



HAL
open science

Génomique de l'hybridation anthropogénique chez les moules *Mytilus* spp. et valeur sélective des hybrides

Alexis Simon

► To cite this version:

Alexis Simon. Génomique de l'hybridation anthropogénique chez les moules *Mytilus* spp. et valeur sélective des hybrides. Sciences agricoles. Université Montpellier, 2019. Français. ⟨NNT: 2019MONTG032⟩. ⟨tel-02481299⟩

HAL Id: tel-02481299

<https://theses.hal.science/tel-02481299v1>

Submitted on 17 Feb 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.



HAL Authorization

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

Filière EERGP - Écologie, Évolution, Ressources Génétiques, Paléobiologie
Champ disciplinaire Sciences de l'Évolution et de la Biodiversité

École doctorale GAIA

Unité de recherche ISEM

Génomique de l'hybridation anthropogénique chez les moules *Mytilus* spp. et valeur sélective des hybrides

Présentée par Alexis SIMON

Le 12/12/2019

Sous la direction de Nicolas BIERNE

Devant le jury composé de

Florence DÉBARRE, chargée de recherche, CNRS

Oscar GAGGIOTTI, professeur, Université de Saint-Andrews, Écosse

Maud TENAILLON, directrice de recherche, CNRS

Patrice DAVID, directeur de recherche, CNRS

John WELCH, professeur adjoint, Université de Cambridge, Grande-Bretagne

Nicolas BIERNE, directeur de recherche, CNRS

Rapportrice

Rapporteur

Examinatrice

Président du jury

Examineur

Directeur



UNIVERSITÉ
DE MONTPELLIER

Résumé

Les activités anthropiques sont en train de créer de nouveaux contacts entre des lignées génétiques différenciées, nous donnant accès à la phase initiale et à des réplicats de contacts secondaires. Le complexe *M. edulis* est composé d'un ensemble d'espèces et de lignées partiellement isolées reproductivement, présentant une grande diversité de contacts secondaires et de flux de gènes, constituant un modèle de choix pour l'étude de l'hybridation et de la spéciation. Nous avons profité de plusieurs événements de contact secondaires, à la fois naturels, anthropiques, et expérimentaux, pour étudier leurs effets sur les patrons génomiques d'admixture. En effet, plusieurs événements d'introduction de l'espèce *M. galloprovincialis* dans plusieurs ports français mais aussi en Norvège, provenant de lignées différentes et s'étant admixées avec les espèces locales, ont été identifiés d'abord à l'aide d'une centaine de marqueurs génétiques informatifs. Les populations portuaires forment des zones de contact à très fine échelle entre les populations introduites à l'intérieur des ports et les populations locales à l'extérieur. Ces populations portuaires présentent une introgression importante de gènes des lignées locales avec lesquelles elles sont en contact. La comparaison de ces événements d'admixture anthropiques avec des cas impliquant d'autres lignées mais aussi avec des croisements expérimentaux ont permis d'identifier des parallélismes dans la distorsion des fréquences alléliques lorsque les mêmes lignées sont impliquées. Afin d'explorer les patrons génomiques produits par ces événements, nous analysons 156 génomes de populations de référence et d'individus admixés. Cette analyse permet de montrer que les corrélations de distorsions entre événements d'admixture sont conservés à l'échelle génomique. Enfin, un modèle théorique permettant de prédire la valeur sélective des hybrides a été développé en utilisant le cadre du modèle géométrique de Fisher. Ce modèle considère les interactions épistatiques à l'échelle du génome, émergeant de la co-adaptation des mutations produites lors de la divergence. Il permet entre autres de s'affranchir des limitations des modèles classiques d'incompatibilités génétiques tout en conservant des prédictions en accord avec la théorie et les données de l'isolement post-zygotique. Ce modèle, simple et flexible, permet de prédire de nombreuses observations empiriques de l'étude de l'hybridation, dont les distorsions de ségrégation chez les moules. Il permet enfin d'expliquer que les populations admixées retrouvent une valeur sélective élevée bien qu'aillant une combinaison d'allèles provenant des deux génomes parentaux.

Abstract

Anthropogenic activities are creating new contacts between genetically differentiated lineages, providing access to the initial phase and replicates of secondary contacts. The complex *M. edulis* contains several species and lineages partially reproductively isolated, exhibiting a large diversity of secondary contacts and levels of gene flow. It constitutes therefore a good model for the study of hybridisation and speciation. We took advantage of several secondary contact events, both natural, anthropogenic and experimental, to investigate their effects on genomic admixture patterns. Indeed, events of *M. galloprovincialis* introduction from two different lineages, implicating admixture between the introduced species and the native species, have been identified in several French ports and in Norway. The introduction has first been identified with a hundred ancestry informative markers. Mussel port populations are forming fine scale contact zones between the inside of the ports and local populations located outside. Those port populations exhibit an important rate of introgression from local lineages with which they are in contact. The comparison of those anthropogenic admixture events with cases implicating other lineages and experimental crosses allowed us to identify parallelisms in allele frequency distortions when similar lineages are hybridising. In the aim of investigating the genomic patterns produced by such events, we analysed 156 genomes of reference and admixed populations, showing the correlation of distortions between events is conserved at the genomic scale. Finally, we developed a theoretical model for the prediction of hybrid fitness using the framework of Fisher's geometric model. This model takes into account genome scale epistatic interactions, emerging from the co-adaptation of mutations produced during divergence between the two lineages. The model is simple and flexible and predicts many empirical observations of hybridization research including segregation distortions in mussels. Finally, it explains that admixed populations retrieve high fitness while being a combination of alleles from the two parental genomes.

Publications en lien direct avec la thèse :

Acceptées ou soumises

Simon, A., N. Bierne et J. J. Welch (2018). « Coadapted Genomes and Selection on Hybrids : Fisher's Geometric Model Explains a Variety of Empirical Patterns ». *Evolution Letters* 2.5, p. 472-498. DOI : 10.1002/evl3.66.

Simon, A. et M. Duranton (2018). « Digest : Demographic Inferences Accounting for Selection at Linked Sites ». *Evolution*. DOI : 10.1111/evo.13504.

Simon, A., C. Fraïsse, T. El Ayari, C. Liutard-Haag, P. Strelkov, J. J. Welch et N. Bierne (2019a). « Local Introgression at Two Spatial Scales in Mosaic Hybrid Zones of Mussels ». *bioRxiv*. DOI : 10.1101/818559. Soumis.

Simon, A. et al. (2019b). « Replicated Anthropogenic Hybridisations Reveal Parallel Patterns of Admixture in Marine Mussels ». *Evolutionary Applications*. DOI : 10.1111/eva.12879. Prépubl.

Publications annexes :

Travaux sur le cancer transmissible chez *Mytilus* spp. (cf. partie 5.5)

Riquet, F., A. Simon et N. Bierne (2017). « Weird Genotypes? Don't Discard Them, Transmissible Cancer Could Be an Explanation ». *Evolutionary Applications* 10, p. 140-145. DOI : 10.1111/eva.12439.

Burioli, E. et al. (2019). « Implementation of Various Approaches to Study the Prevalence, Incidence and Progression of Disseminated Neoplasia in Mussel Stocks ». *Journal of Invertebrate Pathology* 168, p. 107271. DOI : 10.1016/j.jip.2019.107271.

Yonemitsu, M. A. et al. (2019). « A Single Clonal Lineage of Transmissible Cancer Identified in Two Marine Mussel Species in South America and Europe ». *eLife* 8, e47788. DOI : 10.7554/eLife.47788.

Table des matières

Introduction	1
i.1 Spéciation, hybridation et isolement reproductif	1
i.1.1 L'isolement reproductif : un ensemble de mécanismes et de scénarios	1
i.1.2 Les bases génétiques de l'isolement reproductif	5
i.1.3 Génomique de l'isolement reproductif et des flux de gènes	9
i.1.4 Contacts secondaires : expérimentations pour l'étude de la spéciation	10
i.2 Introductions biologiques et hybridation	11
i.2.1 Le processus d'introduction biologique	11
i.2.2 Les introductions en milieu marin	12
i.2.3 La génétique des invasions : la place de l'hybridation	15
i.2.4 Introduction et hybridation : une expérience évolutive	17
i.3 Le modèle d'étude <i>Mytilus</i> : au hasard des rencontres	18
i.3.1 Le complexe d'espèces de la moule bleue : composition, distribution et histoire évolutive	18
i.3.2 Modèle pour l'étude de l'hybridation et de la spéciation	21
i.3.3 <i>Mytilus</i> à l'âge de la génomique	23
i.3.4 Écologie et traits d'histoire de vie liés à l'hybridation et aux introductions	25
i.3.5 <i>M. galloprovincialis</i> une espèce au caractère invasif	26
i.4 Objectifs de la thèse et structure du document	27
1 Comètes d'introggression dans le complexe d'espèces <i>M. edulis</i>	33
1.1 Contexte et résumé de l'étude	33
1.1.1 Obtenir une vision plus précise des liens au sein du complexe	33
1.1.2 Introggressions locales et zones de tension	34
1.1.3 Résumé	34
1.2 Article 1	35
2 Admixture anthropogénique : analyse spatiale et génétique	77
2.1 Contexte et résumé de l'étude	77
2.1.1 Découverte d'introductions intra-européennes	77
2.1.2 Les lignées de <i>M. galloprovincialis</i> peu souvent identifiées	78
2.1.3 Les ports : points chauds et ponts d'invasion	79
2.1.4 Résumé	80
2.2 Article 2	80
3 Analyse des patrons d'admixture à l'échelle génomique	171
3.1 Contexte et résumé de l'étude	171
3.1.1 Des cartes de plus en plus précises de l'admixture	171
3.1.2 Vision génomique des admixtures entre espèces du complexe <i>M. edulis</i>	172

3.1.3	Résumé	172
3.2	Article 3	173
4	Modèle polygénique et prédiction de la valeur sélective des hybrides	207
4.1	Contexte et résumé de l'étude	207
4.1.1	Modèles d'isolement génétique classiques face aux observations	207
4.1.2	Passage à une vision polygénique	208
4.1.3	Résumé	209
4.2	Article 4	209
5	Discussion générale et perspectives	257
5.1	La continuité du complexe d'espèces <i>Mytilus</i>	257
5.2	Les moules des docks	259
5.2.1	Une invasion en devenir ou est-ce déjà le cas?	259
5.2.2	Celle dont il ne faut pas prononcer le nom : l'espèce hybride?	261
5.3	Un impact génétique? Pollution génétique, conservation et gestion	263
5.4	Considérations théoriques	265
5.4.1	Résolution de l'admixture	265
5.4.2	Devenir d'un allèle introduit	266
5.5	Moule des docks, vecteur d'un parasite d'un nouveau genre?	269
5.6	Conclusion	271
	Annexe : <i>Evolution digest</i>	273
	Liste des figures	278
	Liste des tableaux	278
	Liste des encadrés	278
	Glossaire	279
	Bibliographie	281
	Remerciements	303

Introduction

i.1 Spéciation, hybridation et isolement reproductif

La spéciation est un processus graduel menant à la séparation d'une espèce en deux entités distinctes ne pouvant plus échanger de matériel génétique (i.e. deux espèces). Ce processus est généré par un ensemble multifactoriel de mécanismes d'isolement reproductif, rendant complexe sa compréhension et son étude. Durant la spéciation, deux lignées passent par un stade d'isolement reproductif partiel – mis à l'épreuve lors d'hybridations naturelles ou non – correspondant à ce que l'on peut appeler la « zone grise » de la spéciation (De Queiroz 2007 ; Roux et al. 2016). L'isolement reproductif peut être défini comme l'absence de flux de gènes, due à l'inviabilité ou l'infertilité des hybrides ou à la rareté de leur formation (Barton et Hewitt 1985). Le test de cet isolement nécessite des événements d'hybridation entre des individus de deux populations génétiquement distinctes, soit dans la nature, soit au laboratoire.

La spéciation se déroule rarement de manière totalement allopatrique ou instantanée. Ainsi, l'hybridation est un phénomène courant lors de ce processus (Abbott et al. 2013). L'admixture désigne la résultante de l'hybridation à l'échelle populationnelle, i.e. le mélange de matériel génétique au sein d'un même génome. Afin de comprendre ce qui limite le flux génique entre deux populations, il est nécessaire d'étudier la valeur sélective des hybrides ainsi que le résultat d'événements d'admixture ayant été soumis à la sélection. Les zones de contact entre deux populations divergentes, secondaires ou initiales, sont des laboratoires naturels de la spéciation (Hewitt 1988). Dans ces zones de contacts, appelées aussi zones d'hybridation, des ensembles de gènes ayant divergé sont mélangés par l'action de l'hybridation et de la recombinaison, et les nouvelles combinaisons ainsi créées sont testées par la sélection naturelle.

i.1.1 L'isolement reproductif : un ensemble de mécanismes et de scénarios

Les mécanismes

Le flux de gènes est dépendant de facteurs génétiques, phénotypiques, démographiques, spatiaux et environnementaux (Figure 1). L'isolement reproductif peut se définir sur deux axes majoritaires : (i) la période du cycle de vie de l'organisme durant laquelle il s'exprime, souvent

classifiée entre pré-zygotique (avant la fécondation) et post-zygotique (après la fécondation), et (ii) la dépendance à l'environnement, sous la dichotomie intrinsèque et extrinsèque (ou endogène/exogène). Comprendre la spéciation, c'est principalement identifier quelles sont les barrières aux flux géniques, actuels et passés, et comprendre comment elles se sont mises en place au cours du temps. Au cours du XX^{ème} siècle, un nombre important de modèles ont été développés pour tester la contribution des différents mécanismes à la spéciation (Kirkpatrick et Ravigné 2002).

L'isolement pré-zygotique comprend les mécanismes écologiques, sexuels, comportementaux et biologiques limitant les possibilités de fécondation entre les gamètes d'individus de populations différentes (Figure 1b). L'absence d'un événement de reproduction va être un facteur clé du processus d'isolement reproductif, qui va opérer une limitation du flux de gènes sur l'ensemble du génome.

L'isolement post-zygotique repose sur une contre sélection des hybrides, liée ou non à l'environnement (Figure 1a). Certaines combinaisons génétiques créées par la recombinaison des génomes parentaux n'arboreront pas forcément les facteurs génétiques responsables de cet isolement ; elles vont pouvoir survivre et ainsi permettre un certain niveau de flux génique. Ces génotypes hybrides de valeur sélective proche de celle des parents constituent des ponts empruntés par les allèles d'une population parentale pour passer dans l'autre population parentale. L'étude de la valeur sélective d'une diversité de génotypes hybrides est un sujet de recherche clé dans la compréhension de la sélection post-zygotique (Turelli et Orr 2000). Le niveau de flux de gènes va être entièrement dépendant de la force de sélection, des interactions épistatiques possibles et de l'architecture des incompatibilités. De manière simplifiée, un locus contribuant à la contre sélection des hybrides créera une diminution du flux de gènes autour de sa position sur le chromosome, et son influence diminuera avec la distance génétique (Figure 1a). Lorsqu'un nombre suffisant de locus densément positionnés sur le génome combinent leurs effets, la barrière peut devenir génomique (Barton 1983 ; Barton et Bengtsson 1986 ; Kruuk et al. 1999) bien que l'effet reste plus fort autour des locus barrières. Cependant, le flux de gènes reste possible loin des locus barrières via l'hybridation, i.e. tant que les rétro-croisements (*backcrosses*, BC) restent possibles, même s'ils sont en faible fréquence et fortement contre-sélectionnés.

Différents mécanismes d'isolement reproductif peuvent agir de concert. Entre certaines espèces d'oursins par exemple, il existe des incompatibilités gamétiques contribuant à l'isolement reproductif (Lessios 2009). Ces incompatibilités sont notamment liées à une protéine de surface du sperme, *bindin*, nécessaire à la reconnaissance et la fusion avec la membrane ovulaire (Zigler et al. 2005 ; Lessios 2009). S'ajoute souvent à cet isolement pré-zygotique, des réductions de survie ou de capacité reproductive des hybrides, corrélés avec le temps de divergence entre espèces (Lessios 2009).

L'isolement reproductif a été beaucoup étudié chez les souris domestiques. Une zone d'hybridation étroite entre *Mus musculus domesticus* et *M. m. musculus* s'étend du Danemark à

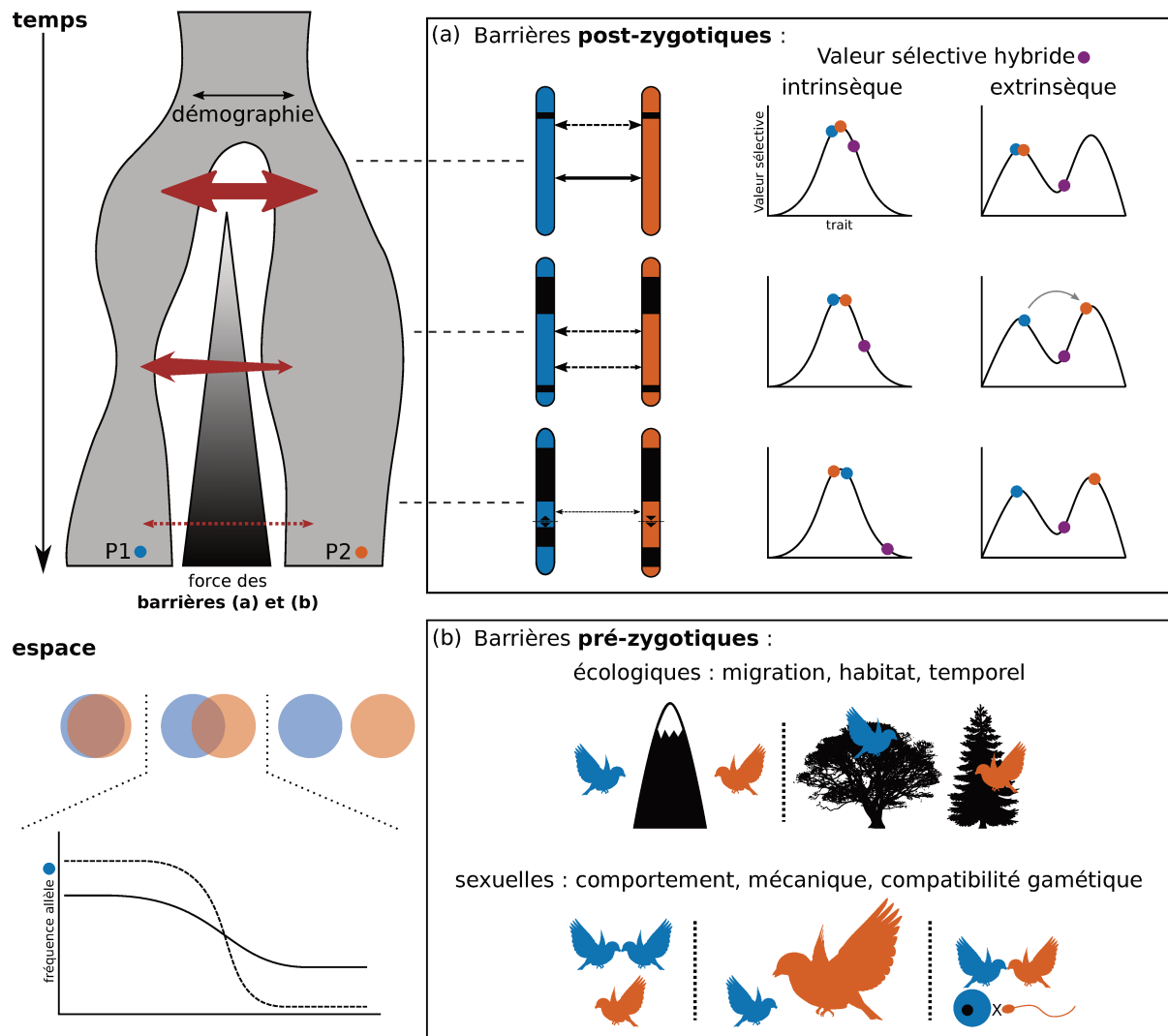


Figure 1 : Durant le processus de spéciation, les barrières aux flux de gènes peuvent être multiples et interagir entre elles. (a) Les barrières post-zygotiques agissent en diminuant la valeur sélective des hybrides, soit de manière intrinsèque, soit en interaction avec l'environnement (extrinsèque). Ces barrières (segments chromosomiques noirs) créent une hétérogénéité des flux de gènes le long du génome et peuvent s'accumuler au cours du temps. Elles peuvent être soit des modificateurs directs des traits associés à la valeur sélective ou des variations structurales limitant la recombinaison ou le déroulement correcte de la méiose. (b) Les barrières pré-zygotiques vont limiter la formation d'hybrides avant la fécondation, et ainsi contribuer à la limitation des flux de gènes. De nombreux mécanismes intrinsèques (e.g. comportement, gamètes) et extrinsèques (vicariance, sélection d'habitat) existent, seulement quelques exemples sont représentés ici. Certains de ces mécanismes peuvent aussi illustrer des barrières post-zygotiques extrinsèques. Globalement, les barrières sont susceptibles de s'accumuler mais aussi d'interagir au cours du temps (e.g. renforcement). Leurs effets sont dépendants du contexte démographique et spatial dans lequel se déroule le contact. Par exemple, les locus participant aux barrières post-zygotiques peuvent présenter des patrons spatiaux différents dans un contexte de zone de tension (en bas à gauche).

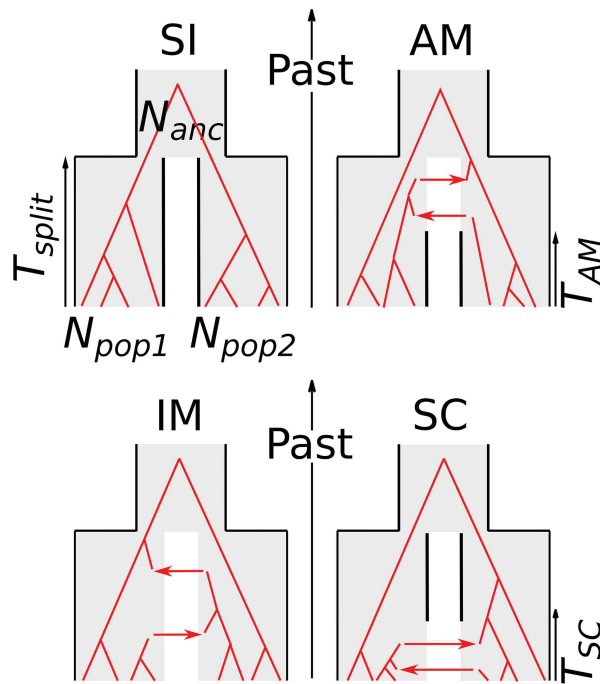


Figure 2 : Scénarios d'isolement classiques utilisés dans les modèles d'inférence démographique. SI : isolement strict ; AM : isolement avec migration ; IM : migration constante ; SC : contact secondaire. T_{split} : nombre de générations depuis l'évènement de spéciation ; T_{AM} : nombre de générations depuis l'arrêt de la migration ; T_{SC} : nombre de générations depuis le début du contact secondaire. L'aspect démographique est contrôlé par les tailles de populations (N) tandis que le flux de gènes est contrôlé par les taux de migration (flèches rouges) qui peuvent être asymétriques. Tirée de Roux et al. (2016).

la Bulgarie, formée lors d'un contact secondaire entre ces deux sous-espèces (Duvaux et al. 2011 ; Macholán et al. 2011). Le maintien de cette zone de tension fait intervenir des mécanismes de choix de partenaire (Smadja et Ganem 2005), ainsi que de la stérilité mâle et des réductions de fertilité femelle en F1 (croisement de première génération ; Storchová et al. 2004 ; Britton-Davidian et al. 2005 ; Larson et al. 2018).

Les scénarios

Les scénarios envisagés pour expliquer la spéciation ont recours à ces différents mécanismes, associés à une histoire de divergence dans des contextes géographiques différents. On a coutume de séparer les scénarios de spéciation en deux grands modes : avec ou sans flux de gènes.

Le scénario le plus simple que l'on puisse envisager stoppant le flux de gènes, est l'allopatrie (pouvant correspondre au modèle SI Figure 2). La spéciation allopatrique fait appel à une première barrière extrinsèque, souvent spatiale comme la vicariance entre deux populations, permettant l'accumulation de mutations dans les deux lignées. Les variations ainsi fixées s'accumulent avec le temps et peuvent contribuer à des incompatibilités (génétiques ou phénotypiques) limitant la reproduction ou la survie des hybrides (Figure 1). La polyploïdisation du génome est un autre mécanisme entraînant un arrêt du flux de gènes, plus souvent observé chez les plantes (Orr 1990 ; voir Ptacek et al. 1994 pour un exemple animal).

Lorsque l'isolement géographique ou pré-zygotique est incomplet, un certain niveau de flux de gènes est possible au cours du processus de spéciation. Le flux de gènes peut être présent depuis le début de la spéciation (Figure 2, IM), seulement au début du processus (AM) ou ré-apparaître après une période d'isolement (SC). Les types de scénarios peuvent

être complexifiés, par exemple en considérant des pulses périodiques de migration (e.g. Christe et al. 2017 ; Duranton et al. 2018).

i.1.2 Les bases génétiques de l'isolement reproductif

Les bases génétiques limitant le flux de gènes entre espèces naissantes sont au cœur des différents scénarios de spéciation. L'accumulation de mutations contribue à la divergence entre populations et constitue la base sur laquelle des mécanismes d'isolement peuvent apparaître (Encadré 1).

La divergence peut se construire sur la seule base du hasard lié au caractère fini des populations, qui correspond à la dérive génétique (Encadré 1). Théoriquement, l'accumulation d'un nombre important de différences liées à l'horloge moléculaire et son maintien, requiert une barrière au flux de gènes et un temps important. La spéciation par effet de fondation est l'un des modèles basé sur la dérive génétique (Gavrilets et Hastings 1996 ; Templeton 2008), bien que certains auteurs le considère comme rare et peu probable (Coyne et Orr 2004).

La sélection naturelle permet d'un autre côté d'envisager des scénarios de spéciation se classant en deux grandes catégories : la spéciation écologique et la spéciation par ordre des mutations (*mutation-order speciation* ; Schluter 2009). La spéciation écologique repose sur la mise en place d'isolement reproductif comme sous-produit de la divergence adaptative (Nosil 2012). Dans chacun des environnements rencontrés par plusieurs populations, différents allèles peuvent être sélectionnés et ainsi mener à des mécanismes d'isolement pré- ou post-zygotiques. Un exemple intéressant de spéciation écologique a récemment été décrit chez des scarabées de marais salants (*Pogonus chalceus*) par Van Belleghem et al. (2018). Dans ce système, deux écotypes co-existent. Le premier a un phénotype de petite taille avec des ailes courtes ayant un comportement de submergement¹ associé aux marais salants soumis aux marées. Le deuxième a un phénotype avec des ailes longues lui permettant une dispersion plus importante et est associé à des environnements inondés de manière saisonnière sans durée prévisible. Van Belleghem et al. (2018) montrent que les allèles conférant les adaptations locales sont anciens (~190 mille ans) et ont été récemment repropagés par la sélection après un évènement de contact secondaire et d'expansion suivant le dernier maximum glaciaire. Le tri spatial généré par ces adaptations contribue à l'isolement pré-zygotique entre les deux écotypes (Van Belleghem et al. 2016).

La spéciation par ordre des mutations est définie comme l'apparition et la fixation d'allèles différents entre populations s'adaptant à une même pression de sélection (Mani et Clarke 1990). Dans ce scénario, l'ordre d'apparition des mutations peut influencer les variants qui vont atteindre la fixation. Par exemple, chez *Mimulus*, il existe des cas de stérilité mâle cytoplasmique, lié à un réarrangement du génome mitochondrial, et d'un allèle nucléaire restaurant

1. Ces scarabées peuvent rester submergés par la marée plusieurs heures et résister à l'hypoxie en coinçant une bulle d'air sous leurs élytres.

Encadré 1 : Bases moléculaires de la différenciation génétique

La **théorie neutre** développée par Motoo Kimura (1968) est maintenant largement acceptée et utilisée comme hypothèse nulle de l'évolution moléculaire. Elle met en avant que « la grande majorité des substitutions ne sont pas causées par de la sélection positive Darwinienne mais par la fixation aléatoire de mutants sélectivement neutres ou quasi neutres » (Kimura 1983, p. 306).

Cette première force évolutive contribuant à la fluctuation des fréquences alléliques, due au tirage aléatoire des génotypes d'une génération à l'autre, est appelée **dérive génétique**. La variation des fréquences alléliques à chaque génération est inversement proportionnelle à la taille efficace de la population (notée N_e). Le polymorphisme neutre présent dans une population dépendant de la durée de vie d'une mutation dans la population, directement lié à N_e . Cependant le taux de substitution neutre n'est pas dépendant de N_e et n'est dépendant que du taux de mutation (μ), correspondant à l'horloge moléculaire.

Tomoko Ohta proposa en 1973 une extension de ce modèle. La théorie presque neutre de l'évolution moléculaire prend en compte qu'une partie des mutations faiblement délétères le sont assez pour se comporter comme neutres dans des populations assez grandes. En effet, la probabilité de fixation d'une mutation étant dépendante de la taille effective de la population (N_e), une mutation doit avoir un coefficient supérieur à $1/N_e$ pour être efficacement contre-sélectionnée. Il en résulte qu'une population de petite taille (ou passant par un goulot d'étranglement démographique) va fixer un nombre plus élevé de mutations faiblement délétères.

La valeur sélective d'une mutation va dépendre de son effet moléculaire et phénotypique. Bien que la distribution des effets de valeur sélective (DFE pour *Distribution of Fitness Effects*) soit dense autour de zéro (i.e. neutre ou quasi neutre), lorsqu'une mutation a des effets délétères ou positifs suffisants pour échapper à l'effet de la dérive génétique, la sélection naturelle va entrer en jeu. La **sélection naturelle** est directionnelle et son effet sur la variation des fréquences alléliques est d'autant plus important que le coefficient de sélection (souvent noté s) est important. Pour échapper à l'effet de la dérive, $|s|$ devra être bien supérieur à $1/N_e$.

