



Marqueurs moléculaires et sélection en milieu marin: une relation indirecte

Nicolas Bierne

► To cite this version:

Nicolas Bierne. Marqueurs moléculaires et sélection en milieu marin: une relation indirecte. Génétique des populations [q-bio.PE]. Montpellier II, 2009. tel-03034218

HAL Id: tel-03034218

<https://hal.science/tel-03034218>

Submitted on 1 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Année 2009

MEMOIRE

présenté devant

**L'UNIVERSITE MONTPELLIER II
SCIENCES ET TECHNIQUES DU LANGUEDOC**

par

Nicolas BIERNE

en vue d'obtenir

L'HABILITATION A DIRIGER DES RECHERCHES

Marqueurs moléculaires et sélection en milieu marin: une relation indirecte

soutenue le 12.10.2009 devant la commission d'examen :

Jacques DAVID
Myriam VALERO
Oscar GAGGIOTTI
Xavier VEKEMANS
Patrice DAVID
François BONHOMME

Professeur, Montpellier SupAgro
Directeur de recherche, CNRS Roscoff
Professeur, Université de Grenoble
Professeur, Université de Lille I
Directeur de recherche, CNRS Montpellier
Directeur de recherche, CNRS Montpellier

Président
Rapporteur
Rapporteur
Rapporteur
Examineur
Examineur

Sommaire

SOMMAIRE	2
REMERCIEMENTS	3
CURRICULUM VITAE	5
DOCUMENT DE SYNTHÈSE	10
Résumé	10
I- Introduction	10
I.1- L'évolution n'est pas l'adaptation seule	10
I.2- L'effet indirect de la sélection n'est pas la sélection	12
II- Polymorphisme et structure génétique des populations marines	14
II.1- Les bivalves marins, champions du polymorphisme génétique	14
II.2- La zone d'hybridation en double mosaïque des moules <i>Mytilus edulis</i> et <i>M. galloprovincialis</i>	17
II.3- Histoire du flux génique dans deux zones d'hybridation marines	22
II.3.1- Spéciation dans les abysses: analyse multi-locus de la divergence et du flux génique entre deux espèces de moules hydrothermales	22
II.3.2- Analyse du contact secondaire entre les moules <i>Mytilus edulis</i> et <i>M. galloprovincialis</i>	24
III- Marqueurs moléculaires et sélection	28
III.1- Relations indirectes	28
III.1.1- La consanguinité générale du génome explique les corrélations hétérozygotie / valeur sélective.	28
III.1.2- Mécanismes d'isolement reproductif en milieu marin	32
III.2- Relations directes	35
III.2.1- Une proportion importante des substitutions en acide aminé sont fixées par la sélection darwinienne positive	35
III.2.2- Effet délétère des mutations introniques polymorphes du gène EF1 α chez les modioles hydrothermales du genre <i>Bathymodiolus</i>	38
III.3- Relations incertaines	38
IV- Marqueurs moléculaires, valeur sélective et environnement	41
IV.1- Le leurre du Fst trop fort : autostop local en environnement variable <i>versus</i> autostop global en population subdivisée.	41
IV.1.1- Hypothèse de l'autostop global en population structurée	43
IV.1.2- Hypothèse de l'autostop local en environnement variable	45
IV.1.3- Une marche chromosomique autour d'un locus outlier soutient l'hypothèse de l'autostop global en population structurée.	48
IV.2- Association génétique / environnement lorsque l'adaptation locale interagit avec des incompatibilités génétiques endogènes.	51
IV.2.1- Le modèle	51
IV.2.2- Résultats	52
Sélection endogène et spécialisation d'habitat modérées.	52
Sélection endogène forte et spécialisation d'habitat faible, les clines endogènes coïncident avec une barrière environnementale.	52
Sélections fortes, les associations génétique/environnement sont observées dans les clines	53
IV.2.3- Discussion	55
V- Conclusion	56
Références citées	59
ANNEXES	66

Remerciements

Quelle corvée cette HDR quand j'ai dû m'y mettre ; manque d'enthousiasme, plus de niaque. Ce n'était pas le moment. Je n'avais dirigé qu'une seule thèse depuis mon recrutement au CNRS cinq ans auparavant, mes projets ANR se faisaient rejeter les uns après les autres et surtout je me retrouvais entraîné dans un jeu de rubik's cube organisationnel, changeant de labo et perdant mes collègues les uns après les autres dans une spirale de remaniements quadriennaux. J'aurais voulu attendre de retrouver mon enthousiasme et une certaine stabilité avant de passer mon HDR. Mais le message de ma hiérarchie était on ne peut plus clair : comme les autres de ton cheptel qui ont le dossier pour, de gré ou de force, ton HDR tu soutiendras. Alors je l'ai écrite. Je vous rassure j'ai aujourd'hui retrouvé de l'enthousiasme, moteur du chercheur, et je remercie ici les personnes qui m'y ont aidé, souvent par le souvenir des moments passés en leur compagnie et que j'ai revécus en préparant cette HDR.

Je remercie les membres du jury, Myriam, Oscar et Xavier, les rapporteurs, et Jacques, le président, pour avoir évalué mon travail et m'avoir encouragé à poursuivre mes travaux dans la direction prise. En période de doute, il est réconfortant de s'entendre dire qu'on fait du bon boulot.

Déjà abondamment remercié dans ma thèse, je remercie à nouveau du fond du cœur mes deux mentors, François et Patrice, pour leur soutien indéfectible. Merci François d'avoir construit la génétique des populations marines en France. La station de Sète en reste un pôle majeur je pense, j'espère. Après la pluie quadriennale vient le beau temps de la science à l'œuvre. Et bravo pour ton Pradinel. Merci Patrice de me donner l'impression de faire une recherche originale. Chaque rencontre avec toi me regonfle d'enthousiasme. Merci aussi de m'encourager à garder le cap alors qu'il pourrait être plus facile de faire autrement. J'ai plus récemment eu la chance d'obtenir un troisième soutien qui s'est révélé m'être très précieux lors des perturbations quadriennales endurées, celui de Nicolas Galtier. Il m'est indispensable d'interagir avec toi et ton équipe, notamment Sylvain Glémin que je remercie aussi au passage. Je retrouve enfin le plaisir d'évoluer dans un environnement scientifique dynamique et convivial.

Il est temps d'aborder les remerciements de la guest star de cette section. Celui sans qui cette HDR n'existerait pas. Le seul, l'unique, du trois étoiles à l'alambique, Matthieu Faure. Ce fut un réel plaisir de t'avoir comme thésard, même si j'ai vite compris que tu te détournerais de la recherche après ta thèse. Les discussions que nous avons eues ensemble m'ont fait beaucoup progresser. Nous partageons cet appétit immodéré pour la critique scientifique. Tu vas maintenant faire un très beau et important métier. Les élèves auront de la chance de t'avoir. Les créationnistes n'ont plus qu'à rentrer aux States avec des profs comme toi.

Je voudrais remercier chaleureusement tous les étudiants qui ont contribué aux travaux présentés ici, Marta Szulkin, Eva Boon, Baptiste Faure, Christopher Sauvage, ainsi que tous les autres qui ont dû me supporter ou me supportent encore, et dont je garde un

excellent souvenir. J'espère avoir l'occasion de collaborer avec vous dans le futur, ce qui est déjà initié pour certains.

Un grand merci à Adam Eyre-Walker qui m'a appris la british touch en sciences et comment rédiger un article scientifique "straight to the point". J'aurais bien continué dans le domaine de la génomique évolutive à la Eyre-Walker mais l'appel de la mer était trop fort.

Merci à mes collaborateurs de la petite communauté des généticiens marins, Pierre Boudry, Pascale Garcia, Didier Jollivet, Sylvie Lapègue, Grant Pogson, Denis Roze, Frédérique Viard. Merci à David McKenzie et Stefano Marras, deux physiologistes qui apprécient et comprennent la démarche évolutive, et dont j'ai beaucoup apprécié la compagnie les midis ou devant une mousse. Merci aussi, encore et toujours, à mes comparses du début que je ne vois pas assez souvent, Tchouffy, Krostif et Clairouche.

Je remercie Marie-Thérèse qui fait tourner le labo à Sète et s'occupe de plus en plus de mes étudiants pour l'aspect bio mol. Je n'ai pas eu la place dans cette HDR de parler des travaux récents que tu as réalisés sur les communautés bactériennes et les scans AFLP. Merci aussi à Frédérique Ceirquera et Erick Desmarais à la plateforme de l'IFR119.

Je fais mes recherches dans ce petit microcosme qu'est la station de Sète. Je remercie l'ensemble du personnel de la station, Yoyo, Katy, Jean-François, Michel, Marie-Odile, et les collègues occupant ce lieu, Monsieur Euzet, Emilie et ses comparses d'Ecolag, et les autres.

Merci à l'ANR d'avoir finalement dénié me financer. Il était temps. « En France, on n'a pas de pétrole mais on a des idées », heureusement que j'avais un bureau éclairé et chauffé pour chercher des idées sans argent.

Enfin, merci à Sandrine et Tilia.

Curriculum Vitae

SITUATION ACTUELLE

2007 **Institut des Sciences de l'Evolution (UMR 5554)**
Montpellier, France
Chercheur CNRS de première classe

CURSUS

1998-2001 **Université de Montpellier 2**
Montpellier, France
Diplôme de docteur : Très Honorable
Spécialité : Biologie des Populations et Ecologie
Directeurs de thèse : François Bonhomme et Patrice David

1995-1996 **Ecole Nationale Supérieure d'Agronomie de Montpellier**
Montpellier, France
Diplôme d'Etudes Approfondies (DEA) (3^{ème}, mention Bien)
Formation doctorale : Biologie de l'Evolution et Ecologie

1993-1996 **Institut National Agronomique Paris-Grignon**
Paris, France
Diplôme d'Ingénieur Agronome.

EXPERIENCES PROFESSIONNELLES

2003-2006 **Génome, Populations, Interactions, Adaptation (UMR 5171)**
Montpellier, France
Chercheur CNRS de deuxième classe

2002-2003 **Centre for the Study of Evolution, University of Sussex**
Brighton, Royaume Uni
Post-doc

1998-2001 **Génome, Populations, Interactions (UMR 5000)**
Montpellier, France
Doctorat (Bourse MNERT)

1996-1998 **IFREMER - Centre Océanologique du Pacifique**
Tahiti, Polynésie française
Service national, volontaire à l'aide technique

LISTE DES PUBLICATIONS

- A1- Faure B, Jollivet D, Tanguy A, Bonhomme F & Bierne N (2009). Speciation in the deep sea: multi-locus analysis of divergence and gene flow between two hybridizing species of hydrothermal vent mussels. **PLoS ONE**, 4: e6485. Annexe 1
- A2- Boon E, Faure MF & Bierne N (2009) The flow of antimicrobial genes through a genetic barrier between *Mytilus edulis* and *M. galloprovincialis*. **Journal of Molecular Evolution**, 68:461-74. Annexe 2
- A3- Faure MF, David P, Bonhomme F & Bierne N (2008) Genetic hitchhiking in a subdivided population of *Mytilus edulis*. **BMC Evolutionary Biology** 8:164. Annexe 3
- A4- Gérard K, Bierne N, Borsa P, Chenuil A & Féral JP (2008) Pleistocene separation of mitochondrial lineages of *Mytilus* spp. mussels from Northern and Southern Hemispheres and strong genetic differentiation among southern populations. **Molecular Phylogenetics and Evolution** 49:84-91.
- A5- Tanguy A, Bierne N, Saavedra C, Pina B, Bachère E, Kube M, Bazin E, Bonhomme F, Boudry P, Boulo V, Boutet I, Cancela L, Dossat C, Favrel P, Huvet A, Jarque S, Jollivet D, Klages S, Lapègue S, Leite R, Moal J, Moraga D, Reinhardt R, Samain J-F, Zouros E, Canario A (2008) Increasing genomic information in Bivalves through new EST collections in four species: development of new genetic markers for environmental studies and genome evolution. **Gene** 408:27-36.
- A6- Bierne N, Tanguy A, Faure M, Faure B, David E, Boutet I, Boon E, Quere N, Plouviez S, Kempainen P, Jollivet D, Moraga D, Boudry P & David P (2007) Mark-recapture cloning: a straightforward and cost-effective cloning method for population genetics of single-copy nuclear DNA sequences in diploids. **Molecular Ecology Notes** 7: 562-566.
- A7- Borsa P, Daguin C & Bierne N (2007) Genomic reticulation indicates mixed ancestry in Southern-Hemisphere *Mytilus* spp. mussels. **Biological Journal of the Linnean Society** 92: 747-754.
- A8- 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: 99-107.
- A9- Sauvage C, Bierne N, Lapegue S & Boudry P (2007). Single Nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*. **Gene** 406: 22-33.
- A10- 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*. **Proceedings of the Royal Society B-Biological Sciences** 273: 1253-1260.
- A11- Bierne N & Eyre-Walker A (2006). Variation in synonymous codon use and DNA polymorphism within the *Drosophila* genome. **Journal of Evolutionary Biology** 19: 1-11.
- A12- Galtier N, Bazin E & Bierne N (2006). GC-biased segregation of noncoding polymorphisms in *Drosophila*. **Genetics** 172: 221-228.
- A13- Bierne N & Eyre-Walker A (2004). The genomic rate of adaptive amino acid substitution in *Drosophila*. **Molecular Biology and Evolution** 21: 1350-1360.
- A14- Bierne N & Eyre-Walker A (2003). The problem of counting sites in the estimation of the synonymous and nonsynonymous substitution rates: Implications for the correlation between the synonymous substitution rate and codon usage bias. **Genetics** 165: 1587-1597.
- A15- Bierne N, Bonhomme F & David P (2003). Habitat preference and the marine-speciation paradox. **Proceedings of the Royal Society B-Biological Sciences** 270: 1399-1406.

- A16- Bierne N, Borsa P, Daguin C, Jollivet D, Viard F, Bonhomme F & David P (2003). Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. **Molecular Ecology** 12: 447-461.
- A17- Bierne N, Daguin C, Bonhomme F, David P & Borsa P (2003). Direct selection on allozymes is not required to explain heterogeneity among marker loci across a *Mytilus* hybrid zone. **Molecular Ecology** 12: 2505-2510.
- A18- Bierne N, David P, Boudry P & Bonhomme F (2002). Assortative fertilization and selection at larval stage in the mussels *Mytilus edulis* and *M. galloprovincialis*. **Evolution** 56: 292-298.
- A19- Bierne N, David P, Langlade A & Bonhomme F (2002). Can habitat specialisation maintain a mosaic hybrid zone in marine bivalves? **Marine Ecology-Progress Series** 245: 157-170.
- A20- Bierne N, Lenormand T, Bonhomme F & David P (2002). Deleterious mutations in a hybrid zone: can mutational load decrease the barrier to gene flow? **Genetical Research** 80: 197-204.
- A21- Huvet A, Balabaud K, Bierne N & Boudry P (2001). Microsatellite analysis of 6-hour-old embryos reveals no preferential intraspecific fertilization between cupped oysters *Crassostrea gigas* and *Crassostrea angulata*. **Marine Biotechnology** 3: 448-453.
- A22- Bierne N, Tsitrone A & David P (2000). An inbreeding model of associative overdominance during a population bottleneck. **Genetics** 155: 1981-1990.
- A23- Bierne N, Bezuart I, Vonau V, Bonhomme F, Bedier E & Aquacop (2000). Microsatellite-associated heterosis in hatchery-propagated stocks of the shrimp *Penaeus stylirostris*. **Aquaculture** 184: 203-219.
- A24- Bierne N, Lehnert A, Bedier E, Bonhomme F & Moore SS (2000). Screening for intron-length polymorphisms in penaeid shrimps using exon-primed intron-crossing (EPIC)-PCR. **Molecular Ecology** 9: 233-235.
- A25- Moore SS, Wilson KJ, Whan V, Bierne N, Lehnert SA, Chu KH, Pongsomboon S & Tansanakajon A (1999) Molecular tools for mapping the Black tiger shrimp (*Penaeus monodon*) genome. **Plant and Animal Genome** 7, 34.
- A26- Vonau V, Ohresser M, Bierne N, Delsert C, Beuzart I, Bedier E & Bonhomme F (1999). Three polymorphic microsatellites in the shrimp *Penaeus stylirostris*. **Animal Genetics** 30: 234-235.
- A27- Bierne N, Launey S, Naciri-Graven Y & Bonhomme F (1998). Early effect of inbreeding as revealed by microsatellite analyses on *Ostrea edulis* larvae. **Genetics** 148: 1893-1906.

CHAPITRES DE LIVRE

- A28- Bierne N, Bonhomme F & David P (2002) Genetics at larval stage in marine bivalves. **Recent Advances in Marine Biotechnology** 10, edited by M. Fingerman and R. Nagabhushanam. Science Publishers, Inc., Enfield (NH), USA and Plymouth, UK.
- A29- Bierne N (2009) Sous-chapitre "Adaptation vs. sélection purifiante: évolution des séquences codantes" du chapitre "Evolution moléculaire" (N. Galtier et L; Duret eds.) et sous-chapitre "Isolement post-zygotique" du chapitre "La spéciation" (V. Ravigné ed.) du livre "Biologie évolutive" (M. Raymond, F. Thomas eds.), sous presse.

ENCADREMENT DE STAGIAIRES

Étudiants	Année	Université	Cursus	Situation actuelle
Master 1				
Lise Dupont	2000	Montpellier 2	Ecologie, Biodiversité, Evolution	Maître de conférences à Paris XII
Annabelle Mordacque	2000	Lille	Biodiversité et Ecosystèmes continentaux et marins	Secteur privé
Louis de Redon	2004	INA-PG	INA-PG	Doctorat (MNHN)
Anaïs Didier	2006	Perpignan	Ecologie fonctionnelle	Doctorat (Fribourg)
Maëva Fortin	2007	Angers	Environnement, Ecologie	M2 Eco-ingénierie des zones humides
Virginie David	2007	Pau	Environnement	M2Pro
Anna Kneidinger	2008	Montpellier 2	Biologie, Ecologie, Evolution	Master in Ecology (Vienne)
Amélie Buguel	2008	Montpellier 2	Ecologie, Biodiversité, Evolution	M1 SVT
Nicolas Glaser	2008	AgroParisTech	Ingénieur agronome	M2 Science du végétal
Marcos Grabiél	2008	Santa Cruz	Evolutionary Biology	Master (U. Santa Cruz)
Master 2				
Marta Szulkin	2001	Varsovie	Ecologie	Post-doc (Oxford)
Matthieu Faure	2004	Montpellier 2	Biologie de l'Evolution et Ecologie	Professeur de SVT
Eva Boon	2005	Groningen	Ecologie et Evolution	Doctorat (Montréal)
Anaïs Didier	2007	Perpignan	Ecologie fonctionnelle	Doctorat (Fribourg)
Joana Do Nascimento	2009	Paris 6	Océanographie	Doctorat (La Rochelle)
Doctorants, collaborations				
Erick Bazin	2006	Montpellier 2	Biologie de l'Evolution et Ecologie	Post-doc (CIRAD)
Baptiste Faure	2008	Montpellier 2	Océanographie	Post-doc (Pen state)
Christopher Sauvage	2008	La Rochelle	Génétique	Post-doc (Université Laval)
Karin Gérard	2008	Aix-Marseille	Océanographie	Post-doc (Chili)
Doctorant, co-direction				
Matthieu Faure	2007	Montpellier 2	Biologie de l'Evolution et Ecologie	Professeur de SVT

Publications avec les étudiants

- Faure B**, Jollivet D, Tanguy A, Bonhomme F & Bierne N (2009). Speciation in the deep sea: multi-locus analysis of divergence and gene flow between two hybridizing species of hydrothermal vent mussels. **PLoS ONE**, 4: e6485.
- Boon E**, **Faure MF** & Bierne N (2009) The flow of antimicrobial genes through a genetic barrier between *Mytilus edulis* and *M. galloprovincialis*. **Journal of Molecular Evolution**, 68:461-74.
- Faure MF**, David P, Bonhomme F & Bierne N (2008) Genetic hitchhiking in a subdivided population of *Mytilus edulis*. **BMC Evolutionary Biology** 8:164.
- Gérard K**, Bierne N, Borsa P, Chenuil A & Féral JP (2008) Pleistocene separation of mitochondrial lineages of *Mytilus* spp. mussels from Northern and Southern Hemispheres and strong genetic differentiation among southern populations. **Molecular Phylogenetics and Evolution** 49:84-91.
- Tanguy A, Bierne N, Saavedra C, Pina B, Bachère E, Kube M, **Bazin E**, Bonhomme F, Boudry P, Boulo V, Boutet I, Cancela L, Dossat C, Favrel P, Huvet A, Jarque S, Jollivet D, Klages S, Lapègue S, Leite R, Moal J, Moraga D, Reinhardt R, Samain J-F, Zouros E, Canario A (2008) Increasing genomic information in Bivalves through new EST collections in four species: development of new genetic markers for environmental studies and genome evolution. **Gene** 408:27-36.
- Bierne N, Tanguy A, **Faure MF**, **Faure B**, David E, Boutet I, **Boon E**, Quere N, Plouviez S, Kemppainen P, Jollivet D, Moraga D, Boudry P & David P (2007) Mark-recapture cloning: a straightforward and cost-effective cloning method for population genetics of single-copy nuclear DNA sequences in diploids. **Molecular Ecology Notes** 7: 562-566.
- 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: 99-107.
- Sauvage C**, Bierne N, Lapegue S & Boudry P (2007). Single Nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*. **Gene** 406: 22-33.
- 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*. **Proceedings of the Royal Society B-Biological Sciences** 273: 1253-1260.
- Galtier N, **Bazin E** & Bierne N (2006). GC-biased segregation of noncoding polymorphisms in *Drosophila*. **Genetics** 172: 221-228.

Activités pérennes dans l'enseignement

Co-organisation avec François Bonhomme et Bruno Guinand et intervention (3h) dans le module d'école doctorale « Génétique et évolution des organismes marins » (25h).

Co-organisation avec Sandrine Maurice et Virginie Ravigné et intervention (18h) dans l'UE de M1 « Génétique et génomique évolutive » (50h).

Participation au jury de soutenance, Master 1 « Biodiversité, Ecologie, Evolution ».

Document de synthèse

Résumé

Ma démarche est de mettre à profit les nouveaux outils de la biologie moléculaire, en association avec l'expérimentation, l'étude de terrain et le développement théorique, pour mettre à jour les mécanismes à l'œuvre au cours de l'adaptation et de la spéciation. Mes modèles biologiques de prédilection sont les organismes marins à larve planctonique car ce sont des organismes qui possèdent une biologie originale (cycle de vie, dynamique des populations, écologie ...) qui doit être prise en compte pour comprendre leur évolution. J'ai une expérience sur l'usage des techniques moléculaires pour l'étude de l'évolution, sur les méthodes modernes d'analyse des informations portées par les séquences des génomes et sur les spécificités écologiques et génétiques des organismes marins. Mes travaux les plus reconnus portent sur (i) la compréhension des corrélations entre l'hétérozygotie individuelle et la valeur sélective, (ii) la génétique des larves planctoniques, (iii) la structure des zones d'hybridation, (iv) les mécanismes d'isolement reproductif en milieu marin et (v) l'évolution adaptative des protéines.

I- Introduction

I.1- L'évolution n'est pas l'adaptation seule

La génétique évolutive est un domaine de la biologie pour lequel existent des modèles mathématiques prédictifs. C'est un atout considérable qui autorise une démarche par tests d'hypothèses et par ajustement des données à des modèles évolutifs. La démarche est la même qu'en physique où l'expérimentation valide la théorie, et l'alimente en retour. La contrepartie est d'avoir un goût pour l'abstraction théorique et d'avoir une bonne connaissance des hypothèses sous-jacentes aux modèles ; deux raisons de se trouver partiellement isolé des autres biologistes qui n'ont souvent pas un goût prononcé pour l'abstraction mathématique (Lynch, 2007) et qui n'ont souvent pas été confronté à la théorie de la génétique des populations. Néanmoins, l'évolution est aussi l'opportunité de raconter une belle histoire dans laquelle l'adaptation est souvent l'acteur principal (Clark, 2006), si bien que tout biologiste se réclame volontiers évolutionniste. En fait, la biologie de l'évolution est la seule science qui soit abordée à la fois par des spécialistes mais aussi par d'autres scientifiques et même par le grand public (Lynch, 2007). En étudiant les caractères phénotypiques (macroscopiques, cellulaires ou moléculaires), il est tentant de faire la déduction qu'ils semblent particulièrement bien appropriés à exercer leurs fonctions dans l'environnement occupé par les organismes étudiés. Il est arrivé à plusieurs reprises qu'une communauté scientifique s'approprie l'évolution, la résume à la sélection naturelle et obtienne un impact significatif jusqu'à irriter des généticiens des populations au point qu'ils écrivent une réponse argumentée. L'article de Gould et Lewontin (1979) sur ce qu'ils ont appelé le programme adaptationniste est un pamphlet bien connu contre le « tout adaptation » des biologistes

britanniques et américains de l'époque. En décrivant avec quelle complexité et apparente harmonie fonctionnent les organismes à tout niveau d'organisation, les biologistes ont trop souvent tendance à y voir la marque de l'adaptation sans s'imaginer qu'une bonne part de hasard, de contrainte et de conflit se cache derrière cette évolution. Par exemple, il n'est pas dans la culture du biologiste non-évolutionniste d'imaginer que des structures biologiques ne servent à rien -amusez vous donc à dire à un biologiste moléculaire que les introns sont inutiles¹. Plus récemment, Clark (2006) a eu son mot au sujet du « tout adaptation » en génomique, Hoekstra et Coyne (2007) sur le « tout régulation » en Evo-Devo, Lynch (2007) sur le « tout adaptation » pour expliquer l'évolution de la complexité, Hughes (2007; 2008) sur le « tout computationnel » en évolution moléculaire et l'incroyable absence de validation expérimentale. En fait, à chaque fois qu'une nouvelle voie de recherche se développe en biologie, il n'est pas rare qu'il soit proclamé qu'elle révolutionnera notre compréhension des processus évolutifs (Hoekstra, Coyne, 2007). Prenons une illustration récente encore peu critiquée et qui nous évitera de reprendre les exemples précédents. Les génomiciens se sont récemment rendu compte qu'il y avait plus de différence dans le nombre de copies des gènes (CNP pour "Copy Number Polymorphism") entre deux humains qu'attendu (Sebat *et al.*, 2004; Redon *et al.*, 2006). Comme nous étions aveugles de nous concentrer sur les polymorphismes nucléotidiques (SNP pour "Single Nucleotide Polymorphism") qui n'ont dévoilé que si peu de trace de l'adaptation... Enfin, nous découvrons les vrais changements qui expliqueront les changements phénotypiques et l'adaptation, bref l'évolution. En réalité il n'y avait pas vraiment d'attendu sur l'abondance des CNPs, et les CNPs restent beaucoup plus rares que les SNPs, seulement le passage à l'échelle des génomes entiers en révèle une quantité substantielle, rien de plus. Un effort important a été alloué à l'étude de ces CNPs qui sont certes techniquement plus compliqués à étudier que les SNPs (Freeman *et al.*, 2006; Jakobsson *et al.*, 2008). Alors qu'une comparaison aux attendus de la génétique des populations démontre que ces polymorphismes sont principalement délétères, quelque fois neutres et très rarement adaptatifs - comme tout autres polymorphismes moléculaires- leur importance dans l'explication des variations phénotypiques et dans l'adaptation est sujet à des spéculations inutilement hâtives² (Perry *et al.*, 2008).

