches in order to understand biodiversity patterns (Briggs, 2007), ecological processes acting at different spatial and temporal scales being often imbricate (Levin *et al.*, 2001). On the other hand, local processes influencing communities structure and biodiversity patterns can blur broader biogeographical patterns if a non-adapted sampling strategy is undertaken (e.g. lack of samples and/or replicates compared to the scope of the study).

Since 1977, taxonomists have described 533 morphological species comprised in 320 genera from vents (Desbruyères *et al.*, 2006b). However, all known vent sites did not undergo the same sampling effort, and biodiversity may be underestimated for poorly explored zones. Moreover, previous biogeographic analyses did not take advantage of the occurrence of sibling species, which is widespread in marine environment (Knowlton, 1993). Added to the identification of morphospecies, molecular systematics is a powerful tool to address biogeographic issues as it tells how species separate or expand in time and space. Comparative phylogeography (i.e. comparison of phylogeographic patterns among multiple co-distributed taxa) has strong parallel with historical biogeography and can help in identifying vicariant events unveiled by the taxonomy (Arbogast & Kenagy, 2001). To date, cryptic species have been reported from hydrothermal vents for many taxa separated by physical geological discontinuities that offset the ridge system (e.g. Won *et al.*, 2003; Hurtado *et al.*, 2004; Young *et al.* 2008). Along the EPR, the presence of cryptic species complexes was highlighted for lepetodrilid species (Johnson *et al.*, 2008; Matabos *et al.*, 2008b). A strong genetic divergence between a northern and a southern lineages was also observed from a set of taxonomic groups suggesting the occurrence of permeable genetic

boundaries around the Equator (e.g. Hurtado *et al.*, 2002; Won *et al*. 2003, Hurtado *et al.*, 2004; Matabos *et al*., 2008b). This hypothesis was recently validated by Plouviez *et al.* (2009) who demonstrated that the recent ridge history of the East Pacific Rise led to a vicariant event that induced genetic divergence across the Equator about 1.3 millions years ago for nearly all the vent taxa.

The Southern East-Pacific Rise (SEPR) is among the fastest spreading ridge centres in the world (150 $mm.an^{-1}$); it is characterized by a high instability in hydrothermal venting often associated with recent effusive volcanism (Auzende *et al.*, 1994; Fouquet *et al.*, 1994; Embley *et al.*, 1998). While this portion of the EPR has been extensively studied from the geological, tectonic and geochemical points of view (Auzende et *al.*, 1994; Fouquet *et al.*, 1994; Fornari & Embley, 1995; Fujioka *et al.*, 1995; Embley *et al.*, 1998), few papers are dealing with faunal composition across vent fields (Van Dover, 2002) and the only studies found in the literature were based on few samples and study sites. Thus first analyses of faunal assemblages along the SEPR described fauna similar to the northern part of the EPR (Geistdoerfer *et al.*, 1995; Halanych *et al.*, 1999). From quantitative samples, (Van Dover, 2002, 2003), proposed that the East Pacific Rise represents a single hydrothermal biogeographic province over at least 31 degrees of latitude, differences relying solely on the relative abundances of the dominant species and the distribution of rare species. This latter author suggested that fields are not far enough to prevent larval and/or adult dispersal in a stepping-stone manner (Van Dover, 2002). However, the author came to these conclusions on the analysis of a single particular habitat, mussel beds. Recently, based on presence/absence of species among all hydrothermal vent fields in the global ocean, Desbruyères *et al.* (2006a) and then Bachraty *et al.* (2009) argued that the Northern EPR (NEPR) and Southern EPR (SEPR) separate on both sides of the Equator as two distinct provinces. Despite a more or less continuous distribution of the vent fauna along the SEPR, assemblages consist in a succession of small patches dominated either by vestimentiferan tubeworms, bivalves, or stalked cirripeds, each of them resembling assemblages typical of 13°N/9°50'N/EPR for tubeworms, 21°N/EPR for bivalves, or the western Pacific back-arc basins for cirripeds (Jollivet *et al.*, 2005). These observations led to the hypothesis that this ridge portion may represent a transition zone between the NEPR fauna and those of the Atlantic or the Western Pacific back arc basins where species may have reached their range limits (Jollivet *et al.*, 2005).

The BIOSPEEDO cruise, held in the south-east Pacific in 2004 (Jollivet *et al.*, 2005), aimed at understanding biodiversity patterns, and processes governing them, along

the SEPR: a poorly explored but important ridge cross-road between biogeographic vent provinces. The objective of the present study was twice:

(1) to test the hypothesis that the SEPR could represent a transition zone between the NEPR and the other biogeographic provinces found in Atlantic and the Western Pacific. For this, macrofaunal community structure were compared at two spatial scales: (1) along the whole East Pacific Rise from 21°N to 17°S using species lists and then (2) along the 7°25'S-21°33'S/SEPR portion using a semi-quantitative dataset from 10 mussel beds at 6 hydrothermal vent fields in order to determine the relative roles of local and regional factors on mussel bed diversities.

(2) to generalize the hypothesis of vicariance across the Equator over nearly all vent habitats. This was assessed by a bar-code analysis of the main gastropod species encountered along the EPR. Gastropods are indeed an important component of hydrothermal fauna: species display various life history strategies and are found in all kinds of habitats (Jollivet, 1996; Mills *et al.*, 2007).

METHODS

Taxa, regions and data sources

Global analysis of communities along the EPR

A global species list of benthic macrofauna including only the main fauna (i.e. gastropods, polychaetes and echinoderms) was generated from data compiled in (Desbruyères *et al.*, 2006b), and completed with the species list of the BIOSPEEDO cruise. Highly mobile epifauna (e.g. crustaceans) have been intentionally excluded as they were not properly sampled during most hydrothermal cruises. A final list of species coming from chimney walls, vestimentiferan tubeworm clusters and mussel beds was thus compiled for nine hydrothermal vent fields: 21°N, 13°N, 9°50N, 7°25S, 14°S, 17°25S, 18°36S and 21°33S. In order to detect more precisely biogeographic relationships, the data was analysed at the species presence-absence level (Desbruyères *et al.*, 2006a, Bachraty et al. 2009).

Variation of the macrofaunal community structure associated with mussel beds along the SEPR

A total of nine semi-quantitative samples of *Bathymodiolus thermophilus* beds were collected in April/May 2004 along the Southern East-Pacific Rise (SEPR) at six vent fields from both volcanic-dominated and tectonic-dominated zones between 7°25'S and 21°33'S

(Table 5.1). Areas driven by volcanic activity consist in dome-shaped ridge crest where hydrothermal conduits resulting from seawater warming are short, shallower and unstable. Conversely, tectonic areas are characterized by a deep axial graben where hydrothermal fluxes are controlled by faults and are more stable through time (Fouquet *et al*., 1994). Distances between neighbouring sites were comprised between 3 km and 891 km, and between 30 m and 147 m within a site (i.e. Grommit). Collections were performed using the hydraulic arm of the French manned-submersible vehicle 'Nautile' (Ifremer). For a brief description of all vent fields see Jollivet *et al.* (2005).

Table 5.1 Characteristics of the sampling units. Macrofauna density corresponds to number of individuals standardize to the estimated surface sampled.

Vent field	Vent sites	Latitude	Longitude	Dive	Nb of mussels	Size range (mm)	Proportion of juvéniles (%)	Density of individuals (ind.m²)	Sampled area (cm²)	Geodynamic environment	Type of venting
7°25'S	Sarah's Spring (SS)	7°25.2445'S	107°47.6386'W	1574	46	0-80	9	3959	687	Volcanic	Moribund
14°S	Lucky Eric (LE)	13°59.2093'S	112°28.9559'W	1592	84	0-170	39	1170	3,581	Tectonic	Diffusion
17°25'S	Oasis BS6 (OBS6)	17°25.3843'S	113°12.2893'W	1579	147	0-170	3	2028	8,438	Volcanic	Diffusion
	Rehu Marka (RM)	17°24.9787'S	113°12.1369'W	1590	91	0-160	2	771	11,081	Volcanic	Diffusion
17°34'S	Susie (Susie)	17°35.4718'S	113°14.8252'W	1588	607	0-170	69	1794	9,888	Tectonic	Chimney
18°36'S	Animal Farm (AF)	18°36.6966'S	113°24.0482'W	1585	379	0-130	56	2249	8,735	Volcanic	Moribund
21°33'S	Grommit (Gr1)	21°33.6695'S	114°17.9049'W	1575	238	10-120	0	647	15,326	Tectonic	Chimney
	Grommit (Gr2)	21°33.6648'S	114°17.9814'W	1577	76	20-120	0	832	4,823	Tectonic	Chimney
	Grommit (Gr3)	21°33.6848'S	114°17.9808'W	1578	430	0-120	9	845	11,353	Tectonic	Chimney

Mussels were washed on deck with filtered seawater and the associated fauna was collected onto 1 mm mesh sieve. The animals were sorted onboard the ship; the retained organisms were fixed in 10% neutral formalin or in absolute ethanol for subsequent molecular analysis. In the laboratory, all specimens were sorted, identified to the lowest possible taxon, i.e. mainly to the species level, under a dissecting microscope and then transferred to 70° ethanol. Data on echinoderm species correspond to those reported in Stöhr & Segonzac (2006). Large and highly motile predators, including zoarcid fishes, squat lobsters (*Munidopsis* spp) and bythograeid crabs were not included in the analysis. Their presence was assessed by screening dives' video records. Meiofauna and vagile epifauna (mainly amphipods and copepods), which were not adequately sampled, were also excluded from the analyses. Mussels < 10 mm, usually deemed to represent an associated fauna at this life history stage rather than a structural component (Van Dover, 2002; Dreyer *et al.,* 2005), were also not included in this study. The proportion of mussel specimens < 10 mm, considered as juveniles, was however regarded here as an indication of the mussel beds' condition (newly settled, mature or senescent mussel beds).

Surface area (SA) of mussel shells was calculated by applying the relationship set between the mass of aluminium foil required to cover the outer surface of a mussel shell and mussel length (Bergquist *et al.*, 2005). One hundred and six mussel shells, representative of the whole size range, were used to determine this relationship. The outer shells of each collection were wrapped in aluminium foil (0.00363 $g.cm^{-2}$), and the mass required to cover the entire shell was then converted to surface area. Bias emerging from folding of the aluminium foil was considered equivalent for each mussel such that comparison between sampling units was valid. The relationship between SA (in cm^2) and shell mussel length (L) was derived by plotting the surface area against the shell length to obtain the following equation:

$$SA = 0.0137\ L^{1.9671}$$

This highly significant relationship ($R^2 = 0.996$; $p = <10^{-3}$) was then applied to each mussel collection. Standardized abundance datasets correspond to the number of individuals divided by the area sampled and is expressed as number of $individuals.m^{-2}$.

Bar-coding of hydrothermal gastropods

The total genomic DNA was extracted using the CTAB (cetyltrimethyl ammonium bromide) procedure from ethanol-preserved specimens. Gastropods were removed from

the shell, homogenized in 600µL of a 60°C preheated 2% CTAB buffer solution (containing 1.4M NaCl, 0.2% 2-mercaptoethanol, 20 mM EDTA, 100 mM Tris-HCl pH 8, 0.1 mg/mL proteinase K) and digested for 2 h at 60°C. Proteins, lipids and carbohydrates were removed using a single chloroform-isoamyl alcohol (24:1) precipitation. DNA was precipitated with 370µL cold 100% isopropanol and stored at −20°C for 1 to 2 h. The tubes were finally centrifuged at 15 000*g* for 5 min. DNA pellets were washed with 70–75% ethanol and resuspended in 20–60µL of TE buffer (10 mM Tris-HCl; 20 mM EDTA pH 8).

For each individual, partial sequences of mitochondrial cytochrome oxidase subunit I gene (mtCOI) were amplified by conducting a "nested PCR". DNA was pre-amplified with the universal primers M1 and M2 (Nelson & Fisher 2000) and used as a target (1 µl of PCR-product) for a second amplification with the universal primers LCO1490 and HCO2198 described by Folmer *et al.* (1994). For both amplifications, the 50 µl amplification mixture contained 3 µl of template DNA, 1x PCR buffer, 2 mM MgCl2, 0.4 µM of each primer, 2.5 µM of each dNTP, 2 U Taq DNA polymerase, and sterile H2O to final volume. Polymerase chain reactions (PCR) were performed as follows: (a) a 3 min initial denaturation step at 94°C, (b) 40 cycles of 45 s of denaturation at 94°C, 45 s of annealing at 50°C and 90 s of elongation at 72°C and, (c) a 7 min final elongation at 72°C.

The PCR products were purified and sequenced on ABI 3100 using BigDye® terminator chemistry (Applied Biosystems) following the manufacturer's protocol. Sequences were proofread and aligned manually using BioEdit Sequence Alignment 7.0.1.

Phylogenetic relationships between specimens were assessed for the wide-range species found at a high number in the sample: the Lepetodrilids, i.e. *Lepetodrilus elevatus*, *L. cristatus*, *L. pustulosus*, *L. tevnianus*, *Gorgoleptis emarginatus* and *G. spiralis*, the Neomphalids, i.e. *Pachydermia laevis* and *Peltospira operculata* and the Neolepetopsid *Eulepetopsis vitrea*. Additional species, found in low proportion during the Biospeedo cruise but well known from the East Pacific Rise were also added, i.e. *Nodopelta heminoda* and *Rhyncopelta concentrica* (Neomphalids), *Clypeosectus delectus* and *Gorgoleptis spiralis* (Lepetodrilids). Details of specimens analysed are presented in Table 5.2. Aligned sequences were analysed by Bayesian inferences. First, the optimal substitution model for COI was selected using the hierarchical likelihood ratio test (hLRTs) implemented in Modeltest 3.07 (Posada and Crandall 1998). Then, bayesian inferences were performed using the MrBayes 3.1.1 program (Huelsenbeck & Ronquist 2003) that uses a Markov Chain Monte Carlo (MCMC) method for exploring the parameter space in a stepwise

fashion. The analysis, involving 6 chains, was run for 10,000,000 generations with a sampling frequency of 1,000 and a burn_in period of 50,000. The trees were unrooted.

Table 5.2 Number of specimens per species and per locality used in the bar-coding analysis

Species	This study	Genebank Accession nos
Eulepetopsis vitrea	21°N (1), 13°N (1), 9°50'N (1), 7°25'S (1), 14°S (1), 17°25'S (1), 17°34' (1), 18°36'(1), 21°33'S (1)	
Lepetodrilus elevatus	21°N (1), 13°N (2), 14°S (1), 17°25'S (1), 17°35'S (1), 21°33'S (4)	EU306401-402 (Wood : 22°N/EPR :21°N-0°)
Lepetodrilus aff. *elevatus*		EU306407-408 (EPR:7°S-38°S)
Lepetodrilus aff. *galriftensis*		EU306413-414 (EPR:9°N-23°S)
Lepetodrilus ovalis	7°25'S (1), 14°S (1), 17°25'S (1), 17°35'S (1), 18°34'S (1), 21°33'S (1)	EU306478-479 (EPR:Monterey Bay/EPR:21°N-38°S)
Lepetodrilus pustulosus	13°N (3)	EU306457-458 (Wood:22°N/EPR:21°N-17°S)
Lepetodrilus aff. *pustulosus*		EU306464-465 (EPR:31°S-38°S)
Lepetodrilus tevnianus		EU306389-390 (EPR:9°N-23°S)
Lepetodrilus aff. *tevnianus*		EU306395-396 (EPR:23°S-31°S)
Lepetodrilus cristatus	13°N (17)	EU306425-426 (EPR:21°N-38°S)
Gorgoleptis emarginatus	7°25'S (2), 14°S (1), 17°25'S (2), 21° 33'S (1)	
Gorgoleptis spiralis	14°S (1), 17°25'S (2), 21° 33'S (3)	
Clypeosectus delectus	17°25'S (2), 21°S (1), 23°S (2)	
Pachydermia laevis	21°N (3), GoC (1), 13°N (1), 17°25'S (2), 17°34'S (1), 18°36'S (2), 21°33'S (2), 31°S (1)	
Peltospira operculata	13°N (4), 14°S (2), 17°25'S (1), 17°34'S (1), 18°25'S (2), 21°33'S (2)	
Peltospira delicata		AY923931
Nodopelta subnoda	13°N (2)	
Rhyncopelta concentrica	13°N (2)	

Published mtCOI sequences from *Lepetodrilus elevatus* (Genbank Accession nos EU306401-EU306402 and EU306407-EI306408 described as *L.* aff. *elevatus* (Johnson *et al.*, 2008)), *L. pustulosus* (Genbank Accession nos EU306457-EU306458 and EU306464-EU306465 for *L.* aff. *pustulosus* (Johnson *et al.*, 2008)), *L. cristatus* (Genbank Accession nos EU306425-EU306426), *L. ovalis* (Genbank Accession nos EU306478-EU306479), *L.* aff. *galriftensis* (Genbank Accession nos EU306413-EU306414) and *L. tevnianus* (Genbank Accession nos EU306389-EU306390 and EU306395-EU306396 for *L.* aff. *tevnianus* (Johnson *et al.*, 2008)) were included. A published sequence of *Peltospira delicata* (Genbank accession number AY923931) was also included.

Biodiversity measurements

For both global and regional analyses, the species richness (S), the average taxonomic distinctness (Δ^+), variation in taxonomic distinctness (Λ^+), and the Jaccard coefficient were calculated from species presence/absence matrice using the DIVERSE subroutine in PRIMER v5 (Clarke & Warwick, 2001).

The average taxonomic distinctness (Δ^+) is the average taxonomic distance between all pairs of species (i and j) in a sampling unit by gracing these distances through a taxonomic tree such as (Warwick & Clarke, 2001):

$$\Delta^+ = \left[\Sigma\Sigma_{i<j}\omega_{ij}\right] / \left[s(s-1)/2\right],$$

where ω_{ij} is the 'distinctness weight' given to the path length linking species i and j in the hierarchical classification, S the total number of species in the sampling unit, and the double summation are over all pairs of species.

The variation in taxonomic distinctness (Λ^+) is a measure of the assemblage complexity, which highlights differences in the structure of the hierarchical taxonomic relationships in the sampling unit (Warwick & Clarke, 2001). In other words, it can be assimilated to a measure of 'evenness' of the distribution of taxa across the hierarchical taxonomic tree (Warwick & Clarke, 2001). Formally, Λ^+ is equal to:

$$\Lambda^+ = \left[\Sigma\Sigma_{i<j}(\omega_{ij} - \bar{\omega})^2\right] / \left[s(s-1)/2\right]$$

where $\bar{\omega}$ corresponds to the average distance between each pair of species and is defined as:

$$\bar{\omega} = \left[\Sigma\Sigma_{i<j}\omega_{ij}\right] / \left[s(s-1)/2\right] \equiv \Delta^+,$$

For these two indices, equal step-lengths between each successive taxonomic level were assumed by setting the path length ω to 100 for two species connected at the highest taxonomic level. Five taxonomic levels were considered: species, genus, family, order and class, so that the weights are ω = 20 (species in the same genus), 40 (same family but different genera), 60 (same order but different families), 80 (same class but different orders) and 100 (different orders). For semi-quantitative collections performed in mussel beds along the SEPR, these two indices were then tested by comparing them to theoretical values of Δ^+ and Λ^+. Expected values were generated for a given number of species (S) after 1000 re-samplings of the dataset for each S value. The species list came from species found in mussel beds along the whole SEPR by adding species from this study and from

Van Dover (2002, 2003). In the context of this study, the main advantage of these two indices is that they are independent of sample size or sampling effort.

Pielou's eveness (J'), Shannon-Weiner diversity index (H') and rarefaction index $ES_{(n)}$ were calculated for each mussel bed sampling unit from species relative abundance matrices using the DIVERSE subroutine in PRIMER v5. The E(Sn) rarefaction index is the Expected Number of Species which estimates the number of species (S) in samples normalized (rarefied) to successively smaller sizes (Hurlbert, 1971). It was originally designed to allow comparisons of species diversity from sampling units of very different size. This method assumes that samples are homogeneous with similar distribution of individuals among species and that species are randomly distributed. These assumptions are rarely observed in practice and rarefaction index will strongly be biased for small samples. In our case, the sampling units were normalized to the lowest common number of individuals found at Grommit (21°33°S), i.e. 600 (Table 5.1).

Multivariate statistical analyses

Prior to analyses, a Hellinger transformation was applied to the response dataset in order to preserve the Hellinger distance, which is more appropriate for community composition data, instead of the Euclidean distance (Legendre & Gallagher, 2001).

Global analysis of communities along the EPR

For global analysis, putative biogeographic breaks were identified using a multivariate regression tree (MRT) (De'ath, 2002) applied on two different datasets. The first one came from the species list recorded in Desbruyères *et al.* (2006b). The species collected during the BIOSPEEDO cruise were added to the second analysis. This method was used to define group of fields that are similar in faunal composition while geographically adjacent. The MRT is a partitioning method of the multivariate data response table (i.e. species matrix) constrained along the axis of a single explanatory variable. Latitude was used as the constraining variable to generate spatially consistent reliable groups. The tree structure is generated by partitioning the whole dataset into mutually exclusive groups and is grown by successive binary splits of the data, each split involving a breakpoint in the explanatory variable. Each split is chosen to minimize the total multivariate sum of squares within the two groups formed, or maximise the R^2 (or equivalently the among-group sum of squares) at each division step. The tree with the lowest cross-validation error was chosen as the best predictive tree. At the end of the procedure, each leaf of the tree is characterized by the

number of sites, the multivariate mean of the sites and the explanatory variable values that define it. The tree was computed using the 'mvpart' library (De'ath, 2002) in the R statistical language (R Development CoreTeam, 2006).

Variation of the macrofaunal community structure associated with mussel beds along the SEPR

To investigate the correlations of faunal composition with explanatory variables, a redundancy canonical analysis (RDA) was performed on standardized data on macrofauna associated with mussel beds along the Southern EPR. This method combines aspects of ordination and regression and uses a permutation procedure to test the significance of the explained variation; thus species abundances have not to be normally distributed (Legendre & Legendre, 1998). The explanatory variables were the latitude from the equator, the geodynamic environment (two-state variable: tectonic *vs* volcanic), the type of venting (three-state variable: moribund, diffusion zone or chimney wall) and the proportion of mussels juveniles (i.e. <10mm). Permutations tests involving 999 permutations were run to test the significance of each and all canonical axes respectively.

In order to group sample units collected from the mussel beds according to their faunal composition, the chord distances was also calculated between sample units from the relative species abundance data. This measure is similar to the Euclidean distance after normalization of the site vectors and is thus more appropriated when the sample area is variable among sampling units. Data were then clustered using a group average-sorting (UPGMA) method as it gave the best correlation between the chord distances and cophenetic matrices (i.e. level at which two sampling sites i and j become member of the same cluster).

RESULTS

Global analysis of the vent fauna along the EPR

The BIOSPEEDO cruise added 18 species to the list reported in Desbruyères *et al.* (2006b) leading to a final list of 96 species along the whole EPR (Appendix S1). Data acquired during the BIOSPEEDO cruise extended the geographic range of most gastropod and polychaete species that had been previously reported from the Northern EPR (Warén & Bouchet, 2001, Desbruyères *et al*. 2006b; Appendix S1). As an example, *Lepetodrilus tevnianus*, only recorded to date at 11°N, was collected at 14°S, 17°25'S and 21°33'S. *L. cristatus*, only recorded at 13°N, 21°N and from the Galapagos, was found at 14°S and

21°33'S. Similarly, *Echinopelta fistulosa* and *Cyathermia naticoides*, only known from the Northern EPR (i.e. 9°N to 21°N), were collected at 21°33'S and 17°25'S respectively. For the first time, benthic individuals of the genus *Laeviphitus* were collected while this taxon had only been found as larvae in this area (Desbruyères *et al.*, 2006b). Furthermore, this cruise led to the discovery of new species, i.e. *Phymorhyncus* n. sp. at 17°25'S and 18°25'S, *Melanodrymia* n. sp. and *Haloceras* sp., a new genus for hydrothermal vents, at 21°33'S for gastropods (Warén, pers. obs.) and *Paralvinella* n. sp., *Glycera* n. sp. and a new Ampharetidae species at 21°33'S, *Sphaerodopsis* n. sp. at 18°25'S and *Nicomache* n.sp. and a new Flabelligeridae species at 17°25'S and 21°33'S for polychaetes (Desbruyères & Hourdez, pers. obs.), as well as species belonging to genera characterizing other ridges systems (i.e. *Eosipho auzendi*, Pacific Antarctic Ridge and *Shinkailepas* sp. from the Mid-Atlantic Ridge and back-arc basins). These new species were only found south of the latitude of 17°25'S. Moreover, the gastropods *Nodopelta heminoda* and *N. subnoda* and the polychaete *Hesiolyra bergi*, highly abundant on the chimney walls of the NEPR vent fields, were virtually nearly absent (very few individuals) from SEPR samples despite a huge sampling effort (9 chimneys extensively sampled).

When analysing only the species dataset reported in the handbook, species richness was twice higher at all NEPR vent fields (i.e. 42 to 62 species) than at any vent field from SEPR (i.e. 13 to 28 species) (Table 5.3). The highest species richness was observed at 13°N and 21°N (i.e. 62 and 51, respectively) and at 17°25'S (i.e. 28) for the southern part. Conversely average taxonomic distinctness was nearly homogeneous along the whole EPR with the highest value (84.48) found at 17°25'S. Adding the species recorded during the BIOSPEEDO cruise increased significantly the number of species recorded in the SEPR to an extent similar to that of the NEPR. The analysis revealed two centres of species diversity: one in the north including 11°N and 13°N and one in the south included vent fields located between 17°25'S and 21°33'S. Species diversity in the SEPR reached 60 at 17°25'S against 62 at 13°N. The average taxonomic distinctness was not too much affected with a highest value still found along the SEPR and associated with the highest variation coefficients of the taxonomic distinctness (Table 5.3).

Table 5.3 Diversity indices for each hydrothermal vent field along the East-Pacific Rise (EPR).
Data are expressed in presence/absence. EPR handbook: data reported in (Desbruyères *et al.*, 2006b) and EPR hb+bs: species collected in the Southern EPR during the BIOSPEEDO cruise were added to the data.

	EPR hand book			EPR hb + bs		
Vent site	**S**	**Δ^+**	**Λ^+**	**S**	**Δ^+**	**Λ^+**
21°N	51	79.79	513.2	51	79.79	513.2
13°N	62	80.94	536.2	62	80.94	536.2
11°N	42	80.22	560.4	42	80.22	560.4
9°50'N	45	79.41	566.3	45	79.41	566.3
7°25S	21	78.73	555.5	35	79.33	519.4
14°S	22	79.73	520.9	34	79.32	557.8
17°S	28	84.48	445.5	60	81.96	484.9
18°S	14	79.3	472.2	38	83.95	398.7
21°S	13	76.5	473.2	52	80.62	507.1

The multiple cross-validations produced a MRT with three groups of fauna for both analyses, explaining 70.3% of the variation in the species dataset when considering only the handbook and 45.4% when adding the new species records (Fig. 5.2). Those trees displayed the lowest cross-validation error and offered sufficient resolution when considering the number of observed hydrothermal fields along the East Pacific Rise (i.e. 9). The first split separated the northern and southern parts of the EPR more or less at the Equator. A second split sub-divided the SEPR into two groups of assemblages including sites on both sides of the latitude of 18°S when considering only species reported in the Handbook, and between 14°S and 17°25'S when adding species collected during the BIOSPEEDO cruise.

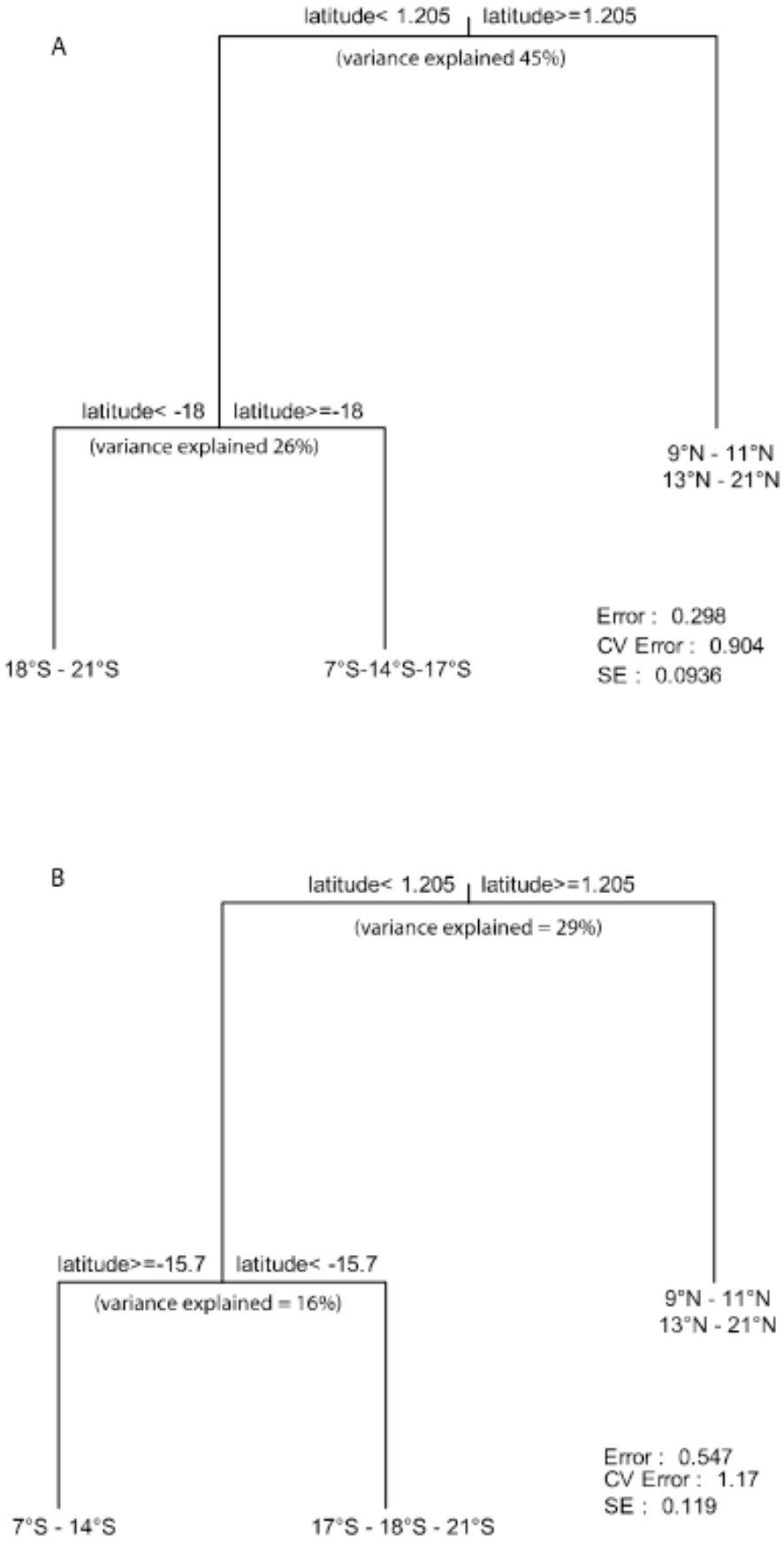


Fig. 5.2 Groups obtained by multivariate regression trees (MRT) analysis of presence/absence data along the EPR. Latitudes are south from the equator. A. Data from the handbook (Desbruyères *et al.*, 2006b). B. Data complemented with species data from the BIOSPEEDO cruise.

