



UNIVERSITE MONTPELLIER II
SCIENCES ET TECHNIQUES DU LANGUEDOC

THESE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE MONTPELLIER II

Discipline : Evolution, Ecologie, Ressources génétiques, Paléontologie

Ecole doctorale : SIBAGHE

Présentée et soutenue publiquement par

Célia GOSSET

Le 18 décembre 2012

Bases génétiques et histoire de la différenciation adaptative chez *Mytilus*

Jury :

Patrice DAVID, Directeur de recherche, CNRS Montpellier
Jean-Christophe SIMON, Directeur de recherche, INRA Rennes
Anne CHENUIL, Chargé de recherche, CNRS Marseille
Sylvie LAPEGUE, Cadre de recherche, IFREMER La Tremblade
Nicolas BIERNE, Directeur de recherche, CNRS Sète

Président
Rapporteur
Rapporteur
Examineur
Directeur de thèse



Résumé

Les bases génétiques et l'histoire de la différenciation adaptative ont été étudiées chez les moules du complexe d'espèces *Mytilus edulis* qui représente un modèle d'étude intéressant pour mieux comprendre comment se propage et se distribue la différenciation adaptative en population structurée. Grâce à la technique AFLP, une approche de génomique des populations (« scan génomique ») a été utilisée pour mesurer la différenciation entre des populations isolées sur des échelles de temps et d'espaces contrastées. Notre objectif était de vérifier si la différenciation génétique n'avait pas une origine plus complexe qu'habituellement proposé. Trois parties s'articulent autour de cette question centrale. La première s'intéresse à la différenciation entre l'Atlantique et la Méditerranée chez l'espèce *M. galloprovincialis* et a montré que la structure génétique était la conséquence d'un différentiel d'introgression avec l'espèce sœur *M. edulis*. Dans la deuxième partie de ce travail nous avons mis en évidence qu'une sélection directe, soit balancée soit intermittente, sur un polymorphisme pré-existant expliquait le niveau de différenciation anormalement élevé d'un gène de l'immunité entre populations de la côte européenne chez *M. edulis*. Enfin, la troisième partie s'est intéressée à revisiter un exemple classique de la littérature de la génétique des populations: le cas des *M. edulis* du détroit de Long Island et a permis de suggérer que la structure observée à très petite échelle spatiale correspondait probablement à un contact secondaire entre des moules européennes introduites et les moules américaines. D'une manière générale nos résultats démontrent que, quelque soit l'échelle à laquelle nous nous plaçons, la différenciation génétique tire son origine d'une histoire souvent plus complexe qu'attendu.

Mots-clés: Différenciation adaptative, adaptation locale, *Mytilus*, scan génomique, sélection, polymorphisme, AFLP

Abstract

The genetic basis and history of adaptive differentiation were studied in the species complex *Mytilus edulis* which is an interesting model system to understand how adaptive differentiation spreads and structure itself in subdivided populations. Using the AFLP technique, a genome scan approach was undertaken to measure differentiation between populations on contrasted spatial and temporal scales. Our objective was to verify whether the origin of genetic differentiation could be more complex than anticipated. This question was addressed in three chapters. The first one focuses on the differentiation between populations of the Atlantic Ocean and the Mediterranean Sea in *M. galloprovincialis*. Our results show that the genetic structure was the result of differential introgression with the sister hybridizing species *M. edulis*. In the second chapter we demonstrated that direct selection on a pre-existing polymorphism explained the unusually high level of differentiation at a defensin locus between populations of the European coast in *M. edulis*. Finally, in the third chapter we revisited a classic example of the population genetics literature: the aminopeptidase cline in *M. edulis* populations of the Long Island Sound. We obtained evidence that the genetic differentiation observed at a very fine spatial scale in the sound was the consequence of a secondary contact between introduced mussels from Europe and indigenous American mussels. Whatever the spatio-temporal scale at which we analyzed genetic differentiation, its origin proved to originate from an unsuspectedly long and complex history.

Keywords: Adaptive differentiation, local adaptation, *Mytilus*, genome scan, selection, polymorphism, AFLP

REMERCIEMENTS

Au terme de ce travail, c'est avec émotion que je tiens à remercier toutes les personnes qui, de près ou de loin, ont contribué à la réalisation de ce projet.

Je tiens tout d'abord à adresser mes remerciements les plus sincères à Nicolas Bierne pour avoir dirigé cette thèse et m'avoir permis de la réaliser dans les meilleures conditions. Je suis consciente de la chance que j'ai eu de pouvoir travailler avec un directeur sachant se rendre disponible pour ses étudiants, même lorsqu'un petit Timéo pointe le bout de son nez. Merci Nico pour ton soutien et la confiance que tu m'as accordé dans cette aventure. Tout au long de ces trois années, tu as su m'orienter en me faisant découvrir ta vision des choses, en m'apprenant à chercher toutes les hypothèses et à voir plus loin.

Je tiens aussi à remercier les rapporteurs Anne Chenuil et Jean-Christophe Simon d'avoir ainsi que les examinateurs Sylvie Lapègue et Patrice David pour avoir accepté d'évaluer cette thèse.

Merci également à François Bonhomme pour son accueil au sein de l'équipe et au sein de la station. Même avec ta casquette de chef, tu sais rester simple, accessible (humainement et pas forcément physiquement !) et souvent drôle avec tes célèbres contrepétories.

Merci à tant d'entre vous pour les discussions scientifiques, et les conseils avisés sur l'avancée de ma thèse, pour réussir une PCR ou une analyse bayésienne...

Merci au groupe des filles de la station ! Certaines passent, d'autres restent mais le café dans le bureau de Marie-T persiste. Honneur à celles qui restent, j'entends par là celles qui seront encore là quand je serais partie. Merci à Marie-T pour les pauses café et pour son écoute. Merci d'avoir supporté mes moments de doute, d'espoir, de peur, de confusion... Merci de m'avoir rassuré quand ça n'allait pas et merci pour ton aide précieuse dans les manip. Qu'est-ce que j'aurais fait sans toi ?!? Merci aussi à Yoyo, Cathy et Marie-Odile pour leur accueil et la chaleur humaine. Ça fait du bien quand on arrive de loin. Et merci à toutes celles qui sont parties ou qui étaient de passage pour les bons moments passés au travail : Lamya, Lolo, Audrey et Cathy B. Merci à toutes pour la bonne ambiance qui règne à la station. Mes

collègues thésardes, je ne vous oublie pas ! Merci d'être là car ma première année « seule » à la station était beaucoup moins agréable que depuis que vous êtes arrivées : Soumaya, Tahani et Christelle (la spécialiste de la chasse aux coquilles).

Et un grand merci à tous les autres : Pierre-Alexandre, Camillo, Jean-François, Christophe (x2), Emilie, David, Michel, Claude, Gérard, Laurent, Hervé, Nicolas, Sophie... que ce soit pour des petites conversations à l'angle d'un couloir, les pauses déjeuner du midi ou les réunions un peu plus arrosées du vendredi soir.

Un immense merci à celles et ceux de mon entourage qui m'ont soutenu. Merci à Bart de m'avoir suivi dans cette aventure et d'avoir supporté mes horaires et mon humeur massacrant lors des grandes étapes de ce travail. Enfin, merci à mes amis qui, malgré la distance et mon incapacité à voir le temps passer, ont toujours été là pour moi.

Le dernier témoignage de ma gratitude ira à Jean-Pierre Geron, conseiller au Service Formation Continue de l'Université du Havre et Professeur de Mathématiques, sans qui je n'aurais jamais eu l'audace de commencer des études universitaires avec un DAEU. Merci d'avoir cru en mes capacités et de m'avoir insufflé la confiance qu'il me manquait.

SOMMAIRE

SOMMAIRE	1
INTRODUCTION	4
1- Génomique des populations « étape 1 » : identifier les locus sous sélection à l'aide des scans génomiques de la différenciation	6
1.1- Les tests d'outliers	6
1.2- Les méthodes moléculaires pour scanner les génomes	7
2- Comment interpréter une différenciation génétique anormalement forte ?	9
2.1- L'adaptation locale	9
2.2- L'hypothèse du « balayage global », comment une mutation avantageuse peut laisser une trace sur la structure génétique	13
2.3- Les barrières génétiques multifactorielles	15
3- Génomique des populations « étape 2 » : l'histoire et les bases génétiques de l'adaptation par une analyse détaillée des locus identifiés	19
4- Contexte de la thèse	21
4.1- Originalité des organismes marins	21
4.2- Un laboratoire naturel pour l'étude de la différenciation, le complexe <i>Mytilus edulis</i>	22
4.3- Objectifs de la thèse	25
 CHAPITRE I : L'introgression différentielle explique les F_{ST} outliers entre les populations d'Atlantique et de Méditerranée chez <i>Mytilus galloprovincialis</i>	27
I.1- Résumé	27
I.2- Differential introgression from a sister species explains high F_{ST} outlier loci within a mussel species	28
 CHAPITRE II: Adaptation d'un peptide antimicrobien à partir d'un polymorphisme pré-existant	51
II.1- Résumé	51
II.2- Evidence for adaptation from standing genetic variation on an antimicrobial peptide in the mussel <i>Mytilus edulis</i>	52

CHAPITRE III: Le projet Long Island.....	74
1- Scans génomiques entre habitats à petite échelle spatiale	74
1.1- Les associations génotype-environnement au locus <i>MPI</i> des balanes	75
1.2- Une distribution des génomes à petite échelle entre <i>M. edulis</i> et <i>M. trossulus</i>	76
1.3- Adaptation au gradient environnemental latitudinal dans une population panmictique d'anguilles.....	77
2- Le cline de la Lap dans le détroit de Long Island	78
2.1- Revue de la littérature existante	78
2.2- Les zones d'ombre	78
3- Structuration génétique dans le détroit de Long Island (Stratégie 1).....	80
3.1- Matériels et méthodes.....	80
3.2- Résultats	84
4- Tentative d'identification du gène de la <i>LAP</i> (Stratégie 2).....	89
5- Discussion	91
5.1- Des <i>M. edulis</i> européennes dans le détroit de Long Island	91
5.2- Des shifts génétiques jumeaux	92
 CONCLUSIONS ET PERSPECTIVES.....	 95
1- Synthèse des résultats	95
2- Perspectives.....	96
2-1 AFLPs ou des marqueurs difficiles à étudier de façon approfondie	96
2-2 Améliorer la couverture d'échantillonnage pour une étude plus approfondie de MGD2	97
2-3 Des croisements pour mieux comprendre les origines de l'introgession.....	97
 ANNEXES	 100
 REFERENCES BIBLIOGRAPHIQUES	 106

INTRODUCTION

L'objectif de la génétique des populations est d'identifier les processus évolutifs qui sous-tendent la diversité génétique dans les populations et sa structure spatiale entre populations (Lewontin 1974). Pour ce faire les généticiens des populations possèdent une batterie de modèles théoriques prédictifs issus d'un siècle de recherche, depuis les travaux des pères fondateurs Wright, Fisher et Haldane, et qui combinent les effets des forces évolutives fondamentales (*i.e.* mutation, dérive, recombinaison, migration et sélection, Table 1). Depuis moins longtemps, les généticiens des populations peuvent étudier les polymorphismes moléculaires et confronter les patrons de structure observés aux prédictions théoriques (Lewontin 1974).

Après 20 ans d'études d'une poignée de locus allozymiques, la génétique des populations s'est dotée dans un laps de temps assez court d'une panoplie de nouveaux marqueurs génétiques permettant une couverture génomique toujours plus grande. Alors qu'on comprend maintenant assez bien comment se structure la diversité génétique neutre, ce n'est que très récemment que l'étude de locus influencés par la sélection est devenue facilement et rapidement réalisable chez n'importe quelle espèce (Luikart *et al.* 2003).

La génétique des populations moderne porte un nouveau nom, « génomique des populations », et elle développe des méthodes d'inférence qui combinent processus neutres et adaptatifs pour l'étude d'un très grand nombre de polymorphismes génétiques de diverses natures. La génomique des populations porte l'espoir de finalement comprendre comment se structure la diversité adaptative dans les populations naturelles (Hedrick 2006). La première étape est d'identifier les locus influencés par la sélection. La deuxième étape est de mieux comprendre l'histoire et les bases génétiques de l'adaptation par une analyse détaillée des locus précédemment identifiés.

Table 1 : Les forces évolutives

Force évolutive	Description
Mutation	La mutation est une modification de l'information génétique héréditaire lors de sa transmission d'une génération à la suivante. Considérée comme une force évolutive peu efficace (mais voir Lynch 2007; Nei 2007), elle est à l'origine des polymorphismes génétiques sur lesquels les autres forces peuvent agir.
Dérive génétique	Cette force est la conséquence de processus stochastiques lors de la reproduction dans une population de taille finie (tirage aléatoire avec remise puisqu'un individu peut se reproduire plus d'une fois ou ne pas se reproduire). Son effet est d'autant plus important que la population est de petite taille.
Migration	La migration est une source de renouvellement de la diversité génétique car elle permet l'introduction de nouveaux variants génétiques ou la réintroduction d'allèles perdus. Elle a tendance à homogénéiser les compositions génétiques entre populations et s'oppose à la dérive génétique et la sélection locale qui tendent à faire diverger les populations.
Recombinaison	La recombinaison conduit à l'apparition, dans une cellule ou dans un individu, de gènes ou de caractères héréditaires dans une association différente de celle observée chez les cellules ou individus parentaux par un brassage lors de la méiose. Elle n'est pas toujours citée comme étant une force évolutive car elle ne modifie pas les fréquences alléliques d'un locus. Cependant elle modifie les fréquences génotypiques multilocus et constitue à la fois une source de nouveau génotype et une force assez efficace pour contrecarrer l'effet de la sélection indirecte (effet de la sélection d'un locus sur les autres locus, liés ou non ; effet Hill-Robertson).
Sélection	La sélection désigne la meilleure survie et reproduction des individus les mieux adaptés à leur environnement (sélection naturelle) ou les mieux adaptés à l'accès à l'accouplement (sélection sexuelle), et par conséquence, le fait que les caractères héréditaires responsables de cette adaptation soient mieux transmis d'une génération à l'autre.

1- Génomique des populations « étape 1 » : identifier les locus sous sélection à l'aide des scans génomiques de la différenciation

Grâce à l'essor de la génomique des populations (Luikart *et al.* 2003; Stapley *et al.* 2010), l'étude de l'adaptation chez les espèces non-modèles est rendue possible par l'analyse d'un très grand nombre de marqueurs moléculaires, une approche qualifiée de scan génomique. L'idée de ces scans est que les forces évolutives non-adaptatives, comme la dérive ou la migration, influent de la même manière sur le polymorphisme de tous les locus du génome, alors que la sélection va agir sur certains locus seulement, en augmentant ou en diminuant la différenciation génétique de ces locus par rapport au reste du génome (Cavalli-Sforza 1966). Ces locus, couramment dénommés « outliers », vont donc montrer des patrons de structure anormaux par rapport à l'attendu neutre. Par conséquent, il est extrêmement important de pouvoir identifier ces locus outlier de manière fiable et des méthodes de plus en plus sophistiquées ont été mises au point dans ce but (Foll & Gaggiotti 2008; Bazin *et al.* 2010).

1.1- Les tests d'outliers

Le premier test a été proposé dès les tous débuts de la génétique des marqueurs moléculaires par Lewontin et Krakauer (1973), le test LK. Ces auteurs ont cherché un moyen d'obtenir la variance du F_{st} , l'indice de Wright classiquement utilisé en génétique des populations pour mesurer la différenciation entre populations, dans le cas d'un modèle neutre afin de pouvoir confronter les données observées à cet attendu neutre. Malheureusement le F_{st} ne suit pas une distribution connue et elle dépend de nombreuses hypothèses du modèle populationnel considéré, que les populations naturelles ne suivent jamais à la perfection (Lewontin 2002). Lewontin et Krakauer (1973) ont cependant constaté que sous l'hypothèse que les fréquences alléliques sont distribuées normalement et que la variance calculée sur les données est très proche de la variance réelle, la statistique $(n-1)F_{st}/\overline{F_{st}}$ (où n est le nombre de populations échantillonnées et $\overline{F_{st}}$ la moyenne du F_{st} sur les locus analysés) devrait suivre une loi du Chi2 à $n-1$ degrés de liberté. Ils ont vérifié que cette approximation fonctionnait sur des données simulées, ce que Whitlock (2008) a refait 35 ans plus tard pour aboutir à la même conclusion que l'approximation à une loi du Chi2 fonctionnait très bien.

Le test LK a néanmoins été immédiatement l'objet de critiques car il existe de nombreuses raisons pour que la variance neutre soit en fait plus grande que prévu dans l'approximation du test LK (Lewontin & Krakauer 1975; Nei & Maruyama 1975; Robertson

1975a, b; Tsakas & Krimbas 1976; Nei & Chakravarti 1977; Nei *et al.* 1977). Le test LK a été suggéré peu robuste aux déviations des hypothèses sous-jacentes en cas de forts taux de mutation, de variation de *Fst* en raison de tailles efficaces inégales entre sous-populations, et de corrélation des fréquences alléliques entre dèmes géographiquement proches ou avec une historique de divergence commune. On pourrait croire que ces critiques ont contribué à ce que le test LK soit resté longtemps peu utilisé mais l'explosion du nombre de citations de cet article fondateur à partir de la fin des années 90 suggère qu'il a simplement fallu attendre que les méthodes moléculaires permettent d'analyser un grand nombre de locus pour que l'approche de Lewontin et Krakauer (1973) deviennent pertinente. En parallèle le développement de l'informatique a permis d'améliorer le test LK, tout d'abord en substituant l'approximation du Chi2 par des simulations informatiques pour obtenir la distribution du *Fst* dans un modèle neutre (Beaumont & Nichols 1996; Vitalis *et al.* 2001; Excoffier *et al.* 2009; Pérez-Figueroa *et al.* 2010; Antao & Beaumont 2011) puis par le développement d'approches probabilistes basées sur les statistiques bayésiennes et l'utilisation de simulations informatiques en substitut de la fonction de vraisemblance (Beaumont & Balding 2004; Foll & Gaggiotti 2008; Foll *et al.* 2010). La robustesse de tous ces tests à des déviations aux modèles sous-jacents reste néanmoins sujette à débat bien qu'elle ait été vérifiée par des études de simulations (Pérez-Figueroa *et al.* 2010; Narum & Hess 2011). Les tests d'outliers devraient donc être considérés comme une première étape pour identifier des locus candidats à la sélection.

1.2- Les méthodes moléculaires pour scanner les génomes

L'identification de gènes (ou de régions du génome) outliers par une approche de génomique des populations nécessite l'utilisation d'un grand nombre de marqueurs moléculaires répartis le plus uniformément possible sur tout le génome. Il s'agit d'avoir une vision globale de la variabilité du génome. Plusieurs types de marqueurs ont récemment été utilisés pour identifier des allèles sous sélection (Luikart *et al.* 2003; Nosil *et al.* 2009) dont notamment les microsatellites, les SNPs (Single Nucleotide Polymorphisms) et les AFLPs (Amplified Fragment Length Polymorphisms).

Grâce aux avancées technologiques des dernières années, l'approche des scans génomique a explosé à l'aube du nouveau millénaire. Luikart (2003) écrivait: « The ideal molecular approach for population genomics should uncover hundreds of polymorphic markers (...) that cover the entire genome in a single, simple and reliable experiment. »

Les marqueurs microsatellites ont été utilisés dès le début des années 90 pour les études de génétique des populations car ils présentaient des avantages (variables, multialléliques et co-dominants), facilitant l'utilisation des tests de détection d'outliers. Ils ont cependant l'inconvénient d'être mono-locus, ce qui nécessite la connaissance préalable d'un grand nombre de marqueurs. Ainsi, le développement de centaines de marqueurs peut vite devenir coûteux et gourmand en temps, notamment chez les espèces non-modèles.

Parce que les SNPs sont bi-alléliques, ils peuvent sous certaines conditions être moins informatifs que les microsatellites, mais cela est compensé dans le cadre des scans génomiques de la différenciation par leur abondance dans le génome. En outre, sous certaines hypothèses les SNPs peuvent être combinés en haplotypes, augmentant la quantité d'information extraite d'une région donnée de l'ADN. Ces haplotypes formés peuvent alors être analysés comme des marqueurs multi-alléliques. Le développement de nouvelles technologies de génotypage a permis l'analyse simultanée de plusieurs centaines voir de plusieurs milliers de ces marqueurs. Néanmoins, cette approche nécessitant une connaissance préalable du génome (séquençage partiel ou complet), avant de pouvoir rechercher des SNPs à grande échelle, a pour l'instant été principalement utilisée chez des organismes modèles ou à intérêt commercial (par exemple Wong *et al.* 2004; Kijas *et al.* 2009; McNally *et al.* 2009).

Avec la technique AFLP (voir Annexe 1, Vos *et al.* 1995; Vuylsteke *et al.* 2007), il a été possible d'obtenir facilement un grand nombre de polymorphismes génétiques à l'échelle du génome sans connaissance préalable. Cette méthode permet de détecter principalement des mutations ponctuelles situées au niveau de sites de coupure d'enzymes de restriction. Bien que ce type de marqueur soit bi-allélique et dominant (présence/absence d'un fragment de restriction), les AFLPs offrent surtout le grand avantage de générer rapidement et à moindre coût un très grand nombre de marqueurs (plusieurs centaines) couvrant l'ensemble du génome, rendant ces marqueurs de plus en plus utilisés pour la recherche de gènes sélectionnés chez de nombreuses espèces non-modèles (Wilding *et al.* 2001; Campbell & Bernatchez 2004; Bonin *et al.* 2006; Murray & Hare 2006; Lallias *et al.* 2007; Egan *et al.* 2008; Nosil *et al.* 2008; Wood *et al.* 2008; Gagnaire *et al.* 2009; Galindo *et al.* 2009; Paris *et al.* 2010; Colbeck *et al.* 2011; Fischer *et al.* 2011; Gagnaire *et al.* 2011; Nunes *et al.* 2011; Tice & Carlon 2011; Hoffman *et al.* 2012).

Aujourd'hui, et depuis la mise au point du premier séquenceur haut débit (Margulies *et al.* 2005), les limitations méthodologiques semblent levées. Les NGS (pour Next Generation Sequencing) ne cessent de se perfectionner (Metzker 2010; Stapley *et al.* 2010), permettant la

détection et l'identification de milliers de SNPs sans connaissance préalable du génome (Hohenlohe *et al.* 2010; Davey *et al.* 2011).

2- Comment interpréter une différenciation génétique anormalement forte ?

Si un effort important a été porté sur les méthodes d'identification des locus outliers et leur détection dans un grand nombre d'espèces (Wilding *et al.* 2001; Bonin *et al.* 2006; Murray & Hare 2006; Gagnaire *et al.* 2009; Tice & Carlon 2011), beaucoup moins d'effort a été consenti pour décliner toutes les interprétations possibles à l'observation d'une structure génétique anormalement forte (Bierne *et al.* 2011). Les locus outliers sont interprétés être sous sélection locale en oubliant d'autres processus qui affectent les génomes tels que les contraintes, le hasard, l'histoire, les effets indirects des autres forces et les conflits que peut engendrer la rencontre de différents backgrounds génomiques. L'adaptation locale est une explication possible, peut-être la plus fréquente, mais d'autres scénarios peuvent aussi expliquer ces locus outliers.

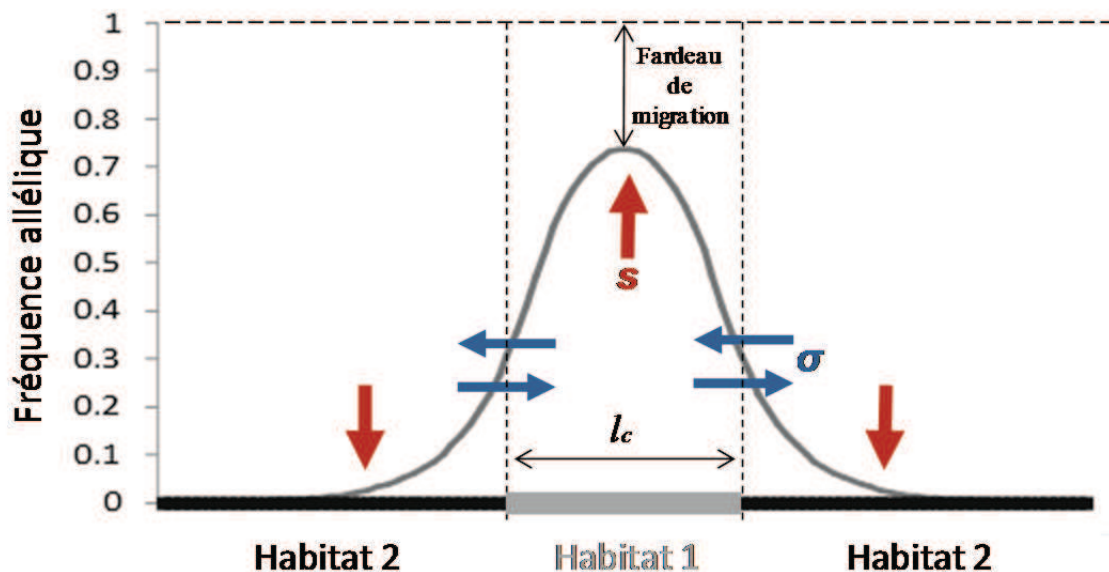
2.1- L'adaptation locale

L'hétérogénéité de l'environnement dans l'aire de répartition géographique des espèces peut exercer des pressions sélectives qui agissent pour maximiser la valeur adaptative des individus dans leur environnement spécifique (Kawecki & Ebert 2004). Ce processus est connu comme l'adaptation locale.

De nombreuses études suggèrent que l'adaptation locale est un phénomène relativement courant (voir par exemple Leimu & Fischer 2008; Hereford 2009; Fraser *et al.* 2011; Sanford & Kelly 2011). Cependant, la capacité d'adaptation n'est pas une conséquence inéluctable même si différentes pressions sélectives sont à l'œuvre. La capacité de réponse à la sélection va être compromise par d'autres mécanismes évolutifs tels que la dérive génétique et le flux génique (Wright 1931; Hedrick 2000; Ronce & Kirkpatrick 2001; Alleaume-Benharira *et al.* 2006).

2.1.1- Les conditions d'existence d'une adaptation locale

L'observation de Mayr (1963) que la divergence phénotypique est souvent corrélée au degré d'isolement des populations suppose que le flux de gènes peut empêcher l'adaptation locale. La migration homogénéise la fréquence des allèles entre habitats et limite la réponse locale à la sélection par un flux continu d'allèles maladaptés (Endler 1977; Slatkin 1985, 1987; Garcia-Ramos & Kirkpatrick 1997; Lenormand 2002; Hartl & Clark 2007; Rasanen & Hendry 2008). Il en résulte des clines de fréquence allélique au niveau de la transition environnementale des deux habitats (Par exemple Koehn *et al.* 1980; Mallet *et al.* 1990; Berry & Kreitman 1993; Powers & Schulte 1998; Lenormand *et al.* 1999; Schmidt & Rand 2001; Hoekstra *et al.* 2004). La largeur de ces clines est généralement plus grande que la transition de l'environnement et elle peut être utilisée pour estimer les effets opposés de la sélection et du flux de gènes (Barton & Gale 1993). L'influx d'allèles maladaptés maintient l'allèle avantageux à une fréquence intermédiaire dans l'habitat où il est favorisé, produisant un fardeau génétique de migration (habitat 1 dans la figure 1).



Si on considère de grandes populations, on peut négliger l'effet de dérive génétique puisque très faible en comparaison des autres forces. Dans ce contexte, les théoriciens ont pu étudier les effets antagonistes de la migration et de la sélection (Slatkin 1973; Nagylaki 1975;

Endler 1977). La figure 1 représente le patron de fréquence d'un allèle avantageux (coefficient de sélection s) dans un patch d'habitat 1 et délétère dans l'habitat 2 qui entoure ce patch. La migration est mesurée par σ , qui représente l'écart-type de la distance parent-descendant dans un environnement à une dimension. A l'équilibre migration/sélection, la fréquence de l'allèle avantageux (et donc le niveau du fardeau de migration) dans l'habitat 1 dépend du rapport σ/s . Slatkin (1973) définit la longueur caractéristique comme étant la taille minimale du patch d'habitat 1 pour que l'allèle favorable dans cet habitat soit maintenu et montre que $lc = \sigma/\sqrt{s}$ (Slatkin 1973). L'existence de lc est liée au concept de « grain de l'environnement » de Levins (1968). Dans un environnement à « grain fin » la variation spatiale de l'environnement se fait à une échelle inférieure à l'échelle de dispersion de l'espèce qui a donc du mal à répondre à la sélection alors que dans un environnement à grains grossiers, la variation spatiale de l'environnement se fait à une échelle supérieure à l'échelle de dispersion si bien que la réponse à la sélection est beaucoup plus facile. Slatkin amende ce concept en précisant que le concept de grain de l'environnement dépend non seulement de la mobilité de l'espèce, mais aussi de la force de la sélection appliquée au caractère. Par conséquent, il est possible d'avoir une situation dans laquelle un environnement est grossier par rapport à un caractère subissant une forte sélection, et fin par rapport à un autre, sur lequel agit une sélection relativement faible (Slatkin 1973).

Finalement, le gene swamping, une forme extrême de fardeau de migration, s'explique par un flux continu de migrants maladaptés, empêchant la réponse à la sélection lorsque la taille du patch est plus petite que lc . Dans ce cas, le fardeau de migration est maximal.

2.1.2- Les limites démographiques dans les petites populations

La dérive génétique va augmenter avec la diminution de la taille efficace des sous-populations (Robertson 1960; Hill 1982). La capacité d'adaptation peut être compromise dans les petites populations en raison d'une diversité génétique réduite (Antonovics 1976; Stockwell *et al.* 2003; Pertoldi *et al.* 2007), limitant la cible de la sélection. Les petites populations peuvent ne pas être bien adaptées à leur environnement car la dérive peut rendre difficile la montée en fréquence du caractère adaptatif (Whitlock 2003), que ce soit par perte de l'allèle avantageux ou par diminution du potentiel adaptatif lié au fardeau génétique. En effet, la fixation d'allèles délétères peut conduire à une valeur adaptative réduite ou à l'extinction de la population (Lynch & Gabriel 1987; Lande 1994; Whitlock *et al.* 2000).

2.1.3- Découplage entre différenciation phénotypique et génétique

De nombreux traits et combinaisons de traits peuvent être impliqués dans l'adaptation locale. Cependant, il semble probable que de nombreux événements adaptatifs, voire la plupart, se produisent par l'évolution de traits polygéniques, plutôt que par la fixation d'une mutation bénéfique unique (Pritchard & Di Rienzo 2010).

Récemment, des études d'association sur l'ensemble du génome humain (Yang *et al.* 2010; Stranger *et al.* 2011) et dans différents organismes modèles (par exemple chez *Arabidopsis thaliana*: Atwell *et al.* 2010; chez *Drosophila melanogaster*: Jumbo-Lucioni *et al.* 2010) ont confirmé de petits changements de fréquences alléliques à de nombreux locus dispersés dans l'ensemble du génome (Mackay *et al.* 2009). Seules quelques études ont porté sur la différenciation des traits et de leurs gènes sous-jacents, et ces études ont conclu qu'il existe des écarts considérables entre ces deux mesures (Hall *et al.* 2007; Luquez *et al.* 2008; Namroud *et al.* 2008). Des prédictions théoriques appuyées par des simulations ont en effet indiqué qu'il peut y avoir des écarts importants entre la différenciation phénotypique et la différenciation des gènes contrôlant le trait, comme le suggèrent Le Corre et Kremer (2003) ou Santure et Wang (2009). D'autres études ont pu montrer que cet écart provient de déséquilibres accumulés par la sélection intergénique qui découple la différenciation au niveau des traits de celle montrée par les gènes (Latta 1998; McKay & Latta 2002; Le Corre & Kremer 2003; Latta 2004; Le Corre & Kremer 2012).

2.1.4- Effet d'un locus sous sélection locale sur son voisinage chromosomique

Il a été démontré théoriquement qu'une adaptation locale pouvait augmenter la structure génétique de locus neutres liés à un locus sous sélection (Petry 1983; Charlesworth *et al.* 1997). Pour un site proche d'un locus sous sélection locale, de fortes différenciations entre populations peuvent être générées. La sélection locale va donc contribuer à la mise en place d'îlots génomiques de différenciation responsables d'une réduction du flux génique entre populations aux régions du génome considérées.

Si l'on s'intéresse à la dynamique de ce processus, l'empreinte laissée sur le chromosome est générée en deux phases. La première phase se caractérise par la montée en fréquence du locus sous sélection et l'entraînement des locus neutres voisins, entraînant une augmentation de la différenciation entre populations par effet d'autostop génétique (Pollinger *et al.* 2005; Bierne 2010). La seconde phase correspond à une diminution de la différenciation par effet du flux génique et de la recombinaison (Miller & Hawthorne 2005). Alors que la

première phase est très rapide, la seconde phase va plus ou moins lentement ré-homogénéiser les fréquences alléliques sous l'effet de la migration et de la recombinaison. Parce que l'homogénéisation est plus rapide pour les locus neutres éloignés du locus sélectionné, le pic de F_{ST} se resserre autour du locus sélectionné. Le système converge vers un état d'équilibre entre migration, dérive et effet indirect de la sélection (Petry 1983; Charlesworth *et al.* 1997) dans lequel la migration efficace décroît de façon inversement proportionnelle à la distance génétique entre le locus sélectionné et le locus neutre. L'attendu théorique propose donc un pic de F_{ST} centré sur le locus sélectionné qui ne génère une différenciation neutre que sur une région chromosomique peu étendue (Charlesworth *et al.* 1997; Barton 2000; Feder & Nosil 2010), ce qui se vérifie dans les scans génomiques en comparant des populations où seule l'adaptation locale intervient, comme par exemple entre les sites à sol composé de serpentine et les sites sans serpentine chez *Arabidopsis lyrata* (Turner *et al.* 2010).

2.1.5- Effet de l'adaptation locale sur la structure génétique et le paradoxe des scans génomiques

L'adaptation locale génère de la structure génétique et peut expliquer les outliers identifiés dans les scans génomiques mais plusieurs arguments suggèrent que la proportion de polymorphismes maintenus par l'adaptation locale dans un environnement hétérogène devrait être faible : (i) les conditions d'existence d'une adaptation locale peuvent être limitées quand l'environnement est à grain fin et que les populations sont de petite taille, (ii) la différenciation qui existe sur le phénotype ne résulte pas forcément en une structure génétique sur les gènes qui contrôlent le trait et (iii) même si la structure est forte sur les gènes sélectionnés, ils n'affectent qu'une petite région chromosomique, et pendant de courtes périodes.

On peut donc se poser la question de pourquoi il semble facile de trouver autant d'outlier quand l'adaptation locale est censée être si rare (Hedrick 2006). Ce paradoxe des scans génomiques peut être expliqué par des modèles alternatifs (Bierne *et al.* 2011).

2.2- L'hypothèse du « balayage global », comment une mutation avantageuse peut laisser une trace sur la structure génétique

Le modèle d'autostop génétique de Maynard Smith et Haigh (1974) montre qu'une mutation favorable qui se fixe dans une population peut entraîner avec elle un allèle neutre d'un locus lié. Les premières études sur les organismes modèles montrent qu'une part importante des

polymorphismes neutres a été affectée par un autostop génétique produit par la fixation d'une mutation favorable à un locus voisin (Nielsen 2005; Li & Stephan 2006). L'effet sur la structure génétique dans une population structurée spatialement a été beaucoup moins étudié (Slatkin & Wiehe 1998; Barton 2000; Bierne 2010; Kim & Maruki 2011) et surtout ces modèles ont été négligés pour interpréter les locus outliers.

Dans les grandes populations, la structure neutre peut être faible malgré des migrations faibles entre sous-populations. Dans ces conditions, la propagation d'une mutation avantageuse augmentera la différenciation des locus neutres qui lui sont liés. Au fur et à mesure que la mutation avantageuse se propage, la recombinaison casse les associations entre le locus sélectionné et le locus neutre si bien que l'effet d'autostop s'estompe et un gradient de fréquence allélique est généré au locus neutre (Barton 2000). L'idée qu'un cline de fréquence allélique puisse être créé par ce processus, ce qu'on pourrait appeler un cline d'autostop, est largement ignorée comme étant une explication possible des clines observés dans les populations naturelles (Bierne 2010). Ce type d'autostop en population structurée laissera une empreinte chromosomique caractéristique sur la diversité et la structure génétique autour du locus sélectionné : (i) comme l'effet d'autostop est plus fort là où apparaît la mutation favorable que loin de cette origine, la vallée de diversité sera de plus en plus étroite au fur et à mesure qu'on s'éloigne du lieu d'origine de la mutation favorable, (ii) la différenciation entre populations sera plus forte pour des coefficients de recombinaison intermédiaire, créant deux pics de différenciation jumeaux de part et d'autre du locus sélectionné (Bierne 2010).

Quand la structure génétique neutre est forte, par exemple parce que les populations sont de petite taille, l'autostop en population structurée aura l'effet inverse, réduisant la différenciation génétique autour du locus sélectionné par rapport au reste du génome (Santiago & Caballero 2005). En fait, on peut comprendre ce processus comme étant de l'introggression adaptative c'est-à-dire le passage d'un allèle avantageux d'un fond génétique dans un autre (Pialek & Barton 1997).

Poursuivant l'étude d'un locus outlier détecté par Faure *et al* (2008), Bierne (2010) a étudié des polymorphismes voisins sur le chromosome. Il a obtenu la preuve qu'un autostop global en population structurée, produit par un locus estimé être situé à 3 Kb en 5' du locus outlier, expliquait la structure forte observée à ce locus entre deux populations de moules de l'espèce *Mytilus edulis*.

2.3- Les barrières génétiques multifactorielles

2.3.1- Les incompatibilités génétiques

Une forte différenciation entre populations peut aussi s'interpréter par l'existence d'incompatibilités génétiques. Dobzhansky (1937) et Muller (1942) ont proposé que les allèles responsables de la dépression hybride (voir encadré sur l'isolement reproducteur) soient avantageux dans leur background génétique habituel mais délétères lorsqu'ils sont associés aux allèles d'une autre espèce. Le modèle impliquant plusieurs locus avec épistasie a été appelé modèle de Dobzhansky-Muller (Orr 1995).