L'évolution ne se résume pas à la sélection naturelle et l'adaptation se démontre en prenant en compte l'ensemble des alternatives possibles (contraintes développementales, biais mutationnels, déséquilibres démographiques, structures des populations, ...). Il n'est pas souhaitable d'extrapoler de l'observation ("pattern") le processus sous jacent ("process") sans vigilance (Hoekstra, Coyne, 2007). Pourquoi exiger cette vigilance sans tolérer le fun de la spéculation ? Car les processus évolutifs sont l'ardu objet d'étude des généticiens des populations si bien qu'ils n'apprécient guère

1 Pourtant ils le sont, même si certains introns ou certains motifs localisés dans les introns servent à quelque chose. Les Eubactéries par exemple ont probablement réussies à éliminer ces introns superflus et à s'en passer.

2 Je ne dis pas que l'étude des CNPs n'est pas un champ de recherches passionnant, je dis qu'il est prématuré de clamer que les CNPs sont plus important que les SNPs en évolution, tout comme il est prématuré de clamer que les polymorphismes de régulation sont plus importants que les polymorphismes en acide aminé (Hoekstra, Coyne, 2007).

qu'on en parle comme d'une déduction facile basée sur une argumentation verbale mal formalisée et appréciant encore moins qu'il soit suggéré que les évolutionnistes aient pu rester aveugles sur un point de leur discipline en ignorant les développements modernes de la science (Lynch, 2007). C'est pour cela qu'à l'adage de Dobzhansky "rien en biologie n'a de sens, si ce n'est à la lumière de l'évolution", il convient d'ajouter le bémol que "rien en évolution n'a de sens, si ce n'est à la lumière de la génétique des populations" (Lynch, 2007). L'évolution est un processus de génétique des populations gouverné par cinq forces fondamentales. La sélection darwinienne n'en est qu'une des cinq mais la seule abordée par les non spécialistes, les quatre autres sont non-adaptatives, uniquement prises en compte par les évolutionnistes, et sont la mutation, la dérive génétique, la migration et la recombinaison. Ceux qui croient que la sélection est la force principale et que les autres forces sont négligeables feraient mieux d'être plus prudents (Lynch, 2007; Nei, 2007). Un aspect qui peut frustrer certains, même parmi les évolutionnistes, est que les attendus théoriques précèdent souvent les expérimentations par plusieurs décennies d'avance. Les expérimentateurs qui mettent à l'épreuve les prédictions appartiennent à de nouvelles générations de chercheurs avec une culture différente étayée par les derniers résultats scientifiques souvent exagérément annoncés, et accueillis, comme des remises en causes drastiques des recherches passées. Mon opinion est que la plupart des modèles théoriques de la génétique des populations, aussi vieux soient ils, n'ont pas encore été illustrés avec le détail, le soin et la quantité d'exemples qu'ils méritent. L'ère de la génomique apporte l'espoir d'une plus grande facilité pour étudier l'adaptation (Luikart *et al.*, 2003; Clark, 2006; Hedrick, 2006). Cependant, si pour les espèces modèles le processus est bien enclenché, la révolution de la génomique pour l'étude des polymorphismes d'adaptation n'est pour l'instant qu'un simple effet d'annonce chez les espèces non modèles.

I.2- L'effet indirect de la sélection n'est pas la sélection

Mon HDR aborde des questions de génétiques des populations avec cet aller-retour entre expérimentation et théorie qui caractérise cette discipline. L'objectif est d'étudier la sélection à l'aide de marqueurs génétiques. C'est un travail long, fastidieux et qui dépend des avancées techniques dans la détection et l'étude des polymorphismes des génomes. 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 - disons 10 ans, soit mon âge en recherche- d'une panoplie de nouveaux marqueurs permettant une couverture génomique toujours plus grande. Cependant, il est probable que la sélection ne se produise à un moment donné que sur quelques locus noyés dans l'immensité du génome³. De ce fait, les polymorphismes adaptatifs ne sont toujours pas facilement accessibles, du moins chez les espèces non modèles. Pour les étudier j'ai d'abord adopté une approche indirecte, profitant des associations statistiques (déséquilibre de liaison) entre marqueurs génétiques et gènes sélectionnés.

3 Il faut préciser que cette vision n'est pas partagée par tous. Pour une vision alternative où la sélection est omniprésente voir par exemple le livre de Mitton JB (1997) *Selection in natural populations* Oxford University Press, New York.

Je me suis placé dans des situations où ces associations étaient maximisées comme dans des croisements de laboratoire (A27, A21, A18), des petites populations d'élevage (A23) ou dans les zones d'hybridation (A16, A15, A10). Je me suis ensuite placé sur une échelle de temps plus large durant laquelle les événements de sélection s'accumulent pour faire une somme appréciable, la divergence entre espèce, avec cette fois une approche directe sur les changements en acide aminé des protéines (A13). Plus récemment, j'ai commencé à essayer de m'affranchir de ces artifices pour étudier la sélection en action dans des populations naturelles plus standards (A3), avec l'idée, partagée par tous, que les techniques moléculaires étaient mûres pour une approche à grande échelle génomique. Mon espoir est d'aboutir prochainement à une étude directe des polymorphismes adaptatifs. Malgré quelques pistes prometteuses je n'ai pas encore identifié de façon certaine de tels polymorphismes dans les espèces que j'étudie.

Le constat auquel j'aboutis est que l'effet indirect de la sélection sur un polymorphisme neutre est trop souvent interprété comme étant équivalent à de la sélection avec un coefficient de sélection atténué. Ce n'est pas un raisonnement souhaitable car il est tout simplement faux. A partir du moment où le polymorphisme étudié est neutre, fut-il fortement lié à un polymorphisme sélectionné, l'effet s'apparente à une dérive génétique efficace (Felsenstein, 1974; Gillespie, 2000a; Charlesworth, 2009), et une migration efficace lorsque les populations sont structurées (Barton, 1979b; Barton, Bengtsson, 1986; Ingvarsson, Whitlock, 2000), modifiées par rapport à la prédiction en absence de sélection indirecte. Les interprétations peuvent être erronées si la simplification est faite que le polymorphisme neutre étudié se comporte comme son voisin sélectionné avec un simple facteur d'atténuation négligeable. C'est pour cela qu'il importe de distinguer la sélection directe de la sélection indirecte (A17). A l'argument « si ce n'est lui (le marqueur) c'est donc son voisin (un gène lié), quelle différence ?⁴ », il faut répondre que la différence est grande car on ne peut se satisfaire de détecter la sélection, il faut identifier sa nature (négative, positive, balancée), son histoire (signature d'une sélection passée ou sélection en action) et comment elle a influencée dans les détails les polymorphismes étudiés. Le fil conducteur de cette HDR est d'illustrer ce propos, l'effet indirect de la sélection n'est pas la sélection, et cette exigence de compréhension du lien indirect entre locus sélectionnés et marqueurs neutres.

4 “These results are consistent with the operation of selection on allozymes or on loci in linkage disequilibrium with them”, “suggesting selection at this locus, or a closely linked one”...

II- Polymorphisme et structure génétique des populations marines

L'évolution n'est pas la sélection seule rappelais-je dans l'introduction. De ce fait, il est difficile d'étudier la sélection lorsqu'on n'a pas une bonne maîtrise des autres forces évolutives à l'œuvre. Le prérequis est donc une bonne connaissance de son modèle biologique. Les organismes que j'étudie sont des organismes marins à dispersion larvaire planctonique, principalement des bivalves. Ils présentent des caractéristiques démographiques qui les distinguent de la plupart des organismes terrestres: très grandes fécondités, très grande taille des populations et dispersion passive à longue distance par une phase larvaire planctonique. La génétique des populations nous prédit donc des diversités génétiques fortes, des différenciations génétiques entre populations faibles, une efficacité de la sélection accrue, une adaptation locale difficile ou coûteuse car les échelles de dispersion surpassent l'hétérogénéité spatiale de l'environnement (A28). Dans ce premier chapitre, je m'intéresserai à la diversité génétique des populations et à la structure génétique entre populations.

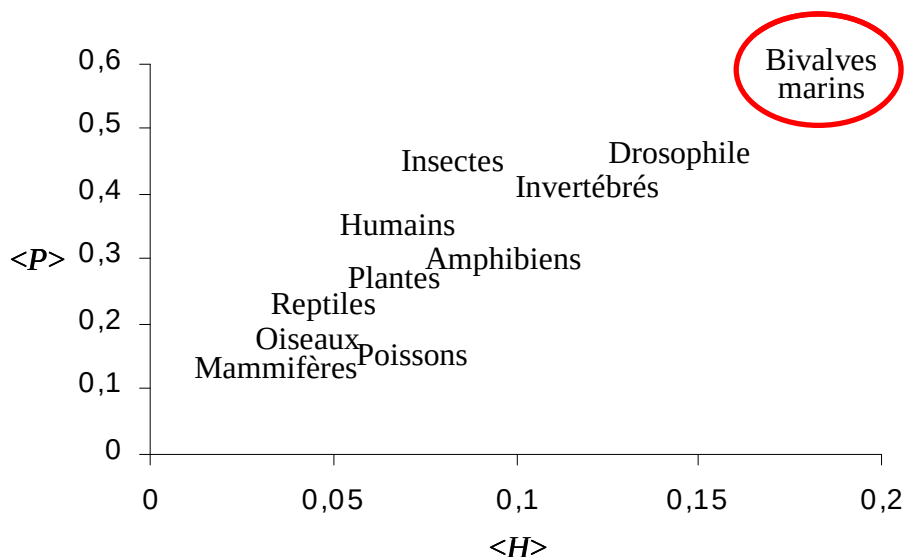


Figure 1 : Représentation schématique de la diversité enzymatique de différents groupes d'animaux ou de plantes. <H> : Hétérozygotie. <P> : Proportion de locus polymorphe. Adapté d'après Nevo (1978).

II.1- Les bivalves marins, champions du polymorphisme génétique

Les génomes des bivalves marins sont encore méconnus mais leur étude est un enjeu important car les caractéristiques originales de ces espèces permettront d'élargir ou de réviser les paradigmes établis en milieu terrestre. Entre autres, les bivalves marins nous procurent une opportunité unique d'explorer les valeurs extrêmes d'un paramètre fondamental en génétique des populations : la taille efficace des populations (N_e). Les diversités enzymatiques observées chez les

Bivalves marins sont parmi les plus élevées des Eucaryotes (Figure 1, Ward *et al.*, 1992; López *et al.*, 2001; Figure 1, Bazin *et al.*, 2006). De fortes diversités ont également été rapportées avec des marqueurs microsatellites (Huvet *et al.*, 2000; Launey *et al.*, 2002) ou introniques (A16, Daguin *et al.*, 2001). Une taille efficace élevée a aussi pour corollaire un fardeau génétique (quantités de mutations délétères ségrégeantes, Bataillon, Kirkpatrick, 2000) qui compte parmi les plus forts jamais estimés chez des animaux (A1, Launey, Hedgecock, 2001). Cependant, quasiment aucune donnée de polymorphisme de séquence d'ADN nucléaire n'était disponible chez les bivalves jusqu'à peu. J'ai coordonné un projet, « Bivalvomix: Génomique évolutive des bivalves marins », qui avait pour but de générer ce type de données chez trois bivalves, les moules *Mytilus*, les huîtres *Crassostrea* et les modioles hydrothermales *Bathymodiolus*. Au même moment au laboratoire, Eric Bazin, Sylvain Glémin et Nicolas Galtier démontraient la surprenante invariance de la diversité nucléotidique de l'ADN mitochondrial pour des groupes animaux présentant des tailles de populations et des diversités allozymiques variables (Bazin *et al.*, 2006). Par exemple, alors que l'hétérozygotie allozymiques des mollusques marins est bien plus forte que celle des mollusques terrestres ($H=0.3$ pour les marins contre $H=0.074$ pour les terrestres), les diversités nucléotidiques mitochondriales sont semblables ($\pi=0.056$ pour les marins contre $\pi=0.078$ pour les terrestres). Ce résultat est interprété comme étant la conséquence de fixations récurrentes de mutations favorables qui diminuent la diversité mitochondriale neutre par autostop génétique du fait de l'absence de recombinaison, un processus appelé « genetic draft »⁵ par Gilliespie (2000b; 2000a). Bien que j'adhère à la conclusion de mes collègues -elle me conforte même dans l'idée que ce processus de genetic draft doit également affecter les gènes nucléaires- le défaut du jeu de données est que la diversité nucléaire est estimée avec des allozymes. Or, la forte diversité allozymiques des mollusques marins n'est pas interprétée par tous comme étant la conséquence d'une grande taille des populations. Il est encore très présent dans la littérature marine que les locus allozymiques sont maintenus polymorphes par la sélection (Karl, Avise, 1992; Mitton, 1997; Lemaire *et al.*, 2000; Pujolar *et al.*, 2005; Fauvelot *et al.*, 2007; Silva, Skibinski, 2009). Il est donc préférable d'étudier la diversité nucléaire avec des données de séquences, notamment dans les régions non-codantes.

Chez l'huître *Crassostrea gigas*, nous avons étudié le polymorphisme nucléotidique de 41 locus nucléaires (A9). Comme attendu, la diversité s'est révélée très élevée avec en moyenne 1 SNP tout le 60 pb dans les portions codantes et 1 SNP tous les 40 pb dans les régions non codantes. C'est un niveau de polymorphisme équivalent aux espèces les plus polymorphes connues (Tableau 1). Plus surprenant a été la mise en évidence d'une forte contribution des polymorphismes non-synonymes avec un ratio du polymorphisme non-synonyme sur polymorphisme synonyme, π_n/π_s , de 16%. Ce ratio est une mesure de la contrainte sélective sur les protéines (Eyre-Walker, Keightley, 1999; Smith, Eyre-Walker, 2002). Normalement, plus la taille efficace de la population est grande, plus ce ratio est petit car la sélection est plus efficace pour éliminer les mutations

⁵ « genetic draft » pourrait éventuellement se traduire en Français par « remous » ou « turbulence » génétique mais le jeu de mot avec « genetic drift » (dérive génétique) n'y est plus.

faiblement délétères (Kimura, 1983). Cependant, il est préférable d'attendre la confirmation de ce résultat avant de spéculer sur son interprétation (par ex. forte fécondité des bivalves).

Tableau 1: Diversité et contrainte sélective chez quelques espèces connues pour être très polymorphes et chez l'Homme.

	Diversité	Contrainte sur les protéines π_n/π_s
<i>Ciona savignyi</i>	1 SNP / 20pb	0.07
<i>Crassostrea gigas</i>	1 SNP / 40pb	0.16
<i>Drosophila melanogaster</i>	1 SNP / 50pb	0.12
<i>Zea mays</i>	1 SNP / 104pb	0.08
<i>Aedes aegypti</i>	1 SNP / 125pb	0.10
Homme	1 SNP / 1000pb	0.22

Nous avons profité de cette étude pour mettre à jour une corrélation forte entre la diversité non-synonyme et le biais d'usage du code (Figure 2), un résultat qui suggère que la sélection agit sur l'usage du code et qu'elle est corrélée à la sélection sur la protéine (A11). Cette observation est elle aussi un attendu des espèces à grande taille efficace des populations. Une forte diversité nucléotidique a également été observée chez les moules du genre *Mytilus* (Faure, 2007) mais l'existence d'une introgression secondaire d'allèles divergents contribue à augmenter la diversité si bien qu'il est difficile de connaître le niveau de diversité attendu en absence d'introgression. Enfin, les *Bathymodiolus* ont dévoilé une diversité à peine plus faible que les espèces côtières, autour de $\pi=0.01$ en moyenne.

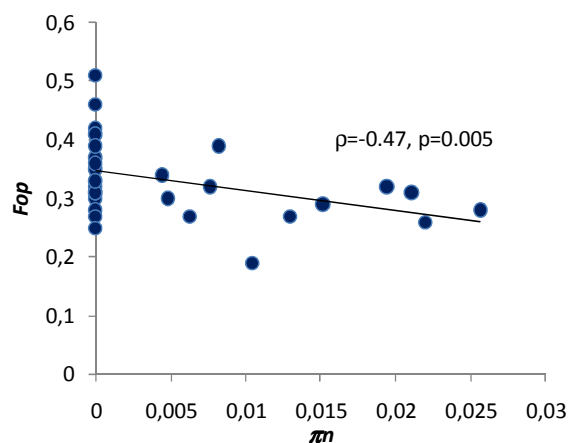


Figure 2: Relation entre la diversité nucléotidique non-synonyme (π) et le biais d'usage du code (F_{op} : fréquence moyenne des codons optimaux) chez *Crassostrea gigas* (d'après A9).

Pour conclure, la diversité nucléotidique observée chez les bivalves est en accord avec des tailles efficaces de populations très grandes. Ça ne veut pas dire que localement et transitoirement

les populations ne puissent pas être de petite taille (Hedgecock, 1994; David *et al.*, 1997), mais à l'échelle de l'espèce un polymorphisme considérable est retenu. Rétrospectivement, ces résultats suggèrent que la forte diversité générale des locus enzymatiques s'explique de la même façon, par la grande taille des populations et non par un maintien du polymorphisme par la sélection.

II.2- La zone d'hybridation en double mosaïque des moules *Mytilus edulis* et *M. galloprovincialis*

Les espèces marines à phase larvaire planctonique sont très souvent associées à de faibles différenciations génétiques sur de grandes distances et les occasions pour qu'une subdivision forte apparaisse entre populations semblent limitées (Palumbi, 1992; Palumbi, 1994). Les généticiens des populations marines ont ainsi eu tendance à se focaliser sur les situations où une structure génétique forte est maintenue. La plupart du temps, ces structures génétiques sont associées à des fronts océaniques qui agissent comme des barrières physiques à la dispersion et sont interprétés comme telles (Burton, 1998; Patarnello *et al.*, 2007). D'autres fois, la structure génétique s'explique par l'existence de mécanismes d'isolement reproductif créant un autre type de barrière au flux génique, les barrières génétiques.

La zone d'hybridation des moules côtières *M. edulis* et *M. galloprovincialis* était un exemple type des paradoxes rencontrés chez les espèces à dispersion planctonique. Quand j'ai commencé mes travaux sur cette zone d'hybridation il y a 10 ans, la structure spatiale de la zone était perçue comme une vaste mosaïque, où des populations des deux formes parentales et des populations hybrides à divers degrés se succédaient sur une très grande distance, entre les côtes Basques et l'Écosse (Skibinski *et al.*, 1983; Coustau *et al.*, 1991; Daguin *et al.*, 2001). Les modèles classiques de zones d'hybridations établis pour des organismes terrestres (Barton, Hewitt, 1985) semblaient impuissants à rendre compte d'une telle situation: comment une structure génétique forte peut-elle se maintenir avec une zone d'hybridation si large et avec autant de mélange au centre? Logiquement, une zone d'hybridation crée une barrière forte quand elle est étroite et quand il y a peu d'hybridation (Barton, 1986; Barton, Bengtsson, 1986). Or chez les moules, il est possible d'observer des populations avec des fréquences alléliques intermédiaires à l'équilibre de Hardy-Weinberg et de liaison, ce qui a amené à la conclusion que l'hybridation était très fréquente⁶. En étudiant la structure spatiale et la composition génétique des populations de cette zone nous avons pu réconcilier l'observation avec les attendus théoriques.

6 L'adjectif « extensive » s'est un jour trouvé collé à « hybridization » dans la littérature sur les zones d'hybridation pour faire « extensive hybridization » (j'utilise le z car l'origine de cette jonction malencontreuse est probablement américaine) et il y est souvent resté attaché ensuite. Si l'hybridation est si fréquente, les marqueurs neutres devraient s'homogénéiser très vite et la zone d'hybridation ne devrait plus être détectable avec de tels marqueurs. Généralement nous n'avons pourtant que des marqueurs neutres à notre disposition, d'où problème!

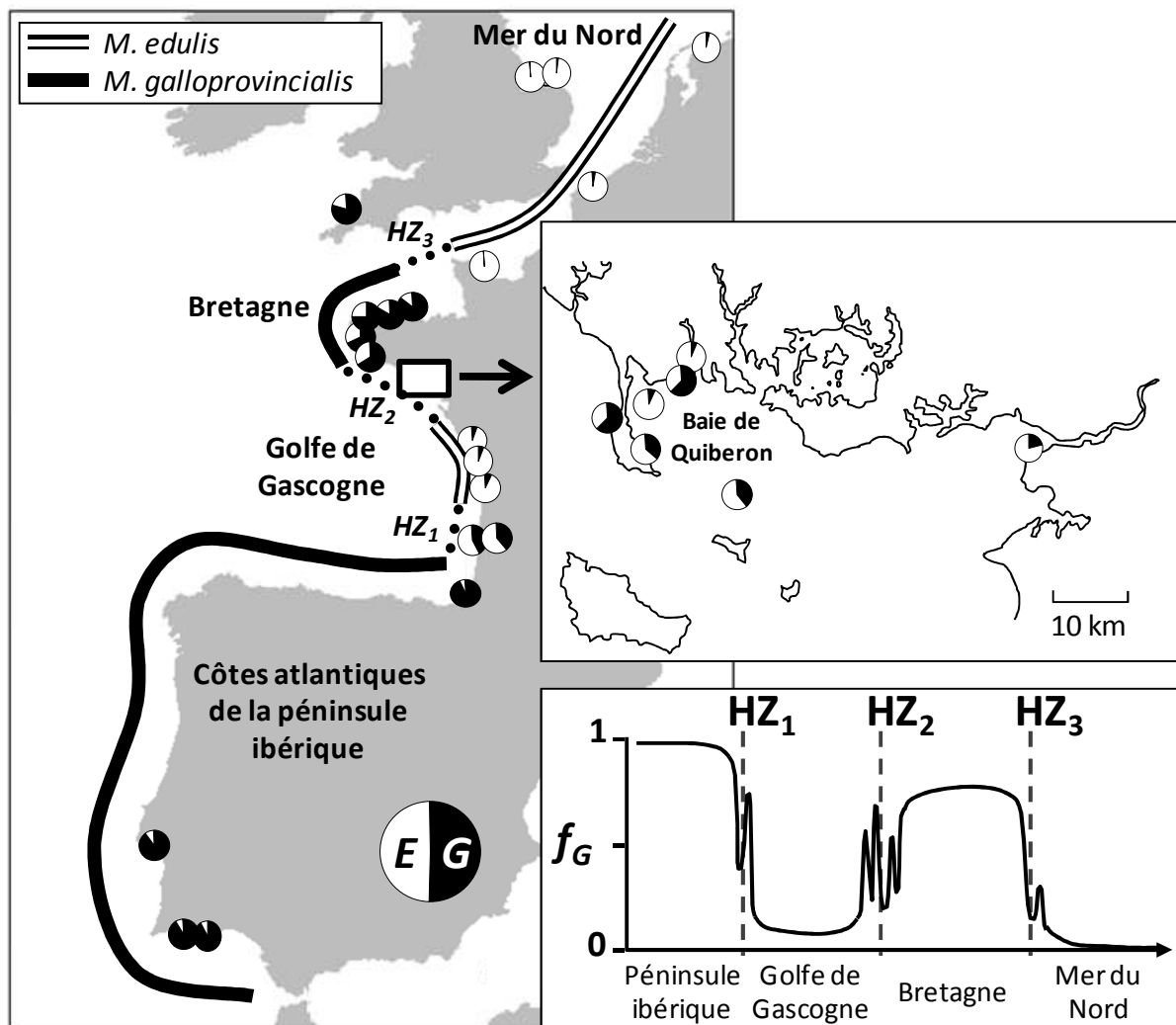


Figure 3 : Structure spatiale de la zone d'hybridation en double mosaïque des moules *Mytilus edulis* et *M. galloprovincialis*. Les camemberts représentent la fréquence des allèles synthétiques E (somme des allèles les plus fréquents dans les populations *M. edulis*), en blanc, et G (somme des allèles les plus fréquents dans les populations *M. galloprovincialis*), en noir, moyennée sur trois locus (d'après A16). A droite de la figure, l'encart du haut est un grossissement de la région de la baie de Quiberon en Bretagne Sud (carré blanc sur la carte de l'Europe), l'encart du bas est un schéma de la structure en double mosaïque représentant des variations spatiales fictive de la fréquence de l'allèle G.