Evidence of species crypticism and vicariance in vent gastropods

Phylogenetic relationships were analysed for 18 species of Neolepetopsidae, Lepetodrilidae and Neomphalidae gastropods along the whole EPR using a 425pb sequence portion of the Cytochrome oxidase subunit I (COI). The tree obtained by Bayesian inferences is presented in Fig. 5.3. All nodes with less than a 95% probability were collapsed to basal polytomy. Within the Lepetodrilidae family, we detected thirteen clades with ten clades involving six species of the genus *Lepetodrilus* sp.. The two species of the genus *Gorgoleptis* sp. and *Clypeosectus delectus* were constituted of one clade each. Pairwise K2P Kimura distances between clades detected within each well-identified morphological species, i.e. *L. elevatus*, *L. pustulosus*, and *L. tevnianus*, ranged between 5.5% and 8.3%. Within cryptic lineages (i.e. clades), pairwise distances ranged from 0 to 1.5% for *L. ovalis*. *Lepetodrilus* aff *galriftensis*, identified by Johnson et al. (2008), grouped with the southern clade (9°50N to 21°33S) of *Lepetodrilus elevatus* (Matabos et al. 2008, Plouviez et al. 2009) and *Lepetodrilus* aff *elevatus* (Johnson *et al.* 2008) with the third clade only found at 21°33S. This latter diverged from the northern and southern clades by 7% and 5.8% respectively. *L. ovalis* and *G. emarginatus* also displayed two clades with a slight divergence of 1.5%. *G. spiralis* and *C. delectus* were homogeneous along the EPR with intra-species distance comprised between 0.01 and 1.5%. *Eulepetopsis vitrea* displayed two distinct mitochondrial lineages separated by a divergence of 1.5% and located on each part of the Equator. The only gastropod species of the genus *Lepetodrilus* for which no geographic break was detected is *L. cristatus*, mostly found on the walls of vent chimneys. Within the Neomphalinae, *Peltospira operculata* and *Pachydermia laevis* exhibited two to three divergent clades. *Peltospira operculata* and *Pachydermia laevis* showed the strongest divergence between the NEPR and SEPR with 7.9% and 12.5% divergence between the two clades respectively. An individual of *P. laevis* sequenced from 31°S beyond the Easter microplate showed 10.8% divergence with the NEPR clade and 11.2% with the SEPR, respectively. Pairwise K2P Kimura distances within clades ranged from 0.2% to 0.6%. In all cases, cryptic lineages within a single morphological species were geographically structured with a northern and southern lineage across the Equator and, in a few cases an additional mt lineage at the most southern vent field (i.e. 21°33S). However the absence of samples between 9°50'N and 17°25'S impede to clearly identify the localisation of the geographic break along the ridge, and as it was impossible to identify lineages on the basis of morphological criteria, the species complexes identified

here were considered as a single widely distributed species in further biogeographical analyses.

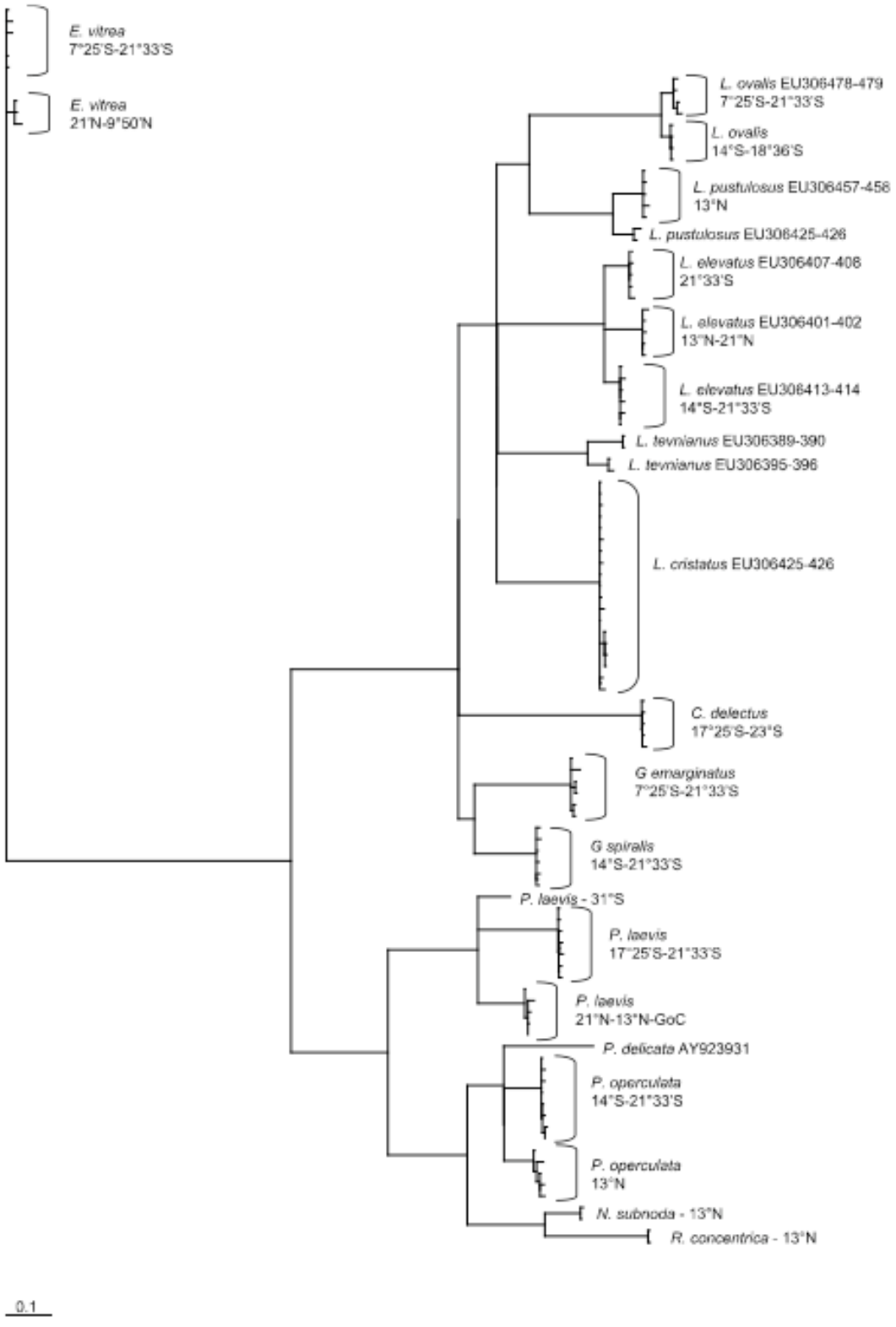


Fig. 5.3 Bayesian tree based on the 111 gastropods COI sequences. Numbers correspond to the posterior probability values of the nodes. Genebank sequences added to the tree were provided by (Johnson *et al.*, 2008). Abbreviations for sites are given in Table 3.

Variations in the macrofaunal community structure of mussel beds along the SEPR

Up to 13,339 individuals belonging to sixtie species were collected in mussel beds along the Southern EPR: twenty-height species of gastropods (nineteen genera and eleven families), thirty species of polychaetes (twenty-two genera and fourteen families) and 2 species of echinoderms (two genera and two families) (Table 5.4). The most abundant species at the regional scale were *Lepetodrilus* sp. (i.e. *L. elevatus* and *L. ovalis*), *Eulepetopsis vitrea*, *Amphisamytha galapagensis* and *Archinome rosacea.* While *Melanodrymia aurantiaca* was only found at 21°33'S, it was highly abundant at that site. Conversely, *Planorbidella* sp. was found in all samples but represented few individuals at each locality.

Table 5.4 List of species collected in 10 mussel beds along the Southern East-Pacific Rise (SEPR).

Taxon	Familily	Species	Sarah's Spring (7°24S)	Lucky Eric (14°S)	Oasis BS6 (17°25S)	Rehu Marka (17°25S)	Susie (17°34S)	Animal Farm (18°36S)	Grommit (1) (21°33S)	Grommit (2) (21°33S)	Grommit (3) (21°33S)
Gastropods	Neoleptopsidae	*Eulepetopsis vitrea*	291	170	20	76	41	449	10	6	21
	Trochidae	*Bathymargarites simplector*	15	0	8	7	0	260	0	0	0
	Sutilizonidae	*Sutilizona theca*	0	0	0	0	0	0	2	0	0
	Lepetodrilidae	*Lepetodrilus elevatus*	15	757	1791	19	1607	2	42	29	102
		Lepetodrilus pustulosus	1877	64	0	0	6	0	0	0	2
		Lepetodrilus ovalis	87	103	63	219	114	218	11	12	57
		Lepetodrilus cristatus	0	25	0	0	2	0	5	10	4
		Lepetodrilus tevnianus	0	17	0	0	4	0	1	2	2
		Gorgoleptis cf. *emarginatus*	73	3	0	17	1	0	0	0	0
		Gorgoleptis spiralis	15	8	32	48	0	10	5	12	15
		Gorgoleptis patulus	0	0	1	0	0	0	0	0	5
		Clypeosectus sp.	58	0	14	23	0	0	0	2	0
	Neomphalidae	*Cyathermia naticoides*	0	0	0	0	0	0	0	0	0
		Melanodrymia auriantica	0	0	0	0	0	0	472	309	458
		Melanodrymia galeronae	0	0	0	0	0	0	1	0	0
		Melanodrymia n. sp.	0	0	0	0	0	0	1	0	0
		Pachydermia cf. *laevis*	0	0	0	33	0	0	18	6	20
		Planorbidella cf. *planispira*	0	3	17	29	2	18	2	2	12
		Neomphalus fretterae	0	0	1	0	0	1	0	0	0
	Peltospiridae	*Nodopelta subnoda*	0	0	0	0	0	0	0	15	0
		Echinopelta fistulosa	0	0	0	0	0	0	0	149	0
		Rhynchopelta concentrica	0	0	0	0	2	0	0	0	0
	Phenacolepapidae	*Shinkailepas* sp.	0	0	0	0	0	0	0	4	0
	Provannidae	*Provanna ios*	0	0	0	9	0	7	0	0	1
	Conidae	*Phymorhyncus* n. sp.	0	0	2	0	0	2	0	0	0
		Phymorhyncus cf. *major*	0	0	0	0	0	2	0	0	0
	Elachisinidae	*Laeviphitus* sp.	0	0	0	1	0	0	0	0	0
	Buccinidae	*Eosipho auzendi*	0	0	0	0	0	6	0	0	4
Polychaetes	Amphinomodae	*Archinome rosacea*	0	0	15	134	0	389	21	70	25
	Dorvilleidae	*Ophryotrocha akessoni*	0	0	4	4	0	22	0	50	15
	Glyceridae	*Glycera* n. sp.	0	0	0	0	0	0	0	0	1
	Hesionidae	*Hesiospina vestimentifera*	0	0	0	52	0	332	1	10	0
		Hesiolyra bergi	0	0	0	0	0	0	1	0	0
		Hesiocaeca cf. *hessleri*	0	0	0	26	0	0	0	0	0
	Nereididae	*Nereis sandersi*	0	0	3	10	0	68	26	46	13

Table 5.4 (continued…)

Taxon	Familily	Species	Sarah's Spring	Lucky Eric	Oasis BS6	Rehu Marka	Susie	Animal Farm	Grommit (1)	Grommit (2)	Grommit (3)
Polychaetes	Phyllodocidae	*Galapagomystides aristata*	0	0	0	0	0	2	0	0	0
	Polynoidae	*Branchinotogluma hessleri*	0	0	0	1	0	0	0	0	0
		Branchinotogluma sandersi	0	0	0	0	0	0	0	0	0
		Levensteiniella kincaidi	15	0	14	18	0	0	0	10	2
		Levensteiniella plicata	0	0	0	1	0	10	1	0	0
		Lepidonotopodium fimbriatum	29	0	0	0	0	0	1	10	3
		Lepidonotopodium williamsae	29	0	4	1	0	0	0	0	0
		Lepidonotopodium atalantae	0	0	1	0	0	0	0	0	0
		Thermiphione tufari	0	0	1	2	0	21	3	4	2
	Serpulidae	*Laminatubus alvinus*	0	0	4	0	0	0	0	0	0
	Siboglinidae	*Oasisia* n. sp.	0	17	1	0	0	0	0	0	0
		Riftia pachyptila	0	0	1	0	4	0	0	0	0
		Tevnia jerichonana	0	0	0	0	0	0	0	2	0
		Branchipolynoe symmytilda	0	3	2	0	0	2	2	0	1
	Maldanidae	*Nicomache* sp. n.	0	0	8	0	0	0	2	6	1
	Spionidae	*Prionospio* sp.	0	0	1	0	0	0	3	6	0
	Flabelligeridae	Flabelligeridae n. sp.	0	0	1	0	0	2	0	2	0
	Sphaerodoridae	*Sphaedoropsis* n. sp.	0	0	0	0	0	40	0	0	0
	Alvinellidae	*Alvinella caudata*	0	0	0	0	0	0	1	0	0
		Paralvinella n. sp.	0	0	0	0	0	0	0	17	0
		Paralvinella pandorae	0	0	0	4	1	0	0	0	0
		Paralvinella grasslei	0	0	0	0	0	0	0	0	0
	Ampharetidae	*Amphisamytha galapagensis*	1455	0	18	36	10	229	14	41	78
		Ampharetidae n. sp.	0	0	0	0	0	0	1	0	1
Echinoderms	Ophiuridae	*Spinophiura jolliveti*	0	0	0	0	0	151	0	0	0
	Ophiocanthidae	*Ophiolamina eprae*	0	0	1	1	0	6	0	0	0

Local species richness varied between 11 and 26 with the lowest values observed at Lucky Eric (14°S), Susie (17°34'S) and Sarah's Spring (7°25'S) (Table 5.5). The highest species diversities, both in terms of species richness and evenness, were observed at the 17°25'S Rehu Marka and 18°36'S Animal Farm sites and at one sample from the 21°33'S Grommit site. However, Shannon indices are relatively low, ranging between 0.47 at Suzie and 2.33 at Rehu Marka (Table 5.5).

Table 5.5 Diversity indices for each quantitative sample collected in mussel beds along the Southern East-Pacific Rise (SEPR).

Vent field	Vent site	N(ind.m²)	S	$ES_{(600)}$	J'	H'(log_e)	Δ^+	Λ^+
7°24S	Sarah's Spring	3959	12	11.64	0.52	1.29	75.45	798
14°S	Lucky Eric	1170	11	10.65	0.51	1.22	67.20	980
17°25S	Oasis-BS6	2028	26	17.44	0.20	0.65	87.02	310
	Rehu Marka	771	24	22.48	0.75	2.33	86.01	374
17°34S	Susie	1794	12	8.85	0.19	0.47	77.09	835
18°36S	Animal Farm	2249	24	19.5	0.71	2.24	88.84	258
21°33S	Grommit (Gr1)	647	25	25.24	0.39	1.26	84.6	431
	Grommit (Gr2)	832	26	25.6	0.68	2.23	86.28	335
	Grommit (Gr3)	845	24	22.2	0.55	1.74	84.2	472

The same trend was observed regarding to average taxonomic distinctness indices. Δ^+ varied between 67.20 for Lucky Eric (14°S) and 88.84 for Animal Farm (18°36'S). The following highest values, observed for collections at Oasis (17°25'S), Rehu Marka (17°25'S) and Grommit (21°33'S), were equivalent. Taxonomic groups were more evenly distributed at sites characterized by the highest average taxonomic distinctness, as suggested by the Λ^+ value. Oasis (17°25'S), Rehu Marka (17°253S), Animal Farm (18°36'S) and one replicate from Grommit (i.e. Gr2, 21°33'S) were significantly higher than the theoretical value estimated by 1000 permutations of the dataset. The sampling unit collected at Lucky Eric (14°S) was significantly smaller (Fig. 5.4) in term of average taxonomic distinctness; and the variation in taxonomic distinctness values observed at Sarah's Spring (7°25'S), Lucky Eric (14°S) and Susie (17°34'S) differed significantly from the theoretical value (Fig. 3).

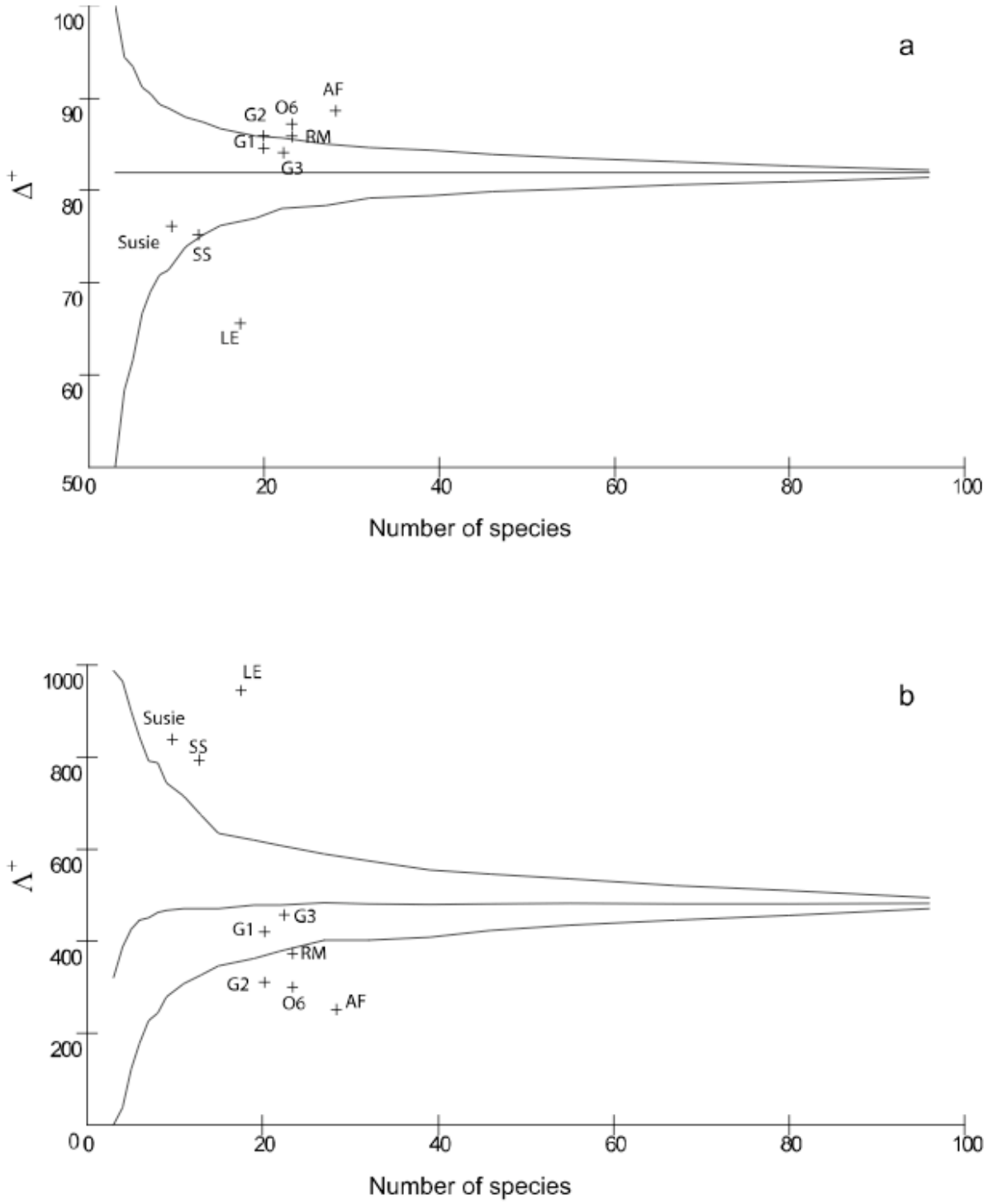


Fig. 5.4 Measured values of Δ^+ (a) and Λ^+ (b) from the semi-quantitative sampling units collected in mussel beds along the SEPR, plotted against the number of species on the simulated 95% confidence interval generated after 1000 resamplings of the same number of species from the total SEPR species list. Sites abbreviations are presented in Table 1.

The cluster analysis shown in Fig. 5.5 indicated a clear separation of sampling sites into three distinct groups plus one isolated site, Sarah's Spring (7°25'S). Sarah's Spring was characterized by low species richness and taxonomic diversity, with the absence of echinoderms and the presence of only four species of polychaetes. This site differs from the others in that the gastropod *Lepetodrilus pustulosus,* almost absent from other samples, was found in a high proportion. *L. pustulosus* and the polychaete *Amphisamytha galapagensis* constituted 85% of the individuals collected. The first group corresponds to the highest values of the Shannon's diversity index (i.e. H') and a high average taxonomic distinctness (Table 5.5) and includes Animal Farm (18°36'S) and Rehu Marka (17°25'S). They were both composed of five and seven dominant species (i.e. >5%). The second group, which includes collections from Lucky Eric (14°S), Oasis BS6 (17°25'S) and Susie (17°34'S), corresponds to the lowest Shannon's diversity index (Table 5.5). They were mainly composed of Lepetodrilid species, with *L. elevatus* present up to 88% and 90% in the sampling units collected at Oasis BS6 and Susie, respectively. Finally, samples collected at Grommit (21°33'S) grouped together with 65% of species similarity. They displayed the same species and taxonomic diversity and contained at least three new species each, and six new species only shared by those samples (i.e. *Melanodrymia* n. sp., *Shinkailepas* sp., *Glycera* n. sp., *Nichomache* n. sp., *Oasisia* n. sp. and the new species of Ampharetidae). While the gastropod *Melanodrymia auriantaca* was completely absent from the other sampling units, it was found at a very high abundance at this site where samples were collected in mussels beds from the chimney wall.

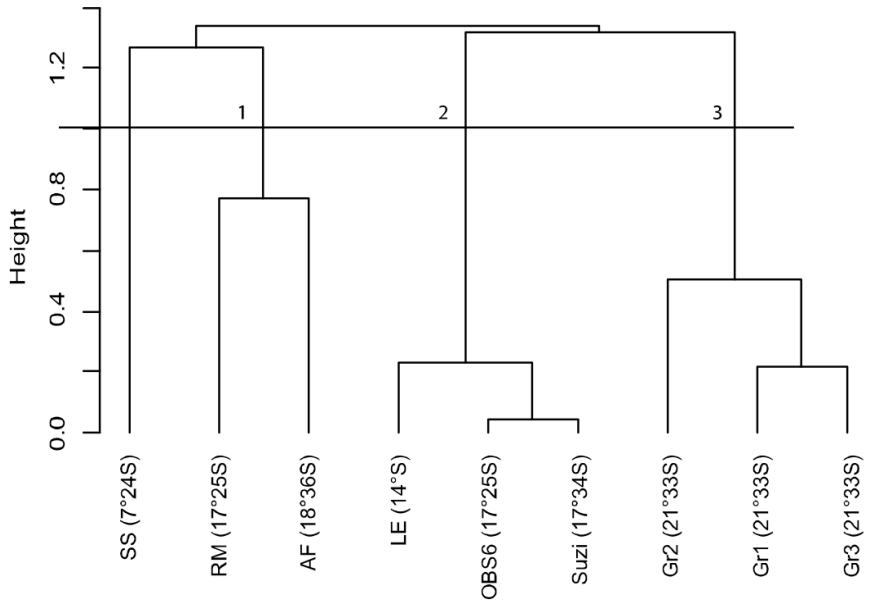


Fig. 5.5 Group-average sorting dendrogram computed from the cord distance among the 9 sampling units collected in mussel beds along the Southern East Pacific Rise. Numbers identify clusters highlighted. Sites abbreviations are presented in Table 1.

Sites and explanatory variables used in the RDA are displayed as an ordination biplot in Fig. 5.6. Results of the RDA showed that the transformed species abundances are significantly correlated with two environmental variables out of four ($R^2adj=0.29$, $p=0.004$). The first axis explained 28.6% ($R^2adj=0.18$, $p=0.003$) of the total variation in species abundance data, and the latitude and the type of venting explained 27.8% ($R^2adj=0.17$) and 26.4% ($R^2adj=0.16$), respectively. However, the power of the test is very low due to the lack of observations and environmental variables in the dataset. This prevents the analysis from reaching a statistical significance when considering each variable alone.

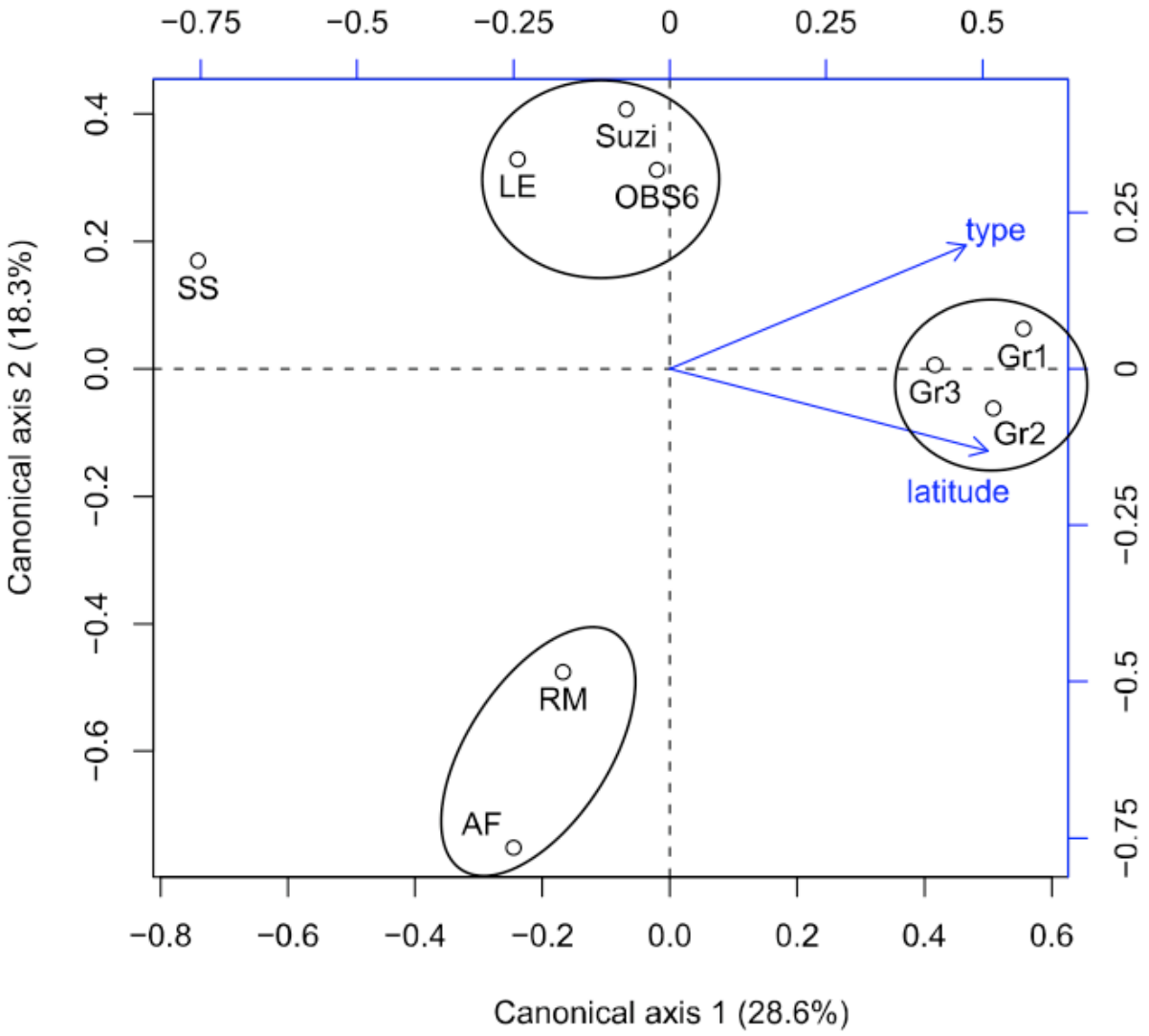


Fig. 5.6 Ordination correlation biplot of the sites and environment variables based on redundancy analysis (RDA) of transformed species abundance data. Abbreviations used for the sites are presented in Table 1. Groups identified with the dendrogram are represented with circles. Species were not represented.

DISCUSSION

Evidence of crypticism in gastropod species

For the Lepetodrilids species, 6.5% divergence was already highlighted for *Lepetodrilus elevatus* separating the northern and southern clades found in sympatry at 9°50N (Matabos *et al.*, 2008b). Johnson *et al*. (2008) also revealed several distinct cryptic species complexes within the genus *Lepetodrilus* with pairwise distances ranging between 4.37 and 31.25%. *Pachydermia laevis* displayed 12.5% divergence between north and Southern EPR, higher than the divergence found with the individual collected on the other side of the microplate. Differentiation between individuals of *Planorbidella planispira* found on both sides of the equator fell in the same range as the Lepetodrilids species with 7.9% divergence. Strong genetic differentiation between Northern and Southern EPR were detected for other divergent taxonomic groups including bivalves and polychaetes (Plouviez *et al.*, 2009). While levels of divergence can be highly variable between two species (Knowlton, 2000), pairwise genetic distances between and among clades highlighted in this study fit inter- and intra-specific levels respectively (Peek *et al.*, 1997). Since the development of molecular techniques, occurrence of sibling species appears to be more widespread in hydrothermal ecosystems, and at larger scope, in the sea (Peek *et al.*, 1997; Knowlton, 2000; Johnson *et al.*, 2008; Matabos *et al.*, 2008b).

