

# Institut de la Génomique Marine

## "Bivalvomix": Génomique évolutive des Bivalves marins

### Rapport final (2004-2006)

Nicolas Bierne, François Bonhomme et Matthieu Faure  
Laboratoire Génome, Populations, Interaction, Adaptation  
UMR 5171 CNRS-IFREMER-Université Montpellier II  
Station Méditerranéenne de l'Environnement Littoral  
1 Quai de la Daurade 34200 Sète  
Tel : +33 (0)4 67 46 33 75  
Fax : +33 (0)4 67 46 33 99  
n-bierne@univ-montp2.fr

Pierre Boudry, Sylvie Lapègue et Christopher Sauvage  
IFREMER, Laboratoire de Génétique et Pathologie  
17390 La Tremblade

Dario Moraga, Isabelle Boutet et Elise David  
LEMAR - UMR6539  
Technopôle Brest-Iroise  
Place N. Copernic  
29280 Plouzané

Didier Jollivet, Arnaud Tanguy et Baptiste Faure  
Laboratoire Evolution et Génétique des Populations Marines  
UMR-CNRS 7127  
Station Biologique de Roscoff, B.P.74  
29682 ROSCOFF cedex

Patrice David  
CEFE - UPR 9056  
1919 route de Mende  
34293 Montpellier cedex 5



# Sommaire

I- INTRODUCTION .....	3
I.1 Cadre scientifique .....	3
I.2 Contexte humain .....	4
II- CARACTERISATION DES GENOMES : EFFETS DES NIVEAUX D'EXPRESSION ET DE LA STRUCTURE EN ISOCHORE SUR LE BIAIS D'USAGE DU CODE DES GENOMES DES BIVALVES MARINS ET DES ANIMAUX EN GENERAL. ....	5
III- UNE METHODOLOGIE INEDITE POUR L'ANALYSE DU POLYMORPHISME DE SEQUENCE : LE CLONAGE PAR MARQUAGE RE-CAPTURE. ....	8
IV- RESULTATS CHEZ LES MOULES DU GENRE <i>MYTILUS</i> . ....	9
IV.1- Effet d'un balayage sélectif dans des populations de moules spatialement structurée. ....	9
IV.2- Introgression adaptative dans la zone hybride de <i>Mytilus</i> . ....	11
IV.3- Absence de barrière, évolution neutre et recombinaison au locus de la Glucanase .....	12
IV.4- Données moléculaires acquises mais non encore analysées .....	14
V- RESULTATS CHEZ LES MODIOLES HYDROTHERMALE DU GENRE <i>BATHYMODIOLUS</i> . ....	14
V.1- Objectifs de l'étude .....	14
V.2- Approche multigénique et effort d'échantillonnage .....	14
V.3- Analyse de la divergence et du polymorphisme .....	15
VI- RESULTATS CHEZ LES HUITRES DU GENRE <i>CRASSOSTREA</i> . ....	23
VI.1- Le modèle <i>Crassostrea gigas</i> / <i>C. angulata</i> .....	23
VI.2- Génétique de la différenciation entre <i>Crassostrea angulata</i> et <i>C. gigas</i> .....	24
VI.3- Analyse des contraintes sélectives .....	27
VII- CONCLUSIONS .....	29
REFERENCES CITEES .....	30

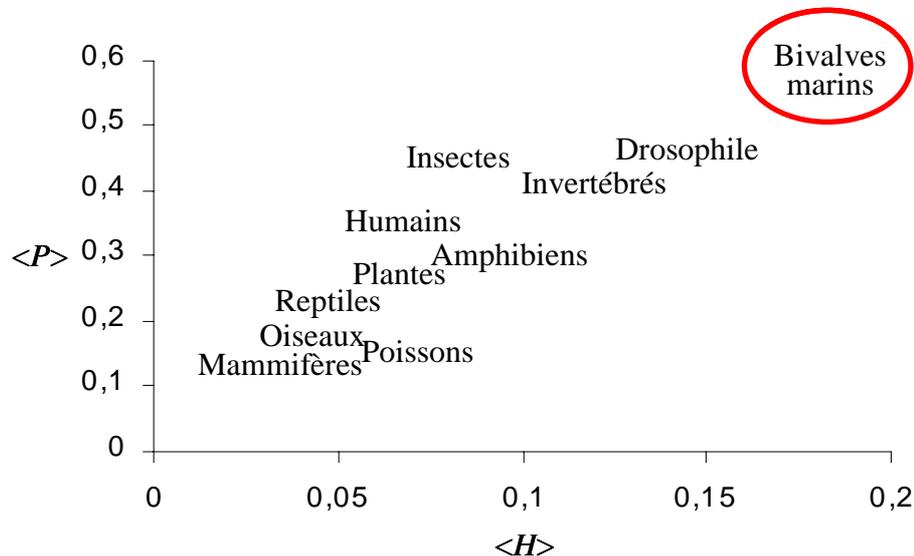
# I- Introduction

## I.1 Cadre scientifique

Les généticiens des populations ont longtemps dû se contenter de l'analyse de marqueurs neutres pour inférer indirectement les processus sélectifs. Les progrès récents des outils de séquençage permettent aujourd'hui de pouvoir analyser les variations d'un grand nombre de gènes et ainsi de pouvoir détecter les gènes voire les acides aminés qui sont sous l'emprise directe de la sélection (Yang & Bielawski, 2000). Pour analyser les processus sélectifs agissant sur les mutations et discriminer les différents facteurs agissant sur leur évolution (sélection ou dérive génétique), il est essentiel de comparer des données de polymorphisme entre génomes d'une même espèce/population et de divergence entre espèces/populations (Kimura, 1983). Récemment, des données de polymorphisme sont devenues disponibles pour un nombre suffisant de gènes (>10) chez quelques espèces modèles comme l'homme (Fay *et al.*, 2001; Bustamante *et al.*, 2005), la drosophile (Bierne & Eyre-Walker, 2004) ou l'arabette (Bustamante *et al.*, 2002). Certaines inférences ont donné des résultats surprenants, suggérant un rôle bien plus important de l'adaptation dans la divergence entre deux espèces que l'attendu de l'hypothèse neutraliste (revue dans Eyre-Walker, 2006). Cependant d'autres résultats suggèrent que la taille efficace des populations ( $N_e$ ) pourrait être un facteur limitant de l'évolution adaptative (Bustamante *et al.*, 2002), même si son rôle sur la variation génétique non-neutre est encore mal comprise (Gillespie, 1999; Bazin *et al.*, 2006). Enfin, le nombre de sites polymorphes dans un gène donné est souvent trop petit pour pouvoir faire des inférences gène par gène (mis à part peut être chez certains gènes de la drosophile) et seule une estimation globale peut être obtenue en compilant les résultats de plusieurs gènes (Bierne & Eyre-Walker, 2004). S'il existait des séquences ultra-denses en sites polymorphes, des estimations fiables pourraient être réalisées à l'échelle du gène.

Les génomes des Bivalves marins sont encore méconnus mais leur étude est un enjeu important car les caractéristiques originales de ces espèces permettront d'élargir ou de réviser les paradigmes établis en milieu terrestre. Entre autre, les Bivalves marins nous procurent une opportunité unique d'explorer les valeurs extrêmes d'un paramètre fondamental en génétique évolutive : la taille efficace des populations ( $N_e$ ). La diversité enzymatique observée chez les Bivalves marins est la plus élevée des Eucaryotes (Figure 1, Solé-Cava & Thorpe, 1991; López *et al.*, 2001). De fortes diversités ont également été rapportées avec des marqueurs microsatellites (Huvet *et al.*, 2000a; Launey *et al.*, 2002) ou introniques (Daguin *et al.*, 2001; Bierne *et al.*, 2003). Une taille efficace élevée a aussi pour corollaire un fardeau génétique (quantités de mutations délétères ségrégeantes, Bataillon & Kirkpatrick, 2000) qui compte parmi les plus forts jamais estimés chez des animaux (Bierne *et al.*, 1998; Launey & Hedgecock, 2001). Cependant, quasiment aucune donnée de polymorphisme de

séquence d'ADN codant n'était disponible chez les Bivalves lorsque notre projet a été initié. De telles données de séquences sont les seules qui pourront simultanément nous permettre d'inférer précisément la contribution relative des forces évolutives agissant sur les gènes de ces espèces et en retour de comprendre les raisons de leur hyper-diversité.



**Figure I.1** : Représentation schématique de la diversité enzymatique de différents groupes d'animaux ou de plantes. <H> : Hétérozygotie. <P> : Proportion de locus polymorphe. Adapté d'après Nevo (1978) et Solé-Cava & Thorpe (1991).

## I.2 Contexte humain

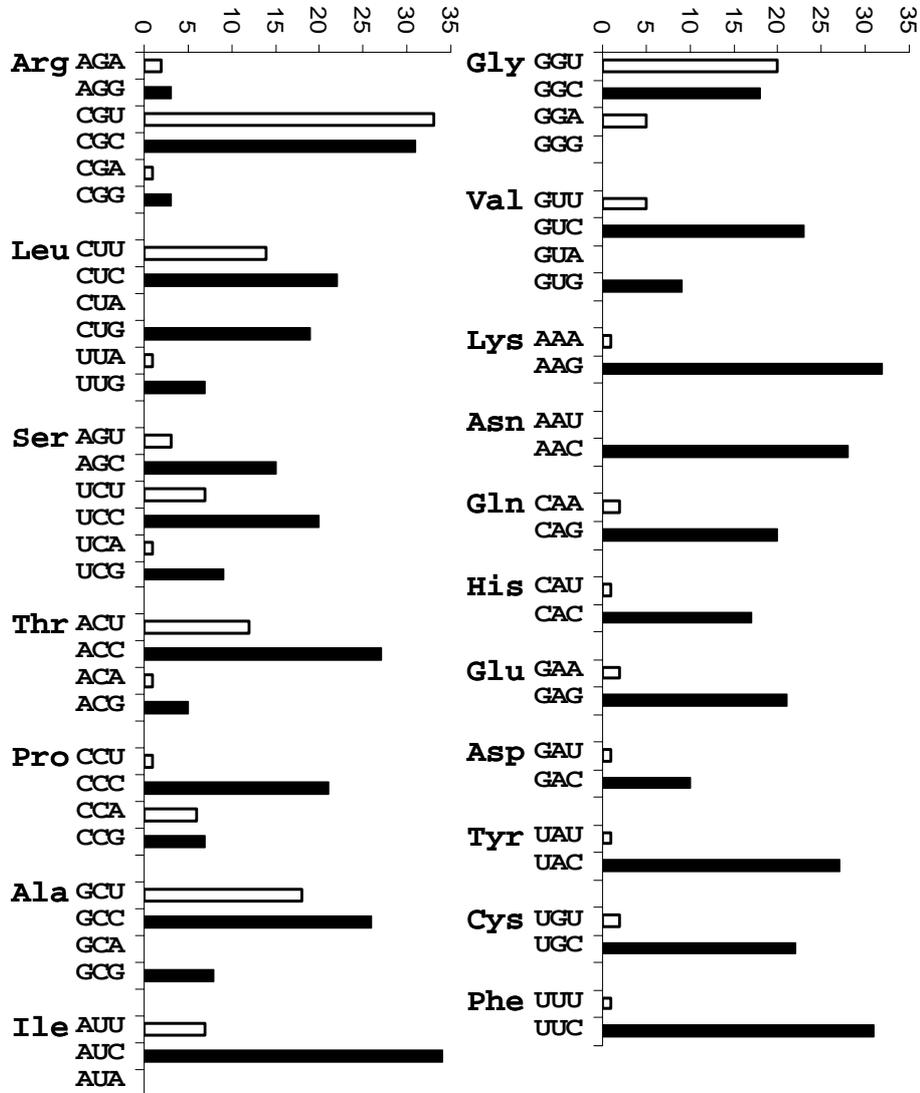
Le projet "Bivalvomix" a regroupé cinq équipes françaises qui se proposaient d'analyser le polymorphisme et la divergence de séquences d'ADN dans trois genres de Bivalves marins: *Mytilus*, *Crassostrea* et *Bathymodiolus*. Ce projet avait pour but de mieux connaître le génome de ces Bivalves, de confronter les hypothèses neutres et sélectives pour expliquer leur diversité, ainsi que de qualifier et quantifier leur diversité fonctionnelle (gènes d'adaptation ou de spéciation). Le projet a débuté en Janvier 2004. Deux thèses directement liées au projet ont démarré en Novembre 2004: une au laboratoire GPIA à Sète (Mathieu Faure) qui se focalise sur les moules du genre *Mytilus* et une autre au laboratoire EGPM à Roscoff (Baptiste Faure) portant sur les modioles hydrothermales du genre *Bathymodiolus* et les huîtres du genre *Crassostrea*. Deux autres thèses ont pu par la suite se rapprocher du projet, une au LEMAR de Brest (Elise David) et une seconde au laboratoire Ifremer LGP de La Tremblade (Christopher Sauvage), toutes deux sur le modèle *Crassostrea*. L'avancement des projets de recherche étant intimement lié aux travaux des thèses, l'état actuel des productions est en accord avec le déroulement de ces thèses qui se termineront à l'automne 2007, c'est-à-dire que les données moléculaires ont été produites et le travail d'analyse de données et de rédaction d'articles scientifiques sont en cours.