- Plutôt qu'une vaste mosaïque chaotique, la zone est plutôt une succession de plusieurs zones d'hybridation indépendantes et étroites. Le long des côtes françaises, il est possible de distinguer trois zones indépendantes (dans les Landes, en sud Bretagne et en Normandie) qui définissent deux patchs de populations parentales enclavées, un patch *M. edulis* dans le Golfe de Gascogne et un patch *M. galloprovincialis* en Bretagne (A16, Figure 3). A petite échelle, au sein de chacune des trois zones d'hybridation, des populations parentales des deux types peuvent alterner rapidement formant une micro-mosaïque (A19, Figure 3). La structure géographique de la zone est donc une mosaïque à double échelle spatiale: une macro-mosaïque créée par des zones indépendantes et des micro-mosaïques au sein des zones.
- La figure 4 montre la distribution de l'indice hybride (nombre d'allèle *galloprovincialis* d'un individu) et le déséquilibre de liaison dans cinq populations caractéristiques: les *M. edulis* de Mer du Nord et du Golfe de Gascogne, les *M. galloprovincialis* du Portugal et de Bretagne et les populations hybrides de la baie de Quiberon. La variance de l'indice hybride est une mesure du déséquilibre de liaison (Barton, Gale, 1993). Bien qu'une fréquence forte d'allèle *edulis* soit observée dans les populations *M. galloprovincialis* de Bretagne, ces populations sont à l'équilibre de Hardy-Weinberg et de liaison. Elles sont introgressées. Si nous n'avions pas échantillonné de populations hybrides « vraies » qui sont la presque superposition des populations parentales locales (Baie de Quiberon, Figure 4) et qui sont finalement assez rares, nous aurions eu plus de mal à interpréter la composition génétique des échantillons de Bretagne. Nous aurions peut-être conclu que l'hybridation était fréquente et serions resté avec notre paradoxe en place. En réalité l'hybridation est rare et elle est très difficile à détecter dans la nature. Par contre les niveaux d'introgression des patchs enclavés de populations parentales sont forts, notamment en Bretagne. Distinguer les deux types de populations, parentales introgressées ou hybrides vraies, est important car il est inutile d'aller étudier les mécanismes d'isolement reproductif dans les populations introgressées. Ces populations sont polymorphes pour les locus neutres mais fixées pour les allèles d'un taxon aux gènes d'isolement. Malheureusement, des auteurs continuent à considérer ces populations comme hybrides et y trient les individus sur la base de leurs génotypes aux marqueurs pensant inférer leurs génotypes aux gènes d'isolement (*i.e.* leur statut spécifique) alors qu'ils sont dans une situation intra-spécifique en fait. Je reviendrai dans le chapitre III.1.2 sur ce problème et sur l'étude des mécanismes d'isolement reproductif.

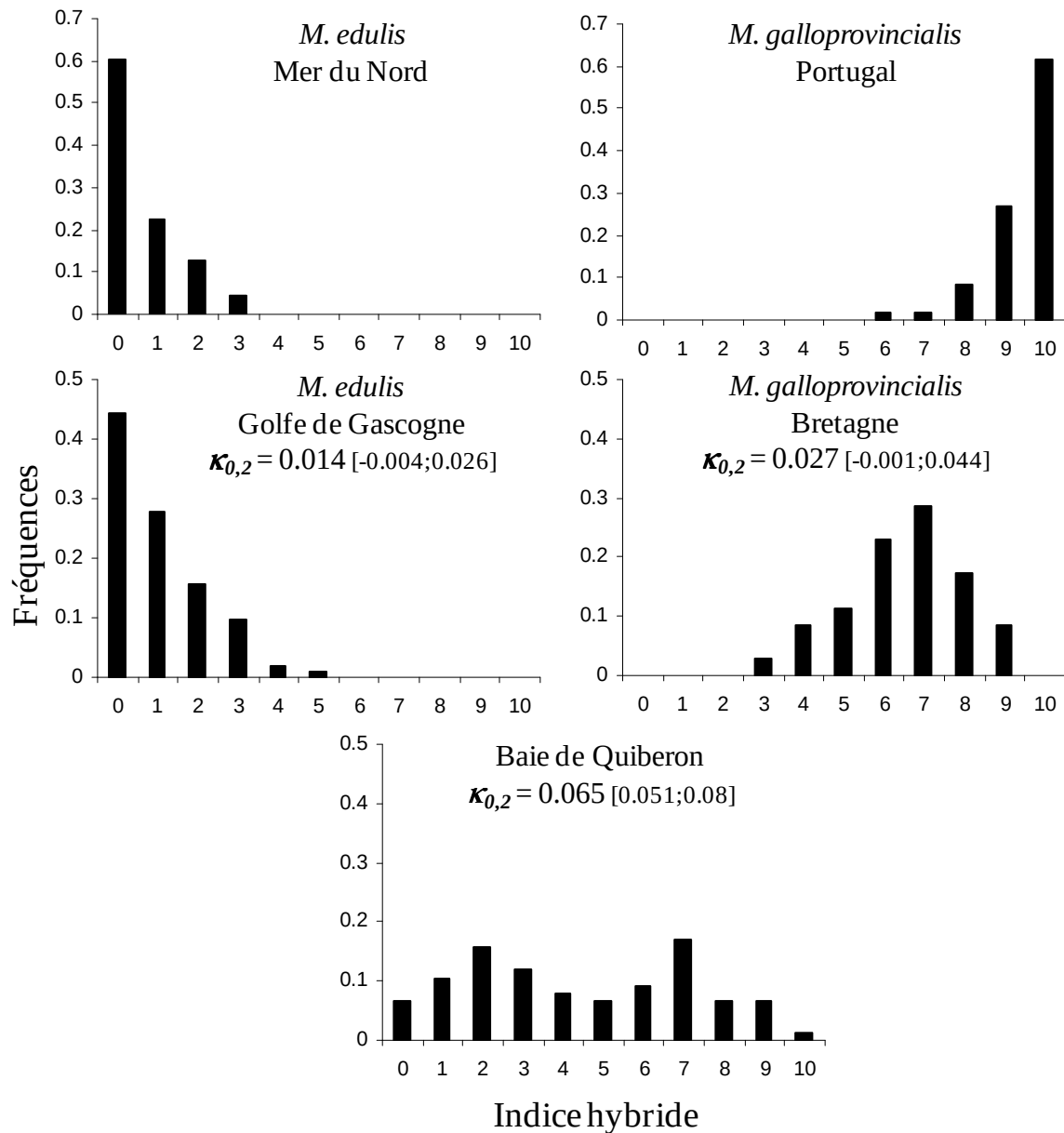


Figure 4 : Distribution de l'indice hybride et $\kappa_{0,2}$ [intervalle de confiance à 95%], le déséquilibre de liaison multi-locus de Barton (2000), dans différents types de populations, les *M. edulis* de Mer du Nord et du Golfe de Gascogne, les *M. galloprovincialis* du Portugal et de Bretagne et les populations hybrides de la Baie de Quiberon (d'après A19).

Finalement, il n'y avait pas ou peu de paradoxe dans la structure génétique forte observée entre *M. edulis* et *M. galloprovincialis*: trois zones indépendantes et étroites dans lesquelles existe peu d'opportunité d'hybridation se cumulent pour former une barrière forte au flux génique. J'ajouterai que les études sur cette zone ont eu tendance à sélectionner les marqueurs les plus

discriminants entre les deux espèces⁷, si bien que la différenciation a été artificiellement gonflée (A4). L'étude de locus choisis aléatoirement dévoile une différenciation moins forte avec beaucoup de polymorphisme partagé (Faure 2007).

Pour l'instant je n'ai pas encore vraiment parlé de sélection. Cependant, la difficulté posée par l'étude de backgrounds génétiques définis par des gènes d'isolement à l'aide de marqueurs neutres est un premier jalon du message principal de cette HDR. A partir du moment où il y a un peu de recombinaison efficace entre génomes (*i.e.* introgression) les corrélations entre locus ne sont pas maximales et le génotype des marqueurs neutres ne renseigne qu'imparfaitement sur le génotype des gènes d'isolement: « gène $\neq$ génome ».

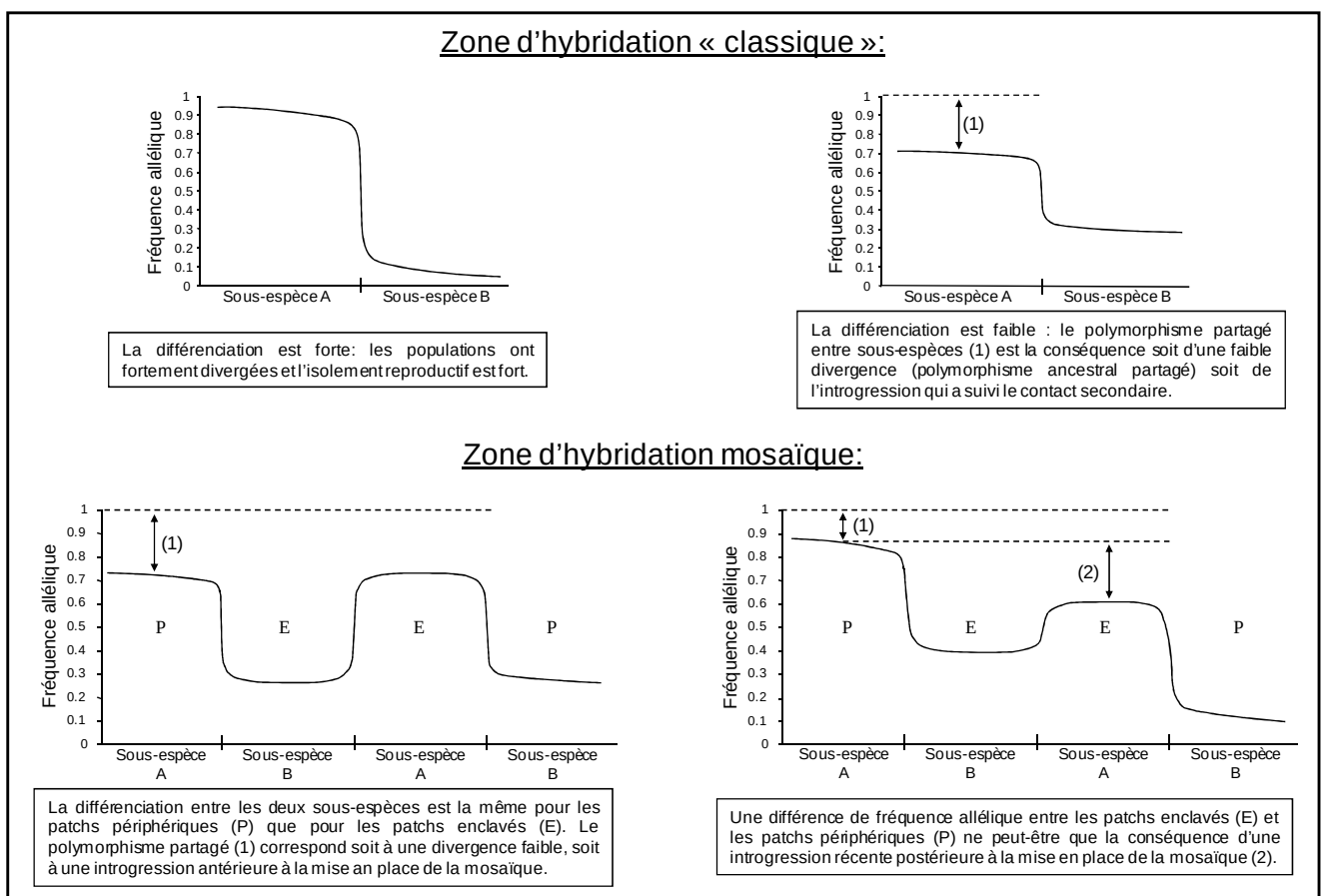


Figure 5 : Introgression *versus* polymorphisme ancestral partagé dans les zones d'hybridation classiques (mono-clinales) et les zones d'hybridation mosaïques (poly-clinales, d'après Faure 2007).

7 C'est ce qu'il est préférable de faire quand on souhaite étudier les mécanismes d'isolement reproductif.

II.3- Histoire du flux génique dans deux zones d'hybridation marines

Nous n'avons généralement pas accès aux gènes d'isolement dans une zone d'hybridation et il est nécessaire d'utiliser des locus marqueurs. L'analyse est souvent compliquée par la difficulté de faire la part entre deux hypothèses qui expliquent toute deux l'existence de polymorphismes partagés entre populations de part et d'autre d'une zone de contact : (i) l'introgession : le locus était bien différencié avant le contact mais des flux géniques postérieurs au contact ont effacé en partie ce contraste, (ii) la faible divergence (ou polymorphisme ancestral partagé) : les fréquences alléliques à ce locus, pour une raison ou pour une autre, était faiblement différenciées au moment du contact (Figure 5).

Ces dernières années, les études sur les échanges géniques entre génomes différenciés ont intégré les développements modernes de la génétique des populations et de la phylogénie appliqués à l'analyse des séquences d'ADN (Hey, Machado, 2003). L'application des méthodes d'analyse traditionnelle n'est pas immédiate et l'analyse de la divergence sans rupture totale du flux génique est assez récente (Wakeley, Hey, 1997; Kliman *et al.*, 2000; Nielsen, Wakeley, 2001; Hey, Nielsen, 2004; Hey, Nielsen, 2007). Cette nouvelle discipline a été baptisée la "génétique des populations de la divergence" ("divergence population genetics", DPG, Kliman *et al.*, 2000). Le projet « Bivalvomix: Génomique évolutive des bivalves marins » avait pour deuxième objectif d'appliquer ces nouvelles méthodes de la DPG à des zones d'hybridation de bivalves marins.

II.3.1- Spéciation dans les abysses: analyse multi-locus de la divergence et du flux génique entre deux espèces de moules hydrothermales

La thèse de Baptiste Faure a en partie porté sur la zone d'hybridation des moules hydrothermales *Bathymodiolus azoricus* et *B. puteoserpentis* (A1, Annexe 1). L'idée était que le milieu hydrothermal était un bon endroit pour faire de la DPG. Tout d'abord, les caractéristiques environnementales originales de l'hydrothermalisme sont la source de spéculations sur l'importance de l'adaptation dans les processus de spéciation. Ensuite, les populations profondes n'ont pas été autant perturbées que les populations de surface par les oscillations climatiques du quaternaire qui ont eu des conséquences importantes sur la structure génétique et la spéciation (Hewitt, 1996 ; 2004). Enfin, comme pour le milieu marin côtier, les opportunités d'isolement géographique semblent limitées. Il est important de préciser ici que les *Bathymodiolus* possèdent une phase de dispersion par larves planctonotrophes. Nous espérons pouvoir reconstruire l'histoire des flux géniques au cours de la divergence des populations et identifier soit la trace d'un contact secondaire soit au contraire d'un flux génique ininterrompu (Nosil, 2008). Le polymorphisme de séquence nucléotidique de sept locus nucléaires et un locus mitochondrial a été ajusté au modèle d'isolement avec migration de Hey et Nielsen (Hey, Nielsen, 2007). L'analyse a montré (i) que les deux espèces avait commencé à diverger il y a 760 mille ans, (ii) que la taille efficace de la population de *B. azoricus* était de deux à cinq fois plus grande que celle de la population de *B. puteoserpentis* et de la population ancestrale, (iii) qu'il existe ou a existé du flux de gènes entre les deux espèces et (iv) que ce flux était principalement asymétrique de *B. azoricus* vers *B. puteoserpentis*. L'asymétrie du flux

génique peut être expliquée de diverses façons. La théorie des zones d'hybridation prédit une barrière génétique asymétrique lorsque les densités de populations (c'est à dire les tailles de populations en modèle discret) sont différentes (Barton, 1986) ou lorsque que la zone d'hybridation est en mouvement (Barton, Hewitt, 1985; Buggs, 2007). Ces deux explications peuvent être valides ici puisque *B. azoricus* est connu pour avoir des populations de plus grandes tailles et plus denses que *B. puteoserpentis* et que les courants abyssaux pourraient favoriser un déplacement de la zone vers le nord. Il est aussi possible que cette asymétrie soit la conséquence de mécanismes d'isolement reproductif générant une asymétrie de l'hybridation, soit générale à l'échelle du génome, soit localement en fonction du ou des gènes d'isolement principalement responsables de la barrière agissant sur les locus étudiés. En y regardant de plus près, un des huit locus (17% du génome) montre une légère tendance pour une asymétrie du flux génique dans le sens opposé des sept autres.

Le décryptage de l'histoire des flux géniques s'est avéré une tâche beaucoup plus délicate qu'attendu pour deux raisons: (i) Tout d'abord le modèle sous-jacent est un modèle avec flux génique constant au cours du temps, homogène sur tous les locus et avec des hypothèses fortes sur la stabilité démographique ou la neutralité des locus (absence de sélection qu'elle soit directe ou indirecte). Or, nos populations dévient indéniablement de ces hypothèses : la barrière est semi-perméable (les flux géniques sont différents entre locus), l'équilibre démographique n'est pas avéré, des indices de sélection directe (A8) ou indirecte (locus SAHH, Annexe 1) ont été identifiés. (ii) Le paramètre inféré par statistique bayésienne est l'âge moyen des événements de migration. La méthode délivre une distribution postérieure de ce paramètre, paramètre qui est une moyenne. Cette distribution est trop facilement interprétée comme la distribution des événements de migration (Nosil, 2008). En fait, la distribution nous donne l'incertitude sur l'estimation du paramètre sous le modèle considéré (statistique bayésienne) mais le paramètre n'est autre que la moyenne des âges des événements de migration. Pour mieux comprendre, prenons l'exemple factice de l'existence de deux contacts secondaires d'intensité et de durée identiques, l'un ancien (disons 500 mille ans) et l'autre plus récent (disons 50 mille ans), une histoire complexe qui n'est pas considérée dans le modèle à juste titre (nous n'en finirions jamais de complexifier les modèles et de ne plus savoir ce qu'on infère). L'âge moyen des événements de migration serait intermédiaire aux deux contacts (275 mille ans), une période durant laquelle il n'y avait pas de contact. Nous pourrions très bien avoir des données qui soutiennent fortement ce résultat si bien que la distribution formerait un pic net centré sur cette valeur.

Appliqué à nos données nous avons estimé que l'âge moyen des événements de migration était autour de 400 mille ans (pour une divergence de 750 mille ans), ce qui n'est pas en accord apparent avec l'hypothèse du contact secondaire. Nous avons alors cherché d'autres indices d'une période sans flux génique dans l'histoire de divergence (A1). L'indice le plus flagrant a été observé sous la forme de recombinaison intragénique au locus GF1B. Comme la recombinaison ne peut se produire qu'entre allèles ségréguant dans la même population, la trace d'une reprise récente de la recombinaison entre deux clades d'allèles divergents est un bon indice d'une introgression secondaire. Nous avons reconstruit la généalogie ainsi que les événements de recombinaison du

locus GF1B (Figure 6). La recombinaison entre deux clades d'allèles divergents n'est observée que dans les derniers 10-20% les plus récents de la généalogie, ce qui permet une estimation très grossière d'une reprise de la recombinaison depuis 50-200 mille ans. A ce stade, il est difficile de donner une réponse tranchée sur la question de l'histoire du flux génique pendant la divergence de ces deux espèces. De mon point de vue cependant, le scénario d'un arrêt complet du flux génique au moins pendant la phase initiale de la divergence et d'un contact secondaire assez récent avec formation d'une barrière semi-perméable, semble plus parcimonieux.

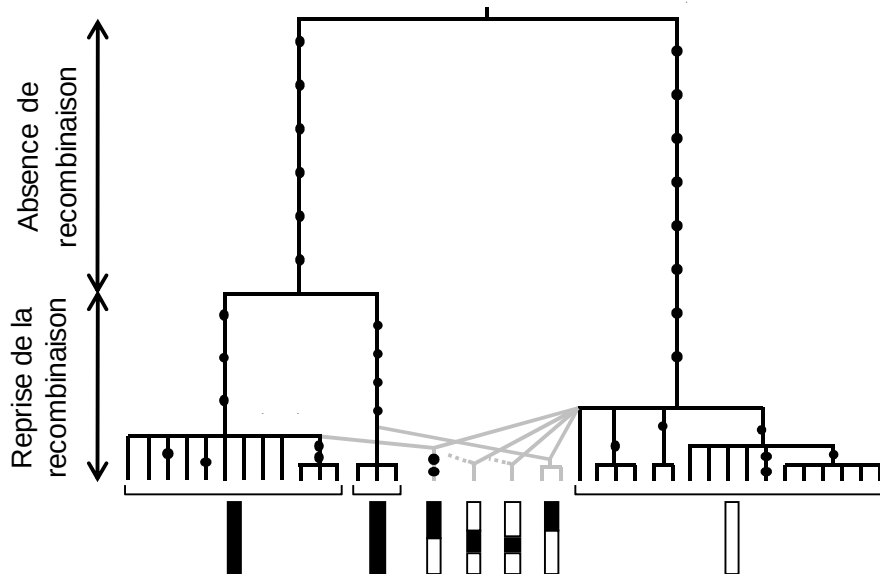


Figure 6 : Généalogie et histoire mutationnelle (points) et recombinaisonnelle (traits gris) du locus GF1B chez *Bathymodiolus azoricus* et *B. puteoserpentis*.

II.3.2- Analyse du contact secondaire entre les moules *Mytilus edulis* et *M. galloprovincialis*

La thèse de Matthieu Faure a porté sur la zone d'hybridation des moules côtières *Mytilus edulis* et *M. galloprovincialis*. Pour cette zone, le scénario d'un contact secondaire ne faisait que peu de doute bien qu'il n'ait jamais été formellement démontré. De plus, une variation forte des niveaux de différenciation était connue mais il n'était pas avéré que cette variation soit la conséquence d'une hétérogénéité de l'introgression après contact (barrière semi-perméable) plutôt qu'une hétérogénéité forte de la différenciation avant contact (Figure 5). La structure mosaïque de la zone (Figure 3) nous a tout de même offert l'opportunité d'identifier l'introgression postérieure à la mise en place de la mosaïque. Pour ce faire, il faut comparer les fréquences alléliques des patches périphériques aux fréquences alléliques observées dans les patches enclavés. Le différentiel de différenciation est interprété comme la conséquence de l'introgression (Figure 5). La figure 7A présente les fréquences alléliques de 13 locus dans les quatre patches des côtes françaises (A2, Annexe 2). La figure 7B montre le différentiel de fréquence allélique entre les populations *M. galloprovincialis* et *M. edulis*,

Δp , entre les patchs enclavés en fonction du Δp entre les patchs périphériques. Tous les points sont en dessous de la première diagonale, montrant ainsi l'existence d'une introgression plus avancée dans les patchs enclavés de la mosaïque. Cependant, les écarts à la première diagonale sont très variables entre locus, dévoilant une hétérogénéité génomique de l'introgression et démontrant la nature semi-perméable de la barrière. Le locus pour lequel le différentiel d'introgression est le plus marqué est le locus *mac-1*, en gras sur la figure 7, un locus dont le comportement surprenant, à savoir une introgression très forte des populations *M. galloprovincialis* de Bretagne contre une introgression faible des *M. galloprovincialis* du Portugal, avait été identifié précédemment (A16).

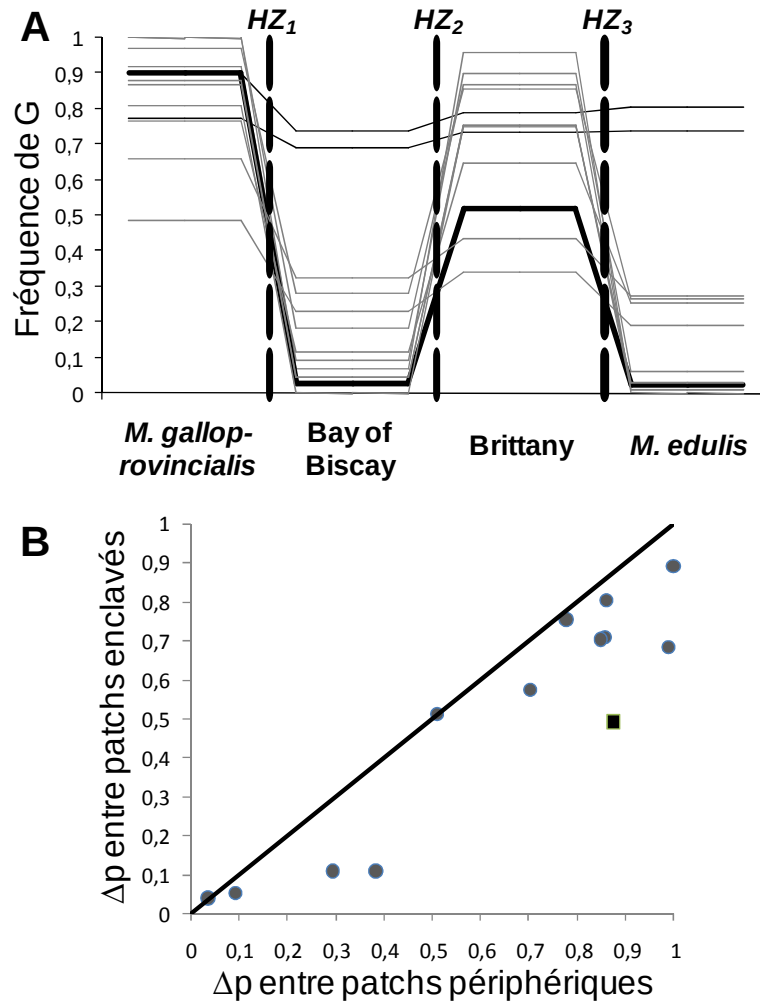


Figure 7 : Différentiels d'introgression entre les patchs enclavés et les patchs périphériques de la zone hybride mosaïque des moules *Mytilus edulis* et *M. galloprovincialis*. (A) Fréquence de l'allèle synthétique G de 13 locus dans les quatre patchs de la zone (A2). Le trait gras correspond au locus *mac-1* qui dévoile une introgression particulièrement forte dans le patch enclavé de Bretagne comparé au patch périphérique. (B) Différentiel de fréquence allélique entre les populations *M. galloprovincialis* et *M. edulis*, Δp , des patchs enclavés en fonction du Δp des patchs périphériques. Le carré noir correspond au locus *mac-1*.