Considérons le cas le plus simple avec deux locus bialléliques de génotype AABB chez l'espèce ancestrale (Figure 2). Dans une population, une mutation 'a' apparaît au locus 1 et atteint la fixation, ce qui donne l'aaBB, qui est fertile et viable. Dans une autre population (géographiquement séparée), une mutation 'b' apparaît au locus 2 et se fixe, ce qui donne AAbb, qui est aussi fertile et viable. Le point essentiel est que, tandis que 'a' et 'b' se comportent normalement dans leurs backgrounds génétiques réciproques, la formation d'hybrides lors d'un contact secondaire va permettre aux deux allèles dérivés de se retrouver combinés dans le même génotype. Les valeurs sélectives de ces combinaisons jusque là jamais soumises à l'action de la sélection naturelle peuvent être délétères et être responsables d'un isolement reproductif.

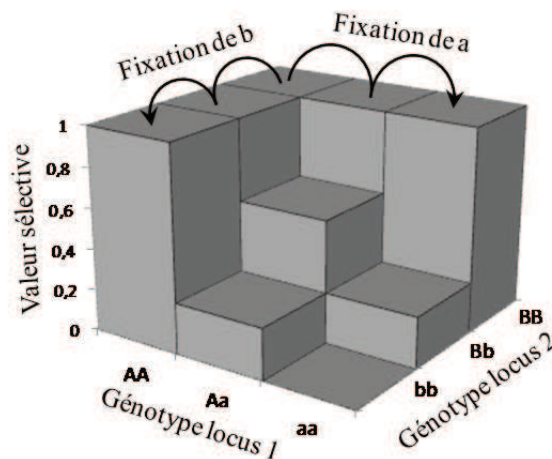


Figure 2 : Modèle Dobzhansky-Muller. Paysage des valeurs sélectives des génotypes pour deux locus bialléliques.

Ces locus d'isolement forment des clines stables à l'équilibre migration-dérive (Gavrilets 1997; Barton & Shpak 2000) et constituent les "verrous" des génomes car ils s'opposent localement à l'homogénéisation complète de ceux-ci (Figure 3).

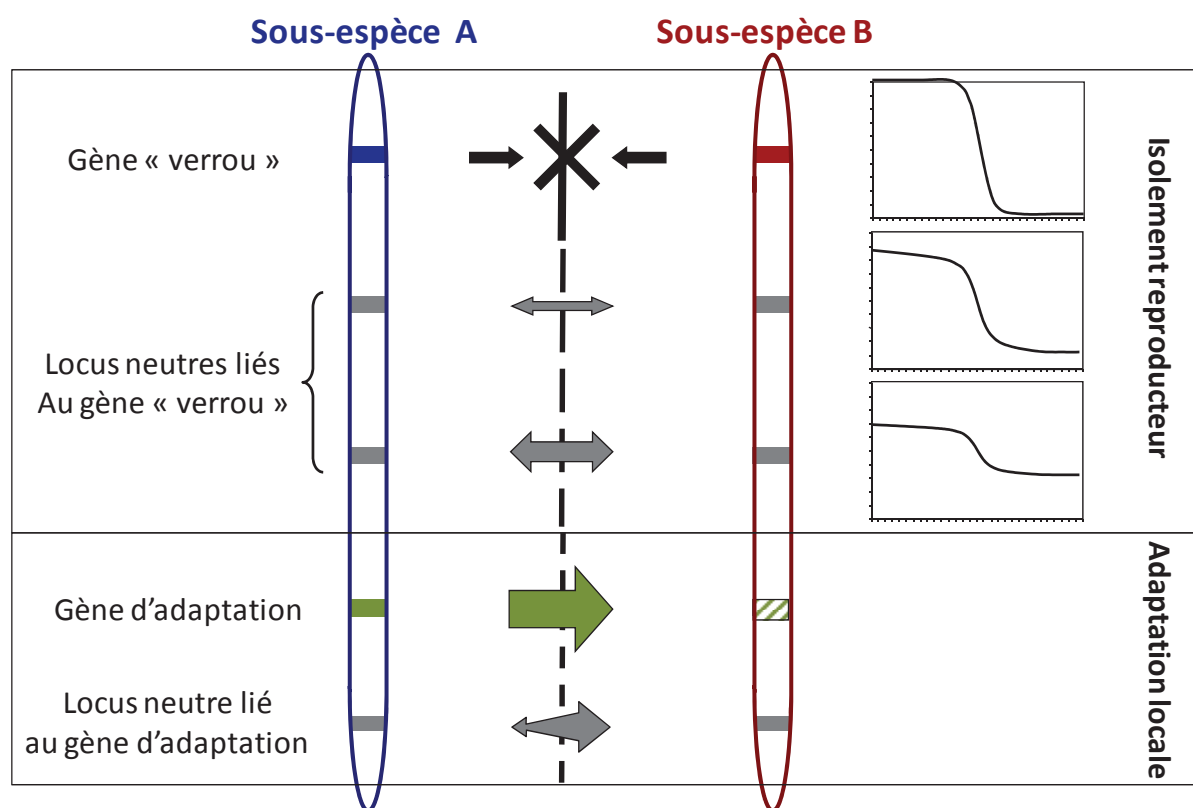


Figure 3 : Schématisation de deux fonds génétiques (représentés chacun par un chromosome) séparés par une barrière semi-perméable au flux génétique.

2.3.2- Que nous apprennent les zones hybrides sur l'interprétation des scans génomiques de la différenciation

Les zones hybrides sont des régions étroites dans lesquelles des populations génétiquement distinctes se rencontrent, se reproduisent et forment des hybrides (Barton & Hewitt 1985). Ces zones agissent comme des barrières dites génétiques, réduisant le flux de gènes neutres. Dans le cas des zones hybrides, plusieurs types de clines peuvent être distingués. Comme expliqué plus haut, ces clines sont maintenus par un équilibre entre la migration et deux formes de sélection, selon que les différents allèles sont maintenus entre environnements (exogène) ou backgrounds génétiques (endogène) (Moore & Price 1993). La sélection exogène est définie lorsque la valeur sélective est liée à l'environnement; la sélection endogène est définie par la valeur sélective des interactions intra-génomiques (comme la sous-dominance ou l'épistasie) indépendamment de l'environnement. Cependant, des modèles empiriques basés sur des simulations ont montré que les clines endogènes et

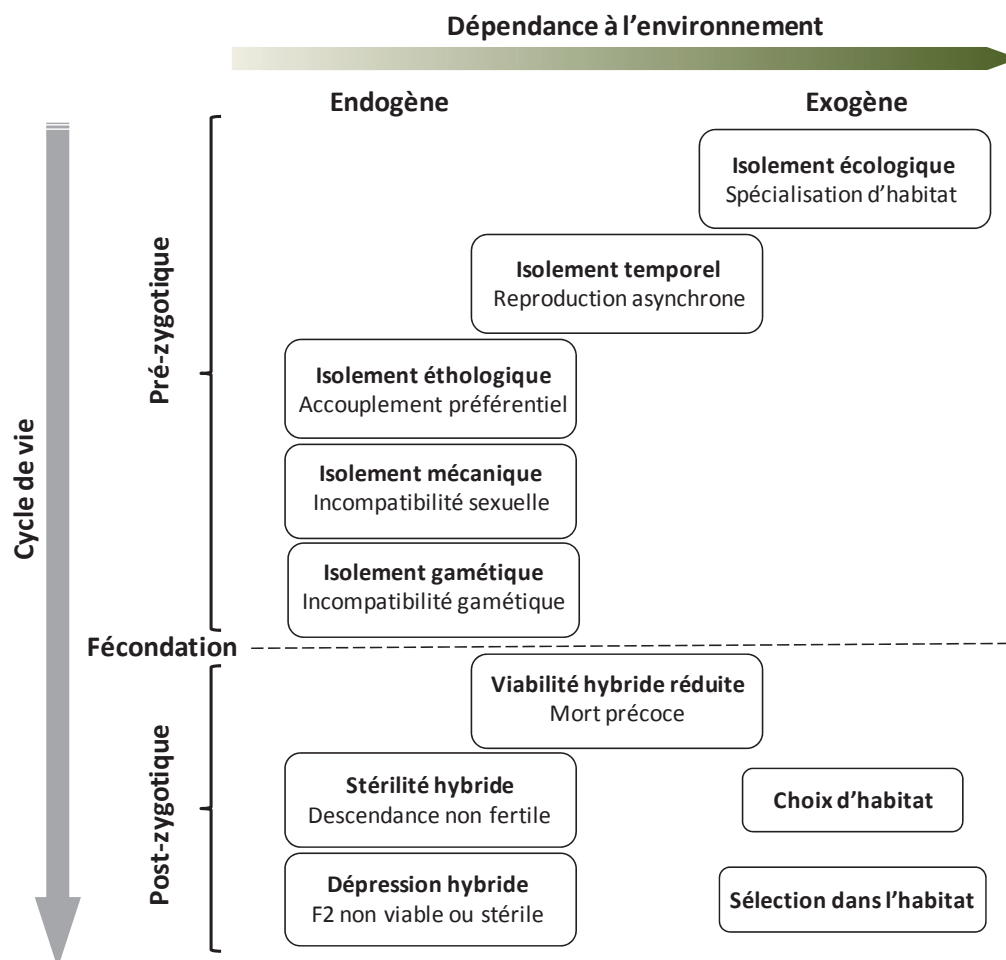
exogènes révèlent des formes semblables (Barton & Gale 1993; Kruuk *et al.* 1999), si bien qu'un cline exogène n'est pas distinguable d'un cline endogène.

Les zones hybrides sont souvent multifactorielles, présentant les deux types de sélection (endogène et exogène, revue dans Bierne *et al.* 2011). La prédominance relative de chaque type a fait l'objet d'un débat dans la littérature des zones hybrides et les études ont conclu que la plupart sont des "zones de tension" (Barton & Hewitt 1981, 1985): le mécanisme principal est la dépression d'hybridation.

Contrairement aux barrières physiques (par. exemple les montagnes ou les rivières pour les organismes terrestres, les fronts océaniques pour les organismes marins...) qui vont réduire le flux de gènes de manière homogène sur le génome, les barrières génétiques affectent différentes régions du génome de façon hétérogène et sont pour cela souvent qualifiées de barrières semi-perméables au flux génique (Harrison 1990). La variance de différenciation entre les locus neutres peut devenir très grande dans les génomes avec des régions complètement isolées et d'autres introgressées plus ou moins librement, car le flux de gène dépend de la densité en gènes d'isolement autour des locus neutres (Barton & Bengtsson 1986). Les échanges aux locus neutres seront réduits s'ils sont liés à des gènes d'isolement (revu dans Barton 1986) et augmentés par effet d'auto-stop génétique si liés à un locus avantageux sous sélection positive (Maynard Smith & Haigh 1974; Faure *et al.* 2008) (Figure 3).

Encadré : L'isolement reproductif

Il est possible de regrouper les mécanismes d'isolement reproductif selon le stade du cycle de vie auquel ils interviennent. Les mécanismes pré-zygotiques, sont ceux qui vont agir avant ou pendant la reproduction. Ils vont favoriser la probabilité de rencontre et de fécondation entre gamètes d'individus de la même espèce. Les mécanismes post-zygotiques agissent après la fécondation. Ce sont des mécanismes qui diminuent la valeur sélective des hybrides ou qui favorisent les génotypes d'une espèce par rapport à l'autre. Les mécanismes d'isolement peuvent aussi être regroupés en fonction de leur dépendance à l'environnement. Ils peuvent être endogènes, c'est à dire gouvernés uniquement par le génotype des individus indépendamment de l'environnement, ou exogènes, c'est à dire résultant d'une interaction entre le génotype des individus et l'environnement dans lequel ils se trouvent. Certains mécanismes sont difficiles à catégoriser de façon claire.



Un exemple de classification des mécanismes d'isolement reproductif est donné dans la figure ci-dessus. Les mécanismes pré- et post-zygotiques, endogènes et exogènes se combinent souvent pour créer l'isolement reproductif entre deux espèces (par exemple : les crapauds sonneurs du genre *Bombina* (Szymura & Barton 1986), les papillons du genre *Heliconius* (Mallet *et al.* 1990), les clams du genre *Mercenaria* (Bert & Arnold 1995), les moules du genre *Mytilus* (Bierne *et al.* 2002b; Bierne *et al.* 2003b).

Ces zones ont été mises en évidence par des analyses moléculaires présentant des degrés très variables d'introgression entre les locus (Harrison 1993) et des études récentes de génomique des populations ont confirmé ce résultat (Wilding *et al.* 2001; Murray & Hare 2006; Lexer *et al.* 2007; Lexer *et al.* 2010; Scascitelli *et al.* 2010; Teeter *et al.* 2010; Colbeck *et al.* 2011; Gagnaire *et al.* 2011; Macholán *et al.* 2011). Ce patron d'introgression a été interprété comme un résultat de la sélection maintenant de solides clines entre populations contiguës ou, plus fréquemment, comme le résultat d'un contact secondaire entre populations précédemment isolées géographiquement (Endler 1977; Barton & Hewitt 1985; Hewitt 1989). Par conséquent, la variation observée des taux d'introgression entre locus (la semi-perméabilité) révèle une caractéristique fondamentale des barrières génétiques, ce qui implique l'isolement reproductif entre les taxons partiellement isolés (Payseur 2010).

3- Génomique des populations « étape 2 » : l'histoire et les bases génétiques de l'adaptation par une analyse détaillée des locus identifiés

La mise en évidence d'un locus outlier n'explique cependant pas les mécanismes sous-jacents à la différenciation observée et les études ayant enquêté de manière plus explicite sur comment la sélection opère sur un locus outlier sont rares dans la littérature. Quelques exemples peuvent être cités.

La pantophysine (*PanI*) a montré des différences très significatives dans les fréquences alléliques chez les populations de morue de l'Atlantique (*Gadus morhua*) lors de son identification par la méthode des RFLP (Restriction Fragment Length Polymorphism). La confirmation de l'action de la sélection a été donnée par l'analyse du polymorphisme de séquence (Pogson 2001) qui montre que la coalescence des allèles dévie très fortement de l'attendu d'un coalescent standard. Deux clades très divergents sont maintenus polymorphes depuis longtemps, probablement plus d'un million d'année, ce qui a conduit Pogson (2001) à la conclusion que la sélection balancée maintenait ce polymorphisme, sans pourtant exclure complètement l'hypothèse que cette divergence longue et sans recombinaison intra-génique pouvait être la conséquence d'une histoire d'isolement géographique suivi d'un contact secondaire. Une analyse moléculaire directe des mutations synonymes et non synonymes a montré que ce gène évoluait sous la pression de la sélection darwinienne positive chez les gadidés (Pogson & Mesa 2004). Enfin, des corrélations entre la fréquence des allèles *PanI* et des variables environnementales, comme la température ou la salinité, ont été identifiées ce qui a permis de suggérer que les conditions environnementales jouent un rôle important dans

la détermination de la distribution des différents génotypes de *PanI* (Case *et al.* 2005). Toutes les études sur *PanI* font l'hypothèse implicite que la sélection agit directement sur ce locus et il n'a jamais été suggéré que *PanI* puisse en fait être un marqueur lié au locus réellement sous sélection.

Colosimo *et al.* (2005) a utilisé des méthodes de clonage afin d'identifier un locus majeur contrôlant le patron de structuration des plaques latérales dans des populations sauvages d'épinoches à trois épines (*Gasterosteus aculeatus*). La reconstitution de l'histoire du gène de l'ectodysplaine (*EDA*) a été réalisée par la cartographie, le séquençage et des études transgéniques montrant que la voie de signalisation joue un rôle clé dans le changement évolutif des populations naturelles et que l'évolution parallèle des phénotypes des plaques latérales des épinoches a eu lieu par la sélection répétée d'allèles *EDA* dérivés d'un haplotype ancestral. Comme pour *PanI*, la divergence entre les clades d'allèles de rivière et de mer du locus *EDA* est très ancienne, datée à plus de 2 millions d'années, ce que Colosimo *et al.* (Colosimo *et al.* 2005) ont interprété comme une preuve que la sélection agissait sur une variation génétique pré-existante.

L'histoire évolutive du locus *K*, qui code pour une protéine β -défensine contribuant à la diversité pigmentaire du pelage du chien domestique (Kerns *et al.* 2007) a été étudiée par scan d'association et cartographie génétique chez les loups d'Amérique du Nord (Anderson *et al.* 2009). Cette étude, basée sur le génotypage de 52 SNPs centrés sur le locus *K*, a analysé la structure haplotypique, inféré les généalogies de ces haplotypes, mesuré les déséquilibres de liaison par paire de sites et estimé le temps de coalescence à l'ancêtre commun le plus récent. Les auteurs ont ainsi pu montrer que la mutation au locus *K* chez les loups provient de l'hybridation avec les chiens domestiques et que l'allèle K^B présente une signature moléculaire de sélection locale dans les habitats forestiers (Anderson *et al.* 2009). Cette introgression adaptative d'un allèle provenant des chiens produit une coalescence anormalement ancienne au locus *K* par rapport au reste du génome.

Alors que les AFLPs se sont révélés être des marqueurs de choix pour les scans génomiques de la différenciation chez les espèces non-modèles, très peu de scans AFLP ont été poursuivis par une étude détaillée des locus outliers. La raison est probablement que l'individualisation des locus AFLP par découpage et séquençage des bandes est difficile et a donné des résultats décevants en dévoilant que le polymorphisme responsable était souvent dû à des éléments transposables (Wood *et al.* 2008; Midamegbe *et al.* 2011; Nunes *et al.* 2011; Paris & Despres 2012). Quelques études ont néanmoins persévéré dans cette stratégie et ont

permis de dévoiler des régions du génome avec de fortes divergences (Wood *et al.* 2008; Nunes *et al.* 2011).

4- Contexte de la thèse

4.1- Originalité des organismes marins

Le cycle de vie benthopélagique d'un animal marin à phase larvaire planctonique est représenté en Figure 4. Pour un généticien des populations, les caractéristiques importantes de ce cycle sont :

- (i) la fécondation externe ; on s'attend donc à l'union au hasard des gamètes, c'est à dire la panmixie, à l'intérieur des populations,
- (ii) la phase larvaire planctonique, qui autorise des taux de migration (m) importants, et les effectifs des populations (N) souvent très grands ; de ce fait on s'attend à des flux géniques (Nm) forts qui assurent une homogénéité génétique sur de longues distances,
- (iii) les fortes fécondités, qui autorisent de forts différentiels de sélection et les effectifs importants qui permettent l'emprise de la sélection même avec des coefficients faibles, deux raisons d'espérer mieux observer les processus de sélection mis en œuvre.

Globalement, ce cycle favorise le brassage génétique intra- et inter-population. Pourtant un nombre croissant d'études a mis à l'épreuve les idées émises a priori au vu du cycle de vie. Par exemple, la taille des populations pourrait être, à petite échelle et transitoirement, plus faible que prévu du fait d'une forte variance du succès reproducteur chez ces organismes très féconds et avec une mortalité très forte dans les stades précoces (Hedgecock et Pudovkin 2011). Ce qui nous intéresse ici c'est qu'il existe de nombreux exemples de structure génétique forte chez les animaux marins attribuée à la sélection (Koehn *et al.* 1984; Hilbish 1985; Ward *et al.* 1992; Huvet *et al.* 2000; Daguin *et al.* 2001; López *et al.* 2001; Launey *et al.* 2002). Comme nous l'avons vu plus haut, le flux génique contrebalance l'effet de la sélection locale et limite l'adaptation dans un environnement hétérogène (Lenormand 2002), surtout si il est à grain fin. Or la plupart des variations environnementales auxquelles font face les animaux marins sont à une échelle inférieure à la dispersion larvaire et donc sont à grain fin. L'adaptation locale devrait donc être plus rare, ou au moins plus coûteuse, en milieu marin. Cette thèse adresse le problème du paradoxe de l'adaptation marine en essayant de mieux comprendre l'origine de ces structures fortes attribuées à la sélection locale.

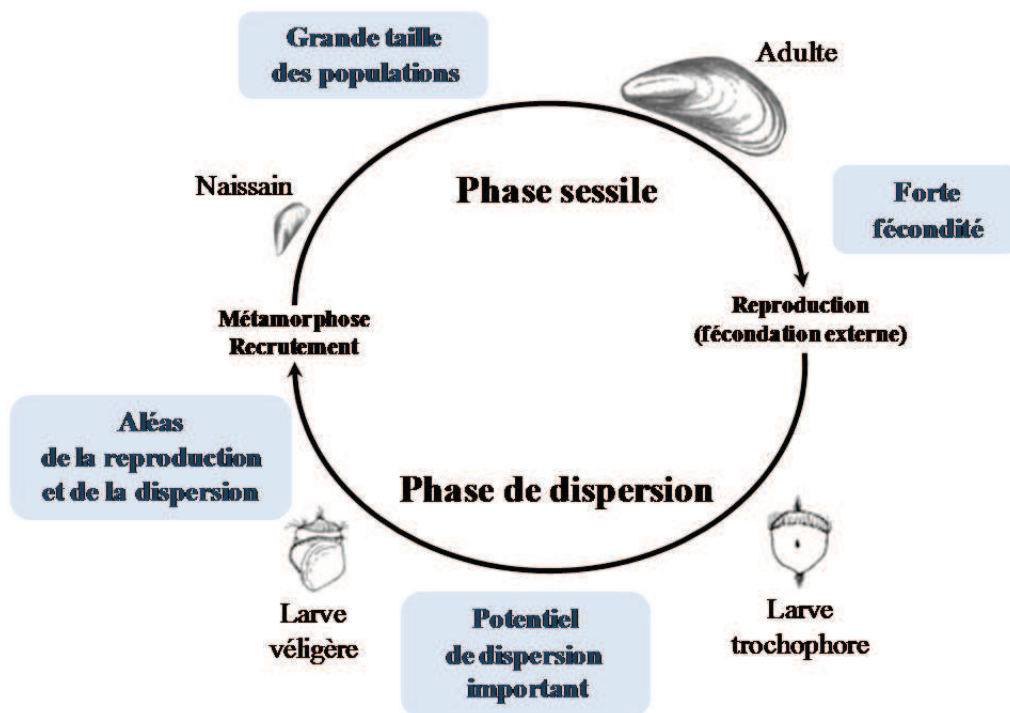


Figure 4 : Cycle de vie schématique des moules du genre *Mytilus*.

4.2- Un laboratoire naturel pour l'étude de la différenciation, le complexe *Mytilus edulis*

Trois espèces de *Mytilus* formant le complexe d'espèces *M. edulis* sont habituellement distinguées (McDonald & Koehn 1988; Varvio *et al.* 1988; Koehn 1991; McDonald *et al.* 1991; Seed 1992) : *M. edulis* Linné (1758), *M. galloprovincialis* Lamarck (1819) et *M. trossulus* Gould (1850). Ce sont des organismes cosmopolites qui offrent la possibilité d'étudier la variation génétique des populations à plusieurs échelles spatiales. Les moules sont connues pour occuper des habitats très variés, depuis les milieux estuariens très abrités jusqu'aux sites côtiers très exposés. De plus, les moules présentent, comme beaucoup d'organismes marins, une phase larvaire planctonique permettant une dispersion à longue distance.

En Europe, ces trois espèces sont présentes et forment deux à deux des zones hybrides. *M. edulis* et *M. trossulus* sont séparées par une zone hybride clinale simple bien documentée sur les côtes de l'Øresund entre la Mer du Nord et la Mer Baltique (Väinölä & Hvilson 1991). Le long des côtes françaises, *M. edulis* et *M. galloprovincialis* forment une zone hybride à la structure inédite (Bierne *et al.* 2002b; Bierne *et al.* 2003c) : il s'agit d'une zone hybride mosaïque formée par une succession de trois zones de contact indépendantes (dans les Landes, en sud Bretagne et en Normandie) qui isolent des patchs introgressés et différenciés de *M. galloprovincialis* en Bretagne et de *M. edulis* en Charente (Figure 5). De

nombreuses études permettent de considérer ces patchs comme panmictiques (Skibinski *et al.* 1983; Coustau *et al.* 1991; Quesada *et al.* 1995c; Daguin *et al.* 2001; Bierne *et al.* 2003c; Kijewski *et al.* 2011; Hilbish *et al.* 2012). Les niveaux de différenciations entre patchs sont très hétérogènes, ils varient non seulement entre locus pour une même zone, mais aussi entre zones pour un même locus. L'hybridation des moules est étudiée au laboratoire depuis une dizaine d'année et les mécanismes d'isolement reproductifs sont connus (Bierne *et al.* 2002a; Bierne *et al.* 2003b; Bierne *et al.* 2006) et multiples. La thèse de Matthieu Faure (2008) a permis de révéler une longue histoire de divergence sans flux génique entre les deux espèces, de l'ordre de 1 à 2 millions d'années, et un contact secondaire assez récent, probablement post-glaciaire. L'hétérogénéité de la différenciation peut donc être interprétée comme le résultat d'une introgression secondaire différentielle entre région du génome (Faure *et al.* 2008; Boon *et al.* 2009).

Mises à part les différenciations interspécifiques et entre patchs enclavés et périphériques de la zone mosaïque, il existe aussi des différenciations génétiques assez marquées au sein même des espèces. Chez *M. galloprovincialis* une différenciation forte a été observée de part et d'autre du front Almeria-Oran, entre les populations méditerranéennes d'une part et les populations atlantiques et de la Mer d'Alboran d'autre part (Quesada *et al.* 1995a; Quesada *et al.* 1995c; Daguin & Borsa 1999).

Chez *M. edulis*, il existe une différenciation transatlantique entre l'Amérique du nord et l'Europe (Varvio *et al.* 1988; Riginos *et al.* 2004). Enfin des différenciations génétiques ont été observées entre populations *M. edulis* de la côte est des Etats Unis (Koehn *et al.* 1976). De part et d'autre du Cap Cod dans le Massachusetts la différenciation a été révélée par plusieurs locus allozymiques mais entre les populations océaniques et du détroit de Long Island la différenciation n'a été révélée à ce jour que par un seul locus, l'aminopeptidase 1. A un niveau intra-spécifique, la différenciation semble être assez hétérogène entre locus même si le nombre de marqueurs analysés était assez faible. Ceci a été pris comme un argument en faveur de la sélection sur certains locus allozymiques (Koehn *et al.* 1980; Mitton 1997). Ainsi le cline de fréquence allélique de l'aminopeptidase 1 dans le détroit de Long Island (Koehn *et al.* 1976; Koehn *et al.* 1980; Koehn *et al.* 1984; Hilbish & Koehn 1985) reste un des exemples les plus célèbres de sélection en environnement hétérogène repris dans les manuels d'évolution (Gillespie 1991; Mitton 1997; David & Samadi 2000).



Figure 5 : Zone d'hybridation mosaïque entre *Mytilus edulis* et *M. galloprovincialis* le long des côtes européennes. Les zones de transition sont représentées par des pointillés (ZH_1 , ZH_2 et ZH_3 pour les zones d'hybridation et ZH_{AOF} pour la zone de transition au niveau du front Alméria/Oran). L'aire de répartition de *M. edulis* est matérialisée par une double ligne ; celle de *M. galloprovincialis* est matérialisée par un trait plein.

Dépassant le contexte de l'hybridation, le complexe d'espèces *M. edulis* représente donc une situation intéressante pour mieux comprendre comment se propage et se distribue la diversité adaptative en population structurée. Ce modèle biologique procure un outil intéressant pour mener des scans de la différenciation entre des populations isolées sur des échelles de temps et d'espaces contrastés.

4.3- Objectifs de la thèse

L'objectif de la thèse est d'étudier les bases génétiques et l'histoire de la différenciation adaptative chez une espèce marine à phase planctonique, les moules du complexe d'espèce *M. edulis*. Pour ce faire une approche de génomique des populations a été choisie en combinant la technique AFLP qui permet d'obtenir facilement la différenciation génétique sur un grand nombre de locus et d'autres types de polymorphismes sur moins de locus mais pour lesquels la poursuite de l'analyse est beaucoup plus facile lorsqu'ils s'avèrent être outliers.

L'adaptation locale serait-elle plus fréquente qu'attendu? La différenciation génétique a-t-elle une origine plus complexe? L'ambition de cette thèse était (i) d'identifier des signatures de la sélection sur les patrons de différenciation des génomes, (ii) d'analyser finement le type de sélection à l'œuvre (directe vs. indirecte, positive vs. disruptive, exogène vs. endogène, en cours vs. passée), et (iii) de quantifier l'importance de ces événements de sélection afin de comprendre les bases génétiques et de reconstituer l'histoire de la différenciation adaptative entre populations de moules du complexe *M. edulis*.

CHAPITRE I : L'introgression différentielle explique les F_{ST} outliers entre les populations d'Atlantique et de Méditerranée chez *Mytilus galloprovincialis*

I.1- Résumé

Les scans génomiques de la différenciation sont devenus un standard de la génétique des populations. Un locus présentant un niveau de différenciation anormalement élevé est traditionnellement interprété comme étant sous sélection locale. Cependant, une structure génétique forte à un locus peut être la conséquence de scénarios plus ou moins complexes combinant l'histoire des populations avec diverses formes de sélection. Ici, nous avons essayé d'identifier la base génétique et l'histoire de la différenciation de deux locus anormalement structurés chez la moule *Mytilus galloprovincialis*. Un scan génomique a été réalisé entre les populations atlantiques et méditerranéennes en utilisant différents types de marqueurs moléculaires (allozymes, microsatellites, polymorphismes d'intron, AFLP). La différenciation globale était faible (F_{ST} moyen ~ 0.03), mais des locus anormalement différenciés ont été identifiés. L'analyse du polymorphisme de séquence d'ADN de deux de ces deux locus nous a permis de montrer que la structure génétique était la conséquence d'un différentiel d'introgression d'allèles provenant de l'espèce sœur *M. edulis*. De façon inattendue c'est la population de Méditerranée qui s'est révélée être la plus introgressée alors que la zone de contact entre les deux espèces se trouve actuellement dans l'Atlantique, le long des côtes françaises et britanniques. L'introgression est historique et implique un contact entre des *M. edulis* et des *M. galloprovincialis* de Méditerranée. Ce contact aurait eu lieu pendant la dernière période glaciaire alors que les populations étaient distribuées plus au sud. Deux hypothèses non mutuellement exclusives permettent d'expliquer le différentiel d'introgression : (i) l'introgression est adaptative et implique que les allèles *edulis* soient favorisés dans les populations de Méditerranée, (ii) l'architecture génétique de la barrière au flux génique est différente entre les deux fonds génétiques *galloprovincialis* et implique que le fond génétique méditerranéen soit plus perméable à l'introgression que le fond génétique atlantique aux régions du génome considérées. Notre étude met en évidence l'importance de l'histoire et de l'architecture génomique de l'isolement reproductif dans la mise en place d'une structure génétique anormalement forte sur certain locus.

I.2- Differential introgression from a sister species explains high F_{ST} outlier loci within a mussel species

Célia C. Gosset^{1,2} and Nicolas Bierne^{1,2}

¹ Université Montpellier 2, Place Eugène Bataillon, 34095 Montpellier, France

² CNRS - Institut des Sciences de l'Evolution UMR5554 Montpellier, France.

Article publié dans la revue *Journal of Evolutionary Biology* (volume 26, issue 1, pages 18-26)

Abstract

Scanning genomes for loci with high levels of population differentiation has become a standard of population genetics. F_{ST} outlier loci are most often interpreted as signatures of local selection, but outliers might arise for many other reasons too often left unexplored. Here, we tried to identify further the history and genetic basis underlying strong differentiation at F_{ST} outlier loci in a marine mussel. A genome scan of genetic differentiation has been conducted between Atlantic and Mediterranean populations of *Mytilus galloprovincialis*. The differentiation was low overall ($F_{ST} = 0.03$), but seven loci (2%) were strong F_{ST} outliers. We then analyzed DNA sequence polymorphism at two outlier loci. The genetic structure proved to be the consequence of differential introgression of alleles from the sister hybridizing species *M. edulis*. Surprisingly, the Mediterranean population was the most introgressed at these two loci, while the contact zone between the two species is nowadays localised along the Atlantic coasts of France and the British Isles. A historical contact between *M. edulis* and Mediterranean *M. galloprovincialis* should have happened during glacial periods. It proved difficult to disentangle two hypotheses: (i) introgression was adaptive, implying *edulis* alleles have been favoured in Mediterranean populations, or (ii) the genetic architecture of the barrier to *edulis* gene flow is different between the two *M. galloprovincialis* backgrounds. Five of the seven outliers between *M. galloprovincialis* populations were also outliers between *M. edulis* and Atlantic *M. galloprovincialis*, which would sustain the latter hypothesis. Differential introgression across semi-permeable barriers to gene flow is a neglected scenario to interpret outlying loci that may prove more widespread than anticipated.

Introduction

Understanding evolutionary processes that maintain genetic diversity in natural populations is an essential goal of population genetics (Lewontin 1974; Rockman 2012). Genome scans now allow identifying loci, or genome regions, with greatly increased differentiation between populations, the so-called “ F_{ST} outlier loci” or “genomic islands of differentiation” (Luikart *et al.* 2003; Storz 2005; Nosil *et al.* 2009). F_{ST} outliers are usually interpreted as signatures of spatially heterogeneous selection (Beaumont 2005), but various alternative hypotheses have also been proposed (Bierne *et al.* 2011). Identifying the genetic basis (Storz & Wheat 2010; Barrett & Hoekstra 2012) and reconstructing the history (Pogson 2001; Colosimo *et al.* 2005; Faure *et al.* 2008) of candidate loci identified in F_{ST} scans is therefore required before to conclude on the evolutionary processes underlying genetic differentiation. Strongly differentiated loci between populations often show very deep coalescences (Schulte *et al.* 1997; Pogson 2001; Colosimo *et al.* 2005; Wood *et al.* 2008) that imply polymorphisms have been maintained for long times. During such long periods populations have likely changed in size and spatial distribution as a consequence of oscillatory climate changes (Hewitt 1996; 2011) and might have evolved partial reproductive isolation mechanisms. Although local adaptation is likely to contribute to the differentiation, it is also likely that multifarious processes jointly act in generating semi-permeable genetic barriers to gene flow (Bierne *et al.* 2011).

Marine species with a pelagic larval phase are associated with low genetic differentiation over long distances (Palumbi 1994; Hellberg 2009) but genetic differentiation is often observed at some specific loci affected by selection (Koehn *et al.* 1976; Lemaire *et al.* 2000; Schmidt & Rand 2001; Riginos *et al.* 2002; Johannesson & Andre 2006; Murray & Hare 2006; Faure *et al.* 2008). Surprisingly, these loci are not that difficult to uncover. As soon as a sufficient number of loci are analysed a few loci often reveal genetic structure at key geographic locations that are increasingly being well characterized (Pelc *et al.* 2009; Bierne *et al.* 2011; Riginos *et al.* 2011), such as for instance the Oresund between the Baltic Sea and the North Sea (Johannesson & Andre 2006). In recent years, increased availability of methods for the simultaneous analysis of large numbers of loci has facilitated a population genomics approach to this problem (Guinand *et al.* 2004; Nielsen *et al.* 2009). However, the focus on local adaptation as the exclusive explanation of F_{ST} outlier loci might be excessive (Bierne 2010; Bierne *et al.* 2011; Roesti *et al.* 2012), especially because marine species have demographic characteristics conducive to generating discrepancies with simple theoretical

models (Eldon & Wakeley 2009; Bierne 2010; Hedgecock & Pudovkin 2011; Tice & Carlon 2011).

The Almeria-Oran front (AOF), an oceanographic front that separates Atlantic from Mediterranean water masses, is a recognised hotspot of genetic structure in the Sea (Borsa *et al.* 1997; Patarnello *et al.* 2007). Although it certainly acts as a natural barrier to larval dispersal, examples exist of species with a planktonic dispersal phase that, from the panel of markers analysed, do not exhibit any genetic break at the AOF (Launey *et al.* 2002; Patarnello *et al.* 2007). This suggests that the oceanographic front itself is not sufficient alone to explain the genetic structure observed in many marine species. When differentiation is observed it is often highly heterogeneous across loci (Quesada *et al.* 1995c; Lemaire *et al.* 2005; Kasapidis *et al.* 2012) suggesting the barrier to gene flow is not only natural but also genetic (Bierne *et al.* 2011). Indeed, genetic barriers to gene flow are often semi-permeable (Harrison 1993) and are expected to coincide with natural barriers to gene flow (Hewitt 1975; Barton 1979). The mussel *Mytilus galloprovincialis* is one emblematic species that exhibit such a genetic break at the AOF. The genetic differentiation across the AOF is especially strong at the allozyme locus *ODH* (Sanjuan *et al.* 1994; Quesada *et al.* 1995c) and the mitochondrial genome (Quesada *et al.* 1995a; Sanjuan *et al.* 1996) while it is modest, although detectable, at other loci (Sanjuan *et al.* 1994; Quesada *et al.* 1995c; Diz & Presa 2008).

We investigated the genomic architecture of differentiation across the AOF, and the history and genetic basis underlying the differentiation of high F_{ST} outlier loci in the mussel *M. galloprovincialis*. We first conducted an F_{ST} scan with a combination of AFLP and nuclear codominant markers (allozymes, microsatellites and length polymorphisms). Secondly, we analysed DNA sequence polymorphisms at two outlier loci. Although we focus on the differentiation across the AOF, we also analysed samples of the sister hybridizing species *M. edulis* with the same loci. We show that the genetic differentiation at five of the seven identified outliers is the consequence of a stronger introgression of heterospecific alleles into the Mediterranean than the Atlantic background of *M. galloprovincialis*.