## **II- Caractérisation des génomes : Effets des niveaux d'expression et de la structure en isochore sur le biais d'usage du code des génomes des Bivalves marins et des animaux en général.**

Chaque acide aminé est codé par un codon de trois nucléotides. Il existe soixante quatre codons pour les vingt acides aminés et le non-sens (terminaison de la traduction ou codon "stop"). De ce fait, le code génétique est dit redondant ou dégénéré. Cependant il a été mis en évidence que ces codons n'étaient pas utilisés avec la même fréquence : certains sont plus utilisés que d'autres et induisent un biais d'usage du code (Eyre-Walker, 2002). Les connaissances dans ce domaine sont encore limitées, seules des études sur des espèces modèles ont été menées jusqu'à présent (Akashi & Eyre-Walker, 1998; Duret, 2002). Ce biais peut être expliqué de deux manières : (i) un biais mutationnel neutre ou (ii) un effet de la sélection. Si certaines mutations sont plus fréquentes que d'autres, la composition en base du génome sera globalement déséquilibrée et se reflétera sur la fréquence des codons synonymes (qui codent pour le même acide aminé). Chez les Mammifères par exemple, le génome est structuré en régions plus ou moins riches en GC qu'on appelle isochores (Eyre-Walker & Hurst, 2001). Un biais d'usage du code peut aussi être induit par une sélection sur l'efficacité de la traduction (Akashi & Eyre-Walker, 1998). Chez la drosophile par exemple, les codons préférés correspondent aux ARN<sub>t</sub> les plus nombreux dans le cytoplasme des cellules (Moriyama & Powell, 1997) et il existe une corrélation positive entre le biais d'usage du code d'un gène et son niveau d'expression (Duret & Mouchiroud, 1999). Le modèle sélectif n'est pas contradictoire avec les résultats contrastés obtenus chez les Mammifères où la sélection sur les codons semble absente et la drosophile où la sélection est active : pour un coefficient de sélection donné, l'efficacité de la sélection dépend de la taille efficace de la population,  $N_e$  (Kimura, 1983). Ainsi chez les Mammifères qui ont un  $N_e$  faible, la dérive génétique l'emporterait sur la sélection alors que chez la drosophile le  $N_e$  serait suffisamment grand pour que la sélection soit efficace. Cependant, il est aussi possible que la structure en isochore des génomes des Mammifères brouille le signal. Chez la drosophile, une structure en isochore existe aussi mais elle est beaucoup plus faible (Akashi *et al.*, 1998). Négligé jusqu'à peu, l'étude de l'ADN non-codant a récemment dévoilé un écart au modèle neutre à l'équilibre similaire à celui observé sur les séquences codantes (Galtier *et al.*, 2006).

Nous avons mené une étude conjointe de l'effet de l'expression et de la structure en isochore sur le biais d'usage du code des génomes animaux à partir de 45 banques EST dont les six disponibles chez les Bivalves marins. Nous avons utilisé les ESTs dans le but d'obtenir simultanément (i) des données sur l'usage du code, (ii) une mesure de l'expression des gènes et (iii) une mesure de la structure en isochore à l'aide de la composition en base de l'UTR 3'. Les séquences sont directement utilisées pour calculer la fréquence d'utilisation des codons synonymes. La redondance d'un gène dans

la banque est utilisée comme une mesure grossière du niveau d'expression (Duret & Mouchiroud, 1999). Enfin, la structure en isochore est appréciée par l'intensité de la corrélation entre la composition en GC de la troisième position des codons (GC3) et la composition en GC de l'UTR 3' (Eyre-Walker & Hurst, 2001).

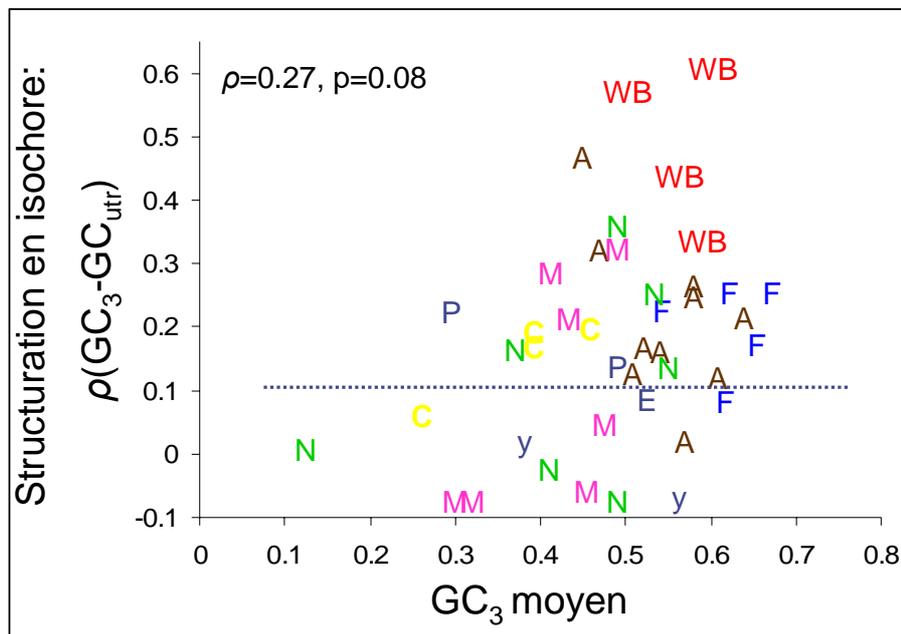


**Figure II.1** : Nombre de fois où un codon est identifié comme étant optimum pour la traduction dans un jeu de donnée de 45 espèces animales. Les histogrammes noirs correspondent aux codons se terminant par un G ou un C et les blancs aux codons se terminant par un A ou un U.

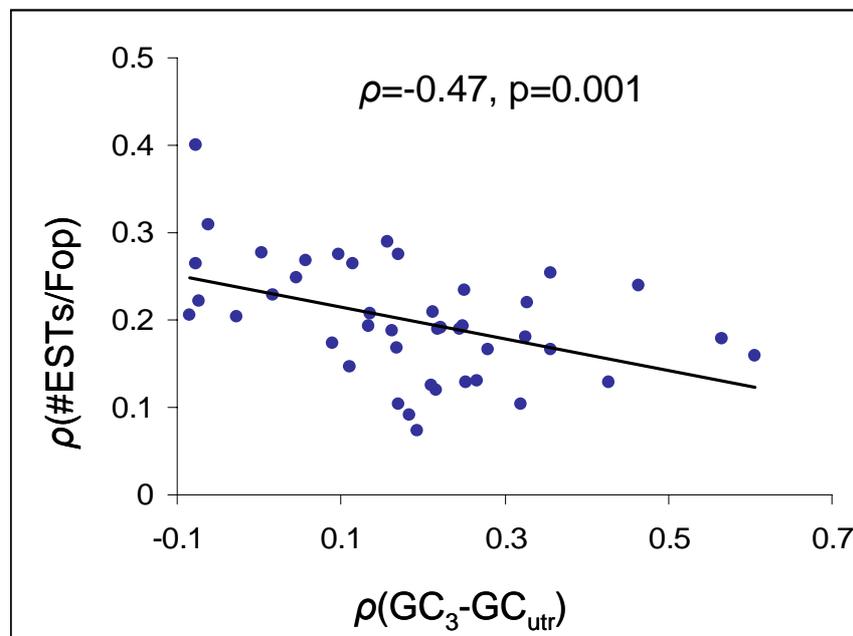
La figure II.1 montre que les codons synonymes qui augmentent significativement en fréquence avec l'expression sont très souvent des codons se terminant par un G ou un C. C'est surtout le cas pour les acides aminés dégénérés deux fois (c.a.d. codés par deux codons synonymes). Il apparaît donc légitime de combiner l'analyse de l'usage du code avec l'analyse de la composition en GC des génomes.

La variation de la composition en base dans un génome n'est pas forcément liée à la variation de la composition en base entre génomes. Pourtant, les auteurs se contentent souvent de décrire un

génomique par sa composition moyenne en GC (Cutter *et al.*, 2006). La figure II.2 montre que le contenu moyen en GC est un mauvais indicateur de la variation en GC et notamment de l'autocorrélation spatiale du GC comme mesurée par la corrélation entre le GC3 et le GC de l'UTR 3'.



**Figure II.2 :** Absence de relation entre le GC3 moyen et la structure en isochore mesurée par la corrélation entre le GC3 et le GC de l'UTR 3'. Le trait pointillé sépare les corrélations significatives [ $\rho(\text{GC3-GC}_{\text{utr}}) > 0.1$ ] des corrélations non significatives au seuil de 5%. WB : Vertébrés à sang chaud, F : Poissons, E : Echinodermes, A : Arthropodes, M : Mollusques, N : Nématodes, P : Plathelminthes, C : Cnidaires, y : Levures.



**Figure II.3 :** Relation négative entre la structure en isochore et l'effet de l'expression sur le biais d'usage du code.

En utilisant notre mesure de la structure en isochores, nous montrons que cette dernière explique en grande partie l'intensité de la corrélation entre le biais d'usage du code et l'expression (Figure II.3).

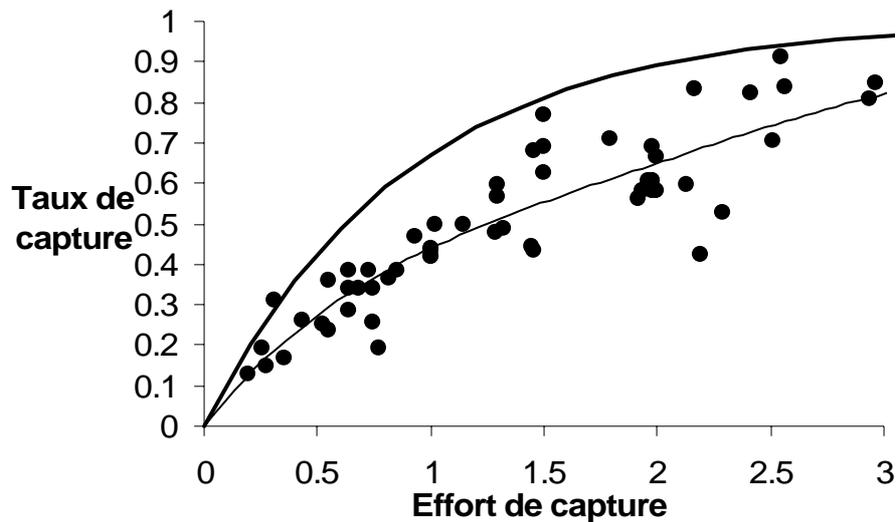
Par contre, l'idée que nous avons de la taille des populations ainsi que les mesures de la diversité génétique actuelle n'expliquent pas du tout cette variation (résultats non montrés). Nous concluons donc que l'effet de l'expression sur le biais d'usage du code est brouillé par la structure en isochores dans les génomes ainsi structurés. L'étude résumée ci-dessus fait l'objet d'un article en préparation.