Malgré les atouts de la structure mosaïque, le manque d'information sur les relations phylogénétiques des allèles ne permettait pas de reconstruire l'histoire de la divergence et du flux génique. Nous avons donc ajusté le modèle d'isolement avec migration aux données de polymorphisme de séquence nucléotidique de quatre locus nucléaires et un locus mitochondrial (Annexe 2). Il n'a fait aucun doute que les hypothèses du modèle étaient violées et l'analyse multi-locus s'est révélée impossible puisque des paramètres supposés identiques entre locus dans le modèle ne l'étaient manifestement pas. Tout d'abord, plusieurs des locus utilisés ont dévoilé des indices de l'effet de la sélection indirecte et parfois peut-être même de la sélection directe (Annexe 2, Annexe 3, Riginos, McDonald, 2003). De ce fait les niveaux de diversité étaient très différents entre locus. Ensuite, la semi-perméabilité de la barrière s'est traduite par des estimations des taux de migration très différents et dans des sens opposés pour les différents locus (Tableau 2). Cependant, malgré ces différences entre locus, les estimations du temps de divergence et de l'âge moyen des événements de migration se sont révélées être très similaires entre locus.

Tableau 2: Estimation bayésienne des paramètres démographiques du modèle d'isolement avec migration de Hey and Nielsen (2007)

Locus	Ne_E	Ne_G	Ne_A	$m_{G \rightarrow E}$	$m_{E \rightarrow G}$	t	$mtime$
<i>Mytilin B</i>	$0.3 \cdot 10^6$	$0.2 \cdot 10^6$	$0.4 \cdot 10^6$	0.01	0	$0.8 \cdot 10^6$	$26 \cdot 10^3$
<i>MGD2</i>	$1,6 \cdot 10^6$	$0.4 \cdot 10^6$	$0.4 \cdot 10^6$	0.04	0	$1.7 \cdot 10^6$	$36 \cdot 10^3$
<i>EF1α</i>	$0.3 \cdot 10^6$	$1.0 \cdot 10^6$	$0.7 \cdot 10^6$	0.0003	0	$1.5 \cdot 10^6$	0
<i>Lysin M7</i>	$0.1 \cdot 10^6$	$0.1 \cdot 10^6$	$0.2 \cdot 10^6$	0	0.005	$1.1 \cdot 10^6$	$12 \cdot 10^3$
<i>mtDNA</i>	$1,5 \cdot 10^6$	$0.6 \cdot 10^6$	$0.6 \cdot 10^6$	0	0.0009	$0.8 \cdot 10^6$	$59 \cdot 10^3$
5-loci	$0.5 \cdot 10^6$	$0.4 \cdot 10^6$	$0.6 \cdot 10^6$	0.002	0.0003	$1.1 \cdot 10^6$	$25 \cdot 10^3$

Ne_E , taille efficace de la population *M. edulis*; Ne_G , taille efficace de la population *M. galloprovincialis*; Ne_A , taille efficace de la population ancestrale; $m_{G \rightarrow E}$, migration de *galloprovincialis* vers *edulis*; $m_{E \rightarrow G}$, migration d'*edulis* vers *galloprovincialis*; t , temps de divergence; $mtime$, âge moyen des événements de migration.

Le temps de divergence est estimé à 1 [0.5, 2] million d'année et l'âge moyen des événements de migration à 25 [0-200] mille ans, en accord avec l'hypothèse du contact secondaire récent. En fait, ces estimations sont conformes à celles que nous aurions pu faire directement sur les mesures de divergence par des mécaniques d'horloge moléculaire car les phylogénies des allèles dévoilent toutes une trace d'une divergence longue sous la forme de deux clades d'allèles divergents. Cependant, suivant les locus l'introgession et la recombinaison ont plus ou moins commencé à effacer la trace de cette divergence (Figure 8). Comme pour le locus GF1B des *Bathymodiolus*, le locus *Mytilin B* des *Mytilus* dévoile une reprise récente de la recombinaison entre allèles divergents (Figure 9).

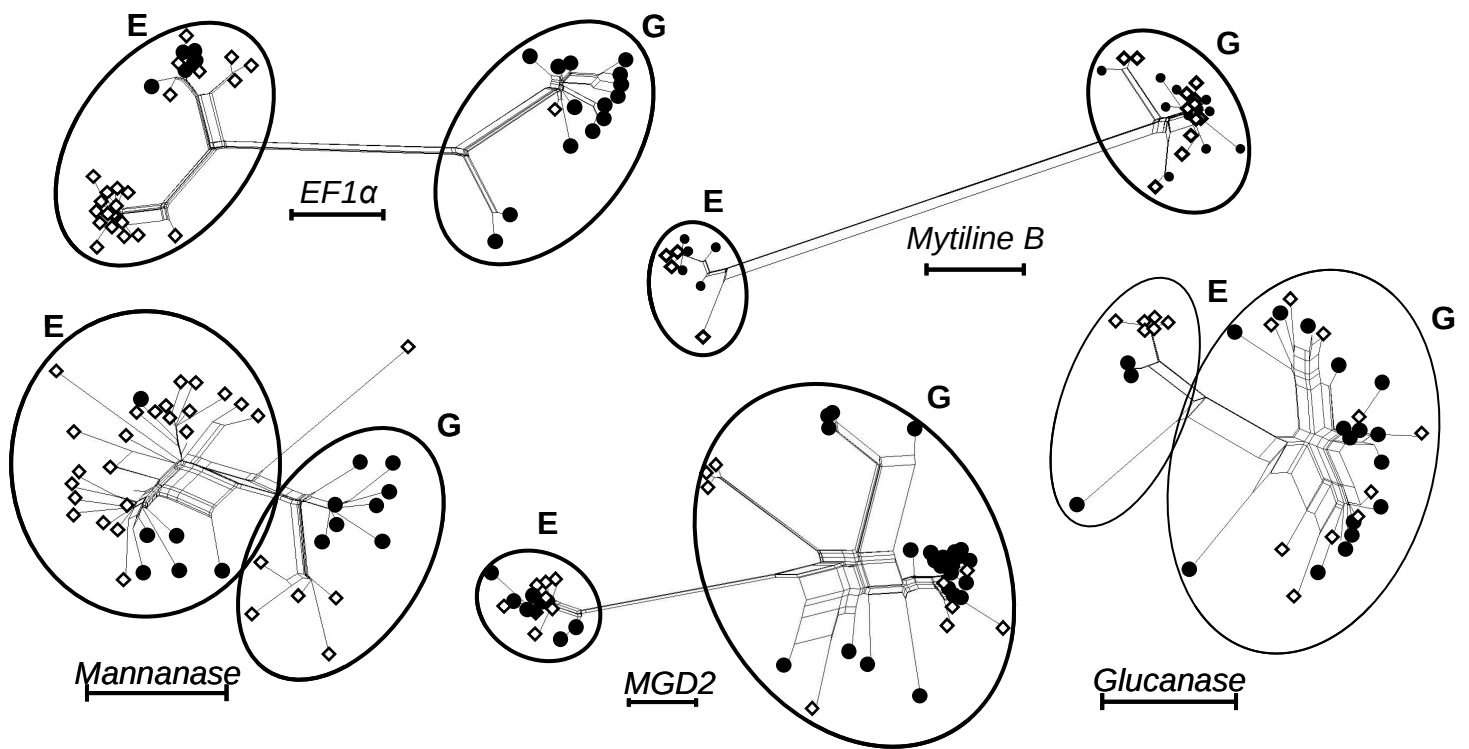


Figure 8: Réseaux phylogénétiques réalisés par la méthode du NeighborNet (Bryant, Moulton, 2004; Huson, Bryant, 2006) sur cinq gènes analysés chez les moules *Mytilus edulis* et *M. galloprovincialis*. Les allèles échantillonnés dans les patches *M. galloprovincialis* sont représentés par des points noirs, ceux échantillonnés dans des patches *M. edulis* sont représentés par des losanges blancs. L'échelle des longueurs de branches indiquée sous les réseaux correspond à une divergence de 1% (d'après Faure 2007).

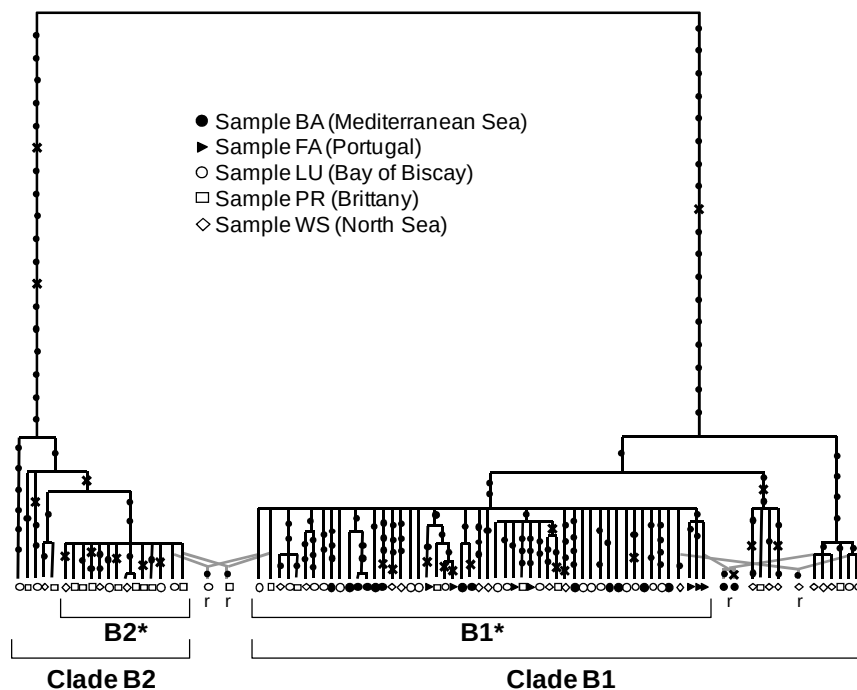


Figure 9: Généalogie et histoire mutationnelle (points : mutations silencieuses, croix : mutations non-synonymes) et recombinaisonnelle (traits gris) du locus *Mytilin B* chez *Mytilus edulis* et *M. galloprovincialis* (d'après A2).

III- Marqueurs moléculaires et sélection

Dans le premier chapitre, j'ai présenté des travaux qui n'avaient pas vraiment pour but d'étudier la sélection mais plutôt de décrire la diversité génétique et sa distribution spatiale, et d'inférer l'histoire des populations. Une fois bien appréhendées les caractéristiques biologiques, écologiques et génétiques du modèle ainsi que l'histoire des populations, il est possible de s'intéresser à la sélection. Comme annoncé en introduction, j'ai principalement adopté une approche indirecte sur des données de polymorphisme. Je présenterai ces approches dans un premier sous-chapitre. Pendant mon post-doc avec Adam Eyre-Walker à l'université du Sussex, je me suis intéressé à une approche directe de l'étude de la sélection sur des données incluant des substitutions dans la divergence entre espèces. Bien qu'en marge du message principal de cette HDR car abordant une échelle de temps différente, ces résultats sont cependant importants pour la compréhension des mécanismes agissant à des échelles de temps plus courtes puisqu'ils renseignent sur la fréquence à laquelle des fixations favorables balayent la diversité des génomes. Dans un troisième sous-chapitre je présenterai des résultats qui laissent suspecter la découverte de polymorphismes adaptatifs dans des peptides antimicrobiens de moules sans toutefois pouvoir le conclure avec assurance.

III.1- Relations indirectes

III.1.1- La consanguinité générale du génome explique les corrélations hétérozygotie / valeur sélective.

Il existe une littérature abondante sur la démonstration d'une corrélation entre l'hétérozygotie individuelle multilocus (le nombre de locus à l'état hétérozygote chez un individu) et des traits reliés à la valeur sélective comme par exemple la croissance ou la survie (David, 1998). La plupart des études des corrélations entre l'hétérozygotie et des traits phénotypiques a été réalisée à l'aide de marqueurs allozymiques. Ceci était à l'origine d'un débat sur les causes de ces corrélations. Il est en effet difficile de trancher entre les hypothèses qui considèrent les allozymes comme l'agent responsable de la corrélation (superdominance directe), et les hypothèses qui les considèrent comme des marqueurs neutres des conditions génétiques responsables de la corrélation (superdominance associative). Mes travaux sur l'huître et la crevette ont permis de montrer que l'hétérozygotie de locus réputés neutres comme les microsatellites pouvait également être corrélée à des traits phénotypiques (A27, A23). Or, dans une grande population panmictique, l'hétérozygotie des marqueurs neutres n'est pas corrélée à celle des autres gènes (Jacquard, 1975; Chakraborty, 1987), y compris ceux qui déterminent la valeur sélective. De telles corrélations ne peuvent de ce fait apparaître que dans des conditions populationnelles bien particulières.

Une approche théorique a permis de montrer que ces corrélations étaient la conséquence de la consanguinité et de sa variation interindividuelle dans les populations (A22). Le coefficient de détermination entre l'hétérozygotie individuelle à un marqueur, H (H prends la valeur 0 si l'individu est homozygote et la valeur 1 si l'individu est hétérozygote), et la valeur sélective, W , peut s'écrire:

$$r^2(H,W) = \frac{E(H)}{1-E(H)} \frac{\sigma^2(f)}{1-E(f)}$$

où $E(H)$ est l'espérance de l'hétérozygotie individuelle (c'est à dire l'hétérozygotie de la population), $\sigma^2(f)$ est la variance et $E(f)$ l'espérance de la consanguinité individuelle. Cette formule est générale, elle est obtenue sans faire d'hypothèse sur la structure de la population. Dans ce modèle la liaison physique des gènes peut être intégrée comme une composante de la variance de consanguinité (Weir *et al.*, 1980). Ce résultat montre l'importance de l'existence de variation interindividuelle de consanguinité pour générer une corrélation. Cependant, à l'équilibre démographique cette variance est très faible (Jacquard, 1975), tout le monde à la même consanguinité (la même hétérozygotie que ce soit aux locus marqueurs ou aux locus de fitness). Cependant, dans les phases initiales d'un goulot démographique, la variance de consanguinité générée par le hasard des unions est grande (Jacquard, 1975) et $r^2(H,W)$ est plus grand qu'à l'équilibre (Figure 10, A22). D'autres façons de générer une variance de consanguinité sont l'autofécondation partielle (Charlesworth, 1991; David, 1999) ou la phase initiale d'un mélange de populations ayant divergé en allopatrie (Tsitrone *et al.*, 2001).

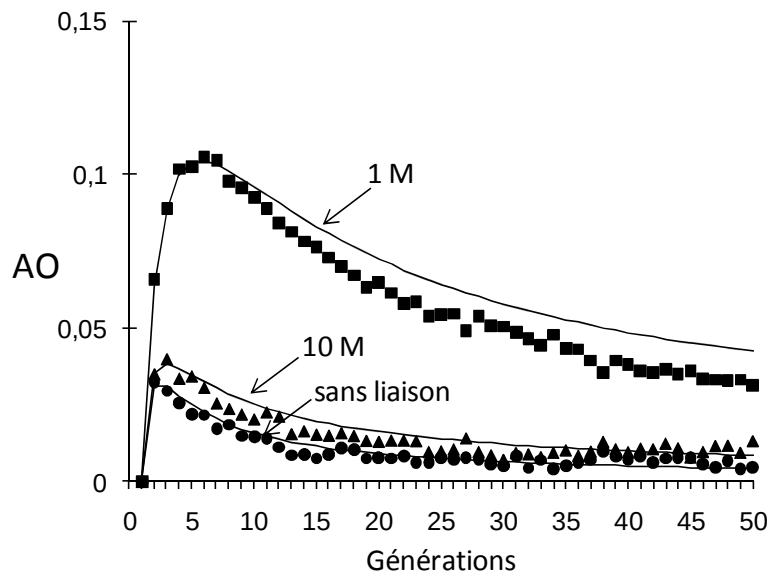


Figure 10 : Effet de l'âge du goulot démographique et de la taille du génome (c'est-à-dire de la liaison physique) sur l'avantage apparent des hétérozygotes à un marqueur neutre (AO). Les résultats de simulations en absence de liaison, pour un génome de 10 Morgans et de 1 Morgan sont représentés respectivement par des cercles, des triangles et des carrés. Les courbes sont les attendus analytiques. La taille de la population est $N=20$, le mode de reproduction est la monogamie, le taux de mutations délétères génomique est $U=1$, le coefficient de dominance est $h=0.3$, le coefficient de sélection contre les homozygotes délétères est $s=0.05$, l'hétérozygotie initiale du marqueur est $H_0=0.9$.

Un résultat qui me tient à cœur est que la liaison physique n'est pas une cause suffisante à l'existence d'une corrélation (A22, Ohta, 1973)⁸ et son effet est faible pour des génomes de taille réaliste (>1 Morgan, A22). Imaginons comme le suggère les études sur le fardeau génétique (ex. A27, revue dans Eyre-Walker, Keightley, 2007) que les variations de valeur sélective sont la conséquence d'un grand nombre de mutations délétères réparties aléatoirement dans le génome. Un marqueur génétique sera sous l'influence d'un grand nombre de délétères non liés (sur les autres chromosomes par exemple) qui auront individuellement un effet faible, et sous l'influence d'un tout petit nombre de délétères liés physiquement qui auront individuellement un effet fort sur le marqueur. Il suffit d'avoir quatre fois plus de délétères non liés pour que la somme des effets faibles des délétères non liés dépasse la somme des effets forts des délétères liés, ce qui est par exemple le cas des espèces avec plus de cinq chromosomes (A22).

Depuis ces travaux, une hypothèse alternative est née dans la littérature. Le point de départ est comme énoncé plus haut que dans une population d'assez grande taille et à l'équilibre démographique, l'homozygotie d'un marqueur génétique renseigne très peu sur l'homozygotie globale du génome, c'est-à-dire la consanguinité individuelle. Au lieu de remettre en cause l'hypothèse d'équilibre et de rechercher des structures de populations permettant de générer ces corrélations, comme nous l'avons fait (A22), il a été suggéré que la dépression de consanguinité ne pouvait pas être la cause des corrélations hétérozygotie / valeur sélective et que certains locus étaient fortement liés à un locus sous sélection balancée (Balloux *et al.*, 2004; Pemberton, 2004; Hansson, Westerberg, 2008). Le paradoxe est que le nombre de locus sous sélection balancée devrait alors être incroyablement grand pour qu'on puisse tomber aussi facilement juste à côté avec une poignée de marqueurs microsatellites. Or, la sélection balancée n'est pas un mode de sélection très répandu dans les génomes (Charlesworth, 2006). L'erreur de cette littérature est qu'une corrélation d'hétérozygotie entre deux locus même liés n'est rien d'autre que de la consanguinité. De plus, la liaison physique n'est pas génératrice de déséquilibre de liaison mais simplement promotrice (pour simplifier, la liaison favorise la genèse et ralentit la dissipation du déséquilibre de liaison). Or, les processus générateurs de consanguinité localement sur le chromosome génèrent inévitablement de la consanguinité globale. Ces processus sont la consanguinité aléatoire (aussi appelée dérive génétique, Hartl, Clark, 1989), la consanguinité systématique, comme par exemple l'autofécondation partielle (Malécot, 1969), ou la structure des populations (Barton, 2000a). Malheureusement, comme la corrélation est souvent faible et que la variance de l'effet entre locus est énorme, on peut avoir l'impression que certains locus sont fortement corrélés au phénotype alors que d'autres ne le sont pas (par ex. Lieutenant-Gosselin, Bernatchez, 2006). Cependant, cette observation ne résiste pas à l'analyse statistique appropriée de David (1997).

⁸ Ohta (1973) disait "detrimentals on different chromosomes will also contribute to increase the associative overdominance, roughly by the amount $L_i/2N_e c = 0.1\%$ per chromosome", alors qu'elle inférait juste avant que les délétères du chromosome du marqueur avait une superdominance associative de 0.3% (3 fois plus seulement).

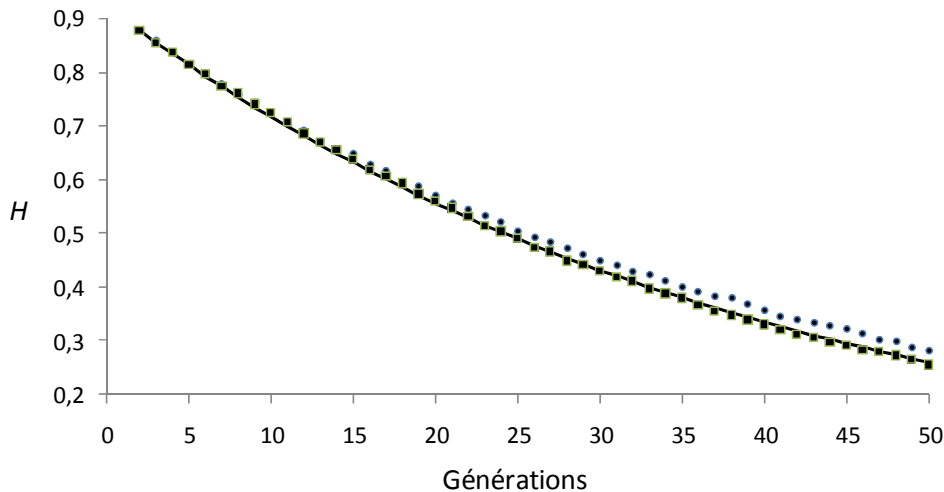


Figure 11 : Evolution de l'hétérozygotie lors d'un goulot démographique soutenu. La courbe est l'attendu neutre, les carrés présentent les résultats de simulations pour un génome de 10 Morgans et les cercles les résultats de simulations pour un génome de 1 Morgan. Les paramètres sont les même qu'en figure 10.

La sélection indirecte n'est pas la sélection. Posons ici un deuxième jalon au message de cette HDR. La sélection observée dans les études des corrélations hétérozygotie / valeur sélective n'est qu'apparente. Si à une génération donnée, le génotype neutre procure une information sur les gènes sélectionnés, ces derniers n'ont pas d'effet directionnel sur l'évolution des marqueurs neutres. En absence de liaison physique, l'hétérozygotie n'est pas maintenue par la sélection (Figure 11). En présence de liaison physique, l'effet des mutations délétères, si elles sont assez densément réparties sur le chromosome, est tout de même d'augmenter légèrement la taille efficace du locus neutre dans une petite population (Figure 11, Ohta, 1971; Latter, 1998; Pamilo, Palsson, 1998; Palsson, Pamilo, 1999). Dans les grandes populations, l'effet est inverse et les délétères liés à un locus neutre diminuent sa taille efficace, c'est la sélection d'arrière plan (background selection, Charlesworth *et al.*, 1993; Charlesworth, 1994).

III.1.2- Mécanismes d'isolement reproductif en milieu marin

Les zones d'hybridation sont souvent présentées comme des laboratoires naturels pour étudier la spéciation (Hewitt, 1988). En effet, elles correspondent à des situations intermédiaires dans lesquelles des différences génétiques se maintiennent, au moins temporairement, malgré l'existence de flux génique. Elles sont comme des arrêts sur image de spéciations en cours, qu'elles se soldent ou non par la formation de deux espèces. Les zones d'hybridation nous donnent ainsi l'opportunité précieuse de pouvoir étudier les mécanismes responsables du maintien de la cohésion génétique des espèces.

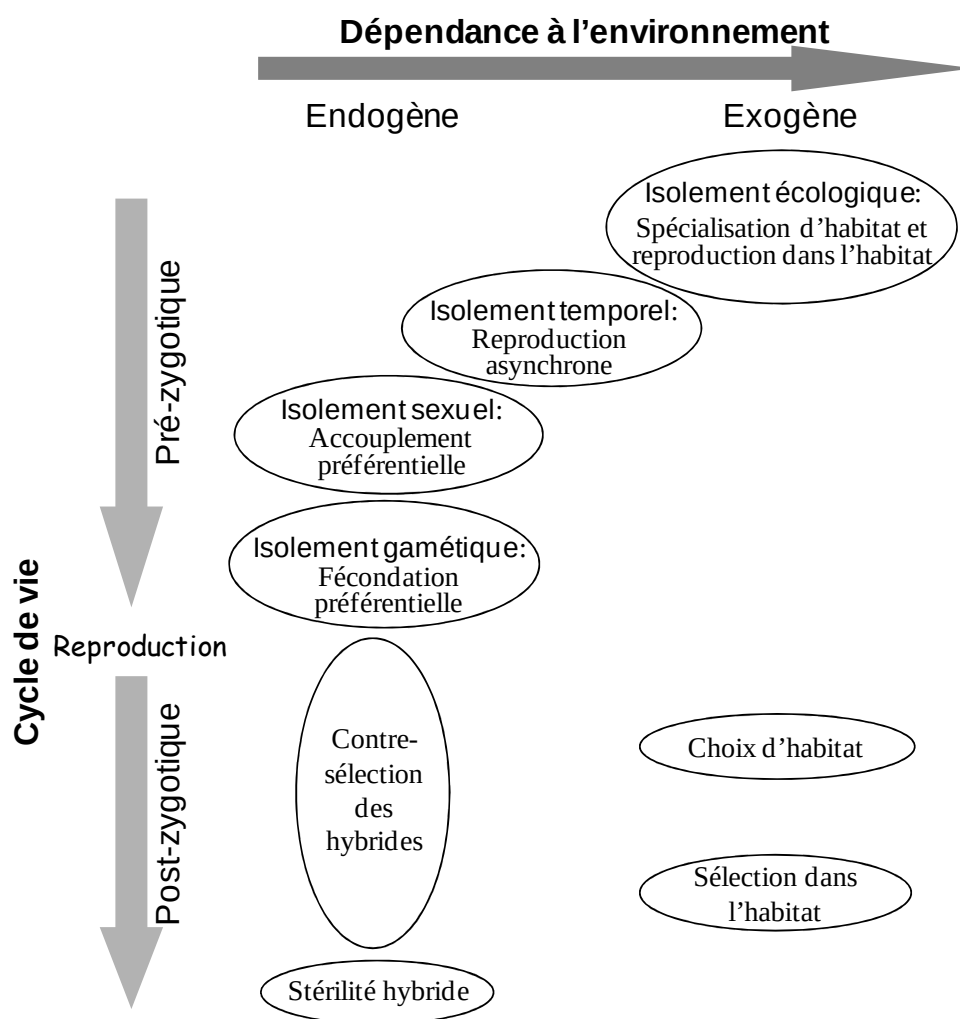


Figure 12 : Classification des mécanismes d'isolement reproductif en fonction de leur dépendance à l'environnement et du stade du cycle de vie au moment duquel ils interviennent

La figure 12 présente un exemple de classification des mécanismes d'isolement reproductif. Pendant ma thèse, je me suis attaqué à l'étude de la plupart de ces mécanismes en profitant d'une avancée technologique, les marqueurs génétiques basés sur l'amplification de l'ADN par PCR. L'avantage était double. Tout d'abord, grâce à ces nouveaux marqueurs, je disposais de locus dont

la neutralité ne faisait que peu de doute, à la grande différence des allozymes. Dans les zones d'hybridation, l'idéal est d'avoir accès aux gènes d'isolement, les seuls formant des clines de fréquence allélique stables dans le temps. Alternativement, l'utilisation de marqueurs neutres est possible si les déséquilibres de liaison l'autorisent et si l'introgression est convenablement prise en compte. Le pire est l'incertitude, des locus qui pourraient bien être intrinsèquement neutres mais pour lesquels le doute existe d'une possible implication directe dans l'isolement reproductif. C'était le cas des allozymes. Le deuxième avantage était d'avoir enfin accès, grâce à la PCR, aux génotypes des stades précoces du cycle de vie, les larves et les jeunes recrues (A15, A28).