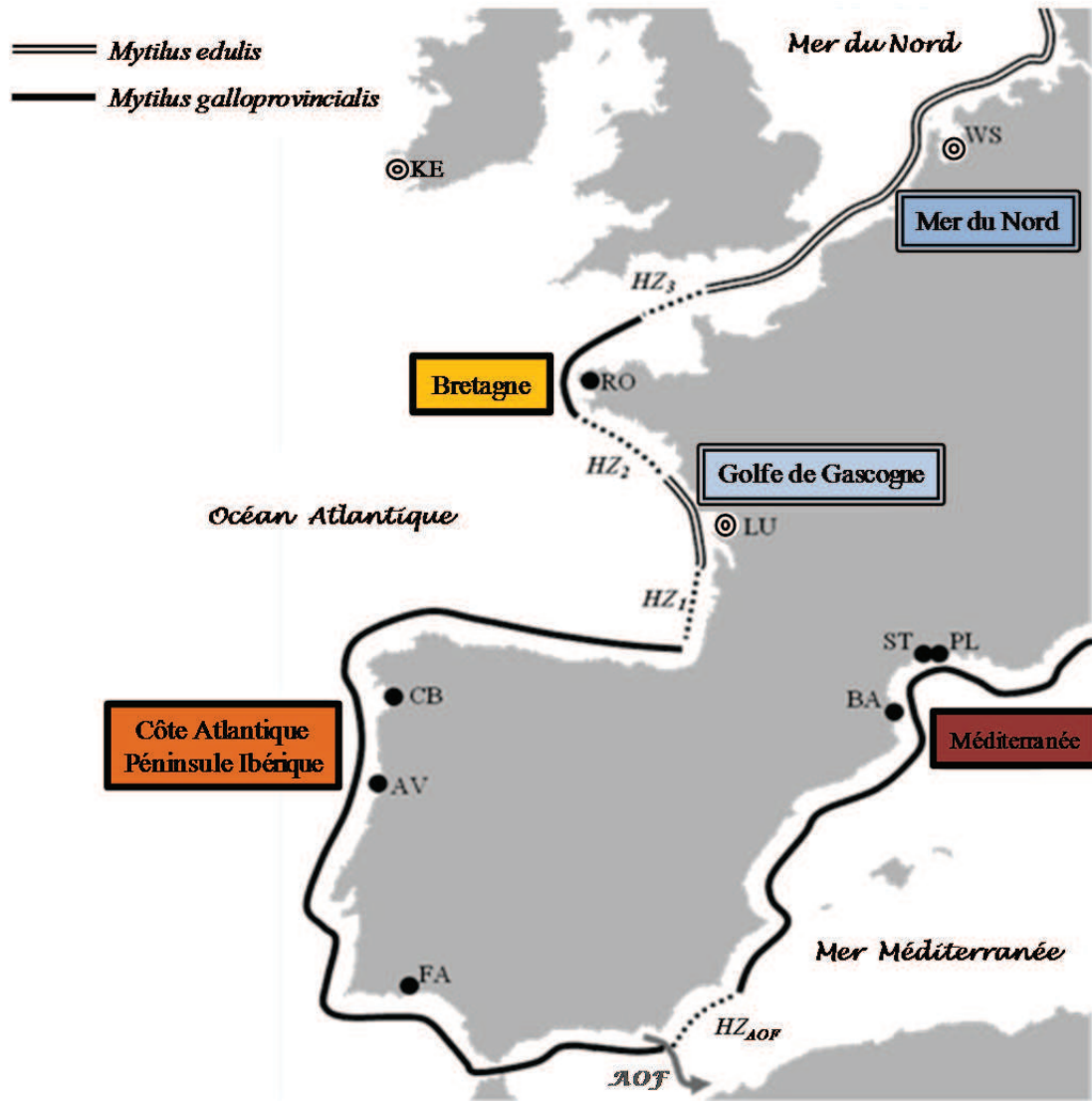


Figure I.1 : Sampling locations of *Mytilus* spp. mussels. Patches of *M. galloprovincialis* are indicated with black lines and patches of *M. edulis* are indicated by black and white lines. White and black dots represent *M. edulis* and *M. galloprovincialis* samples, respectively. The three hybrid zones described by Bierne *et al.* (2003c) and the Almeria-Oran genetic break described in Quesada *et al.* (1995c) are indicated by dashed lines: HZ_1 , between *M. galloprovincialis* of the Atlantic coast of the Iberian Peninsula and *M. edulis* of the Bay of Biscay; HZ_2 , between *M. edulis* of the Bay of Biscay and *M. galloprovincialis* of Brittany; HZ_3 , between *M. galloprovincialis* of Brittany and *M. edulis* of the North Sea; and HZ_{AOF} between *M. galloprovincialis* of the Atlantic coast of the Iberian Peninsula and *M. galloprovincialis* of the Mediterranean Sea. WS: Wadden Sea, Holland, RO: Roscoff, Brittany, France, LU: Lupin, Poitou-Charente, France, CB: Combarro, Iberian Peninsula, AV: Aveiro, Iberian Peninsula, FA: Faro, Algarve, Portugal, BA: Banyuls-sur-mer, Languedoc-Roussillon, France, ST: Sète, Languedoc-Roussillon, France, PA: Palavas, Languedoc-Roussillon, France.

Materials and methods

Sampling sites

We used five geographic samples of 48 individuals that on the basis of previous publications are representative of each of five panmictic patches (Skibinski *et al.* 1983; Coustau *et al.* 1991; Quesada *et al.* 1995c; Daguin *et al.* 2001; Bierne *et al.* 2003c; Kijewski *et al.* 2011; Hilbish *et al.* 2012): (i) samples from RO (Roscoff, Brittany, France), FA (Faro, Algarve, Portugal) and ST (Sète, Languedoc-Roussillon, France) are representative of *M. galloprovincialis* populations of Brittany, Iberian Peninsula and Mediterranean Sea respectively and (ii) samples from WS (Wadden Sea, Holland) and LU (Lupin, Poitou-Charente, France) are representative of *M. edulis* populations of the North Sea and Bay of Biscay respectively (Figure I.1). WS, RO, LU and FA samples were described in Faure *et al.* (2008) and were collected in the same sites as samples Helgoland, Primel, Faro and Brouage described in Bierne *et al.* (2003c).

DNA extraction and AFLP scoring

Total DNA was extracted from gills using the DNeasy Tissue Kit (Qiagen) following the instructions of the manufacturer and extraction quality was checked on agarose gel. DNA concentration was measured for each sample using a NanoDrop8000 Spectrophotometer (Thermo Scientific), and standardized to a DNA concentration of 100ng.μl⁻¹. All AFLP reactions were performed in 96 well plates, containing 4 different sampling sites and at least 30% of replicates from other plates to measure the repeatability of AFLP fingerprints and to perform an error rate analysis. AFLP procedure followed the original protocol by Vos *et al.* (1995), which employs the combination of enzymes *EcoRI* and *MseI*. Standardized genomic DNA was first digested with 2 U of *EcoRI* and 6 U of *MseI* at 37°C for 2h30. Double-stranded *EcoRI* and *MseI* adapters (Table I.1) were ligated to restriction fragments for 2h at 22°C. The preselective PCR amplification parameters were as follows: 5 min at 94°C, 20 cycles of 30 s denaturing at 94°C, 1 min annealing at 56°C and 1 min extension at 72°C, ending with 5 min at 72°C for complete elongation, using two PCR primer pairs: *Eco*+A, *Eco*+C and *Mse*+C (see Table I.1). The selective amplifications were performed with four chosen selective primer pairs *Eco*+ACG, *Eco*+CAG, *Mse*+CGA and *Mse*+CTC. Touchdown cycles were performed to increase PCR specificity by initiating annealing 30 s at 65°C, then reduced by 0.7°C for the next 12 cycles and maintained at 56°C for the remaining 20 cycles.

Selective-PCR products were separated and detected using an Applied ABI Prism 3130XL (Applied Biosystems) and sized by GeneScan™ 500 ROX™ Size Standard (Applied Biosystems).

Table I.1 : Sequences (5'–3') of oligonucleotide adaptors and primers used for AFLP genotyping.

Name	Adaptor/Primer	Sequence
<i>EcoRI</i>		
<i>Eco_adapt1</i>	Adaptor	5'-CTCGTAGACTGCGTACC-3'
<i>Eco_adapt2</i>	Adaptor	5'-AATTGGTACGCAGTCTAC-3'
<i>Eco_A</i>	Preselective primer	5'-GACTGCGTACCAATTCA-3'
<i>Eco_C</i>	Preselective primer	5'-GACTGCGTACCAATTCC-3'
<i>Eco_ACG</i>	Selective primer	5'-GACTGCGTACCAATTCACG-3'
<i>Eco_CGA</i>	Selective primer	5'-GACTGCGTACCAATTCCGA-3'
<i>MseI</i>		
<i>Mse_adapt1</i>	Adaptor	5'-GACGATGAGTCCTGAG-3'
<i>Mse_adapt2</i>	Adaptor	5'-TACTCAGGACTCAT-3'
<i>Mse_C</i>	Preselective primer	5'-GATGAGTCCTGAGTAAC-3'
<i>Mse_CGA</i>	Selective primer	5'-GATGAGTCCTGAGTAACGA-3'
<i>Mse_CTC</i>	Selective primer	5'-GATGAGTCCTGAGTAACCTC-3'

Abbreviation: AFLP, amplified fragment length polymorphism.

GeneMapper® v4.5 software (Applied Biosystems) was used to read the resulting chromatograms. Fragment length classes (bin sets) were created automatically and manually revised for each primer combination in the interval 50–500 base pairs. Only unambiguously identifiable fragments were translated into a binary matrix of peak intensity values for each detected locus at each sample. This matrix was secondary translated into a phenotype matrix of presence (1) absence (0) of the fragments with AFLPScore v1.4b (Whitlock *et al.* 2008). Nomenclature for AFLP scoring thresholds follows Whitlock *et al.* (2008), where 'Locus-calling' and 'Phenotype-calling' are thresholds used to exclude error-prone loci from the AFLP genotypes table, and to determine fragment phenotype (band presence or absence) based on raw chromatogram peak-height data, respectively. Error rates were measured using the method of Hadfield *et al.* (2006), modified for dominant data as implemented in the R package MasterBayes, which estimates the rate of error applying to each of the two alleles that underlie the genotypes scored using AFLP via a Gibbs sampler (Whitlock *et al.* 2008).

Under this method, error rates ϵ_1 and ϵ_2 determine the probability of misscoring an AFLP fragment presence '1' or absence '0' allele, respectively. We also determined a widely

used measure of error (the mismatch error rate) that compared the dissimilarity of pairs of AFLP phenotype profiles that had originated from the same genetic individual (replicated samples).

We estimated allelic frequencies at each AFLP marker with AFLPsurv *v1.0* (Vekemans 2002; Vekemans *et al.* 2002) assuming Hardy-Weinberg equilibrium in each population and using a bayesian method with non-uniform prior distribution of allele frequencies (Zhivotovsky 1999).

Nuclear codominant markers

We compiled data on 14 allozyme loci from Quesada *et al.* (1995c), 6 microsatellite loci from Diz & Presa (2008), 8 length polymorphisms from Daguin *et al.* (2001), Bierne *et al.* (2002b) and Boon *et al.* (2009) to which we have added new results for the Mediterranean sample (see details in Table I.2). Sampling localities for allozymes and microsatellites were slightly different to those used for AFLPs and other codominant nuclear markers (see Table 2 and Figure I.1) but within the attested panmictic patches. We also analyzed three new intron-length polymorphisms. Primers were designed to amplify portions of the endo-beta-1,4-glucanase gene (Xu *et al.* 2001), the endo-beta-1,4-mannanase gene (Xu *et al.* 2002) and the elongation factor 2 gene (from an EST library, Tanguy *et al.* 2008). An analysis of a sample of DNA sequences allowed us to identified indel polymorphisms in each of the three genes, and internal primers were redesigned in conserved regions each side of an indel polymorphism (see primer list in Table 3).

Table I.2 : List of codominants markers used in this study.

Locus name	Marker type	Sampling localities	Reference
<i>Aat-1</i>	Allo	a	(1)
<i>Aat-2</i>	Allo	a	(1)
<i>AP</i>	Allo	a	(1)
<i>Dia</i>	Allo	a	(1)
<i>Est-D</i>	Allo	a	(1)
<i>IDH</i>	Allo	a	(1)
<i>Lap-1</i>	Allo	a	(1)
<i>Lap-2</i>	Allo	a	(1)
<i>MPI</i>	Allo	a	(1)
<i>ODH</i>	Allo	a	(1)
<i>6Pgdh</i>	Allo	a	(1)
<i>Pgi</i>	Allo	a	(1)
<i>Pgm</i>	Allo	a	(1)
<i>Stdh</i>	Allo	a	(1)
<i>Mgμ1</i>	μsat	b	(2)
<i>Mgμ2</i>	μsat	b	(2)
<i>Mgμ3</i>	μsat	b	(2)
<i>Mgμ5</i>	μsat	b	(2)
<i>Mgμ6a</i>	μsat	b	(2)
<i>Mgμ6b</i>	μsat	b	(2)
<i>mac-1</i>	L-P	c	(3)
<i>Glu-5'</i>	L-P	c	(4)*
<i>EFbis</i>	L-P	c	(4)*
<i>DAMP1</i>	L-P	c	(4)*
<i>DAMP2</i>	L-P	c	(4)*
<i>DAMP3</i>	L-P	c	(4)*
<i>Gluca</i>	L-P	c	(5)
<i>Manna</i>	L-P	c	(5)
<i>EF2</i>	L-P	c	(5)
<i>MGD2</i>	L-P	d	(6)
<i>MytilinB</i>	L-P	d	(6)

Marker type: All: allozyme; μsat: microsatellite; L-P: length-polymorphism

Sampling locality: a: Aveiro (Iberian Peninsula) & Palavas (Mediterranean Sea); b: Combarro (Iberian Peninsula) & Palavas (Mediterranean Sea); c: Faro (Iberian Peninsula) & Sète (Mediterranean Sea); d: Faro (Iberian Peninsula) & Banyuls-sur-mer (Mediterranean Sea). Reference: (1): Quesada *et al.* (1995 c); (2): Diz & Presa (2008); (3): Daguin *et al.* (2001); (4): Bierne *et al.* (2002); (5): this study; (6): Boon *et al.* (2009); * this study for the Mediterranean sample.

Table I.3: Sequences (5'–3') of oligonucleotide primers used for the elongation factor 2 gene, the endo-beta-1,4-glucanase gene and the endo-beta-1,4-mannanase gene.

Marker	Forward primer	Reverse primer
<i>EF2</i>	5'-GGAAATCCCATGGGTGATTTAGCGG-3'	5'-GTCAAATAAATACTGAAACACAGTGACTTC-3'
<i>Gluca</i>	5'-ACAGTGTTTCGGCCAATCAGAAATG-3'	5'-TTATAAAGATTCACTTACTTGTTTCG-3'
<i>Manna</i>	5'-TAGGAGTGTGGCTGCATATTGAAGGGG-3'	5'-AATCAGGTATTTACGCTTTGTTTCCC-3'

Outlier loci detection

Following Perez-Figueroa *et al.* (2010), we performed three different methods to identify loci that depart from the expected neutral distribution of F_{ST} . We used the method of Beaumont & Nichols (1996) to obtain the average and 95% confidence interval of single locus F_{ST} as a function of heterozygosity. Simulations were performed with the FDIST2 program (Beaumont & Nichols 1996) for codominants markers and DFDIST program, an extension of FDIST software that allows the use of dominant markers, for the AFLP dataset. Null allele frequencies were first estimated at each locus of the empirical AFLP data set using Zhivotovsky's (1999) Bayesian approach, enabling F_{ST} values to be estimated for each locus (Weir & Cockerham 1984). For both datasets, the outlier threshold was defined by an envelope delimited by the 0.05 and 0.95 quantiles of simulated F_{ST} . A mean 'neutral' F_{ST} value supposedly uninfluenced by selected loci was then calculated after an iterative procedure to remove loci outlying the generated 0.95 quantile.

The second method used was developed to identify markers that show deviation from neutral expectation in pairwise comparisons of diverging populations. The DETSEL program (Vitalis *et al.* 2001; Vitalis 2003; Vitalis 2012) relies on a model in which a common ancestor population split into two daughter populations that diverge independently by random drift after a possible bottleneck event (Vitalis *et al.* 2001; Vitalis 2003). This method provides a parameter of divergence for each population, whose joint distribution under neutral expectations is generated using coalescent simulations. Loci detected outside the resulting confidence envelope of 95% were identified as potentially under selection.

The main difference between FDIST2 and DETSEL lies in the underlying demographic model. FDIST2/DFDIST considers an island model of population structure, where a set of populations with constant and equal deme sizes are connected by gene flow (Wright 1931), while DETSEL considers a pure divergence model, in which an ancestral population splits into two daughter populations.

The third method is a bayesian approach implemented in BAYESCAN v2.01 (Foll & Gaggiotti 2008; Foll *et al.* 2010) to estimate directly the posterior probability that each locus is subject to selection. This method is an extension of the hierarchical-Bayesian approach of Beaumont & Balding (2004), and is based on a logistic regression model. The posterior probability of one locus under selection is estimated by defining two alternative models. Selection is introduced by decomposing locus-population F_{ST} coefficients into a population-specific component (β), shared by all loci and a locus-specific component (α) shared by all the populations using a logistic regression. Departure from neutrality at a given locus is assumed when the locus-specific component is necessary to explain the observed pattern of diversity (α significantly different from 0). The respective posterior probabilities of these two models are estimated using a reversible jump Markov chain Monte Carlo (RJMCMC) approach that takes all loci into account in the analyses through the prior distribution, resolving the problem of multiple testing of a large number of genomic locations.

Recent simulation studies have advocated using two or more outlier detection methods to avoid false conclusions (Luikart *et al.* 2003; Vasemagi & Primmer 2005; Pérez-Figueroa *et al.* 2010). Additionally, using a multiple test correction (Storz & Nachman 2003; Galindo *et al.* 2009) is encouraged but have been criticized for being too conservative and reducing the power of outlier detection methods (Murray & Hare 2006; Galindo *et al.* 2009). Ultimately, the measures taken to control the False Discovery Rate (FDR) depend on the goal of the study. If the purpose of a genome scan is to target strongly differentiated loci for further research, the most appropriate approach is probably the use of more than one method and to combine the outlier lists to avoid the risk of losing interesting candidates.

DNA polymorphism and sequence analysis at outlier loci

DNA sequences were obtained for two outlier loci, *EFbis* and *EF2*. For this analysis we used different primer pairs than those used for length polymorphism in order to amplify longer fragments. The new locus named *EFseq* was amplified with the same reverse primer as *EFbis*, *EFbis*-R (5'-CTCAATCATGTTGTCTCCATGCC-3'), and a newly designed forward primer, *EFseq*-F 5'-AGGCTCCTTCAAGTACGCCTGGG-3'). *EF2seq* was amplified the newly designed primer pair *EF2seq*-R (5'-TTATTGTTGTAGAAATTGTCACCCC-3') – *EF2seq*-F (5'-TAATCAACTTGATTGATTACCTGG-3'). Loci were amplified and cloned following the Mark-Recapture (MR) method (Bierne *et al.* 2007; Faure *et al.* 2007; Faure *et al.* 2008; Boon *et al.* 2009) the ancestor protocol of the tagging technique now widely used

for multiplexing libraries in high-throughput parallel sequencing (Binladen *et al.* 2007; Galan *et al.* 2010). Every individual was PCR-amplified separately with different sets of 5'-tagged primers. PCR products were mixed together (equimolar) and cloned into a pGEM-T vector by using the pGEM-T cloning kit (Promega, Madison, WI, USA). We used a capture effort of two in MR-cloning: 96 clones per population and per gene were sequenced at the Genoscope (Evry) (<http://www.genoscope.cns.fr/>) with the universal primers sp6 and T7 flanking the insert on the plasmid. Sequence alignment was performed with ClustalW (Thompson *et al.* 1994b) in the BioEdit interface (Hall 1999) and verified by eye. Alignment gaps were excluded from the analyses. To give a representation of gene genealogy, phylogenetic reconstructions were obtained with Mega 5.1 (Kumar *et al.* 2001) using the neighbour-joining (NJ) algorithm with number of nucleotide differences.

Results

AFLP screening

422 AFLP loci were scored in total, of which 357 were polymorphic in FA and ST populations. The mismatch error rate, weighted to take into account the different numbers of loci arising from different primer combinations, was 3.6% (Table I.4). A summary of the scoring parameters used, the error rates and the final number of loci retained for each selective primer combination is given in Table I.4.

Table I.4 : Summary of AFLP selective primer combinations used during genotyping. The scoring conditions applied in order to derive AFLP phenotypes and the genotyping error rate determined via AFLP fingerprinting of replicated samples are given for each primer pair. The initial and final numbers of loci refer to the initial number of fragment length classes (bins) considered, and the final number of loci retained for analysis after scoring via AFLPScore (Whitlock *et al.* 2008). The replicates rate refers to the number of replicate pairs of fingerprints from which genotyping error was estimated. Total number of loci=422 of which 357 are polymorphic between the two populations studied here; weighted mean percentage error rates, $\varepsilon_1=12.26$, $\varepsilon_2=0.05$, mismatch=3.59.

Primer pair	Scoring thresholds		Error rates (%)			Number of loci		Replicates rate
	Locus calling	Phenotype calling	ε_1	ε_2	Mismatch	Initial	Final	
<i>Eco</i> _ACG- <i>Mse</i> _CGA	160	32	13.252	0.007	3.881	140	102	32.7
<i>Eco</i> _ACG- <i>Mse</i> _CTC	330	30	8.219	0.084	3.847	133	67	31.7
<i>Eco</i> _CAG- <i>Mse</i> _CGA	340	34	15.224	0.009	4.033	127	103	33.5
<i>Eco</i> _CAG- <i>Mse</i> _CTC	190	30	11.368	0.093	2.983	152	150	33.7

Abbreviation: AFLP, amplified fragment length polymorphism.

Genetic structure and outlier identification

The analysis of genetic structure was conducted separately with the two types of molecular markers (codominant nuclear markers and AFLPs) but the results proved to be very similar. Overall the genetic differentiation was weak, although significant, between Atlantic and Mediterranean populations of *M. galloprovincialis* (global $F_{ST} = 0.03$, $F_{ST} = 0.031$ for AFLPs and 0.025 for codominant markers). Figure I.2 shows F_{ST} at individual loci as a function of their heterozygosity together with the neutral envelope of F_{ST} obtained from simulations with the method of Beaumont and Nichols (1996) for codominant markers (Figure I.2a) and AFLPs (Figure I.2b). The results obtained with F-mtDNA ND2-COIII marker (Kijewski *et al.* 2011) were also added in Figure I.2a to illustrate that F-mtDNA ND2-COIII marker also

exhibited an outlying pattern of population differentiation although F-mtDNA ND2-COIII marker were not included in F_{ST} outlier detection methods due to its singular characteristics (haploid, uniparental inheritance, no-recombining). F_{ST} outliers were detected with the three methods of outlier detection but the results were not totally consistent. The method of Beaumont and Nichols (1996) identified seven outliers (2%), the method of Vitalis *et al.* (2003) identified three outliers that were also identified by the method of Beaumont and Nichols (1996), and the method of Foll and Gaggiotti (2008) identified only 2 outliers that were also detected with the two previous methods. Four loci were detected by only one method (*EF2*, *ODH*, *AFLP_1-187* and *AFLP_1-111*), one locus by two methods (*AFLP_1-218*) and two loci by the three methods (*EFbis* and *AFLP_3-363*). Individual F_{ST} values between Atlantic and Mediterranean populations of *M. galloprovincialis* at these seven loci were all above 0.2 whatever the type of marker.

Allelic frequencies of the *G* allele (defined as being more frequent in *M. galloprovincialis* samples than in reference *M. edulis* sample of the North Sea) in each of the five panmictic patches of *Mytilus* from the Mediterranean Sea to the North Sea (Figure I.1) are presented in Figure I.3. Five outlier loci, among which the two loci identified by every outlier detection methods (*EFbis*, *AFLP_3-363*), revealed a decrease of the *G* allele in the Mediterranean population. *AFLP_3-363* also revealed a surprisingly high frequency of the *G* allele in the *M. edulis* patch of the Bay of Biscay that might deserve further investigation. The two other outlier loci did not exhibit this pattern of a lower *G* allele frequency in the Mediterranean sample (*ODH* and *AFLP_1-218*). *ODH* is a tri-allelic locus and the shift in allele frequency relied on two alleles that are more frequent in *M. galloprovincialis* populations while the frequency of the third allele, which was more frequent in *M. edulis* populations, did not change between Atlantic and Mediterranean samples of *M. galloprovincialis* (Figure I.3). *AFLP_1-218* was the only one outlier locus that did not show strong allele frequency differences between *M. edulis* and *M. galloprovincialis*, so that the assignment of each allele to a species background was unreliable.

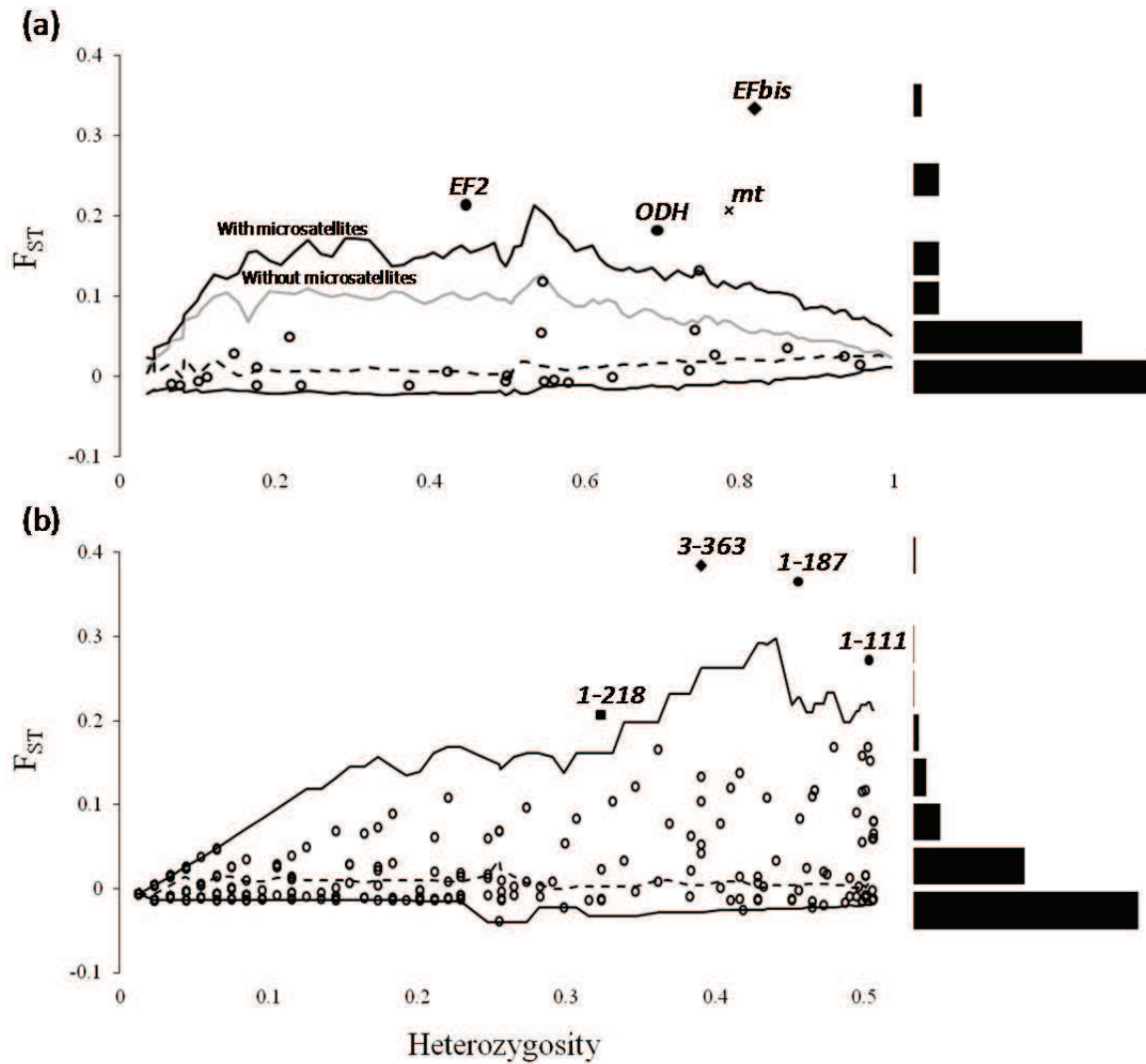


Figure I.2 : F_{ST} values between *M. galloprovincialis* of the Atlantic coast of the Iberian Peninsula and of the Mediterranean Sea plotted against heterozygosity for (a) 31 codominant nuclear loci (14 allozymes, 6 microsatellites, and 11 length polymorphisms); the cross represents the result of F-mtDNA ND2-COIII marker (Kijewski *et al.* 2011) and (b) 357 AFLP loci. Average (dotted line) and 95% confidence envelope (bold lines ; the superior limit of the 95% confidence interval for data excluding microsatellite loci is represented in grey) are the results from simulations performed with the FDIST2 program (a) and the DFDIST program (b). Outlier loci are annotated and represented by black dots (when identified by only one method), square (when identified by two methods) and diamonds (when identified by three methods, including BAYESCAN). On the right side of each scan is represented the observed F_{ST} distribution.

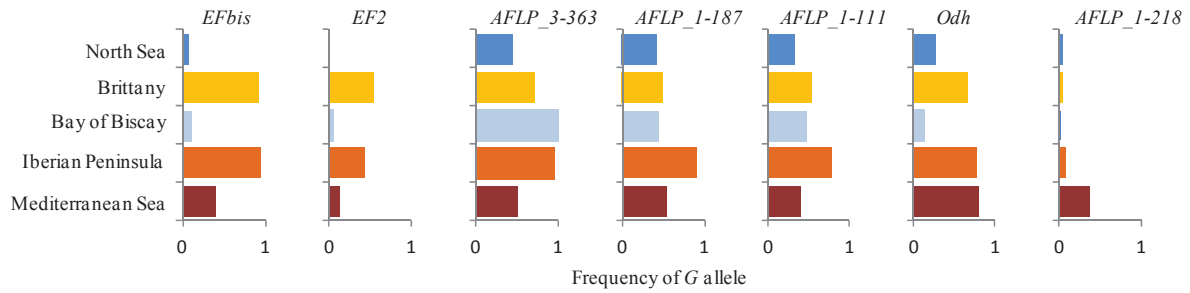


Figure I.3 .: Frequency of the *G* allele, defined as the compound of alleles more frequent in *M. galloprovincialis* samples than in the *M. edulis* reference sample of the North Sea, at the 7 outlier loci, in the five panmictic patches of the studied area (colors referred to the map, see Figure I.1).

Sequence analysis and phylogeny at outlier loci

Two of the seven outlier loci were chosen for subsequent analysis of DNA sequence polymorphisms. These two loci (*EFbis* and *EF2*) were those for which we had reference sequences to design primers and amplify a sufficiently long fragment (*EFseq* and *EF2seq*, respectively). For both loci the reconstructed gene phylogeny (Figure I.4) revealed two highly divergent clades: the first composed of sequences mostly sampled in *M. edulis* populations (*EF2seq* A-clade and *EFseq* A-clade) and the second composed of sequences mostly sampled in *M. galloprovincialis* populations (*EF2seq* B-clade and *EFseq* B-clade). For these two loci, the analysis of DNA sequence allowed to validate the *edulis* origin of A-clade alleles sampled in *M. galloprovincialis* populations and to conclude that the strong differentiation between Atlantic and Mediterranean populations is the consequence of a stronger introgression of *edulis* allele in the latter population.

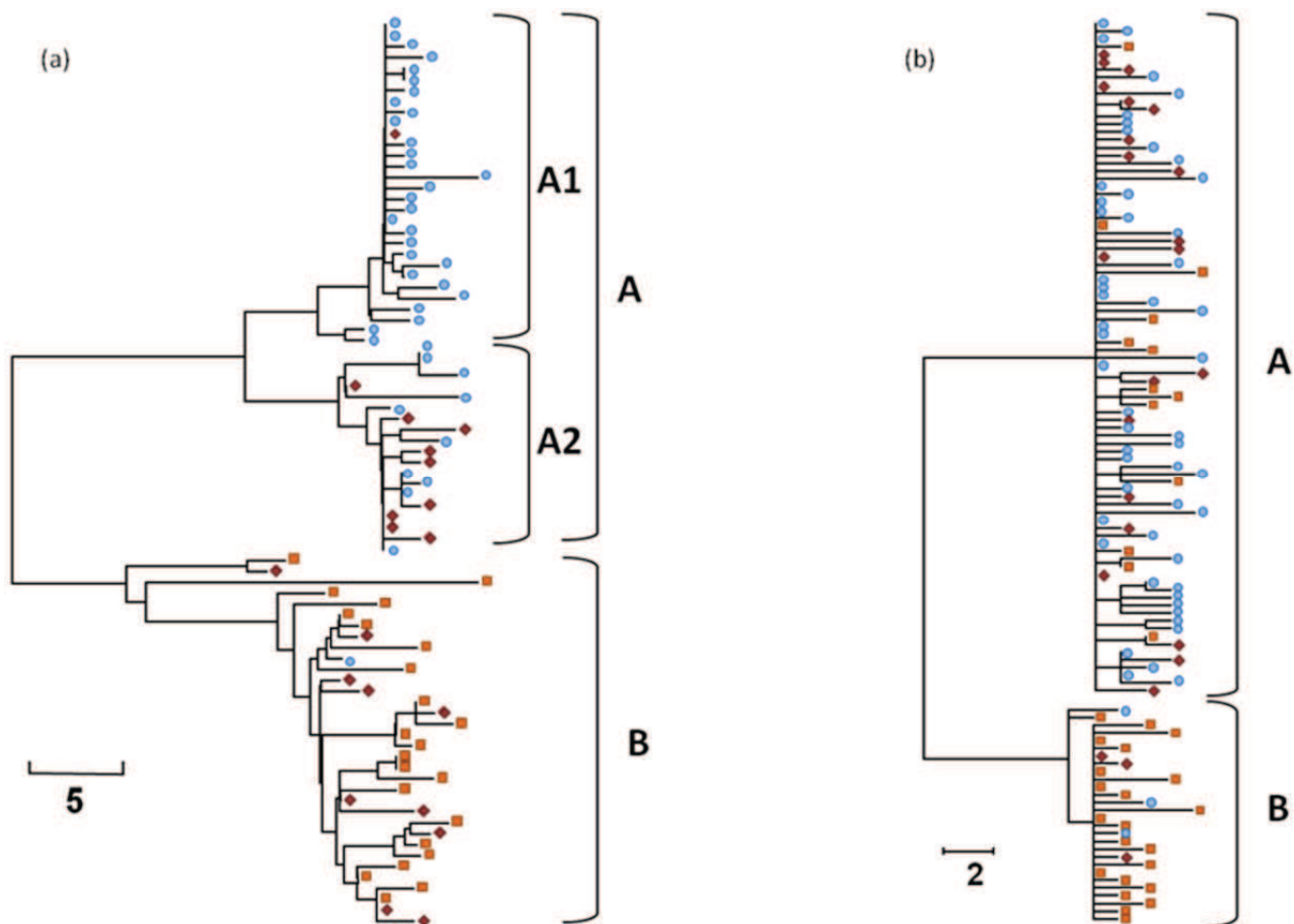


Figure I.4 : Allele phylogeny reconstructed by the Neighbor Joining algorithm on the number of nucleotide differences at (a) *EFseq* and (b) *EF2seq*. Sequences sampled in *M. edulis* populations of the North Sea and the Bay of Biscay are represented by blue dots. Sequences sampled in *M. galloprovincialis* populations of the Atlantic coast of the Iberian Peninsula and Brittany are represented by orange squares and sequences sampled in *M. galloprovincialis* populations of the Mediterranean Sea by red diamonds.

Discussion

The present study aimed to identify further the history and the genetic basis of high F_{ST} outlier loci across the AOF in the marine mussel *Mytilus galloprovincialis*. We first identified seven loci with an outlying level of differentiation between Atlantic and Mediterranean *M. galloprovincialis*. At five of the seven outlier loci, among which the two loci identified as outliers with every outlier detection method, the genetic structure proved to be the consequence of differential introgression of heterospecific alleles from the sister hybridizing species *M. edulis*. Surprisingly, the Mediterranean population was more introgressed than the Atlantic population. We discuss the possible evolutionary processes that can underlie such a pattern of differential introgression.

The genomic architecture of differentiation across the AOF

The AOF is positioned at the Easternmost edge of the Alboran Sea, a 300 km transition zone between superficial Atlantic waters and deep Mediterranean Sea waters (Tintoré *et al.* 1988). Gradients of temperature (1.4°C) and salinity (2 psu), water currents (40 cm/s), and the multiple eddies and gyres across a 2 km zone from the southeast Iberian Peninsula (Almeria) to Algeria (Oran), create a hydrogeographical barrier with biological, geological and chemical consequences (e.g. Sardà *et al.* 2004). Although the AOF is recognized as a hotspot of genetic structure (Quesada *et al.* 1995a; Quesada *et al.* 1995c; Pérez-Losada *et al.* 2002; Duran *et al.* 2004; Lemaire *et al.* 2005), examples exist of species that do not exhibit any genetic shift at the Almeria-Oran front (e.g. Launey *et al.* 2002) while their population sizes and dispersal capabilities are similar to those exhibiting such a shift. Several authors have therefore concluded that the AOF itself cannot be directly responsible for genetic shifts (Borsa *et al.* 1997; Lemaire *et al.* 2005; Patarnello *et al.* 2007; Bierne *et al.* 2011). Tension zones are expected to be trapped by natural barriers to gene flow (Barton 1979) and/or environmental boundaries (Bierne *et al.* 2011), and the AOF is both. It has therefore been hypothesized that the AOF has trapped secondary contact tension zones in many marine species (Bierne *et al.* 2011).

In *M. galloprovincialis* the Atlantic-Mediterranean phylogeographic split has been described with mtDNA-RFLP and sequence data (Quesada *et al.* 1995a; Sanjuan *et al.* 1996; Quesada *et al.* 1998a; Quesada *et al.* 1998b; Hilbish *et al.* 2000), allozymes (Sanjuan *et al.* 1994; Quesada *et al.* 1995c), nuclear marker *mac-1* (Daguin *et al.* 2001), microsatellites (Diz & Presa 2008) and a combination of mitochondrial and nuclear markers (Śmietanka *et al.*

2010; Kijewski *et al.* 2011). The Atlantic-Mediterranean split suggests the existence of two main entities in *M. galloprovincialis*. Quesada *et al.* (1995b) argued that the genetic shift is a consequence of a secondary contact between allopatric populations and that the genetic divergence between Atlantic and Mediterranean populations predates the origin of the AOF, explaining the geographic coincidence of genetic breaks observed with different markers. However, the secondary contact hypothesis does not seem sufficient alone to explain why the genetic differentiation has not dissipated since gene flow has been restored between the two entities. In fact the genetic differentiation is mainly visible with mitochondrial loci and the allozyme locus *ODH*. Our genome scan reveals that these two loci exhibit an outlying level of differentiation (Figure I.2a) and that the average differentiation, although significant, is modest ($F_{ST} = 0.03$). The semi-permeable nature of the barrier to gene flow may well sustain the hypothesis that a genetic barrier is superimposed on the natural and environmental barrier at the AOF and call for a better understanding of the processes that maintain the genetic differentiation at these two and the newly discovered outlying loci.