En ce qui concerne les Bivalves marins, une structure en isochore d'une intensité semblable à celle de la drosophile est observée chez les huîtres du genre *Crassostrea* et les pétoncles de l'espèce *Argopecten irradians*. Par contre, le génome des Mytilidés (*Mytilus* et *Bathymodiolus*) n'est pas structuré en isochores et permet d'espérer étudier plus nettement et sans biais l'effet de la sélection sur l'usage du code. Une étude de l'évolution du biais de codon chez les Mytilidés est initiée à partir de triplet d'orthologues générés à partir de trois bases ESTs (dont les deux générées dans le cadre du projet et du Réseau d'excellence européen Marine Genomics Europe) de *Mytilus edulis*, *Mytilus californianus* et *Bathymodiolus azoricus*.

### **III- Une méthodologie inédite pour l'analyse du polymorphisme de séquence : le clonage par marquage re-capture.**

Le clonage est une étape indispensable mais coûteuse qui permet de séparer les différents allèles d'un même individu diploïde afin de pouvoir les séquencer. Pour réduire les coûts, nous avons mis au point une technique de clonage inédite. Elle permet de ne réaliser qu'une seule réaction de transformation pour plusieurs individus.

La technique consiste à marquer les produits de PCR de chaque individu, ce qui a pour but de les rendre identifiables à la lecture de leur séquence. Le marquage de la séquence lors de l'amplification se fait à l'aide d'une queue constituée de quelques nucléotides ajoutés sur l'extrémité 5' de l'amorce. Ces nucléotides n'ont pas pour but de s'apparier au brin d'ADN complémentaire comme le reste de l'amorce mais de différencier les amorces entre elles. Cette queue sera ensuite lue sur les séquences qui sont réalisées à l'aide d'amorces s'appariant sur le plasmide de part et d'autre de l'insert. Les produits de PCR des différents individus d'un échantillon peuvent ainsi être mélangés lors de la ligation et chaque colonie récupérée à la fin du clonage ne contiendra qu'une seule séquence dont la provenance sera identifiable. Le nombre de séquences différentes récupérées dépendra du nombre d'individus mélangés avant le clonage et de l'effort de ré-échantillonnage des colonies après le clonage. Cette méthode que nous avons baptisée clonage par marquage-recapture a été utilisée dans 57 clonages et nous a permis d'évaluer les taux de capture réalisés en laboratoire (Figure III.1). La technique de clonage par marquage re-capture a fait l'objet d'une publication acceptée dans *Molecular Ecology Notes*.



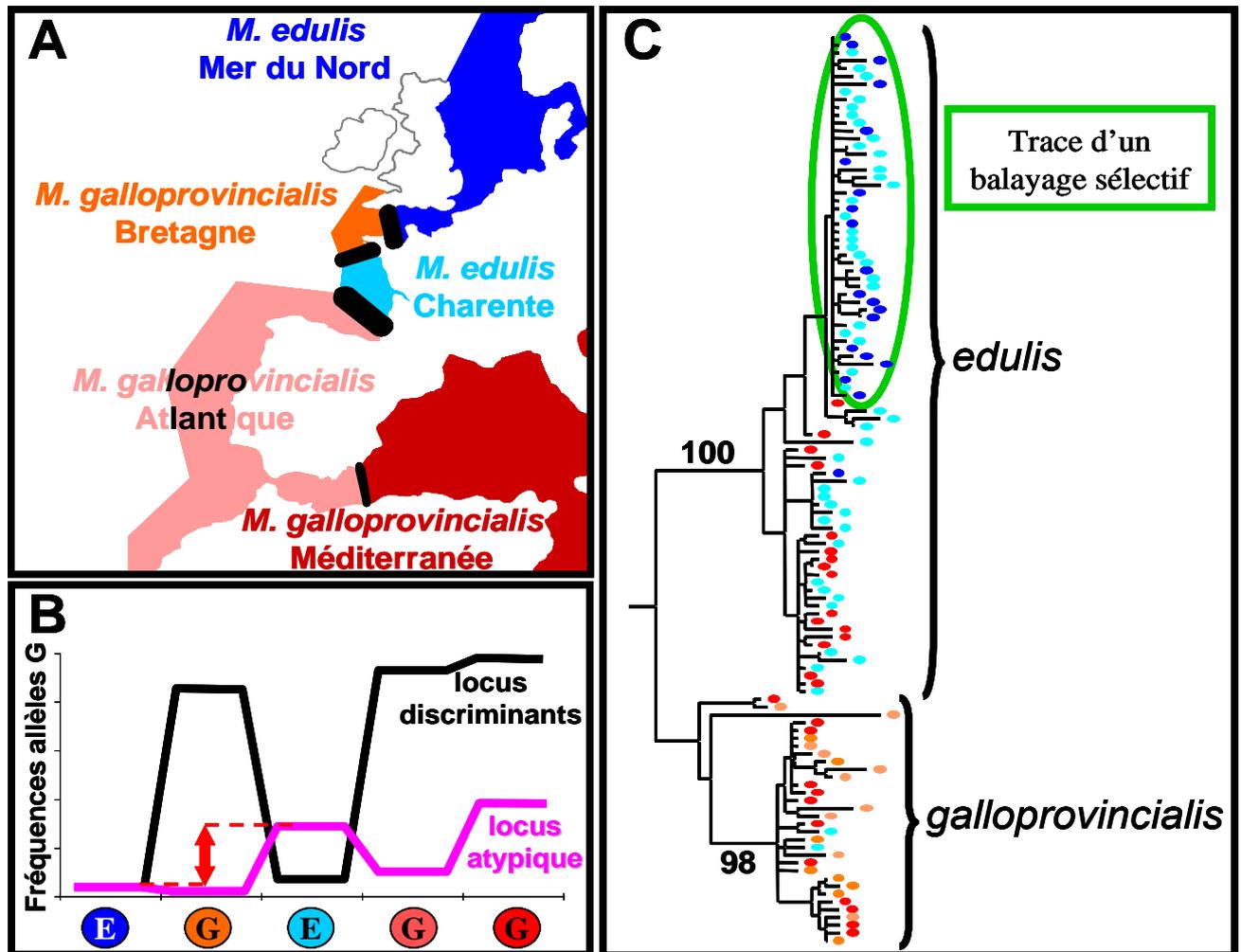
**Figure III.1** : Taux de capture (nombre d'individus capturés / taille de l'échantillon initial) en fonction de l'effort de capture (nombre de réactions de séquençage réalisées / taille de l'échantillon initial). La ligne épaisse est l'attendu sous l'hypothèse d'une distribution uniforme et la ligne fine est une régression polynomiale sur les données.

## IV- Résultats chez les moules du genre *Mytilus*.

### IV.1- Effet d'un balayage sélectif dans des populations de moules spatialement structurée.

Il existe de nombreux exemples, parfois saisissants, de balayage sélectif démontrant indirectement la fixation d'un allèle favorable dans une population. Par contre, il existe peu d'exemple de balayage sélectif établissant le passage d'un allèle favorable au travers d'une barrière au flux génique. Ce type de balayage est pourtant important à distinguer de la situation intra-populationnelle car il correspond à des mutations inconditionnellement favorables quelque soit le contexte génétique et écologique où elles s'expriment. Nous avons recherché la signature d'effets sélectifs sur des allèles introgressés à partir de séquences d'ADN dans la zone hybride des moules *Mytilus edulis* et *M. galloprovincialis* le long des côtes européennes. Le facteur d'élongation  $1\alpha$ , locus présentant des distributions de fréquences atypiques (Bierne *et al.*, 2003), a été choisi pour une étude plus fine. Les données de polymorphisme de 117 séquences d'une portion de ce gène nous ont permis, grâce à l'information généalogique qu'elles contiennent, d'expliquer les différentiels de fréquences alléliques observés en populations (Figure IV.1). Ce différentiel est du à un balayage sélectif en mer du Nord dont l'effet est atténué dans le Golfe de Gascogne suite au passage d'une barrière au flux génique. Un modèle théorique en simulations nous a permis d'illustrer l'effet d'atténuation d'un auto-stop génique au passage d'une barrière. Ces données nous ont également permis de démontrer une introgression forte d'allèles *edulis* dans les populations *M. galloprovincialis* de Méditerranée alors que les

populations d'Atlantique ne sont quasiment pas introgressées. Ce résultat est en désaccord avec la répartition actuelle des patches et suggère que des événements de déplacement des peuplements auraient eu lieu lors des périodes glaciaires.

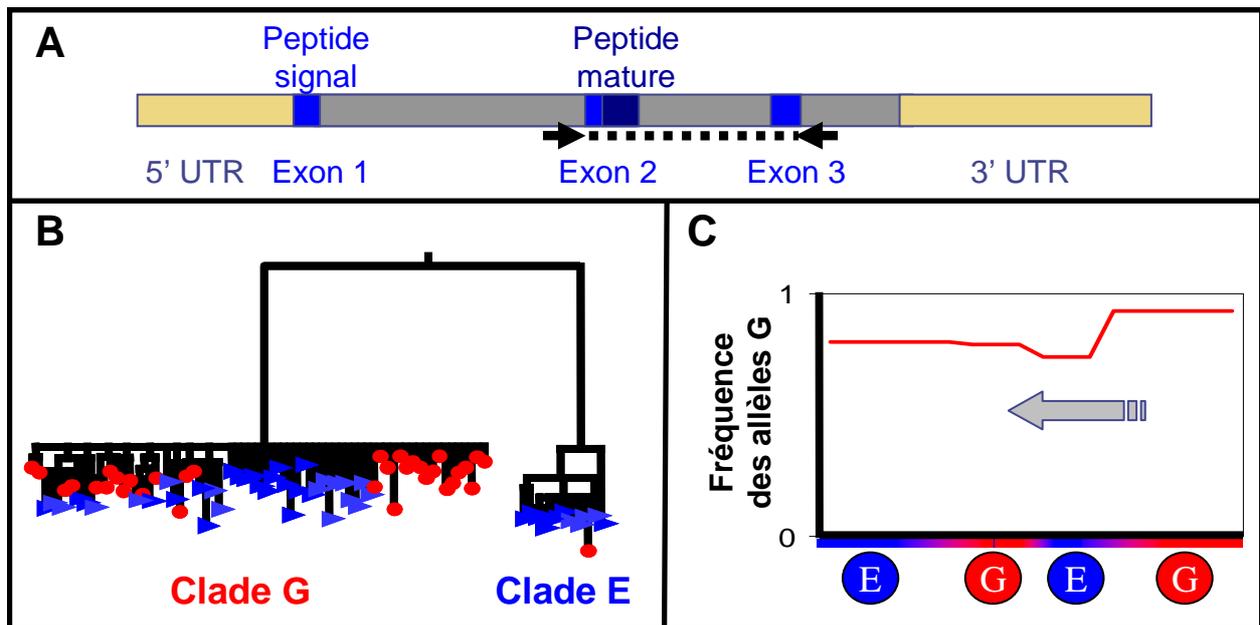


**Figure IV.1 :** Introgression et sélection dans la zone hybride de *Mytilus*. (A) La zone d'hybridation des moules *Mytilus edulis* et *M. galloprovincialis* est une mosaïque où alternent des patches de chaque espèce le long des côtes européennes. Cette structure inédite, constituée de quatre zones hybrides successives indépendantes, nous offre une opportunité originale de détecter un effet de la sélection sur les barrières aux flux de gène. (B) Le patron de structure d'un des marqueurs que nous avons étudié semble montrer l'action de la sélection. En effet, l'allèle G3 du facteur d'élongation (locus EFBis) se trouve en forte fréquence (~30%) dans les populations *M. edulis* de Charente ce qui n'est pas le cas dans les populations *M. galloprovincialis* qui les entourent, ni dans les populations *M. edulis* de Mer du Nord considérées comme les populations *M. edulis* d'origine. (C) La généalogie de 117 séquences d'une portion du gène du facteur d'élongation 1 $\alpha$  nous a permis d'identifier les traces de la sélection sous la forme d'une généalogie en étoile. De plus, la forme globale de la généalogie avec deux clades d'allèles très différenciés permet d'identifier le processus d'introgression d'allèle *edulis* dans les populations *M. galloprovincialis* (points rouges du clade *edulis*) et dans une moindre mesure d'allèle *galloprovincialis* dans les populations *M. edulis* (points bleus du clade *galloprovincialis*).