Par effet de bord dû à la liaison entre sites génomiques proches, l'effet de la sélection naturelle ne va pas seulement influencer la mutation directement sous sélection, mais va aussi entraîner un auto-stop génétique des positions voisines. La **liaison génétique** va alors avoir plusieurs effets possibles sur le polymorphisme. Parmi ces effets, on peut citer les interférences de Hill-Robertson (Hill et Robertson 1966), la sélection d'arrière-plan (Charlesworth et al. 1993) ou encore le balayage sélectif (Maynard-Smith et Haigh 1974).

la fertilité, générant une incompatibilité entre une population de *M. guttatus* et *M. nasutus* (Case et Willis 2008). Une population de *M. guttatus* possède un haplotype cytoplasmique générant de la stérilité mâle, mais compensé par un allèle nucléaire, tandis que l'espèce *M. nasutus* ne possède aucun des deux.

À large échelle génomique, des modifications structurales des génomes peuvent tout d'abord expliquer certaines incompatibilités post-zygotiques. Ces modifications comprennent des fusions, fissions, translocations, inversions ou encore des duplications. Le mécanisme limitant le flux de gènes va souvent être lié au processus de la méiose et de la recombinaison (Piálék et al. 2005 ; Kirkpatrick et Barton 2006). En effet, certaines modifications structurales vont être trop importantes pour permettre des appariements corrects lors de la méiose, rendant les hybrides stériles, comme dans le cas du mulet. De manière moins directe, les modifications structurales peuvent aussi limiter la recombinaison dans certaines parties du génome, limitant le flux de gènes localement, comme cela peut être le cas lors d'une inversion chromosomique (Kirkpatrick et Barton 2006). À l'image des mutations, les inversions peuvent de plus avoir des effets négatifs à l'état hétérozygote, contribuant à l'isolement post-zygotique (modèle de sous-dominance). Finalement, les duplications du génome ou de certains chromosomes peuvent aussi entraîner des incompatibilités soit structurales, soit par résolution divergente (à l'échelle du gène). Dans une résolution divergente, l'une des deux copies d'un gène peut être inactivée. Si cette copie n'est pas la même dans deux populations, une partie des hybrides issus d'un croisement seront dépourvus de toute version active du gène (McGrath et al. 2014).

Dès le début du XX^e siècle, on se demandait déjà comment la combinaison de deux génomes pouvait engendrer une diminution de la valeur sélective, c'est-à-dire être incompatibles, sans l'avoir fait pendant la divergence. Il fallait pouvoir expliquer comment les facteurs génétiques responsables avaient pu se fixer si leur effet était délétère. En effet, le modèle le plus simple de sous-dominance, reposant sur un effet délétère ne s'exprimant que chez les hétérozygotes, est difficilement conciliable avec l'observation de vigueur hybride en F1, et avec la fixation de l'allèle dérivé sous-dominant dans une population (car des effets délétères existent de son apparition à sa fixation). La combinaison des travaux de Bateson, Dobzhansky et Muller ont permis d'envisager un modèle pouvant expliquer les incompatibilités lors d'un croisement et leur évolution (Encadré 2 ; Bateson 1909 ; Dobzhansky 1936 ; Muller 1942). Ce modèle permet d'expliquer la fixation d'allèles incompatibles sans passer par une vallée de valeur sélective. Cependant, il requiert souvent chez les organismes diploïdes l'hypothèse de récessivité des allèles dérivés afin d'expliquer (i) la vigueur ou l'asymétrie sexuelle (règle de Haldane) en F1 et (ii) la dépression hybride des F2s, BCs ou du sexe hétérogamétique F1 (Turelli et Orr 2000).

Ces prédictions étaient associées dès sa conception à la recherche des bases génétiques des incompatibilités (Dobzhansky 1936), avec comme idée prévalente le concept de « gène de spéciation » (Orr et al. 2004). Ces gènes seraient des éléments clés, contribuant de manière importante à l'isolement reproductif, et les trouver pourrait nous indiquer quelles fonctions ils occupent et si certaines fonctions sont plus susceptibles d'être impliquées dans le processus

Encadré 2 : Incompatibilité de Bateson-Dobzhansky-Muller (BDMI)

Dans le modèle BDMI, deux nouveaux allèles incompatibles (A et B) fixent indépendamment au sein de deux populations différentes. Ainsi, au cours de la fixation les deux allèles dérivés ne se retrouvent jamais associés dans un même génome, la combinaison délétère n'est alors jamais testée et ne peut pas être contre-sélectionnée. Ce processus permet d'expliquer la fixation d'allèles réduisant la valeur sélective des hybrides sans que les populations portant ceux-ci passent par une vallée de valeur sélective. L'exemple de la Figure 3 considère deux locus par simplicité, cependant il est possible de considérer des interactions épistatiques entre un nombre plus élevé de locus.

Lors d'un nouveau contact entre les deux populations, les hybrides produits portant les allèles incompatibles seront contre-sélectionnés et le flux de gènes sera limité autour des positions de ces deux locus. Aucun flux de gènes n'est possible si les allèles dérivés incompatibles sont dominants, létaux (ou stérilisants) et ont totalement fixé au sein des deux populations (aucun F1 n'aura de descendance). Au contraire si la réduction de valeur sélective n'est que partielle, des génotypes recombinants pourront être produits et le devenir des allèles dérivés dépendra de facteurs démographiques, spatiaux et biologiques (Figure 1).

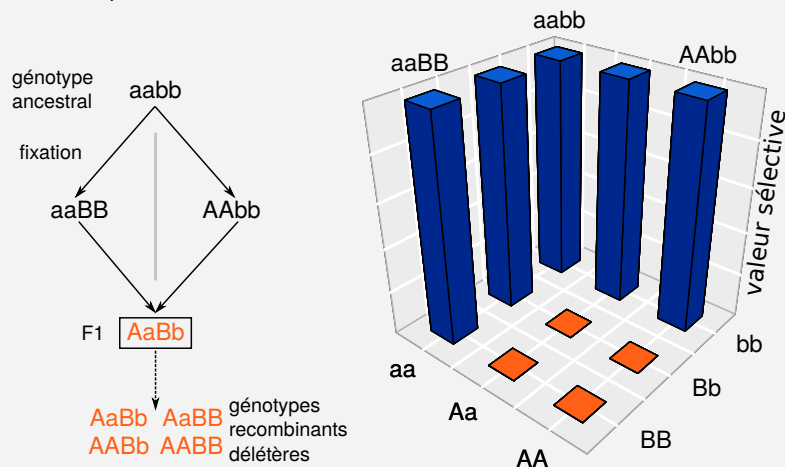


Figure 3 : Le modèle d'incompatibilité bi-locus de Bateson-Dobzhanski-Muller. À gauche, deux populations divergent sans contact et fixent un nouvel allèle à deux locus différents. Ces deux allèles sont incompatibles. La remise en contact de ces populations produit d'abord un génotype hétérozygote (F1), puis des génotypes recombinants si quelques F1s arrivent à se reproduire. Les combinaisons alléliques oranges faisant intervenir des allèles incompatibles vont être contre sélectionnés, ici un cas extrême de paysage de valeur sélective est présenté à droite.

de spéciation. L'aspect expérimental de cette recherche a été beaucoup développé chez des modèles où l'on peut facilement créer des lignées consanguines et faire des croisements, comme les drosophiles ou les plantes (Fishman et Willis 2001 ; Orr et Irving 2001 ; Orr et al. 2004). La cartographie des incompatibilités passe par la création de « lignées d'introgession », i.e. des individus ne portant qu'une région limitée d'un chromosome non-spécifique, chez qui nous pouvons mesurer la fitness (A. W. Davis et Wu 1996 ; Noor et al. 2001 ; Moyle et

Nakazato 2008 ; Corbett-Detig et al. 2013 ; Guerrero et al. 2017). Ce type de manipulation, tout comme la modélisation, devient rapidement irréalisable lorsque l'on commence à considérer des interactions épistatiques à plus de quelques locus.

Malgré cette vision centrée sur les « gènes à fort effet », les données génétiques accumulées au cours des quarante dernières années ont montré que quelques locus pouvaient difficilement expliquer la spéciation à eux seuls. Par exemple, Szymura et Barton (1986) ont estimé, à partir de l'analyse de la zone de tension entre *Bombina bombina* et *B. variegata*, que le coefficient de sélection global contre les hybrides devrait être étalé sur plusieurs centaines de locus à faible effet. De la même façon, Barton et Hewitt (1981) montrent que l'isolement post-zygotique chez des sauterelles (*Podisma pedestris*) repose sur environ 150 segments chromosomiques, chacun de faible effet sur l'inviabilité des hybrides.

L'accès à une résolution toujours plus importante des génomes et les capacités accrues de nombre d'échantillons, depuis l'avènement des méthodes de séquençage nouvelle génération, a révolutionné notre capacité de cartographie des locus contribuant à l'isolement et de détection d'effets plus faibles (e.g. Janoušek et al. 2012 ; L. M. Turner et Harr 2014). De façon simplifiée, la recherche de ces régions est l'équivalent d'une étude d'association génomique (GWAS pour *Genome Wide Association Study*) en utilisant la valeur sélective comme trait dans des croisements. Ces nouvelles capacités, associées à la réalisation que peu de gènes de spéciation avaient été trouvés en une vingtaine d'années de recherches (Morán et Fontdevila 2014), ont finalement ravivé une vision plus polygénique de la stérilité hybride chez des organismes modèles comme les drosophiles (Morán et Fontdevila 2014) ou les souris domestiques (Larson et al. 2018).

i.1.3 Génomique de l'isolement reproductif et des flux de gènes

Durant la révolution technologique du séquençage haut débit, la recherche des gènes de spéciation s'est transposée vers l'échelle génomique, pour produire une vision plus lisse intégrant les aspects polygéniques, la démographie et les effets de la recombinaison. La génomique des populations sur de plus en plus d'espèces (modèles ou non) a permis de réaliser que le flux de gènes pouvait être fortement variable le long du génome. Cette transformation s'est traduite par la caractérisation d'« îlots génomiques de différenciation », définis comme des régions génomiques dans lesquelles le flux génique est fortement réduit (T. L. Turner et al. 2005 ; Harr 2006 ; Ellegren et al. 2012). L'hypothèse qui sous-tend cette recherche est que ces îlots doivent contribuer d'une façon ou d'une autre à l'isolement reproductif.

Plusieurs hypothèses ont été développées pour expliquer les patrons hétérogènes observés le long des génomes (Harrison et Larson 2016 ; Ravinet et al. 2017 ; J. B. W. Wolf et Ellegren 2017). Tout d'abord les pics de différenciation peuvent se former via des mécanismes de sélection liés, comme des balayages sélectifs ou de la sélection d'arrière plan, qui associés à des variations du taux de recombinaison produisent des patrons hétérogènes de diversité génétique

(Burri 2017; Encadré 1). L'hétérogénéité de la divergence peut aussi être liée à l'érosion différentielle, due au flux de gènes lors d'un contact secondaire entre régions contribuant ou non à l'isolement reproductif (Bierne et al. 2013). Enfin un autre modèle propose que la sélection hétérogène post-spéciation pourrait être à l'origine des îlots de différenciation (Cruickshank et Hahn 2014). Dans les faits, l'ensemble de ces processus agissent de concert et il est nécessaire de prendre en considération l'histoire démographique ainsi que différentes mesures de divergence (relatives ou absolues) afin de pouvoir interpréter les données (e.g. Duranton et al. 2018). La prise en compte de ces patrons est particulièrement importante dans l'inférence des histoires démographiques, car ils peuvent être des facteurs confondants dans la détection des régions sous sélection (voir Annexe).

L'avancée des méthodes génomiques a un rôle important à jouer dans le domaine de la génomique de la spéciation (Campbell et al. 2018). En effet l'accès à des assemblages en contigs, scaffolds ou chromosomes de plusieurs mégabases permet la détection plus précise de variations structurales, importantes dans le processus de spéciation (cf. partie i.1.2). Par exemple, Todesco et al. (2019) montrent que les écotypes du tournesol (*Helianthus* spp.) sont associés à de nombreux haplotypes non-recombinants, permettant la cohésion d'allèles co-adaptés. D'un autre côté, l'information haplotypique permet aussi d'étudier les flux de gènes et l'histoire évolutive de façon plus précise. Par exemple, Duranton et al. (2018) profitent de la puissance des génomes phasés pour identifier les blocs génomiques (*tracts*) d'introgression du bar atlantique dans le loup méditerranéen (*Dicentrarchus labrax*).

i.1.4 Contacts secondaires : expérimentations pour l'étude de la spéciation

En dehors des croisements de laboratoire permettant de comprendre les mécanismes de l'isolement reproductif, la première source d'expérimentations naturelles à large échelle sont les zones d'hybridation. Les zones d'hybridation sont définies comme des régions étroites dans lesquelles des populations génétiquement distinctes se rencontrent, se reproduisent et engendrent des hybrides (Barton et Hewitt 1985). L'histoire de formation des zones d'hybridation actuelles est étroitement liée aux variations de distribution géographique des espèces durant les oscillations climatiques du Quaternaire (Hewitt 2000, 2011). Les scénarios les plus communs comprennent des périodes d'isolement dans des refuges lors des périodes de glaciation, suivies par des expansions géographiques menant à un contact secondaire entre lignées génétiques maintenant différenciées (Hewitt 2011).

Les zones d'hybridation sont le plus souvent des zones de tension maintenues par une balance entre migration et contre-sélection des hybrides (Barton et Hewitt 1985, 1989). Elles apparaissent donc comme une mise en situation idéale pour l'étude de l'isolement reproductif en conditions naturelles et permettent l'étude de la mise en place et le maintien des barrières aux flux de gènes. La force d'une barrière au flux génique va dépendre d'une balance entre la

sélection s'appliquant sur un ensemble de régions génomiques et la recombinaison tendant à casser les associations (Barton 1983 ; Kruuk et al. 1999 ; Abbott et al. 2013). La formation de ces associations entre régions contribuant à l'isolement – mais aussi avec des régions pouvant être liées à de la sélection exogène – est appelée couplage (*coupling* ; Barton 1983 ; Bierne et al. 2011 ; Abbott et al. 2013).

Les zones d'hybridation sont le résultat d'histoires phylogéographiques parfois complexes et largement variables, importantes dans la compréhension des patrons actuels (Hewitt 2011). Les progrès des méthodes d'inférence démographiques à partir de données génomiques permettent d'avoir aujourd'hui accès à des scénarios bien plus précis de l'histoire des contacts secondaires (Gutenkunst et al. 2009 ; Excoffier et al. 2013 ; Rougeux et al. 2017 ; Tóth et al. 2017).

Malgré une connaissance accrue de l'histoire évolutive sous-jacente aux contacts secondaires, la continuité du processus évolutif efface les traces laissées par les mécanismes à l'œuvre lors de l'initiation du contact, et les observations actuelles peuvent être confondues avec d'autres processus (Szymura et Barton 1986 ; Bierne et al. 2013 ; Bertl et al. 2018). De plus, les zones d'hybridation actuelles sont en grande majorité stables et les événements d'introgression qui peuvent nous intéresser ont déjà eu lieu. Il apparaît alors très intéressant de pouvoir observer des contacts secondaires récents, souvent créés par des perturbations anthropiques (e.g. modifications d'habitats, changements de distribution liés au changement climatique ou aux introductions). Ces contacts secondaires récents peuvent potentiellement nous donner accès aux dynamiques de la phase initiale. Les introductions biologiques impliquant de l'hybridation peuvent alors devenir des études de cas pour la biologie évolutive.

i.2 Introductions biologiques et hybridation

Une introduction est avant tout une modification de l'aire de distribution naturelle d'une espèce due aux activités humaines (que ce soit volontaire ou accidentel). Cette modification de distribution est alors une opportunité pour les biologistes d'étudier les aspects écologiques et évolutifs liés aux changements d'habitats, aux interactions biotiques et à l'adaptation (Thompson 2014, p. 7). Au cours de cette thèse, je me suis principalement intéressé aux conséquences évolutives d'une introduction, et notamment sur l'utilisation d'une introduction comme expérience d'un contact secondaire récent.

i.2.1 Le processus d'introduction biologique

Une espèce introduite², pour s'établir, doit passer un ensemble de barrières à la colonisation pérenne d'un nouveau environnement (Figure 4). Ces barrières peuvent être dispersives (e.g. être capable d'utiliser un vecteur), démographiques (e.g. avoir une pression de propagule suffisante), ou encore écologiques (e.g. survivre dans un nouvel environnement). Blackburn et

2. Synonymes : non native, non indigène, alien

al. (2011) définissent une espèce invasive comme une espèce ayant au moins une population autonome dans la nature et pouvant émettre des migrants, menant à sa propagation (Figure 4 « spread » et D1). Cette définition ne tient pas compte de la notion d'impact environnemental ou économique dans la typologie d'une introduction biologique.

En parallèle de cette typologie, Blackburn et al. (2014) ont proposé une classification des espèces introduites suivant leur impact environnemental (Figure 5, qui peut aussi être étendu à l'impact économique ou sociétal). L'introduction d'une espèce non indigène peut engendrer dans certains cas des conséquences écologiques et évolutives importantes (Mooney et Cleland 2001 ; Ehrenfeld 2010). Les conséquences biologiques majeures des invasions qui sont régulièrement rapportées sont l'extinction d'espèces, la perte de biodiversité ou encore la « pollution génétique » (Rhymer et Simberloff 1996 ; Carlton et al. 1999 ; Bellard et al. 2016). Bien que ces effets soient observés lors d'évènements d'introduction biologique, il existe un biais dans ce qui est rapporté et dans les termes utilisés, qui entraîne souvent un amalgame entre espèce non natives et effets négatifs (M. A. Davis et al. 2011 ; Thompson 2014).

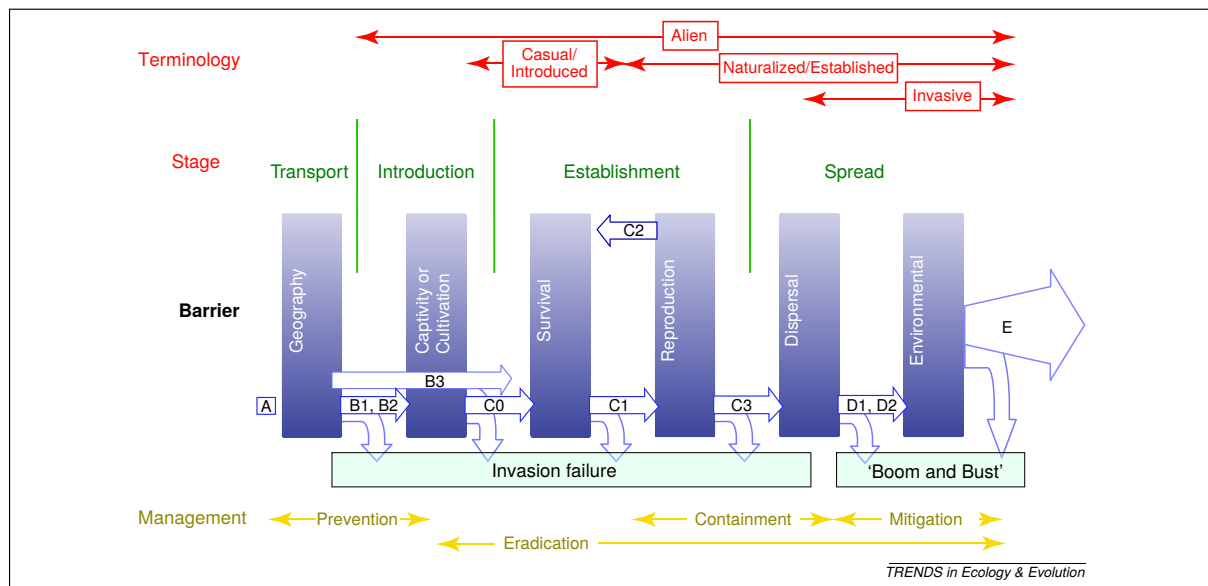
Le succès parfois fulgurant d'une espèce introduite dans un environnement dans lequel elle n'a pas d'histoire évolutive est une question majeure des recherches en invasion biologique et est appelé le « paradoxe de l'invasion » (Sax et Brown 2000). Très succinctement, un ensemble d'hypothèses ont été proposées permettant d'expliquer ce phénomène incluant des modalités de reproduction adaptées (e.g. clonalité ou autofécondation), des forts taux de fécondité, l'absence des prédateurs et pathogènes natifs, ou encore la disponibilité de niches écologiques. Comme indiqué dans la partie i.1.4, les contacts récents peuvent devenir une source importante d'informations sur les processus à l'œuvre durant la phase initiale d'un contact secondaire.

i.2.2 Les introductions en milieu marin

Dans son ouvrage fondateur sur l'étude des invasions, Elton (1958, p. 94) faisait déjà remarquer les remaniements d'aire de distribution en milieu marin : « In contrast to land and fresh waters the sea seems still almost inviolate. Yet big changes in the distribution of species have already begun as a result of human actions during the last hundred years. » Depuis 200 ans, le taux d'espèces introduites en milieu marin a augmenté exponentiellement (Ruiz et al. 2000). Un groupe d'experts liste actuellement plus de 2000 espèces marines introduites (Ahyong et al. 2018)

Les deux vecteurs majeurs des introductions marines sont le transport maritime et l'aquaculture (Figure 6 ; Molnar et al. 2008). Il existe d'autres vecteurs d'introduction moins importants mais pouvant localement représenter une source majeure, comme dans le cas du canal de Suez, ayant une influence importante sur la quantité d'espèces introduites en Méditerranée orientale (Nunes et al. 2014).

Le transport maritime permet le transport d'espèces marines non indigènes via plusieurs mécanismes. Le premier est l'encrassement biologique (*bio-fouling*) impactant les parties ex-



Category	Definition
A	Not transported beyond limits of native range
B1	Individuals transported beyond limits of native range, and in captivity or quarantine (i.e. individuals provided with conditions suitable for them, but explicit measures of containment are in place)
B2	Individuals transported beyond limits of native range, and in cultivation (i.e. individuals provided with conditions suitable for them, but explicit measures dispersal are limited at best)
B3	Individuals transported beyond limits of native range, and directly released into novel environment
C0	Individuals released into the wild (i.e. outside of captivity or cultivation) in location where introduced, but incapable of surviving for a significant period
C1	Individuals surviving in the wild (i.e. outside of captivity or cultivation) in location where introduced, no reproduction
C2	Individuals surviving in the wild in location where introduced, reproduction occurring, but population not self-sustaining
C3	Individuals surviving in the wild in location where introduced, reproduction occurring, and population self-sustaining
D1	Self-sustaining population in the wild, with individuals surviving a significant distance from the original point of introduction
D2	Self-sustaining population in the wild, with individuals surviving and reproducing a significant distance from the original point of introduction
E	Fully invasive species, with individuals dispersing, surviving and reproducing at multiple sites across a greater or lesser spectrum of habitats and extent of occurrence

Figure 4 : Typologie du processus d'invasion biologique, d'après Blackburn et al. (2011). Les auteurs proposent de diviser le processus d'invasion biologique en différentes étapes, chacune contrainte par des barrières à l'établissement d'une espèce introduite. L'échec de l'accès à l'étape suivante se traduit par l'échec de l'invasion. Une terminologie est associée à ces stades, montrant l'évolution du caractère non natif (ou alien) jusqu'à la classification en espèce invasive.

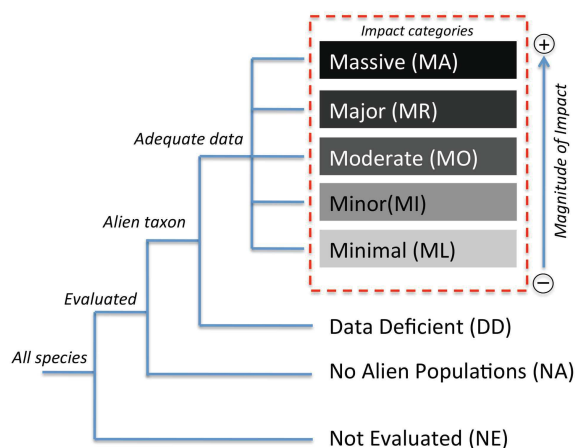


Figure 5 : Catégories de la classification de l'impact des espèces introduites. Tirée de Blackburn et al. (2014)

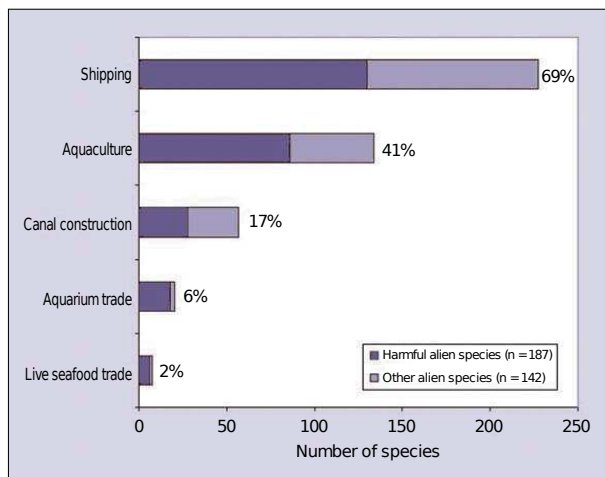


Figure 6 : Proportion d'espèces introduites en milieu marin par les vecteurs les plus importants, parmi 329 espèces introduites analysées dans l'étude. La proportion d'espèces ayant un fort impact écologique est indiquée en foncé pour chaque catégorie. Tirée de Molnar et al. (2008).

ternes des navires, mais aussi les coffres de prise d'eau offrant un environnement plus protégé (Coutts et Dodgshun 2007 ; J. Drake 2007 ; Sylvester et al. 2011). Le second mécanisme est la dispersion via les eaux de ballast, qui a été reconnu comme un vecteur majeur et est maintenant le mécanisme le plus contrôlé (Bailey 2015). Les eaux de ballast peuvent à la fois permettre la dispersion d'espèces de faible taille, mais aussi les phases larvaires ou microscopiques d'espèces macroscopiques (Levings et al. 2004 ; Flagella et al. 2007). Ces vecteurs sont d'importants mécanismes de dispersion longue distance, cependant à une échelle régionale la navigation de plaisance et les liaisons maritimes courtes (e.g. les ferrys) sont de plus en plus considérées comme des vecteurs majeurs de dispersion secondaire (voir primaire dans certains cas) (Wasson et al. 2001 ; Mineur et al. 2008 ; Clarke Murray et al. 2011).

L'introduction d'espèces non natives via l'aquaculture regroupe des transports volontaires mais aussi non intentionnels et sont qualifiées d'« associées » ou « accompagnatrices ». Un exemple ayant contribué à ces deux types d'introduction est celui de l'huître creuse (*Crassostrea gigas*) introduite initialement depuis le Japon dans plus de 70 pays pour sa culture (Ruesink et al. 2005). Ces introductions ont souvent été opérées pour compenser des stocks naturels des espèces natives en décroissance suite à la surexploitation ou à des épisodes de mortalité (e.g. *Ostrea edulis* en Europe). Ce type de bivalve ayant une reproduction externe et une phase larvaire planctonique, les huîtres creuses ont pu facilement échapper aux infrastructures de culture et s'établir dans le milieu naturel. Ces introductions pour l'aquaculture ont contribué au déplacement d'épibiontes accompagnateurs comme les macro-algues. Mineur et al. (2007) montrent par exemple expérimentalement que 57 espèces d'algues peuvent être retrouvées après transfert d'huîtres, parmi lesquelles une grande partie ne sont pas natives en Europe.

L'une des difficultés de l'étude des introductions marines réside dans l'existence d'espèces cryptiques et cryptogéniques. Une espèce cryptique représente l'absence de classification spécifique d'une lignée malgré la présence d'une divergence génétique importante. Les espèces cryptiques sont considérées communes en milieu marin (Appeltans et al. 2012 ; Viard et al. 2016 ; Chenuil et al. 2019), dû à des limitations de classification taxonomique mais aussi à *a priori* que les larges distributions géographiques sont considérées normales pour de nombreux

organismes marins (Knowlton 1993). Il n'apparaît donc pas rare d'observer des invasions cryptiques, qui peuvent être interspécifiques ou intraspécifiques (i.e. lors de l'introduction d'une lignée génétiquement différenciée sans véritable isolement reproductif; Morais et Reichard 2018). Les invasions cryptiques sont particulièrement difficiles à identifier, même avec l'utilisation de la génétique, puisqu'elles requièrent une connaissance de l'ensemble des populations du taxon considéré et une résolution suffisante pour identifier des unités taxonomiques différenciées. En lien avec ces difficultés, la recherche en invasions marines s'est très largement concentrée sur des cas d'introduction longue distance, souvent trans-océaniques et facilement identifiables.

Dans certains cas, le caractère natif ou introduit d'une espèce ne peut pas être déterminé avec certitude, on la qualifie alors de cryptogénique (Carlton 1996, 2009). Les espèces cryptogéniques sont souvent des espèces cosmopolites, dont la distribution peut résulter d'introductions anciennes dans des régions où la pression anthropique a été constante, conduisant à l'absence de preuve de leur caractère introduit.

La combinaison de ces catégories cryptique et cryptogénique, ainsi que l'existence de complexes d'espèces, rend la quantification des introductions en milieu marin difficile. La génétique est donc un outil très important permettant de meilleures déterminations et études des événements d'invasion, que ce soit en milieu marin comme en milieu terrestre.

i.2.3 La génétique des invasions : la place de l'hybridation

Le domaine de la génétique des invasions a maintenant une cinquantaine d'années (Bock et al. 2015). Durant cette période, la génétique a tout d'abord servi d'outil pour identifier et caractériser l'origine des espèces non natives. Elle a également permis d'étudier les processus évolutifs à l'œuvre lors d'une invasion, tels que l'adaptation à un nouvel environnement ou la réponse à un événement de fondation.

L'un des champs d'étude du domaine de la génétique des invasions est la variation et la diversité génétique, et le rôle relatif du processus d'adaptation lors d'une colonisation et de la pré-adaptation (Sakai et al. 2001). Cette question part du principe que la diversité génétique est proportionnelle à la capacité de réponse à la sélection naturelle (Fisher 1930; Sakai et al. 2001). Partant de cela, l'observation de l'existence de goulots d'étranglement démographiques et d'une perte de diversité lors de certaines invasions a conduit à un paradoxe génétique des invasions (Kolbe et al. 2004). Cependant, les événements de fondation, notamment en milieu marin, sont loin d'être la règle lors d'invasions, et la relation de causalité entre diversité génétique et le succès de l'invasion reste incertaine (Viard et al. 2016).