Differential introgression explains most outlier loci

The shift in allele frequency at six of the seven outlier loci was associated with a strong decrease of *galloprovincialis* alleles in the Mediterranean sample (Figure I.3). The phylogenetic analysis of DNA sequences at *EFbis* and *EF2* (Figure I.4) allowed us to validate secondary introgression as the explanation of the pattern at least for these two loci. Unexpectedly, the Mediterranean population proved to be the most introgressed while it is geographically isolated from *M. edulis* populations nowadays.

The first issue is to understand how *edulis* alleles could have reached the Mediterranean population. They could have transited at low frequency through Atlantic populations. They could also have been introduced following farming of *M. edulis* in Mediterranean waters or incidental introduction by boats. In these two cases, selection should have favored these heterospecific alleles in Mediterranean populations for their frequency to increase quickly and in a differential proportion to the rest of the genome. Alternatively, introgression could be historical and implies a direct contact between *M. edulis* and Mediterranean *M. galloprovincialis*. This might have well occurred during glacial periods while population distribution was shifted southward, making a contact possible somewhere at the entrance or within the Mediterranean Sea. Evidence for this latter hypothesis can be found in the results obtained at the *EFbis* locus in *M. edulis* populations (Faure *et al.* 2008; Bierne 2010). Indeed,

the genetic signature of a selective sweep has been observed in *M. edulis* populations in the form of a star-shaped genealogy (the A1 clade in Figure I.4, see Faure *et al.* 2008). Fitting a hitchhiking model to the results obtained on consecutive SNPs distributed along a 5Kb long region around *EFbis*, Bierne (2010) managed to infer that A1 alleles increased from a frequency of 33% before the selective sweep to a frequency of 97% in *M. edulis* populations of the North Sea and a frequency of 57% in populations of the Bay of Biscay. A1 alleles represent 37% of the *edulis* alleles found in Mediterranean *M. galloprovincialis* (Figure I.5), which corresponds roughly to the frequency of A1 before the selective sweep in *M. edulis*. One might thus easily speculate that introgression between *M. edulis* and Mediterranean *M. galloprovincialis* happened before the intraspecific sweep in *M. edulis* and that hitchhiking has subsequently modified allele frequencies in *M. edulis* while the two species were no longer in contact. As the selective sweep dates back to about 25000 years ago, the contact should have happened before. Using complete female and male mitochondrial genomes, Smietanka *et al.* (2010) identified three historical introgression events between *M. edulis* and *M. galloprovincialis*, at 350, 200 and less than 100 thousand years ago. It is therefore likely that several contacts have occurred between these two species, probably during interglacial periods, in a geographic context that cannot be inferred from the present distribution of the two species. It is interesting to emphasize that the younger event is a complete capture of the *edulis* female genome in Atlantic *M. galloprovincialis* (Quesada *et al.* 1998a), which illustrates that differential introgression needs not be always in the same direction depending on the genome regions.

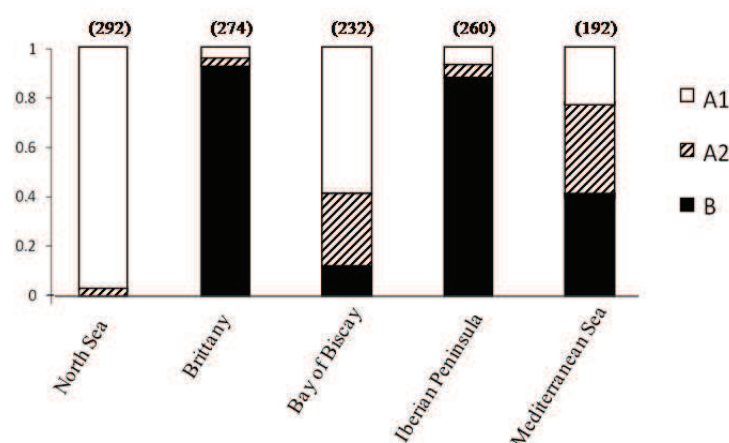


Figure I.5 : Allele frequencies of clade A1 alleles (white), clade A2 alleles (striped) and clade B alleles (black) at the *EFbis* locus deduced from size-allele frequencies. The number of sampled alleles is given within brackets above each histogram.

Heterospecific alleles have therefore probably introgressed Mediterranean *M. galloprovincialis* populations during a historical contact between these two entities, but the second issue is to understand why they have remained confined in the Mediterranean background at some genomic regions while the differentiation is modest on the rest of the genome. Secondary introgression has been detected at other loci with a similar frequency of *edulis* alleles in Atlantic and Mediterranean populations (Boon *et al.* 2009) and we detected no tendency for the Mediterranean population to be more introgressed at the genome level -it is only at outlier loci that this pattern is observed. We offer two hypotheses to explain differential introgression. (i) Introgression could be adaptive, implying *edulis* alleles were favored in the Mediterranean Sea (the “adaptive introgression” hypothesis). (ii) Or the genetic architecture of the barrier to gene flow could be different between the two *M. galloprovincialis* backgrounds, implying the Mediterranean background was permeable to *edulis* introgression at the genome regions considered while the Atlantic background is not (the “semi-permeable barrier” hypothesis). To sustain the latter hypothesis, we verified if the outlier loci identified between the two *M. galloprovincialis* populations were also outliers, or in the upper bound of the differentiation distribution, between Atlantic *M. galloprovincialis* and *M. edulis*. Except for one outlier locus (*AFLP_1-218*) and the F-mtDNA ND2-COIII marker, outliers between *M. galloprovincialis* populations were also outliers between *M. edulis* and Atlantic *M. galloprovincialis* (Figure I.6), suggesting that the genome of Atlantic *M. galloprovincialis* resists introgression from both *M. edulis* and Mediterranean *M. galloprovincialis* at most of the regions identified in our genome scan. While this observation is in accordance with the “semi-permeable barrier” hypothesis it is not sufficient proof. Local adaptation could well have favored the same alleles in *M. edulis* and Mediterranean *M. galloprovincialis* responding to a similar selective pressure, or have repeatedly targeted the same genes at different time scales, during species divergence and during population divergence within species. The *ODH* locus for instance is outlier in both comparisons but the differentiation within *M. galloprovincialis* is independent of introgression. Distinguishing hypotheses on the basis of an F_{ST} scan analysis alone seems difficult. A decisive test would require walking on the chromosome toward the direct target of selection: under the adaptive introgression hypothesis we should find a genome region fixed for *edulis* alleles near the outlier loci identified in this study, while under the semi-permeable barrier hypothesis introgressed alleles should not necessarily reach fixation in the *M. galloprovincialis* background.

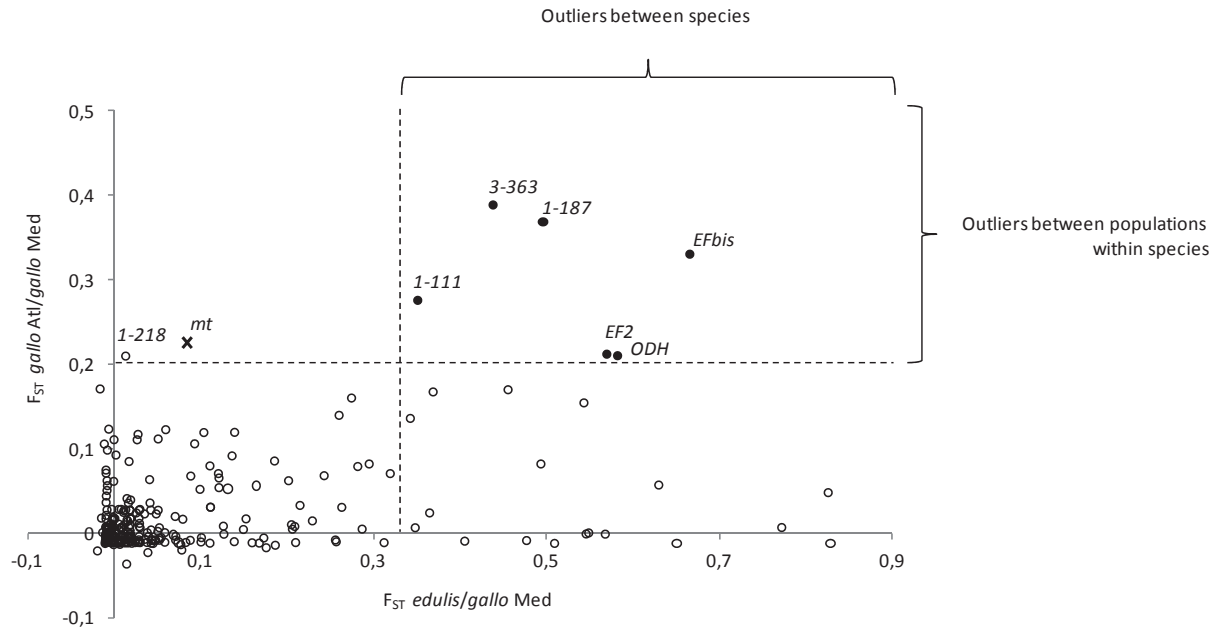


Figure I.6 : Scatterplot of the genetic differentiation between *M. edulis* (North Sea) and Atlantic *M. galloprovincialis* as a function of the differentiation between Atlantic *M. galloprovincialis* and Mediterranean *M. galloprovincialis*. Dashed lines represent outlier F_{ST} thresholds for both comparisons. Dots represent AFLP markers and codominant nuclear markers and the cross represents the result of F-mtDNA ND2-COIII marker (Kijewski *et al.* 2011). Outliers identified in the two comparisons are in black.

Conclusion

Our study highlights the importance of history, and probably also of the genetic architecture of barriers to gene flow, in the genesis of strong genetic structure at specific loci and genome regions. It becomes increasingly documented that adaptive polymorphisms can easily cross species boundaries (Arnold 2004; Anderson *et al.* 2009; Vekemans 2010; Dasmahapatra *et al.* 2012; Staubach *et al.* 2012) and this source of new adaptation might be more common than traditionally recognized (The FroSpects Gregynog Workshop 2012). Furthermore, natural populations are often subdivided into partially isolated genetic backgrounds (Charlesworth *et al.* 2003). These backgrounds can sometimes be recognized as taxonomic entities (ecotypes, races, subspecies or species) but not always. They might have remained invisible to preliminary phylogeographic analysis with a handful of molecular markers if the genetic barrier to gene flow affects unsampled portions of the genome (Bierne *et al.* 2011). Ignoring these subdivisions and focusing on a local scale without analyzing outlying polymorphisms in sister (sub)species might lead to imprecise interpretations. Gagnaire *et al.* (2011) have recently showed that F_{ST} outliers identified within a Pacific eel species were also strongly differentiated between species. The slow leaking of heterospecific alleles in genomic island of differentiation revealed a within species population structure that was otherwise invisible with neutral markers. We here provide another example of how the hybridization with a foreign genome can influence within species population structure, and concur to the recommendation of widening the spatial and temporal scales at which one should investigate and interpret F_{ST} scans.

Acknowledgements

We acknowledge Michel Cantou for sampling. This study was supported by the Agence National de la Recherche (Hi-Flo project ANR-08-BLAN-0334) and the project Aquagenet (SUDOE, INTERREG IV B). This is article 2012-XXX of Institut des Sciences de l'Evolution de Montpellier.

CHAPITRE II: Adaptation d'un peptide antimicrobien à partir d'un polymorphisme pré-existant

II.1- Résumé

L'identification de locus anormalement différenciés par les scans génomiques constitue une approche intéressante pour détecter des locus candidats montrant une différenciation génétique anormale mais n'expliquent pas les causes responsables d'une telle structure. Pour comprendre comment la sélection a agit, il est nécessaire d'étudier ces locus plus en détail. Un polymorphisme leucine/arginine (L31R) anormalement structuré entre populations *M. edulis* et situé dans la boucle active d'une défensine (le peptide antimicrobien MGD2) a été étudié en détail. Nous avons analysé les polymorphismes de séquences d'ADN, les fréquences des allèles et la différenciation des polymorphismes étroitement liés de chaque côté de L31R, ainsi que les déséquilibres de liaison par paires et de troisième ordre. Aucune différenciation anormale n'a été détectée aux locus voisins de L31R, comme c'est le cas pour la majorité du génome, démontrant que la sélection agit directement sur la boucle active du peptide. Trois hypothèses permettent d'expliquer l'absence de structure dans le voisinage chromosomique de L31R: (i) un autostop génétique a eu lieu mais a ensuite été effacé par la migration et la recombinaison, (ii) la sélection était assez faible et la migration assez forte pour empêcher l'effet d'autostop à l'échelle chromosomique considérée ou (iii) la sélection a agit sur un polymorphisme pré-existant à l'équilibre de liaison avec son fond génétique. Comme attendu pour ces trois hypothèses, l'équilibre de liaison est respecté entre L31R et ses polymorphismes voisins dans toutes les populations étudiées. Cependant, des déséquilibres de liaison ont été mesurés dans certaines populations entre paires de locus situés de part et d'autre de L31R, générant un patron complexe mieux expliqué par l'action de la sélection sur un polymorphisme pré-existant. Sans toutefois être en mesure de bien comprendre l'histoire sélective de ce locus chez *M. edulis* à cause des événements d'introgession avec son espèce sœur *M. galloprovincialis*, nous pouvons avancer quelques suppositions. Ainsi, la sélection pourrait être soit balancée, maintenant des allèles à des fréquences différentes selon le pathosystème localement rencontré par les moules, ou intermittente, entraînant des fluctuations sporadiques dans la fréquence des allèles.

II.2- Evidence for adaptation from standing genetic variation on an antimicrobial peptide in the mussel *Mytilus edulis*

Célia Gosset^{1,2}, Joana Do Nascimento^{1,2}, Marie-Thérèse Augé^{1,2} and Nicolas Bierne^{1,2}

¹ Université Montpellier 2, Place Eugène Bataillon, 34095 Montpellier, France

² CNRS - Institut des Sciences de l'Évolution UMR5554 Montpellier, France.

Article soumis dans la revue *Molecular Ecology*

Abstract

Genome scans of population differentiation identify candidate loci for adaptation but provide little information on how selection has influenced the genetic structure of these loci. Following a genome scan we investigated the nature of the selection responsible for the outlying differentiation observed between populations of the marine mussel *Mytilus edulis* at a Leucine/Arginine polymorphism (L31R) in the antimicrobial peptide MGD2. We analysed DNA sequence polymorphisms, allele frequencies and population differentiation of closely linked polymorphisms each side of L31R, and pairwise and third-order linkage disequilibria. An outlying level of population differentiation was observed at L31R only, while no departure from panmixia was observed at neighbour loci, as at most of the genome, providing convincing evidence that selection affects L31R directly. Three hypotheses can explain the lack of differentiation in the direct neighbourhood of L31R: (i) hitchhiking has occurred but migration and recombination subsequently erased the signal, (ii) selection was weak enough and recombination strong enough to prevent hitchhiking effect at a very small chromosomal scale or (iii) selection acted on a pre-existing polymorphism (i.e. standing variation) that was at linkage equilibrium with its background. Linkage equilibrium was observed between L31R and neighbour polymorphisms in every populations analysed, as expected under the three hypotheses. However, linkage disequilibrium was observed in some populations between pairs of loci localised each side of L31R, generating a complex pattern of third-order linkage disequilibria which is best explained by the hypothesis of selection on a pre-existing polymorphism. We were not able to fully understand the long and complex history of selection at this locus which is further complicated in *M. edulis* by introgression from the sister hybridizing species *M. galloprovincialis*. However, we hypothesise selection could be either balanced, maintaining alleles at different frequencies depending on the pathosystem locally faced by mussels, or intermittent, resulting in sporadic fluctuations in allele frequency.

Introduction

Scanning genomes for loci with high levels of population differentiation (Fst scans) has become a standard of population genetics (Luikart *et al.* 2003; Stapley *et al.* 2010). Fst outliers are recognised to be nothing else than candidate loci on which subsequent and detailed investigations are required (Bierne *et al.* 2013). One would not only want to validate the effect of selection on these loci (i.e. to refute the false positive hypothesis) but also to understand how selection has affected them. For instance we often would like to know if they are themselves under direct selection or if they are neutral polymorphisms linked to selected loci, and if the latter we would like to have access to selected loci and their function. Another important issue is the age of adaptive polymorphisms, if they are recent or if they have been maintained for long periods (e.g. surviving several glacial cycles, Bierne *et al.* 2013). However few studies have pursued the analysis of F_{ST} -outlier loci in non-model species (Pogson 2001; Faure *et al.* 2008; Gosset & Bierne 2013) so that we have many lists of outliers in many species but have little idea of their origin, history and genetic basis.

In the present study we have pursued an Fst scan between populations of the marine mussel *Mytilus edulis* that was first conducted by Faure *et al.* (2008) with eleven loci. Faure *et al.* (2008) showed that European populations of *M. edulis* represent a single panmictic unit although one locus, called *EFbis*, proved to be outlier. A detailed analysis of the polymorphism at the *EFbis* locus revealed that the differentiation was the consequence of a selective sweep of different magnitude in every population (Bierne 2010). Here, we added nine nuclear codominant markers to the eleven of Faure *et al.* (2008) and 385 AFLP markers. This scan confirmed that most of the genome does not depart from panmixia in European populations of *M. edulis*, and allowed us to identify five new high Fst outlier loci -four AFLP markers and a Leucine/Arginine polymorphism at the 31st amino acid position (*L31R*) of the mature antimicrobial peptide MGD2 (Mitta *et al.* 2000; Boon *et al.* 2009). We analysed DNA sequence polymorphism of a 700bp-long fragment flanking each side of *L31R* on a sample set of 39 chromosomes. From these data and previously published data (Boon *et al.* 2009) we identified four indel polymorphisms, two each side of *L31R*, that were used to analyse allele frequencies and linkage disequilibrium on a large sample size ($n=224$ mussels, or 448 chromosomes). This allowed us to validate that selection was acting directly on *L31R* and to obtain evidence from a detailed analysis of linkage disequilibria that this polymorphism predated the last action of selection.

Materials and Methods

Samples

Mytilus edulis samples were collected at five localities in Europe: LU (Lupin, Charente-maritime, France) in the Bay of Biscay, CA (Calais, Pas-de-Calais, France) in the English Channel, WS (Wadden Sea, Holland), CLEY (Cley, Norfolk, United Kingdom) in the North Sea and KEN (Kenmare, Kerry, Ireland) on the Atlantic Coast of Ireland (see Figure II.1). We also used two samples of the sister hybridizing species *M. galloprovincialis*: FA (Faro, Algarve, Portugal) and PR (Primel, Finistère, France) in the Atlantic Coast of the Iberian Peninsula and Brittany, respectively. Sample size was 48 individuals per population with the exception of CLEY for which the sample size was 32. These samples have been analysed previously at least with three length-polymorphism markers (*mac-1*, *EFbis* and *Glu-5'*, Bierne *et al.* 2003; Faure *et al.* 2008; Gosset & Bierne 2013).

AFLP typing

DNA extraction and quantification, and the AFLP experiments (preselective and selective PCR amplifications) followed standard protocols previously described in Gosset and Bierne (2013). Selective-PCR products were separated and detected using an Applied ABI Prism 3130XL (Applied Biosystems) and sized by GeneScan™ 500 ROX™ Size Standard (Applied Biosystems). GeneMapper® v4.5 software (Applied Biosystems) was used to read the resulting chromatograms. Fragment length classes (bin sets) were created automatically and manually revised for each primer combination in the interval 50-500 base pairs. Only unambiguously identifiable fragments were translated into a binary matrix of peak intensity values for each detected locus at each sample. This matrix was secondary translated into a phenotype matrix of presence (1) absence (0) of the fragments with AFLPScore v1.4b (Whitlock 2008). Nomenclature for AFLP scoring thresholds follows Whitlock *et al.* (2008), where 'Locus-calling' and 'Phenotype-calling' are thresholds used to exclude error-prone loci from the AFLP genotypes table, and to determine fragment phenotype (band presence or absence) based on raw chromatogram peak-height data, respectively. Error rates were measured using the method of Hadfield *et al.* (2006), modified for dominant data as implemented in the R package MasterBayes, which estimates the rate of error applying to each of the two alleles that underlie the genotypes scored using AFLP via a Gibbs sampler (Whitlock 2008). Under this method, error rates ϵ_1 and ϵ_2 determine the probability of

misscoring an AFLP fragment presence '1' or absence '0' allele, respectively. We also determined a widely used measure of error (the mismatch error rate) that compared the dissimilarity of pairs of AFLP phenotype profiles that had originated from the same genetic individual (replicated samples). We estimated allelic frequencies at each AFLP marker with AFLPsurv v1.0 (Vekemans 2002; Vekemans *et al.* 2002) assuming Hardy-Weinberg equilibrium in each population and using a bayesian method with non-uniform prior distribution of allele frequencies (Zhivotovsky 1999).

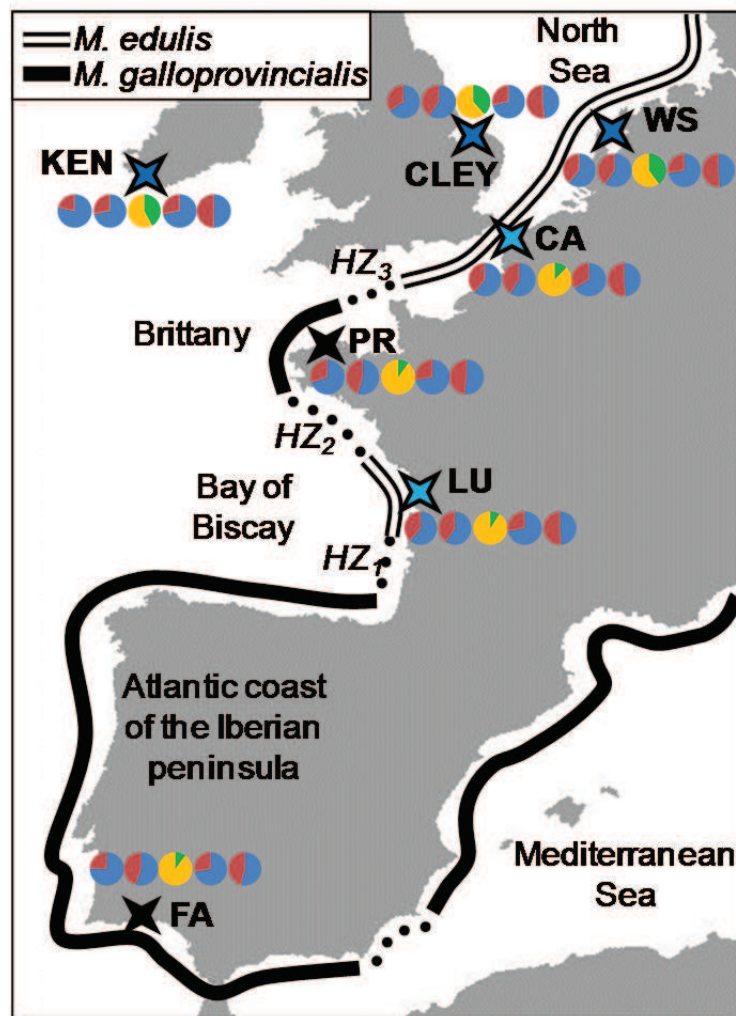


Figure II.1 : Localities of *Mytilus* spp. samples. Sample sites are indicated by blue and black stars for *M. edulis* and *M. galloprovincialis* respectively. Light blue stars are for southern *M. edulis* populations ($p_R \sim 0.1$) and dark blue stars are for northern *M. edulis* populations ($p_R \sim 0.4$). The three hybrid zones described by Bierne *et al.* (2003) are indicated by dashed lines: HZ1 between *M. galloprovincialis* of the Atlantic coast of the Iberian Peninsula and *M. edulis* of the Bay of Biscay, HZ2 between *M. edulis* of the Bay of Biscay and *M. galloprovincialis* of Brittany and HZ3 between *M. galloprovincialis* of Brittany and *M. edulis* of the North Sea. Patches of *M. galloprovincialis* are indicated with continuous black lines and patches of *M. edulis* are indicated by black and white continuous lines.

Nuclear codominant markers

We used the compilation of published data and the data of Faure *et al.* (2008) composed of five allozyme and six length-polymorphism loci, to which we have added new results on four intron-length polymorphisms reported in Gosset and Bierne (2013). Here we analyzed five additional polymorphisms in the MGD2 gene. Their positions are given in Figure II.2. One was previously described by Boon *et al.* (2009). It is a 3bp (one codon) indel in exon 2. Another one was identified as being potentially interesting from the dataset of sequences published by Boon *et al.* (2009). It was a Leucine/Arginine polymorphism at position 31 of the mature peptide (*L31R*), in the loop thought to play a key role in the antimicrobial activity (Romestand *et al.* 2003). It is a radical change that requires two nucleotide substitutions in the codon. The R allele was found surprisingly frequent in a sample of the North Sea but the sample size (28 chromosomes) did not provide enough power to validate the observed differentiation to be unexpected (Boon *et al.* 2009). The three other polymorphisms were developed during the course of the present study using indels in intron 2 and intron 3 of the MGD2 gene. One indel of intron 3 (locus 4 in Figure 2), is ~1Kb long (with an additional nested indel of 187bp). Allele-specific PCR primers were designed that selectively amplify only one allele (Table 1). Homozygous genotypes amplify with one of the two primers only, while heterozygous genotypes amplify with the two primers (Bierne 2010). This technique was used for genotyping the four indels (one in exon 2, one in intron 2 and two in intron 3 of MGD2) as well as the *L31R* polymorphism in exon 3 using the two consecutive nucleotide polymorphisms responsible for the amino acid change at the corresponding codon.

DNA polymorphism

The *MGD2* gene sequences published by Mitta *et al.* (2000) and Boon *et al.* (2009) together with unpublished sequences of intron 3, were used to design primers in intron 2 and intron 3 flanking exon 3 that contains the sequence of the mature peptide (Figure 2). MGD2 is highly polymorphic in *Mytilus edulis* and *M. galloprovincialis* (Boon *et al.* 2009) and conserved regions are scarce. Furthermore a polymorphic ~1000pb long indel was discovered in intron 3 during the course of this work (locus 4 in Figure 2) which has prevented us to analyse a sequence length as long as we had hoped. We identified a small conserved region in the last third of intron 2 which allowed us to design the forward primer: MgD2Intron2_910-F (5'-ATTAAAYAACTGAGCACATACAG-3'). We took advantage of a 6bp indel

polymorphism at the end of intron 3 (locus 5 in Figure 2), the frequency of which has proved to be ~0.5 in every population, to selectively amplify the two alleles of heterozygous individuals at this indel (see Bierne 2010). The allele-specific PCR reverse primer sequences are given in Table II.1 (3231in-R and 3231trans-R). As the frequency of this indel was ~0.5 in every population, this procedure to amplify and directly sequence hemizygous DNA segments from a heterozygous template did not introduce any sampling bias. Because of the ~1Kb indel, some sequences were ~1700bp long and others were ~700bp long. Small alleles were sequenced with the two PCR primers while long alleles were sequenced with four primers, the two PCR primers and two internal primers, allowing a forward/reverse double check at each nucleotide position outside of the 1Kb indel.

Table II.1 : List of primers used to genotype each of the five loci in the MGD2 gene.

Locus	Forward primers	Reverse primers
Indel-20	20in-F: TGGTTTGTGCATCATGATGATG	1550-R: ACTTGCA CAGTATCCCCACAGTA
	20trans-F: TGGTTTGTGCATCATGACGGAC	
Indel-730	730in-F: GATGACTCTGAAAATATCAGCTGA	1654-R: TACTAGATTCAGAAACGTTCTACTG
	730trans-F: GAGATGACTCTGAAAATATCAACG	
L/R34	1556_L-F: GATACTGTGCAAGTTGGTTCAGATT	1654-R: TACTAGATTCAGAAACGTTCTACTG
	1556_R-F: GATACTGTGCAAGTTGGTTCAGACG	
indel-1510	1510-F: AACAAATTATGCCTGTCATCAACAC	1967-R: CCGCATCTCCACATTAGGAAGAAGT
		2919-R: GTCTGGAACATTTGGTGTCTACTAAG
Indel-3231	1510-F: AACAAATTATGCCTGTCATCAACAC	3231in-R: GAAGAGGACACATTATAATTGTTTAA
		3231trans-R: GAAGAGGACATATTATAATTGTTAAC

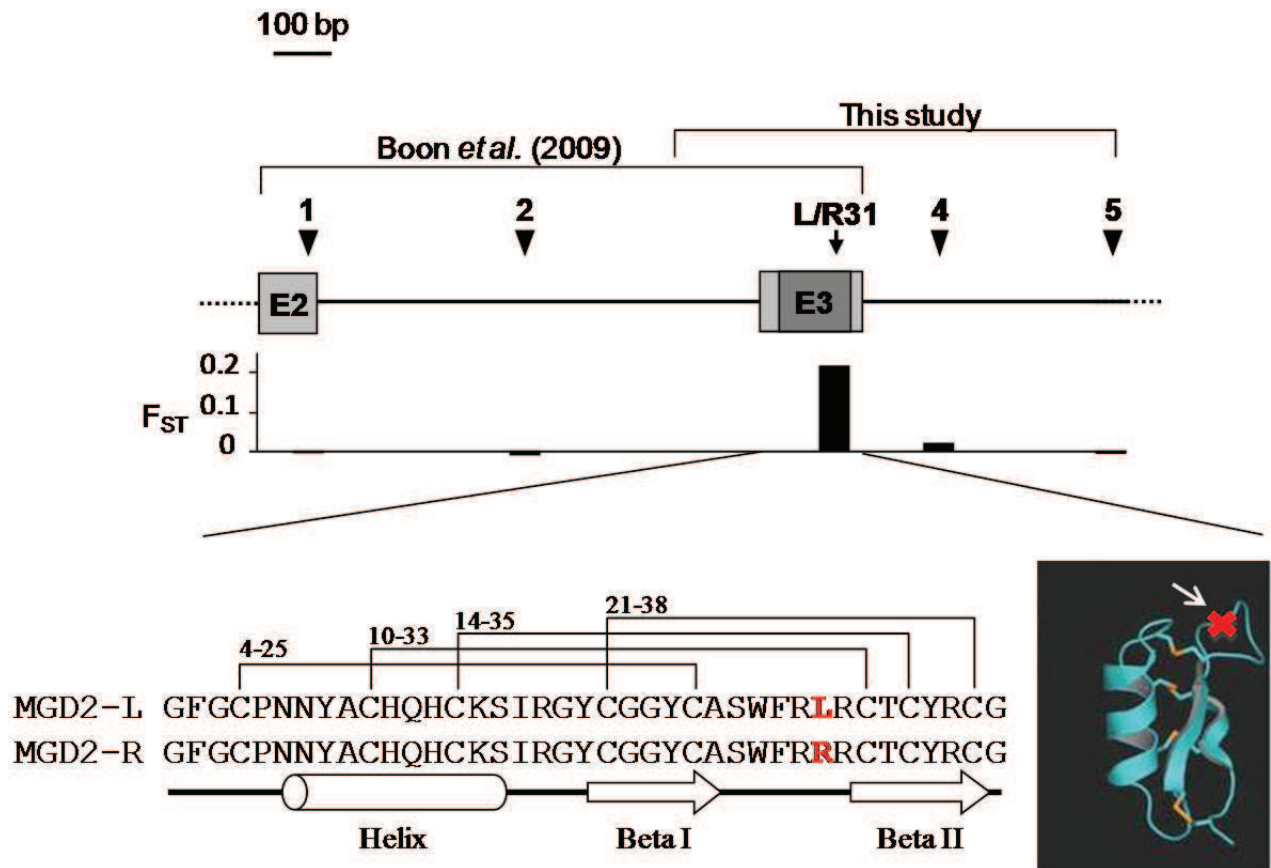


Figure II.2 : Structure of the MGD2 gene, genetic differentiation in *M. edulis*, and mature peptide structure. Positions of the five polymorphisms analysed are mapped on the gene. Indel polymorphisms are depicted by triangles and *L31R* by an arrow. Below the gene structure is given the global F_{ST} values in *M. edulis* for each of the five polymorphisms. Zoom in the mature peptide primary, secondary and tertiary structure with the *L31R* polymorphism in red. Lines joining cysteines represent disulfide bonds. The structure of the peptide is schematized below the alignment.

Outlier loci detection

We used two different methods to identify loci that depart from the expected neutral distribution of F_{ST} . We used the method of Beaumont & Nichols (1996) to obtain the average and 95% confidence interval of single locus F_{ST} as a function of heterozygosity. Simulations were performed with the FDIST2 program (Beaumont & Nichols 1996) for codominant markers and DFDIST program, an extension of FDIST software that allows the use of dominant markers, for the AFLP dataset. We also used the method of Foll and Gaggiotti (2008) implemented in BAYESCAN v2.01 (Foll & Gaggiotti 2008; Foll *et al.* 2010). This method is an extension of the hierarchical-Bayesian approach of Beaumont & Balding

(2004), and is based on a logistic regression model. The posterior probability of one locus to be under selection is estimated by defining two alternative models. Selection is introduced by decomposing locus-population F_{ST} coefficients into a population-specific component (β), shared by all loci and a locus-specific component (α) shared by all the populations using a logistic regression. Departure from neutrality at a given locus is assumed when the locus-specific component is necessary to explain the observed pattern of diversity (α significantly different from 0). The respective posterior probabilities of these two models are estimated using a reversible jump Markov chain Monte Carlo (RJMCMC) approach that takes all loci into account in the analyses through the prior distribution, resolving the problem of multiple testing of a large number of genomic locations.

Analysis of DNA sequence polymorphism

Sequence alignment was performed with ClustalW (Thompson *et al.* 1994) in the Bioedit interface (Hall 1999) and verified by eye. For each sequence, indel composition was registered before excluding alignment gaps from further statistical analyses. We used DNAsp (Rozas & Rozas 1999) to compute basic population genetic parameters such as number of polymorphic sites (S), levels of nucleotide diversity estimated from the number of polymorphic sites θ_w (Watterson 1975), or from pairwise differences, θ_π (Nei 1987); Hudson *et al.*'s (1992) estimator of F_{ST} ; the minimum number of recombination, R_M , estimated by the method of Hudson and Kaplan (1985); and linkage disequilibrium as measured by the ZnS statistic (Kelly 1997). DNAsp was also used to compute Tajima's D (1989), a well-known statistic that proved efficient to detect departures of the allele frequency spectrum from the expectation at mutation-drift equilibrium. Departure from the neutral expectation at mutation-drift equilibrium was tested by coalescent simulations conditional on the number of segregating sites.

As already reported by Boon *et al.* (2009) with another portion of the gene the number of recombination breakpoints proved far too high to reconstruct a comprehensible genealogy for the entire length of the sequence. To represent an overview of the genetic relationships among alleles we inferred a network with the NeighborNet algorithm (Bryant & Moulton 2004) in the software SplitsTree4 (Huson & Bryant 2006).

Linkage disequilibrium

Sample size is an important issue in the analysis of linkage disequilibrium and we restricted the analysis of linkage disequilibria to *L31R* and the four neighbour polymorphisms for which we had a large sample size (448 chromosomes). Furthermore, samples for which the homogeneity of genotypic frequencies cannot be rejected at each of the five polymorphisms were pooled together. Linkage disequilibrium, *D*, was computed by the method of Black and Krafur (1985) in GENETIX4.2 (Belkhir *et al.* 2004) and tested with an exact test with genepop'007 (Rousset 2008). Third-order linkage disequilibria that include the *L31R* polymorphism were calculated from pairwise linkage disequilibria according to the decomposition approach of Gorelick and Laubichler (2004). However, third-order linkage disequilibria are difficult to comprehend and for a better understanding of the results we computed the estimated pairwise disequilibrium in the population of chromosomes carrying the L allele at *L31R* on the one hand and in the population of chromosomes carrying the R allele at *L31R* on the other hand.

Results

AFLP screening

385 AFLP loci were scored in total, of which 366 were polymorphic in *M. edulis* populations. The mismatch error rate, weighted to take into account the different numbers of loci arising from different primer combinations, was 4.37%. A summary of the scoring parameters used, the error rates and the final number of loci retained for each selective primer combination is given in the Table II.2.

Table II.2 : Summary of AFLP selective primer combinations used during genotyping. The scoring conditions applied in order to derive AFLP phenotypes and the genotyping error rate determined via AFLP fingerprinting of replicated samples are given for each primer pair. The initial and final numbers of loci refer to the initial number of fragment length classes (bins) considered, and the final number of loci retained for analysis after scoring via AFLPScore (Whitlock et al. 2008). The replicates rate refers to the number of replicate pairs of fingerprints from which genotyping error was estimated. Total number of loci=385 of which 366 are polymorphic between the two populations studied here; weighted mean percentage error rates, $\epsilon_1=15.87$, $\epsilon_2=0.09$, mismatch=4.37.

Primer pair	Scoring thresholds		Error rates (%)			Number of loci		Replicates rate
	Locus calling	Phenotype calling	ϵ_1	ϵ_2	Mismatch	Initial	Final	
<i>Eco</i> _ACG- <i>Mse</i> _CGA	170	41	16.704	0.084	4.952	101	91	27.5
<i>Eco</i> _ACG- <i>Mse</i> _CTC	290	30	16.901	0.118	4.787	131	61	29.8
<i>Eco</i> _CAG- <i>Mse</i> _CGA	340	40	15.169	0.063	4.455	127	91	25.5
<i>Eco</i> _CAG- <i>Mse</i> _CTC	230	45	15.344	0.115	3.767	152	142	27.5

Abbreviation: AFLP, amplified fragment length polymorphism.

Genetic structure and outlier identification between M. edulis populations of the North Sea and the Bay of Biscay

The analysis of genetic structure was conducted separately with the two types of molecular markers (codominant nuclear markers and AFLPs). Overall the genetic differentiation was weak and did not depart from panmixia at most loci (global $F_{ST} = 0.019$, $F_{ST} = 0.022$ for AFLPs, $F_{ST} = 0.004$ for codominant markers). Figure II.3 shows F_{ST} at individual loci as a function of their heterozygosity together with the neutral envelope of F_{ST} obtained from

simulations with the method of Beaumont and Nichols (1996) for codominant markers (Figure II.3a) and AFLPs (Figure II.3b). Six F_{ST} outlier loci were detected with the two methods of outlier detection, two codominant loci (*EFbis* as already reported by Faure *et al.* 2008 and Bierne 2010, and *L31R* the focus of the present paper) and 4 AFLP loci. Five additional AFLP loci were detected as outliers with the method of Beaumont and Nichols (1996) but not with the method of Foll and Gaggiotti (2008). The important results were (i) that *L31R* proved to be a strongly supported high F_{ST} outlier with every method and (ii) that a broad sampling of the genome with AFLP markers confirmed that most of the genome did not depart from panmixia in European *M. edulis*.