## IV.2- Introgression adaptative dans la zone hybride de *Mytilus*.

Nous nous sommes intéressés à l'évolution d'un gène d'immunité codant pour un peptide antimicrobien, la Mytiline B (Mitta *et al.*, 2000). Notre hypothèse de départ était que la sélection aurait pu favoriser l'échange d'allèles après un contact secondaire (i) si ce gène était soumis à une forme de sélection stabilisatrice comme c'est parfois le cas sur ce type de fonction ou (ii) si un pathogène co-adapté à une espèce envahissait l'autre espèce favorisant le passage des gènes d'immunités à travers la barrière interspécifique.

La généalogie obtenue (Figure IV.2) nous dévoile (i) une introgression forte et asymétrique des allèles *galloprovincialis* dans les populations *M. edulis* alors que l'introgression neutre est d'ordinaire asymétrique dans l'autre sens d'*edulis* vers *galloprovincialis* (Bierne *et al.*, 2006) et (ii) une généalogie en étoile caractéristique d'un balayage sélectif. Ces résultats sont donc en accord avec l'hypothèse d'une introgression adaptative. Cependant, les résultats ne sont pas en accord avec une sélection directe sur la portion du gène séquencée puisque le balayage est imparfait. De plus, des tests d'évolution de la protéine ne révèlent pas d'excès de substitutions non-synonymes dans la divergence. Il reste cependant possible que la substitution adaptative ait eu lieu en amont de nos séquences dans la partie régulatrice du gène.



**Figure IV.2 :** Introgression adaptative dans la zone hybride de *Mytilus*. (A) Structure génomique du gène de la Mytiline B. Les portions beiges correspondent aux UTRs, les grises aux introns et les bleus aux exons (la portion correspondant au peptide mature est en bleu foncé). Le trait pointillé représente la région analysée et les flèches correspondent aux amorces. (B) Arbre de maximum de vraisemblance des séquences de la Mytiline B. (C) Fréquences alléliques dans les différents patches de la zone hybride obtenues par typage d'un indel diagnostique entre les deux clades.

### IV.3- Absence de barrière, évolution neutre et recombinaison au locus de la Glucanase

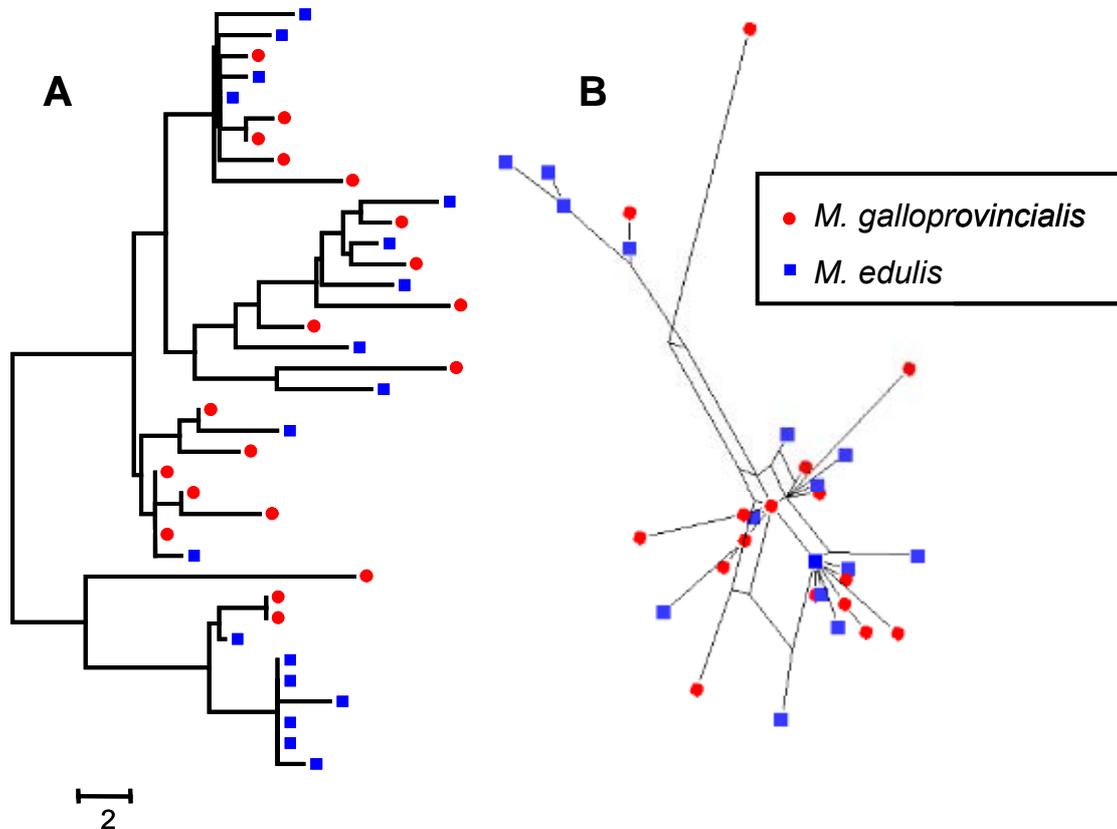
Le troisième locus que nous avons étudié est une portion du gène codant pour une enzyme cellulolytique (endo $\beta$ -1,4-glucanase). Les résultats de l'analyse de la Glucanase se distinguent de ceux précédemment obtenus sur deux points principaux :

(i) Tout d'abord, aucune différenciation génétique significative n'est observée entre les populations des deux espèces (Figure IV.3). Il semble donc que l'introgession ait complètement homogénéisé cette portion du génome. Cependant, contrairement au résultat obtenu sur la Mytiline B, l'introgession observée ici ne semble pas être due à l'action d'une sélection positive. En effet, la généalogie observée sur la figure IV.3 ne montre pas de coalescence en étoile, signature typique d'un balayage sélectif. De plus, les tests d'écart à la neutralité ne sont pas significatifs (exemple : D de Tajima = -1.2,  $p > 0.1$ ).

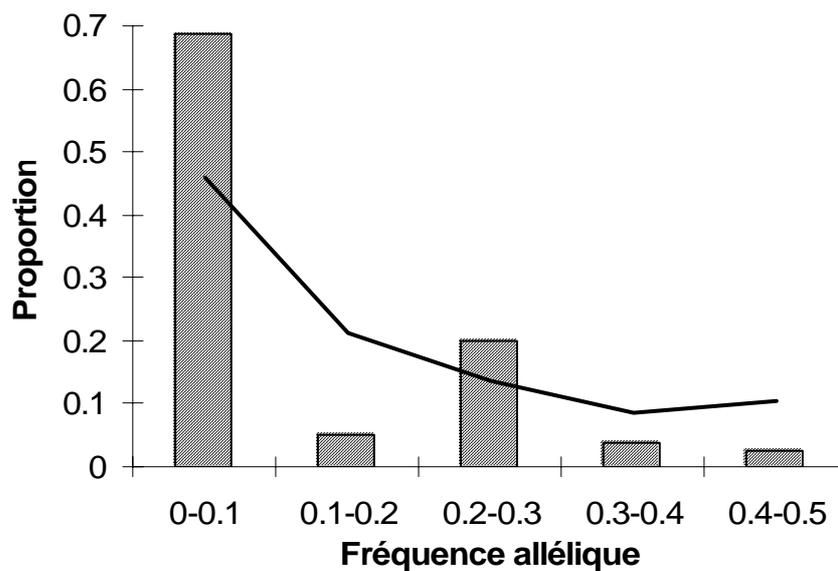
(ii) Ensuite, la phylogénie des allèles ne fait pas apparaître de manière évidente deux clades de séquences très divergentes, indice d'un épisode de divergence longue en allopatrie, comme c'était le cas pour les précédents locus analysés. Les faibles valeurs de déséquilibre de liaison observés au locus de la Glucanase [la moyenne du  $R^2$  de Hill et Robertson (1968) n'est que de 0.06] suggèrent que la recombinaison ait pu brouiller le signal de cette divergence passée. Des simulations de coalescences neutres sans recombinaison avec le nombre de sites polymorphes observés montrent que le déséquilibre de liaison observé (0.06) est inférieur à l'intervalle de confiance à 95% des données simulées (95%IC=[0.08, 0.45]). La méthode de Hudson et Kaplan (1985) identifie un nombre minimal,  $R_m$ , de 8 points de recombinaison. De plus, la représentation en réseau réticulé de recombinaison (Huson & Klopper, 2005) illustre la présence de ces événements de recombinaison (Figure IV.3).

Le locus de la Glucanase se présente donc comme panmixtique à l'échelle des deux espèces. De plus, ce locus ne s'écarte pas de l'attendu neutre à l'équilibre démographique. La barrière interspécifique n'est donc pas active sur la portion du génome où se trouve ce locus. Le schéma de génomes suffisamment verrouillés pour conserver une structure forte sur tous les compartiments génomiques à l'exception de ceux ayant subi un événement de sélection positive est donc remis en question. Est-il possible que le modèle d'un isolement long suivi d'un contact secondaire soit erroné ? Un scénario réputé moins probable mais possible suppose que la barrière interspécifique se soit créée en présence de flux génique (spéciation parapatrique). Alternativement, la divergence a bien eu lieu mais ses traces sur la généalogie du locus de la Glucanase ont été effacées par la recombinaison suite à des événements d'introgession. En accord avec cette hypothèse nous avons détecté une influence importante de la recombinaison intragénique. Nous avons aussi remarqué un nombre important de mutations ségrégeant à une fréquence intermédiaire autour de 25% (Figure IV.4). Cet effet est malheureusement masqué par un excès de mutations rares pouvant être expliqué de diverses façons. Des études supplémentaires sont nécessaires pour explorer ces scénarios alternatifs. Plusieurs

polymorphismes de longueur d'insertion/délétions ont été repérés sur ce locus et permettront un typage à grande échelle. Le développement de méthode d'analyse avec recombinaison sera également exploré.



**Figure IV.3** : (A) Phylogramme non raciné obtenu par la méthode du Neighbor-Joining sur les nombres de différences nucléotidiques. (B) Réseau réticulé de recombinaison obtenu par la méthode de décomposition par bipartition sur le nombre de différences nucléotidiques (Huson & Klopper, 2005).



**Figure IV.4** : Distribution observée (histogramme) et attendue sous le modèle neutre à l'équilibre démographique (ligne) des fréquences alléliques.

#### **IV.4- Données moléculaires acquises mais non encore analysées**

Trois locus supplémentaires ont été envoyés au CNS d'Evry en Juillet 2006 et les résultats des séquences arrivent au moment de la rédaction du présent rapport (30 Octobre 2006). Il s'agit de portions des gènes codant pour l'endo-1,4-mannanase, la défensine MGD2, et facteur d'élongation 2 (EF2). A la fin de la thèse en cours sur le modèle *Mytilus*, six locus auront été analysés. Chaque locus s'avère avoir une histoire plus compliquée que prévu et l'analyse de la diversité en devient très complexe. L'augmentation du temps consacré à l'analyse justifie donc que le nombre de locus étudié ait été revu à la baisse. Malgré cette diminution du jeu de données, ces 6 marqueurs constitueront déjà un bel échantillon du génome et nous permettront d'avoir une image de la variation des flux génique entre locus.

### **V- Résultats chez les modioles hydrothermale du genre *Bathymodiolus*.**

#### **V.1- Objectifs de l'étude**

L'intérêt des *Bathymodiolus* réside dans le fait qu'elles vivent dans le milieu hydrothermal profond et que leurs populations sont ainsi soumises à des phénomènes d'extinction/recolonisation fréquents et récurrents propices à entraîner des goulots d'étranglement dans les populations et à favoriser des balayages sélectifs. De ce fait, une des hypothèses de départ de ce projet est que les diversités nucléotidiques et alléliques observées devraient être plus faibles que celles des espèces côtières (*Mytilus* et *Crassostrea*). Un second objectif, plus spécifique au genre, est de rechercher la trace d'événements d'hybridation entre espèces proches dans l'Atlantique (Ride médio atlantique) comme dans le Pacifique (Dorsale est Pacifique).