Biogeographic entities along the EPR

Barcoding of hydrothermal species (Johnson *et al.*, 2008), population genetics studies (e.g. Chevaldonné *et al.*, 2002; Won *et al*., 2003; Hurtado *et al.*, 2004; Matabos *et al*., 2008b) and comparative phylogeography (Plouviez *et al*., 2009) along the EPR converge to the existence of a population genetic break on both sides of the equator. Presence of topographic barriers in relation with tectonic events (i.e. formation of transform faults, microplates, bathymetric inflation) favoured the role of vicariance events and subsequent population differentiation at hydrothermal vents (Tunnicliffe & Fowler, 1996). Several topographic features, related to the ridge history, have been proposed to explain the divergences observed within morphological species. The formation of the Bauer microplate, 17-6 Mya (Matabos *et al.*, 2008b), and the nearly simultaneous fossilization of the Mathematician ridge (Plouviez *et al*., 2009) were proposed to account for the divergence observed in the split of *Lepetodrilus elevatus* into two cryptic lineages. Four out of six Lepetodrilid species display the same range of genetic divergence (i.e. 5.2 to 9.2;

Johnson *et al.*, 2008; Matabos *et al*, 2008b; this study) and may thus share the same past history. However, levels of divergence are variable among species with two sets of values ranged from 6-12% (an 'old' vicariant event involving *L. elevatus, L. pustulosus, L. tevniatus, P. operculata, P. laevis*) to 1.5-3% ('recent'effect involving *L. ovalis*, *E. vitrea*, *G. emarginatus*), respectively. Tests on shared divergence times were performed on seven hydrothermal vent species along the EPR by Plouviez *et al*. (2009) and revealed the occurrence of a common vicariant event ~1.3 Mya for all studied species, including *L. elevatus* for which another older speciation event is also detected around 11.5 Mya. These latter authors therefore proposed that the formation of transform faults around 1 to 2 Mya between 4°S and 14°S, may be responsible for this vicariant break and the emergence of a transition zone in this area.

Analyses of community composition expressed in terms of presence/absence support the molecular dataset. Recently, Bachraty *et al.* (2009) suggested that the EPR is separated into two different biogeographic provinces: the NEPR *vs* the SEPR. Our MRT analyses confirmed the presence of a geographic break at the equator separating the EPR fauna into two biogeographic entities. However, species lists and molecular data between 9°N and 7°25'S are lacking and further studies centered on the Hess Deep are needed to exactly position the break. The two evolutionary lineages of *Lepetodrilus elevatus* were found in sympatry at 9°50N (Matabos *et al.*, 2008b) were and able to hybridize so other cryptic complex species may also overlap around the equator. This is especially the case for *Lepetodrilus ovalis* and *Gorgoleptis emarginatus* which showed a consistent but weak divergence between two distinct mitochondrial lineages that overlap over nearly all the latitudinal gradient, with one lineage dominating the northern part of the EPR. *A contrario*, the gastropod *E. vitrea*, which displays the same level of divergence exhibits a complete geographic separation of its two cryptic lineages at 7°25'S (Plouviez *et al.* 2009). One explanation could be the differences in life history strategies that can influence dispersal abilities. Some species may have indeed a dispersal potential to cross the barrier and meet secondarily. Genome homogenisation between the previously isolated clades is then depending on the strength of selection against hybrids (genetic barrier).

Added to the occurrence of cryptic species from both parts of the EPR, the BIOSPEEDO cruise brought at least ten new species for hydrothermal vents (three gastropods, six polychaetes and one echinoderm; Stöhr & Segonzac, 2006) and expands the range of several species previously described from the NEPR. The wide geographic range of most vent species confirms studies previously conducted by Van Dover (2002) in

the same area. However, while the first studies based on megafauna observations (Geistdoerfer *et al.*, 1995; Halanych *et al.*, 1999) and the quantitative analysis of communities associated with mussel beds (Van Dover, 2002, 2003) concluded that the fauna was similar between the Northern and Southern EPR, results from the present study highlighted several latitudinal changes in community composition between 17°S and 21°S. New species were found beyond 17°S and some species warrant further studies as their identifications were uncertain (Hourdez & Desbuyères; A. Warén, pers com.). In the listed taxa, two species, *Shinkailepas* sp. and *Eosipho auzendi*, represented by two and more individuals respectively, belong to a genus so far known only from the Mid-Atlantic Ridge and the Northwest Pacific back-arc basins for *Shinkailepas* sp. (Desbruyères *et al.*, 2006b). The presence of *Eosipho* was already reported at Rehu Marka by (Van Dover, 2002). Added to the presence of at least five cryptic species (i.e. *Lepetodrilus elevatus*, *L. pustulosus*, *L. tevnianus*, *Planorbidella planispira*, *Pachydermia laevis*), the East-Pacific Rise appears to be a more complex biogeographic zone than it was first suggested. When taking into account only the species recorded in Desbruyères *et al.* (2006b), a geographic break was established at 18°S and the highest species diversity was found at 13°N. When adding species collected on the SEPR during the BIOSPEEDO cruise, species diversity was not different anymore between NEPR and SEPR vent sites and the geographic break along the SEPR appears positioned between 14°S and 17°25'S. The hypothesis of Bachraty *et al*. (2009) that the NEPR could be a centre of dispersal, is probably due to the over-sampling processed on the northern part of the EPR. Observations and molecular data led to the assumption that the high species diversity observed on this part of the SEPR is the result of the overlap of several distinct biogeographic provinces. The new species of *Phymorhyncus* sp. had been already found at 31°S (A. Warén, pers. obs.) in concordance with the presence of the gastropod *Eosipho auzendi*, described from the Pacific Antarctic ridge. Individuals of *Shinkailepas* sp., a genus only described on the Mid-Atlantic Ridge and in the back-arc basins, were encountered at 21°33'S. Furthermore the average taxonomic distinctness for the sampling units collected in mussel beds located between 17°25'S and 21°33'S (i.e. Oasis BS6, Rehu Marka, Animal Farm and Grommit) deviated significantly from the theoretical number of expected species, reinforcing the hypothesis of "abnormal" hot spot of diversity in this area. The southern part of the SEPR (i.e. between 17°25'S and 21°33'S) could be thus a transition (mixing) zone between the NEPR, the Pacific Antarctic ridge, and possibly the back-arc basins of the western Pacific, supporting first observations made during the BIOSPEEDO cruise (Jollivet *et al.*, 2005). This

assumption is supported by the co-occurrence of NEPR-like *Alvinella* colonies and *Chorocaris* sp. swarms at 21°33'S, shrimp swarms being one of the most striking characteristic of the Mid-Atlantic Ridge, the Indian Ridge and the Northwest Pacific back-arc basins (Van Dover *et al.*, 2001a; Jollivet *et al.*, 2005).

Mussel beds community structure and composition along the Southern East-Pacific Rise

The latitude gradient in the global MRT analysis accounted for 45.4% of the variation in the presence/absence species dataset when adding species collected during the BIOSPEEDO cruise, suggesting additional role of regional and local factors on the community composition. Diversity associated with mussel beds was thus analysed along the SEPR to better understand processes influencing community structure at the regional scale. Diversity associated with mussel beds has been extensively studied along the East-Pacific Rise (Van Dover & Trask, 2000; Van Dover, 2002; Turnipseed *et al.*, 2003; Van Dover, 2003; Bergquist *et al.*, 2005; Dreyer *et al.*, 2005; Van Dover & Doerries, 2005). Most of those studies were conducted from a quantitative sampling including vagile epifauna and large macrofauna juveniles (e.g. crabs). This difference can account for the lower species richness observed in this study compared to previous studies performed along the SEPR (Van Dover, 2002, 2003). Whatever the sampling strategy used, diversity recorded along the SEPR was always higher than values reported from the NEPR (Van Dover, 2003; Dreyer *et al.*, 2005). Results from the canonical analysis showed that latitude combined with the type of venting at a regional scale significantly influenced the community structure and composition associated with mussel beds along the SEPR. The Oasis (17°25'S), Rehu Marka (17°24'S), Animal Farm (18°36'S) and Grommit (21°33'S) sites were characterized by a high diversity when compared to the most northern sites. At Sarah's Spring (7°25'S) and Lucky Eric (14°S) sampling localities, mussels were distributed in a scattered fashion which could explain the low diversity observed. Indeed, the low diversity observed was related to a low number of polychaete species in the sampling units. This is further supported by the significant deviation of the variation in taxonomic distinctness from theoretical values for those sampling units.

The significant influence of the type of venting highlighted the role of local processes on the patterns observed. Edifice life spans are highly variable within a single vent field in response to stochastic events altering fluid discharge and mineralisation processes. Variations in fluid discharge intensity greatly influence physico-chemical

parameters and biotic interactions between species (Mullineaux *et al.*, 2003) which, will in turn greatly impact the community structure and composition. Species composition is thus highly variable in time and space within a single site and strong differences are observed according to the successional stage of the community at the time of sampling (Jollivet, 1996; Sarrazin *et al.*, 1997; Shank *et al.*, 1998). On the EPR, productivity, species richness and abundances of invertebrates associated with mussel beds appeared to be lower at waning and older sites than at young and active sites (Van Dover, 2003; Dreyer *et al.*, 2005). In this study, mussel bed's age was unknown, and the temporal state of mussel beds was assessed via demographic structure of mussel populations and the proportion of juveniles. However mussel demographic structure did not reflect *in situ* observations of the assemblages (i.e. presence of decaying vestimentiferan tubes at Sarah's Spring) and no correlation has been found between community structure and the mussel size frequency distribution (unpublished data) or the proportion of mussels' juveniles. This latter was high at the moribund site Animal Farm (i.e. 56%, Table 5.1) while a low proportion of juveniles characterized the large and dense mussel beds developing in the waxing site Grommit (21°S) (i.e. between 0 and 9%). Thus proportion of mussel juveniles does not appear to be a good proxy to characterize mussel beds conditions.

Analyses showed however that the type of venting had a significant influence when combined with the latitude. First, the lack of sampling units along the gradient can impede the test to reach the level of significance when considering the type of habitat alone. On the other hand, this result could reflect the intermingling role of local and regional factors. Indeed, processes acting at different spatial scales are often imbricate (Levin *et al*., 2001) such that it is difficult to identify their relative role in the absence of replicate samples. Thus, although the influence of the geodynamic environment was not significant, the type of venting is strongly related to the regional dynamic of the ridge (Fouquet *et al*., 1994). All chimneys were found in vent fields characterized by a tectonic environment (i.e. Grommit-21°33'S and Susie-17°34'S). Areas driven by volcanic activity are more unstable in time and space than tectonic areas (Fouquet *et al.*, 1994). Those differences, by controlling the extinction rate of sites, will impact the extinction/recolonisation rate of the vent fauna, promote habitat heterogeneity and reduce distances between sites (Jollivet *et al.* 1999). While unstable environments are expected to reduce species diversity (Juniper & Tunnicliffe, 1997), some authors suggested that the resulting small distances between vent sites should account for bursts of gene flow (Jollivet *et al.* 1999) and thus result in greater

rate of site recolonisation promoting a higher species diversity (Turnipseed *et al.*, 2003; Van Dover & Doerries, 2005). In this study, species and taxonomic diversities were equivalent in the two types of vent dynamics, suggesting that differences resided in species composition. As an example, three gastropod species (i.e. *Melanodrymia aurantiaca*, *Echinopelta fistulosa* and *Nodopelta subnoda)*, known to colonize active black smokers' chimney walls (Jollivet, 1996), were only collected at the Grommit site (21°33'S) where musselbeds developed at the base and on the wall of a black smoker. Similarly, scavenger species (i.e. *Phymorhyncus* spp., *Eosipho auzendi*, *Spioniphura jolliveti*) were mainly found in sites located in volcanic areas (i.e. Oasis BS6, Rehu Marka and Animal Farm).

Nonetheless, while all sampling units collected in the tectonic 21°33'S vent field (i.e. Gommit) were similar in species composition, strong differences were observed at the volcanic 17°25'S area (i.e. Oasis BS6 and Rehu Marka). In unstable volcanic areas, the high rate of spatio-temporal changes in fluid fluxes will result in the intermingling of different successional stages within a site, thus enhancing the variance among mussel beds in a given area. Thus, while found on dome-shaped ridge, the apparent moribund sites Sarah's Spring and Animal Farm displayed large differences in community structure. A low diversity characterized the sample collected at Sarah's Spring with the overwhelming presence of the gastropod *Lepetodrilus elevatus* (48%) and the polychete *Amphisamytha galapagensis* (37%). Lepetodrilid species are grazers and are among the first species to colonize a new site, thus not representative of waning sites and an indicator of a vent re-opening conditions. Indeed, the presence on this site of small specimens of the tubeworms *Tevnia jerichonana* and *Oasisia* sp. (D. Jollivet, pers. obs.) suggests the starting of a new site closed to an ancient vestimentiferan community. Those results illustrate the importance of regional dynamic on local processes.

Latitudinal gradient and venting type explained less than half of the variation in the species data, suggesting that other local factors (i.e. biotic and abiotic factors) independent of the type of venting may account for a large proportion of the variation observed. Small scale habitat heterogeneity (i.e. within a single animal aggregation) can highly influence species composition in response to species environmental tolerances and biotic interactions. Given the high variability in physical and chemical conditions at the centimetre scale within a single colony (Le Bris *et al*., 2006), even a small sampled area can integrate a wide range of physico-chemical conditions. For example, at the Grommit site (21°33'S), two peltospirids species (i.e. *Melanodrymia aurantiaca* and *Echinopelta fistulosa*), usually associated with alvinellid colonies (Jollivet, 1996) were collected in

mussels from active black smokers' chimney wall. On the other hand a sampling area smaller than a metre square may not be representative of the diversity of the whole community. Species distribution can vary between the centre and periphery of a single mussel bed (Bergquist *et al.*, 2005) in response to high variability in environmental conditions within one metre (Sarrazin *et al.*, 2006). Indeed, mobile gastropod species associated with alvinellid colonies, tubeworm clumps, bivalves and suspension-feeders, are not exclusive to a specific zone but rather occupy specific microhabitats that are present in more than one megafaunal community (Mills *et al.*, 2007; Matabos *et al.*, 2008a). Biotic interactions, including competition, predation and facilitation, may also play an important role on community structure.

Limitations and issues

The recent discovery of cryptic species following the emergence of bar-coding approaches raises the problem of species identification and thus the assessment of the vent biodiversity. This appears to be a common issue in the sea. Ubiquist species are frequent occurrences in marine invertebrates and vertebrates and often found over a wide range of latitudes from the poles to the tropics (Knowlton, 1993). But good species identification is needed to understand the vent community history, colonisation pathways, and origin(s) of the fauna. Those observations reinforce the importance of comparative phylogeography to complement analytical approaches of historical biogeography (Arbogast & Kenagy, 2001). Correct assessments of species boundaries are fundamental to test biogeographic hypotheses (Palumbi, 1996). Furthermore, species identifications are not always done by specialists, which lead to awkward comparisons between taxa lists. Valid comparisons across studies require a standardisation process for the identification of organisms. This raises another issue related to species recording and species distribution areas and the validity of lists available in the literature. Individuals identified by acknowledged specialists and recorded in a museum with a vouch number appear to be a safe standard for biodiversity assessment and record. However, ecological studies are more numerous than systematics studies and lots of species lists are provided by ecologists (Van Dover, 2002, 2003; Governar *et al*., 2005; Matabos *et al*. 2008a), so that it is still awkward to assess the real biodiversity associated with that part of the EPR. Differences in sampling methods (i.e. quantitative *vs* qualitative) can also account for the differences observed. Sampling devices have been independently developed in the different labs and may thus not be comparable. Finally, the status of rare species also differs between systematic and

ecological studies and should be seriously considered in biogeographic approaches. Most ecological studies focus on a particular vent habitat, and may thus underestimate the 'real' local biodiversity. To our knowledge, the sediment habitat surrounding vents has been greatly neglected, so that many species may have never been collected.

This study not only reinforces the existence of two biogeographic provinces along the East Pacific Rise separating the North from the South EPR, but also suggests that the 'abnormal' high species diversity associated with the southern part of the SEPR (17°S-21°S) could result from an overlap of several biogeographic entities. Further comparisons are needed to confirm this assumption. Additional samples from the poorly explored Pacific Antarctic Ridge area could help in the understanding of biodiversity patterns associated with the Pacific hydrothermal vent systems. Combining species diversity analyses with comparative phylogeography will help to unveil colonisation pathways that have led to the present species distribution. On the other hand, the lack of standards in species classification for deep-sea biogeography and the cross-validation of species' identification can explain the divergences observed when comparing studies conducted in the same parts of the ocean. The analysis of diversity associated with mussel beds was a first step towards a greater understanding on the role of global, regional and local factors on biodiversity patterns. So far, combination of the latitudinal gradient and the venting significantly influenced the community structure but more replicates and explanatory variables are needed to ascertain such an assumption. This study will help designing future macroecology studies along the EPR aiming to understand processes influencing the distribution of biodiversity.

AKNOWLEDGMENTS

We thank the crew and pilots of the RV L'Atalante and the DSV Nautile for their assistance and technical support during the cruise BioSpeedo'04. We also thank the sequencing/genotyping platform GENOMER for mt COI sequences acquisition. This work was financially supported by the French programme Dorsales (INSU, Ifremer, CNRS), and the GDR Ecchis. A grant form the European SYNTHESIS Project funded the collaboration with the Swedish Museum of Natural History. This study also represents a contribution to the ANR Biodiversité 'Deep Oases' (ANR-06-BDIV-005).

REFERENCES

Arbogast, B.S. & Kenagy, G.J. (2001) Comparative phylogeography as an integartive approach to historical biogeography. *Journal of Biogeography*, **28**, 819-825.

Auzende, J.-M., Ballu, V., Batiza, R., Bideau, D., Cormier, M.-H., Fouquet, Y., Geistdoerfer, P., Lagabrielle, Y., Sinton, J. & Spadea, P. (1994) Activité magmatique, tectonique et hydrothermale actuelle sur la dorsale est pacifique entre 17°S et 19°S. (campagne naudur). *Comptes Rendus de l'Académie des Sciences, série II*, **319**, 811-818.

Bachraty, C., Legendre, P. & Desbruyeres, D. (2009) Biogeographic relationships among deep-sea hydrothermal vent faunas at global scale. *Deep-Sea Research Part I Oceanographic Research Papers*, **56**, 1371-1378.

Bergquist, D.C., Fleckenstein, C., Knisel, J., Begley, B., Macdonald, I.R. & Fisher, C.R. (2005) Variations in seep mussel bed communities along physical and chemical environmental gradients. *Marine Ecology Progress Series*, **293**, 99-108.

Briggs, J.C. (2006) Proximate sources of marine biodiversity. *Journal of Biogeography*, **33**, 1-10.

Briggs, J.C. (2007) Marine biogeography and ecology: Invasions and introductions. *Journal of Biogeography*, **34**, 193-198.

Chevaldonné, P., Jollivet, D., Desbruyères, 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. *Cahiers de Biologie Marine*, **43**, 367-370.

Clarke, K.R. & Warwick, R.M. (2001) *Change in marine communities: An approach to statistical analysis and interpretation*, 2nd edition PRIMER-E Ltd edn, Plymouth, UK.

De'ath, G. (2002) Mulitvariate regression trees: A new technique for modeling species - environment relationships. *Ecology*, **83**, 1105-1117.

Desbruyères, D., Hashimoto, J. & Fabri, M.-C. (2006a) Composition and biogeography of hydrothermal vent communities in western pacific back-arc basins. *Geophysical monograph series*, **166**, 215-234.

Desbruyères, D., Segonzac, M. & Bright, M. (2006b) *Handbook of deep-sea hydrothermal vent fauna.*, second edition edn. Denisia.

Dreyer, J.C., Knick, K.E., Flickinger, W.B. & Van Dover, C.L. (2005) Development of macrofaunal community structure in mussel beds on the northern east pacific rise. *Marine Ecology Progress Series*, **302.**

Embley, R.W., Lupton, J.E., Massoth, G.J., Urabe, T., Tunnicliffe, V., Butterfield, D.A., Shibata, T., Okano, O., Kinioshita, M. & Fujioka, K. (1998) Geological, chemical, and biological evidence for recent volcanism at 17.5°s: East pacific rise. *Earth and Planetary Science Letters*, **163**, 131-147.

Folmer, O., Black, M.B., Hoeh, W.R., Lutz, R.A. & Vrijenhoek, R.C. (1994) DNA primers for the amplification of mitochondrial cytochrome oxydase subunit 1 from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294-299

Fornari, D.J. & Embley, R.W. (1995) Tectonic and volcanic controls on hydrothermal processes at the mid-ocean ridge: An overview based on near-bottom and submersible studies. *Seafloor hydrothermal systems* (ed. by S.E. Humphris, R.A. Zierenberg, L.S. Mullineaux and R.E. Thomson). American Geophysical Union, Washington DC.

Fouquet, Y., Auzende, J.-M., Ballu, V., Batiza, R., Bideau, D., Cormier, M.-H., Geistdoerfer, P., Lagabrielle, Y., Sinton, J. & Spadea, P. (1994) Variabilité des manifestations hydrothermales actuelles le long d'une dorsale ultra-rapide: Exemple de la dorsale est pacifique entre 17°s et 19°s (campagne naudur). *Comptes Rendus de l'Académie des Sciences, série II*, **319**, 1399-1406.

Fujioka, K., Massoth, G.J., De Ronde, C., Tanaka-Nakano, A., Gendron, J., Kinoshita, M., Nishizawa, A., Okano, O., Paradis, G., Roe, K.J., Shinohara, M., Shitashima, K., Tsunogai, U. & Ytow, N. (1995) Active hydrothermal systems of a super-fast spreading ridge, southern east pacific rise (13°58's to 18°26s). *InterRidge News*, **4**, 7-10.

Geiger, D.L. & Thacker, C.E. (2005) Molecular phylogeny of vetigastropoda reveals non-monophyletic scissurellidae, trochoidea, and fissurelloidea. *Molluscan Research*, **25**, 47-55.

Geistdoerfer, P., Auzende, J.-M., Batiza, R., Bideau, D., Cormier, M.-H., Fouquet, Y., Lagabrielle, Y., Sinton, J. & Spadea, P. (1995) Hydrothermalisme et communautés animales associées sur la dorsale du pacifique oriental entre 17°s et 19°s (campagne naudur, décembre 1993). *Comptes Rendus de l'Académie des Sciences, série II*, **320**, 47-54.

Halanych, K.M., Tieger, M., O'mullan, G.D., Lutz, R.A. & Vrijenhoek, R.C. (1999) Brief description of biological communities at 7°s on the east pacific rise. *International Ridge-Crest Research: Biological studies*, **8**, 23-27.

Hawksworth, D.L. & Kalin-Arroyo, M.T. (1995) Magnitude and distribution of biodiversity. *Global biodiversity assessment* (ed. by V.H. Heywood), pp. 107-192. University Press, Cambridge.

Heads, M. (2005) Towards a panbiogeography of the seas. *Biological Journal of the Linnean Society*, **84**, 675-723.

Hurtado, L.A., Lutz, R.A. & Vrijenhoek, R.C. (2004) Distinct patterns of genetic differentiation among annelids of eastern pacific hydrothermal vents. *Molecular Ecology*, **13**, 2603-2615.

Hurtado, L.A., Mateos, M., Lutz, R.A. & Vrijenhoek, R.C. (2002) Molecular evidence for multiple species of *oasisia* (annelida: Siboglinidae) at eastern pacific hydrothermal vents. *Cahier de Biologie Marine*, **43**, 377-380.

Johnson, S.B., Warén, A. & Vrijenhoek, R.C. (2008) DNA barcoding of lepetodrilus limpets reveals cryptic species. *Journal of Shellfish Research*, **27**, 43-51.

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. & Party, T.S.S. (2005) The biospeedo cruise: A new survey of hydrothermal vents along the south east pacific rise from 7°24s to 21°33s. *InterRidge News*, **13**, 20-26.

Juniper, S.K. & Tunnicliffe, V. (1997) Crustal accretion and the hot vent ecosystem. *Philosophical Transactions of the Royal Society of London. Series A*, **355**, 459-474.

Knowlton, N. (1993) Sibling species in the sea. *Ann. Rev. Ecol. Syst.*, **24**, 189-216.

Knowlton, N. (2000) Molecular genetic analyses of species boundaries in the sea. *Hydrobiolgia*, **420**, 73-90.

Le Bris, N., Govenar, B.W., Le Gall, C. & Fisher, C.R. (2006) Variability of physico-chemical conditions in 9°50'n epr diffuse flow vent habitats. *Marine Chemistry*, **98**, 167-182.

Legendre, P. & Legendre, L. (1998) *Numerical ecology*. Elsevier.

Levin, L.A., Etter, R.J., Rex, M.A., Gooday, J., Smith, C.R., Pineda, J., Stuart, C.T., Hessler, R.R. & Pawson, D. (2001) Environmental influences on regional deep-sea species diversity. *Annual Review of Ecology and Systematics*, **32**, 51-93.

Little, C.T.S. & Vrijenhoek, R.C. (2003) Are hydrothermal vent animals living fossils? *Trends in ecology and evolution*, **18**, 582-588.

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

Matabos, M., Le Bris, N., Pendlebury, S. & Thiébaut, E. (2008a) Role of physico-chemical environment on gastropod assemblages at hydrothermal vents on the east pacific rise (13°n/epr). *Journal of the Marine Biological Association of the United Kingdom*, **88**, 995-1008.

Matabos, M., Thiébaut, E., Le Guen, D., Sadosky, F., Jollivet, D. & Bonhomme, F. (2008b) Geographic clines and stepping-stone patterns detected along the east pacific rise in the vetigastropod lepetodrilus elevatus reflect species crypticism. *Marine Biology*, **153**, 545-563.

Micheli, F., Peterson, C.H., Mullineaux, L.S., Fisher, C.R., Mills, S.W., Sancho, G., Johnson, G.A. & Lenihan, H.S. (2002) Predation structures communties at deep-sea hydrothermal vents. *Ecological monographs*, **72**, 365-382.

Mills, S.W., Mullineaux, L.S. & Tyler, P.A. (2007) Habitat associations in gastropod species at east pacific rise hydrothermal vents (9°50'n). *Biological Bulletin*, **212**, 185-194.

Monge-Najera, J. (2008) Ecological biogeography: A review with emphasis on conservation and the neutral model. *Gayana*, **72**, 102-112.

Mullineaux, L.S., Fisher, C.R., Peterson, C.H. & Schaeffer, S.W. (2000) Tubeworm succession at hydrothermal vents: Use of biogenic cues to reduce habitat selection error? *Oecologia*, **123**, 275-284Mullineaux, L.S., Mills, S.W. & Goldman, E. (1998) Recruitment variation during a pilot colonization study of hydrothermal vents (9°50'n, east pacific rise). *Deep Sea Research II*, **45**, 441-464.

Mullineaux, L.S., Peterson, C.H., Micheli, F. & Mills, S.W. (2003) Successional mechanism varies along a gradient in hydrothermal fluid flux at deep-sea vents. *Ecological monographs*, **73**, 523-542.

Palumbi, S.R. (1996) What can molecular genetics contribute to marine biogeography? *Journal of Experimental Marine Biology and Ecology*, **203**, 75-92.

Peek, A.S., Gustafson, R.G., Lutz, R.A. & Vrijenhoek, R.C. (1997) Evolutionary relationships of deep-sea hydrothermal vent and cold-water seep clams (bivalvia: Vesicomyidae): Results from the mitochondrial cytochrome oxidase subunit i. *Marine Biology*, **130**, 151-161.

Sancho, G., Fisher, C.R., Mills, S.W., Micheli, F., Johnson, G.A., Lenihan, H.S., Peterson, C.H. & Mullineaux, L.S. (2005) Selective predation by the zoarcid fish *thermarces cerberus* at hydrothermal vents. *Deep Sea Research I*, **52**, 837-844.

Sarrazin, J. & Juniper, S.K. (1999) Biological characteristics of a hydrothermal edifice mosaic community. *Marine Ecology Progress Series*, **185**, 1-19.

Sarrazin, J., Juniper, S.K., Massoth, G.J. & Legendre, P. (1999) Physical and chemical factors influencing species distributions on hydrothermal sulfide edifices of the juan de fuca ridge, northeast pacific. *Marine Ecology Progress Series*, **190**, 89-112

Sarrazin, J., Robigou, V., Juniper, S.K. & Delaney, J.R. (1997) Biological and geological dynamics over four years on a high-temperature sulfide structure at the juan de fuca ridge hydrothermal observatory. *Marine Ecology Progress Series*, **153**, 5-24.

Sarrazin, J., Walter, C., Sarradin, P.-M., Brind'amour, A., Desbruyères, D., Briand, P., Fabri, M.-C., Van Gaever, S., Vanreusel, A., Bachraty, C. & Thiebaut, E. (2006) Community structure and temperature dynamics within a mussel assemblage in the southern east pacific rise. *Cahiers de Biologie Marine*, **47**, 483-490.

Shank, T.M., Fornari, D.J., Von Damn, K.L., Lilley, M.D., Haymon, R.M. & Lutz, R.A. (1998) Temporal and spatial patterns of biological community development at nascent deep-sea hydrothermal vents (9°50'n, east pacific rise). *Deep Sea Research II*, **45**, 465-515.

Spalding, M.D., Fox, H.E., Allen, G.R., Davidson, N., Ferdaña, Z.A., Finlayson, M., Halpern, B.S., Jorge, M.A., Lombana, A., Lourie, S.A., Martin, K.D., Mcmanus, E., Omolnar, J., Recchia, C.A. & Robertson, J. (2007) Marine ecoregions of the world: A bioregionalization of coastal and shelf areas. *BioScience*, **57**, 573-583.

Stöhr, S. & Segonzac, M. (2006) Two new genera and species of ophiuroid (echinodermata) from hydrothermal vents in the east pacific. *Species diversity*, **11**, 7-32.

Tunnicliffe, V. (1988) Biogeography and evolution of hydrothermal-vent fauna in the eastern pacific ocean. *Proceedings of the Royal Society of London B*, **233**, 347-366.

Tunnicliffe, V. (1991) The biology of hydrothermal vents: Ecology and evolution. *Oceanography and Marine Biology. An Annual Review*, **29**, 319-407.

Tunnicliffe, V. & Fowler, C.M.R. (1996) Influence of sea-floor spreading on the global hydrothermal vent fauna. *Nature*, **379**, 531-533.