L'observation d'une phase de retard (*lag phase*) dans de nombreuses invasions, correspondant à un décalage entre le premier établissement et la phase d'expansion, a souvent été utilisée comme preuve d'un manque d'adaptation au nouvel environnement. Les hypothèses génétiques pouvant expliquer ce temps d'attente postulent qu'un temps suffisamment long est

nécessaire, pour que de nouvelles mutations apparaissent, ou que celles-ci ou la variation déjà présente soient sélectionnées, ou bien pour que la diversité génétique apportée par d'autres populations (ou espèces) introduites soit suffisante pour permettre l'adaptation (Bock et al. 2015). De la même façon que pour la diversité génétique, le rôle de l'adaptation dans le succès d'invasion n'est pas encore très bien compris (Facon et al. 2006 ; Viard et al. 2016).

Ainsi, il est apparu tôt dans la littérature que l'hybridation pouvait avoir un rôle important dans le processus d'invasion, étant considérée comme une augmentation instantanée de la diversité génétique et des capacités d'adaptation (E. Anderson et Stebbins 1954 ; Ellstrand et Schierenbeck 2000).

Les termes hybridation et admixture, plutôt génériques en biologique, peuvent faire référence, dans le domaine des invasions, aussi bien au croisement entre deux populations d'une même espèce qu'au croisement entre plusieurs espèces. L'admixture entre plusieurs populations d'une même espèce (parfois qualifiée d'hybridation intraspécifique) est possible lorsque les invasions impliquent des sources multiples. C'est le cas des lézards (*Anolis sagrei*) introduits en Floride qui présentent une diversité génétique supérieure dans cette aire d'introduction comparée à l'aire native, due à des introductions multiples (Kolbe et al. 2004). L'hybridation interspécifique peut aussi être présente lors d'une introduction, soit entre plusieurs espèces introduites, soit avec des espèces natives proches. Abbott (1992) conclue que 7% des espèces introduites ont mené à de l'hybridation interspécifique dans les îles anglo-saxonnes.

Il est communément suggéré que l'admixture augmente les chances de succès de colonisation des espèces introduites (voir Rius et Darling 2014 pour une synthèse, voir Facon et al. 2008 pour un exemple). Cependant, les bénéfices possibles des admixtures et leurs liens causaux avec le caractère invasif sont débattus (Dlugosch et Parker 2008 ; Chapple et al. 2013 ; Rius et Darling 2014 ; Barker et al. 2018). Dans un premier temps, il est souvent difficile de savoir si l'admixture précède ou non la phase d'invasion. De plus, le résultat immédiat de l'hybridation, surtout interspécifique, est difficilement prévisible et dépend non seulement de facteurs biotiques et abiotiques, mais aussi du degré d'isolement reproductif entre lignées et du niveau d'hybridation (cf. partie i.1). Par exemple, la valeur sélective des hybrides peut être importante dans la première génération (F1) et décroître dans des stades plus avancés d'hybridation (dépression d'hybridation ; voir chapitre 4). Déterminer les scénarios où l'hybridation aura un effet positif et leur probabilité s'avère en réalité difficile. Les effets de l'admixture sur la valeur sélective dans la perspective des invasions ont surtout été étudiés chez les plantes (Ellstrand et Schierenbeck 2000 ; Rieseberg et al. 2007 ; Hovick et Whitney 2014), et beaucoup moins chez des espèces animales.

L'hybridation, souvent considérée pour sa capacité à augmenter la diversité génétique dans le processus d'invasion, peut aussi influencer les facteurs démographiques. Par exemple, elle peut permettre de contourner des effets de densité dépendance (effets Allee) lors de l'introduction en permettant au génotype invasif de s'appuyer sur la démographie du génotype envahi (Mesgaran et al. 2016).

i.2.4 Introduction et hybridation : une expérience évolutive

Les biologistes évolutifs sont la plupart du temps contraints aux inférences indirectes à partir des variations génomiques actuelles, de l'ADN ancien ou d'évolution expérimentale (Bock et al. 2015). Les introductions peuvent être considérées d'un point de vue scientifique comme des fenêtres sur les processus écologiques et évolutifs, et constituer de véritables expériences grandeur nature menant à des avancées dans la compréhension de l'adaptation, de la plasticité phénotypique, de la démographie, des compétitions, du rôle de la diversité génétique ou de l'hybridation (Bock et al. 2015).

Les introductions menant à l'hybridation, soit entre une espèce non indigène et une espèce native, soit entre deux espèces non indigènes, représentent de nouveaux contacts secondaires entre lignées génétiquement différenciées. Ce type d'évènement peut être qualifié d'« hybridation anthropogénique ». Comme indiqué dans la partie i.1.4, les contacts récents peuvent devenir une source importante d'informations sur les processus à l'œuvre durant la phase initiale d'un contact secondaire (Bouchemousse et al. 2016). L'aspect contemporain peut également permettre d'évaluer les effets de la dynamique d'avance et de la démographie sur les processus évolutifs. La formation d'une zone d'hybridation lors d'une invasion donne la possibilité de résoudre certains paradigmes de la génétique des invasions. En effet, les méthodes classiques d'étude des zones de tension peuvent s'appliquer pour tester par exemple les hypothèses d'introgession adaptative, souvent considérées dans le domaine de la biologie des invasions (Colautti et al. 2017).

La rencontre de deux lignées différenciées peut aboutir à de nombreux résultats suivant les contextes démographiques, écologiques et spatiaux (Allendorf et al. 2001). L'espèce native peut potentiellement être absorbée par la vague d'avance de l'espèce introduite et mener à ce que certains auteurs qualifient d'« extinction par hybridation » (Rhymer et Simberloff 1996). Dans certains cas, l'hybridation peut aboutir à un évènement de spéciation, notamment dans des situations de polyploïdisation (dite spéciation allopolyploïde). Par exemple dans le genre *Spartina*, *S. anglica* est le résultat de l'hybridation entre une espèce introduite, *S. alterniflora*, et l'espèce native *S. maritima* (Ferris et al. 1997, voir Anttila et al. 2000 pour un autre exemple). Certains allèles locaux peuvent introgresser la population invasive, de manière neutre (Currat et al. 2008), ou liés à de l'adaptation locale contribuant à l'invasion (Schierenbeck et Ellstrand 2009). La situation opposée d'introgession d'allèles invasifs dans la lignée native est aussi possible (Fitzpatrick et al. 2009), bien qu'elle soit théoriquement plus rare (Currat et al. 2008).

Figure 7 : Photos des trois espèces de l'hémisphère nord du complexe *M. edulis*. De gauche à droite : *M. edulis*, *M. galloprovincialis* et *M. trossulus*. Source : Joop Trausel et Frans Sliker, Natural History Museum Rotterdam



i.3 Le modèle d'étude *Mytilus* : au hasard des rencontres

i.3.1 Le complexe d'espèces de la moule bleue : composition, distribution et histoire évolutive

Parmi les complexes d'espèces intéressants pour leurs particularités d'isolement reproductif partiel, l'un des plus étudiés dans le milieu marin est celui de la moule bleue (*M. edulis*, Figure 7). Il est composé de trois espèces dans l'hémisphère nord, *M. edulis sensu stricto* (Linné, 1758), *M. galloprovincialis* (Lamarck, 1819) et *M. trossulus* (Gould, 1850); ainsi que de trois espèces dans l'hémisphère sud, *M. planulatus* (Lamarck, 1819), *M. platensis* (d'Orbigny, 1842) et *M. chilensis* (Hupé, 1854). Ces espèces à distributions antitropicales sont interfécondes et produisent des hybrides viables. Elles ont une histoire évolutive de divergence longue ponctuée par des événements de flux de gènes. Au cours de ma thèse, je me suis principalement intéressé aux espèces de l'hémisphère nord. Le genre *Mytilus* regroupe deux espèces supplémentaires, *M. californianus* et *M. unguiculatus* (anciennement *M. coruscus*), plus éloignées phylogénétiquement et communément non comprises dans le complexe d'espèces *M. edulis*. Cependant *M. californianus* pourra être utilisée comme groupe externe lors d'analyses génomiques.

La délimitation actuelle des différentes espèces du complexe *M. edulis* ne s'est faite que relativement récemment grâce aux études génétiques (Varvio et al. 1988; Koehn 1991; McDonald et al. 1991). En effet, malgré de nombreux efforts de classification sur des bases morphologiques avec des approches multidimensionnelles, la forte variabilité intraspécifique, l'intergradation importante entre espèces ainsi que l'absence de critères morphologiques discriminants n'ont jamais permis d'obtenir des résultats aussi puissants qu'avec la génétique (cf. références dans Kartavtsev et al. 2005). Ainsi, de nombreuses études se sont basées sur quelques marqueurs génétiques afin d'identifier les espèces et hybrides. L'avancée des méthodes a permis d'observer un nombre non négligeable d'introgessions aussi bien locales, dans certaines populations, que globales, sur l'ensemble de l'aire de répartition de l'espèce (Fraïsse et al. 2016a), mettant en perspective l'utilisation d'un nombre limité de marqueurs pour l'identification.

Il est habituellement postulé que *M. edulis* et *M. galloprovincialis* ont divergé de *M. trossulus* il y a environ 3.5 millions d'années (Ma), sur la base de l'ouverture du détroit de Bering qui aurait permis la colonisation de l'océan Atlantique par l'ancêtre commun à ces trois espèces (Rawson et T. J. Hilbish 1995). Grâce à des modèles démographiques incluant les flux de gènes ayant eu lieu durant les oscillations glaciaires du Quaternaire, Roux et al. (2014) et Fraïsse et al. (2018b) ont estimé la divergence entre *M. edulis* et *M. galloprovincialis* à environ 2.5 Ma. Deux hypothèses ont été proposées pour expliquer la divergence initiale entre *M. edulis* et *M. galloprovincialis* (Daguin 2000). La première propose une formation de *M. galloprovincialis* en Méditerranée au Pléistocène due aux épisodes glaciaires (Barsotti et Meluzzi 1968). La seconde propose que *M. galloprovincialis* ait dérivé d'une ramification de *Mytilus* ayant persisté dans la mer Téthys (~20 Ma) plutôt que du groupe *M. trossulus*-*M. edulis* (Vermeij 1992). Quel que soit le scénario de divergence initial, s'en est suivi une histoire de flux génique ponctuel, de remises en contact, au gré des cycles glaciaires du Quaternaire qui ont redistribué les répartitions géographiques de ces espèces (Hewitt 2000).

Chaque espèce ayant une large aire de distribution sur plusieurs continents, il existe des sous-populations – que l'on peut aussi qualifier de lignées génétiques – au sein de chacune d'elles (Figure 8). La différenciation intraspécifique est le résultat à la fois de périodes d'isolement durant les cycles glaciaires, d'isolement par la distance, ainsi que d'introggression locale d'allèles des autres espèces (Fraïsse et al. 2016a). La Figure 9 est une vision schématique des différentes populations de *Mytilus* dans l'hémisphère nord ainsi que de leurs introggressions naturelles.

L'une des différenciations les plus importantes pour la compréhension de cette thèse est la divergence entre les lignées méditerranéennes (med.) et atlantiques (atl.) de *M. galloprovincialis* (Sanjuan et al. 1994; Quesada et al. 1995b,c). Cette divergence est le résultat d'évènements de vicariance entre l'Océan Atlantique et la mer Méditerranée durant les dernières oscillations glaciaires du Quaternaire (Quesada et al. 1995b). Ces deux lignées forment deux zones d'hybridation parallèles au nord et au sud de la Méditerranée, au niveau du front océanique Alméria-Oran (El Ayari et al. 2019; Figure 8). La zone d'hybridation du nord est un cline de 150 km de large, tandis que la zone d'hybridation du sud est une large zone d'hybridation mosaïque s'étendant sur 600 km de large entre Oran et la baie de Bejaia en Algérie. Une partie de la divergence observée entre ces deux lignées est due à l'introggression d'allèles *M. edulis* en atl. avec lesquelles elles sont en contact plus au nord (Bierne et al. 2003b; Fraïsse et al. 2016a). Plus étonnamment, quelques introggressions d'allèles *M. edulis* sont observées en Méditerranée suggérant un contact ancien entre *M. edulis* et la lignée méditerranéenne de *M. galloprovincialis* (Gosset et Bierne 2013; Fraïsse et al. 2016a).

M. edulis et *M. trossulus* possèdent des lignées américaines et européennes (Figures 8 et 9). La population de *M. trossulus* de la mer Baltique est très particulière puisqu'elle est fortement introggressée par *M. edulis*. Par exemple, la mitochondrie femelle de *M. edulis* a complètement remplacé celle de *M. trossulus* dans cette population. Malgré l'originalité de

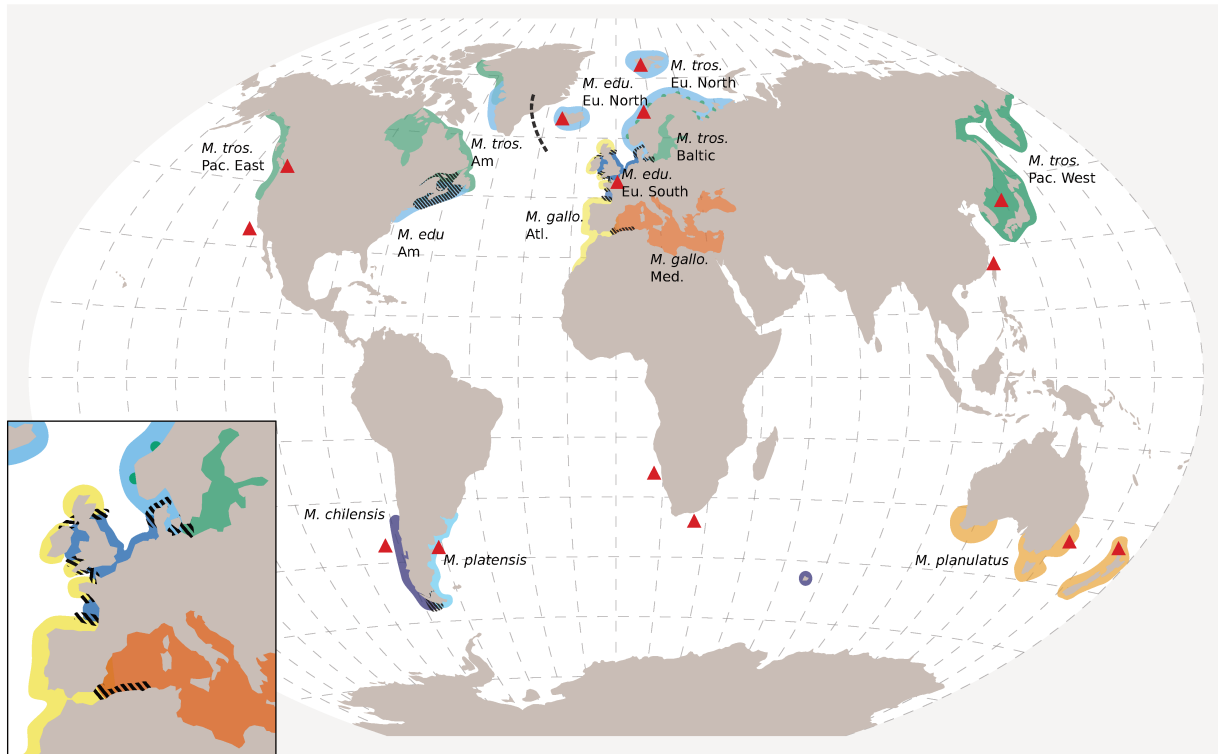


Figure 8 : Distribution des espèces et lignées génétiques du complexe *M. edulis* à l'échelle mondiale. Les distributions représentent un consensus de la littérature et des connaissances acquises tout au long de cette thèse, cependant certains points peuvent être sujets à discussions. Les aires hachurées représentent les zones d'hybridation connues. Les introductions de *M. galloprovincialis* sont représentées par les triangles rouges (Encadré 3 et Tableau 1). Les *M. edulis* américaines et nord-européennes partagent la même couleur mais sont séparées par la ligne en pointillés au niveau du Groenland.

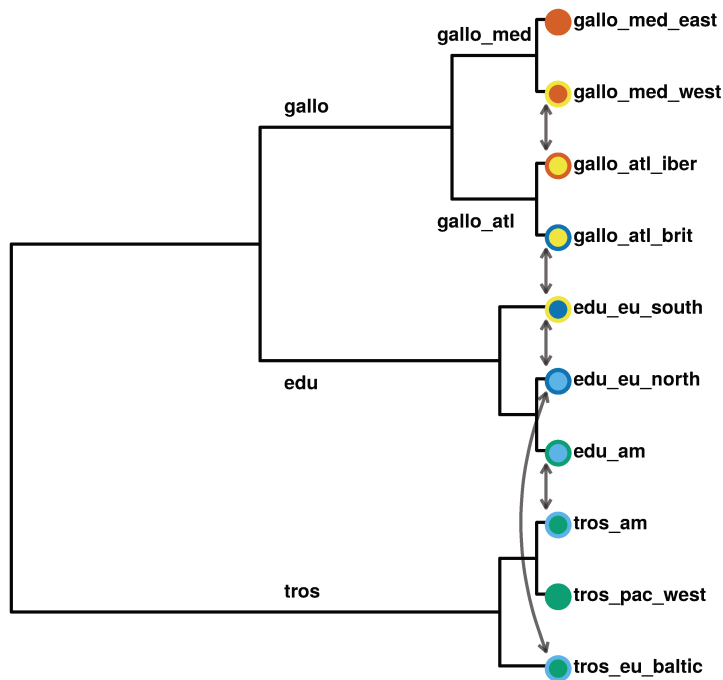


Figure 9 : Dendrogramme des relations au sein du complexe d'espèces *M. edulis* de l'hémisphère nord. La couleur intérieure des cercles représente l'espèce focale tandis que la couleur externe représente la présence d'introgression. Les flèches représentent les flux de gènes ayant eu lieu entre les populations ou espèces.

cette population, le reste des *M. trossulus* de la Scandinavie et de Russie semblent beaucoup moins impactées par des introgressions *M. edulis*. Au sein de l'Europe, nous définissons ici deux lignées de *M. edulis*, séparées par une zone d'hybridation le long des côtes du Danemark (non publié). La lignée qualifiée de *M. edulis* sud-européenne regroupe les populations de la côte Atlantique, de la Manche, d'une partie de la mer du Nord et des îles anglo-saxonnes. La lignée qualifiée de *M. edulis* nord-européenne regroupe les populations de la côte Norvégienne, du Kattegat (où elle est en contact avec les *M. trossulus* de la mer Baltique) et du nord de la Scandinavie (incluant la Russie). La signature génétique de ces *M. edulis* nord-européennes est intermédiaire entre les sud-européennes et les *M. edulis* américaines (cf. [chapitres 1 à 3](#)).

Les lignées de l'hémisphère sud sont beaucoup moins centrales à ce travail de thèse. Les hypothèses d'émergence de ces espèces reposent sur des migrations trans-équatoriales lors d'évènements de fluctuations climatiques du Pléistocène (Daguin [2000](#); T. J. Hilbish et al. [2000](#)). Très succinctement, *M. planulatus* (Océanie) est une espèce proche de *M. galloprovincialis*, ayant été isolée pendant au moins 100 000 ans en hémisphère sud, avant des contacts récents, suite à des introductions anthropogéniques (Popovic et al. [2019](#)). *M. chilensis* et *M. platensis* (Amérique du Sud) sont phylogénétiquement plus proches de *M. edulis* (Zbawicka et al. [2019](#)). Ces deux dernières espèces forment potentiellement une zone d'hybridation au niveau du détroit de Magellan ([Figure 8](#)).

i.3.2 Modèle pour l'étude de l'hybridation et de la spéciation

En 1966, Hubby, Lewontin et Harris démontrèrent que les variants moléculaires peuvent être caractérisés par des techniques d'électrophorèses séquentielles, ouvrant ainsi la porte à des avancées majeures en génétique des populations (Harris [1966](#); Hubby et Lewontin [1966](#);

Charlesworth et al. 2016). Seulement quatre ans plus tard, la première étude s'intéressant aux variations alléliques entre 24 populations de *M. edulis* sur la côte Est des États-Unis était présentée, et publiée deux ans plus tard (Milkman et Beaty 1970; Koehn et Mitton 1972). Ce développement rapide de la méthode des allozymes pour *M. edulis*, ainsi que la découverte d'une forte diversité génétique et de patrons de changements clinaux de fréquence allélique, ont largement contribué à son utilisation importante comme modèle d'étude de la spéciation, de la génétique des populations, de l'hybridation et de l'adaptation en milieu marin (Koehn et al. 1976; Skibinski et al. 1978a; Koehn et al. 1980; Varvio et al. 1988). En Europe, la première étude d'allozymes chez *M. edulis* par Ahmad et Beardmore (1976) s'est intéressée à la différenciation entre *M. edulis* et la « Padstow mussel », une population de moules de Cornouailles à la morphologie inhabituelle, qui s'est avérée plus tard être *M. galloprovincialis*. Depuis un peu moins de 60 ans, l'ensemble des techniques de génétique puis génomique disponibles ont été utilisées chez les moules, allant des allozymes aux génomes complets, en passant par les microsatellites, les RFLPs³, les SNPs⁴ le RADseq⁵, ou encore le RNAseq⁶ (voir l'ensemble des références dans Zbawicka et al. 2019, p. 2, ainsi que Romiguier et al. 2014; Fraïsse et al. 2016a, 2018b et Popovic et al. 2019 pour l'utilisation du RNAseq en génomique des populations de moules). L'ensemble de ces techniques a permis de décrire la diversité génétique très élevée des espèces de *Mytilus* (*M. trossulus* était la deuxième espèce la plus polymorphe des 75 espèces animales étudiées en RNAseq par Romiguier et al. 2014, et *M. edulis* et *M. galloprovincialis* les 14^e et 15^e), les zones d'hybridation multiples existantes et les flux de gènes ayant impacté ces différentes espèces (Fraïsse et al. 2016a, 2018b).

L'une des caractéristiques des zones de contact entre les différentes espèces du complexe est l'existence de « zones d'hybridation mosaïques ». Ce terme, proposé par Harrison (1986), désigne une alternance de types parentaux dans l'espace, séparés par des zones de transition ou zones d'hybridation. Le groupe *Mytilus* présente des zones d'hybridation mosaïques à large échelle très étudiées le long des côtes ouest-européennes, du Pays Basque jusqu'au nord des îles anglo-saxonnes (Figure 8). Ces alternances d'espèces ont permis l'étude de l'hybridation et de l'introgession en milieu marin (Skibinski et Beardmore 1979; Coustau et al. 1991; Quesada et al. 1995a; Bierne et al. 2002b; Gosset et Bierne 2013; Fraïsse et al. 2016a). Une autre zone d'hybridation particulièrement étudiée est celle séparant les *M. trossulus* de la mer Baltique des *M. edulis* de la mer du Nord (Figure 8; Väinölä et Hvilsom 1991; Strelkov et al. 2017).

Différentes barrières reproductives sont connues pour le complexe d'espèces *M. edulis*. Tout d'abord, un certain nombre de barrières pré-zygotiques ont pu être observées comme de la préférence d'habitat (Bierne et al. 2003a; Katolikova et al. 2016), de la reconnaissance des gamètes (Bierne et al. 2002a; Rawson et al. 2003; Springer et Crespi 2007; Klibansky et McCartney 2014) et de l'asynchronie reproductrice (Gardner et Skibinski 1990; Bierne et al.

3. Restriction Fragment Length Polymorphisms

4. Single Nucleotide Polymorphisms

5. Restriction site Associated DNA sequencing

6. séquençage ARN

2003a). Malgré ces barrières, on observe de manière fréquente des hybrides dans les zones de sympatrie. Ceci pourrait être dû à une variabilité intraspécifique de la période de reproduction fortement liée aux conditions environnementales (Seed 1976), et pouvant peut-être engendrer des chevauchements partiels ou complets de la reproduction, ou encore à la possibilité d'une érosion des barrières de fertilisation dans des populations introgressées (Slaughter et al. 2008). Les conséquences de barrières post-zygotiques, se traduisant par de la sélection contre les hybrides, sont observées expérimentalement au stade larvaire (Bierne et al. 2002a, 2006) mais sont plus difficiles à observer en milieu naturel. Un argument supplémentaire pouvant nous indiquer l'existence de barrières post-zygotiques est la stabilité des zones d'hybridation connues entre les différentes espèces de *Mytilus* spp., puisque la concordance de ces observations avec la théorie des zones de tension requiert une balance entre la migration et la contre-sélection des hybrides (Barton et Hewitt 1985, 1989). Les mécanismes génétiques créant ces barrières post-zygotiques pourraient comprendre des incompatibilités de type BDMI (Bierne et al. 2006), mais sont aussi conciliables avec des modèles plus généraux d'incompatibilité génomique (Simon et al. 2018).

Les moules du complexe *M. edulis*, bien que se situant à la limite haute de la zone grise de la spéciation en terme de divergence, montrent entre certaines espèces un niveau de migration contemporaine équivalent à ce qui peut être observé entre populations d'une même espèce (Figure 10, Roux et al. 2016). Associée à l'observation de nombreuses zones hybrides, il n'est donc pas étonnant de conclure que l'introggression – qu'elle soit ancienne, contemporaine ou locale – est omniprésente au sein du complexe *M. edulis* (Fraïsse et al. 2016a et chapitre 1). En prenant l'ensemble des espèces et populations du complexe d'espèces *M. edulis*, nous pouvons avoir une continuité extraordinaire représentative de tous les stades de la spéciation, en faisant un modèle d'étude de choix.

i.3.3 *Mytilus* à l'âge de la génomique

Le complexe *M. edulis* regroupe des espèces à forts enjeux environnementaux et économiques. Les moules sont des espèces clé des écosystèmes intertidaux par leur rôle de producteur secondaire, leur position à l'interface entre le milieu pélagique et benthique (filtration et bioaccumulation, Kautsky et Evans 1987), ainsi que leur rôle d'ingénieur des écosystèmes (Borthagaray et Carranza 2007). Les espèces du complexe *M. edulis*, majoritairement *M. edulis* et *M. galloprovincialis*, sont aussi cultivées à l'échelle mondiale (mytiliculture). Leur valeur économique est importante avec plus de 280 000 tonnes produites en 2016 (FAO 2018). Bien que la mytiliculture n'atteigne pas le degré « d'avancement technologique » d'autres types d'aquaculture, des programmes de sélection commencent à voir le jour (Nguyen et al. 2014). De plus, la valeur économique des moules ne s'arrête pas à leur attrait culinaire, puisqu'elles sont aussi très utilisées dans la surveillance toxicologique (Beyer et al. 2017), ou comme source de molécules anti-bactériennes (Tincu et Taylor 2004). Malgré leur importance comme modèle

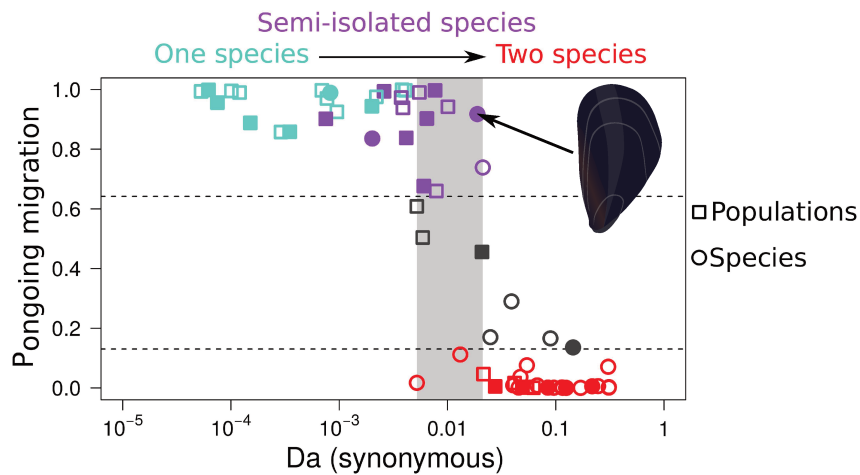


Figure 10 : Zone grise de la spéciation, tirée de Roux et al. (2016). La probabilité de migration contemporaine (P) en fonction de la divergence moléculaire nette synonyme (D_a) montre une zone de transition entre une et deux espèces. La paire d'espèces *M. edulis*/*M. galloprovincialis* est indiquée par la flèche. Son symbole indique que le modèle démographique le plus probable comprend du flux de gènes contemporain hétérogène sur le génome ainsi qu'un N_e hétérogène.

interdisciplinaire, les moules sont loin d'avoir des ressources génétiques à la hauteur d'autres organismes modèles.

Les méthodes de réduction des génomes pour l'étude de la génétique des populations ont pu être appliquées au complexe *M. edulis*. On retrouve notamment des études utilisant le RNAseq ou la capture (Fraisse et al. 2016a; Ribeiro et al. 2018; Popovic et al. 2019), et des études utilisant le RADseq (Araneda et al. 2016; Wilson et al. 2018; Zane 2019). Bien que ce type d'étude permette un gain significatif de puissance par rapport aux méthodes classiques de la génétique des populations, la disponibilité d'un génome augmente le type d'analyses possibles (variation des taux de recombinaison, blocs d'ascendance, annotation, variations structurales, etc.).

Le développement d'un génome de référence chez *Mytilus* spp. a encore un temps de retard par rapport à d'autres modèles biologiques, notamment dû à la complexité de leur génomes. Les bivalves en général ont des génomes complexes, plutôt grands, avec une quantité importante d'éléments répétés et hautement polymorphes (Saavedra et Bachère 2006; Zhang et al. 2012; Wang et al. 2017; Uliano-Silva et al. 2018). Les moules *Mytilus* spp. ne font pas exception avec un génome estimé à $\sim 1,6$ Gb, $\sim 43\%$ d'éléments répétés et un taux de GC à $\sim 32\%$ (Murgarella et al. 2016; Gerdol et al. 2019). Ces caractéristiques rendent l'assemblage difficile et le meilleur assemblage actuel de *M. galloprovincialis*, utilisant pourtant des méthodes de séquençage de long fragments, est encore largement fragmenté (10 577 scaffolds, N50 des contigs de 71,42 Kb; Gerdol et al. 2019).