#### **V.2- Approche multigénique et effort d'échantillonnage**

Pour distinguer les effets sélectifs des effets démographiques passés, il était nécessaire d'utiliser une approche multi-génique sur une dizaine de gènes. Nous avons donc mis au point des marqueurs génomiques : Cytochrome Oxydase 1 (CO1), Facteur d'élongation 1 $\alpha$  (EF1 $\alpha$ ), Heat shock protein 70 (HSP70), Glucanase (Glu), Adenosyl homo-cystéinase (AHC), Sulfotransférase (ST1), Ferritine (GF1), Lysozyme (LYS), collagène (DEC-1) à partir de séquences étiquettes obtenues à l'aide d'une banque ADNc pleine longueur.

Pour les gènes DEC-1 et Glu, seules les données de divergence ont été obtenues entre les espèces *Bathymodiolus thermophilus* et *B. azoricus*. En revanche pour l'ensemble des autres gènes des valeurs de polymorphisme et de divergence ont pu être estimées entre 2 ou 4 populations (48 individus par population) en fonction de l'effort de séquençage actuel (Tableau V.1).

Gènes	Nombre de séquences	Séquençage	Nombre de populations
COI	192	1 sens	4
EF1	384	1 sens	4
HSP	100	1 sens	2
AHC	416	2 sens	2
GF1	416	2 sens	2
LYS	416	2 sens	2
ST1	416	2 sens	2

**Tableau V.1** : Effort séquençage chez *Bathymodiolus* sp :

### V.3- Analyse de la divergence et du polymorphisme

Les analyses de la divergence entre *B. thermophilus* et *B. azoricus* montrent une valeur élevée pour le gène mitochondrial (COI ; D= 0,116) qui contraste avec les valeurs obtenues pour les gènes nucléaires. En effet, les valeurs de divergence nucléaires relativement homogènes entre locus varient de 0.019 à 0.052 (Tableau V.2). Ceci suggère un taux de mutation 4 fois plus élevé pour le mitochondrial que pour le nucléaire.

Locus	$\pi$ (7°S)	D (7°S)	$\pi$ (Rb)	D (Rb)	D <sub>XY</sub>
LYS	0,009 (56)	-2,03 *	0,016 (41)	-1,63 (NS)	0,046
GF1a	0,015 (35)	-1,96 *	0,012 (57)	-2,37 **	0,026
GF1b	0,009 (48)	-2,36 **	0,018 (25)	-0,24 (NS)	0,039
AHC	0,009 (23)	-1,79 (NS)	0,012 (39)	-1,74 (NS)	0,042
EF1 $\alpha$	0,001 (31)	-2,23 **	0,005 (19)	-2,64 **	0,024
ST1	0,018 (62)	-1,80 *	0,011 (63)	-2,26 **	0,019
mtCOI	0,007 (96)	-0,87 (NS)	0,005 (96)	-1,77 (NS)	0,116
GLU	---	---	---	---	0,034
DEC-1	---	---	---	---	0,035

**Tableau V.2** : Valeurs de diversité nucléotidique ( $\pi$ ) et de divergence (D<sub>XY</sub>) estimées au niveau des populations 7°S (*B. thermophilus*) et Rainbow (*B. azoricus*) pour 6 gènes nucléaires et un gène mitochondrial. D correspond à la statistique de Tajima qui mesure l'écart aux attendus de l'évolution neutre en terme de polymorphisme nucléotidique. Les valeurs entre parenthèses correspondent au nombre de séquences utilisées dans l'analyse.

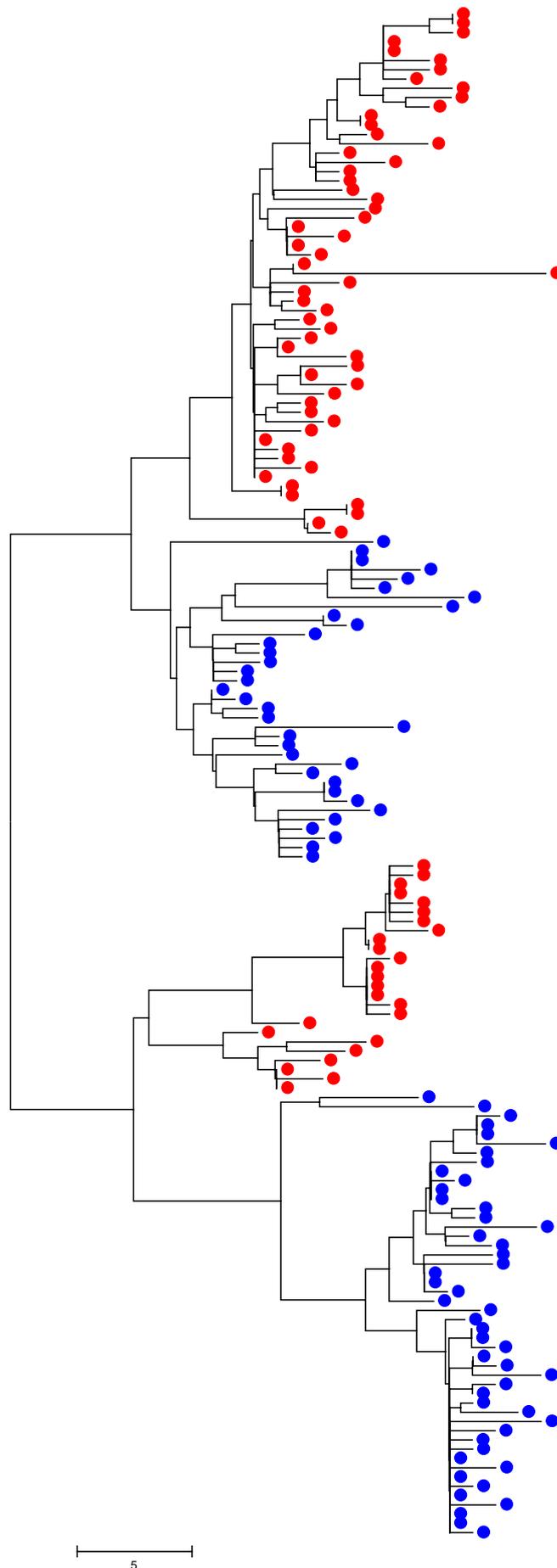
Les valeurs de polymorphisme pour *Bathymodiolus* sont environ 2 fois moins élevées que celles trouvées chez les autres Bivalves et varient entre 0.001 et 0.018 (Tableau V.2). Les valeurs de diversité obtenues pour le gène mitochondrial ne sont que légèrement supérieures à celles des gènes nucléaires ( $\pi = 0.021$ ), ce qui semble en accord avec l'action combinée d'un taux de mutation plus élevé et d'une taille efficace attendue 4 fois plus faible que pour le nucléaire.





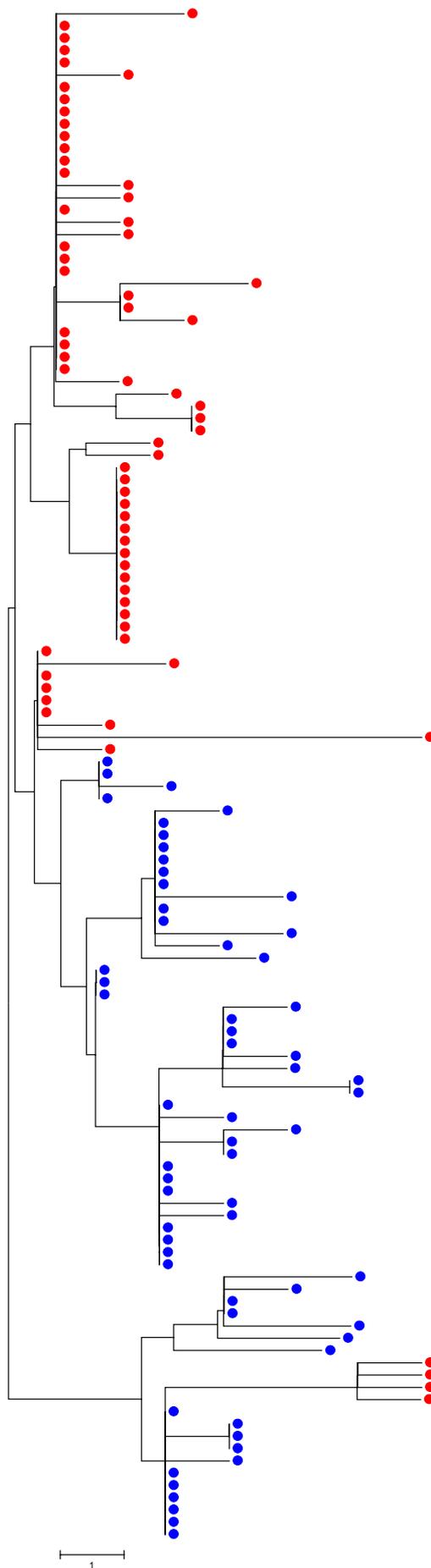
**Figure V.1.c** : Arbre GF1 en NJ avec nombre de différences observées

Rouge : *azoricus* Rb  
Bleu : *thermophilus* 7S



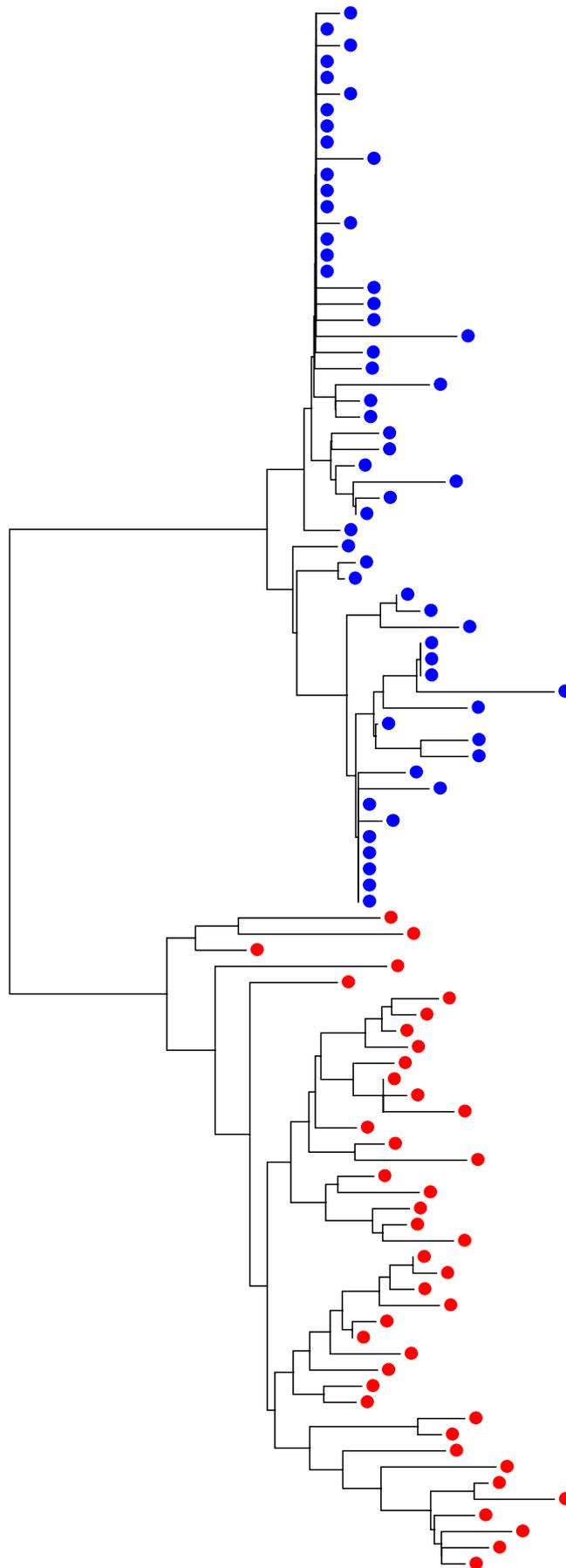
**Figure V.1.d** : Arbre **Sulfotransferase** en NJ avec nombre de différences observées