Tunnicliffe, V., Mcarthur, A.G. & Mchugh, D. (1998) A biogeographical perspective of the deep-sea hydrothermal vent fauna. *Advances in Marine Biology*, **34**, 353-442.

Turnipseed, M., Knick, K.E., Lipcius, R.N., Dreyer, J.C. & Van Dover, C.L. (2003) Diversity in mussel beds at deep-sea hydrothermal vents and cold seep. *Ecology letters*, **6**, 518-523.

Tyler, P.A., German, C.R., Ramirez-Llodra, E. & Van Dover, C.L. (2003) Understanding the biogeography of chemosynthetic ecosystems. *Oceanologica Acta*, **25**, 227-241.

Van Dover, C., German, C.R., Speer, K.G., Parson, L.M. & Vrijenhoek, R.C. (2002) Evolution and biogeography of deep-sea vent and seep invertebrates. *Science*, **295**, 1253-1257.

Van Dover, C., Homphris, S.E., Fornari, D.J., Cavanaugh, C.M., Collier, R., Goffredi, S.K., Hashimoto, J., Lilley, M.D., Reysenbach, A.L., Shank, T., Von Damm, K., Banta, A., Gallant, R.M. & Vrijenhoek, R.C. (2001a) Biogeography and ecological setting of indian ocean hydrothermal vents. *Science*, **294**, 818-823.

Van Dover, C. & Trask, J.L. (2000) Diversity at deep-sea hydrothermal vent and intertidal mussel beds. *Marine Ecology Progress Series*, **195**, 169-178.

Van Dover, C.L. (2002) Community structure of mussels beds at deep-sea hydrothermal vents. *Marine Ecology Progress Series*, **230**, 137-158.

Van Dover, C.L. (2003) Variation in community srtucture within hydrothermal vent mussel beds of the east pacific rise. *Marine Ecology Progress Series*, **253**, 55-66.

Van Dover, C.L. & Doerries, M.B. (2005) Community structure in mussel beds at logatchev hydrothermal vents and a comparison of macrofaunal species richness on slow- and fast-spreading mid-ocean ridges. *Marine Ecology*, **26**, 110-120.

Via, S. (2001) Sympatric speciation in animals: The ugly duckling grows up. *Trends in Ecology & Evolution*, **16**, 381-390.

Warén, A. & Bouchet, P. (2001) Gastropoda and monoplacophora from hydrothermal vents and seeps: New taxa and records. *Veliger*, **44**, 116-231

Warwick, R.M. & Clarke, K.R. (2001) Practical measures of marine biodiversity based on relatedness of species. *Oceanography and Marine Biology. An Annual Review*, **39**, 207-231.

Won, Y., Young, C.R., Lutz, R.A. & Vrijenhoek, R.C. (2003) Dispersal barriers and isolation among deep-sea mussel populations (mytilidae: *Bathymodiololus*) from eastern pacific hydrothermal vents. *Molecular Ecology*, **12**, 169-184.

Young, C.R., Fujio, S. & Vrijenhoek, R.C. (2008) Directional dispersal between mid-ocean ridges: Deep-ocean circulation and gene flow in ridgeia piscesae. *Molecular Ecology*, **17**, 1718-1731.

Références Bibliographiques

A

Adams DK, Mullineaux LS (2008) Supply of gastropod larvae to hydrothermal vents reflects transport from local larval sources. *Limnology and Oceanography* **53**, 1945-1955.

Akey JM, Eberle MA, Rieder MJ*, et al.* (2004) Population history and natural selection shape patterns of genetic variation in 132 genes. *Plos Biology* **2**, 1591-1599.

Arbogast BS, Kenagy GJ (2001) Comparative phylogeography as an integartive approach to historical biogeography. *Journal of Biogeography*, **28**, 819-825.

Arellano SM, Young CM (2009) Spawning, development and the duration of larval life in a deep-sea cold-seep mussel. *Biological Bulletin* **216**, 149-162.

Arnaud-Haond S, Bonhomme F, Blanc F (2003) Large discrepancies in differentiation of allozymes, nuclear and mitochondrial DNA loci in recently founded Pacific populations of the pearl oyster *Pinctada margaritifera. Journal of Evolutionary Biology* **16**, 388-398.

Auzende J-M, Ballu V, Batiza, R, Bideau D, Cormier M-H, Fouquet Y, Geistdoerfer P, Lagabrielle Y, Sinton J, Spadea P (1994) Activité magmatique, tectonique et hydrothermale actuelle sur la dorsale est pacifique entre 17°S et 19°S. (campagne naudur). *Comptes Rendus de l'Académie des Sciences, série II*, **319**, 811-818.

Audzijonyte A, Vrijenhoek RC (in press) When gaps realy are gaps : statistical phylogeography of hydrothermal vent invertebrate. *Evolution*.

Avise JC (1987) Identification and interpretation of mitochondrial DNA stocks in marine species. In : *Proceedings of stock the identification Workshop*. Kump L, Nakamura L (eds) National Oceanographic and Atmospheric Administation, Panama City, FL.

Avise JC (1992) Molecular population structure and the biogeographic history of regional fauna: a case history with lessons for conservation biology. *Oikos* **63**, 62-76.

Avise JC (1994) *Molecular markers, Natural History and Evolution*. New York: Chapman & Hall.

Avise JC (1998) The history and purview of phylogeography: a personal reflexion. *Molecular Ecology* **7**, 371-379.

Avise JC (2000) *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, MA, USA.

Avise JC, Arnold J, Ball RM*, et al.* (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* **18**, 489-522.

Avise JC, Ball RM (1990) *Principles of genealogical concordance in species concepts and biological taxonomy.* Oxford Survey of Evolutionnary Biology **7**, 45-67.

B

Bachraty C, Legendre P, Desbruyeres D (2009) Biogeographic relationships among deep-sea hydrothermal vent faunas at global scale. *Deep-Sea Research Part I-Oceanographic Research Papers* **56**, 1371-1378.

Baines JF, Chen Y, Das A, Stephan W (2002) DNA sequence variation at a duplicated gene: Excess of replacement polymorphism and extensive haplotype structure in

the Drosophila melanogaster bicoid region. *Molecular Biology and Evolution* **19**, 989-998.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37-48.

Barton N, Bengtsson BO (1986) The Barrier to Genetic Exchange between Hybridizing Populations. *Heredity* **57**, 357-376.

Barton NH (1979) Gene flow past a cline. *Heredity* **43**, 333-339.

Barton NH (1986) The Effects of Linkage and Density-Dependent Regulation on Gene Flow. *Heredity* **57**, 415-426.

Barton NH (2000) Estimating multilocus linkage disequilibria. *Heredity* **84**, 373-389.

Barton NH, Hewitt GM (1981) The Genetic-Basis of Hybrid Inviability in the Grasshopper Podisma-Pedestris. *Heredity* **47**, 367-383.

Barton NH, Hewitt GM (1985) Analysis of Hybrid Zones. *Annual Review of Ecology and Systematics* **16**, 113-148.

Barton NH, Hewitt GM (1989) Adaptation, speciation and hybrid zones. *Nature* **341**: 497-503.

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.

Bazin E, Glémin S, Galtier N (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science* **312**, 570-572.

Becker BJ, Fodrie FJ, McMillan PA, Levin LA (2005) Spatial and temporal variation in trace elemental fingerprints of mytilid mussel shells: A precursor to invertebrate larval tracking. *Limnology and Oceanography* **50**, 48-61.

Becker BJ, Levin LA, Fodrie FJ, McMillan PA (2007) Complex larval connectivity patterns among marine invertebrate populations. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 3267-3272.

Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* **152**: 763-773.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 4563-4568.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) Genetix 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, Adaptations, UMR 5000, Université de Montpellier 2. Available at: www.genetix.univ-montp2.fr/ genetix/genetix.htm

Bengtsson J (1989) Interspecific Competition Increases Local Extinction Rate in a Metapopulation System. *Nature* **340**, 713-715.

Bergquist DC, Fleckenstein C, Knisel J, Begley B, Macdonald IR, Fisher CR (2005) Variations in seep mussel bed communities along physical and chemical environmental gradients. *Marine Ecology Progress Series*, **293**, 99-108.

Bermingham E, Avise JC (1986) Molecular Zoogeography of Fresh-Water Fishes in the Southeastern United-States. *Genetics* **113**, 939-965.

Bermingham E, Martin AP (1998) Comparative mtDNA phylogeography of neotropical freshwater fishes: testing shared history to infer the evolutionary landscape of lower Central America. *Molecular Ecology* **7**, 499-517.

Bert TM, Arnold WS (1995) An Empirical-Test of Predictions of 2 Competing Models for the Maintenance and Fate of Hybrid Zones - Both Models Are Supported in a Hard-Clam Hybrid Zone. *Evolution* **49**, 276-289.

Bert TM, Harrison RG (1988) Hybridization in Western Atlantic Stone Crabs (Genus Menippe) - Evolutionary History and Ecological Context Influence Species Interactions. *Evolution* **42**, 528-544.

Bierne N, Borsa P, Daguin C*, et al.* (2003) Introgression patterns in the mosaic hybrid zone between Mytilus edulis and M-galloprovincialis. *Molecular Ecology* **12**, 447-461.

Bierne N, Tanguy A, Faure M*, et al.* (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*.

Birky CW (1995) Uniparental Inheritance of Mitochondrial and Chloroplast Genes - Mechanisms and Evolution. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 11331-11338.

Birky CW (2001) The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models. *Annual Review of Genetics* **35**, 125-148.

Black MB, Lutz RA, Vrijenhoek RC (1994) Gene Flow among Vestimentiferan Tube Worm (*Riftia pachyptila*) Populations from Hydrothermal Vents of the Eastern Pacific. *Marine Biology* **120**, 33-39.

Bodmer WF, Cavalli-Sforza LL (1968) A migration matrix model for the study of genetic drift. *Genetics* **59**, 565-592.

Bohrmann G, Chin C, Petersen S*, et al.* (1999) Hydrothermal activity at Hook Ridge in the Central Bransfield Basin, Antarctica. *Geo-Marine Letters* **18**, 277-284.

Boutet I, Tanguy A, Le Guen D, Piccino P, Hourdez S, Legendre P, Jollivet D (2009) Global depression in gene expression as a response to rapid thermal changes in vent mussels. *Proceedings of the Royal Society of London Series B*. doi:10.1098/rspb.2009.0503

Breton S, Beaupre HD, Stewart DT, Hoeh WR, Blier PU (2007) The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? *Trends in Genetics* **23**, 465-474.

Briggs JC (2006) Proximate sources of marine biodiversity. *Journal of Biogeography*, **33**, 1-10.

Briggs JC (2007) Marine biogeography and ecology: Invasions and introductions. *Journal of Biogeography*, **34**, 193-198.

Broughton RE, Harrison RG (2003) Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets. *Genetics* **163**, 1389-1401.

Burton RS (1998) Intraspecific phylogeography across the Point Conception biogeographic boundary. *Evolution* **52**, 734-745.

Burton RS, Lee BN (1994) Nuclear and Mitochondrial Gene Genealogies and Allozyme Polymorphism across a Major Phylogeographic Break in the Copepod Tigriopus-Californicus. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 5197-5201.

Byers JE, Pringle JM (2006) Going against the flow : retention, range limits and invasions in advective environments. *Marine Ecology Progress Series* **313**, 27-41.

C

Cannon GA, Pashinski DJ (1997) Variation in mean currents affecting hydrothermal plumes of the Juan de Fuca Ridge. *Journal of Geophysical Research* **102**: 24965-24976.

Cao LQ, Kenchington E, Zouros E (2004) Differential segregation patterns of sperm mitochondria in embryos of the blue mussel (Mytilus edulis). *Genetics* **166**, 883-894.

Carling MD, Brumfield RT (2007) Gene sampling strategies for multi-locus population estimates of genetic diversity (theta). *PLoS One* **2**, e160.

Carney SL, Formica MI, Divatia H, Nelson K, Fisher CR, Schaeffer SW (2006) Population structure of the mussel "*Bathymodiolus*" *childressi* from Gulf of Mexico hydrocarbon seeps. *Deep-Sea Research* I **53**: 1061-1072.

Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289-1303.

Chevaldonné P, Desbruyères D, Childress JJ (1992) Some Like 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.

Chevaldonné P, Jollivet D, Desbruyères D, Lutz RA, Vrijenhoek RC (2002) Sister-species of eastern Pacific hydrothermal vent worms (Ampharetidae, Alvinellidae, Vestimentifera) provide new mitochondrial COI clock calibration. *Cahiers De Biologie Marine* **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. An annual review* **30**, 337-441.

Clarke KR, Warwick RM (2001) *Change in marine communities: An approach to statistical analysis and interpretation*, 2nd edition PRIMER-E Ltd edn, Plymouth, UK.

Collin R (2001) The effects of mode of development on phylogeography and population structure of North Atlantic *Crepidula* (Gastropoda: Calyptraeidae). *Molecular Ecology* **10**, 2249-2262.

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, Le Pennec M, Desbruyeres D (1999) Evidence of a sexual pause in Bathymodiolus azoricus (Bivalvia : Mytilidae) from hydrothermal vents of the Mid-Atlantic Ridge. *Journal of the Marine Biological Association of the United Kingdom* **79**, 1149-1150.

Cormier M-H (1997) The ultrafast East Pacific Rise: instability of the plate boundary and implications for accretionary processes. *Philosophical Transactions of the Royal Society of London* **355**, 341-367.

Cowen JP, Fornari DJ, Shank TM, *et al.* (2007) Volcanic Eruptions at East Pacific Rise near 9°50'N. *EOS, Transactions of the American Geophysical Union* **88**, 81-83.

Cox BC (2001) The biogeographic regions reconsidered. *Journal of Biogeography* **28**, 511-523.

Cox CB, Moore PD (2000) Drawing lines in the water. In Biogeography : an ecological and evolutionary approach. Blackwell Science, London, 238-261.

Craddock C, Hoeh WR, Lutz RA, Vrijenhoek RC (1995) Extensive Gene Flow among Mytilid (*Bathymodiolus thermophilus*) Populations from Hydrothermal Vents of the Eastern Pacific. *Marine Biology* **124**, 137-146.

Craddock C, Lutz RA, Vrijenhoek RC (1997) Patterns of dispersal and larval development of archaeogastropod limpets at hydrothermal vents in the eastern Pacific. *Journal of Experimental Marine Biology and Ecology* **210**, 37-51.

Cunha RL, Castilho R, Ruber L, Zardoya R (2005) Patterns of cladogenesis in the venomous marine gastropod genus Conus from the Cape Verde islands. *Systematic Biology* **54**, 634-650.

Cunningham CW, Collins TM (1994) Developing model systems for molecular biogeography : vicariance and interchange in marine invertebrates. Wagner GP, DeSalle R (eds). In : Molecular ecology and evolution : approaches and applications. Birkhauser Velag, Basel, Switzerland.

Cunningham CW, Collins TM (1998) Beyond area relationships : extinction and recolonization in molecular marine biogeography. In : Molecular Approaches to Ecology and Evolution. DeSalle R, Schierwater B (eds), *Birkhäser Verlag Basel*, 297-321.

D

Dalziel AC, Stewart DT (2002) Tissue-specific expression of male-transmitted mitochondrial DNA and its implications for rates of molecular evolution in Mytilus mussels (Bivalvia : Mytilidae). *Genome* **45**, 348-355.

De'ath G (2002) Mulitvariate regression trees: A new technique for modeling species - environment relationships. *Ecology*, **83**, 1105-1117.

Desbruyères D, Alayse-Danet AM, Ohta S (1994) Deep-sea hydrothermal communities in southwestern Pacific back-arc basins (the North Fiji and Lau Basins): Composition, microdistribution and food web. *Marine Geology* **116**, 227-242.

Desbruyères D, Almeida A, Biscoito M*, et al.* (2000) Distribution of hydrothermal vent communities along the Northern Mid-Atlantic Ridge. Dispersal vs. Environmental control : a review. *Hydrobiologia* **440**, 201-216.

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

Desbruyères D, Hashimoto J, Fabri MC (2006a) Composition and biogeography of hydrothermal vent communities in western pacific back-arc basins. *Back-Arc Spreading Systems: Geological, Biological, Chemical, and Physical Interactions* **166**, 215-234.

Desbruyères D, Laubier L (1991) Systematics, Phylogeny, Ecology and Distribution of the Alvinellidae (Polychaeta) from Deep-Sea Hydrothermal Vents. *Ophelia*, 31-45.

Desbruyères D, Segonzac M, Bright M (2006b) Handbook of deep-sea hydrothermal vent fauna. Desbruyères D, Segonzac M, Bright M (eds), *Denisia 18*, 516-517.

Dixon DR, Lowe DM, Miller PI*, et al.* (2006) Evidence of seasonal reproduction in the Atlantic vent mussel Bathymodiolus azoricus, and an apparent link with the timing of photosynthetic primary production. *Journal of the Marine Biological Association of the United Kingdom* **86**, 1363-1371.

Donoghue MJ, Moore BR (2003) Toward an integrative historical biogeography. *Integrative and Comparative Biology* **43**, 261-270.

Dooh RT, Adamowicz SJ, Hebert PDN (2006) Comparative phylogeography of two North American 'glacial relict' crustaceans. *Molecular Ecology* **15**, 4459-4475.

Doyle JJ, Dickson E (1987) Preservation of plant samples for DNA restriction endonuclease analysis. *Taxon* **36,** 715–722.

Doyle RF (1985) Biogeographical studies of rocky shores near Point Conception, California. PhD dissertation, University of California, Santa Barbara, CA.

Dreyer JC, Knick KE, Flickinger WB, Van Dover CL (2005) Development of macrofaunal community structure in mussel beds on the northern east pacific rise. *Marine Ecology Progress Series*, **302.**

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7 **doi:10.1186/1471-2148-7-214**.

Duda TF, Palumbi SR (1999) Population structure of the black tiger prawn, Penaeus monodon, among western Indian Ocean and western Pacific populations. *Marine Biology* **134**, 705-710.

E

Eakins B, Lonsdale P (2003) Structural patterns and tectonic history of the Bauer microplate, Eastern Tropical Pacific. *Marine Geophysical Researches* **24**, 171-205.

Embley RW, Lupton JE, Massoth GJ, Urabe T, Tunnicliffe V, Butterfield DA, Shibata T, Okano O, Kinioshita M, Fujioka K (1998) Geological, chemical, and biological evidence for recent volcanism at 17.5°s: East pacific rise. *Earth and Planetary Science Letters*, **163**, 131-147.

Edwards SV, Beerli P (2000) Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* **54**, 1839-1854.

Emerson BC, Paradis E, Thebaud C (2001) Revealing the demographic histories of species using DNA sequences. *Trends in Ecology & Evolution* **16**, 707-716.

Emsley MG (1965) Speciation in *Helioconus* (Lep. Nymphalidae): morphology and geographic distribution. *Zoologica (NY)* **50**, 191-254.

Endler J (1982) Pleistocen forest refuges : fact or fancy. In Biological diversification in the Tropics. Prance GT (ed), Colombia University Press, New York, 641-657.

Engle VD, Summers JK (1999) Latitudinal gradients in benthic community composition in Western Atlantic estuaries. *Journal of Biogeography* **26**, 1007-1023.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* **1**, 47-50.

F

Faure B (2008) Etude multi-locus des pressions évolutives s'excerçant sur des bivalves marins en situation d'hybridation : cas des moules hydrothermales du genre *Bathymodiolus* des dorsales médio-atlantiques et est-pacifique. Thèse de Doctorat, Université Montpellier 2.

Faure B, Chevaldonné P, Pradillon F, Thiébaut E, Jollivet D (2007) Spatial and temporal dynamics of reproduction and settlement in the Pompeii worm *Alvinella pompejana* (Polychaeta: Alvinellidae). *Marine Ecology Progress Series* **348**, 197-211.

Faure B, Jollivet D, Tanguy A, Bonhomme F, Bierne N (2009) Speciation in the deep sea: multi-locus analysis of divergence and gene flow between two hybridizing species of hydrothermal vent mussels. *PLoS One* **4**, e6485.

Faure MF, David P, Bonhomme F, Bierne N (2008) Genetic hitchhiking in a subdivided population of Mytilus edulis. *Bmc Evolutionary Biology* **8**.

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinean selection. *Genetics*, **155**, 1405-1413.

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* **17**, 368-376.

Felsenstein J (2006) Accuracy of coalescent likelihood estimates: Do we need more sites, more sequences, or more loci? *Molecular Biology and Evolution* **23**, 691-700.

Feral JP (2002) How useful are the genetic markers in attempts to understand and manage marine biodiversity? *Journal of Experimental Marine Biology and Ecology* **268**, 121-145.

Fialamedioni A, Metivier C, Herry A, Lepennec M (1986) Ultrastructure of the Gill of the Hydrothermal-Vent Mytilid Bathymodiolus Sp. *Marine Biology* **92**, 65-72.

Filatov DA, Charlesworth D (1999) DNA polymorphism, haplotype structure and balancing selection in the leavenworthia PgiC locus. *Genetics* **153**, 1423-1434.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3**, 294-299.

Fornari DJ, Embley RW (1995) Tectonic and volcanic controls on hydrothermal processes at the mid-ocean ridge: An overview based on near-bottom and submersible studies. *Seafloor hydrothermal systems* (ed. by S.E. Humphris, R.A. Zierenberg, L.S. Mullineaux and R.E. Thomson). American Geophysical Union, Washington DC.

Fouquet Y, Auzende JM, Ballu V*, et al.* (1994) Variability of the Present-Day Hydrothermal Activity Along an Ultra-Fast Spreading Ridge - East Pacific Rise between 17-Degrees-C and 19-Degrees-S (Naudur Cruise). *Comptes Rendus De L Academie Des Sciences Serie Ii* **319**, 1399-1406.

France SC, Hessler RR, Vrijenhoek RC (1992) Genetic differentiation between spatially-disjunct populations of the deep-sea, hydrothermal vent-endemic amphipod *Ventiella sulfuris*. *Marine Biology* **114**, 551-559.

Francheteau J, Armijo R, Cheminee JL*, et al* (1990) 1 Ma East Pacific Rise Oceanic-Crust and Uppermost Mantle Exposed by Rifting in Hess Deep (Equatorial Pacific-Ocean). *Earth and Planetary Science Letters* **101**: 281-295.

Francheteau J, Armijo R, Cheminee JL*, et al.* (1990) 1 Ma East Pacific Rise Oceanic-Crust and Uppermost Mantle Exposed by Rifting in Hess Deep (Equatorial Pacific-Ocean). *Earth and Planetary Science Letters* **101**, 281-295.

Fretter V (1988) New archaeogastropod limpets from hydrothermal vents; superfamily *Lepetodrilacea*. II. Anatomy. *Philosophical Transactions of the Royal Society of London Series B* **318**, 33-82.

Fu YX, Li WH (1993) Statistical test of neutrality of mutations. *Genetics* **133**: 693-709.

Fujio SZ, Imasato N (1991) Diagnostic Calculation for Circulation and Water Mass Movement in the Deep Pacific. *Journal of Geophysical Research-Oceans* **96**, 759-774.

Fujioka K, Massoth GJ, De Ronde C, Tanaka-Nakano A, Gendron J, Kinoshita M, Nishizawa A, Okano O, Paradis G, Roe KJ, Shinohara M, Shitashima K, Tsunogai U, Ytow N (1995) Active hydrothermal systems of a super-fast spreading ridge, southern east pacific rise (13°58's to 18°26s). *InterRidge News*, **4**, 7-10.

Fukushima T (2000) Deep sea benthology : history and present state. MBRIJ Annual Report, 74-82.

Fustec A, Desbruyères D, Juniper SK (1987) Genetic differentiation between spatially-disjunct populations of the deep-sea, hydrothermal vent-endemic amphipod *Ventiella sulfuris*. *Biological Oceanography* **4**, 121-164.

G

Gagnière N, Jollivet D, Boutet I, *et al* (submitted) Insights metazoan evolution from Alvinella pompejana cDNAs. *Genome Research*.

Galtier N, Depaulis F, Barton NH (2000) Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. *Genetics* **155**, 981-987.

Garnier-Géré P, Dillmann C (1992) A Computer-Program for Testing Pairwise Linkage Disequilibria in Subdivided Populations. *Journal of Heredity* **83**, 239-239.

Geiger DL, Thacker CE (2005) Molecular phylogeny of vetigastropoda reveals non-monophyletic scissurellidae, trochoidea, and fissurelloidea. *Molluscan Research*, **25**, 47-55.

Geistdoerfer P, Auzende J-M, Batiza R, Bideau D, Cormier M-H, Fouquet Y, Lagabrielle Y, Sinton J, Spadea P (1995) Hydrothermalisme et communautés animales associées sur la dorsale du pacifique oriental entre 17°s et 19°s (campagne naudur, décembre 1993). *Comptes Rendus de l'Académie des Sciences, série II*, **320**, 47-54.

Gilpin M (1991) The genetic effective size of a metapopulation. *Biological Journal of the Linnean Society* **42**, 165-175.

Glinka S, Ometto L, Mousset S, Stephan W, De Lorenzo D (2003) Demography and natural selection have shaped genetic variation in Drosophila melanogaster: A multi-locus approach. *Genetics* **165**, 1269-1278.

Goldstien SJ, Schiel DR, Gemmell NJ (2006) Comparative phylogeography of coastal limpets across a marine disjunction in New Zealand. *Molecular Ecology* **15**, 3259-3268.

Goudet (1999) PCA-GEN, a program to perform principal component analysis (PCA) on genetic data, Version 1. 2.

Graham JH, Freeman DC, McArthur ED (1995) Narrow Hybrid Zone between 2 Subspecies of Big Sagebrush (Artemisia-Tridentata, Asteraceae) .2. Selection Gradients and Hybrid Fitness. *American Journal of Botany* **82**, 709-716.

Grassle JP (1985) Genetic differentiation in populations of hydrothermal vent mussels (*Bathymodiolus thermophilus*) from the Galapagos Rift and 13°N on the East Pacific Rise. *Biological Society of Washington* **6**, 429-442.

Graur D, Li WH (1991) Neutral Mutation Hypothesis Test. *Nature* **354**, 114-115.

Griffiths RC, Tavare S (1994a) Sampling Theory for Neutral Alleles in a Varying Environment. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **344**, 403-410.

Griffiths RC, Tavare S (1994b) Simulating Probability-Distributions in the Coalescent. *Theoretical Population Biology* **46**, 131-159.

H

Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P (2005) Multilocus patterns of nucleotide variability and the demographic and selection history of Drosophila melanogaster populations. *Genome Research* **15**, 790-799.

Haffer J (1969) Speciation in Amozonian forest birds. *Science* **165**, 131-136.

Halanych KM, Tieger M, O'mullan GD, Lutz RA, Vrijenhoek RC (1999) Brief description of biological communities at 7°s on the east pacific rise. *International Ridge-Crest Research: Biological studies* **8**, 23-27.

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.

Hannington MD, Jonasson IR (1995) Physical and chemical processes of seafloor mineralization at mid-ocean ridges. In Seafloor Hydrothermal Systems. Humphris, SE, Zierenberg, RA, Mullineaux, LS, Thompson, RE (eds). Geophysical Monograph. American Geophysical Society Washington, 91: 115-157.

Harpending HC, Batzer MA, Gurven M*, et al.* (1998) Genetic traces of ancient demography. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 1961-1967.

Harr B, Kauer M, Schlotterer C (2002) Hitchhiking mapping: A population-based fine-mapping strategy for adaptive mutations in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 12949-12954.

Harrison RG (1990) Hybrid zones: windows on evolutionary processes. *Oxf. Surv. Evol. Biol.* **7**: 69-128.

Harrison S, Hastings A (1996) Genetic and evolutionary consequences of metapopulation structure. *Trends in Ecology & Evolution* **11**, 180-183.

Hartl DL, Clark AG (2007) Inbreeding, Population, and Migration. In Principes of population genetics, Sinauer AD (ed), Sinauer, 257-315.

Hasegawa M, Kishino H, Yano T (1985) Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**: 160-174.

Hashimoto J, Ohta S, Gamo T, *et al* (2001) Hydrothermal vents and associated biological communities in the Indian Ocean. *InterRidge News* **10** : 21-22.

Hawksworth DL, Kalin-Arroyo MT (1995) Magnitude and distribution of biodiversity. *Global biodiversity assessment* (ed. by V.H. Heywood), pp. 107-192. University Press, Cambridge.

Haymon RM, Fornari DJ, Edwards MH*, et al.* (1991) Hydrothermal Vent Distribution Along the East Pacific Rise Crest (9-Degrees-09'-54'n) and Its Relationship to Magmatic and Tectonic Processes on Fast-Spreading Midocean Ridges. *Earth and Planetary Science Letters* **104**, 513-534.

Haymon RM, Fornari DJ, Vondamm KL*, et al.* (1993) Volcanic-Eruption of the Mid-ocean Ridge Along the East Pacific Rise Crest at 9-Degrees-45-52'n - Direct Submersible Observations of Sea-Floor Phenomena Associated with an Eruption Event in April, 1991. *Earth and Planetary Science Letters* **119**, 85-101.

Heads M (2005) Towards a panbiogeography of the seas. *Biological Journal of the Linnean Society*, **84**, 675-723.

Hessler RR, Lonsdale PF (1991) Biogeography of Mariana Trough hydrothermal vents. *Deep Sea Research* **38**, 185-199.

Hessler RR, Smithey Jr WM, 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.