Malgré ces difficultés, les informations tirées pour le moment de la génomique sont très intéressantes et confirment des observations de la génétique de ces espèces. Les moules *Mytilus* possèdent une forte diversité, avec des taux d'hétérozygotie élevés. S'ajoute à cela l'observation

de variations structurales importantes produisant des variations de présence/absence affectant ~24% des gènes (sur 60 338 au total ; Gerdol et al. 2019).

i.3.4 Écologie et traits d'histoire de vie liés à l'hybridation et aux introductions

L'écologie et les traits d'histoire de vie associés à la reproduction et liés à la répartition des espèces étudiées sont importants afin d'interpréter les événements d'hybridation et de flux de gènes. Les moules du complexe d'espèces *M. edulis* possèdent une large niche au sein de l'écosystème côtier. Leur distribution géographique s'étant des zones tempérées aux zones subarctique et subantarctique, et est largement dépendante des températures (Seed 1976). Localement, on retrouve les moules majoritairement en milieu intertidal et subtidal supérieur, mais pouvant aller jusqu'à des profondeurs d'environ 40 m pour certaines populations (e.g. certains gisements profonds de *M. edulis* en Normandie, France). Les moules se distribuent également sur un large gradient de salinité, d'exposition aux vagues et de substrats (Seed 1976). Bien qu'apparemment peu sélectives sur leur habitat, les moules peuvent en réalité présenter des patrons très tranchés de distribution, notamment dans des zones de sympatrie entre espèces. La distribution en sympatrie de *M. edulis* et *M. galloprovincialis* est corrélée avec la salinité, l'exposition aux vagues ou la hauteur de marnage (Gardner 1994 ; Bierne et al. 2002b), ce qui crée des patrons micro-géographiques de variation de la distribution. Par exemple, on peut observer des différences de distribution d'espèces à quelques centaines de mètres entre un milieu exposé et un milieu abrité (Comesaña et Sanjuan 1997). De la même manière, *M. edulis* et *M. trossulus* vivant en sympatrie en Mer Blanche, ont respectivement des associations marquées entre des substrats rocheux de fond et des substrats algaux (Katolikova et al. 2016).

Certains traits d'histoire de vie et caractéristiques des moules marines les rendent particulièrement enclines aux introductions. La stratégie de reproduction des moules, mais aussi de beaucoup d'invertébrés marins en général, repose sur la production d'une grande quantité de descendants permettant de compenser la faible probabilité de survie des larves (qualifiée de stratégie r). Leur reproduction comporte un stade pélagique planctotrophe d'environ trois à quatre semaines, pouvant s'étendre dans le temps jusqu'à dix semaines suivant les conditions environnementales (Seed 1976). Durant cette période, les larves sont soumises aux courants marins et peuvent ainsi être dispersées sur de longues distances. Il n'est donc pas étonnant de pouvoir observer des vitesses de propagation allant jusqu'à 120 km/an lors d'une introduction (Branch et Steffani 2004). Cependant, ce type de dispersion ne peut pas expliquer les observations de dispersions très longue distance et inter-continentales (Encadré 3). Les facteurs anthropogéniques sont donc responsables d'un nombre important d'introductions via les eaux de ballast (Geller et al. 1994) ou le bio-fouling. Le bio-fouling par les organismes fixés sur des vecteurs anthropogéniques, représente une cause majeure d'introductions. Les moules peuvent

se fixer sur des coques de bateaux (Apte et al. 2000 ; Casoli et al. 2016), mais peuvent aussi être retrouvées sur des débris marins dérivants (J. A. Miller et al. 2017 ; Miralles et al. 2018 ; Węśławski et Kotwicki 2018).

En addition à ces caractéristiques biologiques, l'action volontaire humaine a aussi engendré des introductions. À l'instar de la culture des huîtres, la mytiliculture a engendré de nombreux déplacements des stocks au cours du dernier siècle, sans réel contrôle de l'origine du fond génétique ou des espèces associées. Par exemple, l'introduction de *M. galloprovincialis* en Colombie-Britannique (Canada) est liée à l'importation de cette espèce à des fins de mytiliculture (Tableau 1). La moule bleue est donc particulièrement encline aux translocations via les deux voies majeures d'introduction en milieu marin : le transport maritime et l'aquaculture (Molnar et al. 2008).

Cependant, une introduction réussie ne se limite pas à l'arrivée de propagules non indigènes dans un nouvel environnement. Il est nécessaire que les individus introduits établissent une population suffisante pour se propager et perdurer dans le temps. Les premières barrières à la colonisation sont des effets Allee démographiques et génétiques, qui peuvent être évités soit par une pression de propagule forte, soit via l'hybridation avec une espèce native (Mesgaran et al. 2016). Les espèces *Mytilus* spp. présentent des solutions à ces deux problèmes, i.e. une fécondité élevée, une colonisation en forte densité et un isolement reproductif partiel au sein du complexe d'espèces.

Le second élément limitant l'établissement d'une espèce est la compétition interspécifique, puisqu'il faut soit la gagner, soit trouver une niche non occupée. Dans le cas de *Mytilus* spp., il apparaît que *M. galloprovincialis* est une espèce particulièrement compétitrice, capable de supplanter des espèces locales (Gardner et al. 1993 ; Branch et al. 2008 ; B. L. Lockwood et Somero 2011). De plus, les moules présentent une grande tolérance aux perturbations environnementales, à la fois naturelles et anthropiques. En effet, elles sont particulièrement robustes à la pollution et sont largement utilisées comme modèle dans les études de toxicologie marine (Roberts 1976 ; Beyer et al. 2017 ; Mlouka et al. 2019). Par conséquent, il n'est pas étonnant de retrouver ces espèces dans les ports, qui sont à la fois des plateformes tournantes du trafic maritime, des nouveaux environnements favorisant la colonisation opportuniste (Bishop et al. 2017), et des milieux anthropisés et perturbés.

i.3.5 *M. galloprovincialis* une espèce au caractère invasif

M. galloprovincialis est classée comme espèce invasive et est retrouvée dans de nombreuses régions tempérées des deux hémisphères (Bonham et al. 2019). Les deux lignées de *M. galloprovincialis*, atlantique et méditerranéenne, sont invasives. L'Encadré 3 tente de dresser une liste exhaustive des régions d'introduction connues et de la lignée introduite. Les cas de *M. edulis* et *M. trossulus* y sont discutés.

L'Afrique du Sud et la Namibie sont les seules régions envahies répertoriées ici qui étaient

précédemment dépourvues d'espèces du complexe *M. edulis* (Tableau 1). Dans ces conditions seul l'aspect écologique est considéré. Cependant, la distribution cosmopolite des moules *Mytilus* spp. et leur isolement reproductif partiel entraîne la rencontre régulière entre lignées génétiques différenciées et la production d'hybrides.

Différents cas de figures ont déjà été observés. L'introduction de *M. galloprovincialis* sur la côte Ouest des États-Unis (Tableau 1), région native de *M. californianus* et *M. trossulus*, a majoritairement eu un impact écologique et un impact limité sur la cohésion génétique des espèces impliquées. En effet, malgré le remplacement probable de *M. trossulus* dans le Sud de son aire de répartition (B. L. Lockwood et Somero 2011), aucune introgression majeure entre ces deux espèces n'a été observée à ce jour (Saarman et Pogson 2015 ; Simon et al. 2019a).

D'autres introductions ont cependant mené à des évènements d'admixture entre *M. galloprovincialis* et l'espèce native. Par exemple, l'introduction des deux lignées de *M. galloprovincialis* en Australie a engendré une admixture avec les moules natives *M. planulatus* (correspondant à la lignée de l'hémisphère sud de *M. galloprovincialis*). En Amérique du Sud, une population échantillonnée à Puerto Madryn (Argentine) est elle aussi admixée entre les *M. galloprovincialis* introduites et les moules natives *M. platensis* (correspondant à la lignée de l'hémisphère sud de *M. edulis*). Cependant, ces observations d'introduction longue distance semblent être les cas les plus visibles. En effet, les introductions dans des régions où *M. galloprovincialis* peut être considérée native, comme dans l'ensemble de l'Europe, sont rarement étudiées car la distinction Atlantique/Méditerranée n'est pas toujours reconnue ou examinée (mais voir Mathiesen et al. 2016). La description actuelle la plus détaillée d'un évènement de contact anthropogénique intra-européen est ce travail de thèse, décrivant l'admixture entre la lignée méditerranéenne de *M. galloprovincialis* et *M. edulis* (chapitres 2 et 3).

Les introductions de *M. galloprovincialis*, invasives ou non, entraînent donc des contacts secondaires récents pouvant être utilisés comme des expériences grandeur nature dans un contexte démographique et spatial particulier. De plus, les multiples introductions souvent observées rendent ce contexte variable, à la fois vis-à-vis des conditions ci-dessus mais aussi du fond génétique local en interaction.

i.4 Objectifs de la thèse et structure du document

Ce projet de thèse repose sur la première observation du fond génétique *M. galloprovincialis* de Méditerranée dans le port de plaisance de Cherbourg (Normandie, France) en 2003, dans une région habitée par *M. edulis* (chapitre 1). De plus, les quelques individus génotypés possédaient un patron d'ascendance génétique original, montrant une admixture entre les *M. galloprovincialis* med. et *M. edulis* locales. Enfin la proportion d'ascendance *M. edulis* était bien plus importante que celle observée dans les zones hybrides naturelles (atteignant 35%), comme si les barrières reproductives avaient sautées dans cette population. Deux approches complémentaires ont alors nourri ce projet de recherche : (i) l'utilisation de nos connaissances

Tableau 1 : Introductions connues de *M. galloprovincialis*. Les références suivantes ne sont pas représentatifs de la première observation mais plutôt de l'identification grâce aux données génétiques.

Origine	Régions d'introduction	Références
Atlantique	Afrique du Sud et Namibie	Grant et Cherry (1985) Daguin et Borsa (2000) Branch et Steffani (2004)
	Australie	Gérard et al. (2008) Westfall et Gardner (2010) Popovic et al. (2019)
	Norvège et Svalbard	Mathiesen et al. (2016) Simon et al. (2019b)
Méditerranée	Pacifique Nord-Est : Oregon, Californie, Mexique	McDonald et Koehn (1988) Geller et al. (1994) Daguin et Borsa (2000) Saarman et Pogson (2015)
	Pacifique Nord-Ouest : Russie, Japon, Corée, Chine	Wilkins et al. (1983) Lee et Morton (1985) McDonald et al. (1990) Daguin et Borsa (2000) Skurikhina et al. (2001) Kartavtsev et al. (2005) P. M. Brannock et al. (2009) P. Brannock et T. Hilbish (2010) Simon et al. (2019a)
	Australie	Gérard et al. (2008) Westfall et Gardner (2010) Popovic et al. (2019)
	Chili	Daguin et Borsa (2000) Westfall et Gardner (2010) Oyarzún et al. (2016) Larraín et al. (2018)
	France : Manche et côte Atlantique	Simon et al. (2019b)
	Indéterminée	Nouvelle-Zélande
Pacifique Nord-Est : Puget Sound et Colombie Britannique		Heath et al. (1995) A. S. Anderson et al. (2002) Crego-Prieto et al. (2015)

Encadré 3 : la diaspora *Mytilus*

M. galloprovincialis : Cette espèce a été largement introduite pour être cultivée dans des régions dont elle n'est pas native (e.g. en Asie ou en Colombie-Britannique), mais elle a aussi été introduite par inadvertance via les eaux de ballast ou le biofouling des navires.

Le Tableau 1 présente les zones d'introduction connues et l'origine potentielle de *M. galloprovincialis* à l'échelle mondiale, souvent vérifiée par des données génétiques.

Bien qu'aucune donnée génétique ne permette de connaître l'origine de *M. galloprovincialis* au Mexique (Bahía de Todos Santos), il apparaît très probable que ce soit la continuité de la population méditerranéenne présente en Californie (McDonald et Koehn 1988 ; Ramírez et Cáceres-Martínez 1999). De plus, la culture ayant été initiée et continuant avec du naissain local, il est peu probable qu'une autre source de *M. galloprovincialis* ait été utilisée (Cáceres-Martínez 1997).

Concernant l'introduction en Colombie-Britannique et le Puget Sound, il est difficile de trancher sur une origine car il existe une discontinuité de données entre le point le plus proche au Sud où des *M. galloprovincialis* méditerranéennes ont été observées et cette région (Saarman et Pogson 2015). Il est tout à fait possible qu'une autre source ait été utilisée pour l'ensemencement des cultures dans cette région.

De même, en Nouvelle-Zélande les données actuelles ne permettent pas de connaître l'origine atlantique ou méditerranéenne de l'introduction. Sachant que les deux fonds génétiques sont présents en Australie, les deux hypothèses ne peuvent pas être écartées et les deux fonds peuvent aussi y être présents.

Les introductions observées en Norvège et au Svalbard par Mathiesen et al. (2016) sur la base de quelques allèles *M. galloprovincialis* ont pu être vérifiées dans cette thèse (Simon et al. 2019b). Cependant, nous concluons que les individus observés proviennent de populations admixées entre *M. galloprovincialis* atl. et la lignée nord-européenne de *M. edulis*.

M. edulis : Il existe un faible nombre de rapports d'introduction de *M. edulis*. Cette espèce a pu aussi être introduite en Colombie-Britannique pour des essais de culture (Heath et al. 1995 ; Crego-Prieto et al. 2015). Cependant l'hypothèse que les allèles *M. edulis* observés dans cette région proviennent de *M. galloprovincialis* atl. introduites n'est pas à exclure. Si *M. edulis* a été introduite sur la côte Ouest du Canada, il est probable qu'elles soient originaires de la côte Atlantique (i.e. des *M. edulis* américaines). Cette espèce a aussi été rapportée de manière temporaire en Méditerranée par Casoli et al. (2016), mais sans colonisation durable.

M. trossulus : Loch Etive (Écosse) est la seule zone d'introduction potentielle de *M. trossulus* rapportée à ce jour (Beaumont et al. 2008). Il a été proposé par Zbawicka et al. (2010), sur la base de la divergence de leur génome mitochondrial, que cette population était une relique post-glaciaire. Cependant l'ensemble des populations de *M. trossulus* n'étant pas analysées dans ce travail, il est difficile d'éliminer l'hypothèse d'introduction.

du complexe d'espèces *M. edulis* afin de décrire et comprendre un évènement d'introduction de *M. galloprovincialis* impliquant de l'hybridation et (ii) l'utilisation d'évènements d'hybridation anthropogénique comme expérience évolutive.

D'autre part, comme l'aurait été une parfaite anticipation, Bierne et al. (2002a) avaient produit des croisements en laboratoire entre *M. edulis* du nord de la France et des *M. galloprovincialis* méditerranéennes. Ces croisements comprenaient des F1, des F2 et des BC vers *M. galloprovincialis*. Bierne et al. (2006) montrent sur ces mêmes croisements qu'il existe des incompatibilités dispersées sur le génome, se conformant à un modèle de BDMI avec dominance des allèles ancestraux sur les allèles dérivés (Turelli et Orr 1995), qui doivent fortement contribuer au maintien de barrières fortes entre ces deux espèces. Cependant, une analyse préliminaire de ces croisements avec un plus grand nombre de marqueurs avait révélé des distorsions sur l'ensemble du génome vers une augmentation de l'hétérozygotie en F2 difficilement conciliable avec l'action de quelques BDMI. Combinée à la réalisation toujours plus présente que l'on sous-estime sûrement l'impact des interactions épistatiques polygéniques dans l'isolement reproductif (cf. i.1.2), nous trouvons important d'explorer d'autres modèles d'incompatibilité entre espèces partiellement isolées reproductivement.

Partant de ces constats, la thèse s'est construite sur trois grands axes de recherche :

1. Le premier était la description de cette introduction si particulière observée à Cherbourg et l'augmentation de l'effort d'échantillonnage dans toute la région, à la fois dans les ports et dans les populations naturelles, afin de déterminer l'étendue du phénomène. Pour cela, il a fallu développer un jeu de marqueurs informatifs sur l'ascendance génétique, permettant un génotypage fiable et à très grande échelle. L'objectif était d'abord d'augmenter nos connaissances globales sur les relations au sein du complexe d'espèces *M. edulis*, pour ensuite utiliser celles-ci dans un but de description de l'introduction et de suivi d'une possible invasion. Nous souhaitons décrire la distribution spatiale de ces moules introduites et leurs interactions avec les espèces natives, en utilisant la génétique des populations comme outil.
2. Le second axe de recherche était l'utilisation de ce contact secondaire récent comme une expérience évolutive sans précédent dans le complexe *M. edulis*. De plus, nous pouvions compter sur d'autres évènements d'admixture, à la fois d'autres introductions, mais aussi des contacts plus anciens et des croisements de laboratoire, pour explorer différentes facettes du processus d'admixture. L'objectif était dans un premier temps de profiter des nombreux évènements d'admixture indépendants, entre fonds génétiques différents, afin d'explorer les parallélismes entre contacts. Dans un deuxième temps, l'objectif était d'observer comment la sélection contre les incompatibilités se résolvait à l'échelle génomique dans ces différents contextes. Nous souhaitons explorer un continuum de scénarios entre (i) la présence de distorsions majeures vers l'un ou l'autre des parents sur la majorité du génome, marque d'effets sélectifs forts sur les combinaisons produites par l'hybridation, et (ii) une admixture homogène à l'échelle du génome, dans lequel des allèles parentaux sont remélangés et produisent une grande diversité de combinaisons viables et fertiles. Enfin, en lien avec le premier axe, nous souhaitons nous intéresser aux mécanismes en place dans les zones de contacts entre les différents fonds génétiques du complexe d'espèces

M. edulis.

3. Le dernier axe de recherche a porté sur l'aspect théorique de l'isolement reproductif, notamment la sélection contre les hybrides. L'objectif était d'explorer une vision polygénique de l'isolement reproductif et de pouvoir modéliser des interactions épistatiques à un ordre de grandeur supérieur à deux ou trois locus. Pour y parvenir sans recourir à une explosion des paramètres, qui aurait eu lieu avec un modèle BDML, nous souhaitons utiliser le modèle géométrique de Fisher pour proposer des prédictions sur la sélection à l'oeuvre dans les premières générations d'hybridation – et en premier lieu dans les F1, F2 et BC afin d'avoir accès à des données de croisements pour validation.

Le chapitre 1 a d'abord permis de développer les marqueurs nécessaires à la réalisation d'un génotypage sur de nombreux individus et à faibles coûts en recherchant un jeu de marqueurs informatifs sur l'ascendance entre espèces du complexe *M. edulis* ainsi qu'entre lignées de la même espèce (*M. galloprovincialis* atl. vs *M. galloprovincialis* med., *M. edulis* eu. vs *M. edulis* am., *M. trossulus* eu. vs *M. trossulus* am.). Ces marqueurs ont permis d'analyser les patrons d'introgression locaux naturels au sein du complexe et l'identification de nouvelles lignées.

Le chapitre 2 présente ensuite les résultats du génotypage haut-débit à large et fine échelle spatiale permettant dans un premier temps de caractériser la nature du fond génétique introduit, puis dans un second temps de décrire la répartition géographique de celui-ci. Il développe ensuite les premières observations des patrons de distorsion (i.e. des écarts par rapport à la moyenne génomique) à l'échelle des 77 marqueurs informatifs analysés.

À la suite de ces premières conclusions, le chapitre 3 se concentre sur l'échelle génomique en analysant 156 génomes de faible couverture contenant un échantillonnage représentatif de populations de référence et de plusieurs populations admixées entre fonds génétiques différents et dans des proportions variables.

Dans un objectif de compréhension théorique des phénomènes observés dans les chapitres précédents, le chapitre 4 a pour ambition de proposer un modèle généraliste prédisant la valeur sélective des hybrides en fonction de leur composition génomique globale, en utilisant le modèle géométrique de Fisher comme processus génératif des co-adaptations génomiques lors de la divergence.

Chapitre 1

Comètes d'introgression dans le complexe d'espèces *M. edulis*

1.1 Contexte et résumé de l'étude

1.1.1 Obtenir une vision plus précise des liens au sein du complexe

Avant le début de cette thèse, il y avait un écart important, dans le nombre de marqueurs et la taille des échantillons utilisés, entre les études du complexe d'espèces *M. edulis*. D'un côté, on retrouvait un nombre important d'études de génétique des populations basées sur un nombre restreint de marqueurs historiques, supposés diagnostiques entre espèces (e.g. Glu-5', EFbis, Mac1) ou sur des séquences mitochondriales. Ces études avaient malheureusement plusieurs limites : (i) elles sous-estimaient beaucoup l'effet de l'introgression locale et de la présence d'allèles ségrégeant en faible fréquence dans presque toutes les populations du complexe ; (ii) elles sous-estimaient l'existence de différenciation entre certaines lignées du complexe ; et (iii) elles n'avaient pas souvent un échantillonnage représentatif de l'ensemble des entités génétiques du complexe d'espèces. D'un autre côté, Fraïsse et al. (2016a) ont proposé la première étude de génomique des populations du complexe d'espèces *M. edulis*. Cependant, pour offrir une base pour l'étude du complexe, ils ont dû se limiter à quelques populations de référence, plutôt représentatives des points extrêmes de différenciation entre lignées et de quelques points intermédiaires, mais non représentatifs de la distribution incroyablement continue des transitions entre clusters génétiques (cf. chapitre 5). Ainsi, dans un premier temps, ce chapitre vise à combler ce manque dans la littérature, i.e. de pouvoir décrire de manière fiable et à large échelle spatiale les processus à l'œuvre au sein du complexe d'espèces *M. edulis*.

1.1.2 Introgressions locales et zones de tension

Dans la littérature, l'étude des zones d'hybridation revient souvent à l'échantillonnage dans la zone de contact de quelques populations admixées, de quelques populations bordant la zone de contact et d'une population un peu plus distante de chaque côté (e.g. T. Hilbish et al. 2003 ; Gosling et al. 2008 ; Stuckas et al. 2009 ; McKenzie et al. 2016 ; Strelkov et al. 2017), mais rarement à l'échantillonnage de populations n'ayant pas eu d'influence récente de flux de gènes, de multiple transects ou de zones de contact équivalentes. Bien entendu, ce type de système n'est pas toujours disponible et chaque modèle d'étude présentera des caractéristiques différentes, mais lorsqu'elle est disponible, l'image globale peut être importante pour comprendre les patrons locaux.

L'objectif de cette étude était de profiter du nombre important de contacts entre espèces et lignées du complexe *M. edulis* afin d'observer à une large échelle les patrons d'introgressions locales. Fraïsse et al. (2016a) avaient identifié un certain nombre de marqueurs avec une différenciation extrême entre populations très éloignées, dont une partie était due à de l'introgression locale (e.g. de l'introgression *M. edulis* chez *M. galloprovincialis* atl.). Un sous jeu de données a été constitué à partir de ces marqueurs afin de pouvoir augmenter l'effort d'échantillonnage. Le second objectif était d'échantillonner le cœur des zones de tension pour regarder le comportement de ces marqueurs extrêmes. L'hypothèse était que des variants bi-stables asymétriques, ou des introgressions adaptatives, montreraient des déviations des clines génomiques au sein des zones d'hybridation.

1.1.3 Résumé

Nous étudions ici l'ensemble des espèces du complexe *M. edulis* présentes en hémisphère nord en utilisant des populations de référence correspondant à celles de Fraïsse et al. (2016a) et des populations soit que l'on sait introgressées (e.g. la mer Baltique), soit qui se situent dans une zone de contact secondaire.

Nous confirmons dans cette étude les résultats que l'introgression est répandue dans le complexe d'espèces *M. edulis* (Fraïsse et al. 2016a). De plus, augmenter la taille des échantillons par rapport à l'étude précédente, permet de montrer que les marqueurs choisis pour être fixés ou fortement informatifs, présentent toujours des variants ségrégeant à faible fréquence à l'échelle de l'espèce. De plus, ce chapitre permet d'établir un point de référence de la structure génétique au sein du complexe *M. edulis*.

L'observation de l'ensemble des populations permet de mettre en évidence des « comètes d'introgression » présentes dans de nombreuses populations. Ces comètes sont définies comme des introgressions locales. Les différents patrons pointent vers différents temps d'introgression, ayant eu lieu lors d'un contact précédent ou rapidement après contact. Nous concluons que dans les zones de tension, l'observation à l'échelle locale montre des clines génomiques globalement concordantes. Étant donné la structure mosaïque des zones de contact dans le complexe

M. edulis, des introgressions locales peuvent être stoppées à la barrière suivante et, bien que bloquées transitoirement, peuvent produire quelques déviations au sein de la zone de tension.

1.2 Article 1

Cet article a été soumis au *Journal of Evolutionary Biology* pour être potentiellement inclus dans une édition spéciale sur la spéciation en milieu marin.

A. Simon, C. Fraïsse, T. El Ayari, C. Liutard-Haag, P. Strelkov, J. J. Welch et N. Bierne (2019a). « Local Introgression at Two Spatial Scales in Mosaic Hybrid Zones of Mussels ». *bioRxiv*. DOI : 10.1101/818559. Soumis

Local introgression at two spatial scales in mosaic hybrid zones of mussels

Alexis Simon¹, Christelle Fraïsse^{1,2}, Tahani El Ayari¹, Cathy Liautard-Haag¹, Petr Strelkov^{3,4}, John J. Welch⁵ and Nicolas Bierne¹

¹ISEM, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France.

²Institute of Science and Technology Austria, Am Campus 1, Klosterneuburg 3400, Austria.

³St. Petersburg State University, Universitetskaya Emb. 7/9, St. Petersburg 199034, Russia.

⁴Laboratory of Monitoring and Conservation of Natural Arctic Ecosystems, Murmansk Arctic State University, Kapitana Egorova Str. 16, Murmansk 183038, Russia.

⁵Department of Genetics, University of Cambridge, Downing St. Cambridge, CB23EH, UK.

Abstract

When the ranges of closely-related lineages are large, and overlapping, we can often study introgression at many “replicated” contacts, with different locations and spatial scales. Here we analysed multiple contact zones of the *M. edulis* complex of marine mussel species, which represent a mosaic distribution of heterogeneously differentiated, semi-isolated genomes. Our aim was to contrast ongoing introgression at the heart of hybrid zones, with past introgression between similar parental populations, at increasing distance from the contact. Using a panel of ancestry-informative SNPs derived from a previous genomic study, we first confirm, with a broader sampling, that local introgression, affecting one but not all of the populations compared, is both widespread and heterogeneous across the genome. Some outlier loci show patterns of complete introgression in certain populations, and an absence of introgression in others. Genomic cline analyses reveal a globally high concordance among loci at a local scale, albeit with signals of asymmetric introgression at a few loci. Enhanced local introgression at specific loci is consistent with the early transfer of adaptive variants after contact, possibly including asymmetric bi-stable variants, or less loaded alleles. Given the mosaic structure of the *M. edulis* complex, with a succession of genetic barriers to gene flow, variants with enhanced introgression through one barrier can be trapped, maybe transiently, at the next barrier, confining introgression locally. This makes the *Mytilus* complex an ideal model of the heterogeneous porosity of species barriers.

1 Introduction

Divergence between species is almost always accompanied by hybridisation (Abbott et al., 2013), and so the study of speciation is intertwined with the study of introgression. Speciation researchers are interested in both contemporary introgression, which can inform us about the nature of incompatibilities between lineages, and in historic introgression, which can inform us about the history of the speciation process. However, the comparison of ongoing and ancient introgression rates is difficult, and therefore rarely done (Bouchemousse, Liautard-Haag, Bierne, & Viard, 2016; Chaturvedi et al., 2019). Estimating patterns of introgression across the genome can reveal regions of reduced gene flow, due to genetic incompatibilities, divergent selection or pre-zygotic isolation. However, these patterns are also confounded by factors such as linked-selection, and variation in mutation and recombination rates (Ravinet et al., 2017). The study of enhanced introgression in one population, when compared to a reference parental population (sometimes called 'local introgression', Fraïsse, Belkhir, Welch, and Bierne, 2016) is an efficient approach to isolating the effect of introgression from other factors (Green et al., 2010; Martin, Davey, & Jiggins, 2015). In addition, the study of multiple cases of contact between similar lineages allows us to study the effect of the same barrier in multiple demographic and ecological contexts (Abbott et al., 2013), and to test the repeatability of introgression patterns (Harrison & Larson, 2016; Simon et al., 2019).

Many hybrid zones result from secondary contacts, and multiple outcomes can be expected depending on intrinsic and extrinsic factors. The absence of strong enough genetic or migration (i.e. physical) barriers, as well as demographic unbalance, can result in massive introgression. In this scenario, only strongly selected loci would resist introgression while most of the genome tends to homogenise (introgression swamping). If the barrier to gene flow is strong enough, only adaptive and compatible alleles (and flanking hitchhiker loci) can cross it without delay (Nicholas H. Barton, 1979a; Faure, David, Bonhomme, & Bierne, 2008). If the variation is bi-stable (Nicholas H. Barton & Turelli, 2011) such that one parental genotype is fitter but heterozygous or recombinant genotypes are unfit, the spread of the fittest genotype is hindered by the barrier (Nicholas H. Barton, 1979a; Piálek & Barton, 1997). Stochastic processes, such as random drift, or variable migration rates, can free the spreading wave from the barrier trap, but the delay can be very long (Piálek & Barton, 1997). Most of the time, in a setting where genetic and migration barriers are coupled, secondary contact will result in a 'porous barrier' with a slow erosion of the differentiation at a rate that depends on recombination rates and the density of barrier loci (Nicholas H. Barton & Bengtsson, 1986). This has been observed in many recent genome papers (Aeschbacher, Selby, Willis, & Coop, 2017; Duranton et al., 2018; Gagnaire et al., 2018; Martin, Davey, Salazar, & Jiggins, 2019; Roesti, Moser, & Berner, 2013; Schumer et al., 2018).