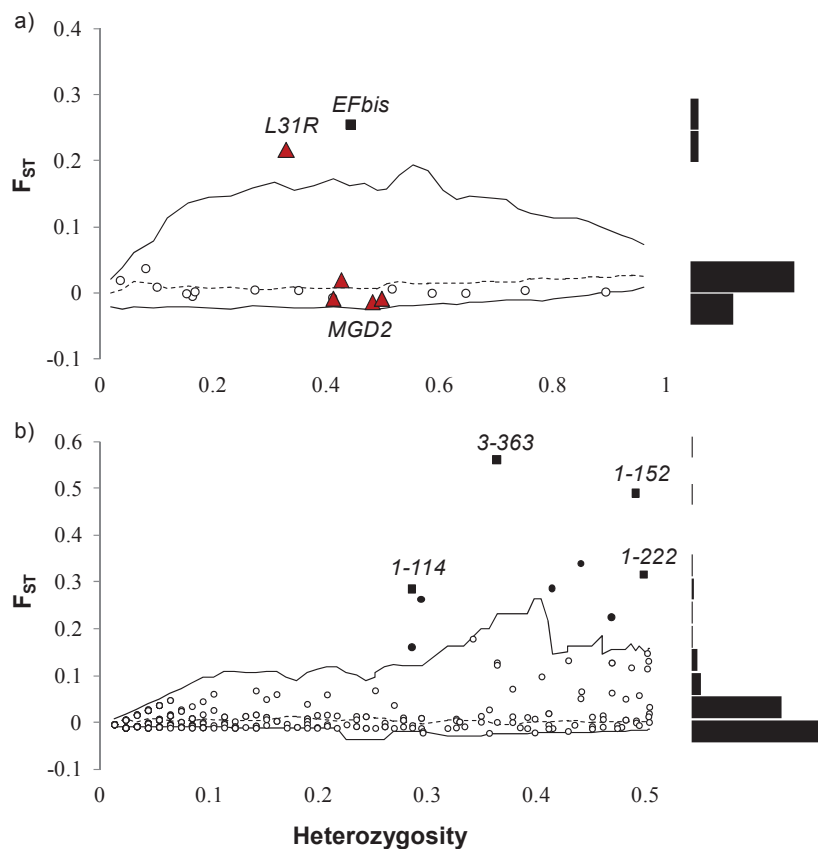


Figure II.3 : F_{ST} scans. F_{ST} values between *M. edulis* samples of the North Sea and the Bay of Biscay plotted against heterozygosity for (a) 19 codominant nuclear loci (5 allozymes and 14 length polymorphisms) and (b) 366 AFLP loci. Average (dotted line) and 95% confidence envelope (bold lines) are the results from simulations performed with the FDIST2 program (a) and the DFDIST program (b). Triangles: the five loci in the MGD2 gene. Outlier loci are annotated and represented by black squares (or triangle for *L31R*) when also identified as outlier with BAYESCAN. On the right side of each scan is represented the observed F_{ST} distribution.

Sequence analysis at MGD2 in the neighbourhood of the Fst outlier locus L31R

Basic descriptors of polymorphism at the *MGD2* gene portion analysed are presented in Table II.3. As previously reported by Boon *et al.* (2009), although on a different portion of the gene (see Figure II.2), a high level of nucleotide diversity, a low level of differentiation between populations, and many intragenic recombination breakpoints were observed. A network inferred with the NeighborNet algorithm depicts the genetic relationship among alleles in Figure II.4. Even though highly reticulated, alleles amplified with the same reverse primer tend to cluster together indicating linkage disequilibrium. Coalescence simulations revealed that the level of linkage disequilibrium as revealed by ZnS was significantly lower than expected in the absence of recombination (Table II.3). ZnS was lower in northern populations where the R allele is more frequent (WS and KEN) than in the South (CA and LU), an observation we will come back below with indel polymorphism data on a more substantial sample size. The network of Figure II.4 also shows that haplotypes carrying the R allele do not cluster together as would have been expected in a “hard sweep” model (Pennings & Hermisson 2006).

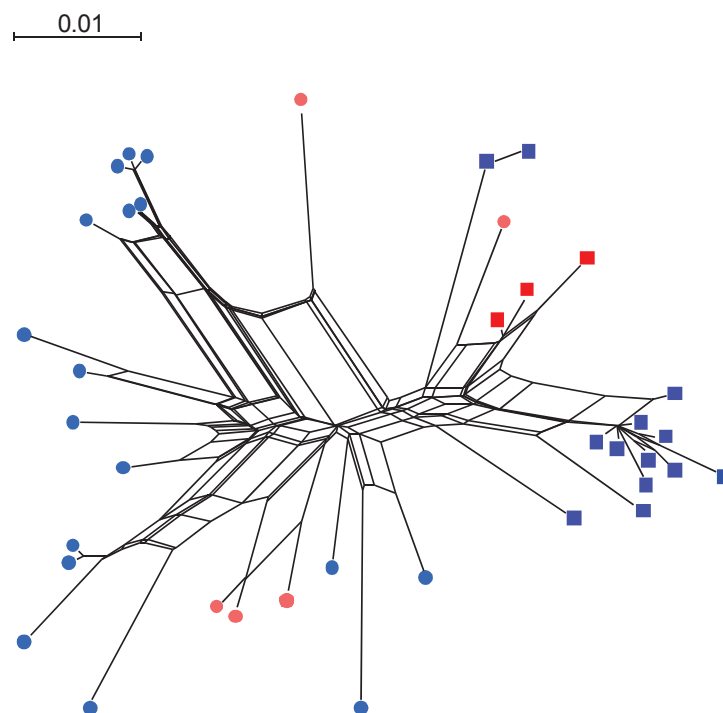


Figure II.4 : Neighbor network for the *MGD2* portion studied. Reticulation of the branches indicates uncertainties due to homoplasia (recombination or multiple hits). Sequences amplified with reverse primer 3231in-R are represented by circles and sequences amplified with the reverse primer 3231trans-R are represented by squares. Sequences with the L allele at *L31R* are in blue and sequences with the R allele at *L31R* are in red.

Table II.3 : Sample sizes, population differentiation, molecular diversities, linkage disequilibrium and Tajima's D for the *MGD2* locus in *M. edulis* samples.

Samples	n	F_{ST}	f_R	S	θ_W	θ_π	R_m	ZnS	D
South (LU, CA)	19	-0.01	0.05	84	0.043	0.038	4	0.119*	-0.55
North (WS, KEN)	20	0.005	0.4	114	0.057	0.043	8	0.075***	-1.22
All <i>M. edulis</i> samples	39	0.005	0.23	141	0.06	0.041	10	0.050***	-1.2

n , sample size; F_{ST} , estimate of F_{ST} (Hudson et al. 1992); f_R , frequency of the R allele at L31R in the sample of DNA sequences; S , number of polymorphic sites; θ_W , nucleotide diversity estimated from the number of polymorphic sites (WATTERSON 1975); θ_π nucleotide diversity estimated from the average number of pairwise differences (NEI 1987); R_m , estimation of the minimum number of recombination (HUDSON and KAPLAN 1985); ZnS , linkage disequilibrium (Kelly 1997); D , TAJIMA's D (TAJIMA 1989); * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Allele frequencies and linkage disequilibria at closely linked polymorphisms flanking L31R

From the DNA sequences analysed in this study and those of Boon *et al.* (2009) we defined and analysed five polymorphisms along the *MGD2* gene: *L31R*, two on its 5' side (left), in exon 2 and intron 2 and two on its 3' side (right) in intron 3 (Figure II.2). 224 *M. edulis* and 96 *M. galloprovincialis* mussels have been genotyped. The five loci did not depart from Hardy-Weinberg equilibrium in any population. Allele frequencies at these five polymorphisms are given as pie charts in Figure II.1 and global F_{ST} values among *M. edulis* populations in Figure II.2. The surprising homogeneity of allele frequencies between the two species has already been described in Boon *et al.* (2009) at the indel of exon 2 and with DNA sequences, and is confirmed here with the other four polymorphisms analysed here, including *L31R* and polymorphisms of intron 3. The increase in the R allele frequency at *L31R* in *M. edulis* is visible between southern populations (LU and CA) where the frequency is $p_R=0.1$ and northern populations (WS, CLEY and KEN) where the R allele frequency is $p_R=0.4$. Samples were grouped accordingly. Figure II.5 shows pairwise linkage disequilibria in the two groups of samples. D was arbitrarily defined to be positive when significantly different from zero. A significant D was found between the two polymorphisms on the 5' side (left) of *L31R* and between the two polymorphisms on the 3' side (right) of *L31R* in the two groups. An absence of significant D was observed between *L31R* and any of the four neighbour polymorphisms in the two groups. Finally the four pairwise disequilibria between one polymorphism on the 5' side (left) and one polymorphism on the 3' side (right) were found significantly different from zero in the southern group while they were low and not

significantly different from zero in the northern group (Figure II.5). The four third-order LD involving *L31R* and two polymorphisms at each side of *L31R* were computed and found to be low in the southern group ($D_{1,3,4}=0.013$, $D_{1,3,5}=0.010$, $D_{2,3,4}=0.000$, $D_{2,3,5}=0.000$), where the R allele frequency is low, and high in the northern group ($D_{1,3,4}=0.025$, $D_{1,3,5}=0.031$, $D_{2,3,4}=0.019$, $D_{2,3,5}=0.057$), where the R allele frequency is more balanced. As third-order LD are not easily understandable we present the results as estimated average pairwise disequilibria between one polymorphism of the 5' side (left) and one polymorphism of the 3' side (right) of *L31R* in the population of chromosomes carrying the L allele and in the population of chromosomes carrying the R allele (Figure II.6). Figure II.6 shows that once sorted according to their background at *L31R* pairwise disequilibria are very similar in the two groups of populations, positive in the Leucine background and negative in the Arginine background. As the R allele is rare in the southern group ($p_R=0.1$), the net D is positive and the third-order LD is low. In the northern group, however, the R allele frequency is balanced ($p_R=0.4$) which results in a net pairwise D close to zero, but this is due to a balanced mixture of L-carrying chromosomes with a positive D and R-carrying chromosomes with a negative D, and hence a strong third-order LD.

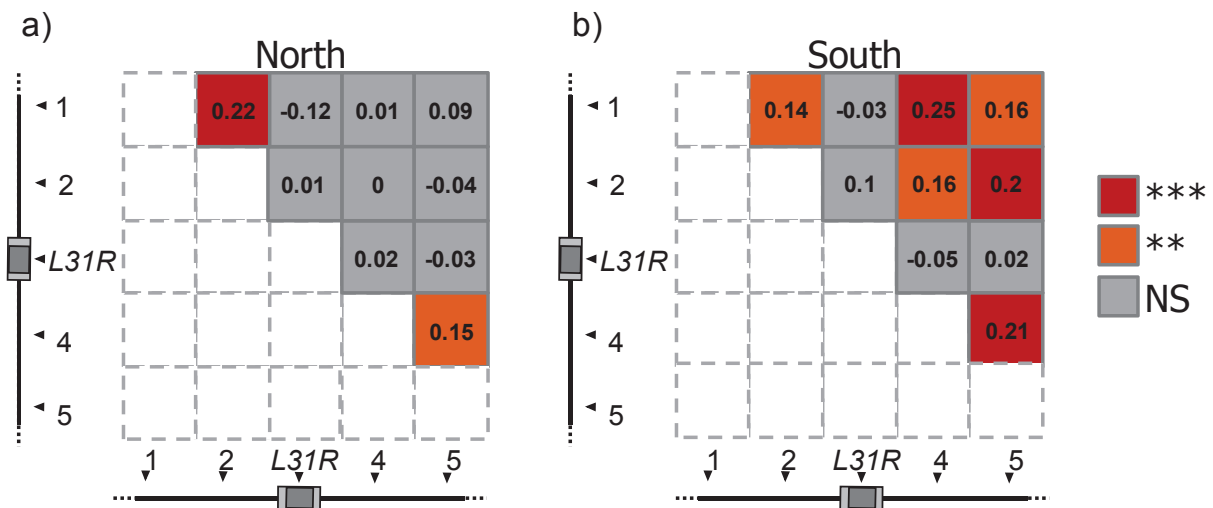


Figure II.5 : Pairwise linkage disequilibria. (a) Linkage disequilibrium in the northern group of samples (WS, CLEY, KEN) with a moderate frequency of the R allele at *L31R* ($p_R \sim 0.4$), (b) linkage disequilibrium in the southern group of samples (LU, CA) with a low frequency of the R allele at *L31R* ($p_R \sim 0.1$). D is polarised to be positive when significant. Grey: negative D values; black: positive non-significant D values; orange: significant positive D values with $0.01 < p < 0.05$; red: significant positive D values with $p < 0.01$.

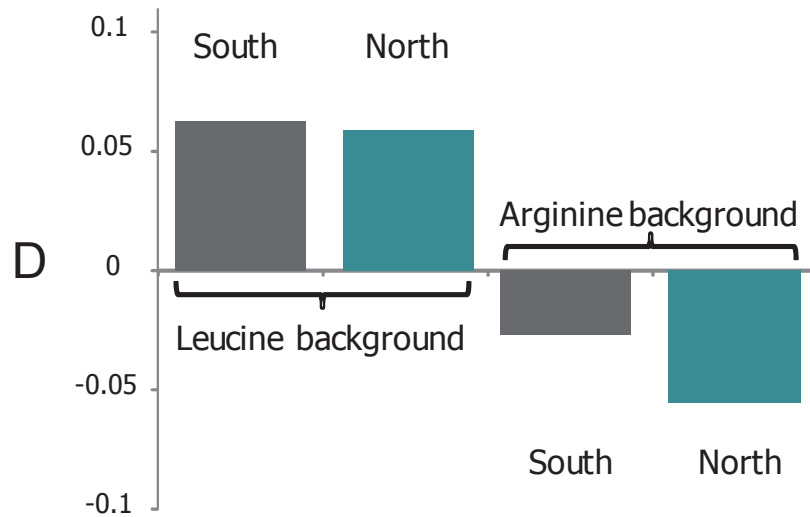


Figure II.6 : Pairwise linkage disequilibria according to the background allele at *L31R*. The average of the four pairwise disequilibria between one polymorphism on the 5' side (left) and one polymorphism on the 3' side (right) of *L31R* in the population of chromosomes carrying the L allele (left) and in the population of chromosomes carrying the R allele (right) in the southern group (light blue) and the northern group (dark blue) of *M. edulis* samples. D is polarised as in Figure II.5.

A close inspection of allele frequencies and linkage disequilibria thus suggest that a simple change in allele frequency at *L31R* without any recombination is sufficient to explain (i) the absence of allele frequency change at closely linked loci because linkage equilibrium is observed between *L31R* and any of the four neighbour polymorphisms, and (ii) the modification of pairwise D at loci localised each side of *L31R* between the southern and the northern groups of populations, because they are in antagonist phase between L-carrying and R-carrying chromosomes.

Discussion

We studied the nature of selection that has affected a Leucine/Arginine polymorphism (*L31R*) in a mussel defensin. This polymorphism revealed an outlying level of genetic differentiation between *Mytilus edulis* populations in Europe (Figure 3). Surprisingly, no differentiation was observed at closely linked loci each side of *L31R*. On the one hand this result allowed us to refute an indirect hitchhiking effect on *L31R* and to conclude selection affects it directly. On the other hand, this calls for an explanation to why the differentiation of linked polymorphism is not affected by selection and to how selection resulted in the preservation of the polymorphism in every population although with different allele frequencies.

A subtle effect of selection on the neighbourhood of L31R: evidence for selection on a pre-existing polymorphism

We can propose three hypotheses to explain the lack of differentiation in the direct neighbourhood of *L31R*: (i) hitchhiking has occurred but migration and recombination subsequently erased the signal, and the barrier effect (Charlesworth *et al.* 1997) of *L31R* on its background is now negligible, (ii) selection was weak enough and recombination strong enough to prevent hitchhiking effect at a very small chromosomal scale or (iii) selection acted on a pre-existing polymorphism (*i.e.* standing variation) that was at linkage equilibrium with its background. Under hypothesis (i) hitchhiking would have had resulted in a sweep of the genetic diversity (Kaplan *et al.* 1989; Fay & Wu 2000) that could have been expected to produce a detectable reduction of diversity overall, even if swept and unswept lineages were redistributed in every population by migration and recombination (Bierne 2010). We observed a high level of nucleotide diversity, whatever the background at *L31R* (Figure 4), and no significant departure from the neutral mutation-drift equilibrium (Table 2). If hitchhiking occurred it is either old or did not extend beyond exon 3 –*i.e.* recombination was strong in the face of selection (hypothesis ii). In both cases, one should invoke a very high rate of recombination around exon 3 of MGD2, which at first sight was supported by identified breakpoints of recombination (Table 2), reticulated network (Figure 4), a lower level of linkage disequilibrium than expected under complete linkage, and linkage equilibrium between *L31R* and the neighbour loci analysed. However, significant linkage disequilibrium was observed between loci each side of *L31R* in the southern group of *M. edulis* populations (Figure 5b), and also in *M. galloprovincialis* populations which have

exactly the same genetic composition at MGD2 than the southern group of *M. edulis* populations. It is difficult to explain this observation if exon 3 was a recombination hotspot or it would need to be a hotspot in the north and a coldspot in the south. LD patterns at neighbour loci between *M. edulis* populations that are differentiated at *L31R* revealed a difference between these two groups of populations (Figure 5). Neighbour loci have therefore been affected by selection on *L31R* but in a complex and subtle manner –their diversity and differentiation has not been perturbed but their LD pattern has. The absence of differentiation at neighbour loci does not require recombination. If the adaptive change in allele frequency at *L31R* did not affect neighbour loci this is because there was linkage equilibrium between *L31R* and its close neighbourhood. This is compelling evidence that selection acted on a pre-existing polymorphism with different backgrounds of LD pattern between the two alleles. We do not need other explanation than a change in the R allele frequency between the northern and southern group of *M. edulis* populations to explain the results, no need for an unusually high and differential rate of recombination.

The mystery of the initial state

Although our results are best explain by selection on a pre-existing polymorphism, it remains difficult to understand how *L31R* initially came to be at linkage equilibrium with its neighbourhood while LD being maintained each side of *L31R*. Outlying level of differentiation and distorted LD patterns have been showed to be detectable signatures of selection from standing genetic variation (Pennings & Hermisson 2006; Sabeti *et al.* 2007; Chen *et al.* 2010), the so-called soft sweeps (Hermisson & Pennings 2005) that do not leave as strong valleys of diversity and distortion of the allele frequency spectrum as selection from new mutations (hard sweeps) could do (Pennings & Hermisson 2006). However, soft sweeps are usually expected to produce elevated level of linkage disequilibrium between the allele that increase in frequency and its closely linked background (Sabeti *et al.* 2002; Pennings & Hermisson 2006) while this was not the pattern we observed. Against any anticipation, we found *L31R* to be at linkage equilibrium with its background in every population (Figure 5). Although the R allele is most probably the derived allele, as the L allele is found in other bivalve species analysed to date (Gonzalez *et al.* 2007), we cannot know whether the R allele increased in frequency in the northern group of *M. edulis* populations or if the L allele increased in frequency in the southern group and in *M. galloprovincialis* populations, or if a shift occurred in both groups. The surprising homogeneity of allele frequency between *M.*

edulis populations of the English Channel and the Bay of Biscay and *M. galloprovincialis* populations might be an argument for the low frequency of the R allele to be the initial state and that the R allele increased in the northern group of *M. edulis* populations. However, this is a speculative and *ad hoc* explanation. One might well argue for the reverse, that selection is maintaining the R allele at a similar frequency in the two species if they face a similar pathosystem. The built up of the background LD might also be taken as an argument to orientate selection (Stephan *et al.* 2006), but a more parsimonious hypothesis to explain the observed LD in *Mytilus* is secondary introgression (Bierne *et al.* 2003; Boon *et al.* 2009). It is therefore preferable to simply acknowledge that we cannot infer the direction of selection, if direction there really is. Although we obtained compelling evidence that selection acted on a pre-existing polymorphism, we are left with the fact that we cannot easily understand how *L31R* has come to be at linkage equilibrium with its close background, the L background in strong positive LD and the R background in negative LD (Figure 5 and 6). One may invoke gene conversion (Hamblin & Di Rienzo 2000; Tennessen 2005; Pennings & Hermisson 2006), but it should here have worked at a very small scale in exon 3 for the conversion track to remain undetected. In any case the history of the MGD2 gene has been long and complex with a period of isolation in allopatric populations and secondary introgression (Boon *et al.* 2009). Theoretical predictions need to assume an initial state that is nearly always mutation-drift equilibrium while in true natural populations selection on a pre-existing polymorphism can operate on very different initial state. Frustrating as it might be, this initial state and/or the evolutionary trajectory that resulted in this state are probably no longer accessible to a reconstruction from genetic data analysis, and this is something we should have in mind when applying ideally-conceived test of selection. Our results nonetheless suggest that third-order LD could possibly be interesting for identifying some type of selection from standing variation.

Divergent, weakly positive, balanced or intermittent selection?

As we obtained evidence of direct selection on *L31R*, we can finally discuss why allele frequencies are far from fixation in every population ($p_R=0.1$ in the southern group and $p_R=0.4$ in the northern group) and very similar at a large spatial scale within each group (Figure 1). If divergent selection explains the differentiation at *L31R*, a migration-selection balance should be invoked to explain that the R allele did not fix in the northern group and the L allele did not fix in the southern group and in *M. galloprovincialis*. Under this

hypothesis, either the grain of the environmental variation that affects this polymorphism is coarse and selection should be very weak, or the grain is fine and there exist varying proportion of alternative habitats (pathosystems) between the two regions. It is difficult however to imagine how this could result in so similar allele frequencies between populations of the same group at such a large spatial scale. Furthermore, the barrier to gene flow between the two mussel species is strong and one could have expected a higher frequency of the L allele in *M. galloprovincialis* if this allele was positively selected in the south. Isolation has also been identified between *M. edulis* populations of the Bay of Biscay, enclosed in the middle of two patches of *M. galloprovincialis* populations in the mosaic hybrid zone (Figure 1), and populations of the English Channel and the North Sea (Bierne *et al.* 2003; Faure *et al.* 2008). The migration-selection balance hypothesis therefore does not easily account for the similarity observed between the population of Calais in the English Channel) and the population of Lupin in the Bay of Biscay. Similarly the hypothesis that one allele became weakly positively selected and is on its way to fixation (the global sweep hypothesis, Bierne 2010) would not predict similar allele frequencies between the two species, and *M. edulis* populations of the Bay of Biscay and the English Channel. We therefore hypothesise selection on *L3IR* must be either (i) balanced, maintaining the frequency of alleles at a given value in each group, or maybe (ii) intermittent, turned-on according to a new modification of the pathosystem and turned-off while new defence solutions evolve or/and because of a further modification of the pathosystem, resulting in sporadic fluctuations in allele frequency.

In the classical gene-for-gene relationship, balancing selection is due to the opposing effects of the advantage to resist a disease and the cost the resistant allele often confers in the absence of pathogens. However, balancing selection at the gene level can also be achieved by divergent selection at the phenotypic level (Le Corre & Kremer 2012). Antimicrobial peptides are involved in first line immune defences and display activity against a broad spectrum of microbes (Zaslloff 2002). However, genetic variation at antimicrobial peptide gene is associated with the bacterial load, and interacts epistatically with other compartments of the immune system, including intracellular signalling that link pathogen recognition with effector activity (Lazzaro *et al.* 2004). *L3IR* might well be a QTL involved in an immune phenotype that integrates genetic variation of many immunity genes. It is also possible that selection acts on a suite of antimicrobial peptides that cope with the local microbial community (Tennessen *et al.* 2009). Interestingly, Gueguen *et al.* (2009) have found evidence for a synergistic antimicrobial activity of a homologous defensin in oyster with another antimicrobial peptide. Finally, the coefficient of selection may vary in time, according to

epizootics, causing intermittent partial sweeps spaced by episodes of (nearly) neutral evolution. One such example could be the CCR5- Δ 32 non-functional allele that confers resistance to HIV infection in humans and that has been supposed to have increased from a frequency of 5% in south-eastern Europe to a frequency of 0.16 in northern Europe through smallpox epidemics (Galvani & Slatkin 2003; Novembre *et al.* 2005). It has also been hypothesised that some susceptibility alleles for autoimmune diseases might have been maintained in human populations due to past selective processes (Fumagalli *et al.* 2011).

Conclusion

We have found evidence that direct selection on a pre-existing polymorphism explains the outlying level of genetic differentiation observed at a Leucine/Arginine polymorphism in a mussel defensin. This polymorphism was not chosen entirely randomly as it has been identified as being potentially interesting from the data obtained in a previous study (Boon *et al.* 2009). This candidate locus has nonetheless been included in an F_{ST} scan as any other loci we had at disposition. We can therefore consider that the scenario of adaptation we have partially reconstructed could well be representative of how adaptation is likely to generate some outliers identified in F_{ST} scans. Indeed, genome scans in human have revealed an overrepresentation of immunity genes (Nielsen *et al.* 2007; Fumagalli *et al.* 2011). Contrary to the previous outlier locus studied in detail in *M. edulis* (Faure *et al.* 2008; Bierne 2010), that was a neutral polymorphism affected by the fixation of an unconditionally favourable mutation at a hypothetical locus of unknown function linked to the outlier, we here have a more concrete story to tell as we have studied an adaptive polymorphism in an immunity gene. In addition, for MGD2 selection coefficients are likely to vary in space and our results are in closer accordance to the usual interpretation of F_{ST} outlier loci. Without knowledge of the gene function, however, one would have probably tried to identify variation of the macro-environment (temperature, salinity, pH etc.) correlated to the genetic differentiation at *L3IR*, and would have found some (e.g. temperature). Finally, contrary to the usual view that divergent selection is maintaining the differentiation at outlier loci, our results suggest selection is not necessarily active on *L3IR* at present, but that once again we might in fact be witnessing the ghost of historical selection.

Acknowledgements.

The authors are grateful to the IFR119 “Montpellier Environnement Biodiversité” for providing access to their sequencing platforms. We are very much indebted to Patrice David and Pierre-Alexandre Gagnaire for constructive discussions. This work was funded by the Agence National de la Recherche (Hi-Flo project ANR-08-BLAN-0334) and the project Aquagenet (SUDOE, INTERREG IV B). This is article 2013-XXX of Institut des Sciences de l'Evolution de Montpellier.

CHAPITRE III: Le projet Long Island

Jusqu'à maintenant dans cette thèse, nous nous sommes intéressés à des échelles d'espace et de temps pour lesquelles nous pouvions facilement comprendre, au moins *a posteriori*, que la structure détectée puisse être la conséquence d'un contact entre génomes différenciés. Pour finir, nous voulions nous intéresser à une échelle spatio-temporelle correspondant réellement à un équilibre migration/sélection. Plutôt que de partir à l'aveugle sur une zone géographique nouvelle, nous avons plutôt choisi de poursuivre l'étude des populations de moules du détroit de Long Island aux Etats Unis qui avaient marqué la littérature de la génétique des populations au début des années 80 avec la découverte d'un cline de fréquence allélique à un locus allozymique, le locus de la leucine amniopeptidase (*LAP*, Koehn *et al.* 1980).

1- Scans génomiques entre habitats à petite échelle spatiale

L'étude de la sélection locale en environnement variable et de son rôle dans le maintien du polymorphisme et la spéciation sont des préoccupations majeures de la génétique évolutive (Williams 1966; Hedrick 1986). Notamment, lorsque l'échelle spatiale de la dispersion surpasse celle de l'hétérogénéité environnementale, les conditions du maintien d'un polymorphisme par le processus d'adaptation locale sont contraignantes car à chaque génération un flux important de migrants maladaptés provenant d'autres habitats s'oppose à la sélection des génotypes adaptés localement (Lenormand 2002). De ce fait, ce n'est pas surprenant si les exemples avérés d'adaptation restent rares, même à l'ère de la génomique (Hedrick 2006). Alors que les organismes marins dispersent à plus longue distance que les organismes terrestres il existe pourtant un certain nombre d'exemples d'une différenciation entre microhabitats à petite échelle spatiale en milieu marin, comme chez le loup *Dicentrarchus labrax* (mer/lagune : Lemaire *et al.* 2000), la balane *Semibalanus balanoides* (haut/bas de l'estran : Schmidt & Rand 2001), la daurade *Sparus aurata* (mer/lagune : Chaoui *et al.* 2012), l'anguille américaine *Anguilla rostrata* (Gagnaire *et al.* 2012), la morue *Gadus morhua* (côtier/pélagique : Sarvas & Fevolden 2005), le bigorneau *Littorina saxatilis* (haut/bas de l'estran : Butlin *et al.* 2008) pour qui le résultat est moins surprenant puisque cette espèce ne possède pas de phase larvaire, le fucus *Fucus spiralis* (haut/bas de l'estran : Billard *et al.* 2010) et enfin la moule *Mytilus edulis* (Koehn *et al.* 1976; Koehn *et al.* 1980),

un exemple qui a été largement repris dans les manuels sur l'Evolution (Gillespie 1991; Mitton 1997; David & Samadi 2000) pour illustrer l'équilibre migration/sélection.

1.1- Les associations génotype-environnement au locus *Mpi* des balanes

Les premières études de Schmidt et Rand (1999) sur les balanes (*Semibalanus balanoides*) ont montré une forte structuration entre microhabitats de la zone intertidale (zone de balancement des marées) pour le locus *Mpi* (mannose-6-phosphate isomérase) et aucune structuration pour la *Gpi* (glucose-6-phosphate isomérase) et l'ADN mitochondrial dans les populations du Maine (Etats-Unis d'Amérique). Par la suite, Schmidt et Rand (2001) ont montré que les fréquences génotypiques dans les microsites exposés à de faibles et forts niveaux d'ensoleillement à marée basse étaient relativement homogènes lorsque les individus de *S. balanoides* étaient échantillonnés juste après la phase de fixation, alors que les juvéniles subissant une période de mortalité post-fixation montraient une différenciation significative au locus *Mpi* entre microsites. Ces études représentent donc un bel exemple qui illustre dans une population reproductivement peu structurée, le rôle puissant de la sélection récurrente d'adaptations différentielles à des microhabitats. De plus, à cause du rôle fonctionnel supposé des allozymes, il a été proposé que la sélection agisse directement sur le locus *Mpi*.

Dans le golfe du St Laurent (Canada), une discontinuité génétique a été observée sur une courte distance entre les sites au nord et les sites au sud de l'estuaire de la rivière Miramichi, non seulement pour le locus de la *Mpi* mais aussi pour le locus de la *Gpi* (Holm & Bourget 1994; Veliz *et al.* 2004) et à deux des six locus microsatellites étudiés (Dufresne *et al.* 2002). Ainsi au niveau du golfe du St Laurent, la sélection semble agir à une échelle supérieure aux microhabitats, alors qu'aucune variation génétique n'avait été détectée à mésoéchelle dans le Maine (Schmidt *et al.* 2000). Les résultats obtenus dans cette région présentent également d'autres différences surprenantes avec ceux obtenus dans le Maine : l'association génotype-environnement avec la *Mpi* semble être inversée, puisque l'allèle supposé favorable dans les microsites chauds du Maine est plus fréquent au nord, où les eaux sont plus froides, qu'au sud de la rivière Miramichi. De plus, le locus *Gpi* montre également un décalage des fréquences géniques d'amplitude similaire à celui de la *Mpi* alors qu'aucune différence n'a été observée entre les microsites du Maine à ce locus. De plus, les variations des fréquences alléliques au locus *Mpi* sont identiques entre le nord et le sud de la rivière Miramichi et les deux micro-habitats du Maine (changement de fréquence de l'allèle *S* de 30% à 50%). Une autre étude menée par Rand *et al.* (2002) dans le Rhode Island (Etats-Unis

d'Amérique) est venue alimenter l'incertitude sur la validité d'une simple hypothèse de sélection directe sur la *Mpi*. A Rhode Island, une variation inverse et non significative des fréquences génotypiques de *Mpi* a été observée entre microhabitats (Rand *et al.* 2002). Par contre la *Gpi* a montré une variation des fréquences génotypiques entre microhabitats. Ainsi, Bierne *et al.* (2011) ont proposé que la *Mpi*, la *Gpi* et les deux microsatellites dévoilant de la structure de part et d'autre de l'estuaire de la Miramichi seraient plus vraisemblablement des marqueurs de deux pools génétiques différenciés et partiellement isolés, non seulement par l'adaptation locale mais aussi probablement par des mécanismes d'isolement reproductifs endogènes (indépendant de l'environnement) pré- et/ou post-zygotique.

1.2- Une distribution des génomes à petite échelle entre *M. edulis* et *M. trossulus*

Un exemple probable de la façon dont l'interaction entre la sélection locale et des mécanismes d'isolement reproductifs endogènes performants peut se faire, est celui d'un modèle présentant beaucoup de parallèles avec les balanes: deux zones hybrides entre *Mytilus edulis* et *M. trossulus* en Europe et en Amérique. En Europe, *M. edulis* et *M. trossulus* forment des zones hybrides à l'entrée de la Mer Baltique et dans certains fjords Norvégiens (Riginos & Cunningham 2005) et *M. trossulus* occupe la Baltique et les fonds des fjords, là où la salinité est faible. En Nouvelle Ecosse *M. edulis* et *M. trossulus* forment une zone hybride mosaïque dans laquelle les *M. edulis* occupent les estuaires et *M. trossulus* les côtes océaniques. Pendant longtemps on a supposé que l'adaptation locale était le mécanisme principal du maintien de ces zones hybrides mais des études récentes ont révélé des incompatibilités gamétiques et une dépression d'hybridation entre les deux espèces (Miranda *et al.* 2010). Bierne *et al.* (2011) ont suggéré que la barrière endogène était la barrière principale et qu'elle se serait couplée avec un polymorphisme d'adaptation locale de façon différente lors des deux contacts secondaires en Europe et en Amérique. Une structure mosaïque à fine échelle spatiale, entre sites battus par les vagues et sites abrités, a aussi été observée et étudiée dans la zone hybride Européenne entre *M. edulis* et *M. galloprovincialis* (Bierne *et al.* 2002b; Bierne *et al.* 2003c). En comparant la composition des nuages larvaires et des jeunes recrues pendant la saison de reproduction Bierne *et al.* (2003b) ont pu montrer que cette structure était la conséquence d'un choix d'habitat des larves lors de leur fixation. En dehors des zones hybrides, les moules occupent l'ensemble des microsites (océan/estuaire, battu/abrité), c'est seulement dans les zones de contacts, où le couplage entre backgrounds endogènes et adaptation locale s'opère que la structure est observée sur les marqueurs neutres

(Bierne *et al.* 2011). L'exemple des moules, et probablement aussi l'exemple des balanes, montre que même à très petite échelle spatiale, où la migration distribue les génotypes dans tous les micro-habitats à chaque génération, l'hypothèse d'une interaction entre deux génomes différenciés et coadaptés est une hypothèse sérieuse à prendre en compte, surtout lorsque les locus sous sélection sont détectés assez facilement en analysant un nombre restreint de locus (Bierne *et al.* 2011).

1.3- Adaptation au gradient environnemental latitudinal dans une population panmictique d'anguilles

Pour finir sur les exemples avant de présenter notre étude sur les populations de moules du détroit de Long Island, il est intéressant de présenter les résultats obtenus dans le cas d'une adaptation locale probablement « pure », c'est à dire non couplée avec une zone hybride. L'anguille américaine *Anguilla rostrata* est un poisson migrateur amphihalin qui peuple les hydrosystèmes continentaux des côtes du nord-est Américain et se reproduit sur une aire de ponte unique dans la mer des Sargasses. Il a été montré à plusieurs reprises que l'ensemble des individus de l'espèce constituait une population panmictique (Côté *et al.* in press). Si une structure est observée à un locus entre les populations de rivières présentant des contrastes environnementaux, elle ne semble pouvoir n'être expliquée que par une sélection locale récurrente à chaque génération. Gagnaire *et al.* (2012) ont entrepris de détecter de tels locus en menant un scan génomique à haute densité. En partant d'une analyse (pré-scan) de 13 000 SNPs entre une population du sud et une population du nord de l'aire de distribution, Gagnaire *et al.* (2012) ont ensuite analysé 57 SNPs candidats sur la base du pré-scan sur un ensemble de 16 populations le long du gradient latitudinal. Cette seconde analyse a permis d'identifier 8 SNPs seulement sur les 57 candidats qui présentaient une structure corrélée au gradient environnemental, et parmi ces 8 locus 2 seulement se sont révélés être stables alors que les 6 autres sont probablement des polymorphismes transitoires pour lesquels un des deux allèles devrait ultimement fixer (Gagnaire *et al.* 2012). Cette étude constitue non seulement un très bel exemple d'étude du maintien du polymorphisme en environnement variable mais pourrait aussi être considérée comme illustrant la difficulté avec laquelle les locus d'adaptation locale devraient être identifiés dans un contexte d'adaptation locale pure. Pourtant, Williams *et al.* (1973) avaient identifié pas moins de quatre locus allozymiques (*MDH*, *SDH*, *PHI* et *ADH*) présentant une structure génétique significative entre rivières parmi les cinq locus qu'ils avaient étudié. Le locus *MDH* fait partie des 8 locus identifiés être

sous sélection et des deux locus inférés être à l'équilibre, dans l'étude de Gagnaire *et al.* (2012). Ces résultats sont très troublants mais suggèrent que la facilité avec laquelle des locus allozymiques ont été identifiés comme répondant à la sélection locale ne reflète pas forcément la vraie proportion de locus sous sélection dans un génome. L'argument selon lequel les locus allozymiques sous sélection ont été trouvés trop facilement chez les espèces marines (Bierne *et al.* 2003a) n'est donc pas suffisant pour soutenir et généraliser l'hypothèse du couplage à l'ensemble des exemples. Reprendre l'étude des populations de moules de Long Island semblait donc indispensable d'autant plus que certaines incohérences, comme nous allons le voir tout de suite, avaient été négligées dans l'interprétation du cline de la *LAP*.