Hessler RR, Smithey WM (1983) The distribution and community structure of megafauna at the Galapagos Rift hydrothermal vents. In hydrothermal processes at seafloor

spreading centres. Rona PA, Bostrom K, Laubier L, Smith KL Jr (eds), Plenum, New York.
Hewitt GM (2001) Speciation, hybrid zones and phylogeography – or seeing genes in space and time. *Molecular Ecology* **10**, 537-549.
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. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 2785-2790.
Hey R, Baker E, Bohnenstiehl D*, et al.* (2004) Tectonic/volcanic segmentation and controls on hydrothermal venting along Earth's fastest seafloor spreading system, EPR 27 degrees-32 degrees S. *Geochemistry Geophysics Geosystems* **5**.
Hickerson MJ, Dolman G, Moritz C (2006a) Comparative phylogeographic summary statistics for testing simultaneous vicariance. *Molecular Ecology* **15**, 209-223.
Hickerson MJ, Meyer CP (2008) Testing comparative phylogeographic models of marine vicariance and dispersal using a hierarchical Bayesian approach. *Bmc Evolutionary Biology* **8**.
Hickerson MJ, Stahl E, Takebayashi N (2007) msBayes: Pipeline for testing comparative phylogeographic histories using hierarchical approximate Bayesian computation. *Bmc Bioinformatics* **8**.
Hickerson MJ, Stahl EA, Lessios HA (2006b) Test for simultaneous divergence using approximate Bayesian computation. *Evolution* **60**, 2435-2453.
Ho SYW (2009) An examination of phylogenetic models of substitution rate variation among lineages. *Biology letters*. doi:10.1098/rsbl.2008.0729
Hoelzer GA (1997) Inferring phylogenies from mtDNA variation: Mitochondrial-gene trees versus nuclear-gene trees revisited. *Evolution* **51**, 622-626.
Howell N, Bogolin Smejkal C, Mackey DA, Chinnery PF, Turnbull DM, Herrnstadt C (2003) The pedigree rate of sequence divergence in the human mitochondrial genome: there is a différence between phylogenetic and pedigree rates. *The American Journal of Human Genetics* **72**, 659-670.
Hudson RR (1983) Testing the constant-rate neutral allele model with protein sequence data. *Evolution* **37**, 203-217.
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, Aguade M (1987) A Test of Neutral Molecular Evolution Based on Nucleotide Data. *Genetics* **116**, 153-159.
Hudson RR, Slatkin M (1991) Pairwise comparisons of mitochondrial-DNA sequences in stable and exponentially growing populations. *Genetics* **129**, 555-562.
Hudson RR, Slatkin M, Maddison WP (1992) Estimation of Levels of Gene Flow from DNA-Sequence Data. *Genetics* **132**, 583-589.
Hudson RR, Turelli M (2003) Stochasticity overrules the "three-times rule": Genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution* **57**, 182-190.
Hurtado L, Lutz R, Vrijenhoek R (2004) Distinct patterns of genetic differentiation among annelids of eastern Pacific hydrothermal vents. *Molecular Ecology* **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. *Cahier de Biologie Marine*, **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. *Applied and Environmental Microbiology* **69**, 2058-2064.

J

Jennings WB, Edwards SV (2005) Speciational history of Australian grass finches (Poephila) inferred from thirty gene trees. *Evolution* **59**, 2033-2047.

Johnson SB, Waren A, Vrijenhoek RC (2008) DNA barcoding of Lepetodrilus limpets reveals cryptic species. *Journal of Shellfish Research* **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 *Biological Bulletin* **210**, 140-157.

Jollivet D, Dixon LRJ, Desbruyères D, Diwon DR (1998) 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 United Kingdom*, **78**, 113-130.

Jollivet D, Party TSS (2005) The biospeedo cruise: A new survey of hydrothermal vents along the south east pacific rise from 7°24s to 21°33s. *InterRidge News*, **13**, 20-26.

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. Thèse de doctorat, Université de Bretagne Occidentale, Brest.

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, Desbruyères D, Bonhomme F, Moraga D (1995) Genetic Differentiation of Deep-Sea Hydrothermal Vent Alvinellid Populations (Annelida, Polychaeta) Along the East Pacific Rise. *Heredity* **74**, 376-391.

Jollivet D, Empis A, Baker MC*, et al.* (2000) Reproductive biology, sexual dimorphism, and population structure of the deep sea hydrothermal vent scale-worm, *Branchipolynoe seepensis* (Polychaeta : Polynoidae). *Journal of the Marine Biological Association of the United Kingdom* **80**, 55-68.

Jolly MT, Viard F, Gentil F, Thiebaut E, Jollivet D (2006) Comparative phylogeography of two coastal polychaete tubeworms in the Northeast Atlantic supports shared history and vicariant events. *Molecular Ecology* **15**, 1841-1855.

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: *Mammalian Protein Metabolism* (ed. Munro HN), pp 21-123. Academic Press, New York.

Juniper SK, Tunnicliffe V (1997) Crustal accretion and the hot vent ecosystem. *Philosophical Transactions of the Royal Society of London. Series A*, **355**, 459-474.

K

Karl SA, Schutz S, Desbruyeres D, Lutz R, Vrijenhoek RC (1996) Molecular analysis of gene flow in the hydrothermal vent clam (Calyptogena magnifica). *Molecular Marine Biology and Biotechnology* **5**, 193-202.

Kelly NE, Metaxas A (2007) Influence of habitat on the reproductive biology of the deep-sea hydrothermal vent limpet Lepetodrilus fucensis (Vetigastropoda : Mollusca) from the Northeast Pacific. *Marine Biology* **151**, 649-662.

Kervévan (1991) Modélisation tridimentionnelle de la convection hydrothermale en zone de rift océanique. Ph.D. dissertation, Université de Bretagne Occidentale, Brest, France.

Kimura M (1953) "Stepping stone" model of population. *Annual Report Nationnal Institute of Genetics. Japan* **3**, 62-63.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **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.

Knowlton N (1993) Sibling species in the sea. *Ann. Rev. Ecol. Syst.*, **24**, 189-216.

Knowlton N (2000) Molecular genetic analyses of species boundaries in the sea. *Hydrobiologia* **420**, 73-90.

Knowlton N, Weigt LA (1998) New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society of London Series B-Biological Sciences* **265**, 2257-2263.

Kojima S, Fujikura K, Okutani T (2004) Multiple trans-Pacific migrations of deep-sea vent/seep-endemic bivalves in the family Vesicomyidae. *Molecular Phylogenetics and Evolution* **32**, 396-406.

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.

Kuhner MK, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* **149**, 429-434.

Kumar S, Balczarek KA, Lai Z-C (1996) Evolution of the *hedgehog* gene family. *Genetics* **142**, 965-972.

Kuo CH, Avise J (2005) Phylogeographic breaks in low-dispersal species: the emergence of concordance across gene trees. *Genetica* **124**, 179-186.

Kureth CL, Rea DK (1981) Large-Scale Oblique Features in an Active Transform-Fault, the Wilkes Fracture-Zone near 9-Degrees-S on the East Pacific Rise. *Marine Geophysical Researches* **5**, 119-137.

L

Larson RL, Searle RC, Kleinrock MC*, et al.* (1992) Roller-bearing tectonic evolution of the Juan Fernandez microplate. *Nature* **356** 571-575.

Laubier L (1986) Des oasis au fond des mers. Monaco: Le Rocher (Sciences et Découvertes).

Lavery L, Moritz C, Fielder DR (1995) Changing patterns of population structure and gene flow at different spatial scales in *Birgus latro* (the coconut crab). *Heredity* **74**, 531-541.

Leberg PL (1992) Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis. *Evolution* **44**, 477-494.

Le Bris N, Govenar BW, Le Gall C, Fisher CR (2006) Variability of physico-chemical conditions in 9°50'n epr diffuse flow vent habitats. *Marine Chemistry*, **98**, 167-182.

Le Pennec M (1988) Alimentation et reproduction d'un Mytilidae des sources hydrothermales profondes du Pacifique oriental. *Oceanologica Acta* **8**, 181-190.

Le Pennec M, Beninger PG (1997) Ultrastructural characteristics of spermatogenesis in three species of deep-sea hydrothermal vent mytilids. *Canadian Journal of Zoology-Revue Canadienne De Zoologie* **75**, 308-316.

Le Pennec M, Hily A, Lucas A (1984) Structures gonadiques particulières d'un Mytilidae profound des sources hydrothermales du Pacifique oriental. *Comptes Rendus de l'Académie des Sciences de Paris, Série III* **299**, 725-730.

Legendre P, Legendre L (1998) *Numerical ecology*. Elsevier.

Lessa EP, Cook JA, Patton JL (2003) Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary *Proceedings of the National Academy of Sciences of the United States of America* **100**, 10331-10334.

Levin LA, Etter RJ, Rex MA, Gooday J, Smith CR, Pineda J, Stuart CT, Hessler RR, Pawson D (2001) Environmental influences on regional deep-sea species diversity. *Annual Review of Ecology and Systematics* **32**, 51-93.

Levin R (1970) Extinction. *Lectures on Mathematics in the Life Sciences* **2**, 75-107.

Lewontin RC, Krakauer J (1973) Distribution of Gene Frequency as a Test of Theory of Selective Neutrality of Polymorphisms. *Genetics* **74**, 175-195.

Li W-H, Graur D (1991) Fundamentals of molecular evolution. Sinauer Associates, Inc. Sunderland, Massachusetts.

Little CTS, Vrijenhoek RC (2003) Are hydrothermal vent animals living fossils? *Trends in ecology and evolution*, **18**, 582-588.

Liu HP, Mitton JB, Wu SK (1996) Paternal mitochondrial DNA differentiation far exceeds maternal mitochondrial DNA and allozyme differentiation in the freshwater mussel, Anodonta grandis grandis. *Evolution* **50**, 952-957.

Liu LL, Foltz DW, Stickle WB (1991) Genetic Population-Structure of the Southern Oyster Drill Stramonita (= Thais) Haemostoma. *Marine Biology* **111**, 71-79.

Lonsdale P (1977) Clustering of suspension-feeding macorbenthos near abyssal hydrothermal vents at oceanic spreading centres. *Deep-sea Research* **24**, 857-863.

Lorion J, Duperron S, Gros O, Cruaud C, Samadi S (2009) Several deep-sea mussels and their associated symbionts are able to live both on wood and on whale falls. *Proceedings of the Royal Society B-Biological Sciences* **276**, 177-185.

Lupton J (1998) Hydrothermal helium plumes in the Pacific Ocean. *Journal of Geophysical Research-Oceans* **103**, 15853-15868.

Lupton JE, Craig H (1981) A Major He-3 Source at 15-Degrees-S on the East Pacific Rise. *Science* **214**, 13-18.

Lutz RA (1988) Dispersal of organisms at deep-sea hydrothermal vents : A review. *Oceanologica Acta*. Hydrothermalism, Biology and Ecology Symposium, Paris, 4-7 November 1985, Proceedings 23-29.

Lutz RA, Fritz LW, Cerrato RM (1988) A Comparison of Bivalve (Calyptogena-Magnifica) Growth at 2 Deep-Sea Hydrothermal Vents in the Eastern Pacific. *Deep-Sea Research Part a-Oceanographic Research Papers* **35**, 1793

Lutz RA, Jablonski D, Rhoads DC, Turner RD (1980) Larval Dispersal of a Deep-Sea Hydrothermal Vent Bivalve from the Galapagos Rift. *Marine Biology* **57**, 127-133.

Lutz RA, Shank TM, Fornari DJ, *et al* (1994) Rapid growth at deep-sea vents. *Nature* **371**, 663-664.

M

Macdonald KC, Scheirer DS, Carbotte SM (1991) Mid-Ocean Ridges: Discontinuities, Segments and Giant Cracks. *Science* **253**, 986-994.

Maddison WP (1997) Gene trees in species trees. *Systematic Biology* **46**, 523-536.

Magde LS, Sparks DW (1997) Three-dimensional mantle upwelling, melt generation and melt migration beneath segmented slow-spreading ridges. *Journal of Geophysical Research* **102**, 20571–20583.

Maggs CA, Castilho R, Foltz D*, et al.* (2008) Evaluating signatures of glacial refugia for north atlantic benthic marine taxa. *Ecology* **89**, 108-122.

Malécot G (1950) Quelques schémas probabilistes sur la variabilité des populations naturelles. *Annales de l'Université de Lyon Sciences* **13**, 37-60.

Mammerickx J, Herron E, Dorman L (1980) Evidence for two fossil spreading ridges in the southeast Pacific. *Geological Society of America Bulletin* **91**, 263-271.

Mammerickx J, Klitgord KD (1982) Northern East Pacific Rise - Evolution from 25 My Bp to the Present. *Journal of Geophysical Research* **87**, 6751-6759.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, 209-220.

Marcus J, Tunnicliffe V, Butterfield DA (2009) Post-eruption succession of macrofaunal communities at diffuse flow hydrothermal vents on Axial Volcano, Juan de Fuca Ridge, Northeast Pacific. *Deep-Sea Research Part Ii-Topical Studies in Oceanography* **56**, 1586-1598.

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.

Martinsen GD, Whitham TG, Turek RJ, Keim P (2001) Hybrid populations selectively filter gene introgression between species. *Evolution* **55**, 1325-1335.

Matabos M, Le Bris N, Pendlebury S, Thiebaut E (2008a) Role of physico-chemical environment on gastropod assemblages at hydrothermal vents on the East Pacific Rise (13 degrees N/EPR). *Journal of the Marine Biological Association of the United Kingdom* **88**, 995-1008.

Matabos M, Thiebaut E, Le Guen D*, et al.* (2008b) Geographic clines and stepping-stone patterns detected along the EPR in the vetigastropod *Lepetodrilus elevatus* reflects species crypticism. *Marine Biology* **153**, 545-563.

McDonald JH, Kreitman M (1991b) Adaptive Protein Evolution at the Adh Locus in *Drosophila*. *Nature* **351**, 652-654.

McDonald JH, Kreitman M (1991a) Neutral Mutation Hypothesis Test - Reply. *Nature* **354**, 116-116.

McLean JH (1988) New Archaeogastropod limpets from hydrothermal vents; superfamily *Lepetodrilacea* I. systematic descriptions. *Philosophical Transactions of the Royal Society of London* **319**, 1-32.

Metaxas A (2004) Spatial and temporal patterns in larval supply at hydrothermal vents in the northeast Pacific Ocean. *Limnology and Oceanography* **49**, 1949-1956.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics* **21**, 1087-1092.

Micheli F, Peterson CH, Mullineaux LS, Fisher CR, Mills SW, Sancho G, Johnson GA, Lenihan HS (2002) Predation structures communties at deep-sea hydrothermal vents. *Ecological monographs*, **72**, 365-382.

Mills SW, Mullineaux LS, Tyler PA (2007) Habitat associations in gastropod species at east Pacific rise hydrothermal vents (9 degrees 50 ' N). *Biological Bulletin* **212**, 185-194.

Monge-Najera J (2008) Ecological biogeography: A review with emphasis on conservation and the neutral model. *Gayana*, **72**, 102-112.

Moore WS (1995) Inferring Phylogenies from Mtdna Variation - Mitochondrial-Gene Trees Versus Nuclear-Gene Trees. *Evolution* **49**, 718-726.

Moritz C, Faith DP (1998) Comparative phylogeography and the identification of genetically divergent areas for conservation. *Molecular Ecology* **7**, 419-429.

Morjan CL, Rieseberg LH (2004) How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology* **13**, 1341-1356.

Mullineaux LS, Fisher CR, Peterson CH, Schaeffer SW (2000) Tubeworm succession at hydrothermal vents: Use of biogenic cues to reduce habitat selection error? *Oecologia*, **123**, 275-284Mullineaux, L.S., Mills, S.W. & Goldman, E. (1998) Recruitment variation during a pilot colonization study of hydrothermal vents (9°50'n, east pacific rise). *Deep Sea Research II*, **45**, 441-464.

Mullineaux LS, Kim SL, Pooley A, Lutz RA (1996) Identification of archaeogastropod larvae from a hydrothermal vent community. *Marine Biology* **124**, 551-560.

Mullineaux LS, Mills SW, Goldman E (1998) Recruitment variation during a pilot colonization study of hydrothermal vents (9 degrees 50 ' N, East Pacific Rise). *Deep-Sea Research Part Ii-Topical Studies in Oceanography* **45**, 441-464.

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, Peterson CH, Micheli F, Mills SW (2003) Successional mechanism varies along a gradient in hydrothermal fluid flux at deep-sea vents. *Ecological monographs*, **73**, 523-542.

Mullineaux LS, Wiebe PH, Baker ET (1995) Larvae of benthic invertebrates in hydrothermal vent plumes over Juan-De-Fuca Ridge. *Marine Biology* **122**, 585-596.

Myers AA, Giller PS (1988) Analytical Biogeography : an integrated approach to the study of animal and plant distributions. Chapman & Hall, London.

N

Naar DF, Hey RN (1989) Speed Limit for Oceanic Transform Faults. *Geology* **17**, 420-422.

Naar DF, Hey RN (1991) Tectonic evolution of the Easter microplate. *Journal of Geophysical Research* **96**, 7961–7993.

Nei M, Kumar S (2001) Molecular evolution and phylogenetics. Oxford University Press, UK.

Neigel JE, Avise JC (1986) Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. Karling S & Nevo E (eds). In: *Evolutionary Process and Theory*. Academic Press, New York.

Neigel JE, Avise JC (1993) Application of a Random-Walk Model to Geographic Distributions of Animal Mitochondrial-DNA Variation. *Genetics* **135**, 1209-1220.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: A Markov chain Monte Carlo approach. *Genetics* **158**, 885-896.

Nosil P, Harmon LJ, Seehausen O (2009) Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution* **24**, 145-156.

O

Olu-Le Roy L, 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.

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

P

Page RDM, Holmes EC (1998). Measuring genetic change, In Molecular evolution : a phylogenetic approach. Blackwell Science, London, 153.

Palumbi SR (1994) Genetic-Divergence, Reproductive Isolation, and Marine Speciation. *Annual Review of Ecology and Systematics* **25**, 547-572.

Palumbi SR (1996) What can molecular genetics contribute to marine biogeography? *Journal of Experimental Marine Biology and Ecology*, **203**, 75-92.

Palumbi SR (2003) Population genetic, demographic connectivity, and the design of marine reserves. *Ecological Applications* **13**, 146-158.

Pamilo P, Nei M (1988) Relationships between Gene Trees and Species Trees. *Molecular Biology and Evolution* **5**, 568-583.

Pamilo P, Viljakainen L, Vihavainen A (2007) Exceptionally high density of NUMTs in the honeybee genome. *Molecular Biology and Evolution* **24**, 1340-1346.

Pannell JR, Charlesworth B (2000) Effects of metapopulation process on measures of genetic diversity. *Philosophical Transactions of the Royal Society of London* **355**, 1851-1864.

Passamonti M, Scali V (2001) Gender-associated mitochondrial DNA heteroplasmy in the venerid clam Tapes philippinarum (Mollusca Bivalvia). *Current Genetics* **39**, 117-124.

Pasteur N, Pasteur G, Bonhomme F, Catalan J, Britton-Davidian J (1987) *Manuel technique de génétique par électrophorèse de protéines*. Paris: Lavoisier.

Pechenik JA (1999) On the advantages ans disadvantages of larval stages in benthic marine invertebrate life cycles. *Marine Ecology Progress Series* **177**, 269-297.

Peek AS, Gustafson RG, Lutz RA, Vrijenhoek RC (1997) Evolutionary relationships of deep-sea hydrothermal vent and cold-water seep clams (bivalvia: Vesicomyidae): Results from the mitochondrial cytochrome oxidase subunit i. *Marine Biology*, **130**, 151-161.

Peters JL, Zhuravlev Y, Fefelov I, Logie A, Omland KE (2007) Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (Anas spp.). *Evolution* **61**, 1992-2006.

Piálek J, Barton NH (1997) The spread of an advantageous allele across a barrier: The effects of random drift and selection against heterozygotes. *Genetics* **145**, 493-504.

Piccino P, Viard F, Sarradin PM*, et al* (2004) Thermal selection of PGM allozymes in newly founded populations of the thermotolerant vent polychaete Alvinella

pompejana. *Proceedings of the Royal Society of London Series B-Biological Sciences* **271**, 2351-2359.

Plouviez S (2006) Divergence et polymorphisme nucléotidiques chez les Lepetodrilidae, complexe d'espèces hydrothermales partiellement sympatriques inféodées à des microhabitats différents. Rapport de Master 2. Université Montpellier 2.

Plouviez S, Daguin C, Hourdez S, Jollivet D (2008) Juvenile and adult scale-worms, *Branchipolynoe seepensis*, in Lucky Strike hydrothermal vent mussels are genetically unrelated. *Aquatic Biology* **3**, 79-87.

Plouviez S, Le Guen D, Lecompte O, Lallier FH, Jollivet D (2010) Determining gene flow and influence of selection across the equatorial barrier of the East Pacific Rise in the tube-dwelling polychaete *Alvinella pompejana*. *BMC Evolutionary Biology* **10**, 220.

Plouviez S, Shank TM, Faure B, Daguin-Thiébaut C, Viard F, Lallier FH, Jollivet D (2009) Comparative phylogeography among hydrothermal vent species along the East Pacific Rise reveals vicariant processes and population expansion in the South. *Molecular Ecology* **18**: 3903-3917.

Pluzhnikov A, Donnelly P (1996) Optimal sequencing strategies for surveying molecular genetic diversity. *Genetics* **144**, 1247-1262.

Pogson GH, Mesa KA, Boutilier RG (1995) Genetic Population-Structure and Gene Flow in the Atlantic Cod Gadus-Morhua - a Comparison of Allozyme and Nuclear Rflp Loci. *Genetics* **139**, 375-385.

Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology & Evolution* **16**, 37-45.

Pradillon F, Gaill F (2003) Oogenesis characteristics in the hydrothermal vent polychaete Alvinella pompejana. *Invertebrate Reproduction & Development* **43**, 223-235.

Pradillon F, Le Bris N, Shillito B, Young CM, Gaill F (2005) Influence of environmental conditions on early development of the hydrothermal vent polychaete Alvinella pompejana. *Journal of Experimental Biology* **208**, 1551-1561.

Pradillon F, Shillito B, Young CM, Gaill F (2001) Deep-seaecology - Developmental arrest in vent worm embryos. *Nature* **413**, 698-699.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

Przeworski M, Wall JD, Andolfatto P (2001) Recombination and the frequency spectrum in Drosophila melanogaster and Drosophila simulans. *Molecular Biology and Evolution* **18**, 291-298.

R

Rambaut A, Bromham L (1998) Estimating divergence dates from molecular sequences. *Molecular Biology and Evolution* **15**, 442-448.

Rand DM (2001) The units of selection on mitochondrial DNA. *Annual Review of Ecology and Systematics* **32**, 415-448.

Raulfs EC, Macko SA, Van Dover CL (2004) Tissue and symbiont condition of mussels (*Bathymodiolus thermophilus*) exposed to varying levels of hydrothermal activity. *Journal of the Marine Biological Association of the United Kingdom* **84**, 229-234.

Rawson PD, Hilbish TJ (1998) Assymetric introgression of female and male lineage mitochondrial DNA (mtDNA) haplotypes within the European hybrid zone between *Mytilus edulis* and *Mytilus galloprovincialis*. *Evolution* **52**, 100-108.

Raymond M, Rousset F (1995) GENEPOP, version 1.2: population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**: 248-249.

Reid JL (1997) On the total geostrophic circulation of the South-Pacific ocean: flow patterns, tracers, and transports. *Progress in Oceanography* **39**, 263-352.

Reid RA (1980) Selfish DNA in "petite" mutants. *Nature* **285**, 620.

Renard V, Hekinian R, Francheteau J, Ballard RD, Backer H (1985) Submersible Observations at the Axis of the Ultra-Fast-Spreading East Pacific Rise (17-Degrees-' to 21-Degrees-30's). *Earth and Planetary Science Letters* **75**, 339-353.

Robertson A, Hill WG (1984) Deviations from Hardy-Weinberg proportions: sampling variances and use in estimation of inbreeding coefficients. *Genetics*, **107**, 703-718.

Rodríguez F, Olivier JL, Marín A, Medina JR (1990) The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology* **142**, 485-501.

Rolan-alvarez E, Zapata C, Alvarez G (1995) Distinct genetic subdivision in sympatric and sibling species of the genus *Littorina* (Gastropoda, Littorinidae). *Heredity* **74**, 1-9.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496-2497.

Rusby RI, Searle RC (1995) A History of the Easter Microplate, 5.25-Ma to Present. *Journal of Geophysical Research-Solid Earth* **100**, 12617-12640.

S

Sadosky F, Thiebaut E, Jollivet D, Shillito B (2002) Recruitment and population structure of the vetigastropod Lepetodrilus elevatus at 13 degrees N hydrothermal vent sites on East Pacific Rise. *Cahiers de Biologie Marine* **43**, 399-402.

Samadi S, Quemere E, Lorion J*, et al.* (2007) Molecular phylogeny in mytilids supports the wooden steps to deep-sea vents hypothesis. *Comptes Rendus Biologies* **330**, 446-456.

Sancho G, Fisher CR, Mills SW, Micheli F, Johnson GA, Lenihan HS, Peterson CH, Mullineaux LS (2005) Selective predation by the zoarcid fish *thermarces cerberus* at hydrothermal vents. *Deep Sea Research I*, **52**, 837-844.

Sarrazin J, Juniper SK (1999) Biological characteristics of a hydrothermal edifice mosaic community. *Marine Ecology Progress Series*, **185**, 1-19.

Sarrazin J, Juniper SK, Massoth GJ, Legendre P (1999) Physical and chemical factors influencing species distributions on hydrothermal sulfide edifices of the juan de fuca ridge, northeast pacific. *Marine Ecology Progress Series*, **190**, 89-112

Sarrazin J, Robigou V, Juniper SK, Delaney JR (1997) Biological and geological dynamics over four years on a high-temperature sulfide structure at the juan de fuca ridge hydrothermal observatory. *Marine Ecology Progress Series*, **153**, 5-24.

Sarrazin J, Walter C, Sarradin P-M, Brind'amour A, Desbruyères D, Briand P, Fabri M-C, Van Gaever S, Vanreusel A, Bachraty C, Thiebaut E (2006) Community structure and temperature dynamics within a mussel assemblage in the southern east pacific rise. *Cahiers de Biologie Marine*, **47**, 483-490.

Schierup MH, Hein J (2000) Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156**, 879-891.

Schlager W (2008) Earth's layers, their cycles and Earth system science. *Austrian Journal of Earth Sciences* **101**, 4-16.

Shank TM, Fornari DJ, Von Damm KL *et al* (1998) Temporal and spatial patterns of biological community development at nascent deep-sea hydrothermal vents (9°50'N, East Pacific Rise). *Deep Sea Research Part II: Topical Studies in Oceanography* **45**, 465-515.

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* **28**, 25-35.

Shriner D, Nickle DC, Jensen MA, Mullins JI (2003) Potential impact of recombination on sitewise approaches for detecting positive natural selection. *Genetical Research* **81**, 115-121.

Skibinski DOF, Gallagher C, Beynon CM (1994) Mitochondrial-DNA Inheritance. *Nature* **368**, 817-818.

Slatkin M (1977) Gene flow and genetic drift in a species subject to frequent local extinction. *Theoretical Population Biology* **12**, 253-262.

Smith CI, Farrell BD (2005) Range expansions in the flightless longhorn cactus beetles, Moneilema gigas and Moneilema armatum, in response to Pleistocene climate changes. *Molecular Ecology* **14**, 1025-1044.

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology* **15**, 4261-4293.

Spalding MD, Fox HE, Allen GR, Davidson N, Ferdaña ZA, Finlayson M, Halpern BS, Jorge MA, Lombana A, Lourie SA, Martin KD, Mcmanus E, Omolnar J, Recchi, CA, Robertson J (2007) Marine ecoregions of the world: A bioregionalization of coastal and shelf areas. *BioScience*, **57**, 573-583.

Stöhr S, Segonzac M (2006) Two new genera and species of ophiuroid (echinodermata) from hydrothermal vents in the east pacific. *Species diversity*, **11**, 7-32.

Sunnucks P (2000) Efficient genetic markers for population biology. *Trends in Ecology & Evolution* **15**, 199-203.

T

Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596-1599.

Thiel M, Macaya EC, Acuna E*, et al.* (2007) The Humboldt Current System of northern and central Chile. *Oceanography and Marine Biology, Vol 45* **45**, 195-344.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**: 4673-4680.

Thomson RE, Roth SE, Dymond J (1990) Near-inertial motion over a mid-ocean ridge: effects of topography and hydrothermal plumes. *Journal of Geophysical Research* **95**: 12961-12966.

Thorrold SR, Zacherl DC, Levin LA (2007) Population Connectivity and Larval Dispersal Using Geochemical Signatures in Calcified Structures. *Oceanography* **20**, 80-89.

Ting CT, Tsaur SC, Wu CI (2000) The phylogeny of closely related species as revealed by the genealogy of a speciation gene, Odysseus. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5313-5316.

Tunnicliffe V (1988) Biogeography and Evolution of Hydrothermal-Vent Fauna in the Eastern Pacific-Ocean. *Proceedings of the Royal Society of London Series B-Biological Sciences* **233**, 347-366.

Tunnicliffe V (1991) The biology of hydrothermal vents: ecology and evolution. *Oceanography and Marine Biology. An annual review* **29**, 319-407.

Tunnicliffe V, Botros M, Deburgh ME*, et al.* (1986) Hydrothermal Vents of Explorer Ridge, Northeast Pacific. *Deep-Sea Research Part a-Oceanographic Research Papers* **33**, 401-412.

Tunnicliffe V, Embley RW, Holden JF*, et al.* (1997) Biological colonization of new hydrothermal vents following an eruption on Juan de Fuca Ridge. *Deep-Sea Research Part I-Oceanographic Research Papers* **44**, 1627-1644.

Tunnicliffe V, Fowler CMR (1996) Influence of sea-floor spreading on the global hydrothermal vent fauna. *Nature*, **379**, 531-533.

Tunnicliffe V, Juniper SK, de Burgh ME (1985) The hydrothermal vent community of Axial seamount, Juan de Fuca Ridge. *Bulletin of Biological Socity of Washington* **6**, 453-464.

Tunnicliffe V, McArthur AG, McHugh D (1998) A biogeographical perspective of the deep-sea hydrothermal vent fauna. *Advances in Marine Biology, Vol 34* **34**, 353-442.

Turnipseed M, Knick KE, Lipcius RN, Dreyer JC, Van Dover CL (2003) Diversity in mussel beds at deep-sea hydrothermal vents and cold seep. *Ecology letters*, **6**, 518-523.

Tyler P, Young C (2003) Dispersal at hydrothermal vents: a summary of recent progress. *Hydrobiologia* **503**, 9-19.

Tyler PA, German CR, Ramirez-Llodra E, Van Dover CL (2003) Understanding the biogeography of chemosynthetic ecosystems. *Oceanologica Acta*, **25**, 227-241.

Tyler PA, Pendlebury S, Mills SW*, et al.* (2008) Reproduction of gastropods from vents on the East Pacific Rise and the Mid-Atlantic Ridge. *Journal of Shellfish Research* **27**, 107-118.

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

V

Van Dover C, Trask JL (2000) Diversity at deep-sea hydrothermal vent and intertidal mussel beds. *Marine Ecology Progress Series*, **195**, 169-178.

Van Dover CL (2002) Community structure of mussels beds at deep-sea hydrothermal vents. *Marine Ecology Progress Series*, **230**, 137-158.