In the marine environment, and especially for broadcast spawners with a dispersive larval phase, migration barriers are potentially rare and species are likely to be subdivided by hybrid

zones rather than by geographic isolation. Nevertheless, where dispersal barriers do exist – such as at well-known bio-geographic boundaries characterised by oceanic fronts and environmental shifts – genetic breaks tend to coincide with them. This is well illustrated in the complex of marine mussels, comprising the three species *M. trossulus*, *M. edulis* and *M. galloprovincialis*. This *Mytilus* species complex is subdivided into a mosaic of semi-isolated genetic clusters, by numerous hybrid zones (Bierne, Borsa, et al., 2003; El Ayari, Trigui El Menif, Hamer, Cahill, & Bierne, 2019; Fraïsse, Belkhir, et al., 2016; Riginos & Cunningham, 2005; Strelkov, Kattolikova, & Väinölä, 2017). In the northern hemisphere all three species form a geographic mosaic delimited by multiple hybrid zones (Fraïsse, Belkhir, et al., 2016). In addition, each species is itself subdivided into infra-specific lineages, sometimes by geographic isolation (East and West Atlantic in *M. edulis* and *M. trossulus*, Riginos and Henzler, 2008; Varvio, Koehn, and Väinölä, 1988) and sometimes by hybrid zones (Atlantic and Mediterranean lineages in *M. galloprovincialis*, El Ayari et al., 2019; Quesada, Zapata, and Alvarez, 1995). Therefore, this complex provides us with numerous post-glacial hybrid zones with a continuum of divergence between the interacting taxa. Although introgression can sometimes be extensive, both intrinsic and extrinsic mechanisms of reproductive isolation maintain narrow admixture zones in which early generation hybrids are continuously produced. Finally, genetic differentiation is highly heterogeneous across the genomes, with a gradient from virtual panmixia (Boon, Faure, & Bierne, 2009) to differentially fixed loci (Fraïsse, Belkhir, et al., 2016). Precisely because of this heterogeneity, however, it is challenging to identify diagnostic loci for the *Mytilus* complex, even for the species *M. edulis* and *M. galloprovincialis*. We took advantage of a population genomics analysis of around 1300 more than 5 Kb long contigs (Fraïsse, Belkhir, et al., 2016) to develop a panel of ancestry-informative markers. This panel allowed us to study introgression between parental populations of the mosaic, as well as within hybrid zones using genomic cline analysis (Gompert & Buerkle, 2011). After a verification that the SNP panel was effective, we confirm that local introgression is both widespread and heterogeneous across the genome. At some loci identified as outliers in (Fraïsse, Belkhir, et al., 2016), heterospecific alleles have sometimes fixed, or nearly so, in one parental patch close to a hybrid zone while they remain nearly absent from other population patches farther from the zone. Genomic clines suggest high concordance in the heart of hybrid zones, although a few loci do depart from the genomic average, and demonstrate asymmetric introgression. Unlike outliers that exhibit enhanced local introgression at a large scale, the introgression of these genomic cline outliers does not extend outside of the admixture zones.

Table 1: Definition of reference populations.

Level 1	Level 2	Level 3	comment
<i>M. trossulus</i> (tros)	America (am)	-	Saint Lawrence
	Europe (eu)	Baltic (baltic)	Baltic sea
	Pacific (pac)	East (east)	from Californian coast
<i>M. edulis</i> (edu)	America (am)	-	Long Island region
	Europe (eu)	external (ext)	-
		local (int)	-
		North (north)	Russia and Scandinavia
<i>M. galloprovincialis</i> (gallo)	Atlantic (atl)	external (ext)	-
		local (int)	-
	Mediterranean (med)	West (west)	-
		East (east)	-
		Black Sea (bs)	-

Abbreviations used in the manuscript and figures are combinations of text in parentheses (e.g., edu_eu_north).

2 Materials and methods

2.1 Sampling

Mytilus spp. individuals were sampled from 58 locations, including several known hybrid zones (Figure 1, Table 1). Sampling sites are located on the American Pacific coast, the American and European North Atlantic coasts, and in the Mediterranean, Baltic, North, Barents and Black seas. 441 individuals were newly genotyped and 72 genotypes by sequencing (GBS) were extracted from the published dataset of Fraïsse, Belkhir, et al. (2016).

2.2 Assay design

We aimed to genotype ancestry informative loci, across a large number of samples, in a cost-effective manner. For this purpose we used an Illumina BeadXpress® assay with Veracode™ technology (GoldenGate Genotyping Assay). We designed an assay of 384 SNPs (being the multiplexing limit of the technology).

Loci were selected, prior to genotyping, based on their ancestry informativeness, using the published results of (Fraïsse, Belkhir, et al., 2016). Briefly, this database was produced via a target enrichment method on the three species and multiple populations of the *Mytilus* species complex (Fraïsse, Belkhir, et al., 2016). It contains 1269 contigs sequenced for 72 individuals from eleven different locations <http://www.scbi.uma.es/mytilus/index.php>.

Markers with a minimum allele frequency of 0.05 and a maximum missing data percentage of 50% were retained. Coverage was estimated as a mean computed on three populations

Table 2: Number of SNPs per type of comparison. The SNP added to the dataset (extra) corresponds to an introgressed amino-acid changing variant (see main text).

Comparison	Number of SNPs genotyped	Number of SNPs retained
edu / all	2	2
edu_am / edu_eu	39	22
within edu_am	17	5
edu_eu / all	51	16
gallo_eu / all	12	4
gallo_atl / gallo_med	27	16
gallo / edu	100	65
gallo_med / all	6	1
tros_am / tros_eu	34	12
tros / all	96	68
Total	384	211 + 1 extra

(two Atlantic *M. galloprovincialis* and one Atlantic *M. edulis*). Contig regions with especially high coverage (> 300 reads) were excluded to avoid repeated elements. Regions of the database produced from cDNA were blasted against a draft genome (Murgarella et al., 2016) to exclude regions close to intron/exon limits as flanking sequences in 3' and 5' of the SNP were needed to design primers for PCR amplification. For the same reason, SNPs close to the start and end of the contigs were also excluded. An ADT score, produced by Illumina Assay Design Tool, quantifies the expected amplification success, and was used to filter the SNPs with the most probable design success (ADT score > 0.4). Finally, the most differentiated SNPs (using F_{ST}) between any population from Fraïsse, Belkhir, et al. (2016) were retained. After this filtering, a few contigs of special interest in mussels were rescued (e.g. immune genes or ecologically relevant traits like adhesion protein). Finally, SNPs were classified in terms of comparison informativeness, and selected to constitute a final 384 SNPs dataset, balanced across comparisons (Table 2). The additional SNP ("extra") correspond to an amino-acid variant introgressed into *M. galloprovincialis* Atlantic from *M. edulis* detected in Fraïsse, Belkhir, et al. (2016) (contig gi_385288268_emb_Contig56466, annotated for a tumor necrosis factor).

2.3 Genotyping and filtration

Genotyping of the 441 individuals was performed with the BeadXpress® (hereafter BXP) technology.

Among the 384 SNPs genotyped, 252 were readable (clusters of homozygotes/heterozygotes well defined) and 132 were lost. This low rate of successful amplification is expected in such highly polymorphic species. 40 additional SNPs were removed due to a differentiation between the BXP and GBS typing within populations. The threshold of missing data per SNP and per

individual was set to 10%. One marker (190) was rescued as the missing data was mostly due to low amplification in *M. trossulus* (overall 12% missing data).

The genotypes for the 77 GBS individuals were retrieved from the calling in Fraïsse, Belkhir, et al. (2016).

Two replicated individuals between the GBS and BeadXpress were present in the dataset. They showed mismatch levels of, respectively, 2.83% and 2.36% between the two experiments, and this was due entirely to a well-known heterozygote assignment bias in GBS experiments. The two replicated GBS individuals were removed from further analyses.

Four samples were removed from the analyses prior to filtering as they have already proven to be cancerous individuals (Riquet, Le Cam, Fonteneau, & Viard, 2016, July 2015) (Por_40 from population 34; Arsud_05, Arsud_07 from population 38; and Ret_04 from population 19).

Reference populations were defined on previous knowledge of the *M. edulis* species complex (Fraïsse, Gunnarsson, Roze, Bierne, & Welch, 2016; Simon et al., 2019). We used three levels of differentiation representing the species level (L1), the ocean basin or continent (L2) and regional groupings (L3) (Table 1). We used the GBS samples from Fraïsse, Belkhir, et al. (2016) as references and previously untyped individuals were assigned to a reference group if they belonged to the same GBS population or were close in geographic distance from it. We used a preliminary Admixture analysis at the species level (L1) using $K = 3$ and default settings (30 independent runs merged with CLUMPAK (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015)) to correct the reference groups for migrants or hybrids, as sympatry and hybridisation is a common phenomenon in the *M. edulis* species complex. Individuals assigned to those populations, and not filtered out, constitute the “reference dataset”.

Due to suspicious levels of *M. trossulus* ancestry in locations devoid of this species, which could indicate the presence of a transmissible cancer (Riquet et al., 2016, July 2015), two additional samples from the Mediterranean sea were removed (MTP_05 and Collo_10 from populations 48 and 52 respectively).

A more stringent filtration was additionally applied to the edu_am population when used as a reference, given the presence of European ancestry in the Long Island Sound (populations 15 and 16, Figure S10, also described in Simon et al., 2019). Filtering yielded 514 individuals genotyped at 212 markers.

Hardy-Weinberg equilibrium was tested for the remaining markers within putative panmictic clusters outside of known hybrid zones (pegas 0.11). Exact tests were performed using 1000 bootstraps and p-values were adjusted for false discovery rate using the Benjamini-Yekutieli method (Benjamini & Yekutieli, 2001).

We used a partial genetic map produced in Simon et al. (2019) based on an F2 cross between Mediterranean *M. galloprovincialis* and *M. edulis*, genotyped at a subset of the markers studied here (Bierne, Bonhomme, Boudry, Szulkin, & David, 2006; Bierne, David, Boudry, & Bonhomme, 2002; Simon, Bierne, & Welch, 2018). In addition to this genetic map, we

extrapolated genetic distances from physical distance for markers present on the same contig using a recombination rate of 2 cM/Mb (rounded from the estimate in Bierne, 2010).

2.4 Population structure

A principal components analysis (PCA) was performed using markers on different physical contigs (retaining the one with least missing data). This filtration was to avoid strong biases due to physical linkage, and led to a final set of 160 markers. The genotype data was centred and scaled using the *adegenet* R package (Jombart, 2008), with the replacement of missing data by the mean allele frequencies. Following the PCA, a dimensional reduction method, UMAP (Diaz-Papkovich, Anderson-Trocme, & Gravel, 2018; McInnes & Healy, 2018), was applied to the first 11 principal components. This threshold was chosen based on the expectation of 12 panmictic populations (level L3 of the reference groups, Diaz-Papkovich et al., 2018). This method was performed using the python package *umap-learn* (McInnes & Healy, 2018) and the R wrapper package *umap* (Konopka, 2019).

The *Structure* software was used to provide Bayesian estimates of ancestry with an admixture model. *Structure* was run with the admixture and linkage model (LINKAGE = 1). The dataset was filtered with the following steps: (i) markers out of Hardy-Weinberg equilibrium were removed; (ii) one marker per physical contig was retained (keeping the one with least missing data); and (iii) the genetic map was used to produce linkage information for the retained markers, with the assumption that markers absent from the genetic map were unlinked (for lack of further information).

For the global *Structure* analysis, 20 independent Monte Carlo Markov Chains (MCMC) runs of 20,000 burn-in iterations followed by 80,000 steps were performed to estimate model parameters for each K between 4 and 8. The standard deviation for the α prior was set to 0.05 for better mixing of the chains. The *Clumpak* software (Kopelman et al., 2015) was used to investigate and aggregate *Structure* outputs with an MCL threshold of 0.9. Only major clusters are presented in the results.

2.5 Hybrid zones analyses

Each hybrid zone was defined by parent 1 (P1), parent 2 (P2) and central populations (Table 3). Each parental population was classified as either “local” or “peripheral”, according to observed levels of introgression or geography. In some cases the “local” parental population was not available in our sampling design and in these cases, the peripheral population was used in its place (Table 3).

Table 3: Grouping of populations for each hybrid zone considered in this study.

Hybrid zone	species	P1 peripheral	P1 local	Central	P2 local	P2 peripheral
Øresund	edu_am/tros_am	18-NovyMost 19-Retinskoye		20-Oresund-Helsinborg 21-Oresund-Raa	22-Tvarminne	08-StLawrence-CBD 09-StLawrence-TAD
Scotland	edu_eu/gallo_atl	26-Hollande 27-Calais 28-Barfleur		24-Aberdeen	31-Roscoff 32-Guillec	40-Faro 41-Dahkla
Brittany	edu_eu/gallo_atl	26-Hollande 27-Calais 28-Barfleur	35-Gascogne 36-Biscay 37-Aiguillon	33-Kerbihan 34-Pornichet	31-Roscoff 32-Guillec	40-Faro 41-Dahkla
Aquitaine	edu_eu/gallo_atl	26-Hollande 27-Calais 28-Barfleur	35-Gascogne 36-Biscay 37-Aiguillon	38-Arcachon	39-Biarritz	40-Faro 41-Dahkla
Algeria	gallo_atl/gallo_med	40-Faro 41-Dahkla		42-Nador 43-Arzew 44-Moustaganem 45-SidiLakhdar 46-ElMarsa 47-Ghouraya 48-Tipaza 49-SidiFredj 50-Bejaia	51-Ziama 52-Collo	55-Bizerte 56-Heraklion 57-Croatia
North European <i>M. edulis</i>	edu_am/edu_eu	26-Hollande 27-Calais 28-Barfleur		18-NovyMost 19-Retinskoye		10-Lgls-SHM 11-Lgls-QNT 12-Lgls-BYY 13-Lgls-OST 14-Lgls-BFM 15-Lgls-MYC 16-Lgls-SPM 17-Boston

For each hybrid zone, three analyses were carried out: (i) computation of hybrid indexes with the R package `introgress` (Gompert & Buerkle, 2010), (ii) a local Bayesian clustering analysis using `Structure`, and (iii) a genomic clines analysis.

`Structure` analyses for each hybrid zone used the same parameters as for the combined data set, and the admixture with linkage model. The filtration of markers was similar to the global analysis. Each subset was filtered for irrelevant genetic backgrounds in the hybrid zone studied. For example, *M. trossulus* individuals were removed from the Northern *M. edulis* (Russia) reference populations for the Øresund hybrid zone study. Each hybrid zone was studied for a K of 2 and 3, to make sure there was no hidden substructure in the subset of individuals considered.

Genomic clines are used to detect markers deviating from the average genomic expectation given the distribution of hybrid indexes. The program `bgc` was used to estimate genomic cline parameters with a Bayesian method (Gompert & Buerkle, 2011, 2012) for each hybrid zone considered. Datasets were prepared with custom R scripts. For each hybrid zone, fixed markers were removed as they are uninformative. The same filtered marker sets as for the `Structure` analyses were used. Four independent chains of 200 000 iterations including 20 000 burn-in iterations with a thinning of 20 were performed. We used the ICARrho model for linked loci with the previously generated genetic map. The R package `rhdf5` (Fischer, Pau, & Smith, 2019) was used to read the MCMC outputs. Convergence was assessed using the method and R code of Vehtari, Gelman, Simpson, Carpenter, and Bürkner (2019).

For each hybrid zone, we performed two analyses called “local” and “peripheral”. The local analysis considers only the “central” population as admixed and uses the local P1 and P2 populations as parental references (Table 3). In the case of a missing local parental population, the peripheral corresponding one was considered (e.g. in the Øresund hybrid zone). The peripheral analysis considers both the central and local parental populations as admixed, while taking the peripheral P1 and P2 populations as parental references.

Loci exhibiting extreme deviations from the neutral genetic background were determined by two methods. First, we estimated locus-specific posterior distributions for the cline parameters α and β . Loci were classified as having “excess ancestry” if the 95% quantiles of these distributions did not include 0 (Gompert & Buerkle, 2012).

3 Results

We use a set of 212 ancestry informative markers to investigate the population genetics of several hybrid zones, and other introgressed populations in the *Mytilus* species complex. We start by showing that our target species and populations are identifiable with our SNP panel. We also identify previously uncharacterised lineages or admixed clusters. Then we analyse the differentiation and introgression patterns in the hybrid zones, at two spatial scales.

3.1 Power of the SNP panel and clustering results

Our markers were taken from a previous GBS study (Fraïsse, Belkhir, et al., 2016), and so we first tested if the markers continued to be informative in our larger sample. To do this, we correlated F_{ST} values, between the complete dataset (BeadXpress and GBS), and the GBS genotypes considered alone (Table S2). Results showed good delimitation (i.e. high F_{ST} correlations) between species and between known semi-isolated lineages within a species (e.g. Atlantic ocean, Almeria-Oran front), while comparisons between less separated entities were less successful (i.e. low F_{ST} correlations). Overall, however, the assay design produced a strong enrichment in high F_{ST} markers compared to the distributions of Fraïsse, Belkhir, et al. (2016), showing that our markers are ancestry informative (Figure S1).

Although the markers are informative, another consequence of increasing the sample size was that truly diagnostic markers became rare or absent at the species level (Table 4). While our dataset contained many allele frequency differences greater than 0.7, only two markers showed fixed differences between species. Indeed, when considering all reference individuals of one species, only the marker 147 was fixed between *M. edulis* and *M. galloprovincialis* (Contig H_L1_abysse_Contig244 position 6092 in Fraïsse, Belkhir, et al., 2016) and only the marker 015 was fixed between *M. edulis* and *M. trossulus* (Contig Contig17324_GA36A position 1089 in Fraïsse, Belkhir, et al., 2016). No marker was fixed between *M. galloprovincialis* and *M. trossulus*.

Using our ancestry informative markers, five genetic clusters could be defined without ambiguity in a Principal Component Analysis (Figure S2). This is visualised, and combined with a Structure analysis in Figure 2 (see also Figure S4). Figure 2, presents a clear picture of population structure in the *Mytilus* complex. First, the three species are clearly differentiated, and so are known genetic clusters within each species: (i) European and American *M. edulis*, (ii) Atlantic and Mediterranean *M. galloprovincialis*, (iii) Baltic and American/North-European *M. trossulus*. We observed that the Baltic population of *M. trossulus* (22) is introgressed by *M. edulis*, as previously described (Fraïsse, Belkhir, et al., 2016), and this cluster can be identified with Structure with $K = 8$ (Figure S3).

The remaining genetic clusters in Figure 2 are labelled A-E and represent admixed populations (see also S2). For example, the hybrid zones of Brittany, Aquitaine and Scotland all involve Atlantic *M. galloprovincialis* and European *M. edulis*, and all are found in group C (Figure 2). By contrast, the Algeria hybrid zone involves Atlantic and Mediterranean *M. galloprovincialis*, and is group D (Figure 2).

Group E contains a small group of individuals sampled in the leisure marina of Cherbourg in 2003, on the French coast of the English Channel (29-Cherbourg). These individuals exhibited an admixture, largely between Mediterranean *M. galloprovincialis* (66% ancestry) and European *M. edulis*, (29% ancestry). We note that a nearby population, in Barfleur (population 28, around 30 km from Cherbourg), is composed exclusively of European *M. edulis*, the

Table 4: Number of markers in three categories of allele frequency difference (AFD), for the comparison between the three species. The allele frequencies are computed using all individuals with the L1 level in the reference dataset.

comparison	$AFD > 0.5$	$AFD > 0.7$	$AFD = 1$
<i>M. edulis</i> / <i>M. galloprovincialis</i>	79	62	1
<i>M. edulis</i> / <i>M. trossulus</i>	108	78	1
<i>M. trossulus</i> / <i>M. galloprovincialis</i>	142	115	0

expected local genetic background.

Group A comprises North-European *M. edulis*. Those *M. edulis* individuals found in Russia (populations 18 and 19), in sympatry with *M. trossulus*, exhibit intermediate ancestries between South-European *M. edulis* (edu_eu_ext and int) and the American cluster of *M. edulis* (edu_am; Figures 2 and S10). Additionally, they do not differentiate on a secondary axis on a classic PCA (Figure S2). The ancestry intergradation observed in Figure S10 for populations 18 and 19 cannot be attributed to unaccounted *M. trossulus* ancestry, because those individuals do not exhibit introgression from this species in the global Structure analysis (Figure 2). A similar admixture is also visible in the Edinburgh population (23), which has additional Atlantic *M. galloprovincialis* ancestry due to its localisation close to the Scottish hybrid zone. Finally, group C corresponds to admixed individuals of the Øresund hybrid zone between the newly identified North-European *M. edulis* and the Baltic *M. trossulus* and populations 20-21 in Figure 2).

Structure analysis shows some further patterns. First, on the West coast of the USA (populations 01 to 07, Figure 2), we observe introduced Mediterranean *M. galloprovincialis* individuals, and a single F1 hybrid with the native *M. trossulus* parent (population 06). This is consistent with the report of Saarman and Pogson (2015).

Second, *M. edulis* samples from the Long Island region (USA, East coast, populations 10 to 16) were mainly assigned to the American *M. edulis* cluster (Figure 2), but there also appears to be some infraspecific ancestry coming from Europe. An analysis considering only *M. edulis* samples, Figure S10, shows more clearly that individuals from the most southern populations of the Long Island Sound sometimes have higher European ancestry than other American populations (e.g. Boston, population 17).

3.2 “Comets” of introgression

We investigated the allele frequency differences between an intergradation of populations both between *M. edulis* and *M. galloprovincialis* (Figure 3), and between *M. edulis* and *M. trossulus* (Figure 4). Because of their visual signature (Figure 3A), we define “comets” as loci with introgression in some, but not all, populations (Staubach et al., 2012). These comets of introgression were previously identified as within-species outlier loci (Fraisie, Belkhir, et al., 2016).

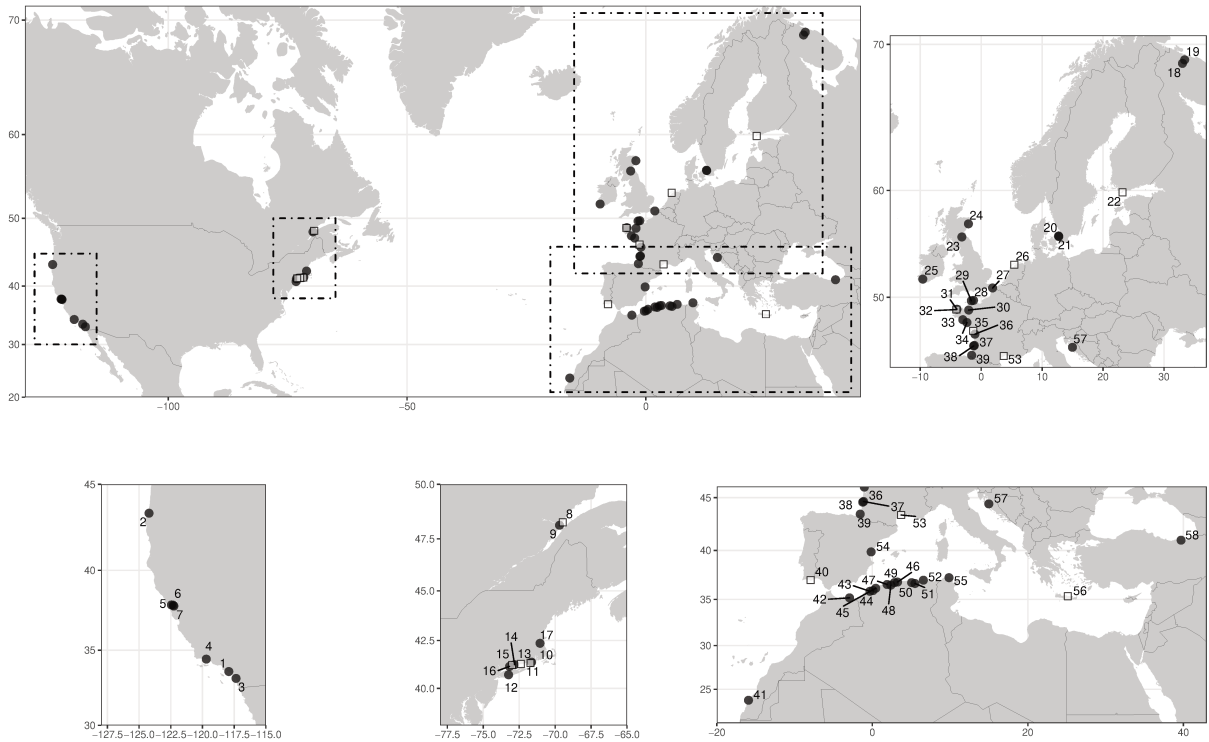


Figure 1: Maps of populations sampled. White squares: GBS. Black circles: BXP. 01-California-NEW, 02-Oregon-Coos, 03-California-Carlsbad, 04-SantaBarbara, 05-California-TBR, 06-California-BRK, 07-California-LAM, 08-StLawrence-CBD, 09-StLawrence-TAD, 10-Lgls-SHM, 11-Lgls-QNT, 12-Lgls-BYY, 13-Lgls-OST, 14-Lgls-BFM, 15-Lgls-MYC, 16-Lgls-SPM, 17-Boston, 18-NovyMost, 19-Retinskoye, 20-Oresund-Helsingborg, 21-Oresund-Raa, 22-Tvarminne, 23-Edinburgh, 24-Aberdeen, 25-Kenmare, 26-Wadden-sea, 27-Calais, 28-Barfleur, 29-Cherbourg, 30-Dinard, 31-Roscoff, 32-Guillec, 33-Kerbihan, 34-Pornichet, 35-Gascogne, 36-Biscay, 37-Aiguillon, 38-Arcachon, 39-Biarritz, 40-Faro, 41-Dahkla, 42-Nador, 43-Arzew, 44-Moustaganem, 45-SidiLakhdar, 46-EIMarsa, 47-Ghouraya, 48-Tipaza, 49-SidiFredj, 50-Bejaia, 51-Ziama, 52-Collo, 53-Thau, 54-Nules, 55-Bizerte, 56-Heraklion, 57-Croatia, 58-Trabzon

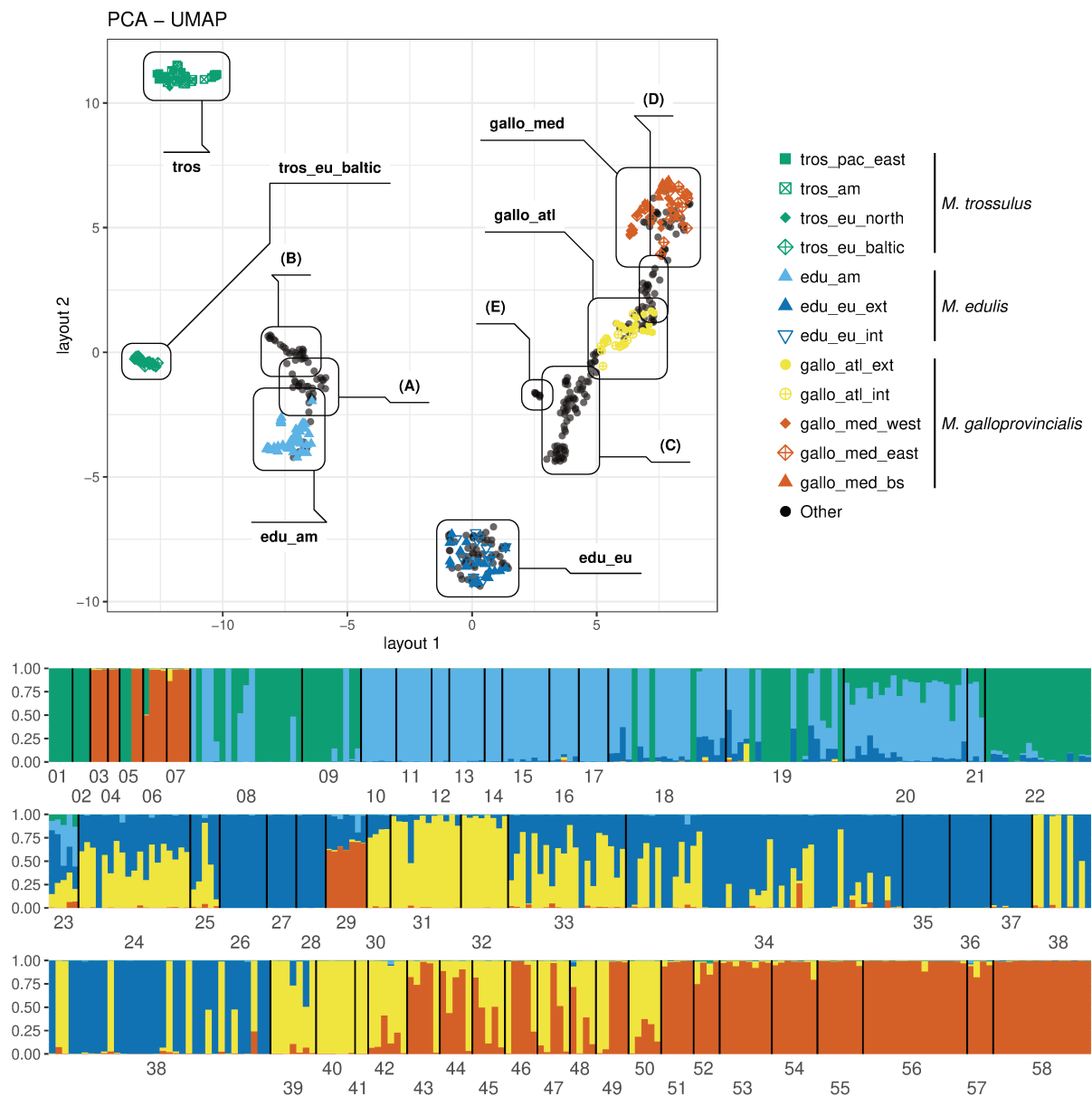


Figure 2: **top**: PCA-UMAP using the first 11 principal components. The reference level L3 (Table 1) is colour and shape coded. Note that this representation does not conserve distances and is designed to maximise groupings between similar entities, see Figure S2 for PCAs. Annotations show five groups of interest discussed in the main text: (A) North-European *M. edulis*; (B) Oresund hybrid zone; (C) Brittany, Aquitaine and Scotland hybrid zones (D) Algerian hybrid zone and (E) the port of Cherbourg. **bottom**: Ancestry composition of each individual in the dataset obtained with Structure for the major mode of $K=5$. Populations are mainly ordered geographically. See Figure 1 caption for the names associated with the population numbers. 01-07: American Pacific coast; 08-17: American Atlantic Coast; 18-19: North of Russia; 20-22: Oresund hybrid zone and Baltic *M. trossulus*; 23-24: Scotland hybrid zone; 25: Ireland 26-28: South-European *M. edulis* peripheral (ext); 29: port of Cherbourg; 30-32: *M. galloprovincialis* Atlantic local (int, North of Brittany); 33-34: Brittany hybrid zone; 35-37: South-European *M. edulis* local (int, Bay of Biscay); 38: Aquitaine hybrid zone; 39: *M. galloprovincialis* from Biarritz; 39-41: *M. galloprovincialis* Atlantic peripheral; 42-50: Algerian hybrid zone; 51-52: *M. galloprovincialis* Med local; 53-54: North-Mediterranean *M. galloprovincialis*; 55-57: *M. galloprovincialis* Mediterranean peripheral; 58: *M. galloprovincialis* Mediterranean from the Black Sea.

Because this study uses ancestry-informative SNPs, we lacked a neutral baseline required for outlier tests. Nevertheless, we confirmed a signal that is so strong as to make fully neutral interpretations unlikely.