Pour étudier les gènes d'isolement avec des marqueurs neutres dans les populations naturelles, je me suis placé dans des populations avec de forts déséquilibres de liaison, dans lesquels les gènes d'isolement sont polymorphes. Comme je l'ai expliqué plus haut, ces populations se sont révélées être rares (chapitre II.2), mais nous les avons heureusement découvertes dans la Baie de Quiberon en Bretagne sud. Cette approche est très puissante pour mettre à jour des mécanismes impliquant des différences phénotypiques entre les deux espèces et nous avons découvert :

- (i) un **asynchronisme partiel de ponte** (A15),
- (ii) un **choix d'habitat** au moment de la fixation des larves (A15, A19),
- (iii) une **sélection directionnelle** en faveur des génotypes *M. galloprovincialis* (A19).

Cependant, cette approche est très problématique lorsqu'on s'intéresse au phénotype des hybrides. Tout d'abord parce que les hybrides qui nous intéressent sont probablement très rares dans la nature. Sinon la barrière génétique au flux génique neutre ne serait pas efficace (Barton, Hewitt, 1985; Barton, Bengtsson, 1986) et la zone d'hybridation n'aurait pas été découverte avec des marqueurs neutres. Ensuite, parce que la caractérisation des génotypes hybrides avec quelques marqueurs est très peu précise. S'il est assez efficace de discerner un pool d'individus plutôt *edulis* et un pool d'individus plutôt *galloprovincialis* et de comparer leur phénotype, il est peu efficace de discerner un troisième pool d'individus classé comme hybrides car il contiendra beaucoup de génotypes introgressés (c'est-à-dire surtout hybrides aux marqueurs neutres mais principalement homospécifiques pour les gènes d'isolement). Enfin, certaines études ont analysées des populations à l'équilibre de Hardy-Weinberg et de liaison qui n'étaient probablement même pas polymorphes pour les gènes d'isolement (Gardner, 1994; Hilbish *et al.*, 2003; Coghlan, Gosling, 2007). Toutes ces raisons font que la dépression d'hybridation a toujours été suspectée être inexistante chez les *Mytilus*.

Afin d'obtenir les génotypes hybrides désirés (F1, F2, rétrocroisements) dans des proportions importantes et contrôlées, le seul moyen est de réaliser des croisements en laboratoire. Cette approche a permis de dévoiler deux autres mécanismes d'isolement :

- (iv) des **fécondations préférentielles** entre conspécifiques (A18),
- (v) une **dépression d'hybridation**, notamment pendant la phase larvaire (A10).

L'étude des croisements F2 réalisés pendant le stage Erasmus de Marta Szulkin, nous a permis de progresser dans la compréhension des bases génétiques de la dépression d'hybridation. Si les F1 sont hétérozygotes EG sur tout le génome, les F2 peuvent être EE, EG et GG à chaque locus, toutes les combinaisons étant produite en proportion mendéliennes. La figure 13 présente les paysages de fitness bi-locus (Gavrilets, 2004), c'est-à-dire les survies (ici standardisées par celle du meilleur génotype) des différents génotypes bi-locus dans la descendance F2. Les survies sont obtenues en comparant les fréquences génotypiques 36h après la fécondation aux fréquences génotypiques observées à 200 jours. La première tendance est un avantage pour les génotypes les plus *M. galloprovincialis*. Cependant, on remarque que le génotype le moins viable est toujours un génotype homozygote EE à un locus et homozygote GG à l'autre locus. Les deux génotypes EEGG et GGEE sont toujours moins viables que le double hétérozygote (EGEG) qui présente pourtant la même dose d'allèles G.

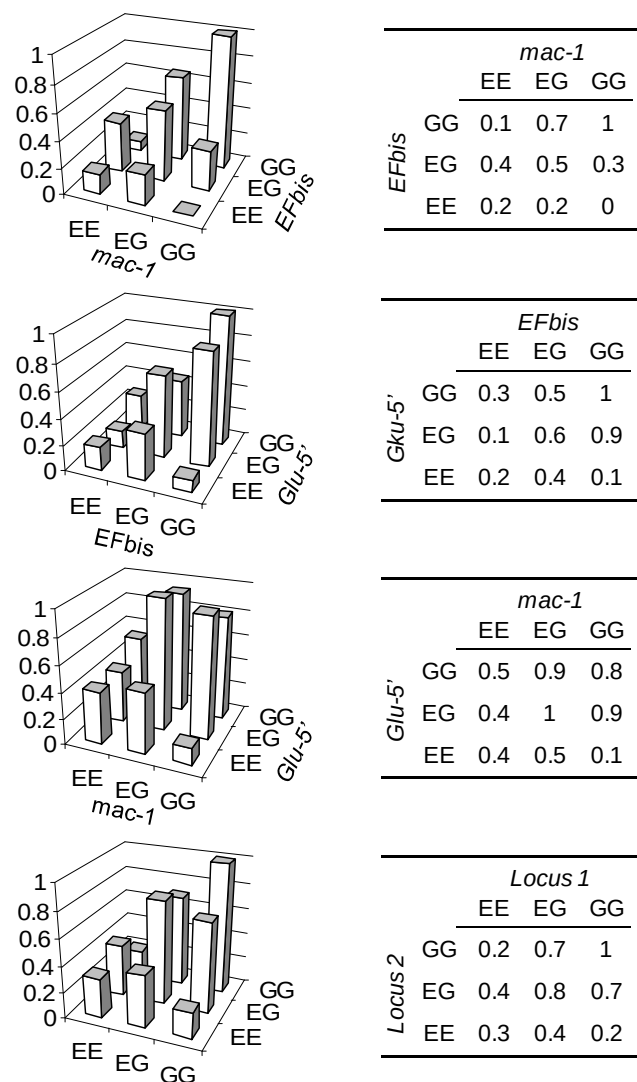


Figure 13 : Paysage de fitness bi-locus : survie normalisées des différents génotypes F2 de *Mytilus* sur trois paires de locus (d'après A10). Le paysage du bas est la moyenne symétrique des trois autres.

En outre, au moins l'un de ces deux génotypes, parfois les deux, est (sont) moins viable(s) que les génotypes EEEE. Comme ici chaque allèle marque l'origine du segment chromosomique qui l'entoure, ces résultats montrent que chaque segment chromosomique pris au hasard comportent au moins une incompatibilité dites « Dobzhansky-Muller » (Orr, Presgraves, 2000) avec un autre segment chromosomique pris au hasard, et que ces incompatibilités sont la conséquence de substitutions récessives. Ces résultats sont en accord avec la théorie de la dominance de l'isolement post-zygotique de Turelli et Orr (2000). De plus, ils suggèrent que le nombre de substitutions Dobzhansky-Muller est très grand dans les génomes de *Mytilus*.

L'expérience a fini par conforter la théorie car un grand nombre d'incompatibilités endogènes distribuées dans le génome est la meilleure façon de créer une barrière génétique efficace au flux génique neutre (Barton, Bengtsson, 1986).

L'absence de dépression d'hybridation entre espèces de *Mytilus* est longtemps restée un fait établi. Comment expliquer alors que nous découvriions finalement une multitude de locus de dépression d'hybridation ? La raison que j'invoque est que précédemment les individus utilisés pour estimer la valeur sélective des hybrides étaient des individus introgressés, c'est-à-dire des hybrides qui ne possèdent les allèles que d'un seul taxon aux gènes d'isolement mais possèdent les allèles des deux taxons pour certains marqueurs. Le troisième jalon de cette HDR est : pour étudier les locus sélectionnés avec les voisins neutres mieux vaut s'assurer que le polymorphisme sélectionné est bien présent dans les populations étudiées ; le polymorphisme du marqueur neutre ne nous l'assure pas.

III.2- Relations directes

III.2.1- Une proportion importante des substitutions en acide aminé sont fixées par la sélection darwinienne positive

Quelle est la proportion des substitutions fixées par la dérive génétique et quelle proportion par la sélection positive ? Pour répondre à cette question très ancienne en évolution moléculaire (Kimura 1983), il est utile de comparer des données de polymorphisme entre génomes d'une même espèce/population et de divergence entre espèces/populations. Cependant, pour pouvoir correctement inférer les forces évolutives qui agissent sur les mutations, il faut pouvoir calibrer les paramètres communs aux deux types de données. Pour cela il faut une référence neutre. McDonald et Kreitman (1991) ont proposé d'utiliser les mutations synonymes, qui ne provoquent pas de changement d'acide aminé dans la protéine, comme référence neutre pour inférer la direction de la sélection agissant sur les mutations non-synonymes (qui, elles, provoquent un remplacement d'acide aminé). Il a été prédit que cette approche, appliquée à un grand nombre de gènes, allait permettre d'évaluer l'importance de la sélection dans l'évolution du génome d'une espèce

(Brookfield & Sharp 1994; Kreitman & Akashi 1995), mais ce n'est que très récemment que des données de polymorphisme sont devenues disponibles sur un nombre suffisant de gènes chez quelques espèces comme l'homme ou la drosophile. Les premières estimations génomiques du taux de substitutions adaptatives sont surprenantes : ~35% entre le chimpanzé et l'homme (Fay et al. 2001) et ~45% entre *Drosophila simulans* et *D. yakuba* (Smith & Eyre-Walker 2002) seraient adaptatives. Cependant ces estimations étaient basées sur un calcul simple où les variations des paramètres évolutifs (taux de mutations ou taille efficace) entre les gènes n'étaient pas prises en compte. J'ai développé une méthode permettant d'analyser par maximum de vraisemblance un grand nombre de gènes sans que le polymorphisme de chacun d'eux ne soit nécessairement fort, tout en prenant en compte les variations génomiques des paramètres évolutifs (A13). Une telle approche globale est indispensable car la variance des estimations gène par gène est très forte. Il a ainsi été montré qu'un modèle à taux variable entre gènes n'est pas significativement meilleur qu'un modèle à taux unique et donc que le taux de substitutions adaptatives est remarquablement constant entre les gènes. La nouvelle méthode a été appliquée sur plusieurs jeux de données utilisant le polymorphisme disponible soit chez *D. melanogaster* soit chez *D. simulans* et diverses données de divergence entre *D. melanogaster*, *D. simulans* et *D. yakuba*. Ceci a permis d'obtenir une estimation fiable et conservatrice du taux de substitutions adaptatives de ~25% dans la lignée *simulans*, soit une substitution adaptative toutes les 600 générations. Ce taux est plus élevé que ne l'avait prédit Kimura dans sa théorie neutraliste de l'évolution moléculaire (Kimura 1983). Cependant, **il reste vrai que la sélection positive n'agit que sur un tout petit nombre de site dans un génome à un moment donné**, l'aiguille de l'adaptation en action est noyée dans la botte de foin du génome.

Au cours de mon travail post-doctoral, j'ai été amené à m'intéresser aux taux de substitutions synonymes. Je me suis alors rendu compte que différentes méthodes d'estimation donnaient des résultats très différents et que la méthode qui donnait des résultats erronés était en fait la méthode reconnue auparavant comme étant la plus fiable, celle de Goldman et Yang (1994) implémentée dans le logiciel PAML (Yang 1999). La méthode de Goldman et Yang surestime la proportion de sites synonymes et donne des résultats incorrects pour les gènes fortement biaisés pour l'usage du code (A14). Cette découverte a été l'occasion de faire le point sur la notion de taux de substitution par site synonyme ou par site non-synonyme (voir encadré 1), une notion dont le côté abstrait est souvent ignoré des utilisateurs de dn , le taux de substitution non-synonyme par site non-synonyme, ds , le taux de substitution synonyme par site synonyme et du ratio dn/ds .

Encadré 1 : Le concept de proportion de synonymie d'une séquence codante.

Ce ne sont pas les sites d'une séquence nucléotidique codant une protéine qui sont synonymes (ne modifiant pas la séquence en acides aminés) ou non-synonymes (modifiant la séquence en acides aminés) mais les changements entre séquences. Supposons que nous observions 10 changements synonymes et 10 changements non-synonymes entre deux séquences. Cela signifie-t-il que les taux de substitutions non-synonymes et synonymes sont identiques ? La réponse est non car la structure du code génétique fait que les possibilités de mutations non-synonymes sont environ quatre fois plus nombreuses que les synonymes. Dans la pratique, pour que les taux de substitutions synonymes et non-synonymes soient directement comparables nous cherchons à les ramener à une valeur par site. Ces taux par site sont virtuels puisque les sites n'ont pas de réalité physique. Appelés ds (parfois Ks) pour les taux synonymes et dn (parfois Ka) pour les taux non-synonymes, ils sont estimés en fonction du choix d'un modèle d'évolution des séquences. Il existe une douzaine de méthodes pour estimer dn et ds (A14). Le modèle d'évolution intervient tout d'abord au niveau de l'estimation du nombre de changements. Il sert alors à corriger pour les substitutions multiples (*i.e.* l'homoplasie). Le modèle intervient aussi pour compter les nombres virtuels de sites synonymes et non-synonymes dans la séquence étudiée. Par exemple, sur les trois mutations possibles en troisième position d'un codon dégénéré deux fois, deux mutations sont des transversions non-synonymes et une mutation est une transition synonyme. Si le ratio transition/transversion, ts/tv , est considéré être égale à 1 comme dans le modèle de substitution le plus simple (Jukes, Cantor, 1969), la position est comptée comme 1/3 synonyme. Si ts/tv est considéré être différent de 1 comme dans le modèle à deux paramètres (Kimura, 1980), la position sera comptée comme $ts/(2tv+ts)$ synonyme. Cette valeur est par exemple de 1/2 pour un ts/tv de 2 et elle tend vers 1 lorsque ts/tv augmente. Ce concept un peu abstrait d'un site a été appelé le concept d'« opportunité mutationnelle » (A14). Il est poussé dans sa sophistication la plus extrême dans le modèle de Goldman et Yang (1994) qui incorpore le biais d'usage des codons en plus du ratio ts/tv . Il est déconseillé d'utiliser ce concept lorsqu'on s'intéresse aux taux d'évolution absolus, par exemple lorsqu'on souhaite comparer les taux d'évolution entre gènes qui ont des compositions en bases différentes. Il est alors préférable d'utiliser un concept « physique » du site en étudiant par exemple le taux de substitution de la troisième position des codons dégénérés quatre fois (A14). Lorsqu'on s'intéresse au ratio $\omega=dn/ds$, c'est un peu différent. Ce qui nous intéresse c'est d'estimer à partir de l'observation des mutations synonymes, la proportion des mutations pouvant être non-synonymes dans le gène étudié avant l'action de la sélection. Dans cette situation, l'utilisation du concept d'« opportunité mutationnelle » dans une version élaborée peut-être préférable. Le modèle de Goldman et Yang (1994) a prouvé sa primauté, d'autant plus que les paramètres sont ajustés de façon élégante par maximum de vraisemblance, permettant de comparer divers modèles emboîtés dans lesquels le paramètre ω peut varier dans les branches d'une phylogénie (Yang, 1998) ou entre positions d'une protéine (Yang *et al.*, 2000).

III.2.2- Effet délétère des mutations introniques polymorphes du gène *EF1α* chez les modioles hydrothermales du genre *Bathymodiolus*

J'expose ici brièvement un résultat obtenu dans le cadre des travaux de thèse de Baptiste Faure. Ce petit travail me permet d'illustrer que le possible effet d'une sélection directe sur le polymorphisme étudié n'est pas une hypothèse écartée d'emblé. D'autant plus que l'effet direct d'une sélection négative très faible sur les mutations polymorphes ressemble beaucoup à l'effet indirect de la sélection positive (balayage sélectif) : on observe un excès de polymorphismes rares par rapport à l'attendu de l'équilibre mutation/dérive (Tajima, 1989). Il m'intéresse donc de trouver comment discriminer les deux. En combinant des données de polymorphisme d'espèces proches et des données de divergence, il nous est apparu possible de distinguer les deux effets car il est très peu probable d'observer un balayage sélectif de même intensité, de même âge et au même endroit du génome chez deux espèces. En étudiant les *Bathymodiolus*, nous avons observé un fort excès de polymorphismes introniques rares au gène *EF1α* dans toutes les espèces étudiées alors que les autres locus ne déviaient pas de l'attendu neutre à l'équilibre mutation-dérive. De plus, la divergence intronique observée à ce gène s'est révélée beaucoup plus faible que pour les autres gènes, en accord avec une action de la sélection purifiante. Il a donc été conclu que les mutations introniques polymorphes du gène *EF1α* étaient sous contraintes sélectives faibles chez les *Bathymodiolus* (A8).

III.3- Relations incertaines

Pendant le stage de master d'Eva Boon et la thèse de Matthieu Faure, nous nous sommes intéressés à deux gènes choisis en raison de leurs rôles dans l'immunité. Trop d'espoir est souvent mis dans cette approche « gènes candidats » et il faut rester très prudent sur l'interprétation des résultats issus de cette approche. Le moindre indice de sélection sur un gène à la fonction connue, et prétendue être très importante dans l'adaptation des espèces, peut facilement glisser vers une interprétation hâtive de la sélection directe malgré l'ubiquité de la sélection indirecte dans les génomes (Li, Stephan, 2006; Charlesworth, 2009).

Les peptides antimicrobiens (PAMs) sont produits par les animaux, les plantes, les champignons et les bactéries pour se défendre contre les attaques microbiennes. Ils agissent sur la structure de la paroi bactérienne et ont généralement un spectre d'action assez large. De plus certains auteurs ont suggéré que leur mode d'action ne permettrait pas facilement aux bactéries d'acquérir des résistances efficaces (Zasloff, 2002). On n'attend donc pas, a priori, que les PAMs participent à la course aux armements entre un hôte et ses microbes pathogènes. En accord avec cette idée, il a été montré que la structure moléculaire des PAMs, stabilisée par des ponts entre cystéines, était restée remarquablement stable dans l'ensemble du vivant (the γ -core signature, Yount, Yeaman, 2004). Cependant, des observations récentes suggèrent que les PAMs pourraient tout de même être des acteurs importants de la course aux armements, notamment chez les

invertébrés qui n'ont comme arme que leur immunité innée. Tout d'abord, il a été montré que les bactéries pouvaient acquérir des résistances aux PAMs (Yeaman, Yount, 2003), parfois rapidement (Perron *et al.*, 2005). La diversité moléculaire des PAMs est très importante avec de nombreuses familles multigéniques et une forte diversité allélique (Tennessen, 2005). La divergence entre espèces suggère une évolution rapide des PAMs (Schlenke, Begun, 2003) avec des indices d'évolution adaptative (Tennessen, 2005). Enfin, la variation génétique observée à certains locus de PAMs a été associée à la variation de la prévalence bactérienne chez la drosophile (Lazzaro *et al.*, 2004). Ainsi, il semble que la sélection soit une force importante de l'évolution de ces molécules. Cependant, la forme que prend cette sélection –positive, disruptive, stabilisatrice, épistatique– n'est pas claire et ne peut être étudiée qu'à petite échelle évolutive, dans le polymorphisme intraspécifique. Pourtant, le polymorphisme des PAMs reste encore peu exploré, probablement du fait de la difficulté engendrée par l'existence de nombreux paralogues proches.

La Mytiline B et la MGD2 (*Mytilus Galloprovincialis* Defensin 2) font partie des PAMs isolés dans les hémocytes des moules du genre *Mytilus* (Mitta *et al.*, 2000). La structure génomique de ces gènes a été caractérisée par Mitta *et al.* (2000). A partir des séquences des introns, des amorces spécifiques des locus de la Mytiline B et de la MGD2 ont pu être définies. Une analyse de la variation nucléotidique et du polymorphisme de longueur de ces locus a été conduite sur des échantillons des populations de la zone d'hybridation mosaïque des moules *Mytilus edulis* et *M. galloprovincialis* (A2, Annexe 2). Ces deux locus ont montré les différenciations génétiques entre les deux espèces les plus faibles de tous les locus analysés jusqu'à maintenant (A2). L'analyse du polymorphisme de séquence a montré les résultats suivant :

- Deux clades d'allèles très divergents ont été observés à la Mytiline B (Figure 9) en accord avec l'histoire de divergence en allopatrie et d'introgression secondaire. Deux substitutions d'acides aminés et une insertion de deux acides aminés ont été observées dans les séquences du peptide mature entre ces deux clades, donnant des sites candidats à l'action directe de la sélection. Les tests de neutralité ont démontré des traces de l'action de la sélection sur ce locus, notamment lorsqu'on regarde les deux clades séparément. Pourtant, en y regardant de plus près les généalogies ne sont pas en accord avec l'hypothèse de la sélection directe car des lignées semblent avoir survécues au balayage sélectif. Alternativement, la sélection est directe mais si faible que la fixation n'a pas encore eu lieu. Quoiqu'il en soit, les résultats ne permettent pas de conclure définitivement.
- Le locus MGD2 est incroyablement homogène entre les deux espèces puisqu'aucune différenciation significative n'est observée. Ce locus présente un très fort taux de recombinaison intra-génique qui a pour conséquence d'empêcher la construction d'une généalogie simple. Une analyse en fenêtre coulissante a permis de mettre en évidence que la recombinaison est beaucoup plus forte dans la partie 3' du locus, dans l'exon codant pour le peptide mature. De plus, la diversité nucléotidique ainsi que le D de Tajima présente une variation forte le long de la séquence (Figure 14). Globalement négatif, le D de Tajima devient positif sur une zone d'environ 400pb située au début de l'intron 3. Ce type d'observation est généralement accepté comme une preuve

sérieuse de la sélection balancée (Hudson *et al.*, 1987; Wheat *et al.*, 2006). Pourtant, le pic de diversité n'est pas localisé sur l'exon codant pour le peptide mature. Plusieurs polymorphismes candidats ont tout de même été identifiés aux alentours de ce pic de diversité, comme des remplacements ou une délétion d'acide aminé dans le peptide signal ou un élément transposable dans l'intron. Le problème est que l'introggression secondaire peut également générer de la diversité et un D de Tajima positif. Il faut alors expliquer par la sélection la faible diversité et le D de Tajima négatif de la portion 3' du locus. Dans le peptide mature a été observée une mutation en fréquence intermédiaire (Leu31Arg) dans la boucle suspectée active du peptide (Romestand *et al.*, 2003). Ce site polymorphe est de plus un des rares sites à montrer une structuration importante entre les populations (Figure 14), bien que la significativité de cette structure ne résiste pas à une correction de Bonferroni.

Aucune preuve irréfutable de l'impact d'une sélection directe sur ces deux PAMs n'a donc été identifiée. Certaines observations laissent cependant la voie ouverte à de futures analyses.

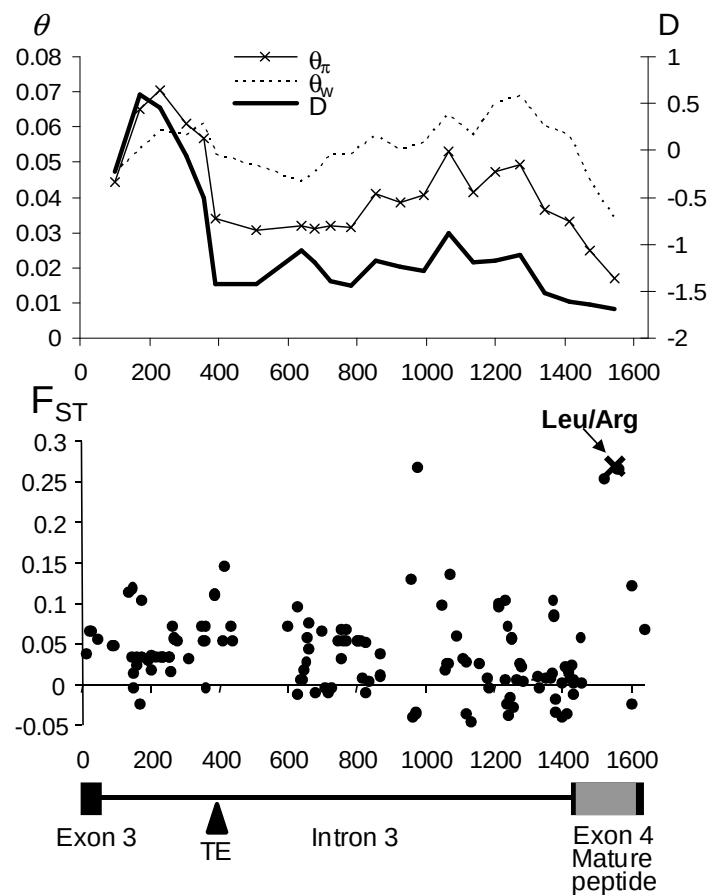


Figure 14 : Analyse en fenêtre coulissante de la diversité nucléotidique, du D de Tajima et du Fst, réalisées sur des données de polymorphisme de séquence de la défensine MGD2 échantillonnées dans les populations *M. edulis* et *M. galloprovincialis*.