2- Le cline de la Lap dans le détroit de Long Island

2.1- Revue de la littérature existante

Chez *M. edulis*, le cline de l'aminopeptidase dans le détroit de Long Island a été très étudié dans les années 80. Koehn et ses collègues ont étudié le rôle de cette enzyme initialement baptisée *LAP* (pour leucine amino-peptidase) dans l'osmorégulation des moules. Les moules utiliseraient cette amino-peptidase pour moduler leur pression osmotique en ajustant la concentration en acides aminés libres dans la cellule. Le polymorphisme électrophorétique a révélé plusieurs allèles chez *Mytilus edulis* dont trois assez fréquents, *LAP*⁹⁴, *LAP*⁹⁶ et *LAP*⁹⁸. L'allèle *LAP*⁹⁴ présente une constante catalytique légèrement plus forte que les deux autres allèles. Dans le détroit de Long Island près de New York aux Etats Unis, la fréquence de l'allèle *LAP*⁹⁴ passe de 55% dans les populations océaniques à 15% dans le fond du détroit où la salinité est faible (Koehn *et al.* 1980) alors qu'aucune structure n'avait été observée pour d'autres locus entre les populations océaniques et celles du détroit. De plus, il a été montré une mortalité plus forte des génotypes homozygotes pour l'allèle *LAP*⁹⁴ dans des conditions de faibles salinités, que ce soit dans la nature et au laboratoire (Hilbish & Koehn 1985). Ces études semblent ainsi réunir des preuves suffisantes pour démontrer que la sélection locale affecte le locus *LAP* directement.

2.2- Les zones d'ombre

Pourtant, plusieurs questions mériteraient d'être soulevées. La première concerne la forme du cline. Comment expliquer que la fréquence de *LAP*⁹⁴ ne chute pas à zéro mais se stabilise à

une fréquence de 15% dans le fond du détroit ? Une explication pourrait être qu'un influx de migrants maladaptés maintient la fréquence à un équilibre entre la migration et la sélection (Lenormand *et al.* 1999; Lenormand 2002). Cependant, la forme du cline n'est pas du tout en accord avec cette hypothèse car un long plateau à une fréquence stable de 15% est observé. D'ailleurs tout comme cet allèle n'est pas complètement éliminé dans le détroit, il n'atteint pas non plus la fixation dans les populations océaniques mais une fréquence de 55%, et ceci sur des distances de plusieurs centaines de kilomètres (Koehn *et al.* 1984), alors qu'aucun avantage n'a été trouvé à LAP^{94} dans les conditions océaniques. Si LAP^{94} est délétère à basse salinité et neutre à salinité océanique comme suggéré par Koehn et ses collègues, le détroit de Long Island ainsi que toutes les zones sous influence estuarienne devraient agir comme des puits pour cet allèle et l'éliminer progressivement de l'ensemble de la population. Il aurait donc fallu des conditions particulières pour expliquer que cet allèle ait atteint une fréquence aussi forte. Enfin, à environ 200 kilomètres plus au nord, au niveau du Cap Cod dans le Massachusetts, une chute de fréquence de l'allèle LAP^{94} est observée en l'absence de cline de salinité. L'écart de fréquence est exactement le même, de 55% au sud de Cap Cod à 15% au nord et dévoile un shift génétique concordant au locus *GPI* (Koehn *et al.* 1976). Enfin, le même saut de fréquence est observé plus au nord, au niveau du Cap Cod. Plus au nord encore, à partir de la Nouvelle Ecosse commence la zone d'hybridation mosaïque entre *M. edulis* et *M. trossulus* (Koehn *et al.* 1984; Bates & Innes 1995). *M. trossulus* est une espèce réputée plus tolérante aux salinités faibles que *M. edulis* (Qiu *et al.* 2002). L'allèle LAP^{94} a une fréquence de 40-60% chez les *M. trossulus* de l'Atlantique Ouest (Groupe III, Varvio *et al.* 1988), tout comme chez les *M. trossulus* du Pacifique (McDonald & Koehn 1988; McDonald *et al.* 1990). En Europe, LAP^{94} a une fréquence de 10% dans les populations *M. edulis* de la Mer du Nord (Bulnheim & Gosling 1988; Varvio *et al.* 1988; McDonald *et al.* 1990) et de moins de 20% dans les populations *M. trossulus* de la Mer Baltique (McDonald & Koehn 1988; Väinölä & Hvilsum 1991; Wenne & Skibinski 1995), une mer à la salinité de 10‰ seulement. En regardant plus en détail la forme du cline et les fréquences alléliques de diverses populations ou espèces, il apparaît donc très difficile d'expliquer le cline LAP^{94} par de la sélection locale directe. De plus, les associations génotypes/mortalités ne sont pas la preuve d'une sélection directe, et de légères différences d'activités entre allèles allozymiques est un résultat courant qui n'est pas une preuve de différence de valeur sélective (Hartl *et al.* 1985).

Il y aurait donc des conditions particulières au maintien du cline. Ces constatations soulèvent la question de l'origine du patron de polymorphisme observé. Est-ce bien la

sélection locale sur LAP^{94} qui en est à l'origine ? Autrement dit, la sélection agit-elle directement sur LAP^{94} ou y a-t-il un processus plus complexe à l'œuvre ?

Afin de mieux comprendre les bases génétiques de ce cline, nous avons utilisé deux stratégies. La première consiste à étudier un grand nombre de marqueurs afin de détecter d'autres locus anormalement structurés et observer la variation des fréquences le long du transect de Long Island pour pouvoir mettre en évidence des clines coïncidents avec celui de la LAP . La seconde stratégie se concentre sur l'étude du polymorphisme de la LAP pour tenter de relier le polymorphisme allozymique mis en évidence dans les études précédentes à un polymorphisme nucléotidique permettant d'identifier le processus à l'œuvre dans le patron de différenciation observé.

3- Structuration génétique dans le détroit de Long Island (Stratégie 1)

Cette stratégie repose sur l'hypothèse que si les populations du détroit et de l'océan représentent deux pools géniques partiellement isolés, nous nous attendons à trouver d'autres locus différenciés. Dans cette optique, nous avons réalisé un scan génomique AFLP pour la recherche d'outliers.

3.1- Matériels et méthodes

3.1.1- Echantillonnage

Un transect de 12 points d'échantillonnage de 48 individus a été réalisé le long du détroit de Long Island, sur les côtes du Connecticut les 19 et 20 septembre 2009 (Table III.1). Deux échantillons supplémentaires ont été récoltés plus au nord. Un échantillon de Boston permet de disposer d'une référence *M. edulis* du nord du cap Cod ; celui de l'embouchure du Saint-Laurent permet de disposer d'une référence *M. trossulus* pour mesurer une éventuelle introgression dans la zone.

Table III.1 : Détail des points d'échantillonnage sur le transect de Long Island.

Population	Lieu	Commentaire
1	Southport Marina N41°06'26.0'' W73°21'17.2''	Grande marina à proximité de la plage de Southport. Faune fixée mature, grosses ascidies, grosses huîtres et moules.
2	Southport Harbor N41°07'45.0'' W73°17'09.2''	Petites moules sous les pontons ; beaucoup de balanes.
3	Milford Yacht Club N41°12'39.3'' W73°03'05.5''	Moules sur le ponton du Yacht Club et à 30m parmi beaucoup de gros individus
4	Branford Harbor N41°15'38.0'' W72°49'14.6''	Pointe rocheuse. Moules dans les fentes, sous les <i>Fucus</i> .
5	Clinton Hammock Road N41°15'30.1'' W72°31'05.4''	Moules sur les rochers, sous les <i>Fucus</i> .
6	Seaside Avenue Westbrook N41°16'27.7'' W72°27'45.1''	Grande plage de sable avec digue rocheuse. Moules dans les fentes sous les <i>Fucus</i> .
7	Shetucket trail (Old Saybrook Town) N41°16'37.9'' W72°24'23.9''	Moules au pied de la digue.
8	Hartlands Drive (Old Saybrook Town) N41°15'47.8'' W72°23'03.1''	Plage de galets proche de l'embouchure de la Connecticut River ; sable un peu vaseux. Moules très éparses.
9	Crescent Beach (Niantic) N41°18'54.9'' W72°12'08.8''	Petite avancée rocheuse au bout d'une petite plage. Petites moules parsemées dans les fissures.
10	Brewer Yacht Yard (Mystic River) N41°20'48.1'' W71°58'03.3''	Grande marina. Moules sous les pontons.
11	Quonochontaug N41°19'57.7'' W71°42'48.0''	Grande plage de sable ; Moules abondantes sur gros rochers de granit à marée basse.
12	Charlestown (Snug Harbor Marina) N41°23'12.9'' W71°31'09.0''	

3.1.2- Extraction de l'ADN et obtention des fragments AFLP

L'ADN total a été extrait des branchies à l'aide du kit DNeasy Tissue (Qiagen) selon les instructions du fabricant et la qualité d'extraction a été vérifiée sur gel d'agarose. Les concentrations d'ADN ont ensuite été mesurées pour chaque échantillon en utilisant un spectrophotomètre NanoDrop8000 (Thermo Scientific), et standardisées à une concentration de 100ng.µl⁻¹. Toutes les réactions AFLP ont été réalisées dans des plaques 96 puits, contenant 4 différents sites d'échantillonnage et d'au moins 30% de répliqués inter-plaques

afin de mesurer la répétabilité des expérimentations et d'effectuer une analyse du taux d'erreur.

La procédure AFLP a été suivie selon le protocole original de Vos *et al.* (1995) qui utilise la combinaison d'enzymes EcoRI et MseI. Le détail de la mise au point des expérimentations de biologie moléculaire est donné en Annexe 1. Finalement, les fragments obtenus ont été séparés et détectés en utilisant un séquenceur ABI Prism 3130XL (Applied Biosystems).

Le logiciel GeneMapper ® v4.5 (Applied Biosystems) a été utilisé pour lire les chromatogrammes obtenus. Les classes de longueur des fragments (« bin sets ») ont été créées par analyse automatique, puis révisées manuellement pour chaque combinaison d'amorces. Seuls les fragments identifiables sans ambiguïté ont été traduits dans une matrice de valeurs d'intensité (« RFU ») pour chaque échantillon. Cette matrice a ensuite été traduite en une matrice binaire de phénotypes de présence/absence des fragments par la procédure implémentée dans AFLPScore v1.4b (Whitlock *et al.* 2008). Cette procédure utilise des seuils (« Locus calling » et « phénotype calling ») appliqués sur les valeurs d'intensité des fragments des échantillons répétés pour détecter les locus montrant des taux d'erreur supérieurs aux seuils fixés. Les taux d'erreur ont été mesurés en utilisant la méthode de Hadfield *et al.* (2006), modifiée pour les données dominantes et implémentée dans le package MasterBayes dans R. Selon cette méthode, les taux d'erreur ϵ_1 et ϵ_2 déterminent la probabilité d'une fausse présence ou la fausse absence d'un fragment AFLP, respectivement. La matrice de données finale est alors construite pour minimiser les taux d'erreur de l'ensemble du jeu de données en excluant les locus montrant des taux d'erreur par paire ($\epsilon_1 - \epsilon_2$) trop élevés, et ciblant un taux d'erreur de mismatch (dissemblance des paires de profils AFLP des échantillons répliqués) inférieur à 5%.

3.1.3- Analyse de génétique des populations

Sur la base de ce jeu de données, la structure génétique globale des populations a été étudiée. Les fréquences alléliques à chaque locus AFLP ont été calculées avec AFLPsurv v1.0 (Vekemans 2002; Vekemans *et al.* 2002) en supposant l'équilibre de Hardy-Weinberg dans chaque population et en utilisant une méthode bayésienne avec une distribution *a priori* non uniforme de la fréquence des allèles (Zhivotovsky 1999).

3.1.4- Structure génétique

Une Analyse Factorielle des Correspondances (AFC, Benzécri 1982) a été réalisée sur le nombre de fragments AFLP présents pour chaque locus et dans chaque localité. Ce type d'analyse multivariée permet de résumer les variables factorielles en une mesure des différences génétiques entre les échantillons et de visualiser graphiquement la ressemblance entre les populations étudiées.

Le nombre de populations de *M. edulis*, leur échelle de répartition spatiale et leurs connections potentielles ont été évalués à l'aide du programme STRUCTURE v2.3.3 (Falush *et al.* 2007). L'algorithme MCMC (Markov Chain Monte Carlo) a été utilisé sous le modèle d'admixture avec les fréquences d'allèles corrélées pour 300000 générations après 100000 étapes de 'burn-in', sans connaissance a priori de l'origine géographique des individus. Les calculs ont été réalisés 10 fois indépendamment pour un nombre de populations choisi de $K = \{1,2,3,4,5\}$ afin de déterminer le K le plus probable par la méthode d'Evanno (2005). La poursuite des analyses dans STRUCTURE (Falush *et al.* 2007) a donc tenu compte du K le plus probable et plusieurs analyses ont été conduites selon différents sous-ensembles du jeu de données afin de déterminer l'origine des patrons d'introgession observés.

3.1.5- Identification des outliers

Pour identifier les locus candidats à la sélection, nous avons utilisé la méthode implémentée dans le logiciel BayeScan v2.01 (Foll & Gaggiotti 2008; Foll *et al.* 2010). Cette méthode est basée sur la comparaison de deux modèles en île dans lesquels les fréquences alléliques des sous-populations sont corrélées à un pool génétique de migrants commun dont ils diffèrent à des degrés divers. La différence de fréquence entre les fréquences alléliques du pool génétique commun et celles de chaque sous-population est mesurée par un F_{ST} spécifique à chaque sous-population. Par conséquent, cette méthode permet de tester des scénarios où la taille efficace et le taux de migration peuvent différer entre les sous-populations. La sélection est alors introduite par la décomposition des composantes du F_{ST} par une régression logistique : une composante du F_{ST} spécifique d'un locus et partagée par les populations (alpha), et une composante spécifique à la population et partagée par tous les locus (bêta). Ainsi, un locus est supposé s'écarter de la neutralité lorsque sa composante spécifique est nécessaire pour expliquer la tendance observée dans les données. Cette méthode conduit à considérer deux scénarios alternatifs pour chaque locus, comprenant ou non la composante alpha, et explorés par MCMC estimant les probabilités postérieures pour chaque scénario.

3.2- Résultats

3.2.1- Traitement des données AFLP

Sur les 385 marqueurs AFLP qui ont été enregistrés, 310 étaient polymorphes dans les populations étudiées. Le taux d'erreur "mismatch" moyen était de 4,37%. Un résumé du traitement des AFLPs est donné pour chaque couple d'amorces utilisé dans la table III.2.

Table III.2 : Résumé du traitement des jeux de données avec AFLPScore (Whitlock *et al.* 2008). Conditions appliquées pour chaque jeu de données issu des 4 couples d'amorces utilisées. Le nombre de locus initial et final se réfère au nombre initial de classes de tailles de fragments (bins) détectés et au nombre final de locus retenus après l'analyse des taux d'erreur. Le taux de réplicats correspond au nombre de paires d'individus répliqués pour estimer les taux d'erreur. Le nombre total de locus retenus est de 385 pour lesquels les taux d'erreur moyens sont $\epsilon_1=15,87$, $\epsilon_2=0,09$, mismatch=4,37.

Couple d'amorces	Seuils		Taux d'erreur (%)			Nombre de locus		Taux de réplicats
	"Locus calling"	"Phenotype calling"	ϵ_1	ϵ_2	Mismatch	Initial	Final	
Eco_ACG-Mse_CGA	170	41	16,704	0,084	4,952	101	91	27,5
Eco_ACG-Mse_CTC	290	30	16,901	0,118	4,787	131	61	29,8
Eco_CAG-Mse_CGA	340	40	15,169	0,063	4,455	127	91	25,5
Eco_CAG-Mse_CTC	230	45	15,344	0,115	3,767	152	142	27,5

Abbreviation: AFLP, amplified fragment length polymorphism.

3.2.2- Structure génétique

L'analyse de la structure génétique des populations *M. edulis* de la côte américaine (détroit de Long Island et Boston) a montré une différenciation génétique globale significative et plutôt forte compte tenu de l'aire géographique étudiée ($F_{ST} = 0,02$). L'étude de la structuration génétique a d'abord été conduite par une AFC qui a montré un regroupement sur l'axe 1 des populations européennes d'une part et américaines d'autre part. Cependant, les populations américaines 1 et 3, situées à l'extrême ouest du transect de Long Island, se détachent du pool américain sur ce même axe (Figure III.1) et dans une moindre mesure la population 2, elle aussi à l'extrême ouest, et les populations océaniques 11 et 12 de l'autre côté du transect. L'axe 2 de l'AFC sépare les deux populations européennes de Mer du Nord et du Golfe de Gascogne, une structure qui était déjà connue (Faure *et al.* 2008). Sur le même axe, les populations 1 et 3 de Long Island se détachent avec des coordonnées semblables à la population européenne de Mer du Nord. Enfin l'axe 3 sépare les populations du cœur du transect (4 à 10) des populations périphériques (1, 2, 11, 12 et Boston) à l'exception de la population 3 qui montre des coordonnées semblables aux populations du cœur du transect.

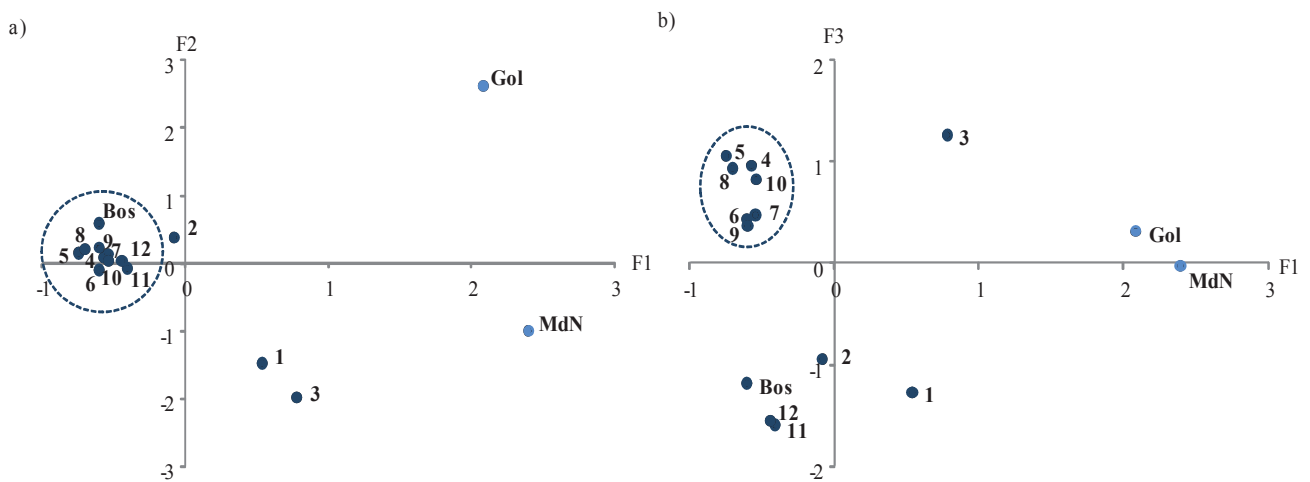


Figure III.1 : AFC sur les fréquences AFLP des populations de Long Island et Boston. Les deux graphiques représentent la répartition des populations en fonction des axes expliquant la plus grande part de variance observée. a) Représentation graphique des axes 1 (27,36%) et 2 (13,6%). b) Représentation graphique des axes 1 (27,36%) et 3 (10,35%). Les points numérotés de 1 à 12 représentent les populations du transect de Long Island depuis Southport jusqu'à Charlestown (Figure III.2). Bos représente la population de Boston; Gol et MdN représentent les populations européennes de *M. edulis* du Golfe de Gascogne et de Mer du Nord, respectivement.

La méthode d'Evanno (2005) a permis d'inférer l'existence de 2 clusters sur la base du jeu de données regroupant les populations européennes et américaines. La poursuite des analyses dans STRUCTURE (Falush *et al.* 2007) a permis de confirmer que les populations 1 et 3 montraient un patron génétique différent des autres populations du détroit, mais proche de celui des populations européennes (Annexe 2). La répartition des fréquences des probabilités d'assignation au cluster 1 pour chaque localité est indiquée dans la figure III.2.

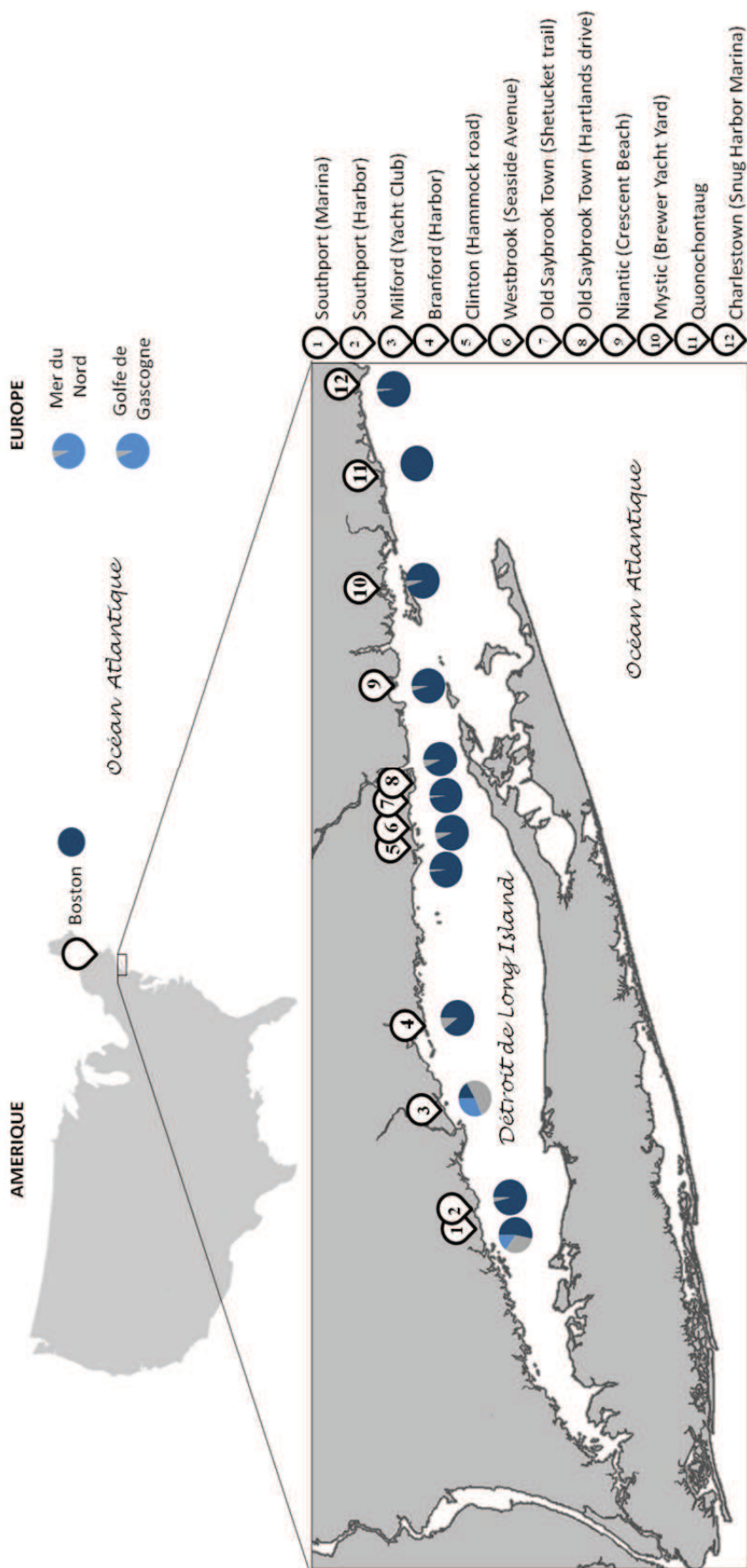


Figure III.2 : Transect de Long Island échantillonné entre le 19 et le 20 septembre 2009 (voir détails dans la table III.1). Les camemberts représentent les fréquences des probabilités d'assignation individuelles Q au cluster 1 pour chaque localité (bleu foncé : $Q > 0.9$; gris : $0.1 < Q < 0.9$; bleu clair : $Q < 0.1$). La localité Boston est indiquée en tant que population de référence.

3.2.3- Identification des outliers

Un scan génomique sur les 13 populations *M. edulis* a été réalisé par la méthode de Foll et Gaggiotti (2008) implémentée dans BAYESCAN et a permis d'identifier 9 locus outliers (3%) sur les 310 marqueurs AFLP polymorphes étudiés (Figure III.3). Ces 9 AFLPs (*1-175*, *3-363*, *3-157*, *1-222*, *2-142*, *3-164*, *2-209*, *1-189* et *2-226*) montrent une probabilité supérieure du seuil décisif de $\text{Log}(\text{PO}) = 2$.

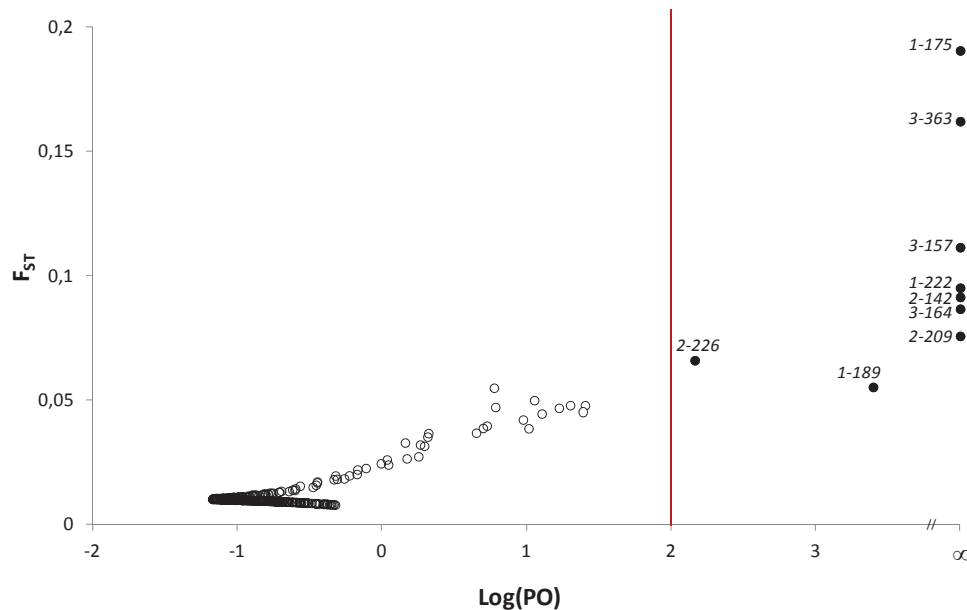


Figure III.3 : Scan génomique issu de l'analyse BAYESCAN de 310 marqueurs AFLP polymorphes sur les 13 échantillons étudiés (les 12 échantillons du transect de Long Island et l'échantillon de Boston) représentant le F_{ST} contre le logarithme de la probabilité bayésienne pour chacun des locus analysés. Le seuil de probabilité décisif est représenté par une ligne verticale rouge à $\text{Log}(\text{PO}) = 2$ et les locus identifiés comme outliers sont noircis et annotés.

Les fréquences alléliques de chacun de ces outliers sont présentées dans la Figure III.4 (voir aussi Annexe 3 pour les spectres de fréquence de l'ensemble des localités *M. edulis* étudiées) pour chaque localité en fonction de la distance à l'est de Southport. Pour les 9 locus représentés, deux patrons de variation géographique se distinguent (Figure III.4). Le patron 1 caractéristique de 3 locus (*3-363*, *2-209* et *1-189*) présente un shift brutal des fréquences alléliques au niveau du point 3 d'échantillonnage. Sur ces trois locus, les fréquences alléliques de la population 3 de Long Island sont très semblables à celles de la population de Mer du Nord en Europe. Le patron 2 montre une forme en U avec une augmentation des fréquences alléliques aux extrémités du transect de Long Island (locus *1-175*, *3-157*, *1-222*, *2-142*, *3-164*, et *2-226*). La coïncidence entre locus n'est pas aussi bonne que dans le patron 1.

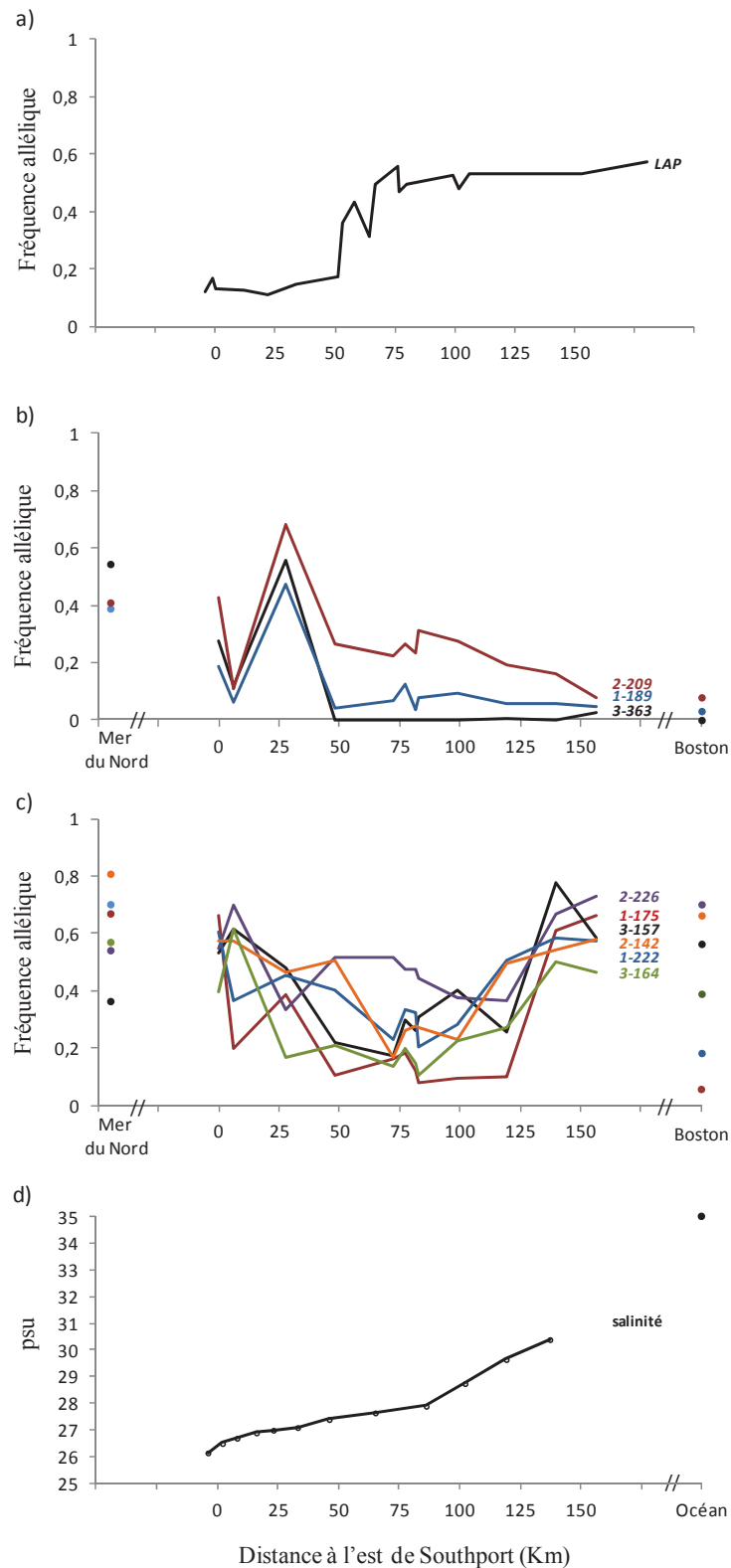


Figure III.4 : Distribution des spectres de fréquences alléliques a) de la *LAP* (d'après Koehn *et al.* 1976; Lassen & Turano 1978), b) des marqueurs AFLP identifiés comme outliers montrant une variation selon le patron 1 et c) selon le patron 2 le long du transect de Long Island et pour les populations de Boston et de Mer du Nord (points colorés).d) Variation de la salinité dans le détroit de Long Island (en unités de salinité pratique : psu ; d'après Gay & O'Donnell 2009).

Cependant, la structure n'est jamais franchement clinale, mais plutôt abrupte et jamais au niveau de la position historique du cline de la LAP. Tous ces locus ont tendance à montrer des fréquences similaires entre les populations périphériques du transect et la population européenne de Mer du Nord. Enfin deux locus (1-175 et 1-122) montrent des fréquences similaires entre les populations du cœur du transect et la population de Boston au nord de Cap Cod. Les analyses d'assignation par STRUCTURE ont été réitérées en éliminant les locus outliers et sont restées globalement inchangées, c'est-à-dire que les populations 1 et 3 du fond du détroit ont toujours un grand nombre d'individus assignés dans le cluster européen.

4- Tentative d'identification du gène de la LAP (Stratégie 2)

De longues expérimentations de protéomique ont été réalisées par électrophorèse bidimensionnelle au laboratoire Innova Proteomics de Rennes, et en parallèle par Jean Mary à Roscoff, dans le but d'identifier les protéines impliquées dans le polymorphisme de la LAP. Deux groupes d'individus de génotype connu ont été utilisés. D'une part, les homozygotes LAP^{94}/LAP^{94} (8 individus) et LAP^{98}/LAP^{98} (12 individus) ont permis au Laboratoire Innova Proteomics de mettre en évidence un spot protéique variable entre les deux échantillons (Figure III.5). D'autre part, Jean Mary a réalisé la même expérience avec les individus typés hétérozygotes LAP^{94}/LAP^{96} et LAP^{96}/LAP^{98} et a obtenu un patron semblable de variabilité. Les spots protéiques ainsi révélés ont été microséquencés et comparés à des banques de données EST. Parmi les meilleurs résultats d'homologie, la peptidase *MI2A* a retenu notre attention pour étudier le polymorphisme de séquence.

Nous avons donc défini des amorces à partir des séquences des bases de données EST et tenté d'amplifier les fragments par PCR. La séquence de la peptidase *MI2A* de la base de données de *Mytilus* nous a permis de mettre en évidence un polymorphisme nucléotidique intéressant qui semblait bien corrélé au génotype allozymique de l'allèle 98, même si de façon imparfaite. Après de multiples essais et le travail de 2 stagiaires, aucun résultat répétable n'a pu être obtenu. Il existe bien un polymorphisme non-synonyme responsable d'un changement de charge de l'enzyme et les fréquences des deux allèles sont équilibrées (30%, 70%) mais nous ne retrouvons pas la structure génétique décrite avec le typage allozymique : pas de structure à Long Island, pas de structure entre Amérique et Europe et mauvaises fréquences alléliques dans les espèces *M. galloprovincialis* et *M. trossulus*. Il semble donc que le gène étudié n'était pas le bon. Une autre peptidase, la peptidase *MI6*,

avait également donné un résultat d'homologie lors de l'interrogation des bases de données même si moins bon que la peptidase *MI2A* et une étude de son polymorphisme a donc été initiée dans un deuxième temps. Un polymorphisme non-synonyme responsable d'un changement de charge de l'enzyme a été identifié mais il ne corrèle pas très bien au génotype obtenu lors du typage allozymique. Une étude en population naturelle est néanmoins en cours sur ce gène. Cette stratégie de recherche du gène de la *LAP* par une approche de protéomique n'a donc pas donné les résultats escomptés.

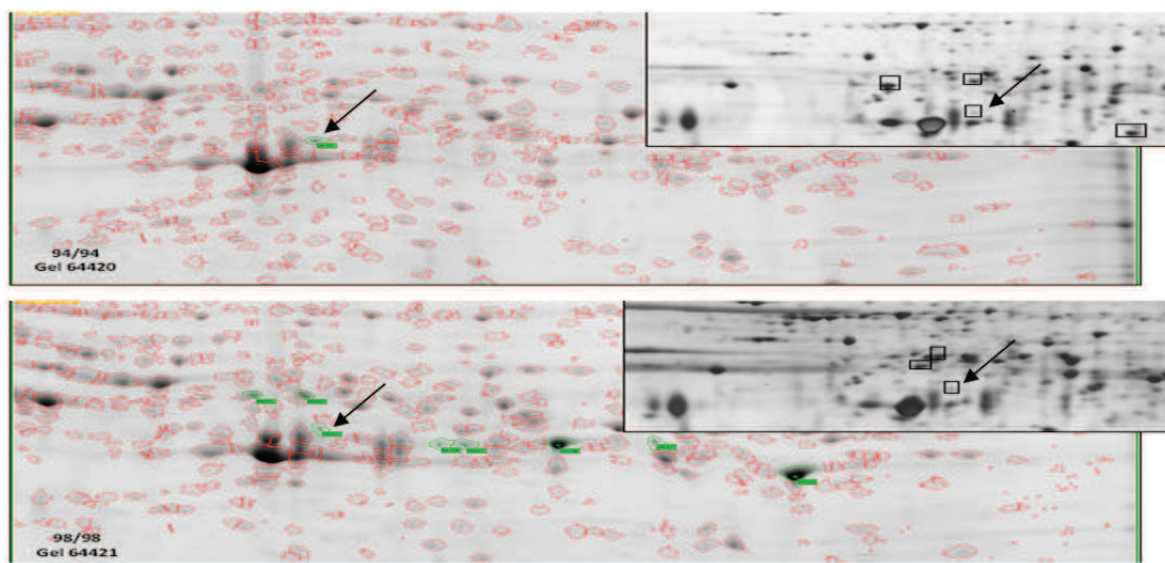


Figure III.5 : Gels obtenus à partir d'électrophorèses bidimensionnelles des protéines de deux groupes d'individus. En couleur, les homozygotes LAP^{94}/LAP^{94} (en haut) et LAP^{98}/LAP^{98} (en bas) ont été analysés au Laboratoire Innova Proteomics de Rennes ; en noir et blanc, les hétérozygotes LAP^{94}/LAP^{96} (en haut) et LAP^{96}/LAP^{98} (en bas) ont été analysés au laboratoire de Roscoff par Jean Mary. Les spots entourés en vert correspondent aux protéines différentiellement exprimées entre les deux échantillons analysés. Les flèches indiquent les spots retenus pour le microséquençage.