The *M. edulis*-*M. galloprovincialis* mosaic hybrid zones (Figure 3) can be seen as a large scale intergradation between the Mediterranean populations (starting in the Black Sea) and the North-European ones. As mentioned above, *M. galloprovincialis* populations are more introgressed by *M. edulis* alleles in the Atlantic than the Mediterranean (see mean hybrid index, hi , in Figure 3A). In addition, the local *M. galloprovincialis* population in Brittany (gallo_atl_brit), surrounded by two patches of *M. edulis* on either side, has a higher level of introgression than the local population in Aquitaine (gallo_atl_aqui). Interestingly, the Black-Sea population (gallo_med_bs) displays a few fixed introgressed alleles from *M. edulis*, contrasting with the rest of the Mediterranean basin (Figure 3A and B6). The South-European *M. edulis*, both external and internal to the mosaic, have nearly as many comets of introgression as *M. galloprovincialis* (Figure 3A and B4); and around half of these extend to the Northern Europe.

For the *M. edulis*-*M. trossulus* comparison, we investigated two intergradations, in America and Europe (Figure 4). As with the *M. edulis*-*M. galloprovincialis* comparison, we observe comets of introgression in both continents. American and North European *M. edulis* populations share some comets (Figure 4A), and almost none are private to either population (though see markers 183 and 087 in Figure 4). Both populations also show introgression from *M. trossulus*, although with different outcomes. Whereas the Baltic *M. trossulus* has been highly introgressed (exhibiting a mean hybrid index of 0.15), American *M. trossulus* does not show strong introgression, except at a few comets.

3.3 Hybrid zones

The hybrid zones studied here, involve four pairs of lineages, as listed in Table 3. We considered three types of populations: (i) “central” populations, from the heart of the hybrid zone; (ii) local parental populations (P1 and P2 local) on each side of the hybrid zone and impacted by direct migration; and (iii) peripheral parental populations (P1 and P2 periph) which are not in direct contact with the focal hybrid zone and so potentially less affected by local introgression.

As a general trend, the local genomic clines (Figure 5, middle panels) exhibited good convergence of the models in the Bayesian analyses (\hat{R} close to 1, Vehtari et al., 2019) with a limited number of excess ancestry markers both for α and β . On the other hand, analyses with peripheral parental populations taken as reference (Figure 5, right panels) exhibited limited convergence between chains, consistent with there being strong distortion of the genomic clines.

Three hybrid zones, in Brittany, Aquitaine and Scotland, involve contact between South-European *M. edulis* and Atlantic *M. galloprovincialis*. The Brittany and Aquitaine zones are on

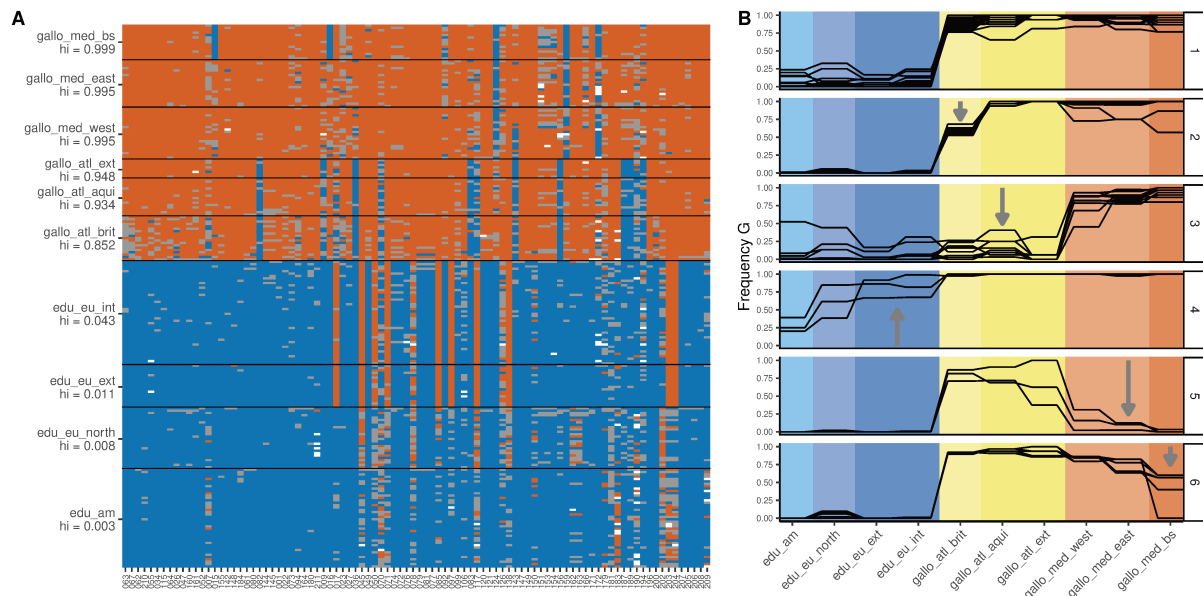


Figure 3: Visualisation of introgression between *M. galloprovincialis* and *M. edulis* the reference dataset. **(A)** Plot of the raw genotypes. Orange: homozygous *M. galloprovincialis*; blue: homozygous *M. edulis*; grey: heterozygous. The orientation of alleles and mean hybrid indexes for each population (hi) were computed with the R package *introgress*. The *edu_eu_north* population was taken as parent 1 (P1), and the *gallo_med_east* and *gallo_med_bs* populations were taken as parent 2 (P2). **(B)** Allele frequency of the *M. galloprovincialis* allele (G) for a few loci selected visually to exemplify six categories found in the dataset: 1) loci informative between *M. edulis* and *M. galloprovincialis*; 2) introgression only in the local *M. galloprovincialis* Atlantic of Brittany; 3) introgression in all *M. galloprovincialis* Atlantic populations; 4) introgression in *M. edulis* European populations; 5) potentially old introgression only in the *M. galloprovincialis* Mediterranean; 6) potentially old introgression only in the *M. galloprovincialis* from the Black Sea.

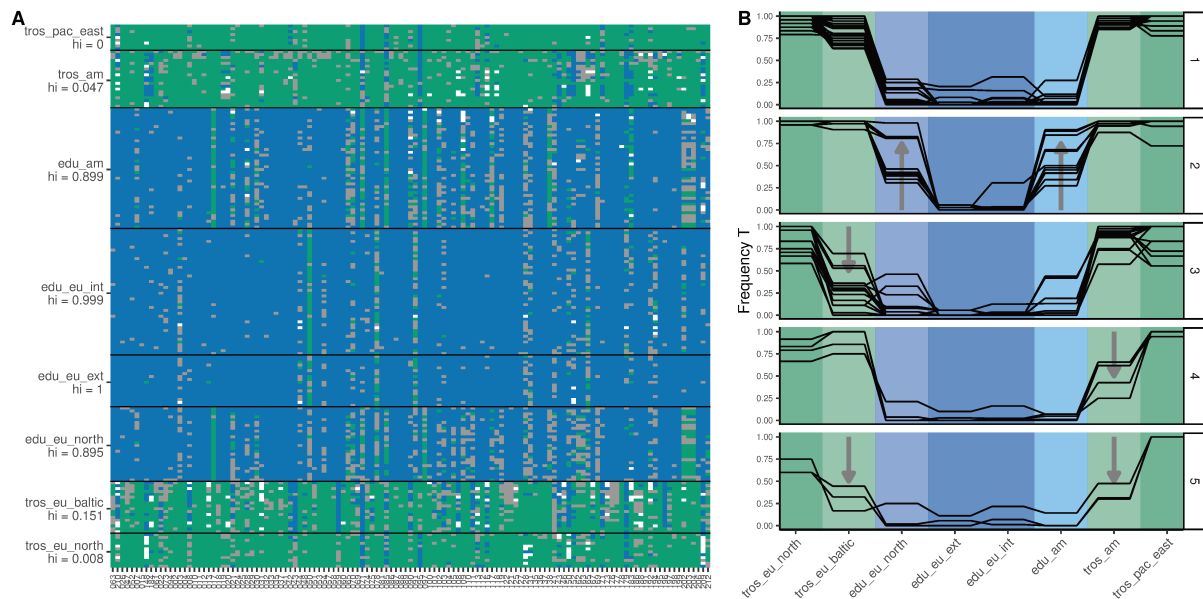


Figure 4: Visualisation of introgression between *M. trossulus* and *M. edulis* using the reference dataset. **(A)** Plot of the raw genotypes. Green: homozygous *M. trossulus*; blue: homozygous *M. edulis*; grey: heterozygous. The orientation of alleles and mean hybrid indexes for each population (hi) were computed with the R package *introgress*. The *tros_pac_east* and *tros_eu_north* populations were taken as parent 1 (P1), and the *edu_eu_ext* and *edu_eu_int* ones were taken as parent 2 (P2). **(B)** Allele frequency of the *M. trossulus* allele (T) for a few loci selected visually to exemplify five categories found in the dataset: 1) loci informative between *M. edulis* and *M. trossulus*; 2) introgression into *M. edulis* populations, either in both American and European or just one of them; 3) introgression into the Baltic *M. trossulus* population; 4) introgression into the East-American *M. trossulus* population; 5) introgression into East-American and Baltic *M. trossulus* populations, also showing a reduced introgression into the North-European *M. trossulus*.

one side of a patch of South-European *M. edulis* on the French Atlantic coast (Bierne, Borsa, et al., 2003). Nevertheless, these two zones exhibit strong differences in their hybrid index distributions, and this is reflected in their genomic clines (Figure 5). On one hand, Brittany presents a broad distribution of hybrid types between the two parental populations and a local *M. galloprovincialis* parental population that is introgressed (P2_local, see also Figure S6). On the other hand, Aquitaine presents only three F1 hybrids between the two parental types and the local *M. galloprovincialis* population (P2_local, including some individuals in sympatry with the *M. edulis* P1_local population) does not present introgression. In Scotland, no local *M. edulis* were sampled, but the central population exhibits variable hybrid indexes, and so resembles Brittany more than Aquitaine.

The Algerian hybrid zone was sampled most extensively (El Ayari et al., 2019). As previously shown with 4 markers (El Ayari et al., 2019), the distribution of hybrid indexes is wide, and the zone extends to around 600 km. While we did not sample Atlantic *M. galloprovincialis* populations close to this zone, no excess ancestry was found when using P1 peripheral, indicating that the peripheral and local populations may be close to each other in genetic composition.

The Øresund hybrid zone, includes the North-European *M. edulis* lineage (Figure 2, populations 20-21). Therefore, we treated such individuals from Russia as the P1 parental population. In this zone, the central population exhibits relatively homogeneous hybrid indexes, and is mainly composed of North-European *M. edulis* introgressed by *M. trossulus* (Figure 2). On the other side of the hybrid zone, the Baltic *M. trossulus* population is introgressed as shown above (Fraïsse, Belkhir, et al., 2016; Väinölä and Strelkov, 2011; Figures 4 and 5).

Finally, to highlight the intermediate character of the North-European *M. edulis* lineage, we carried out an analysis treating as admixed, the North-European *M. edulis* individuals in the Russian populations (18 and 19). As shown by the hybrid index and the Structure analyses, these mussels have an homogeneous admixture of around 60% South-European *M. edulis* and 40% American *M. edulis*.

When considering the correlation of genomic cline parameters (α and β) between hybrid zones, only two correlations proved statistically significant after correcting for multiple tests. The correlated parameters are α and β between the Scotland and Brittany local hybrid zone (Spearman correlation coefficients of 0.38 and 0.36, respectively, with p -values < 0.005), and β between the Brittany and Aquitaine hybrid zone (Spearman correlation coefficient of 0.38, p -value = 0.003).

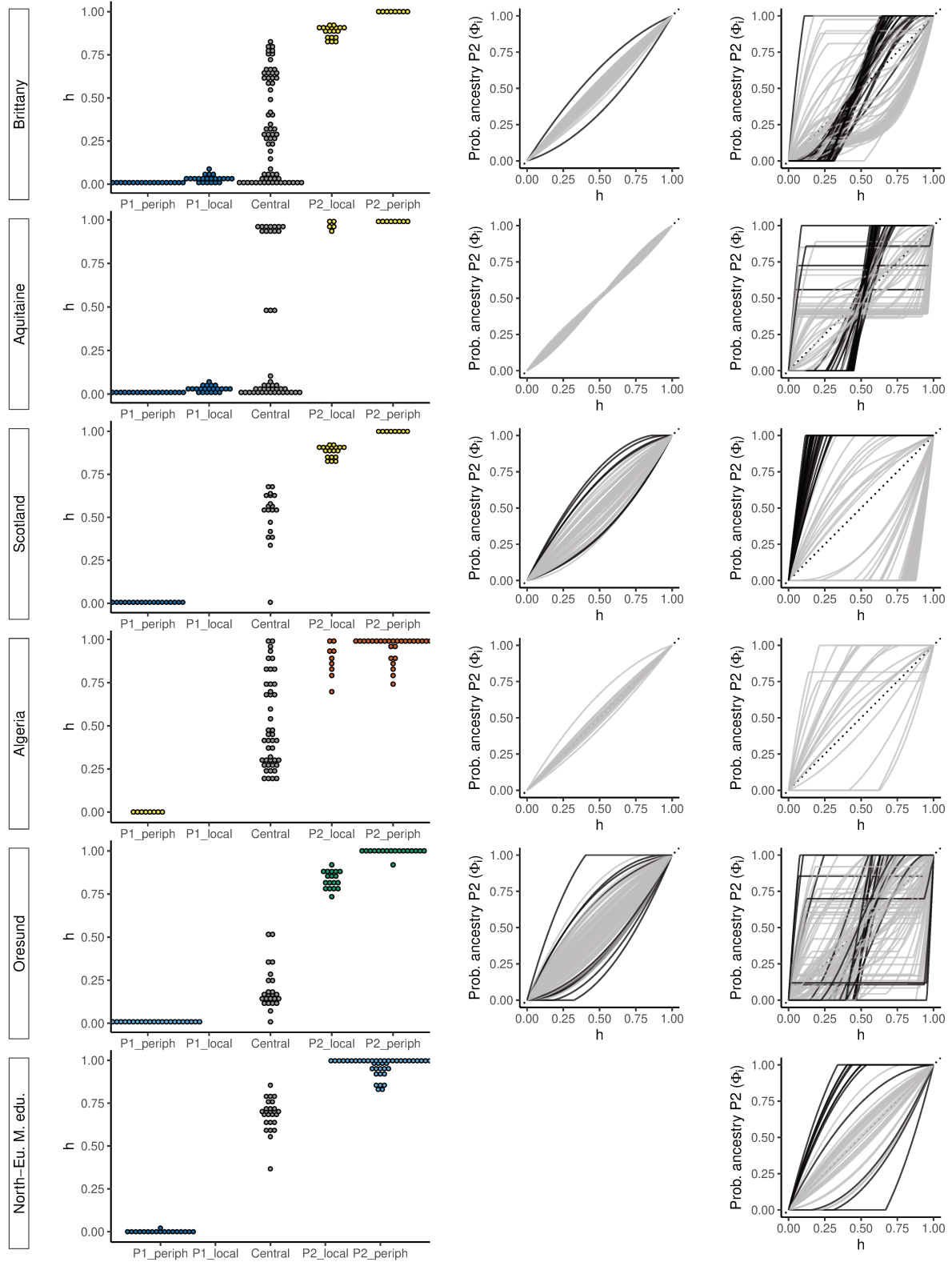


Figure 5: **Left panels:** Hybrid index distributions from introgress, for each group of the hybrid zone. **middle panels:** Genomic cline computed with the local parental groups (P1_local, P2_local or peripheral when not available) with cline presenting an excess in either α or β parameters highlighted in black. **right panels:** Genomic cline computed with external populations as parental groups and other groups as admixed. For genomic cline, only markers with an allele frequency difference > 0.3 between P1_periph and P2_periph are drawn.

4 Discussion

4.1 *Mytilus* mussels are genetically differentiated, but fixed differences are extremely rare

One aim of this study was to develop a panel of ancestry-informative markers for the *M. edulis* complex in the Northern Hemisphere. To this end, we started with 51,878 high-quality SNPs from 1269 contigs (Fraïsse, Belkhir, et al., 2016), and selected SNPs with the greatest discriminatory power. This procedure was successful, in that we were able to discriminate not only individuals of the three species (*M. edulis*, *M. galloprovincialis* and *M. trossulus*) but also partially isolated genetic lineages within species (European and American *M. edulis*; Atlantic and Mediterranean *M. galloprovincialis*; and Baltic and American/North-European *M. trossulus*).

Nevertheless, of 92 ancestry-informative loci between *M. edulis* and *M. galloprovincialis* (Figure 3), only one was a fixed difference. Similarly, there was only one fixed difference among 126 ancestry-informative loci between *M. edulis* and *M. trossulus* (Figure 4). All the other SNPs were found at least once in heterozygous state (grey squares in Figures 3 and 4) and even sometimes as heterospecific homozygotes.

Furthermore, it seems likely that further increasing the sample size would reveal shared polymorphisms even at the remaining two diagnostic markers. Similar failures of diagnostic loci in *Mytilus* have already arisen for the widely used Me15/16 marker (also known as Glu-5') where the "E" allele was observed in *M. galloprovincialis* (Bierne, Borsa, et al., 2003; Borsa, Daguin, Ramos Caetano, & Bonhomme, 1999; Hamer, Korlević, Durmiši, Nerlović, Bierne, et al., 2012; Wood, Beaumont, Skibinski, & Turner, 2003) and the "G" allele in *M. edulis* (Bierne, Borsa, et al., 2003; Kijewski, Wijsman, Hummel, & Wenne, 2009; Luttikhuisen, Koolhaas, Bol, & Piersma, 2002). In *Mytilus* mussels, therefore, it seems prudent to replace single marker diagnostics with multilocus inference; and our results suggest that 5-10 ancestry-informative loci should be sufficient for this purpose.

It is unlikely that this situation is unique to mussels. In the highly divergent *Ciona* tunicate species, *C. robusta* and *C. intestinalis*, loci initially assumed to be diagnostic and used to identify hybrids (Bouchemousse, Lévêque, Dubois, & Viard, 2016; Nydam & Harrison, 2011; Sato, Shimeld, & Bishop, 2014) were subsequently found to harbour shared polymorphisms in a multilocus analysis, and heterozygous genotypes were found in parental populations, both at the initially studied markers and at many of the newly designed markers (Bouchemousse, Liautard-Haag, et al., 2016).

4.2 Local introgression is widespread and has several possible causes

Because our SNPs showed extreme levels of differentiation, they provide a window on patterns of introgression in *Mytilus*. At the broadest level, our study confirms, with greater sampling, the findings of Fraïsse, Belkhir, et al. (2016), that introgression is pervasive in the

complex. In the three species investigated, at least one population or lineage was impacted at some point by introgression. We have also shown that introgression is highly heterogeneous across the genome, with “comets” of heterospecific genotypes at some loci, while others resist introgression altogether (Figures 3 and 4).

The biogeography of the complex also allowed us to study introgressions of a very particular kind. In particular, the *M. galloprovincialis*/*M. edulis* transition is characterised by near-continuous intergradation between the Black Sea and Scandinavia, but with genetic barriers to gene flow at multiple points. This leads to a mosaic distribution in several regions. For example, the isolated *M. edulis* and *M. galloprovincialis* patches on the Atlantic coast of France, are separated by three hybrid zones, in Aquitaine, South Brittany, and Normandy (not sampled in this study) (Bierne, Borsa, et al., 2003; Hilbish et al., 2012). This structure revealed multiple instances of local comets of introgression, that is, introgression events that are localised both genomically, and geographically, with allelic variants crossing one barrier to gene flow, but halted at a subsequent barrier (Figure 3).

Three possible mechanisms could explain these local comets of introgression. First, the introgression could be adaptive (Hedrick, 2013; Pardo-Diaz et al., 2012; Staubach et al., 2012). In this case, the fact that introgressions were halted at a subsequent barrier could be explained either by an environmental difference (such that the allele was only locally beneficial), or by a very strong barrier (in which case the halting would be transient). Alternatively, if the markers are not the direct target of selection – as seems highly plausible (Fraïsse, Belkhir, et al., 2016) – a marker might hitchhike through one barrier, but be halted at a second (N. H. Barton, 2000; Faure et al., 2008). A second possibility is that the introgressions act to reduce genetic load in the recipient population (Kim, Huber, & Lohmueller, 2018). This scenario is similar to adaptive introgression, but emphasises the role of deleterious mutations in the recipient population rather than advantageous mutations in the donor population. Third, and finally, introgressions could involve bi-stable variants in a tension zone (Nicholas H. Barton & Hewitt, 1985). Such variants can move via an asymmetry in parental fitness, while being easily trapped by a density trough or a barrier to dispersion (see Nicholas H. Barton and Turelli, 2011, and a possible example in *Mytilus* in El Ayari et al., 2019). This scenario also implies transience in the structure, because tension zones can move due to genetic drift or changes in the environmental conditions (Piálek & Barton, 1997).

4.3 High overall concordance of barrier loci at hybrid zones

Our analyses of genomic clines, with local parental populations, show consistently high concordance between markers (Figure 5 middle panels). Indeed, the anomalous “excess ancestry” loci observed in Scotland and Øresund, while they could be local comets of introgression, are potentially due to the fact that the P1_local populations were not sampled.

The high levels of concordance provides a striking contrast to results when more distant,

peripheral populations were treated as the parental reference (compare Figure 5 middle and right panels). With distant populations, the proportion of locally introgressed loci was so high, that it became difficult to estimate a "baseline" distribution. The very failure of these analyses further highlights the rampant local introgression in this system.

Even with the local parental populations, we observed a few SNPs that deviated from the genomic average. These loci tend to introgress more than the others in the centre of the zone, but – unlike true comets of introgression – have not escaped the genetic barrier. We suspect that these loci might contribute to bi-stable variation that is asymmetric – i.e. one parental genotype is fitter than the other, but both are fitter than hybrids (Nicholas H. Barton, 1979a; Nicholas H. Barton & Turelli, 2011). Such variation can be trapped by genetic or migration barriers for long periods, but rare events of strong demographic stochasticity can free the "pushed" wave of advance (Piálek & Barton, 1997). Genomic cline outliers with asymmetric introgression are often interpreted as evidence of adaptive introgression, i.e. directional selection (Gompert & Buerkle, 2009). Our results support directional selection in the heart of the hybrid zones, but not in flanking parental populations. If a force is opposing the spread of the adaptive variant, it could be that selection against heterozygous or recombinant hybrid genotypes pushes in the opposite direction to selection for the fitter parental genotype, trapping the variant at the barrier (Nicholas H. Barton, 1979b).

4.4 The timing and context of introgressions

Given the heterogeneity in the introgression patterns we have observed, it seems likely that they were shaped by both contemporary contacts, and historic contacts during the Quaternary period (Hewitt, 2000).

In some cases our data allow us to make inferences about the timing of specific introgression events. For example, the local introgression of *M. edulis* alleles into *M. galloprovincialis* from the Black Sea and Mediterranean, is consistent with ancient contacts between these populations (Fraïsse, Belkhir, et al., 2016; Gosset & Bierne, 2013). The two private introgressions into the Black sea were nevertheless unexpected (Fraïsse, Belkhir, et al. (2016) did not sample these populations, and so our data set was not enriched for discriminatory SNPs), and so such introgressions might be quite numerous in the rest of the genome.

By contrast with these putatively ancient introgressions, it is likely that the private introgressions found in *M. galloprovincialis* from Brittany (Figure 3B2) are relatively recent. This is because, given its position in the mosaic, it is probable that this population only became established during the last post-glacial period.

More recently still, it is likely that the admixed individuals between Mediterranean *M. galloprovincialis* and South-European *M. edulis*, observed in the port of Cherbourg, represent a recent human-mediated introduction. Similar admixed individuals have been observed in other ports in the English Channel and French Atlantic coast (Simon et al., 2019), where these

“dock mussels” form small-scale hybrid zones with the native lineages outside of the port.

In one final case, shared introgressions allow us to make inferences about the historical biogeography of *Mytilus*. In particular, we found that several introgressions from *M. trossulus* to *M. edulis* were shared between *M. edulis* populations in Northern-Europe and America, and the most parsimonious explanation is that these introgressions predated their split. This hypothesis sheds light on the origins of the North-European *M. edulis* population. We have shown that this population is differentiated from both American, and South-European *M. edulis*, and appears as intermediate in the PCA (Figure S2), Structure (Figures 2 and S10) and hybrid index analyses (Figure 5). Given its presence as the parental *M. edulis* population in the Øresund hybrid zone, and the complete absence of American ancestry in the Netherlands, the border between Southern and Northern European *M. edulis* probably falls somewhere near to the Danish coast (see also the previously observed differences between North Sea and populations North of the Kattegat region (Bierne, Daguin, Bonhomme, David, & Borsa, 2003; Stuckas, Stoof, Quesada, & Tiedemann, 2009)). Given the shared introgressions, therefore, the contemporary biogeography could reflect a recolonisation of America after the last glacial maxima, by a proto-North-European *M. edulis*, which was later introgressed by the South-European *M. edulis* (Wares & Cunningham, 2001). Alternatively, American *M. edulis*, having survived in a refugia (Riginos & Henzler, 2008), might have colonised the North-Atlantic and Scandinavia, and then been introgressed by the South-European *M. edulis* in Europe.

5 Conclusion

We developed an ancestry-informative SNP panel powerful enough to classify species and divergent lineages of the *M. edulis* species complex. Following an extended sampling of reference populations, known hybrid zones and populations close to them, we confirm that local introgression is very widespread in the complex. Indeed, most of the markers that are highly differentiated between species pairs, are nonetheless segregating in both. The investigation of genomic clines in hybrid zones shows that loci are consistently concordant at a local scale, although a few asymmetric bi-stable variants might still be trapped by the genetic barrier. Overall our results suggest that asymmetrical parental fitness differences may enhance introgression at some regions of the genome yet successive barriers can prevent or delay propagation.

Acknowledgements: Data used in this work were partly produced through the genotyping and sequencing facilities of ISEM and LabEx CeMEB, an ANR “Investissements d’avenir” program (ANR-10-LABX-04-01) This project benefited from the Montpellier Bioinformatics Biodiversity platform supported by the LabEx CeMEB. We thank Norah Saarman, Grant Pogson, Célia Gosset and Pierre-Alexandre Gagnaire for providing samples. This work was funded by a Languedoc-Roussillon “Chercheur(se)s d’Avenir” grant (Connect7 project). P. Strelkov was supported by the Russian Science Foundation project 19-74-20024. This is article 2020-XXX

of Institut des Sciences de l'Evolution de Montpellier.