IV- Marqueurs moléculaires, valeur sélective et environnement

Dans cette troisième partie, j'aborde l'étude de la sélection en y incorporant le troisième élément nécessaire à l'étude de la sélection en action dans les populations naturelles, l'environnement. Que les écologistes ne se réjouissent pas trop vite : d'une part l'environnement en génétique moléculaire des populations se résume souvent à la géographie (ex : le nord et le sud qui comme tout le monde le sait diffèrent en température⁹), d'autre part, mon message va plutôt dans le sens d'une minimisation de l'importance de l'environnement dans l'explication des patrons de structure génétique analysés avec des marqueurs moléculaires. Vous croyez peut-être qu'une structure plus forte sur un locus que sur les autres implique nécessairement une sélection disruptive ; si de plus, la variation spatiale est clinale et/ou la structure génétique est corrélée à l'environnement, vous pensez l'adaptation locale démontrée. Pourtant si vous analysez des marqueurs neutres vous pouvez obtenir ces résultats avec d'autres sortes de sélection et vous tromper dans l'interprétation car l'effet indirect de la sélection n'est pas la sélection.

IV.1- Le leurre du Fst trop fort : autostop local en environnement variable *versus* autostop global en population subdivisée.

Dès les premières études de génétique moléculaire des populations les généticiens ont recherché l'effet de la sélection sur la structuration de la diversité génétique (Lewontin, Krakauer, 1973; Hedrick *et al.*, 1976). Les premiers marqueurs génétiques étant des enzymes, il semblait concevable que certains allèles soient favorisés sous certaines conditions environnementales et d'autres allèles dans d'autres conditions (Mitton, 1997). De beaux exemples de polymorphismes maintenus par la sélection ont émergé de plusieurs décades de recherches : le polymorphisme fast/slow de la déshydrogénase alcoolique (ADH) de *Drosophila melanogaster* qui forme un cline latitudinal de fréquence allélique (Berry, Kreitman, 1993), la Leucine aminopeptidase (LAP) de *Mytilus edulis* qui forme un cline corrélé à la salinité dans le détroit de Long Island à New York (Koehn *et al.*, 1980), les allèles de résistance aux insecticides chez le moustique *Culex pipiens* dans la région de Montpellier (Lenormand *et al.*, 1999), le gène *Mc1r* responsable du polymorphisme de couleur du pelage chez les souris *Chaetodipus intermedius* (Hoekstra *et al.*, 2004). Cependant, le peu d'exemples émanant de ces longues années de recherches tout comme des arguments théoriques sur le maintien du polymorphisme en environnement variable suggèrent que peu de polymorphismes doivent être maintenus par la sélection locale (Hedrick, 1986; Hedrick, 2006).

Parallèlement, des travaux théoriques ont montré qu'un polymorphisme neutre pouvait être influencé par la sélection agissant sur un gène sélectionné lié sur le chromosome. 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. La conséquence est

⁹ Dans la littérature marine la température se dispute avec la salinité le leadership du paramètre écologique le plus important. "Have the authors not noticed that the North Sea is colder than the Bay of Biscay?" nous rappelait un referee bienveillant pour nous aider à expliquer la structure génétique forte observée à un locus entre ces deux régions. Sommes-nous naïfs quand même, nous ne l'avions même pas remarqué, merci...

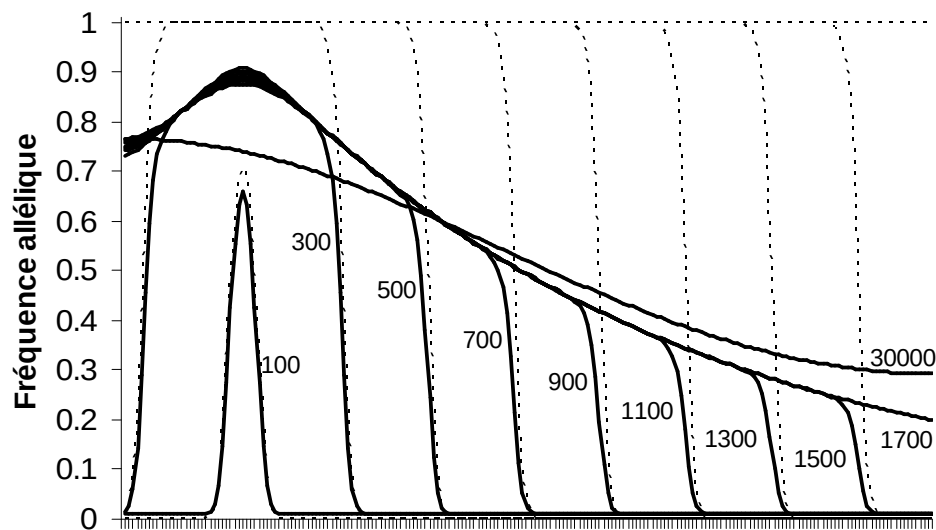
de diminuer la diversité des locus neutres, et ceci d'autant plus que les locus sont proches du gène sous sélection. Ainsi la fixation de la mutation favorable crée une vallée de diversité génétique. De nombreux travaux théoriques ont amélioré les prédictions du modèle initial (Barton, 2000b), des tests ont été construits pour détecter l'autostop, et de nombreux exemples ont été publiés (Nielsen, 2005). Récemment, les techniques de biologie moléculaire à haut débit ont modifié l'échelle de perception de l'évolution des génomes en permettant d'étudier un très grand nombre de locus. 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 *et al.*, 2005; Li, Stephan, 2006), c'est-à-dire qu'ils conservent la trace d'une sélection passée.

Une catégorie d'étude s'est intéressée à la distribution de la différenciation génétique entre populations dans le but d'identifier des locus avec une structure anormalement forte ou anormalement faible (Luikart *et al.*, 2003). Les locus montrant une structure forte seraient sous l'action d'une sélection disruptive (adaptation locale) et les locus montrant une structure faible seraient sous l'action d'une sélection stabilisatrice (Beaumont, 2005). Des méthodes statistiques ont été développées pour identifier ces locus suspectés être sous sélection (Lewontin, Krakauer, 1973; Beaumont, Nichols, 1996; Vitalis *et al.*, 2001; Foll, Gaggiotti, 2008) et elles sont appliquées dans un nombre grandissant d'études (Storz, 2005; Nosil *et al.*, 2009). L'objectif sous jacent est quelque peu différent des études qui recherchent les traces d'une sélection passée puisqu'ici ce sont des polymorphismes maintenus par la sélection qui sont recherchés, c'est-à-dire une sélection en action, cette même sélection dont de nombreux arguments avaient abouti à la conclusion qu'elle devait être rare. Or la proportion de locus identifiés être sous sélection est souvent appréciable (autour de 5-10%). Est-il possible qu'une portion des ces locus soient en fait la conséquence d'un scénario sélectif différent ?

En travaillant avec mon doctorant Matthieu Faure sur un locus présentant une structure anormalement forte entre deux populations de *Mytilus edulis*, nous avons proposé comme explication possible un scénario évolutif alternatif négligé jusqu'alors, il s'agirait en fait d'une signature d'une sélection passée (A3, Annexe 3). Les modèles d'autostop génétique sont principalement restés restreints à la situation simple d'une population panmictique isolée alors que les populations naturelles sont le plus souvent structurées géographiquement ou génétiquement (Charlesworth *et al.*, 2003). Les rares modèles d'autostop en population structurée prédisent que l'effet d'autostop doit diminuer au fur et à mesure que la mutation favorable se propage depuis le deme d'origine, dans lequel la mutation favorable est apparu, vers les autres deme (Slatkin, Wiehe, 1998; Barton, 2000b; Santiago, Caballero, 2005). Sous certaines conditions, un autostop génétique en population structurée peut laisser comme trace sur le polymorphisme neutre une différenciation génétique forte entre populations et la différenciation la plus importante apparaît pour un taux intermédiaire de recombinaison (Slatkin, Wiehe, 1998).

Un modèle d'autostop global en population structurée a d'abord été décrit par Slatkin et Wiehe (1998) et repris sous une forme modifiée par Faure *et al.* (A3, annexe 3). Un balayage sélectif à lieu dans deux sous-populations mais l'effet est plus fort dans la sous-population où apparaît la mutation favorable et moins fort dans la sous-population envahie secondairement après migration. Pour que l'effet soit fort et visible le taux de migration entre les deux sous-populations doit être faible. Ici, je présente un modèle plus général d'un autostop en population spatialement subdivisée dont les résultats ont été brièvement décrits par Barton :

La figure 15 présente la propagation d’une vague d’avance d’une mutation favorable (traits pointillés) ainsi que d’un allèle neutre initialement rare (traits pleins) qui fait de l’autostop avec cette mutation favorable. La mutation apparait dans le trentième dème d’un pas-japonais linéaire de 200 dèmes.



43

Quand l'autostop est terminé, l'allèle favorable a fixé dans tous les dèmes de la métapopulation et un gradient de fréquences alléliques est observé au locus neutre. Au fur et à mesure que la vague d'avance de l'allèle sélectionné s'est propagée, les déséquilibres de liaison entre les deux locus se sont atténués suite à la recombinaison et l'effet d'autostop s'est amoindri. En modélisant une métapopulation plus grande, une sélection plus faible, ou une recombinaison plus forte, il est possible d'atteindre un stade où la vague d'avance se propage sans n'avoir plus aucun effet sur le locus neutre, car l'équilibre de liaison est atteint. C'est ce qui a amené Barton (2000b) à conclure que "this increase will be restricted to the immediate neighbourhood of the birthplace of the new allele". Une fois l'autostop fini, les fréquences alléliques du locus neutre s'homogénéisent lentement sous l'effet de la migration (Figure 15). Cependant, le temps nécessaire pour aboutir à l'homogénéisation par la simple force de la migration peut être très long (Barton *et al.*, 2007).

Le résultat qui mérite de poser un quatrième jalon ici est qu'un cline de fréquence allélique peut-être une trace d'une sélection passée sur un locus neutre. Or, l'observation d'un cline de fréquence allélique est toujours utilisé comme un argument en faveur d'une sélection environnementale en action (et souvent directe). La sélection indirecte n'est pas la sélection, les perturbations engendrées par un balayage sélectif sur la diversité neutre et sa distribution spatiale peuvent être difficile à prédire.

Intéressons nous maintenant à l'empreinte laissée sur le chromosome en faisant varier la recombinaison. La figure 16 présente d'une part la fréquence de l'allèle autostoppé dans deux populations éloignées (dème 30 et dème 170), d'autre part la différence de fréquence allélique entre les deux populations. La fréquence finale, u^* , de l'allèle autostoppé dépend beaucoup de sa fréquence initiale, u_0 . Afin d'avoir une représentation graphique, nous supposons ici u_0 identique quelque soit la valeur de r/s . Nous supposons aussi u_0 petit afin que l'effet soit marqué (lorsque l'allèle neutre est déjà fréquent quand il se fait prendre en autostop, l'effet est moins fort). Concrètement dans la pratique, sur un chapelet de locus le long d'un chromosome u_0 sera différent d'un locus à l'autre, si bien qu'une représentation graphique ne donnera rien. Cependant, un ajustement pourra tout de même être réalisé. De plus, comme nous le verrons plus bas, les locus peuvent être choisis sur des données préliminaires. Sous cette condition (u_0 identique entre locus), l'empreinte chromosomique sur la fréquence de l'allèle autostoppé, u^* , est un pic centré sur le locus sélectionné. L'effet d'autostop étant moins fort dans la population éloignée de la population d'origine de la mutation favorable, le pic est beaucoup plus étroit dans cette population que dans la population d'origine de la mutation. Cependant, les variations de u^* dans les deux populations sont corrélées positivement et convergent vers $u^*=1$ en s'approchant du locus sous sélection. L'empreinte sur la différenciation entre populations n'est pas monotone mais elle est maximale pour des distances intermédiaires, elle prend la forme de deux dômes (courbure concave) de part et d'autre du locus sélectionnés. Loin du locus sous sélection l'effet d'autostop est faible partout et proche du locus sous sélection l'effet d'autostop est fort partout si bien que la structure entre

population est faible. Pour une distance intermédiaire (comme dans la figure 15) l'effet d'autostop est fort dans la population d'origine de la mutation favorable et faible dans la population éloignée, si bien que la différenciation entre population est forte.

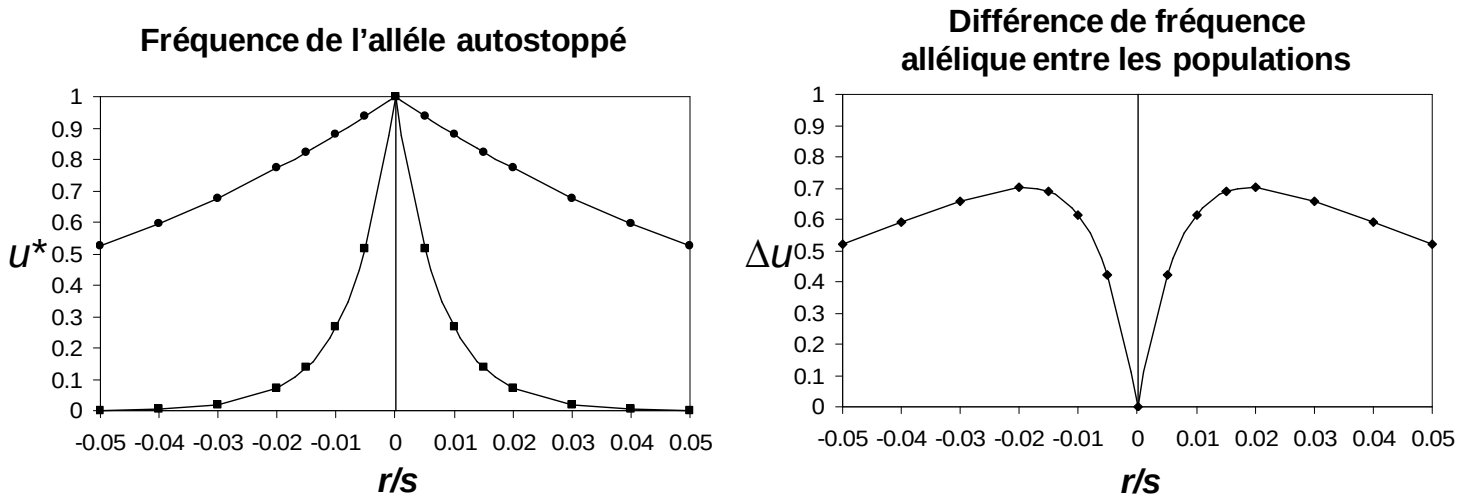


Figure 16 : Effet du ratio r/s (empreinte chromosomique) sur la fréquence de l'allèle autostoppé (à gauche) et sur la différence de fréquence allélique entre populations (à droite). Pas-japonais linéaire de 200 dèmes avec, $p_0=0.001$, $u_0=0.01$, $m = 0.05$, $s = 0.1$.

IV.1.2- Hypothèse de l'autostop local en environnement variable

Il a été démontré théoriquement que l'adaptation locale pouvait augmenter la différenciation génétique de locus neutres liés à un locus sélectionné (Charlesworth *et al.*, 1997). Ici je présente un modèle déterministe en insistant sur les deux étapes se succédant dans un processus d'autostop local.

La figure 17 présente la première phase du processus : la mise en place du cline de fréquence allélique au locus sélectionné (traits pointillés) ainsi que l'effet sur les fréquences d'un allèle neutre initialement rare (traits pleins). C'est l'autostop local *sensu stricto*. Un autostop se produit dans les dèmes où un nouvel allèle adapté localement apparaît par mutation, se propage et se fixe. Sur le locus sélectionné se forme un cline stable, à l'équilibre migration/sélection, au niveau de la frontière environnementale. Sur le locus neutre une différenciation génétique s'est formée. Cette première phase d'autostop locale est en fait la phase qui engendre la structure neutre et ce sera le moment où la structure génétique neutre sera maximale.

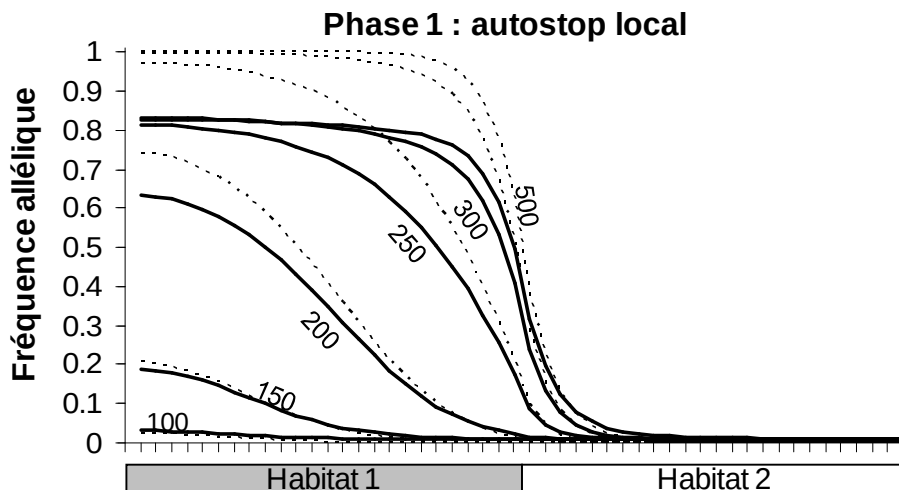


Figure 17 : Autostop local en environnement variable. Fréquence d'un allèle favorable dans l'habitat 1 apparu dans le premier dème (traits pointillés) et fréquence d'un allèle neutre initialement rare qui autostop avec l'allèle favorable (traits pleins). Les générations sont précisées à coté des courbes. Résultats d'un pas-japonais linéaire de 50 dèmes avec, $p_0=0.001$, $u_0=0.01$, $m = 0.3$, $s = 0.05$, $r = 0.001$.

La figure 18 présente la deuxième phase du processus, l'homogénéisation des fréquences alléliques au locus neutre sous l'effet de la migration et de la recombinaison. Ce processus est souvent appelé introgression et la théorie est bien développée (Barton, 1979b; Barton, Bengtsson, 1986). Ainsi les allèles autostopés dans la sous-population du balayage vont pouvoir migrer dans la sous-population non balayée.

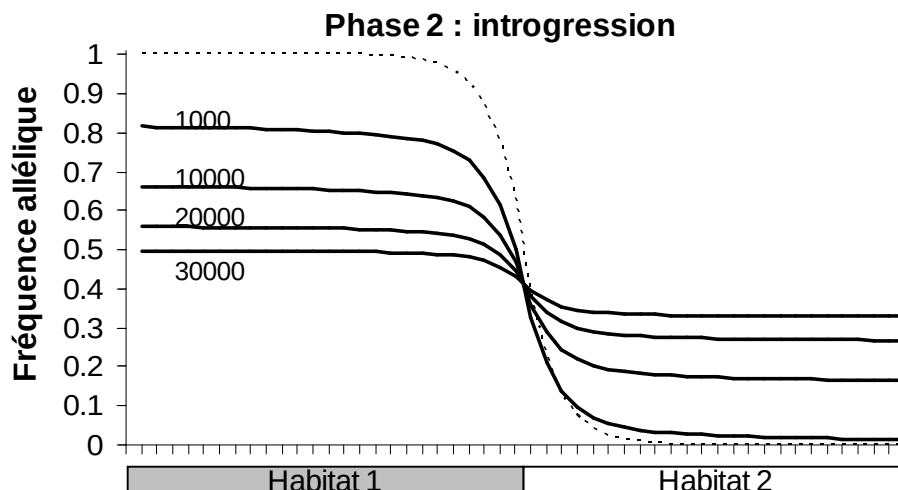


Figure 18 : Introgression d'un locus neutre au travers une barrière génétique créée par un locus d'adaptation locale. Les générations sont précisées à coté des courbes. Résultats d'un pas-japonais linéaire de 50 dèmes avec, $p_0=0.001$, $u_0=0.01$, $m = 0.3$, $s = 0.05$, $r = 0.001$.

Intéressons nous maintenant à l’empreinte laissée sur le chromosome en faisant varier la recombinaison. La figure 19 présente d’une part la fréquence de l’allèle autostoppé dans deux populations des deux habitats éloignés du cline (le locus sélectionné est fixé), d’autre part la différence de fréquence allélique entre les deux habitats, à la fin de l’autostop local *sensu stricto* (génération 500) et après une période d’introggression (génération 30000). A la fin de l’autostop local, une signature d’un balayage sélectif est observée dans la population de l’habitat 1 (où à lieu l’autostop) mais la population de l’habitat 2 n’est pas encore affecté par le processus. L’empreinte sur la différenciation entre populations prend la forme d’un pic (courbure convexe) centré sur le locus sélectionnés. Dans un second temps l’introggression va opérer et ceci d’autant plus vite que le locus neutre sera éloigné du locus sélectionné (Barton, 1979b; Bengtsson, 1985). Les locus éloignés du locus sélectionné vont atteindre très vite l’homogénéisation alors que le temps nécessaire pour atteindre l’homogénéisation va être d’autant plus grand que le locus neutre est proche du locus sélectionné. Aux alentours du locus sélectionné, les variations de u^* dans les deux habitats seront corrélées négativement et convergent vers $u^*=1$ dans l’habitat 1 et vers $u^*=u_0$ dans l’habitat 2. Le pic de différenciation génétique s’encaisse et la zone où une différenciation est observée se raccourcit.

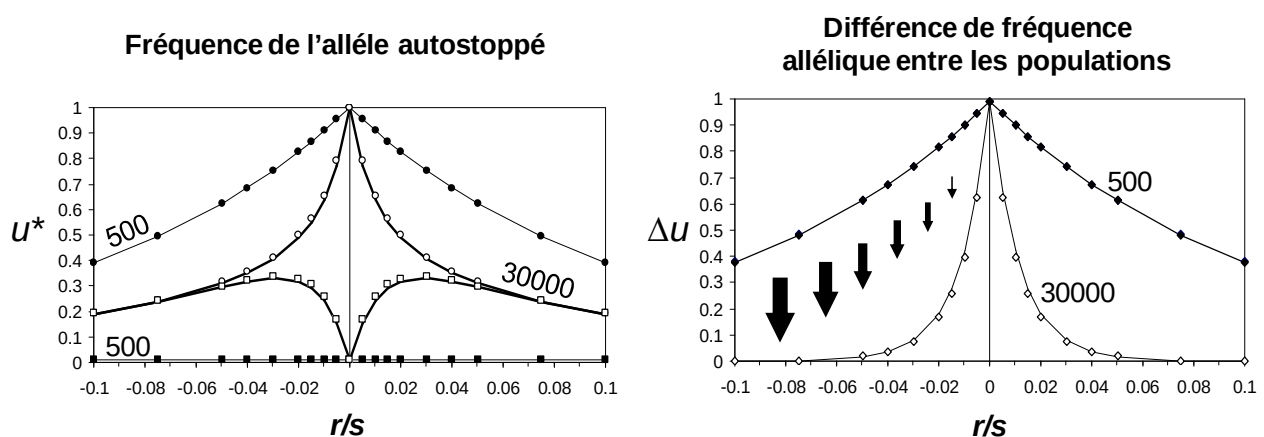


Figure 19 : Effet du ratio r/s (empreinte chromosomique) sur la fréquence de l’allèle autostoppé (à gauche) dans une population de l’habitat 1 (cercles) et une population de l’habitat 2 (carrés) et sur la différence de fréquence allélique entre les populations (à droite), à la fin du processus d’autostop local *sensu stricto* (génération 500, symboles noirs) et après une période d’introggression (génération 30000, symboles blancs). Résultats d’un pas-japonais linéaire de 50 dèmes avec, $p_0=0.001$, $u_0=0.01$, $m = 0.3$, $s = 0.05$.

Nous voyons donc que les attendus sur la signature chromosomique d’un autostop local en environnement variable sont très différents de la signature chromosomique d’un autostop global en population structurée. Cependant, les attendus sur un seul locus sont très semblables : une différenciation génétique accentuée dû à une fréquence forte d’allèle autostoppé dans une population et une fréquence faible dans l’autre. C’est ce que nous avons observé au locus *EFbis* (A3). Pour trancher entre les deux hypothèses, le seul moyen est de marcher sur le chromosome.

IV.1.3- Une marche chromosomique autour d'un locus outlier soutient l'hypothèse de l'autostop global en population structurée.

Lors du stage de M1 d'Anna Kneidinger, nous avons réussi à obtenir la séquence génomique complète du gène *EF1α* chez *Mytilus edulis*, un gène dans lequel un polymorphisme de longueur d'intron, *EFbis*, avait été identifié comme significativement trop différencié entre les populations *M. edulis* de Mer du Nord et du Golfe de Gascogne par rapport aux autres locus analysés (A3). Le gène est long d'environ 5000 pb et il est constitué de 6 exons et 5 introns. La figure 20 représente un schéma à l'échelle du gène avec sa structure en exons et introns. Aux séquences déjà obtenues précédemment (A3), nous avons obtenu un échantillon de 15 séquences dans la région 5' (exon1 à exon 2) et un échantillon de 9 séquences dans la région 3' du gène (exon 4 à exon 6). Les généalogies des échantillons des trois portions du gène sont présentées en figure 20.