Rouge : *azoricus* Rb  
Bleu : *thermophilus* 7S



**Figure V.1.e** : Arbre **Lysozyme** en NJ avec nombre de différences observées

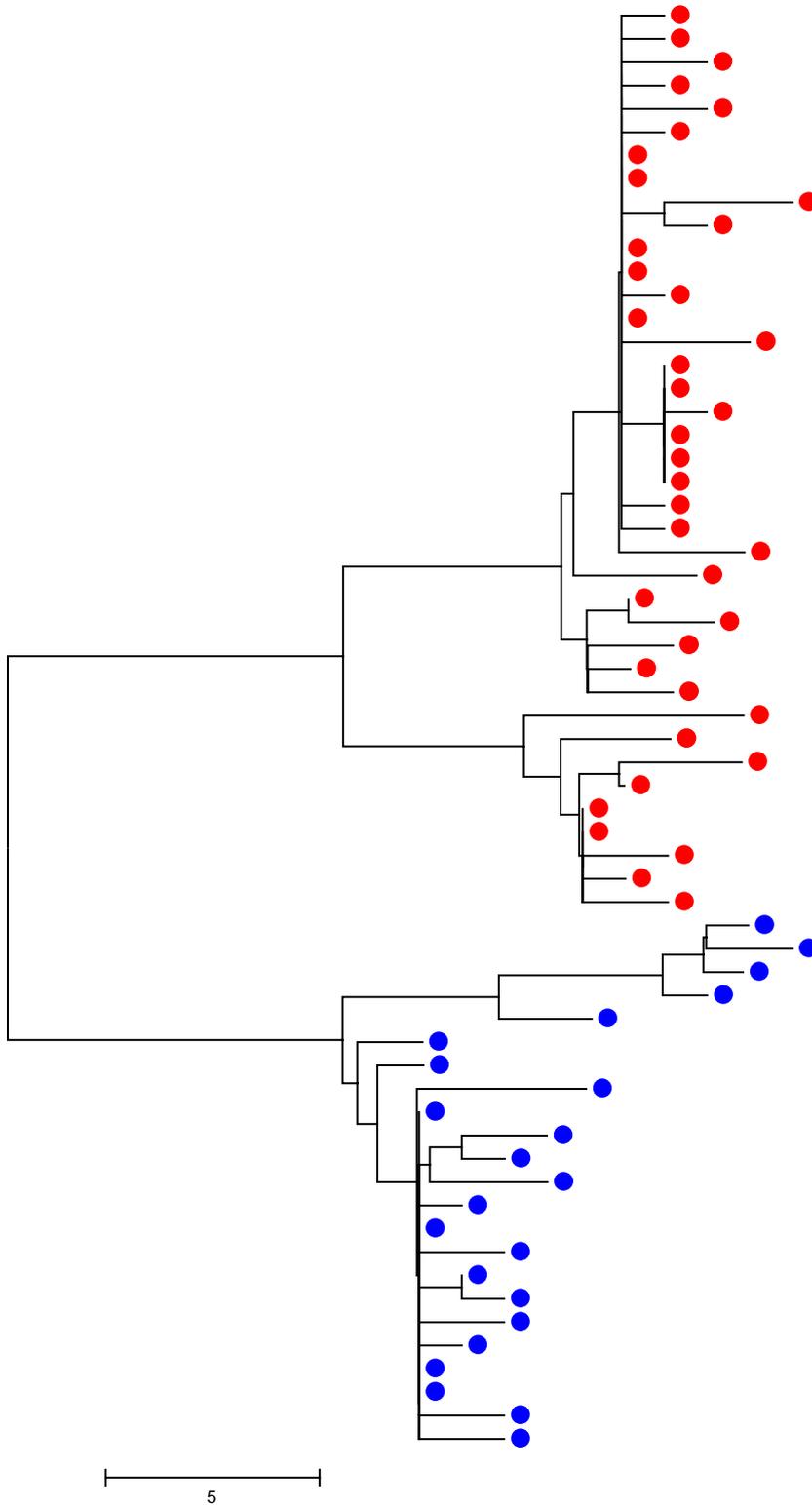
Rouge : *azoricus* Rb  
Bleu : *thermophilus* 7S



5

**Figure V.1.f** : Arbre AHC en NJ avec nombre de différences observées

Rouge : *azoricus* Rb  
Bleu : *thermophilus* 7S



Les phylogénies des allèles séquencés sont représentées figure V.1. Les arbres obtenus présentent des topologies globalement congruentes indiquant la monophylie entre les espèces Pacifique et Atlantique à l'exception de la sulfotransferase qui semble évoluer moins vite que les autres gènes. La barrière reproductive semble donc complète de part et d'autre de l'Isthme de Panama.

Il existe cependant des différences notables entre les gènes. Ainsi, les arbres de COI et d'EF1 présentent une topologie en râteau (ou étoile) que l'on ne retrouve pas chez les autres gènes. Cela pourrait être interprété comme la signature d'un balayage sélectif (EF1 : la séquence comporte un intron de grande taille) ou des contraintes sélectives extrêmement fortes sur le gène (COI). Les premiers alignements de GF1 font clairement apparaître que nous n'amplifions pas un seul gène mais deux paralogues. Les analyses de polymorphisme ont donc été effectuées de façon séparée sur les 2 jeux de paralogues (GF1A et GF1B). L'analyse conjointe de ces gènes dupliqués devrait permettre de tester certaines hypothèses, par exemple sur le relâchement des pressions de sélection et l'évolution de fonction chez les ferritines, ces molécules étant fortement impliquées dans l'adaptation métallique chez les invertébrés marins. Les arbres AHC et Lysozyme présentent des topologies similaires avec des coalescences plus profondes laissant supposer la présence d'allèles ancestraux. Pour ces 2 derniers gènes, des allèles très divergents (supprimés des arbres et de l'analyse) ont été observés dans la population de *B. thermophilus* (7°25S) et pourraient être des allèles introgressés de l'espèce *Bathymodiolus* spp (37°S) située plus au sud à proximité de la dorsale Antarctique. Il est cependant nécessaire d'obtenir les allèles associés à cette dernière espèce pour confirmer l'interprétation. Ces séquences diagnostiques de l'espèce n'ont été en effet obtenues que pour les gènes COI et EF1 $\alpha$ . La sulfotransferase présente divergence inter-spécifique moins nette que pour les autres gènes.

Les tests de Tajima effectués population par population sur les différents gènes réfutent l'hypothèse de l'effet d'un goulet d'étranglement récent au sein des 2 espèces. En effet, certains gènes ne présentent pas d'écart significatif aux attendus d'une évolution moléculaire neutre (cas des gènes codant pour le lysozyme et l'AHC). Cependant, des valeurs significatives de D (Tableau V.2) sont plus fréquemment retrouvées au sein de la population de l'espèce *B. thermophilus* (EPR) et pourrait valider l'idée reçue que cette espèce est une espèce récemment dérivée issue d'une spéciation péripatrique à la fermeture de l'Isthme de Panama.

L'analyse détaillée du locus EF1 $\alpha$  a révélé l'existence de nombreux singletons et la présence d'allèles multiples (>2) chez plusieurs individus. La présence de plusieurs allèles au sein d'un même individu doit être confirmé sur les autres gènes afin de savoir si ceux-ci reflètent des artefacts liés aux méthodes de PCR/clonage utilisées. Cependant, ces excès de singletons s'ils sont avérés, pourraient refléter une accumulation de mutations somatiques liées à la longue durée de vie de ces Bivalves dans un milieu réputé très mutagène.

Le jeu de données obtenu sur 9 locus de *Bathymodiolus* constitue un très bel échantillon du génome, surtout pour cette espèce qui n'avait fait l'objet que de très peu d'études moléculaires préalables. Ces jeux de données seront complétés pour les populations manquantes et les groupes externes par un nouvel envoi de clones à séquencer au Genoscope. En outre, les gènes DEC-1 et GLU seront également traités en terme de polymorphisme pour compléter cette étude. Les nombreuses séquences obtenues et les analyses qui vont suivre permettront de répondre aux interrogations propres à la génomique évolutive des Bivalves mais aussi apporter des réponses aux questions plus spécifiques des modioles hydrothermales, notamment en termes de dispersion et d'hybridation des espèces. La comparaison des valeurs de polymorphisme et de divergence devrait quant à elle apporter des informations précieuses sur l'action de l'environnement en terme d'instabilité dans la propagation et/ou la rétention d'allèles avantageux.

## **VI- Résultats chez les huîtres du genre *Crassostrea*.**

### **VI.1- Le modèle *Crassostrea gigas* / *C. angulata***

L'huître creuse japonaise *Crassostrea gigas* et portugaise *C. angulata* ont été respectivement décrites par Thunberg (1793) et Lamarck (1819). Les populations européennes des deux espèces proviennent d'introduction à partir de populations du Pacifique –*C. angulata* a été introduite, probablement de Taiwan, par les premiers navigateurs portugais au XVIIe siècle alors que *C. gigas* a été introduite du Japon en France pour l'aquaculture au début des années 1970 suite à des épizooties ayant frappées les élevages de *C. angulata* ainsi que de l'huître plate, *Ostrea edulis*. La distinction entre les deux espèces, dont le statut taxonomique reste incertain, n'est pas aisée car il existe de grandes similitudes morphologiques, pas de barrière à l'hybridation au laboratoire ou en milieu naturel (Huvet *et al.*, 2002), peu de différences au niveau chromosomique (Leitão *et al.*, 2004) et peu de marqueurs génétiques les distinguent. Seul l'ADN mitochondrial (O'Foighil *et al.*, 1998; Boudry *et al.*, 1998) et un indel flanquant un locus microsatellite (Huvet *et al.*, 2004) permettent de les distinguer nettement alors que les locus allozymiques (Buroker *et al.*, 1979) et microsatellites (Huvet *et al.*, 2000b) ne dévoilent pas ou peu de différenciation entre les populations des deux espèces. Les raisons des contrastes observés entre les différents compartiments du génome sont donc à rechercher. Dans ce contexte, l'analyse des fréquences alléliques devient insuffisante alors qu'une analyse multi-locus des patrons de coalescence de séquences d'ADN (Wang *et al.*, 1997) peut aider à trancher entre les différents scénarios.

Nous présentons ici les résultats obtenus sur les six gènes qui ont été compilés pour le moment mais les données ont été récoltées pour seize gènes en tout pour ces espèces.

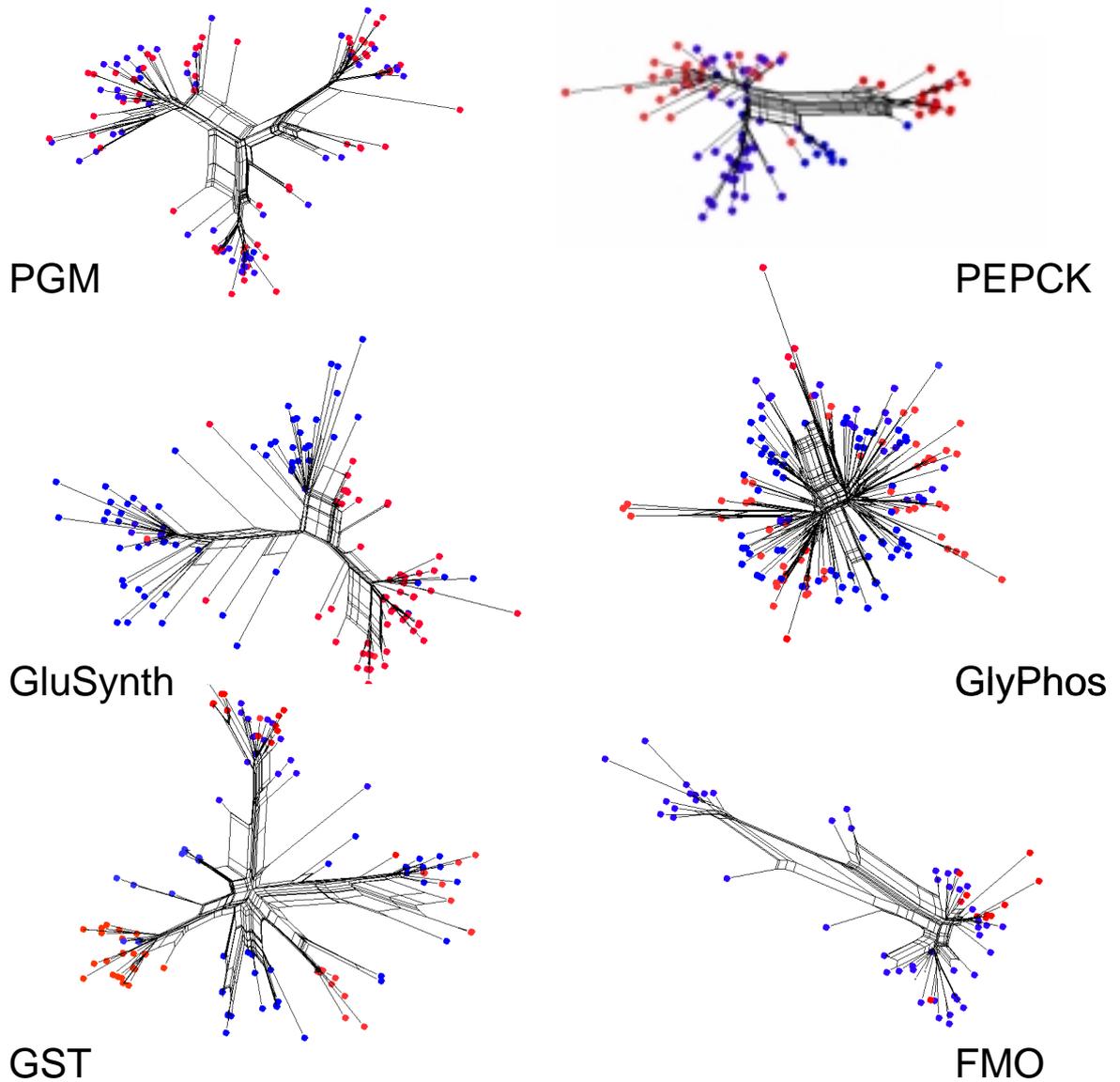
## VI.2- Génétique de la différenciation entre *Crassostrea angulata* et *C. gigas*