5- Discussion

Dans ce chapitre, nous avons étudié la structure génétique des moules à une échelle spatiale restreinte. Les *M. edulis* de Long Island nous ont permis d'étudier un cas d'école qui n'avait pas été repris depuis les publications des années 80 et qui montrait que la *LAP* était le seul locus à montrer un patron clinal parmi les 10 locus étudiés (Koehn *et al.* 1980). Pourtant, la conclusion de ces recherches n'était pas satisfaisante car de nombreux problèmes étaient restés sans réponse. L'étude de nombreux marqueurs dans le détroit de Long Island a permis de mettre en évidence des shifts génétiques sur une proportion appréciable du génome (3%). Cependant ces shifts génétiques ne se situent pas à la même position que celle à laquelle avait été observé le cline de la *LAP* il y a une trentaine d'année. De façon intéressante, les populations du fond du détroit présentent des similitudes avec les populations européennes, que ce soit sur les outliers ou l'ensemble des locus analysés. Les populations océaniques ressemblent aux populations du fond du détroit et aux européennes sur certains locus outliers seulement. Malheureusement, sans connaître la position actuelle du cline de la *LAP*, il va être difficile d'utiliser nos résultats pour interpréter les résultats passés, mais nous avons néanmoins obtenu (i) des shifts génétiques à très petite échelle spatiale et (ii) des preuves qu'un contact secondaire entre des *M. edulis* européennes et des *M. edulis* américaines pouvait expliquer ces shifts et donc la proportion importante de locus affectés.

5.1- Des *M. edulis* européennes dans le détroit de Long Island

Bien que la structure génétique ne soit pas très marquée dans les populations du détroit de Long Island, un F_{ST} de l'ordre de 2% a été mesuré. Cette structure semble davantage concerner les populations 1 et 3 à l'ouest du transect. En effet, l'AFC (Figure III.1) montre que ces 2 populations sont tirées par l'axe 1 en direction des populations européennes en s'écartant du groupe américain. L'analyse de « clustering » bayésien dans STRUCTURE (Falush *et al.* 2007) conduit au même patron de structure globale, assignant les populations 1 et 3 au cluster européen (Annexe 2). En analysant la proportion des probabilités d'assignations individuelles au cluster américain dans les populations de Long Island, ces mêmes populations ont montré une part importante de mélange (32% pour la population 1 et 52% pour la population 3, Figure III.2). Ce détachement du pool génétique américain suggère une introduction dans le fond de Long Island sans savoir si ce phénomène doit être attribué

aux activités humaines récentes, quoique le trafic maritime soit très important dans cette zone géographique.

5.2- Des shifts génétiques jumeaux

L'identification de 9 locus outliers a permis de mettre en évidence des spectres de fréquences alléliques montrant deux patrons distincts (Figure III.4b et III.4c). Les 3 marqueurs exhibant le patron 1 (3-363, 2-209 et 1-189, Figure III.4b) semblent montrer le même patron que celui révélé par la structure génétique globale, à savoir une forte fréquence d'allèles de type européen dans les populations 1 et 3 du détroit de Long Island et formant un shift à l'est de la population 3. Les 6 autres marqueurs identifiés (1-175, 3-157, 1-222, 2-142, 3-164 et 2-226, Figure III.4c) montrent un patron en U avec des shifts plus ou moins marqués qui tendent à regrouper les populations océaniques (11 et 12) et les populations du fond du détroit (1 et 3 mais aussi souvent la population 2, comme c'est également visible sur l'axe 3 de l'AFC), localisées aux deux extrémités du transect et qui sont différenciées des populations centrales du détroit de Long Island..

Les shifts obtenus pour les trois locus du patron 1 s'expliquent bien par une hypothèse de contact secondaire entre un background européen introduit et un background américain indigène. Les fréquences alléliques de ces trois locus sont très similaires entre les populations du fond du détroit et la population de Mer du Nord. Les régions génomiques dans lesquelles se trouvent ces locus auraient résisté à l'introggression d'allèle américain alors que pour le reste du génome l'introggression aurait entamé l'homogénéisation entre les deux backgrounds. Par contre, l'interprétation de la structure observée aux locus outliers présentant le patron 2 est plus problématique. Parmi les locus identifiés, certains montrent un niveau de différenciation plus fort entre populations du détroit (populations 1-3 contre populations 11-12) qu'entre populations américaines et européenne (par exemple 2-209 et 3-157). Une possibilité est que le shift génétique de l'entrée du détroit (entre les populations 9 et 10) soit antérieur à l'introduction de *M. edulis* européennes dans le fond du détroit. Lors de l'introduction, les allèles océaniques des locus d'isolement auraient pu introgresser le background européen. Cette hypothèse suppose donc que les allèles océaniques soient défavorables dans le génome des populations du cœur du détroit plutôt que défavorable dans l'environnement du détroit. Ou alors il faudrait trouver une variable écologique similaire entre l'océan et le fond du détroit. Le shift de l'entrée du détroit serait donc indépendant de l'introduction des *M. edulis* européennes. Cependant, il y a tout de même une tendance à ce

que les populations océaniques ressemblent plus aux européennes qu'aux populations du cœur du détroit sur ces locus outliers. L'histoire de la différenciation entre les génomes américains et européens ne serait donc pas étrangère à l'existence de ce shift génétique, qui serait cependant plus ancien. Ceci pourrait également expliquer la gémellisation des shifts lorsque des moules européennes auraient été introduites par l'homme plus récemment.

Si l'on s'intéresse maintenant à la localisation de ces shifts génétiques par rapport à celui mis en évidence pour la *LAP* dans les années 80, il ne semble y avoir aucune concordance. Dans notre étude, les clines se situent entre 0 et 50 Km et entre 120 et 150 Km alors que le cline de la *LAP* a été initialement rapporté entre 40 et 80 Km à l'est de Southport (Figure III.4). Pourtant, le cline *LAP* a pu se décaler pendant ces 30 dernières années et être aujourd'hui coïncident avec des clines AFLP mis à jour dans notre étude. Cette hypothèse suppose que la *LAP* soit un marqueur endogène d'un fond génétique en interaction avec des adaptations locales. Ce couplage génétique (Bierne *et al.* 2011) pourrait s'être déplacé au cours du temps par la modification des pressions environnementales dans le détroit de Long Island. Ainsi, pour des raisons qui restent à identifier comme par exemple les aléas de remise en contact avec les populations européennes dans le fond du détroit au cours des 30 dernières années ou la perte de certaines associations génétiques pour des raisons sélectives, le cline de la *LAP* pourrait avoir été déplacé vers d'autres clines environnementaux. Cependant, il n'a pas été possible d'obtenir la séquence de la *LAP* pendant cette thèse et il nous est donc impossible de savoir si ce cline a effectivement bougé. Il semble pourtant très vraisemblable que le cline de la *LAP* ait bougé puisque le cas de ce marqueur montre tant de paradoxes facilement explicables par des hypothèses alternatives soulevées ici. Dans cette optique, le cline de la *LAP* aurait donc bien bougé ce qui implique que la *LAP* ne puisse être sous une sélection locale affectée par la salinité puisque cette hypothèse supposerait que le patron de salinité des eaux du détroit de Long Island ait changé.

CONCLUSIONS ET PERSPECTIVES

Le but de cette thèse était d'appréhender le problème du paradoxe de l'adaptation marine en mettant à l'épreuve une des hypothèses implicites de l'interprétation des données à l'origine du paradoxe, qu'une structure génétique forte est la conséquence d'une sélection locale directe sur le locus étudié. Pour ce faire nous avons tenté d'étudier les bases génétiques et l'histoire de la différenciation adaptative chez une espèce marine à phase larvaire planctonique, les moules du complexe d'espèce *Mytilus edulis*. Une approche de génomique des populations nous a permis d'obtenir facilement des mesures de la différenciation génétique sur un grand nombre de locus AFLP et d'autres types de polymorphismes pour lesquels la poursuite de l'analyse est beaucoup plus facile lorsqu'ils s'avèrent être outliers. Tout au long de cette thèse, nous avons enquêté sur l'existence de polymorphismes anormalement structurés entre populations du complexe d'espèces *M. edulis* et avons mis en évidence des locus candidats à une analyse plus poussée. Nos recherches nous ont conduites à prendre en compte des hypothèses trop souvent négligées dans la littérature des scans génomiques, et de mettre en avant plusieurs hypothèses alternatives qui replacent la sélection locale dans un contexte historique plus complexe et dans des situations d'isolement reproductif plus poussé qu'habituellement proposé.

1- Synthèse des résultats

Dans le premier chapitre, l'étude du locus *EFbis* a démontré que la différenciation qu'il exprimait entre les populations atlantique et méditerranéenne de *M. galloprovincialis* était le résultat d'une introgression différentielle avec l'espèce sœur *M. edulis*. De façon inattendue, nous avons découvert que ce sont les populations méditerranéennes qui ont été introgressées alors que le contact actuel entre les deux entités génétiques se trouve au niveau de la population atlantique. Nous avons finalement pu mettre en évidence que le contact entre *M. edulis* et la population méditerranéenne de *M. galloprovincialis* était historique sans pour autant être en mesure de conclure sur la cause de cet événement. En effet, deux hypothèses permettent d'expliquer l'introgression d'un allèle étranger dans un génome. L'introgression peut être adaptative en supposant que l'allèle impliqué soit avantageux dans la population qu'il introgresse. La seconde hypothèse suppose que la barrière au flux génique puisse être érodée aux régions du génome considérées, impliquant que le fond génétique méditerranéen

soit plus perméable que l'atlantique aux allèles *edulis*. L'analyse approfondie du cas Atlantique/Méditerranée chez la moule fournit un exemple de l'importance d'analyser les polymorphismes outliers dans les espèces phylogénétiquement proches pour pouvoir mettre en évidence les polymorphismes introgressés responsables de structure intraspécifique.

Le second chapitre s'est concentré sur un cas d'adaptation au locus MGD2, pour tenter d'expliquer comment la sélection a influencé la structure génétique de ce locus. La mise en évidence de la sélection au locus L31R ne suffisait pas à comprendre l'origine d'une telle structure. Nous avons donc étudié en détail ce locus et montré que la sélection avait agi sur un polymorphisme pré-existant à l'équilibre de liaison avec son voisinage chromosomique. Cependant, l'étude des quatre locus voisins a mis en évidence des déséquilibres de liaison par paires entre ces locus dans certaines populations, rendant d'autant plus complexe l'interprétation des conditions sélectives au locus L31R. Même si nous n'avons pas été en mesure d'expliquer précisément l'histoire de ce locus, nous avons pu soulever la question de la nature de la sélection étant (ou ayant été) à l'œuvre.

La dernière partie de cette thèse s'est intéressée au cas des moules du détroit de Long Island à cause des nombreuses zones d'ombres dans les interprétations proposées dans les études précédentes. L'étude d'un grand nombre de marqueurs AFLP a permis de révéler des shifts génétiques sur une partie appréciable du génome (3%) montrant deux patrons distincts localisés à des positions différentes de celle observée pour le cline de la *LAP* il y a une trentaine d'années. Cette étude a mis en évidence l'existence de shifts génétiques à une très petite échelle spatiale et suggère l'introduction de moules européennes dans le fond du détroit.

2- Perspectives

2-1 AFLPs ou des marqueurs difficiles à étudier de façon approfondie

Le cas du locus 3-363 offre un bon exemple des limites qui peuvent entraver l'analyse détaillée de locus outliers issus de scans génomiques de marqueurs AFLP. Ce locus montre une différenciation anormalement forte dans de nombreuses comparaisons intraspécifiques (en Europe : *M. edulis* de Mer du Nord vs Golfe de Gascogne (données non montrées), *M. galloprovincialis* d'Atlantique vs Méditerranée (Figure I.3); et *M. edulis* en Amérique (entre les populations du fond du détroit de Long Island et les populations centrales et océaniques (Figure III.4b)). Plusieurs tentatives ont été tentées pour isoler une séquence de cette bande

AFLP afin d'obtenir une signature chromosomique et reconstituer l'histoire de ce marqueur, mais nous n'avons pas réussi à obtenir de résultat. Il serait alors très profitable de pouvoir étudier des marqueurs plus faciles à analyser de façon approfondie. L'ère de la génomique offrent désormais la possibilité d'étudier de grand nombre de marqueurs et les limitations méthodologiques semblent levées. Les NGS (pour Next Generation Sequencing) ne cessent de se perfectionner (Metzker 2010; Stapley *et al.* 2010), permettant l'étude de milliers de marqueurs SNPs sans connaissance préalable du génome (Hohenlohe *et al.* 2010; Davey *et al.* 2011).

Sans aller jusqu'à séquencer des génomes entiers, les marqueurs moléculaires « nouvelle génération » offrent de nombreuses possibilités pour étudier les signatures chromosomiques des marqueurs candidats à la sélection en populations naturelles. Une banque BAC (Bacterial Artificial Chromosome) a été scannée et permettra notamment d'obtenir la signature chromosomique autour du gène *EF1α* inféré à -3Kb en 5' du codon start soit -4.8 Kb en 5' du locus *EFbis* (Bierne 2010) initialement détecté comme outlier entre les populations *M. edulis* de Mer du Nord et du Golfe de Gascogne (Faure *et al.* 2008). A partir de cette séquence de grande taille (~100kb) il sera possible d'identifier le locus sélectionné et d'étudier la signature de la sélection sur une échelle chromosomique beaucoup plus large que les 5 Kb actuellement disponibles.

2-2 Améliorer la couverture d'échantillonnage pour une étude plus approfondie de MGD2

Afin de pouvoir mieux comprendre les bases génétiques et l'histoire potentiellement complexe des pressions sélectives et des aléas des barrières génétiques endogènes en jeu dans la différenciation du locus L31R et de son voisinage chromosomique, il serait intéressant d'augmenter l'échantillonnage pour être en mesure de cartographier finement les fréquences alléliques des différents locus analysés dans le cadre de l'étude de MGD2.

2-3 Des croisements pour mieux comprendre les origines de l'introgession

Nous avons vu que l'introgession d'allèles pouvait être source de variabilité entre populations d'une même espèce (exemple d'*EFbis* dans le fond génétique *galloprovincialis* méditerranéen, Chapitre I), que ce soit par l'introgession adaptative d'allèles étrangers dans une population ou par l'érosion des barrières au flux géniques lors de contacts secondaires passés ou plus récents (ex. introgession d'allèles américains dans un background européen

dans le fond du détroit de Long Island). L'étude de croisements intraspécifiques de populations choisies (*M. galloprovincialis* Atlantique X Méditerranée, et *M. edulis* Europe X Long Island) pourrait permettre de mieux appréhender l'histoire et les aléas de remise en contacts de marqueurs de fond génétiques peu différenciés.

ANNEXES

Annexe 1 : Protocole AFLP mis au point pour les espèces de moules du genre *Mytilus* (Extrait du document réalisé à la demande de l'IFR119 "Montpellier Environnement Biodiversité").

Digestion par 2 enzymes de restriction

Site de restriction EcoR1⁸⁰

Site de restriction MseI⁸⁰

Digestion	Ech			
2h30' à 37°C, puis 20' à 65°C	1	20	50	100
Buffer Enz (10X) → 1X	2	40	100	200
BSA (100X) → 1X	0.2	4	10	20
EcoR1 (20U/μl) → 0.1U/μl	0.1	2	5	10
MseI (10U/μl) → 0.1U/μl	0.2	4	10	20
H ₂ O	15.5	310	775	1550
ADN 100ng/μl	2			
Volume total	20			

Il est préférable de travailler avec un ADN standardisé pour éviter les trop grandes différences de concentration des ADN.

Ici la manip fonctionne bien avec 100ng.μl⁻¹ (en fait entre 80 et 120 ng.μl⁻¹ pour faciliter les dilutions sans robot)

Ligation des adaptateurs

Préparation des adaptateurs

Eco Adaptateur 1:	5'-CTC GTA GAC TGC GTA CC-3'	} Utiliser à 10μM*
Eco Adaptateur 2:	5'-AAT TGG TAC GCA GTC TAC-3'	
Mse Adaptateur 1:	5'-GAC GAT GAG TCC TGA G-3'	} Utiliser à 100μM*
Mse Adaptateur 2:	5'-TAC TCA GGA CTC AT-3'	

Préparer autant d'alcots que nécessaire pour toute la manip

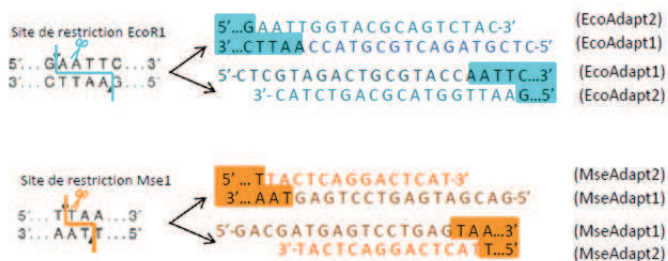
*Adaptateurs		
	Sol mère	Dilution
Eco Adapt 1&2	100μM	1/10
Mse Adapt 1&2	100μM	-

EcoMix = Eco Adapt 1+2 (1/10) 25μl+25μl = 50 μl
MseMix = Mse Adapt 1+2 25μl+25μl = 50μl

5' à 95°, refroidissement à T° ambiante, congélation à -20°C

* Les adaptateurs sont utilisés à des concentrations différentes

Ligation	Ech			
1à 2h à 22°C + 10' à 65°C	1	20	50	100
Eco Mix*	0.5	10	25	50
Mse Mix*	0.5	10	25	50
T4 ligase (20U/μl) → 0.3U/μl	0.2	4	10	20
Ligase Buffer 10X → 1X	2	40	100	200
H ₂ O	1.8	36	90	180
ADN digéré	15			
Volume total	20			



PCR-présélective: 1^{ère} sélection des fragments

Présélective PCR

		Ech			
		1	20	50	100
Tampon 10X → 1X		2	40	100	200
dNTP		0.8	16	40	80
Primer Eco-N		0.32	6.4	16	32
Primer Mse-N		0.32	6.4	16	32
DreamTaq		0.1	2	5	10
H ₂ O		14.46	289.2	723	1446
ADN lig (1/10 TE 0.1X)		2			
Volume total		20			

Programme PCR

	T°C	Temps	Nb cycles
Dénaturation initiale	94	5min	1
Dénaturation initiale	94	30s	
Hybridation	56	1min	20
Elongation	72	1min	
Elongation finale	72	5min	1

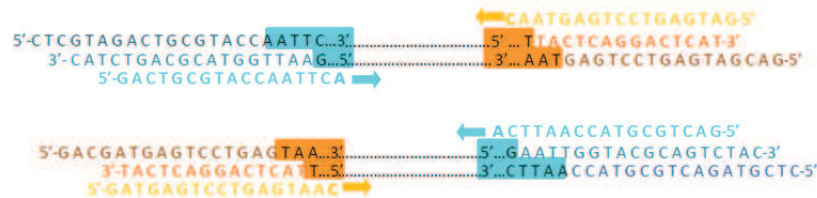
Choisir une base «N» pour ajouter au primer

↓
Diminution du nombre de fragments amplifiés

Exemple



Cas des fragments mixtes EcoR1/Mse1



PCR-sélective: 2^{ème} sélection des fragments

Sélective PCR

		Ech			
		1	20	50	100
Tampon 10X → 1X		2	40	100	200
dNTP		0.8	16	40	80
Primer Eco-NNN *		0.32	6.4	16	32
Primer Mse-NNN		0.32	6.4	16	32
DreamTaq		0.1	2	5	10
H ₂ O		11.46	229.2	573	1146
ADN ps-pcr (1/10 TE 0.1X)		5			
Volume total		20			

Programme PCR

	T°C	Temps	Nb cycles
Dénaturation initiale	94	5min	1
Dénaturation initiale	94	30s	
Hybridation	65*	30s	12
Elongation	72	1min	
Dénaturation initiale	94	30s	
Hybridation	56	30s	20
Elongation	72	1min	
Elongation finale	72	30min	1

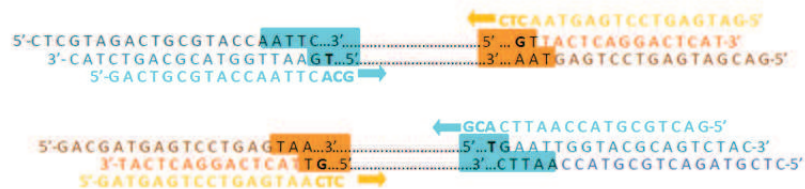
Choisir 2 bases «NN» pour ajouter au primer-N

↓
Diminution du nombre de fragments amplifiés

Exemple



Cas des fragments mixtes EcoR1/Mse1

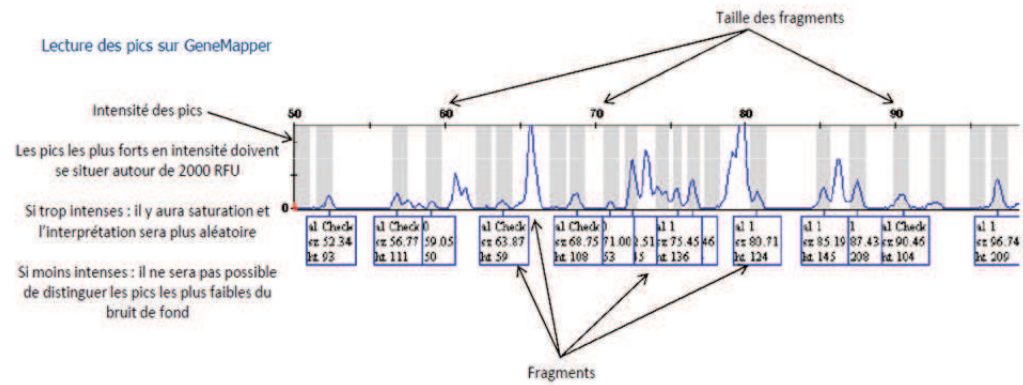


* A ce stade, une des amorces utilisées (Eco-NNN ou Mse-NNN) est marquée avec un fluorochrome

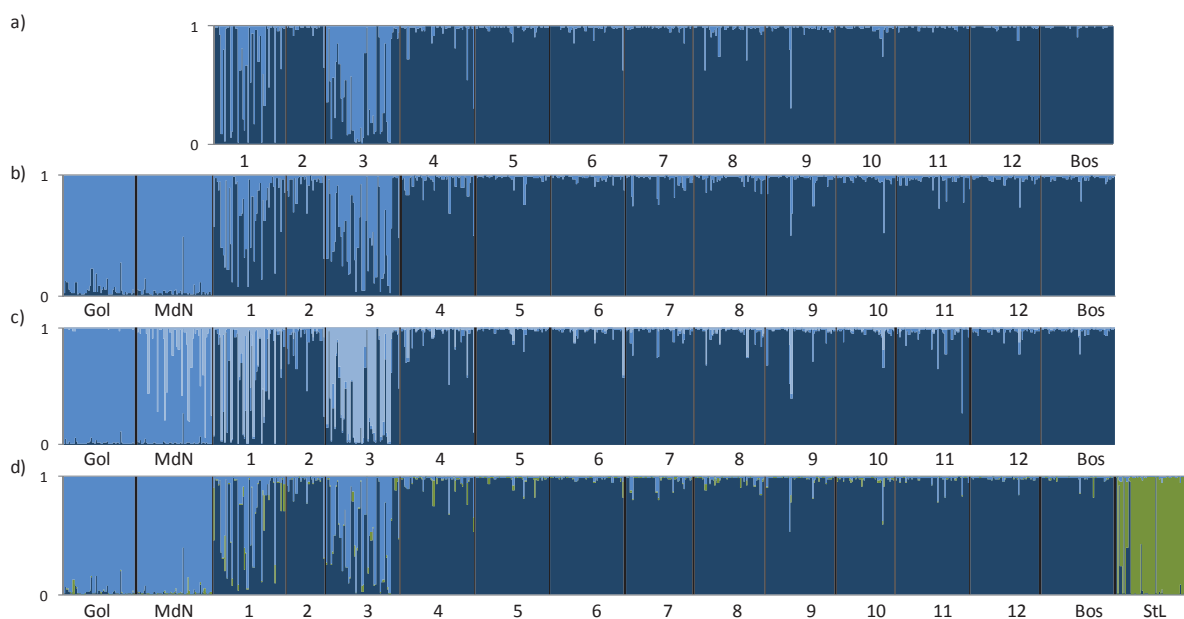
Séquençage des fragments

Séquençage	Ech			
	1	20	50	100
Formamide	13	260	650	1300
Marqueur de taille	0.3	6	15	30
ADN s-pcr (non dilué)	2			
Volume total	15			

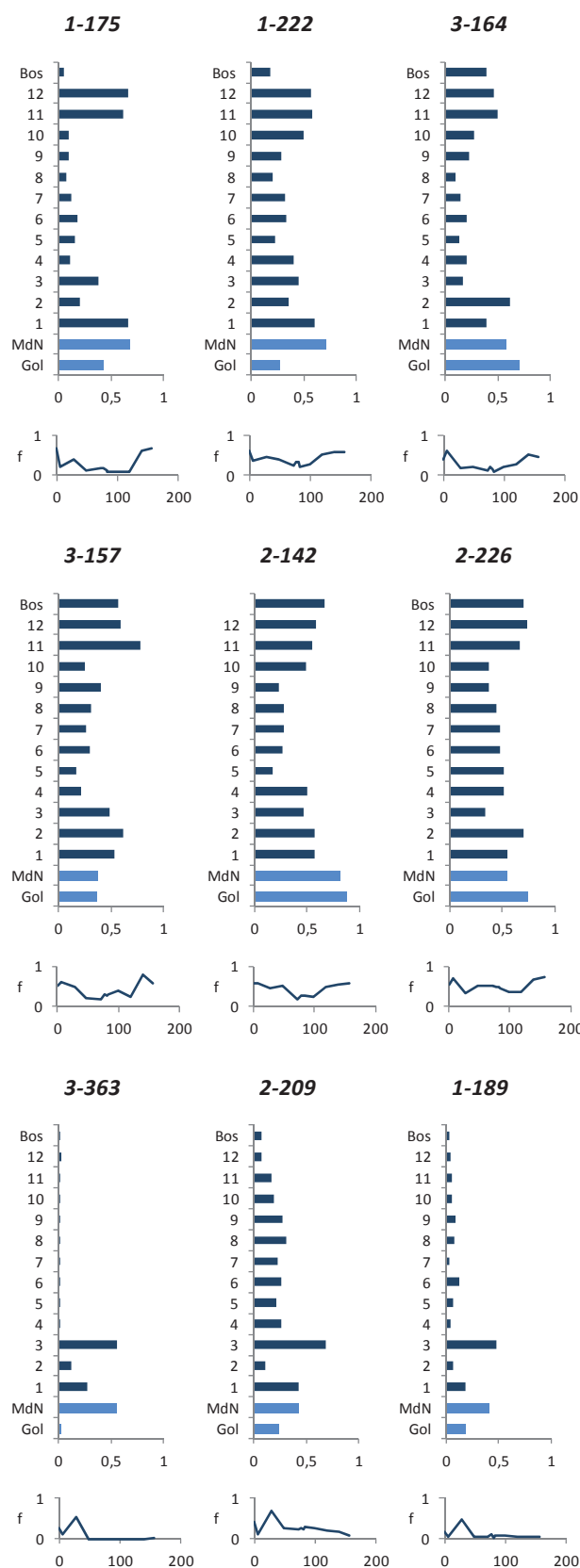
← La quantité de marqueur de taille est à optimiser pour chaque manip



Annexe 2 : Structure des populations inférée par assignation bayésienne dans STRUCTURE v2.3.3 (Falush *et al.* 2007). Représentation graphique des probabilités d'assignation individuelles pour 310 marqueurs AFLP polymorphes où chaque barre représente un individu et chaque couleur représente un cluster. L'analyse a été conduite par un algorithme MCMC (Markov Chain Monte Carlo) sous le modèle d'admixture avec les fréquences d'allèles corrélées pour 300000 générations après 100000 étapes de 'burn-in', sans connaissance a priori de l'origine géographique des individus. Le nombre de clusters le plus probable a été estimé par la méthode d'Evanno (2005). a) K=2 pour les 12 populations du transect de Long Island + Boston. b) K=2 et c) K=3 pour l'ensemble des populations *M. edulis* de cette étude (2 populations européennes : Mer du Nord et Golfe de Gascogne, 13 populations américaines : 12 populations du transect de Long Island + Boston). d) K=3 pour les populations *M. edulis* européennes et américaines et pour la population *M. trossulus* de l'embouchure du fleuve Saint-Laurent (Canada).



Annexe 3 : Fréquences alléliques des 9 locus identifiés comme outliers. Pour chacun des locus, histogramme des variations de fréquence pour l'ensemble des populations *M. edulis* étudiées et spectre des fréquences alléliques dans le détroit de Long Island en fonction de la distance à l'est de Southport (en Km).



REFERENCES BIBLIOGRAPHIQUES

-A-

- Alleaume-Benharira M, Pen IR, Ronce O (2006) Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. *Journal of Evolutionary Biology* **19**, 203-215.
- Anderson TM, vonHoldt BM, Candille SI, *et al.* (2009) Molecular and Evolutionary History of Melanism in North American Gray Wolves. *Science* **323**, 1339-1343.
- Antao T, Beaumont MA (2011) Mchaza: a workbench to detect selection using dominant markers. *Bioinformatics* **27**, 1717-1718.
- Antonovics J (1976) Nature of limits to natural selection. *Annals of the Missouri Botanical Garden* **63**, 224-247.
- Arnold ML (2004) Transfer and origin of adaptations through natural hybridization: Were Anderson and Stebbins right? *Plant Cell* **16**, 562-570.
- Atwell S, Huang YS, Vilhjalmsen BJ, *et al.* (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**, 627-631.

-B-

- Barrett RDH, Hoekstra HE (2012) Molecular spandrels: tests of adaptation at the genetic level (vol 12, pg 767, 2011). *Nature Reviews Genetics* **13**, 70-70.
- Barton N, Bengtsson BO (1986) The barrier to genetic exchange between hybridizing populations. *Heredity* **57**, 357-376.
- Barton NH (1979) The dynamics of hybrid zones. *Heredity* **43**, 341-359.
- Barton NH (1986) The effects of linkage and density-dependant regulation on gene flow. *Heredity* **57**, 415-426.
- Barton NH (2000) Genetic hitchhiking. *Philos Trans R Soc Lond B Biol Sci* **355**.
- Barton NH, Gale KS (1993) Genetic analysis of hybrid zones. In: *Hybrid Zones and the Evolutionary Process* (ed. G HR), pp. 13-45. Oxford University Press, Oxford.
- Barton NH, Hewitt GM (1981) Hybrid zones and speciation. In: *Essays on evolution and speciation* (eds. Atchley WR & Woodruff DS), pp. 109-145. Cambridge University Press, Cambridge.

- Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16**, 113-148.
- Barton NH, Shpak M (2000) The effect of epistasis on the structure of hybrid zones. *Genet Res* **75**, 179-198.
- Bates JA, Innes DJ (1995) Genetic variation among populations of *Mytilus ssp* in eastern Newfoundland. *Marine Biology* **124**, 417-424.
- Bazin E, Dawson KJ, Beaumont MA (2010) Likelihood-free inference of population structure and local adaptation in a bayesian hierarchical model. *Genetics* **185**, 587-602.
- Beaumont MA (2005) Adaptation and speciation: what can FST tell us? *Trends in Ecology & Evolution* **20**, 435-440.
- Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**, 969-980.
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proc. Roy. Soc. Ser. B* **263**, 1619-1626.
- Belkhir, K., Borsa, P., Goudet, J., Chikhi, L., Bonhomme, F. (2004). GENETIX 4.05, software under Windows© for the genetics of populations. University of Montpellier, Montpellier, France.
- Benzécri JP (1982) 2: L'analyse des correspondances. In: *L'analyse des données* (ed. Dunod), Paris.
- Berry A, Kreitman M (1993) Molecular analysis of an allozyme cline: alcohol dehydrogenase in *Drosophila melanogaster* on the East Coast of North America. *Genetics* **134**, 869-893.
- Bert TM, Arnold WS (1995) An empirical test of predictions of two competing models for the maintenance and fate of hybrid zones: both models are supported in a hard-clam hybrid zone. *Evolution* **49**, 276-289.
- Black, W. C., Krafur, E. S. (1985). A FORTRAN program for the calculation and analysis of two-locus linkage disequilibrium coefficients. *Theoretical and Applied Genetics* **70**, 491-496.
- Bierne N (2010) The distinctive footprint of local hitchhiking in a varied environment and global hitchhiking in a subdivided population. *Evolution* **64**, 3254-3272.
- Bierne N, Bonhomme F, Boudry P, Szulkin M, David P (2006) Fitness landscapes support the dominance theory of post-zygotic isolation in the mussels *Mytilus edulis* and *M. galloprovincialis*. *Proc Biol Sci* **273**.

- Bierne N, Bonhomme F, David P (2003a) Genetics at larval stage in marine bivalves. In: *Recent Advances in Marine Biotechnology 10* (eds. Fingerman M & Nagabhushanam R), pp. 239-262. Science Publishers, Inc, Enfield (NH), USA, Plymouth, UK.
- Bierne N, Bonhomme F, David P (2003b) Habitat preference and the marine-speciation paradox. *Proc Biol Sci* **270**.
- Bierne N, Borsa P, Daguin C, *et al.* (2003c) Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. *Molecular Ecology* **12**, 447-461.
- Bierne N, David P, Boudry P, Bonhomme F (2002a) Assortative fertilization and selection at larval stage in the mussels *Mytilus edulis* and *M. galloprovincialis*. *Evolution* **56**.
- Bierne N, David P, Langlade A, Bonhomme F (2002b) Can habitat specialisation maintain a mosaic hybrid zone in marine bivalves? *Marine Ecology-Progress Series* **245**, 157-170.
- Bierne N, Gagnaire P-A, David P (2012) The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. *Current Zoology*, **in press**.
- Bierne N, Tanguy A, Faure M, *et al.* (2007) Mark–recapture cloning: a straightforward and costeffective cloning method for population genetics of singlecopy nuclear DNA sequences in diploids. *Molecular Ecology Notes* **7**, 562–566.
- Bierne N, Roze D, Welch JJ (2013) Pervasive selection or is it...? Why are Fst outliers sometimes so frequent? *Molecular Ecology*, **in press**.
- Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology* **20**, 2044-2072.
- Binladen J, Gilbert MTP, Bollback JP, *et al.* (2007) The Use of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing. *PLoS One* **2**, e197.
- Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Mol Biol Evol* **23**.
- Boon E, Faure MF, Bierne N (2009) The flow of antimicrobial peptide genes through a genetic barrier between *Mytilus edulis* and *M. galloprovincialis*. *J Mol Evol* **68**, 461-474.
- Borsa P, Naciri M, Bahri L, *et al.* (1997) Zoogéographie infra-spécifique de la mer Méditerranée. *Vie Milieu* **47**, 295-305.

- Bryant D, Moulton V (2004) Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. *Mol Biol Evol* **21**, 255-265.
- Bulnheim HP, Gosling E (1988) Population genetic structure of mussels from the Baltic sea. *Helgolander meeresuntersuchungen* **42**, 113-129.
- Butlin RK, Galindo J, Grahame JW (2008) Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 2997-3007.

-C-

- Campbell D, Bernatchez L (2004) Generic Scan Using AFLP Markers as a Means to Assess the Role of Directional Selection in the Divergence of Sympatric Whitefish Ecotypes. *Molecular Biology and Evolution* **21**, 945-956.
- Case RAJ, Hutchinson WF, Hauser L, Van Oosterhout C, Carvalho GR (2005) Macro- and micro-geographic variation in pantophysin (PanI) allele frequencies in NE Atlantic cod *Gadus morhua*. *Marine Ecology-Progress Series* **301**, 267-278.
- Cavalli-Sforza L (1966) Population structure and human evolution. *Proceedings of the Royal Society B: Biological Sciences* **164**, 362-379.
- Chaoui L, Gagnaire PA, Guinand B, *et al.* (2012) Microsatellite length variation in candidate genes correlates with habitat in the gilthead sea bream *Sparus aurata*. *Molecular Ecology* **in press**.
- Charlesworth B, Charlesworth D, Barton NH (2003) The effects of genetic and geographic structure on neutral variation. *Annual Review of Ecology Evolution and Systematics* **34**, 99-125.
- Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research* **70**, 155-174.
- Chen H, Patterson N, Reich D (2010) Population differentiation as a test for selective sweeps. *Genome Research* **20**, 393-402.
- Colbeck GJ, Turgeon J, Sirois P, Dodson JJ (2011) Historical introgression and the role of selective vs. neutral processes in structuring nuclear genetic variation (AFLP) in a circumpolar marine fish, the capelin (*Mallotus villosus*). *Molecular Ecology* **20**, 1976-1987.

- Colosimo PF, Hosemann KE, Balabhadra S, *et al.* (2005) Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* **307**, 1928-1933.
- Côté C, Gagnaire PA, Bourret V, *et al.* (in press) Population genetics of the American eel (*Anguilla rostrata*): $F_{ST} = 0$ and NAO effects on demographic fluctuations of a panmictic species. *Molecular Ecology*.
- Coustau C, Renaud F, Delay B (1991) Genetic characterization of the hybridization between *Mytilus edulis* and *M. galloprovincialis* on the Atlantic coast of France. *Marine Biology* **111**, 87-93.

-D-

- Daguin C, Bonhomme F, Borsa P (2001) The zone of sympatry and hybridization of *Mytilus edulis* and *M. galloprovincialis*, as described by intron length polymorphism at locus mac-1. *Heredity* **86**, 342-354.
- Daguin C, Borsa P (1999) Genetic characterisation of *Mytilus galloprovincialis* Lmk. in North West Africa using nuclear DNA markers. *Journal of Experimental Marine Biology and Ecology* **235**, 55-65.
- Dasmahapatra KK, Walters JR, Briscoe AD, *et al.* (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**, 94-98.
- Davey JW, Hohenlohe PA, Etter PD, *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**, 499-510.
- David P, Samadi S (2000) *La théorie de l'évolution : une logique pour la biologie* Flammarion, Manhecourt.
- Diz AP, Presa P (2008) Regional patterns of microsatellite variation in *Mytilus galloprovincialis* from the Iberian Peninsula. *Marine Biology* **154**, 277-286.
- Dobzhansky T (1937) *Genetics and the origin of species* Columbia University Press, New York.
- Dufresne F, Bourget E, Bernatchez L (2002) Differential patterns of spatial divergence in microsatellite and allozyme alleles: further evidence for locus-specific selection in the acorn barnacle, *Semibalanus balanoides*? *Molecular Ecology* **11**, 113-123.
- Duran S, Pascual M, Estoup A, Turon X (2004) Strong population structure in the marine sponge *Crambe crambe* (Poecilosclerida) as revealed by microsatellite markers. *Molecular Ecology* **13**, 511-522.