Van Dover CL (2003) Variation in community srtucture within hydrothermal vent mussel beds of the east pacific rise. *Marine Ecology Progress Series*, **253**, 55-66.

Van Dover CL, Doerries MB (2005) Community structure in mussel beds at logatchev hydrothermal vents and a comparison of macrofaunal species richness on slow- and fast-spreading mid-ocean ridges. *Marine Ecology*, **26**, 110-120.

Van Dover CL, German CR, Speer KG, Parson LM, Vrijenhoek RC (2002) Evolution and biogeography of deep-sea vent and seep invertebrates. *Science* **295**, 1253-1257.

Van Dover CL, Humphris SE, Fornari D*, et al.* (2001a) Biogeography and ecological setting of Indian Ocean hydrothermal vents. *Science* **294**, 818-823.

Van Dover CL, Jenkins CD, Turnipseed M (2001b) Corralling of larvae in the deep sea. *Journal of the Marine Biological Association of the United Kingdom* **81**, 823-826.

Van Dover CL, Trask JL, Gross J, Knowlton A (1999) Reproductive biology of free-living and commensal polynoid polychaetes at the Lucky Strike hydrothermal vent field (Mid-Atlantic Ridge). *Marine Ecology-Progress Series* **181**, 201-214.

Via S (2001) Sympatric speciation in animals: The ugly duckling grows up. *Trends in Ecology & Evolution*, **16**, 381-390.

Vrijenhoek RC (1997) Gene flow and genetic diversity in naturally fragmented metapopulations of deep-sea hydrothermal vent animals. *Journal of Heredity* **88**, 285-293.

W

Wade MJ, McCauley DE (1988) Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution* **42**, 995-1005.

Wakeley J (2009) The structured coalescent. In : Coalescent theory : an introduction. Camp R (ed), *Roberts and Company*, 165-166.

Wall JD, Andolfatto P, Przeworski M (2002) Testing models of selection and demography in Drosophila simulans. *Genetics* **162**, 203-216.

Wang RL, Wakeley J, Hey J (1997) Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics* **147**, 1091-1106.

Warén A, Bengtson S, Goffredi SK, Van Dover CL (2003) A hot-vent gastropod with iron sulfide dermal sclerites. *Science* **302**, 1007-1007.

Warén A, Bouchet P (2001) Gastropoda and Monoplacophora from hydrothermal vents and seeps; New taxa and records. *Veliger* **44**, 116-231.

Warwick RM, Clarke KR (2001) Practical measures of marine biodiversity based on relatedness of species. *Oceanography and Marine Biology. An Annual Review*, **39**, 207-231.

Watremez P, Kervévan C (1990) Origin of Variations in Hydrothermal Activity - 1st Results of a Simple Numerical-Model. *Comptes Rendus De L Academie Des Sciences Serie Ii* **311**, 153-158.

Watterson GA (1975) Number of Segregating Sites in Genetic Models without Recombination. *Theoretical Population Biology* **7**, 256-276.

Weir B, Cockerham C (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.

Whitlock MC, McCauley DE (1990) Some population genetic consequences of colony formation and extinction: genetic correlations within foundings groups. *Evolution* **44**, 1717-1724.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: Fst not equal 1/(4Nm+1). *Heredity* **82**, 117-125.

Whittam TS, Nei M (1991) Neutral Mutation Hypothesis Test. *Nature* **354**, 115-116.

Wilke T, Davis GM (2000) Intraspecific mitochondrial sequence diversity in Hydrobia ulvae and Hydrobia ventrosa (Hydrobiidae : Rissooidea : Gastropoda): Do their different life histories affect biogeographic patterns and gene flow? *Biological Journal of the Linnean Society* **70**, 89-105.

Wolstenholme DR, Clary DO (1985) Sequence evolution of *Drosophila* mitochondrial DNA. *Genetics* **109**, 725-744.

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

Won YJ, Jones WJ, Vrijenhoek RC (2008) Absence of cospeciation between deep-sea Mytilids and their thiotrophic endosymbionts. *Journal of Shellfish Research* **27**, 129-138.

Wood AR, Turner G, Skibinski DOF, Beaumont AR (2003) Disruption of doubly uniparental inheritance of mitochondrial DNA in hybrid mussels (Mytilus edulis x M-galloprovincialis). *Heredity* **91**, 354-360.

Wood TE, Burke JM, Rieseberg LH (2005) Parallel genotypic adaptation: when evolution repeats itself. *Genetica* **123**, 157-170.

Wright S (1931) Evolution in mendelian populations. *Genetics* **16**, 97-159.

Wright S (1937) The distribution of gene frequencies in populations. *Proceedings of the National Academy of Sciences of the United States of America* **23**, 307-320.

Wright S (1943) Isolation by distance. *Genetics* **28**, 114-138.

Wright S (1977) Evolution and the genetics of populations. Experimental Results and Evolutionary Deductions. University of Chicago Press, Chicago, Vol. 3

Y

Young CR, Fujio S, Vrijenhoek RC (2008) Directional dispersal between mid-ocean ridges: deep-ocean circulation and gene flow in Ridgeia piscesae. *Molecular Ecology* **17**, 1718-1731.

Z

Zacherl DC (2005) Spatial and temporal variation in statolith and protoconch trace elements as natural tags to track larval dispersal. *Marine Ecology-Progress Series* **290**, 145-163.

Zal F, Desbruyères D, Join-Toulmond C (1994) Sexual dimorphism in *Paralvinella grasslei*, a polychaete annelid from deep-sea hydrothermal vents. *Comptes rendus de l'Académie des Sciences Série III Sciences de la vie* **317**, 42-48.

Zierenberg RA, Adams MWW, Arp AJ (2000) Life in extreme environments: Hydrothermal vents. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 12961-12962.

Zink RM (2002) Methods in comparative phylogeography, and their application to studying evolution in the North American aridlands. *Integrative and Comparative Biology* **42**, 953-959.

Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology* **17**, 2107-2121.

Zouros E, Ball AO, Saavedra C, Freeman KR (1994) An Unusual Type of Mitochondrial-DNA Inheritance in the Blue Mussel *Mytilus*. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 7463-7467.

Annexe 1 : Marquage Recapture de séquences

Molecular Ecology Notes (2007) doi: 10.1111/j.1471-8286.2007.01685.x

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 concertedly 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

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

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 or 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

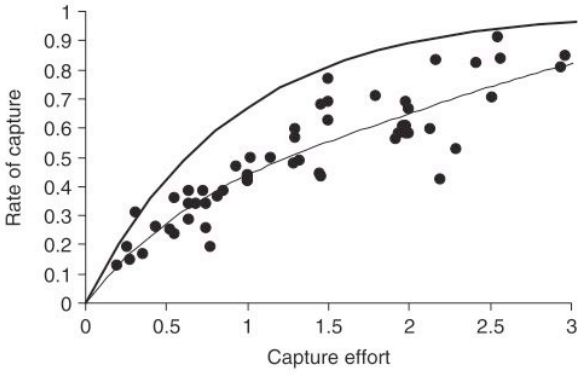


Fig. 1 Rate of capture (number of individuals captured/initial sample size) as a function of the capture effort (number of sequences/initial sample size). The thick line is the expectation based on a uniform distribution and the thin line is a quartic polynomial regression on the data.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

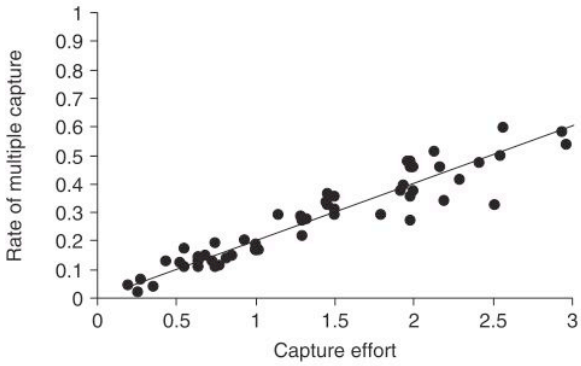


Fig. 2 Rate of multiple captures (number of individuals captured more than once/initial sample size) as a function of the capture effort (number of sequences/initial sample size). The line is a linear regression on the data (slope = 0.2).

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

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

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, 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).

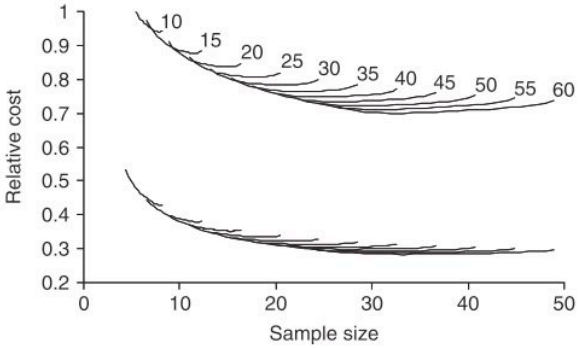


Fig. 3 Estimated cost of MR cloning protocols relative to standard protocols of individual cloning as a function of the sample size for population genetics analysis. The empirically estimated rate of capture of Fig. 1 was used for a gradient of initial sample size (the number of individuals PCR-amplified 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).

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.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

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 polymophisms 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.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Annexe 2 : logiciel Rabouter Séquence

Manuel d'utilisation du programme RabouterSequences

Réalisé par Sophie Plouviez et Matthieu Bruneaux
Equipe Génétique de l'Adaptation en Milieux Extrêmes
Station Biologique de Roscoff

Pourquoi utiliser RabouterSequences?

Le programme RabouterSequences permet de rabouter facilement un grand nombre de séquences forward et reverse en quelques secondes à partir d'un site de coupure que l'on définit. Il peut être très utile pour rabouter des séquences effectuées sur plusieurs individus d'un même gène. En effet, avant de rabouter les séquences forward et reverse, l'utilisateur vérifie les chromatogrammes de ces séquences. La séquence forward est corrigée jusqu'à un site de coupure déterminé appartenant à la zone de recouvrement des deux séquences. Ce site de coupure sera le même pour l'ensemble des séquences. La séquence reverse est ensuite corrigée à partir du site de coupure jusqu'à la fin de la séquence. Le programme RabouterSequences permettra de rabouter chaque séquence forward à sa séquence reverse correspondante.

Avantage de RabouterSequences?

Vous gardez le contrôle sur ce que vous raboutez. Pas de séquences consensus. Le logiciel ne modifie pas les sites de vos séquences.
Pas besoin de changer l'ordre de vos séquences pour que Forward et Reverse soient homogènes. Le logiciel cherche lui même la séquence correspondante.
Rapide et simple d'utilisation (exemple: 600 séquences raboutées en 25 secondes, alors que manuellement cela vous prendrez plusieurs heures voire jours).

Comment utiliser RabouterSequences?

Fichier d'entrée:
Le programme nécessite deux fichiers d'entrée au format fasta (.fas) contenant respectivement toutes les séquences en forward (voir exempleForward.fas) et toutes les séquences en reverse (voir exempleReverse.fas). Les noms des séquences ne doivent pas dépasser 50 caractères. Il est inutile de tronquer les séquences forward et reverse de part et d'autre du site de coupure. Ainsi, si le site de coupure est 500pb, les 500 premiers caractères des séquences du fichier forward seront raboutés aux caractères 501 à la fin des séquences du fichier reverse.
Attention: le programme reconnait automatiquement le format fasta en .fas. Si votre fichier est en .fasta, changer le en .fas. Lorsque le nom des fichiers vous est demandé, il est donc inutile d'inscrire .fas. Dans l'exemple fourni il suffit donc de taper:
Nom du fichier Forward? exempleForward

Nom du fichier Reverse? exempleReverse

Paramètres demandés par le programme:

Il est inutile d'indiquer .fas au logiciel pour le nom du fichier de sortie. Dans l'exemple fourni:
Nom du fichier de sortie? exempleRaboutecoupure110

Le programme compare les premiers caractères des noms des séquences forward aux premiers caractères des séquences reverse. Si les noms sont identiques, il raboute les séquences de part et d'autre du site de coupure. Il faut donc lui définir le nombre de caractères à comparer entre les deux fichiers. Ainsi, dans l'exemple, les 4 premiers caractères permettent d'identifier la séquence complémentaire (e.g. ind1) et le site de coupure a été choisi à 110pb. Dans le fichier de sortie, les séquences raboutées auront le nom ses caractères. Il est possible d'y ajouter jusqu'à 15 caractères pour l'ensemble des séquences. Ainsi, dans l'exemple, nous avons ajouté "Raboute" à toutes les séquences.
Caracteres ajoutes aux noms des sequences raboutees? Raboute
Si vous ne souhaitez pas rajouter de caractères aux noms de séquences raboutées, tapez "non".

Attention: Au sein du même fichier, deux séquences ne doivent pas avoir exactement les mêmes caractères d'identification.

Fichier de sortie:

Deux fichiers de sortie sont founis:

Le premier fichier RabouterJournal (.txt) contient les informations de déroulement du programme: les séquences raboutées (faite), les séquences forward dont aucune séquence reverse n'a été trouvée (forward seule), les séquences reverse dont aucune séquence forward n'a été trouvée (reverse seule).

Le second fichier de sortie est en fasta (.fas) et contient l'ensemble des séquences raboutées.

Compatibilité de RabouterSequences:

RabouterSequences a été compilé pour fonctionner sous Windows. Cependant, son utilisation sous MacOS ou Linux nécessite de recompiler les fichiers sources .c et .h. Ces fichiers peuvent vous être fournis sur demande.

Fichier source .c

```
// programme pour rabouter des sequences
// Sophie et Matthieu
// commence le 26/02/2009
// derniere modification le 01/03/2009

#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include <time.h>
#include "fonctions_rabouterSequences.h"
#define LONGUEUR_MAX_SEQUENCES 5000

int main(int argc, char *argv[])
{

	// variables

	char nomFichierForward[50]={0};		// nom du fichier contenant les sequences forward
	char nomFichierReverse[50]={0};		// nom du fichier contenant les sequences reverse
	char nomFichierSortie[150]={0};		// nom du fichier de sortie
	char nomFichierJournal[50]="rabouterJournal";		// nom du fichier journal
	FILE* fichierForward=NULL;			// pointeur vers le fichier Forward
	FILE* fichierReverse=NULL;			// pointeur vers le fichier Reverse
	FILE* fichierSortie=NULL;		// pointeur vers le fichier de sortie
	FILE* fichierJournal=NULL;			// pointeur vers le fichier journal
	int nombreSequencesForward=0;		// nombre de sequences du fichier Forward
	int nombreSequencesReverse=0;		// nombre de sequences du fichier Reverse
	char nomRecherche[50]={0};			// nom d'une sequence forward dont on cherche la reverse
	char nomCompare[50]={0};		// nom d'une sequence reverse que l'on compare avec celui de la forward
	int nomTrouve=0;				// booleen : nom Reverse trouve correspondant a un Forward
	char sequenceForward[LONGUEUR_MAX_SEQUENCES]={0};		// sequence forward
	char sequenceReverse[LONGUEUR_MAX_SEQUENCES]={0};//		sequence reverse
	int longueurCoupeForward=0;			// longueur de sequence forward a conserver
	int longueurCoupeReverse=0;			// idem reverse
	char ajoutNom[15]={0};			//caracteres ajoutes aux noms des sequences raboutees
```

```
	int caracteresNomSequencesCompares=0;  //nombre de caractères compares des noms de sequences


	// demander les noms de fichiers d'entree et autres
	printf("\nRabouterSequences - programme realise par Sophie Plouviez et Matthieu Bruneaux");
   printf("\nEquipe GAME (Genetique de l'Adaptation en Milieux Extremes)");
   printf("\nStation Biologique de Roscoff");
   printf("\n\nNom du fichier Forward ? ");
	scanf("%s",nomFichierForward);
	printf("Nom du fichier Reverse ? ");
	scanf("%s",nomFichierReverse);
	printf("Nom du fichier de sortie ? ");
	scanf("%s",nomFichierSortie);
	printf("Longueur de coupe Forward ? ");
	scanf("%ld",&longueurCoupeForward);
	printf("Caracteres ajoutes aux noms des sequences raboutees \n(ne rien ajouter: taper non)? ");
	scanf("%s",ajoutNom);
	longueurCoupeReverse=longueurCoupeForward+1;
	printf("Combien de caracteres compares pour les noms de sequences ? ");
	scanf("%ld",&caracteresNomSequencesCompares);

	// preparation du nom des fichiers
	strcat(nomFichierJournal,"_");
	strcat(nomFichierJournal,nomFichierSortie);
	strcat(nomFichierJournal,".txt");
	strcat(nomFichierForward,".fas");
	strcat(nomFichierReverse,".fas");
	strcat(nomFichierSortie,".fas");

	//variables sequences
	char sequenceCoupeForward[LONGUEUR_MAX_SEQUENCES]={0};
	char sequenceCoupeReverse[LONGUEUR_MAX_SEQUENCES]={0};

	// verification des noms de fichiers
	printf("\nLes fichiers sont :");
	printf("\n\tsequences Forward : %s",nomFichierForward);
	printf("\n\tsequences Reverse : %s\n",nomFichierReverse);

	// ouverture des fichiers contenant les sequences et du fichier journal
	fichierForward=fopen(nomFichierForward,"r");
	if (fichierForward==NULL) {
		printf("\nErreur : impossible d'ouvrir le fichier %s\n",nomFichierForward);
		exit(EXIT_FAILURE);
		}
	fichierReverse=fopen(nomFichierReverse,"r");
	if (fichierReverse==NULL) {
		printf("\nErreur : impossible d'ouvrir le fichier %s\n",nomFichierReverse);
```

```
		exit(EXIT_FAILURE);
		}
	fichierSortie=fopen(nomFichierSortie,"w+");
	if (fichierSortie==NULL) {
		printf("\nErreur : impossible d'ouvrir le fichier %s\n",nomFichierSortie);
		exit(EXIT_FAILURE);
		}
	fichierJournal=fopen(nomFichierJournal,"w+");
	if (fichierJournal==NULL) {
		printf("\nErreur : impossible d'ouvrir le fichier %s\n",nomFichierJournal);
		exit(EXIT_FAILURE);
		}

	// journal
	time_t heure;
	struct tm* heureConvertie;
	time(&heure);
	heureConvertie=localtime(&heure);
	fprintf(fichierJournal,"rabouterSequences - fichier journal\n");
	fprintf(fichierJournal,"\n - programme compile le %s a
%s",__DATE__,__TIME__);
	fprintf(fichierJournal,"\n\n\n\n\tDebut du programme :
%s",asctime(heureConvertie));
	fprintf(fichierJournal,"\n\nFichiers :");
	fprintf(fichierJournal,"\n----------");
	fprintf(fichierJournal,"\n  Forward : %s",nomFichierForward);
	fprintf(fichierJournal,"\n  Reverse : %s",nomFichierReverse);
	fprintf(fichierJournal,"\n  sortie  : %s\n",nomFichierSortie);

	// compter le nombre de sequences dans chaque fichier
	nombreSequencesForward=compterSequences(fichierForward);
	printf("\nLe fichier %s contient %ld
sequences",nomFichierForward,nombreSequencesForward);
	nombreSequencesReverse=compterSequences(fichierReverse);
	printf("\nLe fichier %s contient %ld
sequences\n",nomFichierReverse,nombreSequencesReverse);

	// journal
	fprintf(fichierJournal,"\n  Forward : %ld sequences",nombreSequencesForward);
	fprintf(fichierJournal,"\n  Reverse : %ld sequences\n",nombreSequencesReverse);
	//
	fprintf(fichierJournal,"\n\n\nRaboutage des sequences :");
	fprintf(fichierJournal,"\n-------------------------");


	// remise des curseurs au debut des fichiers
	rewind(fichierForward);
	rewind(fichierReverse);

	// lecture d'un nom forward et recherche de la sequence reverse correspondante
```

```
	int continuerLectureForward=1;
	int continuerLectureReverse=1;
	int compteurSequencesLuesForward=0;
	int compteurSequencesTrouvees=0;

	while
(continuerLectureForward&&compteurSequencesLuesForward<nombreSequencesForwar
d) {

	// lecture
	compteurSequencesLuesForward++;
	continuerLectureForward=extraireNomSequence(nomRecherche,sequenceForward,
fichierForward,caracteresNomSequencesCompares);
	extractionChaine(sequenceCoupeForward,sequenceForward,0,longueurCoupeForw
ard-1);

	// comparaison dans l'autre fichier
	continuerLectureReverse=1;
	while(continuerLectureReverse) {

	continuerLectureReverse=extraireNomSequence(nomCompare,sequenceReverse,fi
chierReverse,caracteresNomSequencesCompares);
		nomTrouve=strcmp(nomRecherche,nomCompare);
		if (nomTrouve==0) {
			continuerLectureReverse=0;

	extractionChaine(sequenceCoupeReverse,sequenceReverse,longueurCoupeForward
,LONGUEUR_MAX_SEQUENCES);
			strcat(sequenceCoupeForward,sequenceCoupeReverse);
			if (strcmp(ajoutNom,"non")==0) {
	fprintf(fichierSortie,">%s\n",nomRecherche);}
	else {
			fprintf(fichierSortie,">%s%s\n",nomRecherche, ajoutNom);}
			fprintf(fichierSortie,"%s\n",sequenceCoupeForward);
			compteurSequencesTrouvees++;
			fprintf(fichierJournal,"\n  %s\tfaite",nomRecherche);
			}
		}

	// remonter au debut du fichier reverse
	rewind(fichierReverse);
			int i=0;

	// vider les variables temporaires
	for (i=0; i<LONGUEUR_MAX_SEQUENCES; i++) {
		sequenceCoupeForward[i]=0;
		sequenceCoupeReverse[i]=0;
		}
```

```
	// fin de la boucle pour une sequence forward / retour au debut
	}

	// journal
	fprintf(fichierJournal,"\n\nTotal : %ld sequences raboutees\n",compteurSequencesTrouvees);

	// sequences forward sans reverse
		rewind(fichierForward);
		rewind(fichierReverse);
		int compteurForwardSeules=0;
		continuerLectureForward=1;
		continuerLectureReverse=1;
		compteurSequencesLuesForward=0;

		fprintf(fichierJournal,"\n\n\nSequences Forward seules :");
		fprintf(fichierJournal,"\n-------------------------");

		while (continuerLectureForward&&compteurSequencesLuesForward<nombreSequencesForward) {

		// lecture
		compteurSequencesLuesForward++;

	continuerLectureForward=extraireNomSequenceSeul(nomRecherche,fichierForward,caracteresNomSequencesCompares);

		// comparaison dans l'autre fichier
		continuerLectureReverse=1;
		while(continuerLectureReverse) {

	continuerLectureReverse=extraireNomSequenceSeul(nomCompare,fichierReverse,caracteresNomSequencesCompares);
			nomTrouve=strcmp(nomRecherche,nomCompare);
			if (nomTrouve==0) {
				continuerLectureReverse=0;
				}
			}
		if (nomTrouve!=0) {
			compteurForwardSeules++;
			fprintf(fichierJournal,"\n  %s",nomRecherche);
			}

		// remonter au debut du fichier reverse
		rewind(fichierReverse);
				int i=0;

		// vider les variables temporaires
		for (i=0; i<LONGUEUR_MAX_SEQUENCES; i++) {
```

```
                        sequenceCoupeForward[i]=0;
                        sequenceCoupeReverse[i]=0;
                        }

                // fin de la boucle pour une sequence forward / retour au debut
                }

                // journal
                fprintf(fichierJournal,"\n\nTotal : %ld sequences Forward
seules\n",compteurForwardSeules);

        // sequences reverse sans forward
                rewind(fichierForward);
                rewind(fichierReverse);
                int compteurReverseSeules=0;
                continuerLectureForward=1;
                continuerLectureReverse=1;
                int compteurSequencesLuesReverse=0;

                fprintf(fichierJournal,"\n\n\nSequences Reverse seules :");
                fprintf(fichierJournal,"\n-------------------------");

                while
(continuerLectureReverse&&compteurSequencesLuesReverse<nombreSequencesReverse)
{

                // lecture
                compteurSequencesLuesReverse++;

        continuerLectureReverse=extraireNomSequenceSeul(nomRecherche,fichierRevers
e,caracteresNomSequencesCompares);

                // comparaison dans l'autre fichier
                continuerLectureForward=1;
                while(continuerLectureForward) {

        continuerLectureForward=extraireNomSequenceSeul(nomCompare,fichierForward
,caracteresNomSequencesCompares);
                        nomTrouve=strcmp(nomRecherche,nomCompare);
                        if (nomTrouve==0) {
                                continuerLectureForward=0;
                                }
                        }
                if (nomTrouve!=0) {
                        compteurReverseSeules++;
                        fprintf(fichierJournal,"\n  %s",nomRecherche);
                        }

                // remonter au debut du fichier forward
                rewind(fichierForward);
```

```
                        int i=0;

                // vider les variables temporaires
                for (i=0; i<LONGUEUR_MAX_SEQUENCES; i++) {
                        sequenceCoupeForward[i]=0;
                        sequenceCoupeReverse[i]=0;
                        }

                // fin de la boucle pour une sequence reverse / retour au debut
                }

                // journal
                fprintf(fichierJournal,"\n\nTotal : %ld sequences Reverse
seules",compteurReverseSeules);

                // fin
                time(&heure);
                heureConvertie=localtime(&heure);
                fprintf(fichierJournal,"\n\n\n\t Fin du programme :
%s\n",asctime(heureConvertie));



        // fermeture des fichiers ouverts
        fclose(fichierForward);
        fclose(fichierReverse);
        fclose(fichierSortie);
        fclose(fichierJournal);

        system("PAUSE");
        return EXIT_SUCCESS;
}
```

Fichier source fonction .h

```
#define NOMBRE_MAX_CHAR 80                    // nombres de caracteres maximal lus par ligne de fichier contenant les sequences

int compterSequences(FILE* fichierOuvert) {
        int nombreSequences=0;
        char ligneLue[NOMBRE_MAX_CHAR];
        char* reference=">";

        // lecture des lignes
        while(fgets(ligneLue,NOMBRE_MAX_CHAR,fichierOuvert)!=NULL) {
                if (ligneLue[0]==*reference) {                // on teste si la ligne contient un nom de sequence ('>' en format FASTA)
                        nombreSequences=nombreSequences+1;
                        }
                }
        // renvoie le nombre de sequences
        return(nombreSequences);
        }

void extractionChaine(char* sousChaine, char* chaine, int depart, int fin) {
        int j=0;

        // copie des caracteres un a un (boucle)
        for (j=depart; j<(fin+1); j++) {
                sousChaine[j-depart]=chaine[j];
                }

        //return(sousChaine);
        return;
        }

int extraireNomSequence(char* nomSequence, char* sequenceForward, FILE* fichierOuvert, int longueurComparee) {
        int succes=1;
        int nombreSequences=0;
        char ligneLue[NOMBRE_MAX_CHAR];
        char* reference=">";
        int nomTrouve=0;
        int caractere=0;
        char* testLigne=NULL;

        // recuperation du nom
        do {
        testLigne=fgets(ligneLue,NOMBRE_MAX_CHAR,fichierOuvert);
        if (ligneLue[0]==*reference) {                // on teste si la ligne contient un nom de sequence ('>' en format FASTA)
                nomTrouve=1;
```

```
		extractionChaine(nomSequence,ligneLue,1,longueurComparee); // on extrait le nom de la sequence
		}
	} while ((testLigne!=NULL)&&(nomTrouve==0));

	if (testLigne==NULL) {succes=0;}

	// recul du curseur pour etre en debut de sequence
	fseek(fichierOuvert,-1,SEEK_CUR);

	// recuperation de la sequence
	int i=0;
	int continuer=1;
	while (continuer) {
		caractere=fgetc(fichierOuvert);
		if ((caractere!=EOF)&&(caractere!=10)&&(caractere!=62)) {
			sequenceForward[i]=caractere;
			i++;
				}
		else if ((caractere==EOF)||(caractere==62)) {continuer=0;}
		}
	fseek(fichierOuvert,-1,SEEK_CUR);

	return (succes);
	}

int extraireNomSequenceSeul(char* nomSequence, FILE* fichierOuvert, int longueurComparee) {
	int succes=1;
	int nombreSequences=0;
	char ligneLue[NOMBRE_MAX_CHAR];
	char* reference=">";
	int nomTrouve=0;
	int caractere=0;
	char* testLigne=NULL;

	// recuperation du nom
	do {
	testLigne=fgets(ligneLue,NOMBRE_MAX_CHAR,fichierOuvert);
	if (ligneLue[0]==*reference) {		// on teste si la ligne contient un nom de sequence ('>' en format FASTA)
		nomTrouve=1;
		extractionChaine(nomSequence,ligneLue,1,longueurComparee); // on extrait le nom de la sequence
		}
	} while ((testLigne!=NULL)&&(nomTrouve==0));

	if (testLigne==NULL) {succes=0;}

	// recul du curseur pour etre en debut de sequence
```

```
fseek(fichierOuvert,-1,SEEK_CUR);

return (succes);
}
```

Annexe 3 : Conditions de PCR

- ***Alvinella pompejana***

Amorces utilisées pour l'amplification des gènes nucléaires d'*A. pompejana*
---- signifie que 4 pb ont été ajoutées pour la méthode de Marquage Recapture de séquences

Locus	Sens de l'amorce	Amorce 5'-3'
PGM	Sens	----CCCAGGTGGTCCAAATGC
	Antisens	----GTACAGATGTCAGCCTTCAGGTC
SAHH	Sens	----CTTGTCAACCTCGGCTGTG
	Antisens	----CTGCCTGCTCGTCTGTTAG
GlobX	Sens	----AGTTGACTGAAGAACGCAGAGAAGCGGT
	Antisens	----CGCGAAGATCCTTGAAGTACTCTTGGTA

Conditions de mix de PCR pour un puit

	PGM	SAHH	GlobX
BSA 2%	0,5	0	0
Tampon Taq 10X	2,5	2,5	2,5
MgCl2 25mM	2	2	2
dNTPs 2,5 mM chacun	2,5	0,5	0,5
Amorce Sens 10µM	1	1,2	1,2
Amorce Antisens 10µM	1	1,2	1,2
Taq 5u/µl	0,1	0,1	0,1
ADN	2,5	5	5
Eau ultrapure	12,9	12,5	12,5
Volume final	25	25	25

Conditions d'amplification des gènes nucléaires.
Tm PGM = 57°C, Tm SAHH = 57°C, Tm GlobX = 55°C

Température °C	Temps	Nombre de cycles
94	3 min	1
94	30 s	
Tm	20 s	40
72	2min 30 s	
72	10 min	1