References

- Abbott, R. J., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., ... Zinner, D. (2013). Hybridization and speciation. *J. Evol. Biol.*, *26*(2), 229–246. doi:10.1111/j.1420-9101.2012.02599.x
- Aeschbacher, S., Selby, J. P., Willis, J. H., & Coop, G. (2017). Population-genomic inference of the strength and timing of selection against gene flow. *Proc. Natl. Acad. Sci.*, *114*(27), 7061–7066. doi:10.1073/pnas.1616755114
- Barton, N. H. [N. H.]. (2000). Genetic hitchhiking. *Phil. Trans. R. Soc. Lond. B*, *355*(1403), 1553–1562. doi:10.1098/rstb.2000.0716
- Barton, N. H. [Nicholas H.]. (1979a). Gene flow past a cline. *Heredity*, *43*(3), 333–339. doi:10.1038/hdy.1979.86
- Barton, N. H. [Nicholas H.]. (1979b). The dynamics of hybrid zones. *Heredity*, *43*(3), 341–359. doi:10.1038/hdy.1979.87
- Barton, N. H. [Nicholas H.], & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity*, *56*, 357–376. doi:10.1038/hdy.1986.135
- Barton, N. H. [Nicholas H.], & Hewitt, G. M. (1985). Analysis of Hybrid Zones. *Annu. Rev. Ecol. Evol. Syst.*, *16*, 113–148. doi:10.1146/annurev.es.16.110185.000553
- Barton, N. H. [Nicholas H.], & Turelli, M. (2011). Spatial waves of advance with bistable dynamics: Cytoplasmic and genetic analogues of allee effects. *Am. Nat.*, *178*(3), E48–E75. doi:10.1086/661246
- Benjamini, Y., & Yekutieli, D. (2001). The Control of the False Discovery Rate in Multiple Testing Under Dependency. *Ann. Stat.*, *29*(4), 1165–1188. doi:10.1214/aos/1013699998
- Bierne, N. (2010). The Distinctive Footprints of Local Hitchhiking in a Varied Environment and Global Hitchhiking in a Subdivided Population: The Distinctive Footprints of Local and Global Hitchhiking. *Evolution*, *64*(11), 3254–3272. doi:10.1111/j.1558-5646.2010.01050.x
- Bierne, N., Bonhomme, F., Boudry, P., Szulkin, M., & David, P. (2006). Fitness landscapes support the dominance theory of post-zygotic isolation in the mussels *Mytilus edulis* and *M. galloprovincialis*. *Proc. R. Soc. B*, *273*, 1253–1260. doi:10.1098/rspb.2005.3440
- Bierne, N., Borsa, P., Daguin, C., Jollivet, D., Viard, F., Bonhomme, F., & David, P. (2003). Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. *Mol. Ecol.*, *12*(2), 447–461. doi:10.1046/j.1365-294X.2003.01730.x
- 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. *Mol. Ecol.*, *12*(9), 2505–2510. doi:10.1046/j.1365-294X.2003.01936.x

-
- 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*(2), 292–298. doi:10.1111/j.0014-3820.2002.tb01339.x
- Boon, E., Faure, M. F., & Bierne, N. (2009). The flow of antimicrobial peptide genes through a genetic barrier between *Mytilus edulis* and *M. galloprovincialis*. *J. Mol. Evol.*, *68*(5), 461–474. doi:10.1007/s00239-009-9211-z
- Borsa, P., Daguin, C., Ramos Caetano, S., & Bonhomme, F. (1999). Nuclear-DNA evidence that northeastern Atlantic *Mytilus trossulus* mussels carry *M. edulis* genes. *J. Molluscan Stud.*, *65*(4), 504–507. doi:10.1093/mollus/65.4.504
- Bouchemousse, S., Lévêque, L., Dubois, G., & Viard, F. (2016). Co-occurrence and reproductive synchrony do not ensure hybridization between an alien tunicate and its interfertile native congener. *Evol Ecol*, *30*(1), 69–87. doi:10.1007/s10682-015-9788-1
- Bouchemousse, S., Liautard-Haag, C., Bierne, N., & Viard, F. (2016). Distinguishing contemporary hybridization from past introgression with postgenomic ancestry-informative SNPs in strongly differentiated *Ciona* species. *Mol Ecol*, *25*(21), 5527–5542. doi:10.1111/mec.13854
- Chaturvedi, S., Lucas, L., Buerkle, A., Fordyce, J., Forister, M., Nice, C. C., & Gompert, Z. (2019). Recent hybrids recapitulate ancient hybrid outcomes. *bioRxiv*. doi:10.1101/769901
- Diaz-Papkovich, A., Anderson-Trocme, L., & Gravel, S. (2018). Revealing multi-scale population structure in large cohorts. *bioRxiv*. doi:10.1101/423632
- Duranton, M., Allal, F., Fraïsse, C., Bierne, N., Bonhomme, F., & Gagnaire, P.-A. (2018). The origin and remolding of genomic islands of differentiation in the European sea bass. *Nat. Commun.*, *9*, 2518. doi:10.1038/s41467-018-04963-6
- El Ayari, T., Trigui El Menif, N., Hamer, B., Cahill, A. E., & Bierne, N. (2019). The hidden side of a major marine biogeographic boundary: A wide mosaic hybrid zone at the Atlantic–Mediterranean divide reveals the complex interaction between natural and genetic barriers in mussels. *Heredity*, *122*, 770–784. doi:10.1038/s41437-018-0174-y
- Faure, M. F., David, P., Bonhomme, F., & Bierne, N. (2008). Genetic hitchhiking in a subdivided population of *Mytilus edulis*. *BMC Evolutionary Biology*, *8*(1), 164. doi:10.1186/1471-2148-8-164
- Fischer, B., Pau, G., & Smith, M. (2019). Rhdf5: HDF5 interface to R (Version 2.26.2).
- Fraïsse, C., Belkhir, K., Welch, J. J., & Bierne, N. (2016). Local interspecies introgression is the main cause of extreme levels of intraspecific differentiation in mussels. *Mol. Ecol.*, *25*(1), 269–286. doi:10.1111/mec.13299
- Fraïsse, C., Gunnarsson, P. A., Roze, D., Bierne, N., & Welch, J. J. (2016). The genetics of speciation: Insights from Fisher's geometric model. *Evolution*, *70*(7), 1450–1464. doi:10.1111/evo.12968

- Gagnaire, P.-A., Lamy, J.-B., Cornette, F., Heurtebise, S., Dégremont, L., Flahauw, E., ... Lapègue, S. (2018). Analysis of Genome-Wide Differentiation between Native and Introduced Populations of the Cupped Oysters *Crassostrea gigas* and *Crassostrea angulata*. *Genome Biol Evol*, *10*(9), 2518–2534. doi:10.1093/gbe/evy194
- Gompert, Z., & Buerkle, C. A. [C. A.]. (2010). Introgress: A software package for mapping components of isolation in hybrids. *Mol. Ecol. Resour.*, *10*(2), 378–384. doi:10.1111/j.1755-0998.2009.02733.x
- Gompert, Z., & Buerkle, C. A. [C. A.]. (2011). Bayesian estimation of genomic clines. *Mol. Ecol.*, *20*(10), 2111–2127. doi:10.1111/j.1365-294X.2011.05074.x
- Gompert, Z., & Buerkle, C. A. [C. A.]. (2012). Bgc : Software for Bayesian estimation of genomic clines. *Mol. Ecol. Resour.*, *12*(6), 1168–1176. doi:10.1111/1755-0998.12009.x
- Gompert, Z., & Buerkle, C. A. [C. Alex]. (2009). A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Mol. Ecol.*, *18*(6), 1207–1224. doi:10.1111/j.1365-294X.2009.04098.x
- Gosset, C. C., & Bierne, N. (2013). Differential introgression from a sister species explains high F_{ST} outlier loci within a mussel species. *J. Evol. Biol.*, *26*(1), 14–26. doi:10.1111/jeb.12046
- Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., ... Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science*, *328*(5979), 710–22. doi:10.1126/science.1188021
- Hamer, B., Korlević, M., Durmiši, E., Nerlović, V., Bierne, N., et al. (2012). Nuclear marker Me 15/16 analyses of *Mytilus galloprovincialis* populations along the eastern Adriatic coast. *Cah. Biol. Mar.*, *53*(1), 35–44.
- Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Mol. Ecol.*, *25*(11), 2454–2466. doi:10.1111/mec.13582
- Hedrick, P. W. (2013). Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol Ecol*, *22*(18), 4606–4618. doi:10.1111/mec.12415
- Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*(6789), 907–913. doi:10.1038/35016000
- Hilbish, T. J., Lima, F. P., Brannock, P. M., Fly, E. K., Rognstad, R. L., & Wethey, D. S. (2012). Change and stasis in marine hybrid zones in response to climate warming. *J. Biogeogr.*, *39*(4), 676–687. doi:10.1111/j.1365-2699.2011.02633.x
- Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405. doi:10.1093/bioinformatics/btn129
- Kijewski, T., Wijsman, J. W., Hummel, H., & Wenne, R. (2009). Genetic composition of cultured and wild mussels *Mytilus* from The Netherlands and transfers from Ireland and Great Britain. *Aquaculture*, *287*(3-4), 292–296. doi:10.1016/j.aquaculture.2008.10.048

-
- Kim, B. Y., Huber, C. D., & Lohmueller, K. E. (2018). Deleterious variation shapes the genomic landscape of introgression. *PLoS Genet*, *14*(10), e1007741. doi:10.1371/journal.pgen.1007741
- Konopka, T. (2019). *Umap: Uniform Manifold Approximation and Projection*.
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.*, *15*(5), 1179–1191. doi:10.1111/1755-0998.12387
- Luttikhuisen, P. C., Koolhaas, A., Bol, A., & Piersma, T. (2002). Mytilus galloprovincialis-type foot-protein-1 alleles occur at low frequency among mussels in the Dutch Wadden Sea. *Journal of Sea Research*, *48*(3), 241–245. doi:10.1016/S1385-1101(02)00168-5
- Martin, S. H., Davey, J. W., & Jiggins, C. D. (2015). Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. *Mol. Biol. Evol.*, *32*(1), 244–257. doi:10.1093/molbev/msu269
- Martin, S. H., Davey, J. W., Salazar, C., & Jiggins, C. D. (2019). Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biol*, *17*(2), e2006288. doi:10.1371/journal.pbio.2006288
- McInnes, L., & Healy, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.
- Murgarella, M., Puiu, D., Novoa, B., Figueras, A., Posada, D., & Canchaya, C. (2016). A First Insight into the Genome of the Filter-Feeder Mussel *Mytilus galloprovincialis*. *Plos One*, *11*(3), e0151561–e0151561. doi:10.1371/journal.pone.0151561
- Nydam, M. L., & Harrison, R. G. (2011). Introgression despite substantial divergence in a broadcast spawning marine invertebrate. *Evolution*, *65*(2), 429–442. doi:10.1111/j.1558-5646.2010.01153.x
- Pardo-Diaz, C., Salazar, C., Baxter, S. W., Merot, C., Figueiredo-Ready, W., Joron, M., ... Jiggins, C. D. (2012). Adaptive Introgression across Species Boundaries in Heliconius Butterflies. *PLoS Genet*, *8*(6), e1002752. doi:10.1371/journal.pgen.1002752
- Piálek, J., & Barton, N. H. [Nicholas H.]. (1997). The Spread of an Advantageous Allele Across a Barrier: The Effects of Random Drift and Selection Against Heterozygotes. *Genetics*, *145*(2), 493–504.
- Quesada, H., Zapata, C., & Alvarez, G. (1995). A multilocus allozyme discontinuity in the mussel *Mytilus galloprovincialis*: The interaction of ecological and life-history factors. *Mar. Ecol. Prog. Ser.*, *116*, 99–115.
- Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., ... Westram, A. M. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *J. Evol. Biol.*, *30*(8), 1450–1477. doi:10.1111/jeb.13047

- Riginos, C., & Cunningham, C. W. (2005). Local adaptation and species segregation in two mussel (*Mytilus edulis* × *Mytilus trossulus*) hybrid zones. *Mol. Ecol.*, *14*, 381–400. doi:10.1111/j.1365-294X.2004.02379.x
- Riginos, C., & Henzler, C. M. (2008). Patterns of mtDNA diversity in North Atlantic populations of the mussel *Mytilus edulis*. *Mar. Biol.*, *155*(4), 399–412. doi:10.1007/s00227-008-1038-4
- Riquet, F., Le Cam, S., Fonteneau, E., & Viard, F. (2016). Moderate genetic drift is driven by extreme recruitment events in the invasive mollusk *Crepidula fornicata*. *Heredity*, 1–9. doi:10.1038/hdy.2016.24
- Roesti, M., Moser, D., & Berner, D. (2013). Recombination in the threespine stickleback genome—patterns and consequences. *Mol. Ecol.*, *22*(11), 3014–3027. doi:10.1111/mec.12322
- Saarman, N. P., & Pogson, G. H. (2015). Introgression between invasive and native blue mussels (genus *Mytilus*) in the central California hybrid zone. *Mol. Ecol.*, *24*(18), 4723–4738. doi:10.1111/mec.13340
- Sato, A., Shimeld, S. M., & Bishop, J. D. D. (2014). Symmetrical Reproductive Compatibility of Two Species in the *Ciona intestinalis* (Ascidiacea) Species Complex, a Model for Marine Genomics and Developmental Biology. *Zoological Science*, *31*(6), 369. doi:10.2108/zs130249
- Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., ... Przeworski, M. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, *360*(6389), 656–660. doi:10.1126/science.aar3684
- Simon, A., Arbiol, C., Nielsen, E. E., Couteau, J., Sussarellu, R., Burgeot, T., ... Bierne, N. (2019). Replicated anthropogenic hybridisations reveal parallel patterns of admixture in marine mussels. *Evol. Appl.* doi:10.1111/eva.12879
- Simon, A., Bierne, N., & Welch, J. J. (2018). Coadapted genomes and selection on hybrids: Fisher's geometric model explains a variety of empirical patterns. *Evol. Lett.*, *2*(5), 472–498. doi:10.1002/evl3.66
- Staubach, F., Lorenc, A., Messer, P. W., Tang, K., Petrov, D. A., & Tautz, D. (2012). Genome Patterns of Selection and Introgression of Haplotypes in Natural Populations of the House Mouse (*Mus musculus*). *PLoS Genet.*, *8*(8), e1002891. doi:10.1371/journal.pgen.1002891
- Strelkov, P., Katolikova, M., & Väinölä, R. (2017). Temporal change of the Baltic Sea–North Sea blue mussel hybrid zone over two decades. *Mar. Biol.*, *164*(11). doi:10.1007/s00227-017-3249-z
- Stuckas, H., Stoof, K., Quesada, H., & Tiedemann, R. (2009). Evolutionary implications of discordant clines across the Baltic *Mytilus* hybrid zone (*Mytilus edulis* and *Mytilus trossulus*). *Heredity*, *103*(2), 146–156. doi:10.1038/hdy.2009.37

-
- Väinölä, R., & Strelkov, P. (2011). *Mytilus trossulus* in Northern Europe. *Mar. Biol.*, *158*(4), 817–833. doi:10.1007/s00227-010-1609-z
- Varvio, S.-L., Koehn, R. K., & Väinölä, R. (1988). Evolutionary genetics of the *Mytilus edulis* complex in the North Atlantic region. *Mar. Biol.*, *98*(1), 51–60. doi:10.1007/BF00392658
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2019). Rank-normalization, folding, and localization: An improved R for assessing convergence of MCMC. *arXiv*.
- Wares, J. P., & Cunningham, C. W. (2001). Phylogeography and Historical Ecology of the North Atlantic Intertidal. *Evolution*, *55*(12), 2455–2469. doi:10.1111/j.0014-3820.2001.tb00760.x
- Wood, A. R., Beaumont, A. R., Skibinski, D. O. F., & Turner, G. (2003). Analysis of a nuclear-DNA marker for species identification of adults and larvae in the *Mytilus edulis* complex. *J. Molluscan Stud.*, *69*(1), 61–66. doi:10.1093/mollus/69.1.61

Supplementary information

Table S1: Sampling information per population. pop: population name in the dataset; N: number of individuals genotyped from this population; lat: latitude (in decimal degrees); long: longitude (in decimal degrees); locality: details on the locality.

pop	N	lat	long	locality
01-California-NEW	4	33.609786	-117.92158	Newport Harbor, California, USA
02-Oregon-Coos	3	43.366	-124.217	Coos Bay, Oregon, USA
03-California-Carlsbad	3	33.156302	-117.352779	Carlsbad, California, USA
04-SantaBarbara	2	34.42	-119.698	Santa Barbara, California, USA
05-California-TBR	4	37.872228	-122.451734	Tiburon Point, SF, California, USA
06-California-BRK	4	37.859163	-122.308652	Berkeley Marina, SF, California, USA
07-California-LAM	4	37.802713	-122.257995	Lake Merritt, SF, California, USA
08-StLawrence-CBD	19	48.269414	-69.466146	Cap de Bon Désir, Québec, Canada
09-StLawrence-TAD	10	48.134413	-69.696141	Tadoussac, Québec, Canada
10-Lgls-SHM	6	41.386917	-71.519167	Snug Harbor Marina, Charlestown, Long Island, USA
11-Lgls-QNT	6	41.332694	-71.713333	Quonochontaug, Long Island, USA
12-Lgls-BYY	3	41.346694	-71.967583	Brewer Yacht Yard, Mystic River, Long Island, USA
13-Lgls-OST	6	41.263278	-72.384194	Old Saybrook Town, Hartlands Drive, Long Island, USA
14-Lgls-BFM	3	41.260556	-72.820722	Branford Harbor, Long Island, USA
15-Lgls-MYC	8	41.210917	-73.051528	Milford Yacht Club, Long Island, USA
16-Lgls-SPM	5	41.107222	-73.354778	Southport Marina, Long Island, USA
17-Boston	5	42.36	-71.05	Boston, USA
18-NovyMost	20	68.904723	33.026324	Novy Most, Russia
19-Retinskoye	20	69.114088	33.387851	Retinskoye, Russia
20-Oresund-Helsinborg	21	56.04	12.694	Helsinborg, Oresund, Sweden
21-Oresund-Raa	3	55.999	12.74	Raa, Oresund, Sweden
22-Tvarminne	18	59.841	23.201	Tvarminne, Finland
23-Edinburgh	5	55.95	-3.18	Edinburgh, Scotland, UK
24-Aberdeen	19	57.14	-2.094	Aberdeen, Scotland, UK
25-Kenmare	5	51.88	-9.583	Kenmare, Ireland
26-Wadden-sea	9	53.31	5.424	Wadden Sea, The Netherlands
27-Calais	5	50.974978	1.870983	Calais, France
28-Barfleur	5	49.670616	-1.263	Barfleur, France

Continued on next page

Table S1 – *Continued from previous page*

pop	N	lat	long	locality
29-Cherbourg	7	49.63942	-1.61994	Cherbourg, France
30-Dinard	4	48.633	-2.05	Dinard, France
31-Roscoff	12	48.72	-3.98	Roscoff, France
32-Guillec	8	48.687003	-4.071535	Guillec, France
33-Kerbihan	20	47.576883	-3.021033	Kerbihan, France
34-Pornichet	47	47.26	-2.34	Pornichet, France
35-Gascogne	8	46.325	-1.312	Gulf of Gascogne, France
36-Biscay	7	45.95	-1.032	Bay of Biscay, France
37-Aiguillon	7	44.65	-1.13	Aiguillon, France
38-Arcachon	44	44.589	-1.239	Arcachon, France
39-Biarritz	7	43.483	-1.558	Biarritz, France
40-Faro	7	37.01	-7.93	Faro, Portugal
41-Dahkla	2	23.72	-15.93	Dahkla, Morocco
42-Nador	6	35.166	-2.93	Nador, Morocco
43-Arzew	5	35.85	-0.316	Arzew, Algeria
44-Moustaganem	5	35.93	0.083	Moustaganem, Algeria
45-SidiLakhdar	5	36.165	0.439	Sidi Lakhdar, Algeria
46-ElMarsa	5	36.81	3.244	El Marsa, Algeria
47-Ghouraya	5	36.56	1.899	Ghouraya, Algeria
48-Tipaza	5	36.49	2.381	Tipaza, Algeria
49-SidiFredj	5	36.75	2.846	Sidi Fredj, Algeria
50-Bejaia	5	36.75	5.0567	Bejaia, Algeria
51-Ziama	5	36.66	5.483	Ziama, Algeria
52-Collo	5	37	6.56	Collo, Algeria
53-Thau	8	43.4	3.7	Thau, France
54-Nules	7	39.853	-0.155	Nules, Spain
55-Bizerte	7	37.26	9.86	Bizerte, Tunisia
56-Heraklion	16	35.338	25.1442	Heraklion, Greece
57-Croatia	4	44.43	14.979	Croatia
58-Trabzon	15	41	39.71	Trabzon, Turkey

Table S2: Linear models of F_{ST} comparisons between GBS and both methods combined.

Level	Comparison	slope (p-value)	intercept (p-value)	adjusted r^2	p-value model
L1	edu/gallo	0.93**	2.74e-2* (4.28e-3)	0.96	**
	edu/tro	0.95**	2.07e-2 (7.47e-2)	0.95	**
	gallo/tro	0.93**	3.40e-2* (1.72e-2)	0.92	**
L2	tro am/eu	0.48**	4.74e-2* (5.04e-4)	0.66	**
	edu am/eu	0.85**	1.27e-2 (0.33)	0.85	**
	gallo atl/med	0.90**	1.04e-3* (7.2e-06)	0.87	**
L3	edu eu ext/int	0.46**	1.54e-2 (9.73e-2)	0.59	**
	gallo atl ext/int	0.46**	2.65e-2 (4.84e-3)	0.45	**
	gallo med east/west	0.23**	1.46e-2* (2.28e-2)	0.47	**
	edu intra am	0.39**	-5.94e-3 (0.45)	0.71	**

** : significant with p -value $< 2.2e-16$; * : significant with indicated p -value.

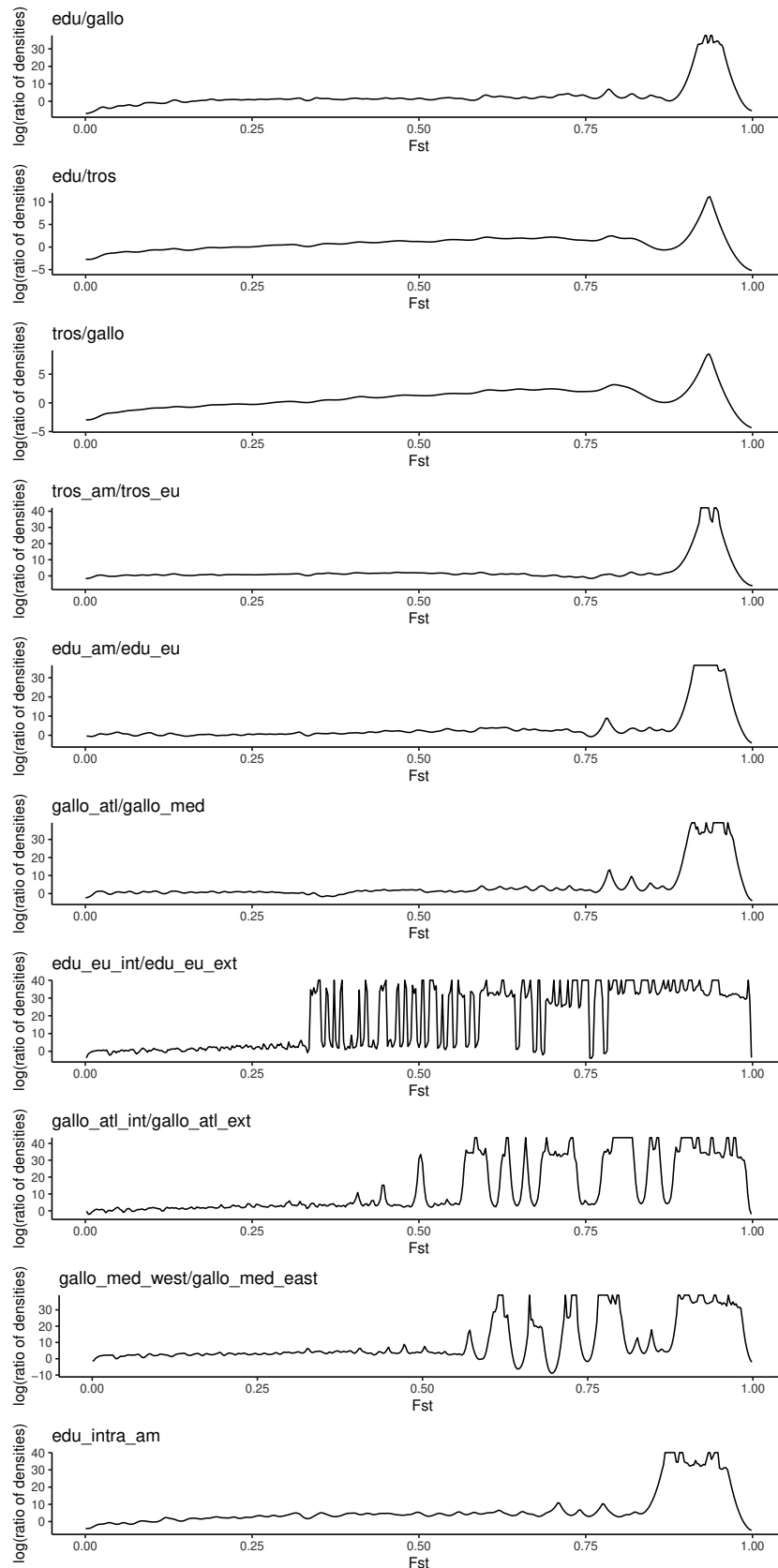


Figure S1: Enrichment in high F_{ST} markers for all comparisons. y-axis: natural logarithm of the ratio of F_{ST} densities between the ancestry informative dataset (present study) and the complete dataset (Fraïsse, Belkhir, Welch, & Bierne, 2016). Densities for F_{ST} between 0 and 1 are presented, so that negative F_{ST} values are omitted.

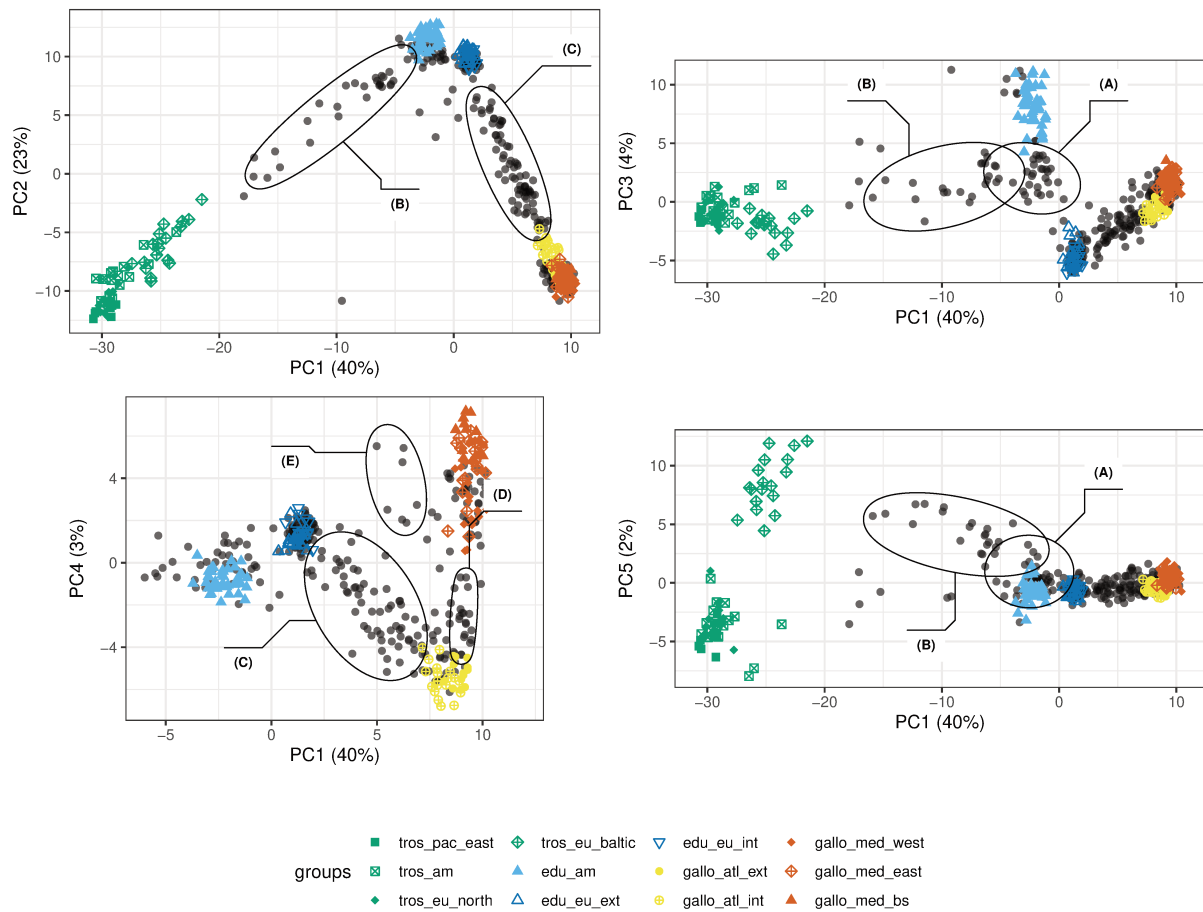


Figure S2: Principal Components Analysis presented for the first 5 components. In the bottom left PCA, *M. trossulus* individuals were cut out from the plot to zoom on the *M. edulis* and *M. galloprovincialis* interesting parts. Annotations show five groups of interest discussed in the main text: (A) North-European *M. edulis*; (B) Oresund hybrid zone; (C) Brittany, Aquitaine and Scotland hybrid zones (D) Algerian hybrid zone and (E) the port of Cherbourg.

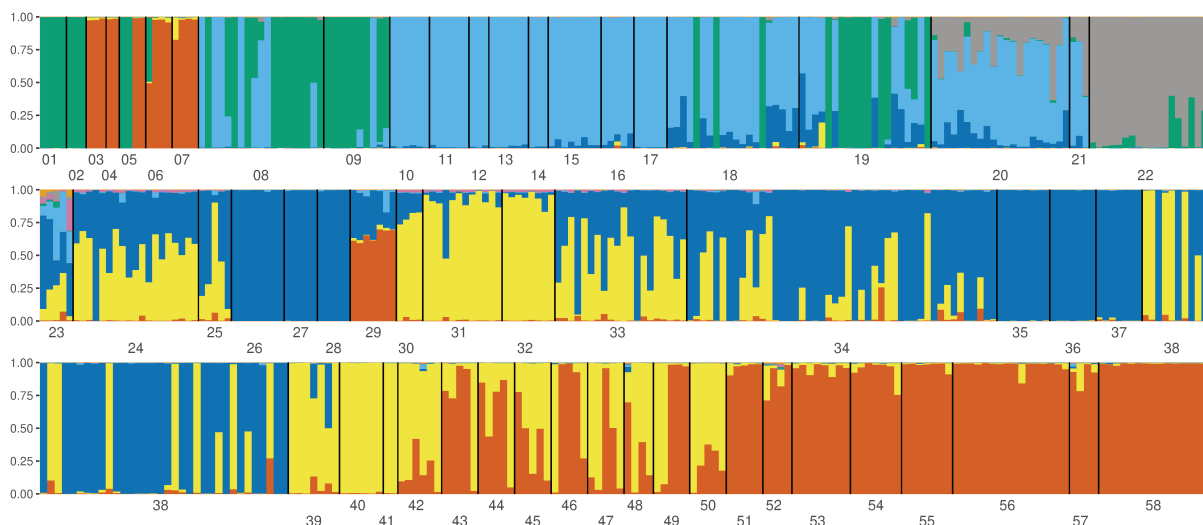


Figure S3: Structure results with $K = 8$, for all populations. The major mode obtained with Clumpak is presented.

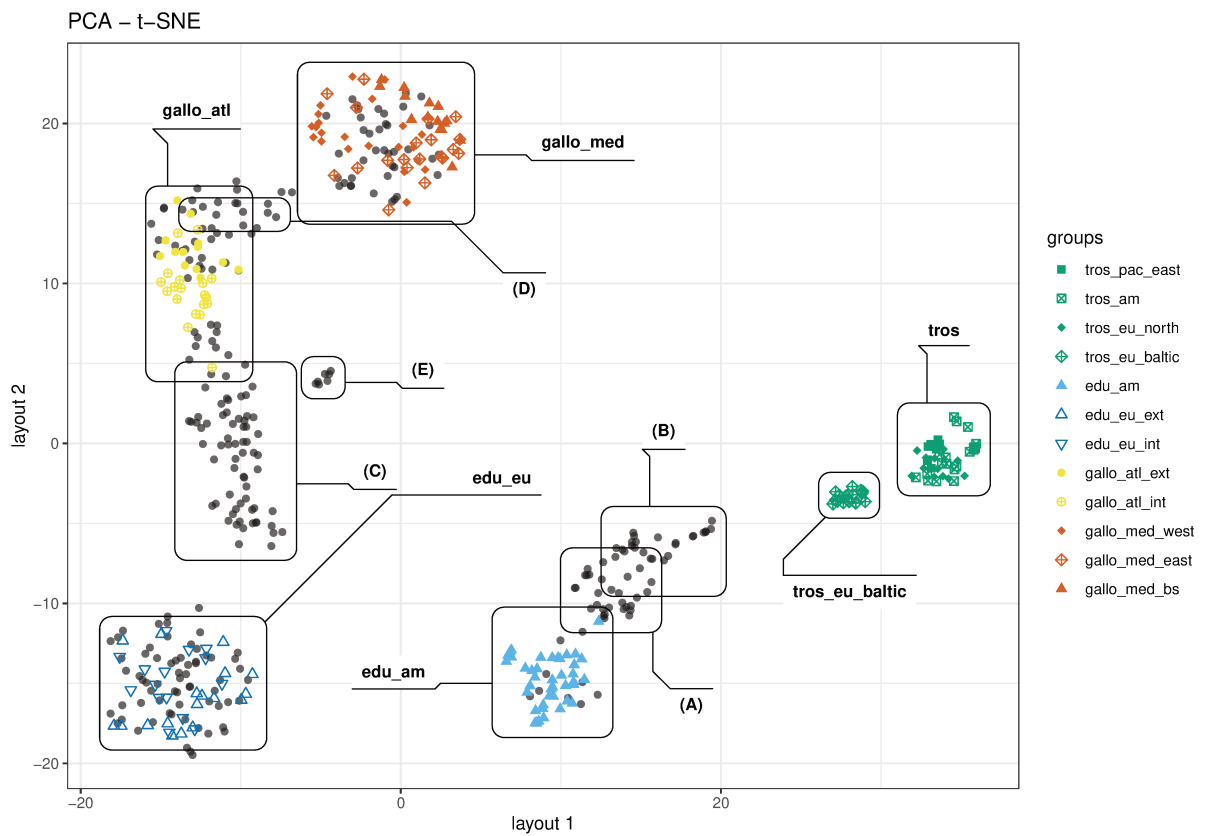


Figure S4: tSNE method applied to the first informative PCA components, as was done for Figure 2.