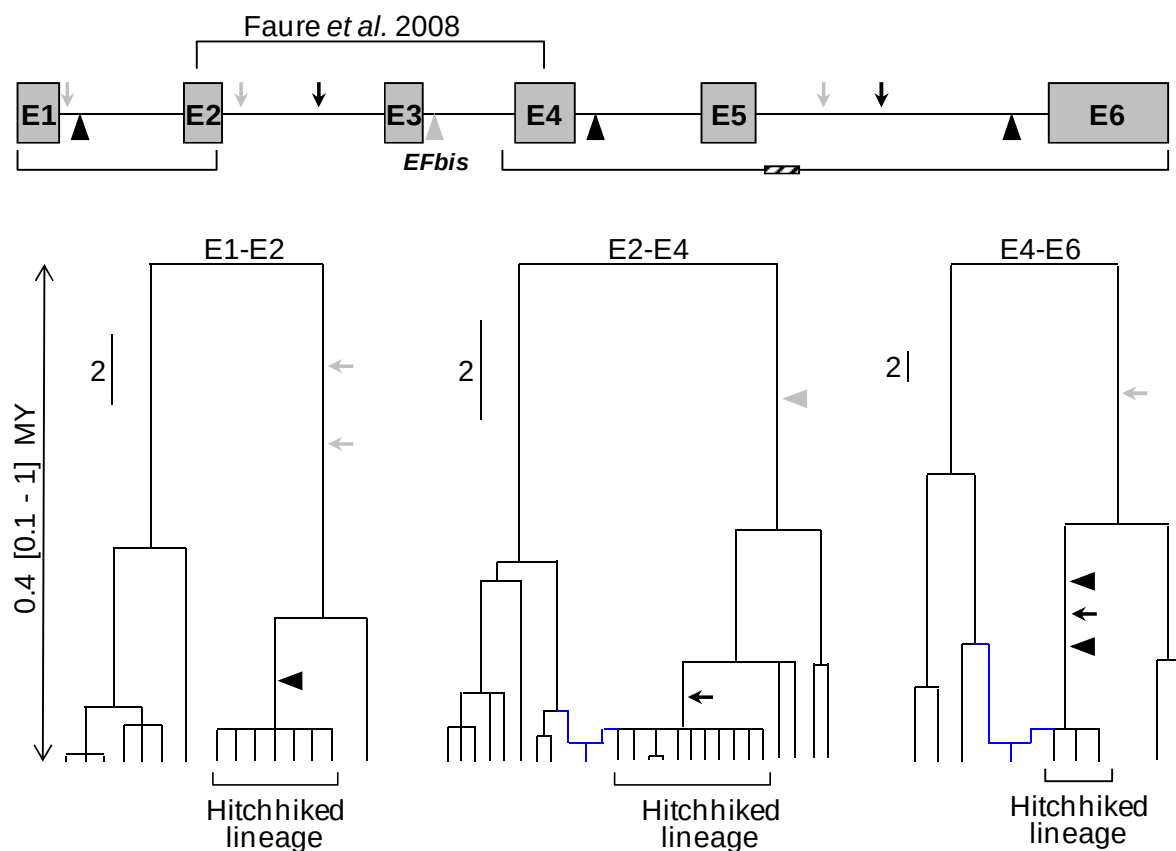


Figure 20 : Schéma à l'échelle de la structure en exons et introns du gène *EF1α* chez *Mytilus edulis* et généalogies des échantillons de trois portions du gène. Les positions dans le gène et dans les généalogies des polymorphismes choisis pour un génotypage extensif sont indiquées par des triangles (polymorphisme d'insertion/délétion) ou des flèches (polymorphisme nucléotidique, SNP). Le petit rectangle hachuré représente une petite région enlevée de l'analyse.

A partir des données de séquence, nous avons défini pour un génotypage extensif huit nouveaux locus en plus d'*EFbis* déjà analysé à grande échelle (A16). Trois correspondent à des

polymorphismes d'insertion/délétion et cinq à des polymorphismes de substitution nucléotidique. Il est important de connaître la position de ces polymorphismes dans la généalogie, cinq sont récents et situés à la base de la lignée autostoppée qui forme un clade avec une forme caractéristique d'étoile et quatre sont plus anciens et situés dans la branche interne de la généalogie. Dit autrement, la différence entre ces deux catégories de mutations peut être aussi comprise comme une différence de fréquence avant l'autostop, u_0 . Les mutations à la base de la lignée autostoppée avaient une fréquence plus faible que les mutations de la branche interne quand l'autostop s'est produit. Comme u_0 est un paramètre important du modèle, les résultats sont attendus être différents pour les deux catégories de polymorphismes.

La figure 21 présente les fréquences de l'allèle autostoppé dans les populations *M. edulis* de Mer du Nord et du Golfe de Gascogne. La fréquence de l'allèle autostoppé augmente vers le côté 5' dans les deux populations alors que la différenciation entre population diminue. Pour les deux hypothèses, que ce soit l'autostop local ou global, les résultats sont attendus être similaires pour la population de la Mer du Nord qui reçoit un balayage sélectif dans les deux scénarios. Les résultats pour cette population suggèrent qu'une fixation d'un allèle favorable à eu lieu en 5' de la région analysée. De façon remarquable, le gradient de fréquence observé dans la population du Golfe de Gascogne pointe l'exacte même zone en 5' de la région analysée pour la fixation d'un allèle favorable, en accord avec l'hypothèse de l'autostop global. Enfin, la variation de la différenciation entre population à une forme de dôme (courbe concave) et la différenciation diminue vers la région 5'.

Il n'est pas possible de comprendre les résultats sous l'hypothèse de l'autostop local en environnement variable. Alors que le gradient de fréquence de l'allèle autostoppé dans la population de Mer du Nord suggère que la mutation favorable est en 5', le gradient de différenciation suggérerait sous l'hypothèse de l'autostop local qu'elle se trouve en 3'. De plus, la variation de la différenciation le long du chromosome ne forme pas un pic mais plutôt un dôme. Les résultats sont par contre en parfait accord avec les attendus de l'hypothèse de l'autostop global en population structurée.

Un modèle d'autostop a donc été ajusté aux données par minimisation des moindres carrés en utilisant les formules suivantes :

$$u_{1,i}^* = u_{0,i} + (1 - u_{0,i})\varepsilon^{\frac{a+\rho x_i}{s}}$$

$$u_{2,i}^* = u_{0,i} + (1 - u_{0,i})\varepsilon^{\alpha \frac{a+\rho x_i}{s}}$$

où $u_{i,j}^*$ est la fréquence de l'allèle autostoppé du locus i dans la population j , $u_{0,i}$ est la fréquence initiale avant autostop du locus i , ε est la fréquence initiale de la mutation favorable (plus exactement la fréquence à laquelle la mutation favorable commence à augmenter en fréquence de façon déterministe), x_i est la position du locus i ($x=0$ est la position du codon « start » du gène *EF1 α*), a est la position du locus sélectionné, ρ est le taux de recombinaison par distance physique, s est le coefficient de sélection et α un coefficient d'atténuation de l'effet d'autostop dans la population 2.

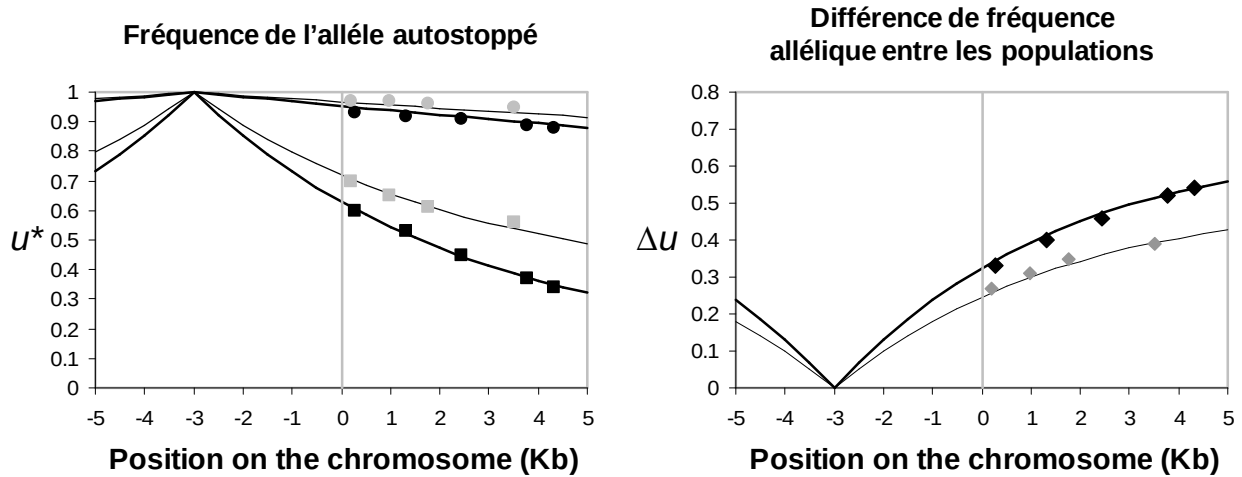


Figure 21 : Fréquence de l'allèle autostoppé (à gauche) dans la population de Mer du Nord (cercle) et la population du Golfe de Gascogne (carré) et différence de fréquence allélique entre les deux populations (à droite). Les symboles noirs correspondent aux polymorphismes situés à la base de la lignée autostoppée alors que les symboles gris correspondent aux polymorphismes situés dans la branche internes des généalogies. Les courbes correspondent à un ajustement au modèle d'autostop global avec une valeur de $u_0=0.1$ et une valeur de $u_0=0.35$.

Nous inférons que le locus sélectionné se trouve à -3Kb en 5' du codon start soit -4.8 Kb en 5' du locus *EFbis* initialement détecté comme outlier (A3). Le coefficient de sélection est inféré être $s=0.011$, le taux de recombinaison local de $\rho=1.7$ cM/Mb et une atténuation de l'effet d'autostop dans la population du Golfe de Gascogne $\alpha=0.1$. Les fréquences initiales des neuf locus sont de 5' vers 3', $u_{0,1}=0.32$, $u_{0,2}=0.1$, $u_{0,3}=0.35$, $u_{0,4}=0.13$, $u_{0,5}=0.33$, $u_{0,6}=0.12$, $u_{0,7}=0.37$, $u_{0,8}=0.11$, $u_{0,9}=0.1$. Comme attendu, les fréquences initiales des mutations des branches internes sont plus fortes que les fréquences initiales des mutations de la base de la lignée autostoppée. Par contre, les différences entre mutations du même type sont très faibles si bien qu'une représentation graphique est possible (Figure 21). Enfin, une horloge moléculaire grossièrement calibrée sur la divergence avec une troisième espèce, *Mytilus trossulus*, nous permet de dater le balayage sélectif à 25000 [5000,100000] ans.

La sélection indirecte n'est pas la sélection. Une structure génétique forte à un locus peut-être la trace d'une sélection passée si le locus est neutre. Une des rares études à avoir recherché dans les détails la nature précise de la sélection responsable d'un F_{st} trop fort nous en procure l'exemple. Ce résultat pose la question de l'importance relative des deux scénarios, autostop local en environnement variable *vs.* autostop global en population structurée, dans l'explication des 5-10% locus outliers généralement observé dans les F_{st} scans.

IV.2- Association génétique / environnement lorsque l'adaptation locale interagit avec des incompatibilités génétiques endogènes.

Une observation m'est apparue difficilement explicable lorsque j'ai commencé à étudier les zones d'hybridations des *Mytilus* : les deux espèces occupent chacune tous les types d'habitats lorsqu'elles sont en allopatrie, mais semblent se spécialiser dans des habitats différents lorsqu'elles sont en contact l'une avec l'autre (zones d'hybridation). Les *M. edulis* seraient-elles meilleures que les *M. galloprovincialis* dans les estuaires bretons ou les lagunes portugaises mais incapables d'y parvenir ? Les *M. galloprovincialis* seraient-elles meilleures que les *M. edulis* sur les côtes rocheuses d'Oléron ou de la Mer du Nord mais incapables d'y parvenir ?

Lorsque les déséquilibres de liaison sont maintenus à une valeur forte, comme dans les zones d'hybridation, les gènes ne vont plus subir uniquement la sélection exercée sur eux mais aussi celle exercée sur d'autres locus. Ceci est pris en compte dans la théorie des clines multilocus (Barton 1983; Kruuk et al. 1999) qui décrit l'interaction entre les différents gènes d'isolement reproductif, et dans la théorie des barrières génétiques (Barton 1979, 1986) qui décrit l'interaction entre les marqueurs neutres et les gènes d'isolement. Cependant, d'autres interactions n'ont pas été étudiées. La théorie des clines multilocus considère des gènes impliqués dans un même mécanisme d'isolement reproductif : il s'agit en général d'une dépression d'hybridation due à des locus sous-dominants. Cependant, l'isolement reproductif ne se borne pas à la sous-dominance : les gènes impliqués dans le choix du partenaire, dans la phénologie, dans l'adaptation à des habitats différents ou le choix d'habitat interviennent également. L'interaction entre différents mécanismes agissant à des stades et à des échelles différentes n'a jamais été étudiée théoriquement. Or l'exemple de la zone d'hybridation des *Mytilus* en France, tout comme d'autres zones hybrides célèbres (Mallet *et al.*, 1990; Bert, Arnold, 1995; MacCallum *et al.*, 1998) montrent bien que plusieurs types de mécanismes peuvent agir simultanément. Je considère ici un modèle dans lequel interagissent deux types de sélection : la dépression d'hybridation (sélection endogène) et la spécialisation d'habitat (sélection exogène).

IV.2.1- Le modèle

Le modèle est très simple. Il s'agit d'un modèle déterministe, haploïde, à trois locus bi-alléliques (Locus 1 avec les allèles E_1 et G_1 , Locus 2 avec les allèles E_2 et G_2 , Locus 3 avec les allèles H_3 et h_3) disposés le long d'un chromosome unique. Deux locus (1 et 2) sont sous sélection endogène (valeurs sélectives $E_1E_2=1$, $E_1G_2=1-s$, $G_1E_2=1-s$, $G_1G_2=1$) et le troisième sous sélection exogène (habitat 1 : $H_3=1$, $h_3=1-t$, habitat 2 : $H_3=1-t$, $h_3=1$). La fitness est multiplicative (par exemple le génotype $E_1G_2h_3$ a dans l'habitat 1 une fitness de $(1-s)(1-t)$). Le modèle autorise la liaison physique mais dans les résultats présentés ici les locus ne sont pas liés. Le modèle est déterministe, il n'y a pas de dérive. La migration, la sélection et la reproduction se font dans cet ordre [NB : les modèles placent souvent la sélection avant la migration car ceci permet des approximations analytiques (Slatkin, 1973); ici (i) le modèle n'est pas développé en analytique, (ii)

cet ordre correspond mieux à la biologie des moules, (iii) pour un même coefficient de sélection, l'hétérogénéité génétique est mieux maintenue quand la sélection opère après la migration]. On considère un chapelet de dèmes successifs de type pas-japonais en une dimension borné par une barrière réfléchissante (les derniers dèmes auto-recrutent la moitié de leurs migrants). Trois types de migration peuvent opérer : (i) pas-japonais classique avec une migration de proche en proche à un taux $m/2$ de chaque côté, (ii) exponentielle (la migration vers le dème i est diminuée par $4|i|$ où i est la position du dème receveur par rapport au dème donneur) et (iii) uniforme (répartie uniformément sur un nombre de dème donné).

Dans un premier modèle, je considère d'une part un contact secondaire classique avec formation d'un cline endogène indépendant de l'environnement. D'autre part, l'environnement est structuré en deux patchs environnementaux. Une nouvelle mutation favorable dans l'un des habitats apparait par mutation et se fixe dans le patch où elle est favorable.

Dans un second modèle, je considère un habitat réparti au hasard sur toute l'aire de répartition. Un dème est assigné à l'habitat 1 avec une probabilité d' $1/2$ et à l'habitat 2 sinon. Les aléas de la colonisation sont modélisés en considérant une condition initiale particulière où un patch de 40 dèmes est fixés pour le génotype $E_1E_2H_3$ d'un côté, un patch de 40 dèmes pour le génotype $G_1G_2h_3$ de l'autre côté et une zone centrale de 70 dèmes fixées pour le génotype $E_1E_2H_3$ avec une probabilité $1/2$ ou pour le génotype $G_1G_2h_3$ sinon.

IV.2.2- Résultats

Sélection endogène et spécialisation d'habitat modérées.

Lorsque les deux types de sélection sont faibles, la structure s'organise indépendamment pour les différents types de locus car les déséquilibres de liaison sont faibles.

Sélection endogène forte et spécialisation d'habitat faible, les clines endogènes coïncident avec une barrière environnementale.

Lorsque la sélection endogène est forte les déséquilibres de liaison commencent à être assez forts pour que les locus endogènes commencent à subir l'effet de la sélection au locus de spécialisation d'habitat même si la sélection exogène est faible. De ce fait les clines des locus endogènes peuvent se déplacer pour venir se caler en coïncidence avec les clines exogènes au niveau de transition environnementale. Cependant, la variation environnementale ne crée aucune mosaïque à petite échelle sur les locus de sélection endogène. La figure 22 montre comment un cline endogène (traits gras) se fait emmener par une sélection exogène faible (traits fins) jusqu'à une frontière environnementale.

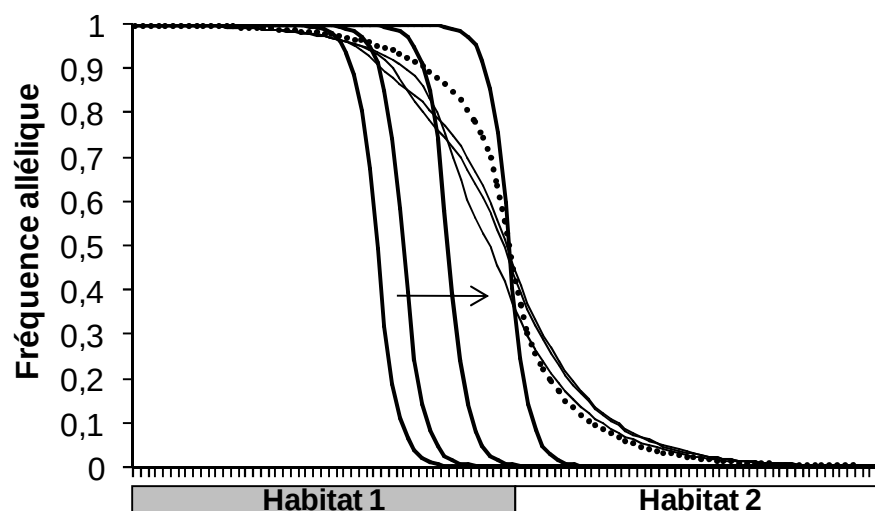


Figure 22 : Déplacement d'un cline endogène vers une frontière environnementale suite à la mise en place d'un cline exogène. En trait gras, cline endogène aux générations 2000, 4000, 6000 et 8000 (de gauche à droite). Cline exogène aux générations 2000, 4000 et 6000 en trait fin et à la génération 8000 en trait pointillé. Un cline endogène est formé par contact secondaire au niveau du dème 30 d'un pas-japonais de 100 dèmes. L'environnement est structuré en deux habitats de 50 dèmes. Une mutation favorable dans l'habitat 1 apparaît, se propage et forme un cline de fréquence allélique. La propagation dans les 20 dèmes à droite du cline endogène est à peine perturbée par ce dernier car la mutation y est favorable, cependant la forme du cline est légèrement déformée. Le cline endogène se déplace alors pour venir se caler en coïncidence du cline exogène et donc de la frontière environnementale. La migration est exponentielle (la migration vers le dème i est diminuée par $4|i|$ où i est la position du dème receveur par rapport au dème donneur), la sélection endogène est $s=0.9$, la sélection exogène est $t=0.02$. Les locus sont non liés.

Lorsqu'existe une sélection environnementale même faible, les clines des locus sous dépression d'hybridation peuvent venir coïncider secondairement avec un cline d'un locus exogène au niveau d'une frontière environnementale. Si l'existence d'une association génétique/environnement est la conséquence d'une adaptation locale, la barrière génétique qui permet la détection de cette association avec des marqueurs génétiques est la conséquence d'une sélection endogène. L'environnement explique la localisation de la structure génétique neutre mais pas la structure. Ce sont les incompatibilités génétiques qui expliquent la structure.

Sélections fortes, les associations génétique/environnement sont observées dans les clines

Quand la spécialisation d'habitat est suffisamment forte, la sélection qui s'exerce indirectement du locus exogène aux locus endogènes va devenir visible sous la forme d'association génétique/environnement aux locus endogènes. La figure 23 représente les variations de fréquences alléliques aux locus de spécialisation d'habitat (trait fin) et de sélection endogène (en gras) pour une condition initiale où les dèmes centraux sont aléatoirement colonisés par la sous-espèce 1 ou la sous-espèce 2.

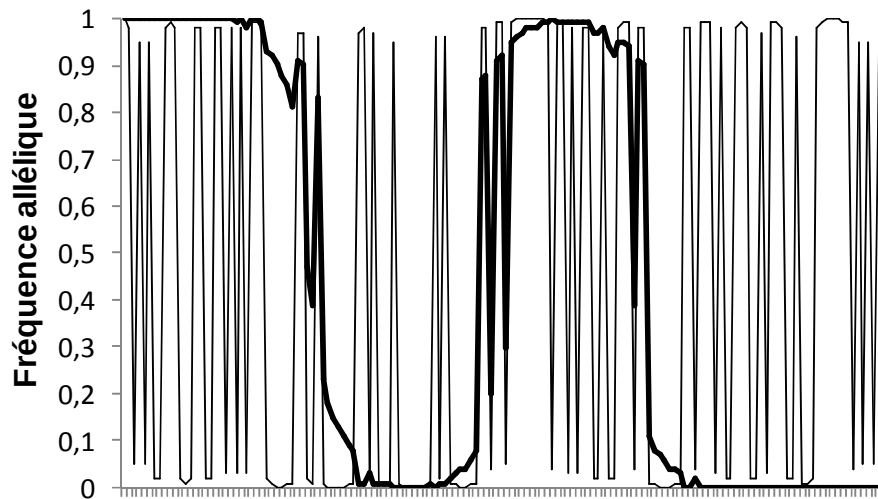


Figure 23 : Clines à l'équilibre de fréquence de l'allèle G_1 (allèle d'un locus de sélection endogène) en trait gras et de l'allèle H_3 (allèle du locus de spécialisation d'habitat) en trait fin, pour une sélection endogène forte ($s=0.9$) et une sélection exogène forte ($t=0.25$). Les conditions initiales sont des dèmes fixés aléatoirement pour l'un ou l'autre des génotypes spécifiques dans les 70 dèmes de la partie centrale d'un pas-japonais de 150 dèmes. Les deux habitats sont répartis uniformément et aléatoirement tout le long de l'aire de répartition

Malgré la force des déséquilibres de liaison, la recombinaison reste suffisante pour que les locus gardent une certaine indépendance. Les clines endogènes s'auto-organisent en fonction du hasard des conditions initiales. Ceci produit une mosaïque à grande échelle pour les gènes de sélection endogène (3 grands clines dans la figure 23). Cependant, ces gènes sont également affectés par l'hétérogénéité environnementale car ils subissent l'influence du locus de spécialisation d'habitat par l'intermédiaire des déséquilibres de liaison. Ils suivent donc des variations à petite échelle, corrélées à celles de l'habitat et à celles des gènes de spécialisation d'habitat. Ces variations à petite échelle ne sont perceptibles qu'au niveau des zones où l'introgression est suffisante (clines) car en-dehors de ces zones, les populations sont fixées pour les locus de sélection endogène. Les variations à petite échelle corrélées à l'habitat sont d'autant plus marquées que la sélection endogène est forte car les déséquilibres de liaison sont, de ce fait, plus forts. Il est remarquable de constater qu'en modélisant une zone d'hybridation à régime de sélection mixte ainsi qu'en simulant les aléas de la remise en contact, nous obtenons la structure en double mosaïque observée dans la zone d'hybridation entre *M. edulis* et *M. galloprovincialis*.

Quand l'environnement est à grain fin (la dispersion surpasse l'hétérogénéité spatiale de l'habitat) et la sélection environnementale assez forte, des associations génétique/environnement peuvent émerger dans les clines des locus de dépression d'hybridation. Alors que le polymorphisme d'adaptation local existe sur toute l'aire de répartition, les associations génétique/environnement ne sont observées que dans les

zones d'hybridation où des déséquilibres de liaison sont maintenus entre locus exogènes, endogènes et neutres. Les sous-espèces (définis par leur génotype endogène) ne sont pas adaptées à des environnements différents, mais possèdent toutes deux un polymorphisme d'adaptation environnementale qui n'est détecté que dans la zone d'hybridation. Sous cette hypothèse, si le contact se refaisait à nouveau, le déséquilibre de liaison pourrait même se créer dans une phase opposée (par ex. les backgrounds *M. edulis* en mode battu et les backgrounds *M. galloprovincialis* en mode abrité).

IV.2.3- Discussion

La modélisation de l'interaction entre sélection endogène (dépression d'hybridation) et sélection exogène (adaptation à un environnement variable) montre tout d'abord qu'une sélection exogène même faible peut attraper les clines endogènes pour les faire coïncider avec le/les cline(s) exogène(s) au niveau d'une frontière environnementale. A ma connaissance, ce résultat n'a jamais été popularisé dans la littérature. Les clines endogènes ne reçoivent aucune force pour rester à une position et il est attendu qu'ils se déplacent sous l'effet de facteurs stochastiques (dérive et migration). Barton a montré que les clines endogènes devaient venir se caler au niveau de régions où la dispersion est réduite soit par la présence d'une barrière physique soit parce que les densités de population sont moindres (Barton, 1979a). Nous trouvons donc ici une autre explication pour la localisation des zones d'hybridation. D'une manière générale, la littérature marine avait déjà négligé l'attendu de Barton. La moindre structure génétique qui coïncide avec un front océanique est interprétée comme la conséquence directe du front océanique qui limiterait la migration. Or, il existe très souvent des espèces au cycle de vie semblable qui ne présente pas de structure génétique au même endroit. Beaucoup de shifts génétiques coïncident aussi avec des frontières environnementales comme par exemple entre la Mer Baltique peu salée et la Mer du Nord (Johannesson, Andre, 2006; Hemmer-Hansen *et al.*, 2007). Ces études mettent fortement en avant la sélection environnementale. L'hypothèse alternative est que la structure génétique est en fait principalement la conséquence d'incompatibilité génétique endogène et qu'elle est venue coïncider secondairement avec une frontière environnementale. Les organismes marins à phase larvaire étant très dispersifs et l'environnement marin apparemment homogène, on s'attend à ce que de nombreuses barrières génétiques soient capturées par quelques barrières physiques ou/et quelques frontières environnementales (en fait, lorsque les masses d'eau se mélangent peu, barrière physique et frontière environnementale coïncident) qui représentent des hot spots de structure génétique (par ex. le petit et grand Belt danois et l'Øresund entre la Mer Baltique et la Mer du Nord, Johannesson, Andre, 2006; le front Almeria-Oran entre la Mer Méditerranée et l'Océan Atlantique, Patarnello *et al.*, 2007). Mon opinion est que l'explication de la localisation de ces hot spots a été interprétée comme l'explication de la structure génétique, alors que la structure s'explique principalement par l'existence de barrière génétique.