L'analyse des séquences a très vite révélé chez *Crassostrea* un impact important de la recombinaison intragénique. De ce fait, l'approche traditionnelle qui consiste à tenter de reconstruire la généalogie des allèles est devenue problématique. Nous avons donc choisi une représentation en réseau réticulé de recombinaison (Huson & Kloepper, 2005) qui permet d'illustrer l'action de la recombinaison dans l'évolution des séquences (Figure VI.1). Les nombres minimaux de points de recombinaison identifiés par la méthode de Hudson et Kaplan (1985) vont de 10 à 29 (Tableau VI.1) ce qui montre qu'un travail sur des portions de séquences plus petites n'est pas envisageable. L'importance de la recombinaison intragénique dépend du taux de recombinaison populationnel (produit de la taille efficace  $N_e$  et du taux de recombinaison,  $r$ ). Cette observation est donc logique compte tenu que la diversité nucléotidique, qui dépend elle du taux de mutation populationnel (produit de la taille efficace  $N_e$  et du taux de mutation,  $u$ ), observé chez *Crassostrea* est également élevée (Tableau VI.1). Enfin, les  $D$  de Tajima bien que toujours négatifs, ne s'écartent pas significativement de l'attendu neutre à l'équilibre démographique. Il faut noter ici que le  $D$  de Tajima qui peut être compris comme une mesure de la déviance du spectre de fréquences alléliques à l'attendu neutre, est réputé peu sensible à la recombinaison (Mousset *et al.*, 2004).

Tableau VI.1 : Diversité nucléotidique,  $D$  de Tajima, déséquilibre de liaison, nombre minimal de points de recombinaison et décomposition de la variance moléculaire pour les six gènes analysés chez *Crassostrea angulata* et *C. gigas*.

	GluSynth	GlyPhos	GST	PEPCK	PGM	FMO
Diversité nucléotidique ( $\pi$ )						
<i>C. gigas</i>	0.025	0.02	0.027	0.011	0.038	0.043
<i>C. angulata</i>	0.014	0.024	0.032	0.018	0.039	0.021
$D$ de Tajima						
<i>C. gigas</i>	-1.63	-1.23	-0.71	-1.78	-1.24	-0.85
<i>C. angulata</i>	-1.67	-1.65	-0.64	-1.11	-1.24	-1.45
Déséquilibre de liaison						
<i>C. gigas</i>	0.022	0.017	0.034	0.023	0.035	0.059
<i>C. angulata</i>	0.032	0.028	0.042	0.028	0.033	0.165
Nombre minimal de points de recombinaison						
	28	33	12	10	18	29
Décomposition de la Variance Moléculaire						
Entre espèces	31%	2.5%	8.5%	17%	0%	10%
Entre populations dans une espèce	3%	1%	1.5%	0.5%	2.5%	4%
Intra population	66%	96.5%	90%	82.5%	97.5%	86%

Dans la mesure où les méthodes utilisant la théorie de la coalescence ne sont pas facilement utilisables car la recombinaison n'a jamais été pris en compte dans les modèles de divergence, la différenciation entre les deux espèces peut-être analysée par les méthodes classiques de la génétique des populations comme l'analyse moléculaire de la variance (Excoffier *et al.*, 1992). La structuration en espèces explique de 0% à 31% de la variance moléculaire (Tableau VI.1). Cependant, nous n'avons pour l'instant pas trouvé de méthode statistique satisfaisante nous permettant de tester si ces différences entre locus impliquent nécessairement la présence d'une barrière semi-perméable au flux génique ou si cette variance entre locus peut être attendue sous l'hypothèse d'un modèle d'isolement partiel à l'équilibre migration-mutation-dérive. En effet, les différences de variance moléculaire expliquées par la structure en espèces sont parfaitement compatible avec la variance inter-locus neutre attendu à l'équilibre migration-mutation-dérive (voir Murray & Hare, 2006), mais les informations phylogénétiques ne sont pas prises en compte de façon satisfaisante (notamment les déséquilibres de liaison). Quoi qu'il en soit, l'effet homogène de la recombinaison sur l'ensemble de la phylogénie montre que si introgression il y a eu, elle est suffisamment ancienne pour avoir permis la recombinaison entre lignées divergentes. De ce fait, l'hypothèse d'un effet important d'une introgression récente en Europe (rappelons que *C. gigas* n'est présent en Europe que depuis une trentaine d'année) peut d'ores et déjà être réfutée. Ce qui ne veut pas dire qu'elle n'a pas eu lieu, mais qu'elle aurait agit sur des populations déjà fortement impactées par le flux génique dans l'aire naturelle de répartition des deux espèces (c.a.d. dans le Pacifique).



**Figure VI.1 :** Réseaux réticulés de recombinaison obtenus par la méthode de décomposition par bipartition sur le nombre de différences nucléotidiques (Huson & Kloepper, 2005) pour les six gènes analysés chez *Crassostrea*. Les points bleus correspondent aux séquences échantillonnées dans les populations de *C. gigas* et les points rouges aux séquences échantillonnées dans les populations de *C. angulata*.

### VI.3- Analyse des contraintes sélectives

L'étude des mutations délétères est classiquement conduite à l'aide de deux approches très différentes : (i) une approche consiste à inférer le nombre et les effets des mutations délétères par une étude globale du génome dans des croisements contrôlés et des mesures de phénotype macroscopique tel que la survie ou la croissance (Falconer & Mckay, 1996; Ritland, 1996), (ii) une autre approche consiste à étudier les variations à l'échelle microscopique des séquences protéiques (Eyre-Walker & Keightley, 1999; Fay *et al.*, 2001). Si la première approche a été énormément employée (Falconer & Mckay, 1996), la deuxième n'a pour l'instant été explorée que chez quelques espèces modèles. Il est probable que les deux approches s'intéressent à des effets délétères très différents, forts pour la première et infimes (non mesurable par l'étude du phénotype macroscopique) pour la seconde. Il est cependant intéressant de poursuivre les efforts d'analyse à l'échelle moléculaire afin de décortiquer plus finement la distribution des effets délétères des mutations.

*Crassostrea gigas* a dévoilé un fardeau génétique considérable qui compte parmi les plus forts des animaux (Evans *et al.*, 2004; Launey & Hedgecock, 2001). Qu'en est-il à l'échelle moléculaire ? Les portions du génome contre-sélectionnées sont-elles porteuses d'une mutation à effet fort ou sont-elles denses en mutations à effet faible ? L'analyse du spectre des fréquences alléliques des mutations non-synonymes par rapport à celui des mutations silencieuses nous donnera une estimation du nombre de mutation délétère par génome (Fay *et al.*, 2001).

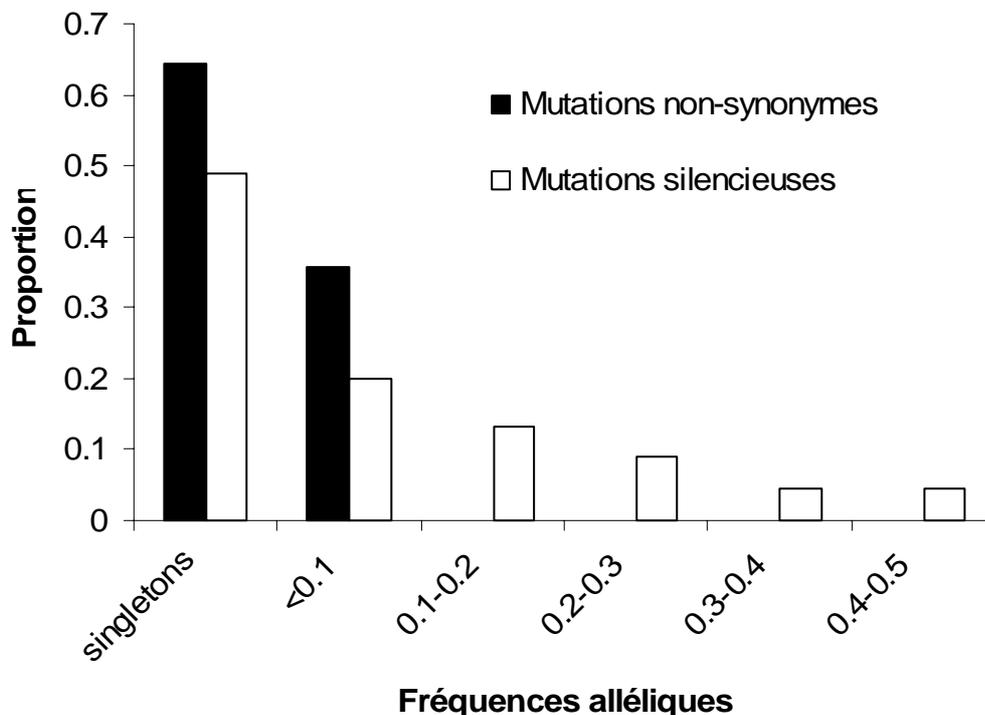


Figure VI.2 : Distribution des fréquences alléliques des mutations non-synonymes et silencieuses (synonymes et introniques) pour les gènes GluSynth, GlyPhos et GST.

La figure VI.2 présente les résultats obtenus pour les gènes GluSynth, GlyPhos et GST. Les mutations non-synonymes ségrégent en fréquence plus faible que les mutations silencieuses, signe qu'elles sont contre-sélectionnées. Une estimation de la proportion des mutations non-synonymes qui sont délétères peut être réalisée en comparant le ratio non-synonymes/silencieux pour plusieurs catégories de fréquences (Fay *et al.*, 2001). Les estimations sont données pour les gènes GluSynth, GlyPhos et GST dans le tableau VI.2.

Tableau VI.2 : Estimation de la proportion délétère dans les mutations non-synonymes polymorphes.

	GluSynth	GlyPhos	GST
C. angulata	92%	85%	75%
C. gigas	60%	82%	71%

En moyenne nous estimons que 78% des mutations non-synonymes sont délétères. Cette proportion est semblable à celle estimée chez l'homme (Fay *et al.*, 2001) ; cependant, comme la diversité est considérablement plus grande chez l'huître que chez l'homme, cette proportion se transforme en ~3500 mutations délétères par génome haploïde chez l'huître contre ~300 chez l'homme. Un fardeau génétique fort est donc détecté à l'échelle moléculaire. La taille génétique du génome de l'huître femelle a été estimé à environ 1000cM (Hubert & Hedgecock, 2004). Nous estimons donc qu'il y a 3.5 mutations délétères par cM en moyenne dans un génome haploïde d'huître. Dans l'étude de Launey et Hedgecock (2001) un marqueur génétique marquait en moyenne 36 cM de génome, soit 126 mutations délétères d'après nos calculs. Il semble donc que ce qui a été inféré comme une mutation délétère à effet fort puisse très bien être en réalité la somme de nombreuses mutations à effet plus faible.