- ***Bathymodiolus thermophilus***

Amorces utilisées pour l'amplification des gènes nucléaires de *B. thermophilus*
---- signifie que 4 pb ont été ajoutées pour la méthode de Marquage Recapture de séquences

Locus	Sens de l'amorce	Amorce 5'-3'
SAHH	Sens	----TAAATCTTGGTTGTGCTCATGGTCATCCA
	Antisens	----TTTGAATGGTCCTTCTTTAGGTAGAC
Lysozyme	Sens	----GTTTCCCCAAAATGTATGAGCTGT
	Antisens	----TAATCTCCGCCTGGACTACCACAATC
Sulfo 1 et 2	Sens	----TCTTTAAAGTCAGGATCACATTGG
	Antisens	----TAAGGCAAAGTGGAACAACGAGACCGC

Conditions de mix de PCR pour un puit

	SAHH	Lysozyme	Sulfo1 et 2
BSA 2%	0,5	0,5	0,5
Tampon Taq 10X	2	2	2
MgCl2 25mM	1,6	1,6	1,6
dNTPs 2,5 mM chacun	2	2	2,4
Amorce Sens 10µM	0,8	0,8	0,8
Amorce Antisens 10µM	0,8	0,8	0,8
Taq 5u/µl	0,1	0,1	0,1
ADN	2	2	2
Eau ultrapure	10,2	10,2	9,8
Volume final	20	20	20

Conditions d'amplification des gènes nucléaires

Température °C	Temps	Nombre de cycles
94	5 min	1
94	35 s	
Tm 60°C	1 min	35
72	2 min	
72	10 min	1

- ***Lepetodrilus elevatus***

Amorces utilisées pour l'amplification des gènes nucléaires et l'amplification et le séquençage des clones mitochondriaux de *L. elevatus*
---- signifie que 4 pb ont été ajoutées pour la méthode de Marquage Recapture de séquences

Locus	Sens de l'amorce	Amorce 5'-3'
Lysozyme	Sens	----GAATGAATCGCTCGTAAAATCCC
	Antisens	----CGCAGGTAGGGGCACGACC
VDG3	Sens	----CTTCTTCTGTAAGTATCCCYGTC
	Antisens	----AAGCCGAACTTCGTCACAA
Textilinine	Sens	----GGGGATCCATGTTGAAGTTTGTCGTCTGCTCCTTGT
	Antisens	----ATCCCACTCCTGCGCACTTGTTCTCGCAGTCG
Plasmide	Sens SP6	CATTTAGGTGACACTATAG
	Antisens T7	GTAATACGACTCACTATA

Conditions de mix de PCR pour un puit

	Lysozyme	VDG3	Textilinine	SP6/T7
BSA 2%	0	1	1	0
Tampon Taq 10X	2,5	2,5	2,5	1,25
MgCl2 25mM	2	2	2	1
dNTPs 2,5 mM chacun	0,5	0,7	0,7	0,35
Amorce Sens 10µM	1,2	1,5	1,5	0,75
Amorce Antisens 10µM	1,2	1,5	1,5	0,75
Taq 5u/µl	0,1	0,1	0,1	0,05
ADN	5	5	5	5
Eau ultrapure	12,5	10,7	10,7	3,35
Volume final	25	25	25	12,5

Conditions d'amplification du gène Lysozyme

Température °C	Temps	Nombre de cycles
94	3 min	1
94	30s	
TD 66-56°C	20s	10
72	2min 30s	
94	30 s	
Tm 56°C	20 s	30
72	2min 30s	
72	10 min	1

Conditions d'amplification du gène VDG3

Température °C	Temps	Nombre de cycles
94	3 min	1
94	45 s	
Tm 57°C	30 s	40
72	1min 15s	
72	10 min	1

Conditions d'amplification du gène Textilinine

Température °C	Temps	Nombre de cycles
94	3 min	1
94	45 s	
Tm 60°C	30 s	40
72	1min 15s	
72	10 min	1

Conditions d'amplification des clones de mtCOI

Température °C	Temps	Nombre de cycles
94	3 min	1
94	1 min	
Tm 55°C	1 min	30
72	1min 15s	
72	10 min	1

Annexe 4 : Code génétique mitochondrial des Alvinellidae

La traduction des séquences d'*A. pompejana* en utilisant le code mitochondrial des Plathelminthes « flatworm » met en évidence six mutations non synonymes dans le polymorphisme du gène de la Cytochrome c Oxydase I (COI). Ces mutations ne sont pas fixées entre les lignées nord et sud, ne correspondent pas à des singletons mais sont plutôt en équi-fréquence dans le jeu de séquences analysées (mutations en fréquence intermédiaire). Elles correspondent toutes à une mutation entre l'acide aminé méthionine (ATG) et un unique codon de l'isoleucine (ATA, Fig. A.1). Cette observation suggère que le codon ATA pourrait coder en réalité pour une méthionine et non pour une isoleucine : une situation déjà rencontrée dans le code génétique de la mitochondrie de la drosophile et des Chordés (e.g. Wolstenholme & Clary 1985). En utilisant le code génétique de la drosophile (plus proche phylogénétiquement), aucune mutation non synonyme n'est présente, ni dans le polymorphisme d'*A. pompejana*, ni dans sa divergence avec *A. caudata*, indiquant une forte sélection purifiante sur le gène mtCOI. Avec le code génétique des Plathelminthes, le test de McDonald Kreitman réalisé entre *A. pompejana* et *A. caudata* est significatif ($p = 0,04$), avec 6 mutations non synonymes dans le polymorphisme d'*A. pompejana* mais aucune dans la divergence fixée entre ces espèces, confirmant que le biais observé est dû à l'usage d'un mauvais code génétique.

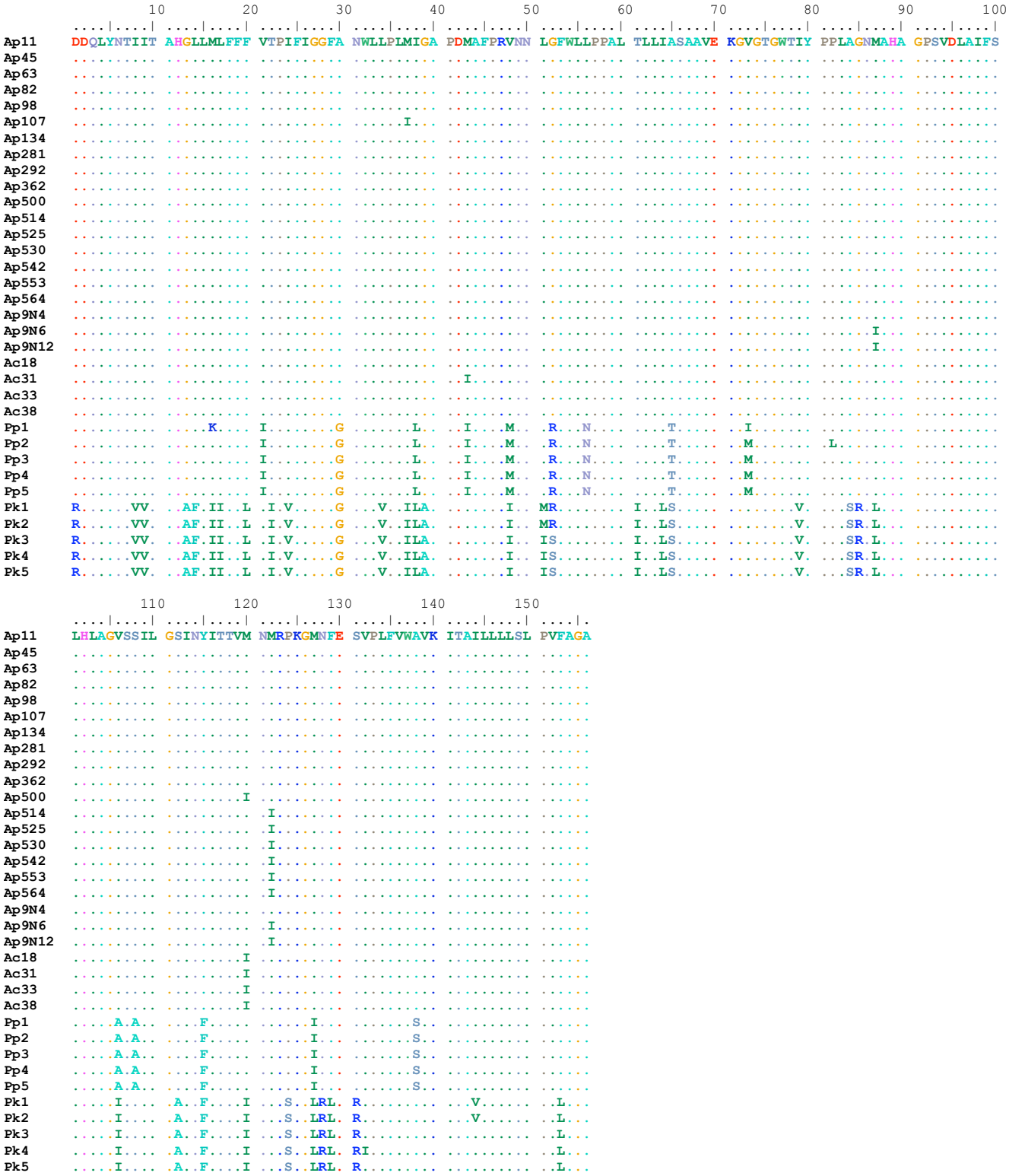


Fig. A.1 Alignement de séquences de mtCOI d'*A. pompejana* (Ap) et *A. caudata* (Ac) traduites en acides aminés en utilisant le code génétique des Plathelminthes.

Nous avons donc cherché à vérifier cette hypothèse en comparant ces séquences d'*A. pompejana* à des séquences de mtCOI d'autres Alvinellidae (*A. caudata*, *Paralvinella pandorae*) et un polychète côtier d'une famille proche (*Pectinaria koreni*). Nous avons effectué un arbre phylogénétique en utilisant un sous-échantillon de 20 séquences d'*A. pompejana*, selon la méthode des plus proches voisins : sur les séquences en acides aminés traduites selon les deux codes génétiques (Fig. A.2). Avec le code génétique des Plathelminthes, un individu *A. pompejana* s'avère identique en acides aminés à plusieurs *A. caudata*. Le code génétique mitochondrial de *Drosophila* montre, quant à lui, que l'ensemble des *Alvinella* présente la même séquence d'acides aminés (râteau), suggérant que le code génétique mitochondrial de *Drosophila* correspondrait mieux à celui des *Alvinella*.

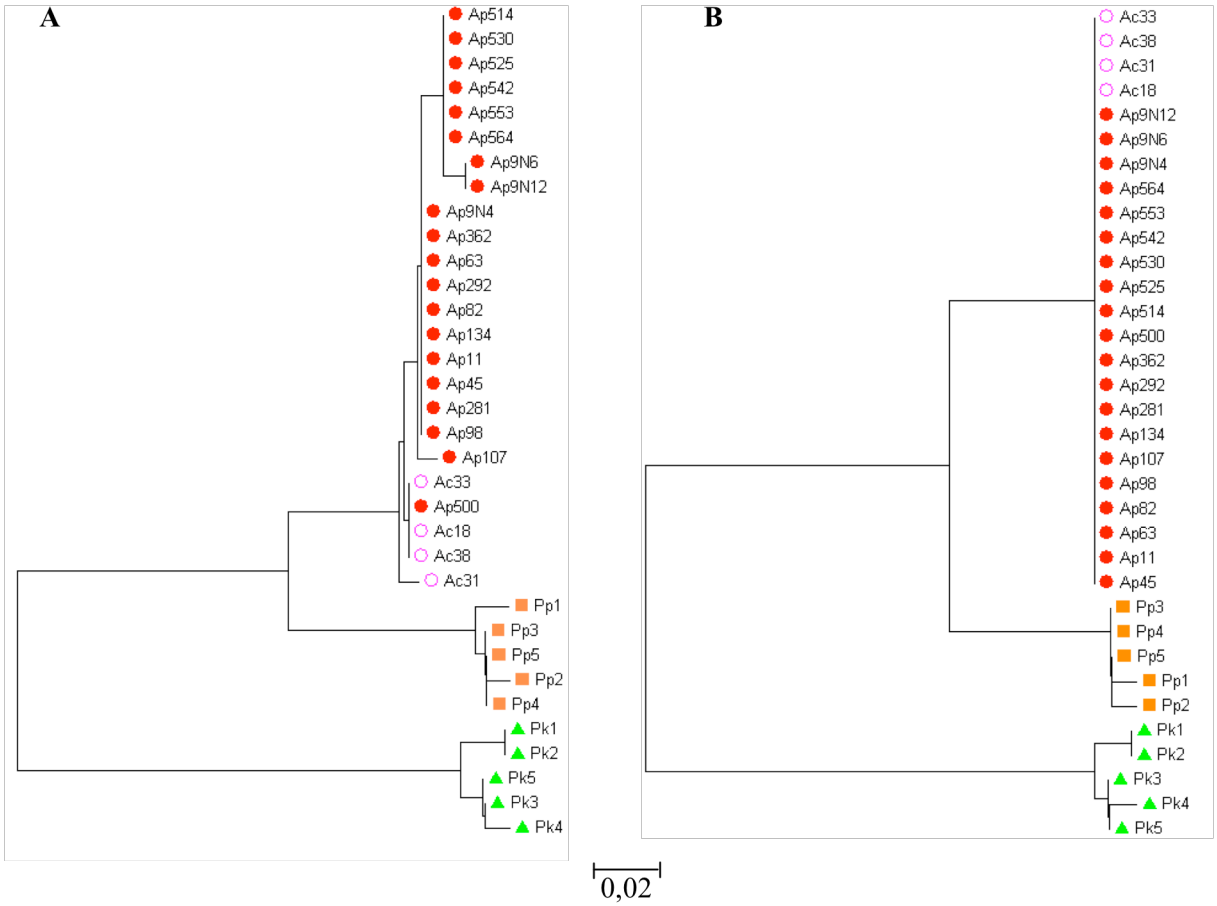


Fig. A.2 Arbres phylogénétiques du gène mtCOI de séquences en acides aminés avec le code des Plathelminthes (A) et de *Drosophila* (B).

Annexe 5 : Article Relation femelles/juvéniles chez *Branchipolynoe seepensis*

Vol. 3: 79–87, 2008
doi: 10.3354/ab00060

AQUATIC BIOLOGY
Aquat Biol

Published online July 17, 2008

OPEN ACCESS

Juvenile and adult scale worms *Branchipolynoe seepensis* in Lucky Strike hydrothermal vent mussels are genetically unrelated

Sophie Plouviez[1,2], Claire Daguin-Thiébaut[1,2], Stéphane Hourdez[1,2], Didier Jollivet[1,2,*]

[1]Université Pierre et Marie Curie-Paris 6, Laboratoire Adaptation et Diversité en Milieu Marin, Roscoff cedex, France

[2]CNRS, UMR7144 Station Biologique de Roscoff, BP 74, Place Georges Teissier, 29682, Roscoff cedex, France

ABSTRACT: The deep-sea Mid-Atlantic Ridge hydrothermal vent mussel *Bathymodiolus azoricus* harbors the polynoid polychaete *Branchipolynoe seepensis* in its palleal cavity. The latter species possesses large, yolky oocytes (~400 µm) and has been hypothesized to brood its developing embryos. The present study used 6 microsatellite markers to assess kinship among adult and juvenile worms from the same host, among worms from mussels collected within a 1 m^2 quadrat, and among samples taken on 2 dates (July 1998 and July 2001) at 2 sites (Tour Eiffel, TE, and Bairro Alto, BA) separated by 900 m at Lucky Strike hydrothermal vents on the Mid-Atlantic Ridge (37° 16′ to 19′ N, 32° 15′ to 18′ W). The hypothesis of genetic relatedness between females and juveniles within a single mussel, or among worms from the same quadrat was rejected. Conversely, great heterozygote deficiencies have been detected at nearly all loci, suggesting possible local inbreeding. There was no apparent genetic differentiation among worms from the 2 spatially separated quadrats collected at the TE and BA sites in 2001. These results suggest that embryonic or juvenile *B. seepensis* are released from their host mussel at a relatively early stage of development. They may settle close to the point of release or be dispersed to nearby sites or beyond.

KEY WORDS: Adult/offspring kinship · Microsatellite · Dispersal · Polychaete · Mid-Atlantic Ridge

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

Long-distance migration has been frequently invoked as an adaptation to ensure population persistence in highly fluctuating and fragmented environments, especially during range expansion (Nichols & Hewitt 1994). In marine environments, dispersal of invertebrates occurs mainly during the larval phase. More specifically in the deep sea, pelagic larval dispersal is considered extremely unfavorable (because of the need to move through the water column) and may be replaced by direct development (Thorson 1936). This idea was however challenged by Young (1994), who reviewed the literature on dispersal in the deep sea, and argued that lecithotrophic and planktotrophic larval stages may be more widespread than previously thought, especially in unstable environments such as hydrothermal vents and cold seeps, where most species possess pelagic larvae (Tyler & Young 1999). Given the level of the vent instability and the oligotrophic nature of bottom seawater surrounding the vents, long-distance dispersal and lecithotrophic larval development are therefore expected to be favored.

The hydrothermal vent scale worms *Branchipolynoe* spp. Pettibone live inside the mantle cavity of mussels belonging to the genus *Bathymodiolus*, which form dense mussel beds at the base of vent chimneys. This association is widespread, being found in vent ecosystems and cold seep areas around the world (Chevaldonné et al. 1998). *Branchipolynoe seepensis*, which lives in the mussel *Bathymodiolus azoricus* at the Azorean triple junction zone on the Mid-Atlantic

*Corresponding author. Email: jollivet@sb-roscoff.fr

© Inter-Research 2008 · www.int-res.com

Ridge, displays unusual reproductive behavior (Jollivet et al. 2000). Females possess a reproductive system with an ovisac containing large yolky oocytes (about 400 µm in diameter) and a pair of spermathecae (Jollivet 1996), indicating that reproduction involves pairing and internal fertilization. Mussels usually harbor only a single, large, sedentary female scale worm. However, for a limited number of observations, the female was associated with either a few juveniles or a male (much smaller in size). This pronounced size dimorphism in favor of females was found in all species of the genus, probably as the result of differing predatory rates between genders or intra-specific competition rather than dwarfism (Jollivet et al. 2000).

These findings raise an intriguing question regarding the dispersal mode of the scale worm and the filial relationships between individuals found inside the same mussel. Mating and offspring-caring behaviors are frequent occurrences in coastal polychaetes that produce large yolky eggs. The terebellids *Ramex californiensis* (maximum oocyte size: 410 µm) and *Thelepus crispus* (maximum oocyte size: 400 µm), brood their larvae inside the maternal tube (Schroeder & Hermans 1975). Brooding inside the host has also been well documented in other commensal polynoids (e.g. *Polynoe antarctica* Kinberg by Fauvel 1916 and *Halosydna brevisetosa* Kinberg by MacGinitie & MacGinitie 1949).

The reproductive features of the scale worm *Branchipolynoe seepensis* are typical of species that brood their offspring. Large lecithotrophic eggs can, however, be viewed either as the result of phylogenetic constraints towards direct development and limited dispersal capabilities or as an advantage for long-distance dispersal in species subjected to recurrent population bottlenecks, especially when the surrounding abyssal waters are poor in food sources for larvae (Young 1994). While previous genetic analyses based on ribosomal and mitochondrial sequences found no genetic differentiation over large spatial scales for either the Atlantic or Pacific *Branchipolynoe* species (Jollivet et al. 1998, Hurtado et al. 2004), suggesting either possible long-distance dispersal or a recent colonization of the sites, Daguin & Jollivet (2005), using highly polymorphic microsatellite markers, revealed that *B. seepensis* displayed genetic breaks along the Mid-Atlantic Ridge between populations ranging from 14 to 35° N. Dispersal of the worm could thus be more limited than previously thought.

The aim of the present study was to perform kinship analyses with the microsatellite markers previously used by Daguin & Jollivet (2005) to determine if the female and juveniles are genetically related inside the same host. Two alternative hypothetical reproductive strategies are possible: (1) females brood the juveniles inside the host; then, the female and juveniles are genetically related as mother/offspring; or (2) eggs and/or juveniles are released from the host and are able to colonize new mussels. In the latter case, the female and juveniles are not expected to show genetic relatedness as mother/offspring. In addition, for the second strategy, 2 alternative scenarios can also be discriminated using kinship analyses: (1) the offspring hatch inside the host, but the juveniles tend to leave their mussel of origin to settle into a new host, and, then, a relationship between juvenile age and kinship is expected within the host with a possible kin structure within the mussel bed at a small spatial scale, or (2) fertilized eggs or larvae are released into the water column to disperse passively with the (deep) bottom currents, and, then, no kinship/size relationship or small-scale spatial structure is expected.

MATERIALS AND METHODS

Sampling. The hydrothermal vent field Lucky Strike is located on the Mid-Atlantic Ridge at the triple junction of the Eurasian, American and African tectonic plates (between 37° 16′ and 37° 19′ N, 32° 15′ and 32° 18′ W, 1700 m deep). A 300 m diameter lava lake governs the hydrothermal system, and the evolution of the associated communities (Desbruyères et al. 2000). The vent sites Tour Eiffel (TE; 37° 17.22′ N, 32° 16.30′ W) and Bairro Alto (BA; 37° 17.37′ N, 32° 17.00′ W) are located on opposite sides of this lake, 900 m apart. Mussel beds *Bathymodiolus azoricus* were sampled on 2 oceanographic cruises: PICO in 1998 and ATOS in 2001. The TE site was sampled twice, once in July 1998 and once in July 2001; and the BA site was sampled once in July 2001. At each site, mussels were collected within a 1 m^2 quadrat using the telemanipulator arms of the submersible 'Nautile' (PICO98 cruise) or the remotely operated vehicle (ROV) 'Victor6000' (ATOS01 cruise), and brought back on board the ship inside a temperature-insulated box. Mussels were opened and carefully examined to collect all *Branchipolynoe seepensis* from the mantle cavity. Scale-worm specimens from a given mussel were then preserved together in 96% ethanol. A total of 299 scale worms was finally obtained out of 117 mussels from the 3 samples (Tour Eiffel 1998 [TE98], Tour Eiffel 2001 [TE01] and Bairro Alto 2001 [BA01]; see Table 1).

Scale-worm biometry. The total length of each scale worm was measured ventrally to the nearest 0.03 mm from the anterior part of the prostomium (evaginated proboscis excluded) to the anus (pygidial appendages excluded) using an ocular micrometer on a binocular microscope (magnification ×25). In addition, chaetigers were counted on each individual in order to determine

Table 1. *Branchipolynoe seepensis* and its host *Bathymodiolus azoricus*. Number of sampled mussels and of female, male and juvenile scale worms at Tour Eiffel (TE98 and TE01) and Bairro Alto (BA01) sites during PICO98 and ATOS01 oceanographic cruises

	B. azoricus	*B. seepensis*			
Site	Total	Females	Males	Juveniles	Total
TE98	36	35	9	24	68
TE01	36	34	8	62	104
BA01	45	50	4	73	127

its developmental stage relative to a fully developed scale worm, which has 20 chaetigerous segments plus 1 achaetous segment. In the present study we report the number of segments as a measure of worm size. The sex of each individual was also determined when possible following the morphological criteria of Jollivet et al. (2000).

DNA extraction and genotyping. Genomic DNA was extracted from the 299 scale worms using a Chelex protocol (Walsh et al. 1991). A small piece of ethanol-preserved tissue (equal to 1 mature female parapodium) was digested in 400 µl of a 5% Chelex100 solution containing 10 mM Tris-HCl, 1 mM EDTA and 25 µl of Proteinase K (20 mg ml^{-1}) and was incubated at 55°C overnight. DNA extracts were vortexed (5 min), centrifuged (10 min at 15 000 × *g*), heated at 95°C for 10 min, and finally stored at –20°C until PCR amplification. These amplifications were carried out for 6 highly polymorphic microsatellite loci (Bs 3E4, Bs 6C3, Bs 3D5, Bs 5H11b, Bs 6D7, Bs 3C8) following the protocol of Daguin & Jollivet (2005). PCR products were then electrophoresed on a denaturing acrylamide 41 cm gel in a Li-Cor NEN Global IR2 DNA analyzer. Individuals that did not amplify at all loci (i.e. 53 ind.) were excluded from the genetic analysis, but scored for estimating null allele frequencies.

Population analysis. The genetic structure of *Branchipolynoe seepensis* was analyzed based on allele frequencies at 6 microsatellite loci using GENETIX 4.05 (Belkhir et al. 2004). Departures from Hardy-Weinberg equilibrium were assessed using the Weir & Cockerham (1984) estimators F and f at each locus for the whole collection and for each sample, respectively, with and without discriminating adults from juveniles. Differentiation tests were performed using the Weir & Cockerham (1984) estimator θ for each locus: (1) between juveniles and adults within samples and (2) between samples without separating individuals into age categories. Departures of F, f and θ from zero were tested by permuting either alleles between genotypes within the whole collection, genotypes within samples, or genotypes between samples, respectively, following the recommendations of Belkhir et al. (2004). Allelic richness (El Mousadik & Petit 1996) was calculated for each locus and site based on the smallest sample size (i.e. at TE98: 54 ind.) with FSTAT 2.9.3 (Goudet 2001). Frequencies of putative null alleles were estimated using both frequencies of individuals that did not yield an amplification (counted as null homozygotes) and the CERVUS 2.0 (Marshall et al. 1998) method, which derives null allele frequencies from heterozygote deficiencies using a Markov-chain Monte Carlo process.

Kinship analysis. In order to establish the marker efficiency in ruling out false parentage, the probability of exclusion was calculated for the 6 microsatellite loci using FSTAT 2.9.3 (Goudet 2001). First, based on allele frequencies characterizing populations, the average level of individual relatedness was estimated at increasing spatial scales within the Lucky Strike hydrothermal vent field from multilocus assignments for: (1) individuals living together within a mussel, (2) individuals from the same quadrat, and (3) the whole sample. The aim of this analysis was to determine whether a kinship structure could be detected. The relatedness coefficient was calculated for each pair of scale worms using KINSHIP 1.2 (Queller & Goodnight 1989). Mean kinship coefficients between the 2 former and the 2 latter spatial scales were tested using Student's t-tests.

Second, several hypotheses of kinship based on probabilities that individuals share alleles by descent through maternal (R_m) or paternal (R_p) inheritance were tested. These hypotheses are: (1) juveniles are direct descendants of the female(s) when located inside the same mussel (R_m = 1, R_p = 0), (2) juveniles living in the same mussel are full siblings (R_m = 0.5, R_p = 0.5). Relative likelihood ratio tests with 1 degree of freedom (Sokal & Rohlf 1981), in which the null hypothesis H_0 (individuals are not related) and the alternative hypothesis H_1 (individuals are related), were used. The r correlation between female/juvenile kinship coefficients and the number of segments of juveniles were tested according to Sokal & Rohlf (1981) to address the age of leaving the mussel of origin.

Finally, the most probable mother was determined for each juvenile from the whole female set using CERVUS 2.0 (Marshall et al. 1998). CERVUS searches only for the most probable mother by excluding the other females using relative likelihood tests (Delta criterion value), similar to those of KINSHIP, at a given threshold value. Parameters used to define the Delta criterion were: 10 000 cycles (number of offspring), 1.000 the proportion of loci typed and 0.010 the proportion of loci mistyped. A search for the most probable mother was first performed inside each 1 m^2 quadrat (TE01 and BA01) with the proportion of candidate parents sampled set to 1.000 under the hypothesis that

juveniles dispersed only to a nearby mussel. A second search for the most probable mother was then performed over the whole female 2001 collection with the proportion of candidate parents set to 0.010 (75 candidate females) under the alternative hypothesis that larva/juvenile dispersal is not limited to the quadrat. The multilocus genotype of a given female was thus compared to that of each juvenile. A female was considered as a potential mother when its genotype was compatible with those of the tested juveniles at all loci (with a relaxed confidence level of 80 and 95%). A relaxed confidence level of 80% was preferred in the global parentage analysis because the alternative 95% confidence is known to be highly stringent and often excludes true kinship in studies where parent and offspring are already known (e.g. in presence of relatives; Marshall et al. 1998). The Poisson likelihood of descent (LOD) distribution was normalized to check whether significant positive assignments fall out of the 95% range of the values, and thus to have more confidence with 'true' kinship.

RESULTS

Biometric analysis

The number of *Branchipolynoe seepensis* per mussel varied from 0 to 13 ind. (Fig. 1), with a median of 1, 2 and 3 ind. mussel^{-1} for TE98, TE01 and BA01, respectively. The lengths of the collected mussels *Bathymodiolus azoricus* ranged from 38.5 to 105 mm. The number of worms per mussel was exactly the same in the TE01 and BA01 quadrats (2.9 worms mussel^{-1}), despite a noticeable difference in the length-frequency distributions of the host. In general, large mussels yielded more scale worms than small ones (data not shown).

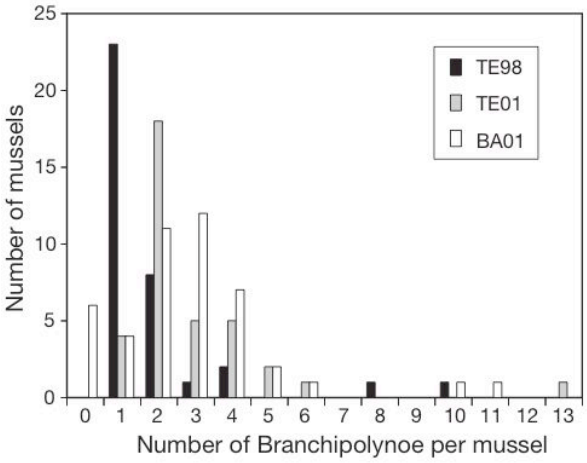


Fig. 1. *Branchipolynoe seepensis* and its host *Bathymodiolus azoricus*. Distribution of the number of individuals per mussel for the sites Tour Eiffel (TE98 and TE01) and Bairro Alto (BA01) during PICO98 and ATOS01 oceanographic cruises

Juveniles, males and females displayed significantly different mean sizes in the 3 samples (Mann-Whitney *U*-test, $p < 0.001$), with the exception of BA01, for which males and females exhibited similar sizes ($p = 0.118$). When genders were compared between samples, the only significant difference was found for females from TE98, which were larger than females from TE01 and BA01 ($p < 0.002$). The number of segments of juveniles ranged from 12 to 20 in our collection. Although there was no difference in the number of segments in juveniles between samples, juveniles from TE01 were significantly larger than those from BA01 ($p = 0.002$).