-E-

- Egan SP, Nosil P, Funk DJ (2008) Selection and genomic differentiation during ecological speciation: isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. *Evolution* **62**, 1162-1181.
- Eldon B, Wakeley J (2009) Coalescence Times and F-ST Under a Skewed Offspring Distribution Among Individuals in a Population. *Genetics* **181**, 615-629.
- Endler JA (1977) *Geographic variation, speciation and clines*. Princeton University Press, Princeton, New Jersey.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.
- Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* **103**, 285-298.

-F-

- Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* **7**, 574-578.
- Faure B, Bierne N, Tanguy A, Bonhomme F, Jollivet D (2007) Evidence for a slightly deleterious effect of intron polymorphisms at the EF1alpha gene in the deep-sea hydrothermal vent bivalve *Bathymodiolus*. *Gene* **406**.
- Faure MF (2008) *Génétique des barrières semi-perméables au flux génique dans la zone d'hybridation des moules *Mytilus edulis* et *M. galloprovincialis**, Montpellier 2.
- Faure MF, David P, Bonhomme F, Bierne N (2008) Genetic hitchhiking in a subdivided population of *Mytilus edulis*. *BMC Evol Biol* **8**, 164-177.
- Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405-1413.
- Feder JL, Nosil P (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* **64**, 1729-1747.
- Fischer MC, Foll M, Excoffier L, Heckel G (2011) Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). *Molecular Ecology* **20**, 1450-1462.
- Foll M, Fischer MC, Heckel G, Excoffier L (2010) Estimating population structure from AFLP amplification intensity. *Molecular Ecology* **19**, 4638-4647.

- Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective. *Genetics* **180**, 977-993.
- Fraser DJ, Weir LK, Bernatchez L, Hansen MM, Taylor EB (2011) Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. *Heredity* **106**, 404-420.
- Fumagalli M, Sironi M, Pozzoli U, *et al.* (2011) Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through Human Evolution. *PLoS Genet* **7**, e1002355.

-G-

- Gagnaire P-A, Minegishi Y, Zenboudji S, *et al.* (2011) Within-population structure highlighted by differential introgression across semipermeable barriers to gene flow in *Anguilla Marmorata*. *Evolution* **65**, 3413-3427.
- Gagnaire PA, Albert V, Jónsson B, Bernatchez L (2009) Natural selection influences AFLP intraspecific genetic variability and introgression patterns in Atlantic eels. *Molecular Ecology* **18**, 1678-1691.
- Gagnaire PA, Normandeau E, Cote C, Hansen MM, Bernatchez L (2012) The Genetic Consequences of Spatially Varying Selection in the Panmictic American Eel (*Anguilla rostrata*). *Genetics* **190**, 725-U703.
- Galan M, Guivier E, Caraux G, Charbonnel N, Cosson JF (2010) A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genomics* **11**, 296-311.
- Galindo J, Morán P, Rolán-Alvarez E (2009) Comparing geographical genetic differentiation between candidate and noncandidate loci for adaptation strengthens support for parallel ecological divergence in the marine snail *Littorina saxatilis*. *Molecular Ecology* **18**, 919-930.
- Galvani AP, Slatkin M (2003) Evaluating plague and smallpox as historical selective pressures for the CCR5-Δ32 HIV-resistance allele. *Proceedings of the National Academy of Sciences* **100**, 15276-15279.
- Garcia-Ramos G, Kirkpatrick M (1997) Genetic models of adaptation and gene flow in peripheral populations. *Evolution* **51**, 21–28.
- Gavrilets S (1997) Single-locus clines. *Evolution* **51**, 979-983.

- Gay P, O'Donnell J (2009) Comparison of the salinity structure of the Chesapeake Bay, the Delaware Bay and Long Island Sound using a linearly tapered advection-dispersion model. *Estuaries and Coasts* **32**, 68-87.
- Gillespie JH (1991) *The causes of molecular evolution* Oxford University Press, Oxford, UK.
- Gonzalez M, Gueguen Y, Desserre G, *et al.* (2007) Molecular characterization of two isoforms of defensin from hemocytes of the oyster *Crassostrea gigas*. *Developmental & Comparative Immunology* **31**, 332-339.
- Gorelick R, Laubichler MD (2004) Decomposing multilocus linkage disequilibrium. *Genetics* **166**, 1581-1583.
- Gosset C, Bierne N (2013) Differential introgression from a sister species explains high Fst outlier loci within a mussel species. *Journal of Evolutionary Biology* **26**, 14-26.
- Gueguen Y, Romestand B, Fievet J, *et al.* (2009) Oyster hemocytes express a proline-rich peptide displaying synergistic antimicrobial activity with a defensin. *Molecular Immunology* **46**, 516-522.
- Guinand B, Lemaire C, Bonhomme F (2004) How to detect polymorphisms undergoing selection in marine fishes? A review of methods and case studies, including flatfishes. *Journal of Sea Research* **51**, 167-182.

-H-

- Hadfield JD, Richardson DS, Burke T (2006) Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Molecular Ecology* **15**, 3715-3730.
- Hall D, Luquez V, Garcia VM, *et al.* (2007) Adaptive population differentiation in phenology across a latitudinal gradient in European Aspen (*Populus tremula*, L.): A comparison of neutral markers, candidate genes and phenotypic traits. *Evolution* **61**, 2849-2860.
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* **41**, 95-98.
- Hamblin MT, Di Rienzo A (2000) Detection of the Signature of Natural Selection in Humans: Evidence from the Duffy Blood Group Locus. *The American Journal of Human Genetics* **66**, 1669-1679.
- Harrison RG (1990) Hybrid zones: windows on evolutionary processes. *Oxf. Surv. Evol. Biol.* **7**, 69-128.

- Harrison RG (1993) Hybrids and hybrid zones: historical perspective. In: *Hybrid zones and the evolutionary process* (ed. Harrison RG), pp. 3-12. Oxford University Press, New York, Oxford.
- Hartl DL, Clark AG (2007). In: *Principals of Population Genetics (4th ed.)*
- Hartl DL, Dykhuizen DE, Dean AM (1985) Limits of adaptation: the evolution of selective neutrality. *Genetics* **111**, 655-674.
- Hedgecock D, Pudovkin AI (2011) Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *Bulletin of Marine Science* **87**, 971-1002.
- Hedrick PW (1986) Genetic-polymorphism in heterogeneous environments - a decade later. *Annual Review of Ecology and Systematics* **17**, 535-566.
- Hedrick PW (2000) *Genetics of populations*, Sudbury, MA.
- Hedrick PW (2006) Genetic polymorphism in heterogeneous environments: the age of genomics. *Annual Review of Ecology, Evolution, and Systematics* **37**, 67-93.
- Hellberg ME (2009) Gene flow and isolation among populations of marine animals. *Annual Review of Ecology, Evolution and Systematics* **40**, 291-310.
- Hereford J (2009) A Quantitative Survey of Local Adaptation and Fitness Trade-Offs. *American Naturalist* **173**, 579-588.
- Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335-2352.
- Hewitt GM (1975) Sex-chromosome hybrid zone in grasshopper *Podisma pedestris* (Orthoptera Acrididae). *Heredity* **35**, 375-&.
- Hewitt GM (1989) The subdivision of species by hybrid zones. In: *Speciation and its consequences* (eds. Otte D & Endler JA), pp. 85-110. Sinauer Ass. Inc., Sunderland, Massachusetts.
- Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. . *Biol. J. Linn. Soc.* **58**, 247-276.
- Hewitt GM (2011) Quaternary phylogeography: the roots of hybrid zones. *Genetica* **139**, 617-638.
- Hilbish TJ (1985) Demographic and temporal structure of an allele frequency cline in the mussel *Mytilus edulis*. *Marine Biology*, 163-171.
- Hilbish TJ, Koehn RK (1985) The physiological basis of natural selection at the *LAP* locus. *Evolution* **39**, 1302-1317.

- Hilbish TJ, Lima FP, Brannock PM, *et al.* (2012) Change and stasis in marine hybrid zones in response to climate warming. *Journal of Biogeography* **39**, 676-687.
- Hilbish TJ, Mullinax A, Dolven SI, *et al.* (2000) Origin of the antitropical distribution pattern in marine mussels (*Mytilus spp.*): routes and timing of transequatorial migration. *Marine Biology* **136**, 69-77.
- Hill WG (1982) Rates of change in quantitative traits from fixation of new mutations. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **79**, 142-145.
- Hoekstra HE, Drumm KE, Nachman MW (2004) Ecological genetics of adaptive color polymorphism in pocket mice: Geographic variation in selected and neutral genes. *Evolution* **58**, 1329-1341.
- Hoffman JI, Clark MS, Amos W, Peck LS (2012) Widespread amplification of amplified fragment length polymorphisms (AFLPs) in marine Antarctic animals. *Polar Biology* **35**, 919-929.
- Hohenlohe PA, Bassham S, Etter PD, *et al.* (2010) Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genet* **6**, e1000862.
- Holm ER, Bourget E (1994) Selection and population genetic structure of the barnacle *Semibalanus balanoides* in the northwest Atlantic and gulf of St-Lawrence. *Marine Ecology Progress Series* **113**, 247-256.
- 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, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**, 583-589.
- Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies *Molecular Biology and Evolution* **23**, 254-267.
- Huvet A, Lapègue S, Magoulas A, Boudry P (2000) Mitochondrial and nuclear DNA phylogeography of *Crassostrea angulata*, the Portuguese oyster endangered in Europe. *Conservation Genetics* **1**, 251-262.

-J-

- Johannesson K, Andre C (2006) Life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Molecular Ecology* **15**, 2013-2029.
- Jumbo-Lucioni P, Ayroles JF, Chambers MM, *et al.* (2010) Systems genetics analysis of body weight and energy metabolism traits in *Drosophila melanogaster*. *BMC Genomics* **11**.

-K-

- Kaplan NL, Hudson RR, Langley CH (1989) The "hitchhiking effect" revisited. *Genetics* **123**, 887-899.
- Kasapidis P, Silva A, Zampicinini G, Magoulas A (2012) Evidence for microsatellite hitchhiking selection in European sardine (*Sardina pilchardus*) and implications in inferring stock structure. *Scientia Marina* **76**, 123-132.
- Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters* **7**, 1225-1241.
- Kelly JK (1997) A test of neutrality based on interlocus associations. *Genetics* **146**, 1197-1206.
- Kerns JA, Cargill EJ, Clark LA, *et al.* (2007) Linkage and segregation analysis of black and brindle coat color in domestic dogs. *Genetics* **176**, 1679-1689.
- Kijas JW, Townley D, Dalrymple BP, *et al.* (2009) A Genome Wide Survey of SNP Variation Reveals the Genetic Structure of Sheep Breeds. *PLoS One* **4**, e4668.
- Kijewski T, Śmietanka B, Zbawicka M, *et al.* (2011) Distribution of *Mytilus* taxa in european coastal areas as inferred from molecular markers. *Journal of Sea Research* **65**, 224-234.
- Kim Y, Maruki T (2011) Hitchhiking Effect of a Beneficial Mutation Spreading in a Subdivided Population. *Genetics* **189**, 213-226.
- Koehn RK (1991) The genetics and taxonomy of species in the genus *Mytilus*. *Aquaculture* **94**, 125-145.
- Koehn RK, Hall JG, Innes DJ, Zera AJ (1984) Genetic differentiation of *Mytilus edulis* in eastern North America. *Marine Biology* **79**, 117-126.
- Koehn RK, Milkman R, Mitton JB (1976) Population genetics of marine pelecypods: IV selection, migration and genetic differentiation in the blue mussel *Mytilus edulis*. *Evolution* **30**, 2-32.

- Koehn RK, Newell RI, Immerman F (1980) Maintenance of an aminopeptidase allele frequency cline by natural selection. *Proc Natl Acad Sci U S A* **77**, 5385-5389.
- Kruuk LE, Baird SJ, Gale KS, Barton NH (1999) A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. *Genetics* **153**.
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) *MEGA: Molecular Evolutionary Genetics Analysis* Pennsylvania State University, University Park, PA.

-L-

- Lallias D, Lapegue S, Hecquet C, Boudry P, Beaumont AR (2007) AFLP-based genetic linkage maps of the blue mussel (*Mytilus edulis*). *Anim Genet* **38**.
- Lande R (1994) Risk of population extinction from fixation of new deleterious mutations. *Evolution* **48**, 1460–1469.
- Lassen HH, Turano FJ (1978) Clinal variation and heterozygote deficit at the lap-locus in *Mytilus edulis*. *Marine Biology* **49**, 245-254.
- Latta RG (1998) Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. *American Naturalist* **151**, 283-292.
- Latta RG (2004) Gene flow, adaptive population divergence and comparative population structure across loci. *New Phytologist* **161**, 51-58.
- Launey S, Ledu C, Boudry P, Bonhomme F, Naciri-Graven Y (2002) Geographic structure in the European flat oyster (*Ostrea edulis* L.) as revealed by microsatellite polymorphism. *Journal of Heredity* **93**, 331-338.
- Lazzaro BP, Scurman BK, Clark AG (2004) Genetic basis of natural variation in *D. melanogaster* antibacterial immunity. *Science* **303**, 1873-1876.
- Le Corre V, Kremer A (2003) Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. *Genetics* **164**, 1205-1219.
- Le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology* **21**, 1548-1566.
- Leimu R, Fischer M (2008) A Meta-Analysis of Local Adaptation in Plants. *PLoS One* **3**.
- Lemaire C, Allegrucci G, Naciri M, *et al.* (2000) Do discrepancies between microsatellite and allozyme variation reveal differential selection between sea and lagoon in the sea bass (*Dicentrarchus labrax*)? *Molecular Ecology* **9**, 457-467.
- Lemaire C, Versini JJ, Bonhomme F (2005) Maintenance of genetic differentiation across a transition zone in the sea: discordance between nuclear and cytoplasmic markers. *Journal of Evolutionary Biology* **18**, 70-80.

- Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution* **17**, 183-189.
- Lenormand T, Bourguet D, Guillemaud T, Raymond M (1999) Tracking the evolution of insecticide resistance in the mosquito *Culex pipiens*. *Nature* **400**, 861-864.
- Levins R (1968) *Evolution in a changing Environment* Princeton University Press, Princeton, New Jersey.
- Lewontin RC (1974) *The Genetic Basis of Evolutionary Change* Columbia University Press, New York.
- Lewontin RC (2002) Directions in evolutionary biology. *Annual Review of Genetics* **36**, 1-18.
- Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of theory of selective neutrality of polymorphisms. *Genetics* **74**, 175-195.
- Lewontin RC, Krakauer J (1975) Testing heterogeneity of F-values. *Genetics* **80**, 397-398.
- Lexer C, Buerkle CA, Joseph JA, Heinze B, Fay MF (2007) Admixture in european *Populus* hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. *Heredity* **98**, 74-84.
- Lexer C, Joseph JA, van Loo M, *et al.* (2010) Genomic admixture analysis in european *Populus spp.* reveals unexpected patterns of reproductive isolation and mating. *Genetics* **186**, 699-U391.
- Li HP, Stephan W (2006) Inferring the demographic history and rate of adaptive substitution in *Drosophila*. *PLoS Genetics* **2**, 1580-1589.
- López JL, Mosquera E, Fuentes J, *et al.* (2001) Two-dimensional gel electrophoresis of *Mytilus galloprovincialis*: differences in protein expression between intertidal and cultured mussels. *Marine Ecology Progress Series* **224**, 149-146.
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Rev. Genet.* **4**, 981-994.
- Luquez V, Hall D, Albrechtsen BR, *et al.* (2008) Natural phenological variation in aspen (*Populus tremula*): the SwAsp collection. *Tree Genetics & Genomes* **4**, 279-292.
- Lynch M (2007) The frailty of adaptive hypotheses for the origins of organismal complexity. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 8597-8604.
- Lynch M, Gabriel W (1987) Environmental tolerance. *American Naturalist* **129**, 283-303.

-M-

- Macholán M, Baird SJE, Dufková P, *et al.* (2011) Assessing Multilocus Introgression Patterns: A Case Study on the Mouse X Chromosome in Central Europe. *Evolution* **65**, 1428-1446.
- Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics* **10**, 565-577.
- Mallet J, Barton N, Lamas G, *et al.* (1990) Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in heliconius hybrid zones. *Genetics* **124**, 921-936.
- Margulies M, Egholm M, Altman WE, *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380.
- Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetics Research* **23**, 23-35.
- Mayr E (1963) *Animal species and Evolution* Harvard University Press, Cambridge.
- McDonald JH, Koehn RK (1988) The mussels *Mytilus galloprovincialis* and *M.trossulus* on the Pacific coast of North America. *Marine Biology* **99**, 111-118.
- McDonald JH, Koehn RK, Balakirev ES, *et al.* (1990) Species identity of the "common mussel" inhabiting the asiatic coasts of the Pacific Ocean. *Soviet Journal of Marine Biology (Biologia morya)* **16 (ou 1)**, 10-18 (ou 13-22).
- McDonald JH, Seed R, Koehn RK (1991) Allozymes and morphometric characters of three species of *Mytilus* in the Northern and Southern hemispheres. *Marine Biology* **111**, 323-333.
- McKay JK, Latta RG (2002) Adaptive population divergence: markers, QTL and traits. *Trends in Ecology & Evolution* **17**, 285-291.
- McNally KL, Childs KL, Bohnert R, *et al.* (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 12273-12278.
- Metzker ML (2010) Sequencing technologies - the next generation. *Nature Reviews Genetics* **11**, 31-46.
- Midamegbe A, Vitalis R, Malausa T, *et al.* (2011) Scanning the European corn borer (*Ostrinia* spp.) genome for adaptive divergence between host-affiliated sibling species. *Molecular Ecology* **20**, 1414-1430.

- Miller JR, Hawthorne D (2005) Durability of marker-quantitative trait loci haplotypes in structured populations. *Genetics* **171**, 1353-1364.
- Miranda MBB, Innes DJ, Thompson RJ (2010) Incomplete Reproductive Isolation in the Blue Mussel (*Mytilus edulis* and *M. trossulus*) Hybrid Zone in the Northwest Atlantic: Role of Gamete Interactions and Larval Viability. *Biological Bulletin* **218**, 266-281.
- Mitta G, Hubert F, Dyrzynda EA, Boudry P, Roch P (2000) Mytilin B and MGD2, two antimicrobial peptides of marine mussels: gene structure and expression analysis. *Developmental & Comparative Immunology* **24**, 381-393.
- Mitton JB (1997) *Selection in Natural Populations*, Oxford University Press edn., New York.
- Moore WS, Price JT (1993) Nature of selection in the Northern Flicker Hybrid Zone and its implications for speciation theory. In: *Hybrid Zones and the Evolutionary Process* (ed. Harrison RG), pp. 196–225. Oxford University Press, Oxford.
- Muller HJ (1942) Isolating mechanisms, evolution and temperature. *Biological Symposia* **6**, 71-125.
- Murray MC, Hare MP (2006) A genomic scan for divergent selection in a secondary contact zone between Atlantic and Gulf of Mexico oysters, *Crassostrea virginica*. *Molecular Ecology* **15**, 4229-4242.

-N-

- Nagylaki T (1975) Conditions for the existence of cline. *Genetics* **80**, 595-615.
- Namroud MC, Beaulieu J, Juge N, Laroche J, Bousquet J (2008) Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology* **17**, 3599-3613.
- Narum SR, Hess JE (2011) Comparison of FST outlier tests for SNP loci under selection. *Molecular Ecology Resources* **11**, 184-194.
- Nei M (1987) *Molecular evolutionary genetics* Columbia University Press, New York.
- Nei M (2007) The new mutation theory of phenotypic evolution. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 12235-12242.
- Nei M, Chakravarti A (1977) Drift variances of FST and GST statistics obtained from a finite number of isolated populations. *Theoretical Population Biology* **11**, 307-325.
- Nei M, Chakravarti A, Tateno Y (1977) Mean and variance of FST in a finite number of incompletely isolated populations. *Theoretical Population Biology* **11**, 291-306.

- Nei M, Maruyama T (1975) Lewontin-Krakauer test for neutral genes - Comment. *Genetics* **80**, 395-395.
- Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D (2009) Population genomics of marine fishes: identifying adaptive variation in space and time. *Molecular Ecology* **18**, 3128-3150.
- Nielsen R (2005) Molecular signatures of natural selection. In: *Annual Review of Genetics*, pp. 197-218.
- Nielsen R, Hellmann I, Hubisz MJ, Bustamante CD, Clark AG (2007) Recent and ongoing selection in the human genome. *Nat Rev Genet* **8**, 857-868.
- Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. *Evolution* **62**, 316-336.
- Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* **18**, 375-402.
- Novembre J, Galvani AP, Slatkin M (2005) The geographic spread of the CCR5 Δ 32 HIV-resistance allele. *PLoS Biol* **3**, e339.
- Nunes VL, Beaumont MA, Butlin RK, Paulo OS (2011) Multiple approaches to detect outliers in a genome scan for selection in ocellated lizards (*Lacerta lepida*) along an environmental gradient. *Molecular Ecology* **20**, 193-205.

-O-

- Orr HA (1995) The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* **139**, 1805-1813.

-P-

- Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics* **25**, 547-572.
- Paris M, Boyer S, Bonin A, *et al.* (2010) Genome scan in the mosquito *Aedes rusticus*: population structure and detection of positive selection after insecticide treatment. *Molecular Ecology* **19**, 325-337.
- Paris M, Despres L (2012) Identifying insecticide resistance genes in mosquito by combining AFLP genome scans and 454 pyrosequencing. *Molecular Ecology* **21**, 1672-1686.
- Patarnello T, Volckaert FAMJ, Castilho R (2007) Pillars of Hercules: is the Atlantic-Mediterranean transition a phylogeographical break? *Molecular Ecology* **16**, 4426-4444.

- Payseur BA (2010) Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources* **10**, 806-820.
- Pelc RA, Warner RR, Gaines SD (2009) Geographical patterns of genetic structure in marine species with contrasting life histories. *Journal of Biogeography* **36**, 1881-1890.
- Pennings PS, Hermisson J (2006) Soft sweeps III: the signature of positive selection from recurrent mutation. *PLoS Genet* **2**, e186.
- Pérez-Figueroa A, García-Pereira MJ, Saura M, Rolán-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology* **23**, 2267-2276.
- Pérez-Losada M, Guerra A, Carvalho GR, Sanjuan A, Shaw PW (2002) Extensive population subdivision of the cuttlefish *Sepia officinalis* (Mollusca: Cephalopoda) around the Iberian Peninsula indicated by microsatellite DNA variation. *Heredity* **89**, 417-424.
- Pertoldi C, Bijlsma R, Loeschcke V (2007) Conservation genetics in a globally changing environment: present problems, paradoxes and future challenges. *Biodiversity and Conservation* **16**, 4147-4163.
- Petry D (1983) The effect on neutral gene flow of selection at a linked locus. *Theoretical Population Biology* **23**, 300-313.
- Pialek 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.
- Pogson GH (2001) Nucleotide polymorphism and natural selection at the pantophysin (Pan I) locus in the Atlantic cod, *Gadus morhua* (L.). *Genetics* **157**, 317-330.
- Pogson GH, Mesa KA (2004) Positive Darwinian selection at the pantophysin (Pan I) locus in marine gadid fishes. *Molecular Biology and Evolution* **21**, 65-75.
- Pollinger JP, Bustamante CD, Flidel-Alon A, *et al.* (2005) Selective sweep mapping of genes with large phenotypic effects. *Genome Research* **15**, 1809-1819.
- Powers DA, Schulte PM (1998) Evolutionary adaptations of gene structure and expression in natural populations in relation to a changing environment: a multidisciplinary approach to address the million-year saga of a small fish. *Journal of Experimental Zoology* **282**, 71-94.
- Pritchard JK, Di Rienzo A (2010) Adaptation - not by sweeps alone. *Nature Reviews Genetics* **11**, 665-667.

-Q-

- Qiu J-W, remblay R, Bourget E (2002) Ontogenetic changes in hyposaline tolerance in the mussels *Mytilus edulis* and *M. trossulus*: implications for distribution. *Marine Ecology Progress Series* **228**, 143-152.
- Quesada H, Beynon CM, Skibinski DOF (1995a) A mitochondrial DNA discontinuity in the mussel *Mytilus galloprovincialis* Lmk: Pleistocene vicariance biogeography and secondary intergradation. *Molecular Biology and Evolution* **12**, 521-524.
- Quesada H, Gallagher C, Skibinski DAG, Skibinski DOF (1998a) Patterns of polymorphism and gene flow of gender-associated mitochondrial DNA lineages in European mussel populations. *Molecular Ecology* **7**, 1041-1051.
- Quesada H, Warren M, Skibinski DOF (1998b) Nonneutral evolution and differential mutation rate of gender-associated mitochondrial DNA lineages in the marine mussel *Mytilus*. *Genetics* **149**, 1511-1526.
- Quesada H, Wenne R, Skibinski DOF (1995b) Differential Introgression of Mitochondrial DNA across Species Boundaries within the Marine Mussel Genus *Mytilus*. *Proceedings of the Royal Society of London Series B Biological Sciences* **262**, 51-56.
- Quesada H, Zapata C, Alvarez G (1995c) A multilocus allozyme discontinuity in the mussel *Mytilus galloprovincialis*: the interaction of ecological and life-history factors. *Marine Ecology-Progress Series* **116**, 99-115.

-R-

- Rand DM, Spaeth PS, Sackton TB, Schmidt PS (2002) Ecological genetics of Mpi and Gpi polymorphisms in the Acorn Barnacle and the spatial scale of neutral and non-neutral variation. *Integrative and Comparative Biology* **42**, 825-836.
- Rasanen K, Hendry AP (2008) Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecology Letters* **11**, 624-636.
- Riginos C, Cunningham CW (2005) Local adaptation and species segregation in two mussel (*Mytilus edulis* X *Mytilus trossulus*) hybrid zones. *Mol Ecol* **14**, 381-400.
- Riginos C, Douglas KE, Jin Y, Shanahan DF, Trembl EA (2011) Effects of geography and life history traits on genetic differentiation in benthic marine fishes. *Ecography* **34**, 566-575.
- Riginos C, Hickerson MJ, Henzler CM, Cunningham CW (2004) Differential patterns of male and female mtDNA exchange across the Atlantic Ocean in the blue mussel, *Mytilus edulis*. *Evolution* **58**, 2438-2451.

- Riginos C, Sukhdeo K, Cunningham CW (2002) Evidence for selection at multiple allozyme loci across a mussel hybrid zone. *Molecular Biology and Evolution* **19**, 347-351.
- Robertson A (1960) A theory of limits in artificial selection. *Proceedings of the Royal Society of London Series B-Biological Sciences* **153**, 235-249.
- Robertson A (1975a) Gene frequency distributions as a test of selective neutrality. *Genetics* **81**, 775-785.
- Robertson A (1975b) Lewontin-Krakauer test for neutral genes - Comment. *Genetics* **80**, 396-396.
- Rockman MV (2012) The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* **66**, 1-17.
- Roesti M, Hendry AP, Salzburger W, Berner D (2012) Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology* **21**, 2852-2862.
- Romestand B, Molina F, Richard V, Roch P, Granier C (2003) Key role of the loop connecting the two beta strands of mussel defensin in its antimicrobial activity. *Eur J Biochem* **270**, 2805-2813.
- Ronce O, Kirkpatrick M (2001) When sources become sinks: Migrational meltdown in heterogeneous habitats. *Evolution* **55**, 1520-1531.
- Rousset, F. (2008). genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.
- Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174-175.

-S-

- Sabeti PC, Reich DE, Higgins JM, *et al.* (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832-837.
- Sabeti PC, Varilly P, Fry B, *et al.* (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913-918.
- Sanford E, Kelly MW (2011) Local Adaptation in Marine Invertebrates. In: *Annual Review of Marine Science, Vol 3* (eds. Carlson CA & Giovannoni SJ), pp. 509-535. Annual Reviews, Palo Alto.
- Sanjuan A, Comesana AS, DeCarlos A (1996) Macrogeographic differentiation by mtDNA restriction site analysis in the SW European *Mytilus galloprovincialis* Lmk. *Journal of Experimental Marine Biology and Ecology* **198**, 89-100.

- Sanjuan A, Zapata C, Alvarez G (1994) *Mytilus galloprovincialis* and *M. edulis* on the coasts of the Iberian peninsula. *Marine Ecology-Progress Series* **113**, 131-146.
- Santiago E, Caballero A (2005) Variation after a selective sweep in a subdivided population. *Genetics* **169**, 475-483.
- Santure AW, Wang JL (2009) The Joint Effects of Selection and Dominance on the Q(ST)-F-ST Contrast. *Genetics* **181**, 259-276.
- Sardà F, Calafat A, Flexas MM, *et al.* (2004) An introduction to Mediterranean deep-sea biology. *Scientia Marina* **68**, 7-38.
- Sarvas TH, Fevolden SE (2005) Pantophysin (Pan I) locus divergence between inshore v. offshore and northern v. southern populations of Atlantic cod in the north-east Atlantic. *Journal of Fish Biology* **67**, 444-469.
- Scascitelli M, Whitney KD, Randell RA, *et al.* (2010) Genome scan of hybridizing sunflowers from Texas (*Helianthus annuus* and *H. debilis*) reveals asymmetric patterns of introgression and small islands of genomic differentiation. *Molecular Ecology* **19**, 521-541.
- Schmidt PS, Bertness MD, Rand DM (2000) Environmental heterogeneity and balancing selection in the acorn barnacle *Semibalanus balanoides*. *Proceedings of the Royal Society of London B Biological Sciences* **267**, 379-384.
- Schmidt PS, Rand DM (1999) Intertidal microhabitat and selection at Mpi: interlocus contrasts in the northern acorn barnacle, *Semibalanus balanoides*. *Evolution* **53**, 135-146.
- Schmidt PS, Rand DM (2001) Adaptive maintenance of genetic polymorphism in an intertidal barnacle: Habitat- and life-stage-specific survivorship of Mpi genotypes. *Evolution* **55**, 1336-1344.
- Schulte PM, Gomez-Chiarri M, Powers DA (1997) Structural and functional differences in the promoter and 5' flanking region of *Ldh-B* within and between populations of the Teleost *Fundulus heteroclitus*. *Genetics*, 759-769.
- Seed R (1992) Systematics, evolution and distribution of mussels belonging to the genus *Mytilus*: an overview. *American Malacological Bulletin* **9**, 123-137.
- Skibinski DOF, Beardmore JA, Cross TF (1983) Aspects of the population genetics of *Mytilus* (Mytilidae; mollusca) in the British Isles. *Biological Journal of the Linnean Society* **19**, 137-183.
- Slatkin M (1973) Gene flow and selection in a cline. *Genetics* **75**, 733-756.

- Slatkin M (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematics* **16**, 393-430.
- Slatkin M (1987) Gene flow and the geographical structure of natural populations. *Science* **236**, 787-792.
- Slatkin M, Wiehe T (1998) Genetic hitch-hiking in a subdivided population. *Genet Res* **71**, 155-160.
- Śmietanka B, Burzyński A, Wenne R (2010) Comparative genomics of marine mussels (*Mytilus spp.*) gender associated mtDNA: rapidly evolving atp8. *Journal of Molecular Evolution* **71**, 385-400.
- Stapley J, Reger J, Feulner PGD, *et al.* (2010) Adaptation genomics: the next generation. *Trends in Ecology & Evolution* **25**, 705-712.
- Staubach F, Lorenc A, Messer PW, *et al.* (2012) Genome Patterns of Selection and Introgression of Haplotypes in Natural Populations of the House Mouse (*Mus musculus*). *PLoS Genetics* **8**, e1002891.
- Stephan W, Song YS, Langley CH (2006) The hitchhiking effect on linkage disequilibrium between linked neutral loci. *Genetics* **172**, 2647-2663.
- Stockwell CA, Hendry AP, Kinnison MT (2003) Contemporary evolution meets conservation biology. *Trends in Ecology & Evolution* **18**, 94-101.
- Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* **14**, 671-688.
- Storz JF, Nachman MW (2003) Natural selection on protein polymorphism in the rodent genus *Peromyscus*: Evidence from interlocus contrasts. *Evolution* **57**, 2628-2635.
- Storz JF, Wheat CW (2010) Integrating evolutionary and functional approaches to infer adaptation at specific loci. *Evolution* **64**, 2489-2509.
- Stranger BE, Stahl EA, Raj T (2011) Progress and Promise of Genome-Wide Association Studies for Human Complex Trait Genetics. *Genetics* **187**, 367-383.
- Szymura JM, Barton N (1986) Genetic analysis of a hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*, near Cracow in Southern Poland. *Evolution* **40**, 1141-1159.

-T-

- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.
- Tanguy A, Bierne N, Saavedra C, *et al.* (2008) Increasing genomic information in bivalves through new EST collections in four species: development of new genetic markers for environmental studies and genome evolution. *Gene* **408**, 27-36.
- Teeter KC, Thibodeau LM, Gompert Z, *et al.* (2010) The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution* **64**, 472-485.
- Tennessen JA (2005) Molecular evolution of animal antimicrobial peptides: widespread moderate positive selection. *J Evol Biol* **18**, 1387-1394.
- Tennessen JA, Woodhams DC, Chaurand P, *et al.* (2009) Variations in the expressed antimicrobial peptide repertoire of northern leopard frog (*Rana pipiens*) populations suggest intraspecies differences in resistance to pathogens. *Developmental and Comparative Immunology* **33**, 1247-1257.
- The FroSpects Gregynog Workshop (2012) Hybridization and speciation. *Journal of Evolutionary Biology* **in press**.
- Thompson JD, Higgins DG, Gibson TJ (1994) ClustalW: improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acid Research* **22**, 4673-4680.
- Tice KA, Carlon DB (2011) Can AFLP genome scans detect small islands of differentiation? The case of shell sculpture variation in the periwinkle *Echinolittorina hawaiiensis*. *Journal of Evolutionary Biology* **24**, 1814-1825.
- Tintoré J, Laviolette PE, Blade I, Cruzado A (1988) A study of an intensive density front in the Eastern Alboran sea: the Almeria-Oran front. *Journal of Physical Oceanography* **18**, 1384-1397.
- Tsakas S, Krimbas CB (1976) Testing heterogeneity of F-values - suggestion and a correction. *Genetics* **84**, 399-401.
- Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV (2010) Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics* **42**, 260-U242.

-V-

- Väinölä R, Hvilsum MM (1991) Genetic divergence and a hybrid zone between Baltic and North Sea *Mytilus* populations. *Biological Journal of the Linnean Society* **43**, 127-148.
- Varvio SL, Koehn RK, Väinölä R (1988) Evolutionary genetics of the *Mytilus edulis* species complex in the North Atlantic region. *Marine Biology* **98**, 51-60.
- Vasemagi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology* **14**, 3623-3642.
- Vekemans X (2002) AFLP-SURV version 1.0. Distributed by the author. Laboratoire de Génétique et Ecologie Végétale, Université Libre de Bruxelles, Belgium.
- Vekemans X (2010) What's good for you may be good for me: evidence for adaptive introgression of multiple traits in wild sunflower. *New Phytologist* **187**, 7-9.
- Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology* **11**, 139-151.
- Veliz D, Bourget E, Bernatchez L (2004) Regional variation in the spatial scale of selection at MPI* and GPI* in the acorn barnacle *Semibalanus balanoides* (Crustacea). *Journal of Evolutionary Biology* **17**, 953-966.
- Vitalis R (2003) DetSel 1.0: a computer program to detect markers responding to selection. *Journal of Heredity* **94**, 429-431.
- Vitalis R (2012) DETSEL: a R-package to detect marker loci responding to selection. In: *Data Production and Analysis in Population Genomics* (eds. Bonin A & Pompanon F). Humana Press, USA.
- Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics* **158**, 1811-1823.
- Vos P, Hogers R, Bleeker M, *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**.
- Vuylsteke M, Peleman JD, van Eijk MJT (2007) AFLP technology for DNA fingerprinting. *Nat Protoc* **2**.

-W-

- Ward RD, Skibinski DOF, Woodwark M (1992) Protein heterozygosity, protein structure, and taxonomic differentiation. *Evolutionary Biology* **26**, 73-159.
- Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor. Pop. Biol.* **7**, 256-276.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.
- Wenne R, Skibinski DOF (1995) Mitochondrial DNA heteroplasmy in European populations of the mussel *Mytilus trossulus*. *Marine Biology* **122**, 619-624.
- Whitlock MC (2003) Fixation probability and time in subdivided populations. *Genetics* **164**, 767-779.
- Whitlock MC (2008) Evolutionary inference from Q(ST). *Molecular Ecology* **17**, 1885-1896.
- Whitlock MC, Ingvarsson PK, Hatfield T (2000) Local drift load and the heterosis of interconnected populations. *Heredity* **84**, 452-457.
- Whitlock R, Hipperson H, Mannarelli M, Butlin RK, Burke T (2008) An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Molecular Ecology Resources* **8**, 725–735.
- Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology* **14**, 611-619.
- Williams GC (1966) *Adaptation and natural selection* Princeton University Press, Princeton.
- Williams GC, Koehn RK, Mitton JB (1973) Genetic differentiation without isolation in american eel, *Anguilla rostrata*. *Evolution* **27**, 192-204.
- Wong GKS, Liu B, Wang J, *et al.* (2004) A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* **432**, 717-722.
- Wood HM, Grahame JW, Humphray S, Rogers J, Butlin RK (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. *Mol Ecol* **17**.
- Wright S (1931) Evolution in Mendelian populations. *Genetics* **16**, 97-159.

-X-

- Xu BZ, Häggglund P, Stålbrand H, Janson J-C (2002) endo- β -1,4-Mannanases from blue mussel, *Mytilus edulis*: purification, characterization, and mode of action. *Journal of Biotechnology* **92**, 267-277.
- Xu BZ, Janson JC, Sellos D (2001) Cloning and sequencing of a molluscan endo-beta-1,4-glucanase gene from the blue mussel, *Mytilus edulis*. *European Journal of Biochemistry* **268**, 3718-3727.

-Y-

- Yang JA, Benyamin B, McEvoy BP, *et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42**, 565-U131.

-Z-

- Zasloff M (2002) Antimicrobial peptides of multicellular organisms. *Nature* **415**, 389-395.
- Zhivotovsky LA (1999) Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology* **8**, 907-913.