Cette analyse préliminaire montre la richesse d'information contenue dans les données de séquences codantes. Nous prévoyons d'utiliser des méthodes plus sophistiquées permettant d'inférer l'effet des mutations sur la fitness (Bustamante *et al.*, 2003). De plus, il sera intéressant d'explorer s'il existe des variations entre gène. En effet, la diversité protéique est en générale trop faible pour faire des inférences gène par gène, mais chez l'huître ce type d'inférence semble possible.

## VII- Conclusions.

Les données moléculaires que nous souhaitons produire dans le cadre du projet Bivalvomix ont dans l'ensemble été obtenues. Ils nous reste un effort à faire dans l'obtention de données de divergence qui se sont révélées plus délicates à obtenir que prévu (problème d'amplification par PCR). Nous sommes maintenant dans la phase d'analyse des données et de rédaction d'articles scientifiques. A ce stade, les résultats majeurs sont les suivants :

- Le génome des Mytilidés est pauvre en GC, il n'est pas structuré en isochore et il existe un effet fort de la sélection pour l'usage des codons synonymes optimaux pour la traduction. Le génome de l'huître ressemble à celui de la drosophile en terme de variation du contenu en base des régions codantes et non-codantes. Ces caractéristiques des génomes de Bivalves ne semblent pas être la résultante d'une taille efficace importante.
- Nous avons mis au point une technique de clonage permettant de produire une grosse quantité de séquence pour l'analyse du polymorphisme.
- L'analyse des premiers gènes chez *Mytilus* a révélé un contraste spectaculaire entre portions du génome avec un fort impact de la sélection sur la généalogie des allèles.
- L'analyse des premiers gènes chez *Bathymodiolus* a permis d'affiner l'étude de la divergence entre les espèces atlantiques et pacifiques ainsi que de dévoiler l'effet possible de balayage sélectif sur certains gènes. Une analyse plus fine des données nous permettra de quantifier les effets.
- L'analyse des premiers gènes chez *Crassostrea* s'est révélée plus complexe que prévu car la forte diversité observée chez cette espèce s'accompagne d'un effet fort de la recombinaison intragénique. Les résultats obtenus montrent que l'introduction récente des populations en Europe n'a probablement pas eu de gros impact sur la diversité génétique et que c'est en fait l'histoire des populations asiatiques que nous cherchons à retracer.
- Nous avons estimé un fardeau de mutations délétères extrêmement fort en terme de nombre de mutations chez l'huître (~3500 mutations délétères par génome haploïde). L'effet des mutations sur la fitness devrait pouvoir être inféré avec l'aide de méthodes statistiques appropriées.

## Références citées

- Akashi, H., and A. Eyre-Walker 1998 Translational selection and molecular evolution. *Curr. Op. Genet. Dev.* **8**: 688-693.
- Akashi, H., R. M. Kliman and A. Eyre-Walker 1998 Mutation pressure, natural selection, and the evolution of base composition in *Drosophila*. *Genetica* **102/103**: 49-60.
- Bataillon, T., and M. Kirkpatrick 2000 Inbreeding depression due to mildly deleterious mutations in finite populations: size does matter. *Genet Res* **75**: 75-81.
- Bazin, E., S. Glemin and N. Galtier 2006 Population size does not influence mitochondrial genetic diversity in animals. *Science* **312**: 570-572.
- Bierne, N., F. Bonhomme, P. Boudry, M. Szulkin and P. David 2006 Fitness landscapes support the dominance theory of post-zygotic isolation in the mussels *Mytilus edulis* and *M. galloprovincialis*. *Proc. R. Soc. B* **in press**.
- Bierne, N., P. Borsa, C. Daguin, D. Jollivet, F. Viard *et al.* 2003 Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. *Molecular Ecology* **12**: 447-462.
- Bierne, N., and A. Eyre-Walker 2004 The genomic rate of adaptive amino-acid substitution in *Drosophila*. *Mol Biol Evol* **21**: 1350-1360.
- Bierne, N., S. Launey, Y. Naciri-Graven and F. Bonhomme 1998 Early effect of inbreeding as revealed by microsatellite analyses on *Ostrea edulis* larvae. *Genetics* **148**: 1893-1906.
- Boudry, P., S. Heurtebise, B. Collet, F. Cornette and A. Gérard 1998 Differentiation between populations of the portuguese oyster, *Crassostrea angulata* and the pacific oyster *Crassostrea gigas* revealed by mtDNA RFLP analysis. *Journal of Experimental Marine Biology and Ecology* **226**: 279-291.
- Buroker, N. E., W. K. Hershberger and K. K. Chew 1979 Populations genetics of the family Ostreidae. I. Intraspecific studies of *Crassostrea gigas* and *Saccostrea commercialis*. *Marine Biology* **54**: 157-169.
- Bustamante, C. D., A. Fledel-Alon, S. Williamson, R. Nielsen, M. T. Hubisz *et al.* 2005 Natural selection on protein-coding genes in the human genome. *Nature* **437**: 1153-1157.
- Bustamante, C. D., R. Nielsen and D. L. Hartl 2003 Maximum likelihood and Bayesian methods for estimating the distribution of selective effects among classes of mutations using DNA polymorphism data. *Theor Popul Biol* **63**: 91-103.
- Bustamante, C. D., R. Nielsen, S. A. Sawyer, K. M. Olsen, M. D. Purugganan *et al.* 2002 The cost of inbreeding in *Arabidopsis*. *Nature* **416**: 531-534.
- Cutter, A. D., J. D. Wasmuth and M. L. Blaxter 2006 The Evolution of Biased Codon and Amino Acid Usage in Nematode Genomes. *Mol Biol Evol* **in press**.
- Daguin, C., F. Bonhomme and P. Borsa 2001 The zone of sympatry and hybridization of *Mytilus edulis* and *M. galloprovincialis*, as described by intron length polymorphism at locus mac-1. *Heredity* **86**: 342-354.
- Duret, L. 2002 Evolution of synonymous codon usage in metazoans. *Curr. Opin. Genet. Dev.* **12**: 640-649.
- Duret, L., and D. Mouchiroud 1999 Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **96**: 4482-4487.
- Evans, F., S. Matson, J. Brake and C. Langdon 2004 The effects of inbreeding on performance traits of adult Pacific oysters (*Crassostrea gigas*). *Aquaculture* **203**: 89-98.
- Excoffier, L., P. E. Smouse and J. M. Quattro 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479-491.
- Eyre-Walker, A. 2002 Codon Usage Bias in *Encyclopedia of Evolution*.
- Eyre-Walker, A. 2006 The genomic rate of adaptive evolution. *TREE* **21**: 569-575.
- Eyre-Walker, A., and L. D. Hurst 2001 The evolution of isochores. *Nature Rev. Genet.* **2**: 549-555.
- Eyre-Walker, A., and P. D. Keightley 1999 High genomic deleterious mutation rates in hominids. *Nature* **397**: 344-347.

- Falconer, D. S., and T. F. C. Mckay 1996 *Introduction to quantitative genetics, 4th ed.* Longman Group Ltd., Harlow, Essex, England.
- Fay, J. C., G. J. Wycoff and C.-I. Wu 2001 Positive and negative selection on the human genome. *Genetics* **158**: 1227-1234.
- Galtier, N., E. Bazin and N. Bierne 2006 GC-biased segregation of noncoding polymorphisms in *Drosophila*. *Genetics* **172**: 221-228.
- Gillespie, J. H. 1999 The role of population size in molecular evolution. *Theor. Pop. Biol.* **55**: 145-156.
- Hill, W. G., and A. Robertson 1968 Linkage disequilibrium in finite populations. *Theoret. Appl. Genet.* **38**: 226-231.
- Hubert, S., and D. Hedgecock 2004 A linkage map of microsatellite DNA markers for the Pacific oyster *Crassostrea gigas*. *Genetics* **168**: 351-362.
- Hudson, R. R., and N. L. Kaplan 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**: 147-164.
- Huson, D. H., and T. H. Klopper 2005 Computing recombination networks from binary sequences. *Bioinformatics* **21**: 159-165.
- Huvet, A., P. Boudry, M. Ohresser, C. Delsert and F. Bonhomme 2000a Variable microsatellites in the Pacific cupped oyster *Crassostrea gigas* and other cupped oyster species. *Anim Genet* **31**: 71-72.
- Huvet, A., C. Fabioux, H. McCombie, S. Lapegue and P. Boudry 2004 Natural hybridization between genetically differentiated populations of *Crassostrea gigas* and *C. angulata* highlighted by sequence variation in flanking regions of a microsatellite locus. *Marine Ecology-Progress Series* **272**: 141-152.
- Huvet, A., A. Gérard, C. Ledu, P. Phélipot, S. Heurtebise *et al.* 2002 Is fertility of hybrids enough to conclude that the two oysters *Crassostrea gigas* and *Crassostrea angulata* are the same species ? *Aquatic Living Resources* **15**: 45-52.
- Huvet, A., S. Lapègue, A. Magoulas and P. Boudry 2000b Mitochondrial and nuclear DNA phylogeography of *Crassostrea angulata*, the Portuguese oyster endangered in Europe. *Conservation Genetics* **1**: 251-262.
- Kimura, M. 1983 *The neutral theory of molecular evolution.* Cambridge University Press, Cambridge, UK.
- Launey, S., and D. Hedgecock 2001 High Genetic Load in the Pacific Oyster *Crassostrea gigas*. *Genetics* **159**: 255-265.
- Launey, S., C. Ledu, P. Boudry, F. Bonhomme and Y. Naciri-Graven 2002 Geographic structure in the European flat oyster, *Ostrea edulis* L, as revealed by microsatellite polymorphism. *J. Hered.* **93**: 331-351.
- Leitão, A., R. Chaves, S. Santos, H. Guedes-Pinto and P. Boudry 2004 Restriction enzyme digestion chromosome banding in *Crassostrea* and *Ostrea* species: Comparative karyological analysis within *Ostreidae*. *Genome* **47**: 781-788.
- López, J. L., E. Mosquera, J. Fuentes, A. Marina, J. Vázquez *et al.* 2001 Two-dimensional gel electrophoresis of *Mytilus galloprovincialis*: differences in protein expression between intertidal and cultured mussels. *Mar Ecol Prog Ser* **224**: 149-146.
- Mitta, G., F. Hubert, E. A. Dyrinda, P. Boudry and P. Roch 2000 Mytilin B and MGD2, two antimicrobial pep-tides of marine mussels: gene structure and expression analysis. *Developmental & Comparative Immunology* **24**: 381-393.
- Moriyama, E. N., and J. R. Powell 1997 Codon usage bias and tRNA abundance in *Drosophila*. *J. Mol. Evol.* **45**: 514-523.
- Mousset, S., N. Derome and M. Veuille 2004 A test of neutrality and constant population size based on the mismatch distribution. *Mol Biol Evol* **21**: 724-731.
- Murray, M. C., and M. P. Hare 2006 A genomic scan for divergent selection in a secondary contact zone between Atlantic and Gulf of Mexico oysters, *Crassostrea virginica*. *Mol Ecol* **15**: 4229-4242.
- O'Foighil, D., P. M. Gaffney, A. E. Wilbur and T. J. Hilbish 1998 Mitochondrial cytochrome oxidase I gene sequences support an Asian origin for the Portuguese oyster *Crassostrea angulata*. *Marine Biology* **131**: 497-503.
- Ritland, K. 1996 Inferring the genetic basis of inbreeding depression in plants. *Genome* **39**: 1-8.

- Solé-Cava, A. M., and J. P. Thorpe 1991 High levels of genetic variation in natural populations of marine lower invertebrates. *Biological Journal of the Linnean Society* **44**: 65-80.
- Wang, R. L., J. Wakeley and J. Hey 1997 Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics* **147**: 1091-1106.
- Yang, Z., and J. P. Bielawski 2000 Statistical methods for detecting molecular adaptation. *TREE* **15**: 496-503.