Population analysis

Multilocus genotypes revealed that both males and females were diploid, with an average number of 3.5 (out of 6) heterozygous loci ind.$^{-1}$. The number of alleles per locus ranged from 4 to 75 alleles for Bs 6D7 in TE01 and Bs 3E4 in BA01, respectively (Table 2). Expected heterozygosities (H_e) varied from 0.460 for Bs 3E4 in TE98 to 0.969 for Bs 3D5 in TE98. The fixation indices *F*, *f* and θ, together with the frequency of the most frequent allele, are presented in Table 2 for each locus. Multilocus *f*-estimates were significantly different from zero in the 3 samples (TE98, $f = 0.193$; TE01, $f = 0.155$; BA01, $f = 0.171$). Only 4 of the 18 single-locus *f* values were not indicative of significant heterozygote deficiencies (Table 2). *F* estimates were significantly different from zero for all loci except Bs 6C3, and θ did not depart from zero for all loci except Bs 3D5. Putative null allele frequency estimates (Table 2) ranged from 0.240 for Bs 3C8 and 0.000 for Bs 6C3, and were very similar with the 2 calculation methods used. The greatest heterozygote deficiencies corresponded to the putative null allele frequencies, but were not correlated with increasing gene diversities.

To avoid bias in their estimation, allele frequencies were checked to see whether they differed between adults and juveniles within samples. If individuals inside a mussel were to be related, some alleles may be more frequent in juveniles than in adults. In our case, frequencies were not significantly different between juveniles and adults ($-0.008 \leq \theta \leq 0.004$). Juveniles, males and females were thus grouped together for the tests of differentiation between samples.

The overall multilocus θ estimate (0.003) was not significantly different from zero ($p > 0.05$ with 1000 permutations), indicating that there was no differentiation among individuals from the 3 quadrats.

Table 2. *Branchipolynoe seepensis*. Number of individuals in samples used for kinship analyses (N), observed number of alleles (N_a), frequency of the most frequent allele (f_a), expected heterozygosity (H_e), allelic richness (R_s) and fixation index (f), at the Tour Eiffel (TE98 and TE01) and Bairro Alto (BA01) sites during PICO98 and ATOS01 oceanographic cruises. Null allele frequencies estimated for each locus from the non-amplified homozygous individual method (A) and the CERVUS 2.0 method (B). Fixation indices (F and θ) for each locus were estimated according to Weir & Cockerham (1984). NS: not significantly different from zero; $^{\bullet}p < 0.05$; $^{\bullet\bullet}p < 0.01$; $^{\bullet\bullet\bullet}p < 0.001$

Parameter	Microsatellite locus					
	Bs 5H11b	Bs 3E4	Bs 3D5	Bs 6C3	Bs 6D7	Bs 3C8
TE98 (N = 54)						
N_a	6	55	11	20	5	5
f_a	0.401	0.086	0.727	0.140	0.673	0.629
H_e	0.667	0.969	0.460	0.920	0.487	0.544
R_s	6.000	46.616	8.999	17.945	3.982	5.000
f	$0.301^{\bullet\bullet\bullet}$	$0.139^{\bullet\bullet\bullet}$	0.050^{NS}	$0.096^{\bullet\bullet\bullet}$	$0.169^{\bullet}$	$0.464^{\bullet\bullet\bullet}$
TE01 (N = 78)						
N_a	5	63	13	24	4	5
f_a	0.505	0.097	0.564	0.138	0.670	0.680
H_e	0.604	0.965	0.645	0.927	0.488	0.499
R_s	4.679	48.049	11.126	20.729	3.992	4.692
f	$0.352^{\bullet\bullet\bullet}$	$0.054^{\bullet\bullet\bullet}$	$0.181^{\bullet\bullet\bullet}$	-0.019^{NS}	$0.172^{\bullet\bullet}$	$0.384^{\bullet\bullet\bullet}$
BA01 (N = 114)						
N_a	5	75	13	20	8	5
f_a	0.500	0.073	0.488	0.112	0.673	0.609
H_e	0.591	0.974	0.711	0.927	0.491	0.562
R_s	4.330	50.870	11.086	17.514	5.597	4.997
f	$0.348^{\bullet\bullet\bullet}$	0.019^{NS}	$0.346^{\bullet\bullet\bullet}$	-0.014^{NS}	$0.117^{\bullet}$	$0.376^{\bullet\bullet\bullet}$
All pop.						
A	0.160	0.066	0.197	0.066	0.208	0.208
B	0.190	0.017	0.151	0.000	0.081	0.235
F	$0.314^{\bullet\bullet\bullet}$	$0.051^{\bullet\bullet\bullet}$	$0.284^{\bullet\bullet\bullet}$	-0.001^{NS}	$0.146^{\bullet\bullet\bullet}$	$0.409^{\bullet\bullet\bullet}$
θ	-0.001^{NS}	0.000^{NS}	$0.018^{\bullet\bullet}$	0.000^{NS}	-0.005^{NS}	-0.003^{NS}

Kinship analysis

The probability of exclusion calculated over the 6 loci was 0.989, providing robust assignments of juveniles for our kinship study. Mean kinship coefficients (R), calculated at different spatial scales (within mussels, within sites and between sites) between females and juveniles, were low, indicating that most scale worms were not related (Fig. 2). These means were not statistically different at the intra-mussel versus intra-site spatial scales (p = 0.458), and at the intra- versus inter-site spatial scales (p = 0.064).

A few pairs of individuals, however, displayed significant kinship coefficients. Out of these related pairs of individuals, the likelihood-ratio test obtained from 1000 simulated pairs indicated that 'true' mother/offspring relationships (R_m = 1, R_p = 0) represented only 5.9% (1/17), 0% (0/34) and 1.5% (1/66) of the individual pairings in TE98, TE01 and BA01, respectively (Table 3). In nearly all cases, the female was not the mother of juveniles living in the same mussel. Moreover, no correlation was found between the female/juvenile kinship coefficient R and the juvenile's number of segments (p = 0.582, Fig. 3). However, 2 juveniles appeared to be related to the female living inside the same mussel. Similarly, none of the males were related to a juvenile inside the same mussel (Table 3). The percentages of full siblings between juveniles represented 0% (0/15), 3.7% (2/54) and 1.6% (1/61) in TE98, TE01 and BA01, respectively (Table 3). These percentages were low (3 pairs) and close to those found for the mother/offspring relationships. Consequently, juveniles were generally not full siblings, and related individuals within mussels were very rare.

Because parents and offspring were apparently not found in the same mussel, we then used the CERVUS software to search for the most probable mother among females living in the same quadrat as the juveniles. Within the BA01 sample, 10 juveniles out of 61 had a multilocus genotype compatible with a female from the same quadrat with a confidence level of 80%. Within the TE01 sample, 7 juveniles out of the 44 displayed a similar degree of kinship with a female from the same quadrat (α = 80%). However, the number of juveniles correctly assigned to a female

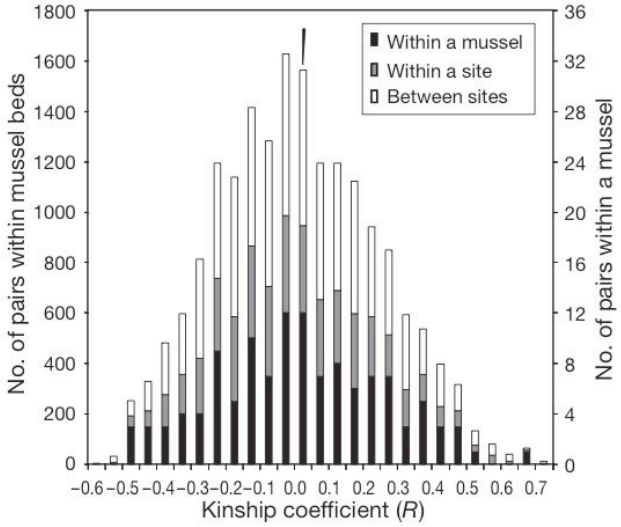


Fig. 2. *Branchipolynoe seepensis*. Superimposed kinship coefficients histograms showing the relatedness distributions at different levels of a spatial hierarchy: within a mussel (right axis), within a site and between sites (left axis). Average kinship coefficients (triangle) belong to the same size class and are 0.016, 0.001 and 0.009, respectively, for the 3 spatial scales

Table 3. *Branchipolynoe seepensis*. Proportion of individuals for which relative likelihood ratios favor the null hypothesis of relatedness (**bold**), relative likelihood ratios and Type II error (*italics*) at the exclusion threshold α = 0.01 for 1000 simulated pairs

Site	Mother/ descendant	Father/ descendant	Full-sibling juveniles
TE98	**1/17 (5.9%)**	**0/4 (0%)**	**0/15 (0%)**
	1.370	1.370	1.130
	0.081	*0.081*	*0.280*
TE01	**0/34 (0%)**	**0/4 (0%)**	**2/54 (3.7%)**
	0.932	0.932	1.060
	0.017	*0.017*	*0.280*
BA01	**1/66 (1.5%)**	**0/4 (0%)**	**1/61 (1.6%)**
	1.220	1.220	1.100
	0.035	*0.035*	*0.290*
Total	**2/117 (1.7%)**	**0/12 (0%)**	**3/130 (2.3%)**

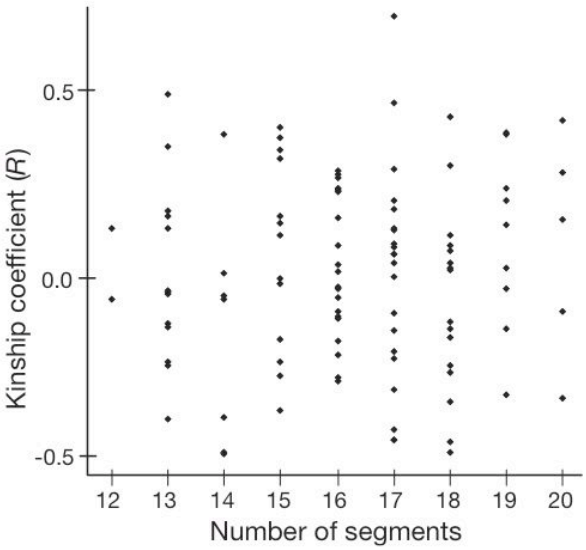


Fig. 3. *Branchipolynoe seepensis*. Relationship between the kinship coefficient (*R*) and the number of segments of juveniles showing an absence of correlation between the age of juveniles and the kinship structure

from the same quadrat was not very different from the number of juveniles correctly assigned to a female from the other quadrat. This number fell to only 2 (1 per quadrat) at a confidence level of 95% and did not correspond to the 'true' intra-mussel offspring/mother associations found in the kinship analyses. Among these 17 assigned juveniles, the 2 juveniles assigned at a 95% threshold and 8 juveniles assigned at a 80% threshold (also containing the juvenile from BA01 significantly related to a female within the same mussel using KINSHIP 1.2) were out of the 95% normalized LOD distribution and represent the most extreme Delta criterion values (Fig. 4). We then performed a search for the most probable mother for our set of juveniles among the whole female 2001 collection, taking into account that we only sampled a very small proportion of females from the entire population. This second analysis did not reveal any significant maternal relationship between juveniles and females within and between the 2 sites and could not exclude the hypothesis that the assigned juveniles may be the offspring of un-sampled females.

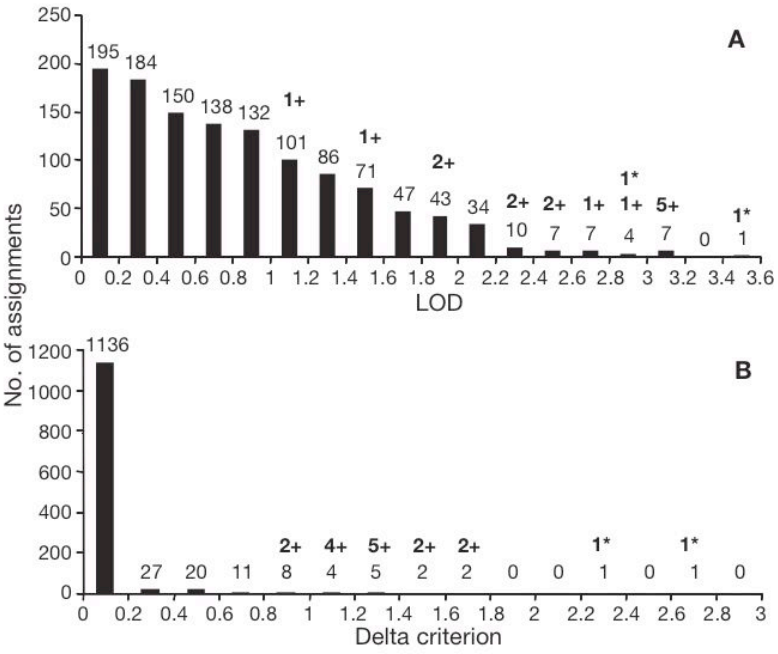


Fig. 4. *Branchipolynoe seepensis*. Distributions of the (A) LOD (likelihood of descent) and (B) Delta criterion values from all juvenile/female pairs obtained within each quadrat. The number of female/juvenile assignments performed for each class is presented above the class. **Bold**: significant assignments obtained from CERVUS (*95%; + 80% threshold). Individuals with a LOD >2.43 are out of the 95% distribution and are the most extreme Delta criterion values, which strengthens their probability to represent 'true' identity by descent. LOD: the best likelihood value that a juvenile is descended from a given female; Delta criterion: the difference between the 2 best LOD obtained from the set of females examined

DISCUSSION

Dispersal strategies of interacting species are influenced both by the proportion of suitable patches and the

level of patch connectivity in a fragmented landscape (Hiebeler 2007). The hydrothermal vent environment constitutes a very unusual landscape in which there are few suitable patches (very few fields), but in which connectivity is very high (numerous vent emissions within a field). Moreover, the rate of local extinction within a field is very high and may lead to bursts of colonization (Slatkin 1977). It is thus interesting to note that, in this kind of landscape, coevolving species are likely to coexist if they adopt a flexible dispersal strategy. This study represents the first attempt to perform parentage analyses in a deep-sea hydrothermal vent commensal species and to test whether brooding inside the host affects dispersal.

If *Branchipolynoe seepensis* females brood their juveniles inside the host, then their progeny must develop inside the host mussel's mantle cavity. The female and juveniles found inside the same host should then be related. However, our relatedness coefficients between juveniles and females were very low. In addition, although significant pairs of 'true' mother/offspring or full siblings inside the same mussel have been detected, these cases were very rare. In the light of our parentage analysis, we can reject the hypothesis of maternal caring and young brooding. The stage at which the offspring are released from the host, however, remains unclear. Tracking the stage at which a propagule is dispersing is not an easy task. Several indirect observations about developmental stage and relatedness or juvenile population structure could, however, help to shed some light on this issue. Juveniles displayed a number of segments ranging from 12 to 20, and this number was not correlated with the level of kinship. Consequently, offspring may be released prior to reaching the minimum developmental stage found in our worm collection. Offspring are thus most likely to be released at a very early stage, as fertilized eggs, larvae (if they exist, down to 10 segments for polychaetes), or very young juvenile stages. This is in agreement with previous studies on polynoid commensals, which, in most cases, have pelagic larvae (Britayev 1991) and tend to be aggressive and territorial towards conspecific individuals (Britayev 1991, Ruff 1991, Britayev & Zamishliak 1996). Dimock (1974) observed a similar pattern *in vivo* in the symbiotic polychaete *Arctonoe pulchra*, which forces congeners to leave the host limpet and move elsewhere in the vicinity. This may explain both genetic unrelatedness, small males and the observed worm distribution inside hosts.

Interestingly, our results also indicated a lack of genetic differentiation between the TE01 and BA01 quadrats. This is in agreement with the low number of juveniles assigned to females from the same quadrat (15.90% for TE01 and 16.39% for BA01), suggesting that a great proportion of juveniles are dispersed outside their quadrat of origin. The absence of apparent differentiation is not surprising given the relatively short distance between the 2 mussel beds (900 m) and because only 1 migrant generation^{-1} is enough to counteract genetic drift (Lewontin 1974). In addition, using the same microsatellite markers, Daguin & Jollivet (2005) suggested that *Branchipolynoe seepensis* are genetically differentiated over large portions of the Mid-Atlantic Ridge. *B. seepensis* may therefore not be able to disperse very far.

Although gene flow does not seem restricted to a given mussel bed but to a rift portion, leading to a certain flexibility in the dispersal distance, both the occurrence of heterozygote deficiencies at all loci and the presence of juveniles significantly related to a nearby female requires explanation. Two alternative hypotheses will be discussed: (1) the very few significant relatedness assignments represent a true mother/offspring kinship, and thus at least a small proportion of juveniles settle near their host of origin (nearby colonization), or (2) the very few significant relatedness assignments are only the result of chance or from a lack of power of the microsatellite markers, and thus *Branchipolynoe seepensis* is not likely to re-enter its host of origin (larval export).

(1) Significant offspring/mother pairs represent a 'true' kinship. On the one hand, based on external positions of nearly all the significant pairs in the LOD and Delta criterion value distributions, one can accept the hypothesis that at least a few juveniles are truly related to a nearby female. The 2 significant offspring/mother pairs found inside the same host indeed fall well outside the 95% LOD distribution together with pairs of individuals sampled in close proximity (same quadrat). Finding 2 significantly related pairs inside a given host does not seem easily explainable from a re-colonization process. Generally speaking, positively buoyant embryos or larvae would be entrained by bottom currents and subsequently passively carried away from the mussel bed. Even if a small proportion of lecithotrophic larvae was trapped within the hydrothermal plume by convective heat fluxes, ca. 200 m above the sea floor (Kim et al. 1994, Kim & Mullineaux 1998, Mullineaux et al. 2005), the chance that larvae would re-enter their mother's mussel or quadrat is close to zero, given the millions of mussels in the several hundred square meters of the Lucky Strike vent field. This may be a strong indication that a small proportion of juveniles stays close to their parents. Although we cannot rule out the hypotheses of null alleles or Wahlund effect, heterozygote deficiencies could also be due to local inbreeding, which would lead to such a within-population structure and also reinforce the nearby settlement hypothesis. As a consequence, offspring may be dispersed over variable

distances depending on such factors as the prevalence of worms in mussels (intra-specific competition) or nutritional limitation (cessation of vent activity). Such flexibility in dispersal distance depends on the distribution of suitable habitats for settlement (Gaines & Roughgarden 1985) and could result in a small mean dispersal distance (e.g. within a single mussel bed) with a long tail of much greater distances (e.g. a whole vent sector), the shape of the distribution being strongly dependent on the selective pressures associated with habitat fragmentation.

(2) Significant kinship coefficients are due to chance. On the other hand, despite a few significant kinship coefficients between individuals from the same host, the CERVUS software failed to find significant pairs under the hypothesis that genotyped individuals represent a very small proportion of the sampled worms in the whole population. Although the probability of exclusion was comparable to those in previous kinship studies (e.g. Avise et al. 2002), Marshall et al. (1998) demonstrated that some unrelated individuals may share at least 1 identical allele locus^{-1} just by chance or homoplasy. In the present study, the number of significant pairs was very low, so the hypothesis that larvae settle near the maternal host was only weakly supported. The high heterozygote deficiencies observed could also be attributable to null alleles, which is not an uncommon phenomenon when using highly polymorphic microsatellite markers (e.g. Taberlet et al. 1996). In our case, the frequencies of null alleles calculated using 2 different methods were almost identical. Consequently, heterozygote deficiencies may be, at least in part, explained by the non-amplification of some alleles. Consequently, if none of the juveniles are related to a nearby female, then the most likely explanation of the lack of differentiation between quadrats is to consider that *Branchipolynoe* sp. is able to disperse sufficiently far away from its host to prevent any re-colonization stage at the scale of the Lucky Strike vent field. Such a strategy contrasts with previous findings about the reproductive biology of this worm (large yolky eggs, pairing, internal fertilization of the oocytes), which usually typifies direct developers in polychaetes (see Jollivet et al. 2000). However, as stated by Young (1994), large yolky eggs may be greatly advantageous for long-distance dispersers that develop slowly in cold oligotrophic zones. From these results, most juveniles were not the offspring of females found within the same mussel, but at least a few individuals were related. The most probable explanation is, therefore, that *B. seepensis* must display a nearby dispersal strategy in which eggs/larvae/juveniles disperse into their immediate vicinity (on a square meter scale), with some kind of dispersal flexibility to colonize new empty patches of host mussels.

Acknowledgements. We are particularly indebted to the captains, chief scientists Daniel Desbruyères (PICO98) and Pierre-Marie Sarradin (ATOS01) and 'Nautile' and 'Victor6000' crews of the PICO98 and ATOS01 oceanographic diving cruises. We are also grateful to Fabienne Chausson, who sampled material for us during PICO98. We thank Carolyn Engel-Gautier, Denis Roze, Myriam Valero and anonymous referees for valuable comments and editorial suggestions on the manuscript. This work was supported by the French programme DORSALES and both the European AMORES (coordinator Daniel Desbruyères: MAST III-CT95-0040) project of the Marine Science and Technology programme and the EU-funded VENTOX project (coordinator David R. Dixon: EVK3 CT-1999-00003).

LITERATURE CITED

Avise JC, Fiumera AC, Jones AG, Walker D, DeWoody JA (2002) Genetic mating systems and reproductive natural histories of fishes: lesson for ecology and evolution. Annu Rev Genet 36:19–45

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) Genetix 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, Adaptations, UMR 5000, Université de Montpellier 2. Available at: www.genetix.univ-montp2.fr/genetix/genetix.htm

Britayev TA (1991) Life cycle of symbiotic scaleworm *Artonoe vittata* (Polychaete: Polynoidae). Ophelia 5:305–312

Britayev TA, Zamishliak EA (1996) Association of the commensal scaleworm *Gastrolepidia clavigera* (Polychaeta: Polynoidae) with holothurians near the coast of South Vietnam. Ophelia 45:175–190

Chevaldonné P, Jollivet D, Feldman RA, Desbruyères D, Lutz RA, Vrijenhoek RC (1998) Commensal scale-worms of the genus *Branchipolynoe* (Polychaeta: Polynoidae) at deep-sea hydrothermal vents and cold seeps. Cah Biol Mar 39:347–350

Daguin C, Jollivet D (2005) Development and cross-amplification of nine polymorphic microsatellite markers in the deep-sea hydrothermal vent polychaete *Branchipolynoe seepensis*. Mol Ecol Notes 5:780–783

Desbruyères D, Almeida A, Biscoito M, Comtet T and others (2000) Distribution of hydrothermal vent communities along the northern Mid-Atlantic Ridge. Dispersal vs. environmental control: a review. Hydrobiologia 440:201–216

Dimock RV Jr (1974) Intraspecific aggression and the distribution of a symbiotic polychaete on its host. In: Vernberg WB (ed) Symbiosis in the sea. South Carolina University Press, Columbia, p 29–44

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

Fauvel P (1916) Annélides Polychètes des Iles Falkland recueillies par M. Rupert Vallentin Esq. (1902-1910). Arch Zool Exp Gén 5:417–482

Gaines S, Roughgarden J (1985) Larval settlement rate: a leading determinant of structure in an ecological community of the marine intertidal zone. Proc Natl Acad Sci USA 82:3707–3711

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (Version 2.9.3). Update from Goudet, 1995. FSTAT (Version 1.2): a computer program to calculate *F*-statistics. J Hered 86:485–486. Available at: www2.unil.ch/popgen/softwares/fstat.htm

Hiebeler DE (2007) Competing populations on fragmented landscapes with spatially structured heterogeneities: improved landscape generation and mixed dispersal strategies. J Math Biol 54:337–356

Hurtado LA, Lutz RA, Vrijenhoek RC (2004) Distinct patterns of genetic differentiation among annelids of eastern Pacific hydrothermal vents. Mol Ecol 13:2603–2615

Jollivet D (1996) Genetic diversity and biodiversity at deep-sea hydrothermal vents: an overview. Biodivers Conserv 5:1619–1653

Jollivet D, Comtet T, Chevaldonné P, Hourdez S, Desbruyères D, Dixon DR (1998) 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, Empis A, Baker MC, Hourdez S and others (2000) Reproductive biology, sexual dimorphism, and population structure of the deep sea hydrothermal vent scale-worm, *Branchipolynoe seepensis* (Polychaete: Polynoidae). J Mar Biol Assoc UK 80:55–68

Kim SL, Mullineaux LS (1998) Distribution and near-bottom transport of larvae and other plankton at hydrothermal vents. Deep-Sea Res II 45:423–440

Kim SL, Mullineaux LS, Helfrich KR (1994) Larval dispersal via entrainment into hydrothermal vent plumes. J Geophys Res 99:12655–12665

Lewontin RC (1974) The genetic basis of evolutionary change. Columbia University Press, New York

MacGinitie GE, MacGinitie N (1949) Natural history of marine animals. McGraw-Hill Book Company, New York

Marshall TC, Slate J, Kruuk L, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations, CERVUS 2.0. Mol Ecol 7:639–655. Available at: http://helios.bto.ed.ac.uk/evolgen/cervus/cervus.html

Mullineaux LS, Mills SW, Sweetman AK, Beaudreau AH, Metaxas A, Hunt HL (2005) Vertical, lateral and temporal structure in larval distributions at hydrothermal vents. Mar Ecol Prog Ser 293:1–16

Nichols RA, Hewitt GM (1994) The genetic consequences of long-distance dispersal during colonization. Heredity 72: 312–317

Queller DC, Goodnight KF (1989) Estimating relatedness using genetic markers. Evolution 43:258–275

Ruff RE (1991) A new species of *Bathynoe* (Polychaeta: Polynoidae) from the Northeast Pacific Ocean commensal with two species of deep-water asteroids. Ophelia 5:219–230

Schroeder PC, Hermans CO (1975) Annelida: Polychaeta. In: Giese AC, Pearse JS (eds) Reproduction of marine invertebrates. Academic Press, New York, p 1–214

Slatkin M (1977) Gene flow and genetic drift in a species subject to frequent local extinctions. Theor Popul Biol 12: 253–262

Sokal RR, Rohlf FJ (1981) Biometry. The principles and practice of statistics in biological research. Freeman, New York

Taberlet P, Friffin S, Goossens B, Questiau S, Manceau V (1996) Reliable genotyping of samples with very low DNA quantities using PCR. Nucleic Acids Res 24:3189–3194

Thorson G (1936) The larval development, growth and metabolism of Arctic marine invertebrates, compared with those of other seas. Medd Gronl 100:1–155

Tyler PA, Young CM (1999) Reproduction and dispersal at vents and cold seeps. J Mar Biol Assoc UK 79:193–208

Walsh PS, Petzger DA, Higuchi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 10:506–513

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358–1370

Young CM (1994) A tale of two dogmas: the early history of deep-sea reproductive biology. In: Young CM, Eckelberger KJ (eds) Reproduction, larval biology and recruitment of the deep-sea benthos. Columbia University Press, New York, p 1–25

Editorial responsibility: Judith Grassle, New Brunswick, New Jersey, USA

Submitted: November 23, 2007; Accepted: May 14, 2008
Proofs received from author(s): July 4, 2008

Titre de thèse : Phylogéographie comparée des espèces hydrothermales de la dorsale du Pacifique oriental

L'histoire démographique des espèces hydrothermales profondes et leurs échanges génétiques le long de la dorsale du Pacifique oriental ont été analysés en combinant des approches classiques d'analyses de fréquences alléliques entre populations avec des méthodes d'analyse plus récentes basées sur la théorie de la coalescence. Une approche de phylogéographie comparée sur sept espèces (un bivalve, trois gastéropodes, trois polychètes), réalisée sur le gène mitochondrial Cytochrome Oxydase I, a permis de révéler une barrière aux flux de gènes, commune à toutes les espèces entre le Nord et le Sud EPR datant d'environ 1,3 Ma et probablement liée à la formation de failles transformantes entre 0°-7°25'N. Cette séparation de faune s'est accompagnée d'une expansion démographique généralisée plus récente (< 0,5 Ma) au Sud. Une approche multi-locus a ensuite été effectuée sur trois taxons cibles aux traits d'histoire de vie contrastés : le polychète *A. pompejana*, le bivalve *B. thermophilus* et le gastéropode *L. elevatus* (comprenant deux espèces cryptiques). Cette approche a permis de confirmer les résultats obtenus par l'approche multi-spécifique. L'hybridation entre lignées divergentes, détectée chez ces trois taxons, suggère une semi-perméabilité de la barrière aux flux de gènes. L'étendue de la zone de contact secondaire dépend du taxon considéré selon ses traits d'histoire de vie, et des forces de sélection agissant au locus considéré. Les patrons de distributions phylogéographique et biogéographique mis en évidence sur les espèces et les communautés s'accordent pour séparer la dorsale du Pacifique oriental en deux provinces biogéographiques différentes de part et d'autre de l'Equateur.

Mots clés : phylogéographie, vicariance, allopatrie, expansion démographique, hybridation, introgression, sélection, milieu hydrothermal

PhD title: Comparative phylogeography of deep-sea hydrothermal vents species along the East Pacific Rise

Demographic history and past gene flow between populations of deep-sea hydrothermal vent species along the East Pacific Rise (EPR) were assessed using a phylogeographic approach combining a classical allele-frequency-based analysis and more sophisticated coalescence-based methods. A comparative phylogeographic COI-based analysis of seven species (one bivalve, three gastropods, three polychaetes) showed the occurrence of a vicariant event associated with the raise of a physical barrier between the Northern and Southern EPR, 1,3 Mya probably due to the formation of transform faults in the equatorial region (0°-7°25S). This fauna separation was then followed by a more recent (< 0.5 Myr) population expansion in the South for all species. A multilocus approach was then performed on three targeted taxa with different life trait histories (the polychaete: *A. pompejana*, the bivalve: *B. thermophilus,* and two cryptic species of the gastropod *L. elevatus*). This study confirmed results obtained from the previous multi-species approach. Hybridization between divergent lineages from each of the three taxa suggested that the barrier is semi-permeable. The range of the secondary contact zone was however variable depending on life-history traits of each species and selective processes acting at a given gene. Phylogeographic and biogeographical patterns derived from species and assemblages are in agreement with the hypothesis of two biogeographical EPR provinces across the Equator.

Keywords: phylogeography, vicariance, allopatry, demographic expansion, hybridization, introgression, selection, deep-sea hydrothermal vents