D'autre part, la modélisation de l'interaction entre sélection endogène et sélection exogène montre que lorsqu'il existe un polymorphisme de spécialisation d'habitat au sein de chaque sous-espèce, il se produit au sein des zones de contact une interaction avec les gènes responsables de la dépression hybride, de telle sorte que les allèles typiques d'une espèce se retrouvent en déséquilibre de liaison avec les gènes de spécialisation d'habitat. Ces derniers sont en quelque sorte recrutés pour participer à l'isolement reproductif en maintenant les deux pools géniques dans des habitats différents. Dans une telle situation, on s'attend à ce que des marqueurs neutres permettent de mettre à jour le polymorphisme de spécialisation d'habitat uniquement dans la zone d'hybridation, seul lieu où les allèles neutres se trouvent associés préférentiellement à des habitats particuliers, grâce aux déséquilibres de liaisons avec des allèles de spécialisation d'habitat.

V- Conclusion

L'effet indirect de la sélection n'est pas la sélection. J'espère avoir montré par les exemples exposés dans cette HDR qu'il existe possiblement un problème général dans l'étude de l'adaptation et de la spéciation à l'aide de marqueurs moléculaires neutres dont toute la valeur informative réside dans leur interaction particulière avec les autres gènes via des déséquilibres de liaison. La complexité des zones d'hybridation réside dans le fait que notre intuition nous pousse souvent au raccourci « gène = génome » ou « gène = espèce »; or, à partir du moment où il y a une hybridation effective, ce n'est pas si simple : le génome est constitué de compartiments hétérogènes soumis à des pressions de sélection différentes (par exemple dépression d'hybridation, sélection d'habitat, hétérosis) voire à aucune sélection (marqueurs neutres), et l'interaction entre compartiments n'est pas triviale. A l'inverse dans une population mono-spécifique bien délimitée, il peut paraître difficile d'imaginer qu'un locus puisse recevoir l'effet de la sélection de tous les délétères d'un génome et que cela n'affecte en rien son évolution. Lorsqu'il existe une variance de consanguinité, c'est pourtant ce que prédit la théorie (A22) et ce qui est observé (A23). Enfin, seule la sélection disruptive semble à première vue capable de générer une différenciation entre population plus forte que l'attendu neutre. Pourtant nous avons vu qu'une structure forte pouvait être la simple signature d'une sélection positive passée. Inversement, seule la sélection balancée semble à première vue capable de générer une différenciation entre population moins forte que l'attendu neutre. Si la structure générale du génome est due à une barrière semi-perméable au flux génique, ce résultat peut s'expliquer par la simple absence de gène d'isolement à proximité du locus ou par une introgression adaptative. Il n'y a pas vraiment d'attendu neutre, tous les locus sont plus ou moins sous l'influence indirecte de la sélection disruptive (barrière génétique semi-perméable).

Pour conclure cette HDR, je vous propose de revisiter un standard de l'adaptation locale marine, le cline de la LAP⁹⁴ dans le détroit de Long Island, un exemple souvent repris dans les manuels sur l'évolution (Gillespie, 1991; Mitton, 1997; David, Samadi, 2000). Koehn et ses collègues ont étudié le rôle d'une aminopeptidase initialement baptisée LAP (pour leucine amino-

peptidase) dans l'osmorégulation des moules. Les moules utiliseraient cette enzyme pour ajuster leur pression osmotique en modulant la concentration en acides aminés libres dans la cellule. Le polymorphisme électrophorétique a révélé plusieurs allèles chez *Mytilus edulis* dont trois assez fréquents, LAP⁹⁴, LAP⁹⁶ et LAP⁹⁸. L'allèle LAP⁹⁴ présente une constante catalytique légèrement plus forte que les deux autres allèles. Dans le détroit de Long Island prêt de New York aux Etats Unis, l'allèle LAP⁹⁴ passe d'une fréquence de 55% dans les populations océaniques à une fréquence de 15% dans le fond du détroit où la salinité est réduite (Figure 24, Koehn *et al.*, 1980).

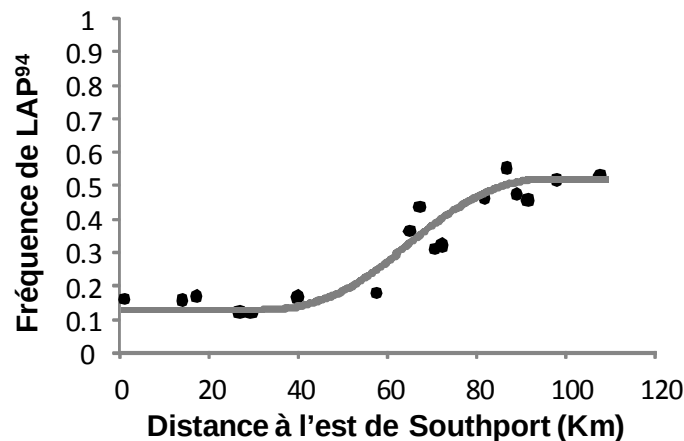


Figure 24 : Cline de l'allèle LAP⁹⁴ de la leucine aminopeptidase dans le détroit de Long Island.

De plus, il a été montré une mortalité plus forte des génotypes homozygotes pour l'allèle LAP⁹⁴ dans des conditions de salinités faibles, que ce soit dans la nature et au laboratoire (Hilbish, Koehn, 1985). Enfin, aucun autre locus analysé jusqu'à maintenant n'a dévoilé de structure entre les populations du détroit et les populations océaniques. Toutes les preuves semblent ainsi réunies pour démontrer que la sélection locale affecte le locus LAP directement.

Pourtant, beaucoup de zones d'ombres mériteraient d'être levées. La première concerne la forme du cline. Comment expliquer que la fréquence de LAP⁹⁴ 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 (Figure 24). La seconde zone d'ombre est qu'aucun avantage n'a été trouvé à LAP⁹⁴ dans les conditions océaniques. 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). Si LAP⁹⁴ 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 la population entière. 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 LAP⁹⁴ est observée. L'écart de fréquence est exactement le même de 55% au sud de Cap Cod à 15% au nord. Cette fois cependant, le shift est coïncidant avec d'autres locus (Koehn *et al.*, 1976). 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 LAP⁹⁴ a une fréquence de 40-60% chez les *M. trossulus* de l'Atlantique Ouest, tout comme chez les *M. trossulus* du Pacifique. En Europe, LAP⁹⁴ a une fréquence de 10% dans les populations *M. edulis* de la Mer du Nord et de 20% dans les populations *M. trossulus* de la Mer Baltique, une mer à la salinité de 10‰ seulement.

En regardant plus en détail la forme du cline et les fréquences alléliques de divers populations ou espèces, il apparaît donc très difficile d'expliquer le cline LAP par de la sélection 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 fitness (Hartl *et al.*, 1985). Je propose de façon beaucoup plus réaliste, que le locus LAP soit principalement un marqueur d'un background endogène en interaction avec des adaptations locales. Pour une raison restant à identifier, comme par exemple l'histoire et les aléas des remises en contact, la petite taille de la poche que représente le détroit ou la perte de certaines associations pour des raisons sélectives, l'introgression des populations du détroit a été beaucoup plus rapide pour d'autres marqueurs que pour la LAP, si bien que la différenciation n'est plus visible.

Afin de démontrer mon hypothèse, j'espère d'une part pouvoir analyser un grand nombre de locus entre le détroit et l'est de Long Island et d'autre part réussir à obtenir la séquence d'ADN codant pour le locus LAP. D'une manière plus générale, je coordonne un projet "Hi-Flo: The genetic basis and history of adaptive differentiation in high gene flow marine species" financé par une ANR blanche, qui propose de mettre à l'épreuve deux hypothèses pouvant résoudre le paradoxe de l'adaptation locale en milieu marin : (i) une longue histoire de conditions hors équilibre (colonisation, expansion, fragmentation) a son importance dans l'existence et le maintien d'adaptations à l'hétérogénéité environnementale, (ii) les espèces marines sont souvent génétiquement sous-structurées par des incompatibilités endogènes cachées, des barrières génétiques, qui viennent coïncider secondairement avec des zones de changement environnemental. La sélection indirecte n'est pas la sélection et son étude reste un champ de recherche qui n'est pas encore épuisé.

Références citées

- Balloux F, Amos W, Coulson T (2004) Does heterozygosity estimate inbreeding in real populations? *Mol Ecol* **13**, 3021-3031.
- Barton NH (1979a) The dynamics of hybrid zone. *Heredity* **43**, 341-359.
- Barton NH (1979b) Gene flow past a cline. *Heredity* **43**, 333-339.
- Barton NH (1986) The effects of linkage and density-dependant regulation on gene flow. *Heredity* **57**, 415-426.
- Barton NH (2000a) Estimating multilocus linkage disequilibria. *Heredity* **84**, 373-389.
- Barton NH (2000b) Genetic hitchhiking. *Philos Trans R Soc Lond B Biol Sci* **355**, 1553-1562.
- Barton NH, Bengtsson BO (1986) The barrier to genetic exchange between hybridising populations. *Heredity* **56**, 357-376.
- Barton NH, Briggs DEG, Eisen JA, Goldstein DB, Patel NH (2007) *Evolution* Cold Spring Harbor Laboratory Press, New York.
- Barton NH, Gale KS (1993) Genetic analysis of hybrid zones. In: *Hybrid zones and the evolutionary process* (ed. HARRISON RG), pp. 13-45. Oxford University Press, New York, Oxford.
- Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16**, 113-148.
- Bataillon T, Kirkpatrick M (2000) Inbreeding depression due to mildly deleterious mutations in finite populations: size does matter. *Genet Res* **75**, 75-81.
- Bates JA, Innes DJ (1995) Genetic variation among populations of *Mytilus ssp* in eastern Newfoundland. *Marine Biology* **124**, 417-424.
- Bazin E, Glemin S, Galtier N (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science* **312**, 570-572.
- Beaumont MA (2005) Adaptation and speciation: what can F_{st} tell us? *Trends in Ecology & Evolution* **20**, 435-440.
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proc. Roy. Soc. Ser. B* **263**, 1619-1626.
- Bengtsson BO (1985) The flow of genes through a genetic barrier. In: *Evolution: essays in honor of John Maynard Smith*. (eds. Greenwood JJ, Harvey PH, Slatkin M), pp. 31-42. Cambridge university Press, Cambridge.
- Berry AJ, 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.
- Bryant D, Moulton V (2004) NeighborNet: an agglomerative algorithm for the construction of planar phylogenetic networks. *Mol Biol Evol* **21**, 255-265.
- Buggs RJ (2007) Empirical study of hybrid zone movement. *Heredity* **99**, 301-312.
- Burton RS (1998) Intraspecific phylogeography across the Point Conception biogeographic boundary. *Evolution* **52**, 734-745.
- Chakraborty R (1987) Biochemical heterozygosity and phenotypic variability of polygenic traits. *Heredity* **59**, 19-28.
- Charlesworth B (1994) The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genet. Res.* **63**, 213-227.
- Charlesworth B (2009) Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* **10**, 195-205.

- Charlesworth B, Charlesworth D, Barton NH (2003) The effects of genetic and geographic structure on neutral variation. *Annual Review of Ecology and Systematics* **34**, 99-125.
- Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289-1303.
- Charlesworth D (1991) The apparent selection on neutral marker loci in partially inbreeding populations. *Genet. Res.* **57**, 159-175.
- Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet* **2**, e64.
- Clark AG (2006) Genomics of the evolutionary process. *Trends Ecol Evol* **21**, 316-321.
- Coghlan B, Gosling E (2007) Genetic structure of hybrid mussel populations in the west of Ireland: two hypotheses revisited. *Marine Biology* **150**, 841-852.
- 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.
- 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.
- David P (1997) Modeling the genetic basis of heterosis: Tests of alternative hypotheses. *Evolution* **51**, 1049-1057.
- David P (1998) Heterozygosity-fitness correlations: new perspectives on old problems. *Heredity* **80**, 531-537.
- David P (1999) A quantitative model of the relationship between phenotypic variance and heterozygosity at marker loci under partial selfing. *Genetics* **153**, 1463-1474.
- David P, Perdieu MA, Pernot AF, Jarne P (1997) Fine-grained spatial and temporal population genetic structure in the marine bivalve *Spisula ovalis*. *Evolution* **51**, 1318-1322.
- David P, Samadi S (2000) *La théorie de l'évolution : une logique pour la biologie* Flammarion, Manchecourt.
- Eyre-Walker A, Keightley PD (1999) High genomic deleterious mutation rates in hominids. *Nature* **397**, 344-347.
- Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. *Nat Rev Genet* **8**, 610-618.
- Faure MF (2007) *Génétique des barrières semi-perméables au flux génique dans la zone d'hybridation des moules *Mytilus edulis* et *M. galloprovincialis**, Montpellier 2.
- Fauvelot C, Lemaire C, Planes S, Bonhomme F (2007) Inferring gene flow in coral reef fishes from different molecular markers: which loci to trust? *Heredity* **99**, 331-339.
- Felsenstein J (1974) The evolutionary advantage of recombination. *Genetics* **78**, 737-756.
- 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.
- Freeman JL, Perry GH, Feuk L, et al. (2006) Copy number variation: new insights in genome diversity. *Genome Res* **16**, 949-961.
- Gardner JPA (1994) The structure and dynamics of naturally occurring hybrid *Mytilus edulis* and *M. galloprovincialis* (Bivalvia, Mollusca) populations: review and interpretation. *Archiv für Hydrobiologie (supplement band)* **99**, 37-71.
- Gavrilets S (2004) *Fitness landscapes and the origin of species* Princeton university press, Princeton and Oxford.
- Gillespie JH (1991) *The causes of molecular evolution* Oxford University Press, Oxford, UK.
- Gillespie JH (2000a) Genetic drift in an infinite population. The pseudohitchhiking model. *Genetics* **155**, 909-919.

- Gillespie JH (2000b) The neutral theory in an infinite population. *Gene* **261**, 11-18.
- Goldman N, Yang Z (1994) A codon based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**, 725-736.
- Gould SJ, Lewontin RC (1979) The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc R Soc Lond B Biol Sci* **205**, 581-598.
- Hansson B, Westerberg L (2008) Heterozygosity-fitness correlations within inbreeding classes: local or genome-wide effects? . *Conservation Genetics* **9**, 73-83.
- Hartl DL, Clark AG (1989) *Principles of Population Genetics*, 2nd edn. Sinauer.
- Hartl DL, Dykhuizen DE, Dean AM (1985) Limits of adaptation: the evolution of selective neutrality. *Genetics* **111**, 655-674.
- Hedgcock D (1994) Does variance in reproductive success limit effective population sizes of marine organisms? In: *Genetics and Evolution of Aquatic Organisms* (ed. Beaumont AR), pp. 122-134. Chapman & Hall, London.
- Hedrick PW (1986) Genetic polymorphism in heterogeneous environments: A decade later. *Annual Review of Ecology and Systematics* **17**, 535-566.
- Hedrick PW (2006) Genetic polymorphism in heterogeneous environments: The age of genomics. *Annual Review of Ecology and Systematics* **37**, 67-93.
- Hedrick PW, Ginevan ME, Ewing EP (1976) Genetic Polymorphism in Heterogeneous Environments. *Annual Review of Ecology and Systematics* **7**, 1-32.
- Hemmer-Hansen J, Nielsen EE, Gronkjaer P, Loeschcke V (2007) Evolutionary mechanisms shaping the genetic population structure of marine fishes; lessons from the European flounder (*Platichthys flesus* L.). *Mol Ecol* **16**, 3104-3118.
- Hewitt GM (1988) Hybrid zones, natural laboratories for evolutionary studies. *TREE* **3**, 158-167.
- 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 (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philos Trans R Soc Lond B Biol Sci* **359**, 183-195; discussion 195.
- Hey J, Machado CA (2003) The study of structured populations--new hope for a difficult and divided science. *Nat Rev Genet* **4**, 535-543.
- Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**, 747-760.
- Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc Natl Acad Sci U S A* **104**, 2785-2790.
- Hilbish TJ, Koehn RK (1985) The physiological basis of natural selection at the *LAP* locus. *Evolution* **39**, 1302-1317.
- Hilbish TJ, Timmons J, Agrawal V, Schneider KR, Gilg MR (2003) Estuarine habitats protect hybrid mussels from selection. *Journal of Experimental Marine Biology and Ecology* **292**, 177-186.
- Hoekstra HE, Coyne JA (2007) The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61**, 995-1016.
- 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.
- Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153-159.
- Hughes AL (2007) Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity* **99**, 364-373.

- Hughes AL (2008) The origin of adaptive phenotypes. *PNAS* **105**, 13193-13194.
- Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**, 254-267.
- Huvet A, Boudry P, Ohresser M, Delsert C, Bonhomme F (2000) Variable microsatellites in the Pacific cupped oyster *Crassostrea gigas* and other cupped oyster species. *Anim Genet* **31**, 71-72.
- Ingvarsson PK, Whitlock MC (2000) Heterosis increases the effective migration rate. *Proc R Soc Lond B Biol Sci* **267**, 1321-1326.
- Jacquard A (1975) Inbreeding: one word, several meanings. *Theor. Pop. Biol.* **7**, 338-363.
- Jakobsson M, Scholz SW, Scheet P, *et al.* (2008) Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* **451**, 998-1003.
- Johannesson K, Andre C (2006) Life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Mol Ecol* **15**, 2013-2029.
- Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: *Mammalian protein metabolism* (ed. Munro HM), pp. 21-132. Academic Press, New York.
- Karl SA, Avise JC (1992) Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. *Science* **256**, 100-102.
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111-120.
- Kimura M (1983) *The neutral theory of molecular evolution* Cambridge University Press, Cambridge, UK.
- Kliman R, Andolfatto P, Coyne JA, *et al.* (2000) The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* **156**, 1913-1931.
- 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.
- Latter BD (1998) Mutant alleles of small effect are primarily responsible for the loss of fitness with slow inbreeding in *Drosophila melanogaster*. *Genetics* **57**, 159-175.
- Launey S, Hedgecock D (2001) High Genetic Load in the Pacific Oyster *Crassostrea gigas*. *Genetics* **159**, 255-265.
- 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. *J. Hered.* **93**, 331-351.
- Lazzaro BP, Scurman BK, Clark AG (2004) Genetic basis of natural variation in *D. melanogaster* antibacterial immunity. *Science* **303**, 1873-1876.
- 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.
- Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology and 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.
- Lewontin RC, Krakauer J (1973) Distribution of gene frequency as test of the theory of the selective neutrality of polymorphism. *Genetics* **74**, 175-195.
- Li H, Stephan W (2006) Inferring the demographic history and rate of adaptive substitution in *Drosophila*. *PLoS Genetics* **2**, e166.

- Lieutenant-Gosselin M, Bernatchez L (2006) Local heterozygosity-fitness correlations with global positive effects on fitness in threespine stickleback. *Evolution* **60**, 1658-1668.
- 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. *Mar Ecol Prog Ser* **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.
- Lynch M (2007) The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc Natl Acad Sci U S A* **104**, 8597-8604.
- MacCallum CJ, Nurnberger B, Barton NH, Szymura JM (1998) Habitat preference in the *Bombina* hybrid zone in Croatia. *Evolution* **52**, 227-239.
- Malécot G (1969) Consanguinité panmictique et consanguinité systématique. *Ann. Génét. Sél. Anim.* **1**, 237-242.
- 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.
- Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23-35.
- Mitta G, Hubert F, Dzynda EA, Boudry P, Roch P (2000) Mytilin B and MGD2, two antimicrobial pep-tides of marine mussels: gene structure and expression analysis. *Developmental & Comparative Immunology* **24**, 381-393.
- Mitton JB (1997) *Selection in natural populations* Oxford University Press, New York.
- Nei M (2007) The new mutation theory of phenotypic evolution. *PNAS* **104**, 12235-12242.
- Nielsen R (2005) Molecular signatures of natural selection. *Ann Rev Genet* **39**, 197-218.
- Nielsen R, Bustamante CD, Clark AG, *et al.* (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biology* **3**, e170.
- Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* **158**, 885-896.
- Nosil P (2008) Speciation with gene flow could be common. *Mol Ecol* **17**, 2103-2106.
- Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Mol Ecol* **18**, 375-402.
- Ohta T (1971) Associative overdominance caused by linked detrimental mutations. *Genet Res* **18**, 277-286.
- Ohta T (1973) Effect of linkage on behavior of mutant genes in finite populations. *Theoretical Population Biology* **4**, 145-162.
- Orr HA, Presgraves DC (2000) Speciation by postzygotic isolation: forces, genes and molecules. *Bioessays* **22**, 1085-1094.
- Palsson S, Pamilo P (1999) The effects of deleterious mutations on linked, neutral variation in small populations. *Genetics* **153**, 475-483.
- Palumbi SR (1992) Marine speciation on a small planet. *Trends in Ecology and Evolution.* **7**, 114-118.
- Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics* **25**, 547-572.
- Pamilo P, Palsson S (1998) Associative overdominance, heterozygosity and fitness. *Heredity* **81**, 381-389.
- Patarnello T, Volckaert FA, Castilho R (2007) Pillars of Hercules: is the Atlantic-Mediterranean transition a phylogeographical break? *Mol Ecol* **16**, 4426-4444.
- Pemberton J (2004) Measuring inbreeding depression in the wild: the old ways are the best. *Trends Ecol Evol* **19**, 613-615.

- Perron GG, Zasloff M, Bell G (2005) Experimental evolution of resistance to an antimicrobial peptide. *Proc. R. Soc. B* **273**, 251–256.
- Perry GH, Yang F, Marques-Bonet T, *et al.* (2008) Copy number variation and evolution in humans and chimpanzees. *Genome Res* **18**, 1698–1710.
- Pujolar JM, Maes GE, Vancoillie C, Volckaert FA (2005) Growth rate correlates to individual heterozygosity in the European eel, *Anguilla anguilla* L. *Evolution* **59**, 189–199.
- Qiu J-W, remblay R, Bourget E (2002) Ontogenetic changes in hyposaline tolerance in the mussels *Mytilus edulis* and *M. trossulus*: implications for distribution. *Mar. Ecol. Prog. Ser.* **228**, 143–152.
- Redon R, Ishikawa S, Fitch KR, *et al.* (2006) Global variation in copy number in the human genome. *Nature* **444**, 444–454.
- Riginos C, McDonald JH (2003) Positive selection on an acrosomal sperm protein, M7 lysin, in three species of the mussel genus *Mytilus*. *Mol Biol Evol* **20**, 200–207.
- 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.
- Santiago E, Caballero A (2005) Variation after a selective sweep in a subdivided population. *Genetics* **169**, 475–483.
- Schlenke TA, Begun DJ (2003) Natural selection drives drosophila immune system evolution. *Genetics* **164**, 1471–1480.
- Sebat J, Lakshmi B, Troge J, *et al.* (2004) Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528.
- Silva EP, Skibinski DOF (2009) Allozymes and nDNA markers show different levels of population differentiation in the mussel *Mytilus edulis* on British coasts *Hydrobiologia* **620**, 25–33.
- 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, Wiehe T (1998) Genetic hitch-hiking in a subdivided population. *Genet Res* **71**, 155–160.
- Smith NGC, Eyre-Walker A (2002) Adaptive protein evolution in *Drosophila*. *Nature* **415**, 1022–1024.
- Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol* **14**, 671–688.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.
- Tennessen JA (2005) Molecular evolution of animal antimicrobial peptides: widespread moderate positive selection. *J Evol Biol* **18**, 1387–1394.
- Tsitroni A, Rousset F, David P (2001) Heterosis, marker mutational processes and population inbreeding history. *Genetics* **159**, 1845–1859.
- Turelli M, Orr HA (2000) Dominance, epistasis and the genetics of postzygotic isolation. *Genetics* **154**, 1663–1679.
- Vitalis R, Dawson KJ, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics* **158**, 1811–1823.
- Wakeley J, Hey J (1997) Testing speciation models with DNA sequence data. In: *Molecular approaches to ecology and evolution* (eds. DeSalle R, Schierwater B). Birkhauser Verlag, Basel.
- Ward RD, Skibinski DOF, Woodward M (1992) Protein heterozygosity, protein structure, and taxonomic differentiation. *Evol. Biol.* **26**, 73–159.
- Weir BS, Avery PJ, Hill WG (1980) Effect of mating structure on variation in inbreeding. *Theor. Pop. Biol.* **18**, 396–429.

- Wheat CW, Watt WB, Pollock DD, Schulte PM (2006) From DNA to fitness differences: sequences and structures of adaptive variants of *Colias* Phosphoglucose Isomerase (PGI). *Mol Biol Evol* **23**, 499–512.
- Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15**, 568-573.
- Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**, 431-449.
- Yeaman MR, Yount NY (2003) Mechanisms of antimicrobial peptide action and resistance. *Pharmacol Rev* **55**, 27-55.
- Yount NY, Yeaman MR (2004) Multidimensional signatures in antimicrobial peptides. *Proc Natl Acad Sci U S A* **101**, 7363-7368.
- Zaslloff M (2002) Antimicrobial peptides of multicellular organisms. *Nature* **415**, 389-395.

Annexes

Annexe 1

Faure B, Jollivet D, Tanguy A, Bonhomme F & Bierne N (2009). Speciation in the deep sea: multi-locus analysis of divergence and gene flow between two hybridizing species of hydrothermal vent mussels. **PLoS ONE**, 4: e6485.

Annexe 2

Boon E, Faure MF & Bierne N (2009) The flow of antimicrobial genes through a genetic barrier between *Mytilus edulis* and *M. galloprovincialis*. **Journal of Molecular Evolution**, 68:461-74.

Annexe 3

Faure MF, David P, Bonhomme F & Bierne N (2008) Genetic hitchhiking in a subdivided population of *Mytilus edulis*. **BMC Evolutionary Biology** 8:164.