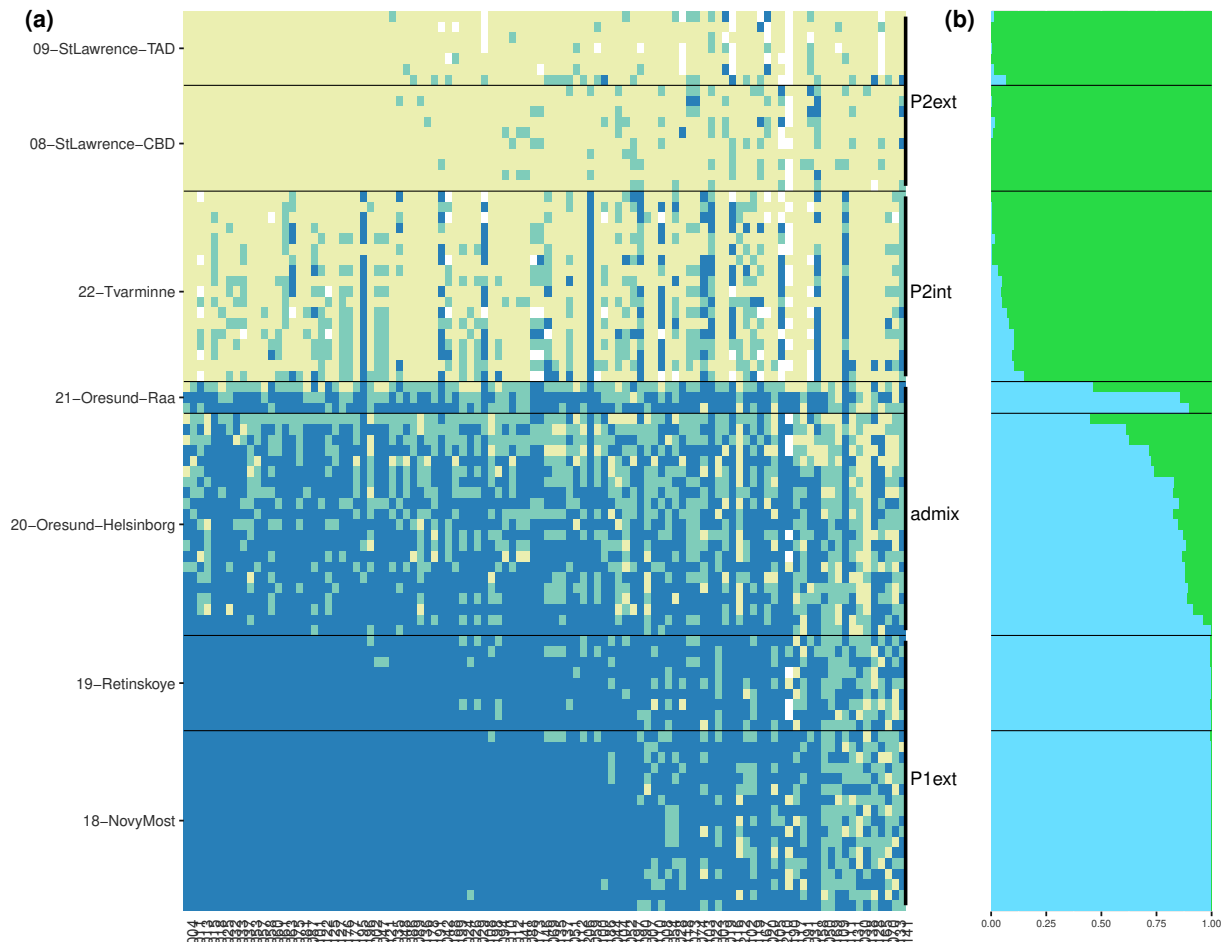


Figure S5: Oresund hybrid zone. (a) introgress plot of genotypes for each individual in line and each marker in column ($AFD > 0.5$). Individuals are ordered by hybrid index in each population while markers are ordered by degree of differentiation. (b) Corresponding Structure ancestry compositions for each individual.

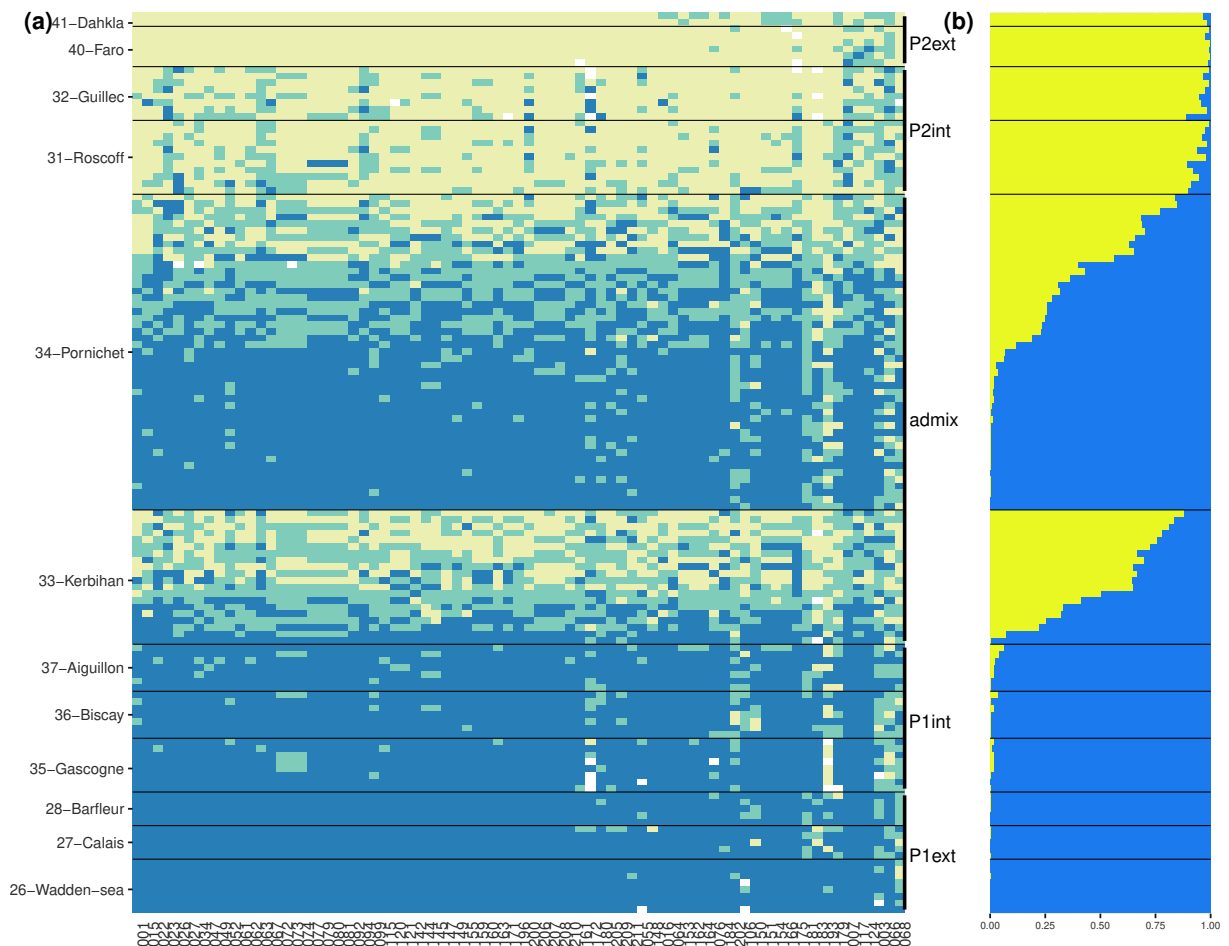


Figure S6: Brittany hybrid zone. (a) introgress plot of genotypes for each individual in line and each marker in column ($AFD > 0.5$). Individuals are ordered by hybrid index in each population while markers are ordered by degree of differentiation. (b) Corresponding Structure ancestry compositions for each individual.

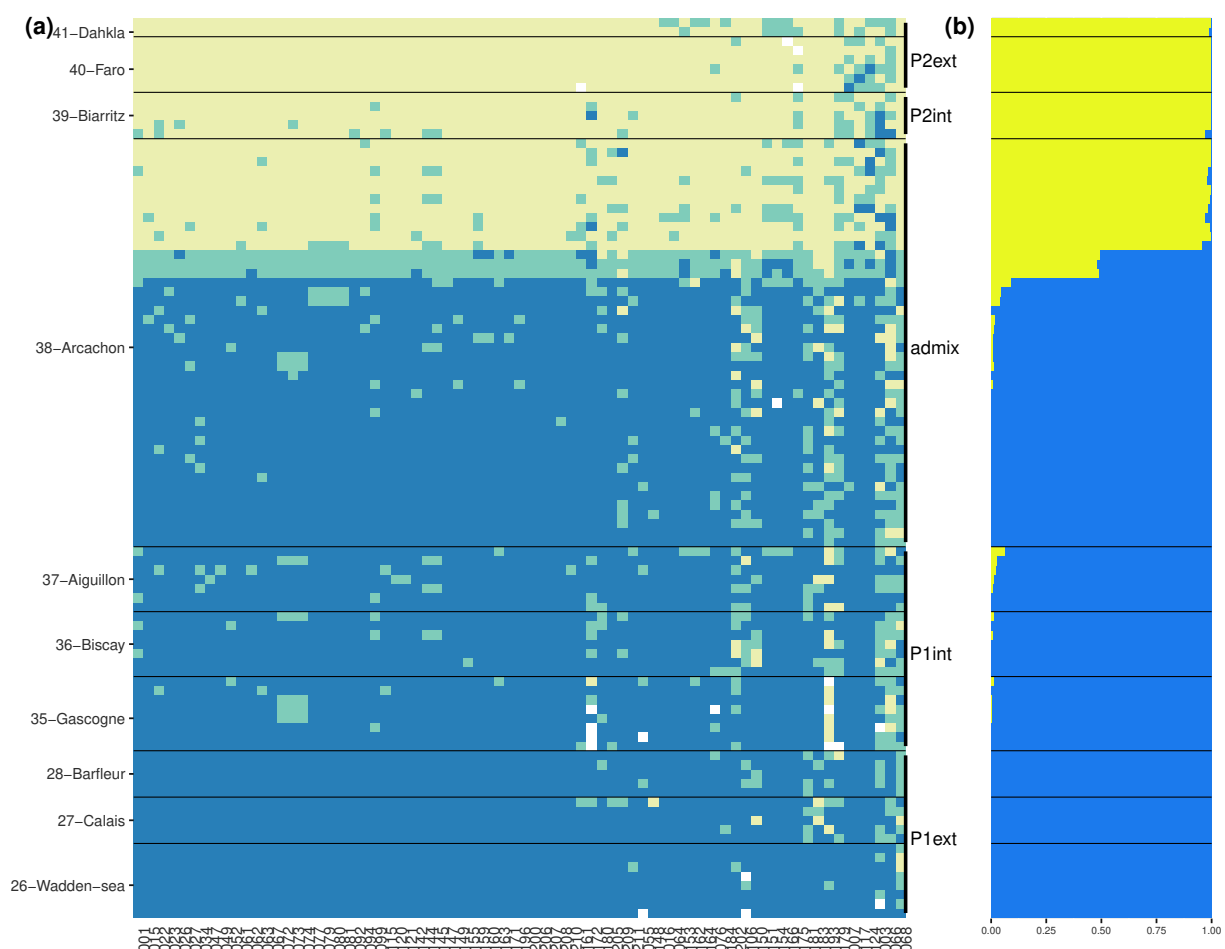


Figure S7: Aquitaine hybrid zone. (a) introgress plot of genotypes for each individual in line and each marker in column ($AFD > 0.5$). Individuals are ordered by hybrid index in each population while markers are ordered by degree of differentiation. (b) Corresponding Structure ancestry compositions for each individual.

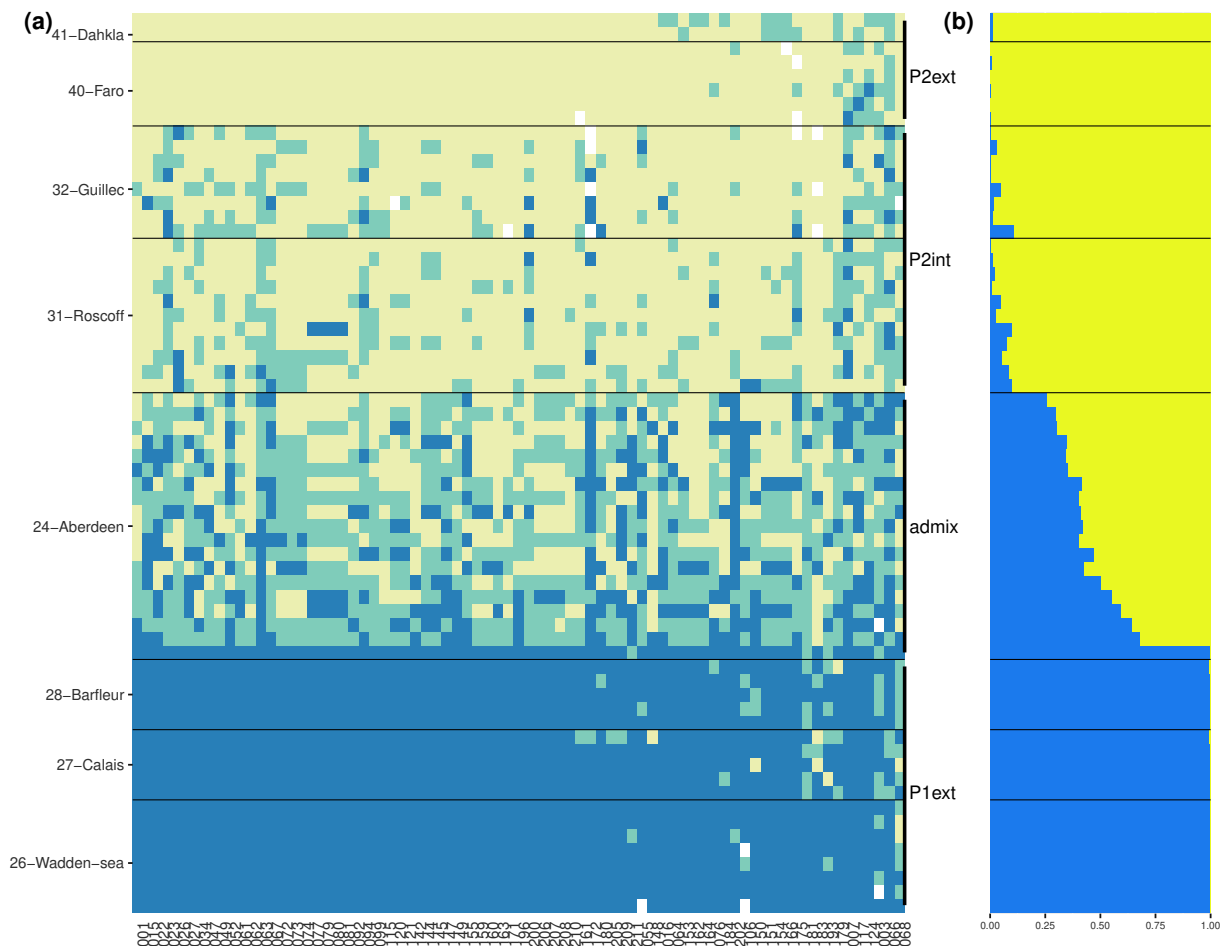


Figure S8: Scotland hybrid zone. (a) introgress plot of genotypes for each individual in line and each marker in column ($AFD > 0.5$). Individuals are ordered by hybrid index in each population while markers are ordered by degree of differentiation. (b) Corresponding Structure ancestry compositions for each individual.

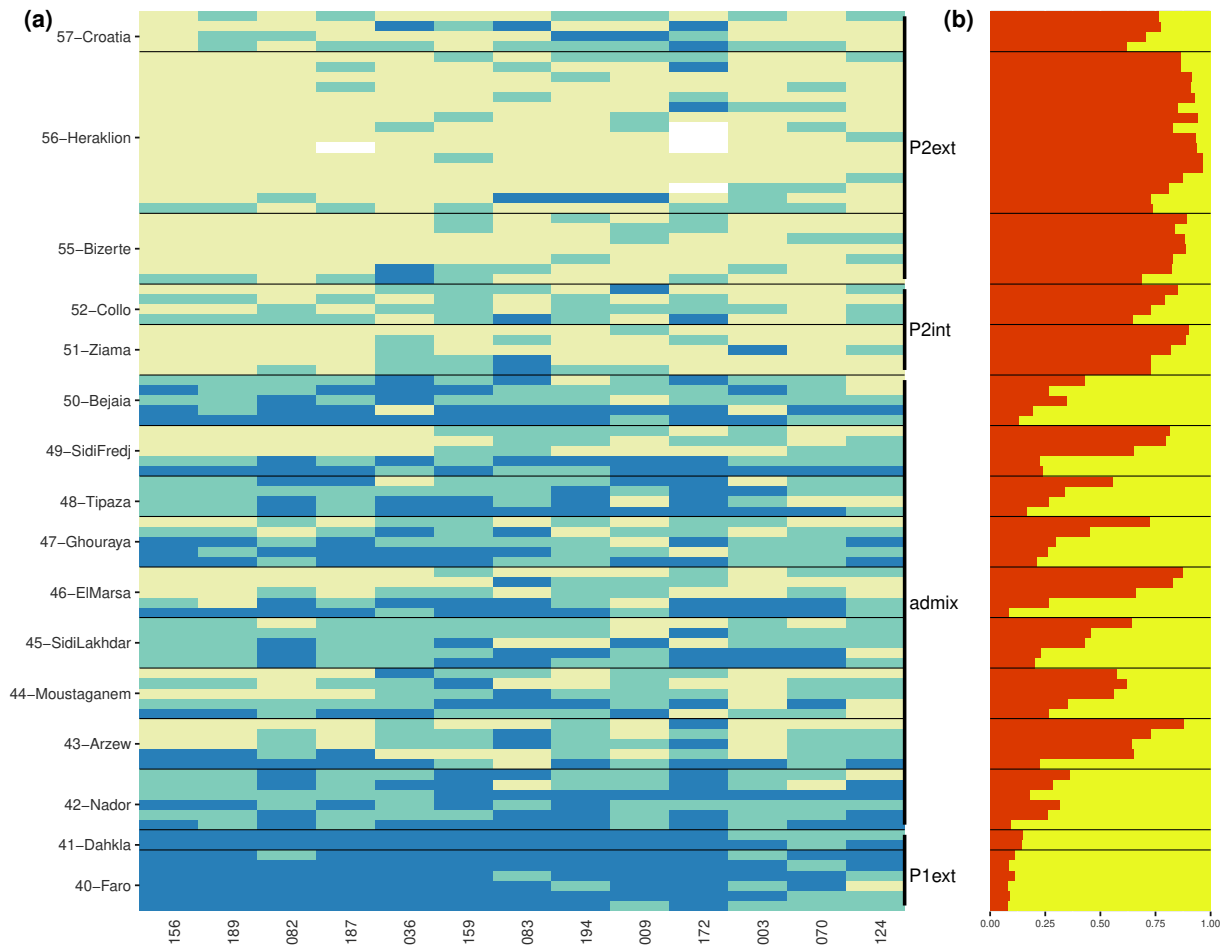


Figure S9: Algerian hybrid zone. (a) introgress plot of genotypes for each individual in line and each marker in column ($AFD > 0.5$). Individuals are ordered by hybrid index in each population while markers are ordered by degree of differentiation. (b) Corresponding Structure ancestry compositions for each individual.

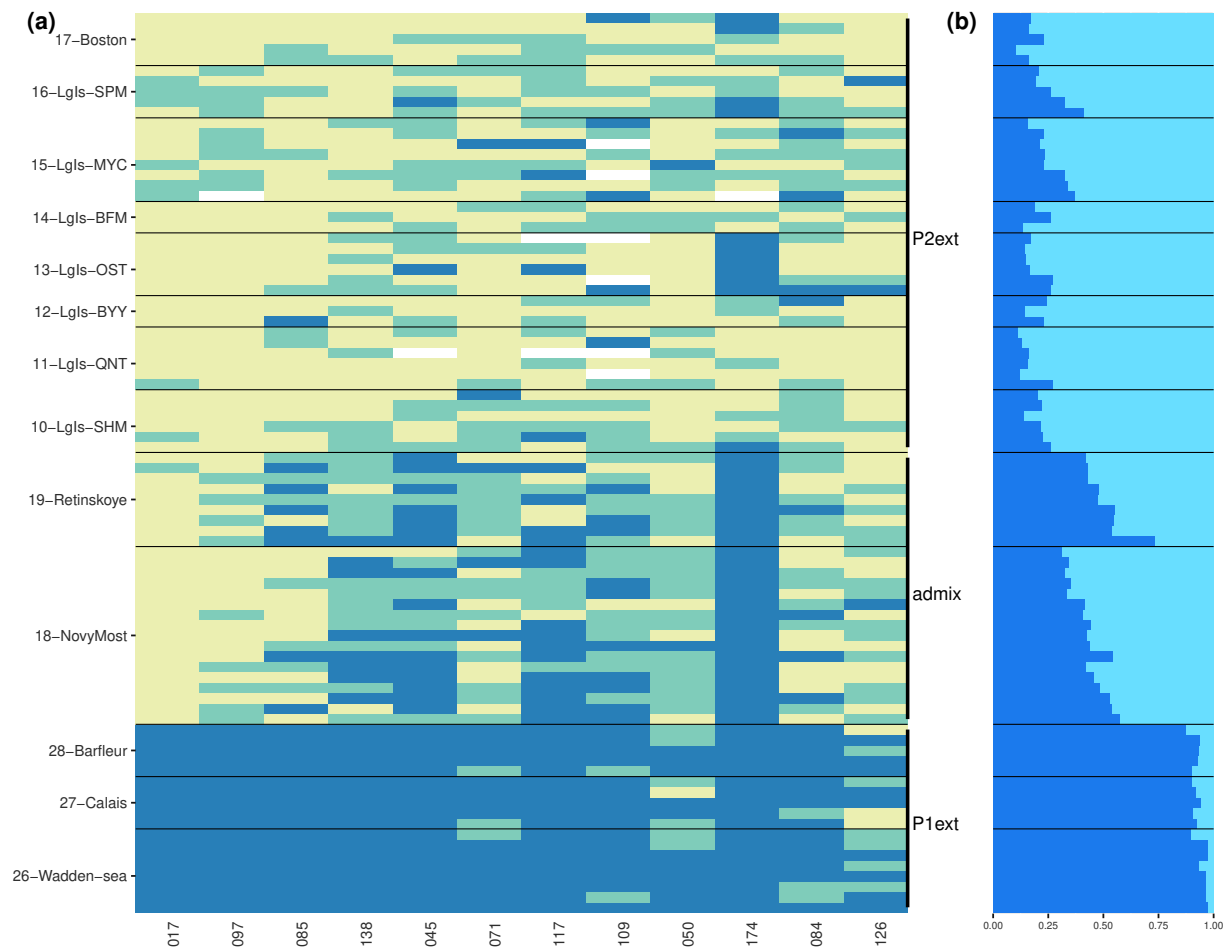


Figure S10: Northern European *M. edulis* hybrid zone. (a) introgress plot of genotypes for each individual in line and each marker in column (AFD > 0.5). Individuals are ordered by hybrid index in each population while markers are ordered by degree of differentiation. (b) Corresponding Structure ancestry compositions for each individual.

Chapitre 2

Admixture anthropogénique : analyse spatiale et génétique

2.1 Contexte et résumé de l'étude

2.1.1 Découverte d'introductions intra-européennes

Dans le chapitre précédent, sept moules échantillonnées dans le port de plaisance de Cherbourg (Normandie, France) ont été identifiées comme des *M. galloprovincialis* admixées, comportant une majorité d'ascendance méditerranéenne. Par la suite, elles seront nommées « moules des docks » (*dock mussels*). Le jeu de marqueurs informatifs développé au chapitre précédent et les connaissances acquises sur le complexe *M. edulis* pouvaient donc être utilisés pour augmenter notre effort d'échantillonnage des côtes françaises de la Manche et de l'Atlantique, dans les ports et les populations naturelles autour de ceux-ci. Nous souhaitons donc combler un manque d'échantillonnage dans des environnements peu explorés habituellement dans les études de génétique des populations des moules, dont les ports et estuaires. Le génotypage de l'ensemble des échantillons s'est distribué sur quatre expériences différentes qui ont permis d'identifier les ports d'intérêt et d'augmenter progressivement l'effort d'échantillonnage en fonction des résultats préliminaires.

En parallèle de ce travail, deux études avaient détecté des allèles *M. galloprovincialis* en mer du Nord (Coolen 2017), ainsi qu'en Norvège et au Svalbard (Mathiesen et al. 2016). Nous souhaitons vérifier dans un premier temps la composition génétique de ces potentielles introductions avec notre jeu de SNPs. La première étude contenait quelques moules des docks, cependant la seconde s'est avérée être une introduction indépendante de la lignée atlantique de *M. galloprovincialis*, présentant elle aussi une admixture avec l'espèce locale.

2.1.2 Les lignées de *M. galloprovincialis* peu souvent identifiées

Une large proportion des invasions marines n'a sûrement pas été détectée, souvent liée à une méconnaissance des espèces cryptiques, c'est-à-dire indifférenciées morphologiquement (Viard et al. 2016). À cela s'ajoute pour le complexe *M. edulis* plusieurs constats ayant retardé la détection de cette introduction. *M. galloprovincialis* med. n'est dans les faits pas recherchée en Europe pour plusieurs raisons :

- La distinction Méditerranée/Atlantique au sein de *M. galloprovincialis*, bien que reconnue depuis longtemps (Quesada et al. 1995c), n'est pas formalisée par une classification en deux espèces ou sous-espèces. La différence entre ces deux lignées est ainsi rarement prise en compte, soit par manque d'une population de référence (d'atlantique ou de méditerranéenne), soit par l'utilisation de marqueurs peu informatifs sur cette différenciation. Cette imprécision provoque souvent un manque d'informations sur la source d'introduction à l'échelle mondiale (cf. Tableau 1), mais aussi une banalisation de la présence de *M. galloprovincialis* dans des régions où la lignée atlantique pourrait être présente naturellement (soit proche de ses distributions naturelles reconnues, soit dans des régions peu étudiées où sa présence naturelle serait probable). Par exemple, l'observation de Skibinski et al. (1978b) d'une population *M. galloprovincialis* atypique dans le port de Swansea, discutée dans l'article de ce chapitre, est passée inaperçue dans la communauté par sa localisation dans une zone d'hybridation entre *M. galloprovincialis* atl. et *M. edulis*.
- De plus, l'histoire naturelle du complexe *M. edulis* étant particulièrement compliquée et encore mal résolue, il est parfois difficile de conclure sur les origines de certaines populations. C'est le cas d'une population de *M. trossulus* en Écosse (Loch Etive) qui est interprétée comme une relique post-glaciaire ayant été coincée dans un loch (Beaumont et al. 2008 ; Zbawicka et al. 2010).
- De manière générale, les ports sont des environnements peu échantillonnés lorsque l'on s'intéresse à la génétique des populations naturelles. De plus, les nombreuses études utilisant *Mytilus* spp. comme modèle pour l'écotoxicologie, parfois dans des milieux anthropisés, caractérisent rarement avec précision les différentes lignées utilisées.
- Enfin, les études spécifiques aux introductions utilisant les méthodes de métagénomique et d'ADN environnemental se heurtent à une limitation technique puisqu'elles ne permettent pas de déterminer à l'aide d'une seule ou deux séquences l'espèce (et encore moins la lignée) de *Mytilus* présente dans l'échantillon. Ce problème est lié à la fois à une divergence limitée entre certaines lignées, mais surtout à une introgression généralisée dans l'ensemble du complexe *M. edulis*. Ainsi ce type d'étude, ciblant pourtant spécifiquement la détection d'événements d'introduction, rapporte la présence d'espèces *Mytilus* spp. (Holman et al. 2019), et ce au sein de régions habitées naturellement par au moins une espèce de *Mytilus*.

Ce chapitre s'intéresse pour l'aspect spatial de l'introduction aux côtes françaises de la Manche et de l'Atlantique. Cette région est naturellement habitée par les deux lignées *M. edulis*

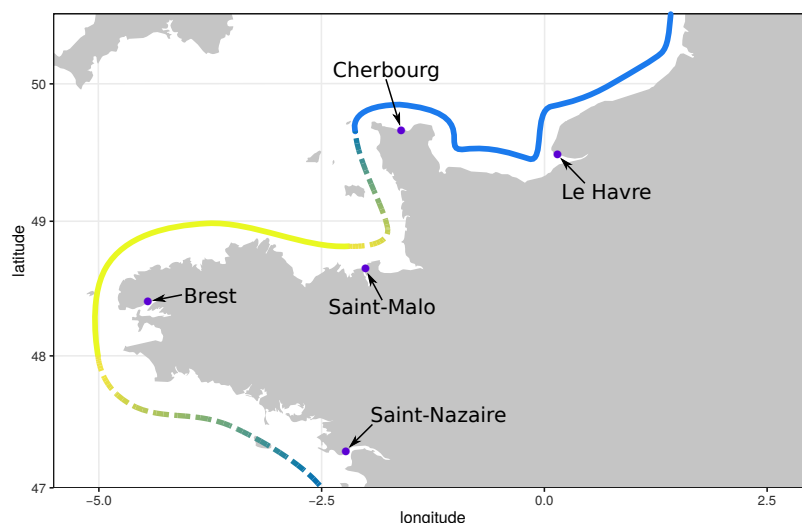


Figure 2.1 : Carte présentant la distribution attendue des deux espèces de moules *M. galloprovincialis* (jaune) et *M. edulis* (bleu) le long des côtes françaises de la Manche et de l'Atlantique, ainsi que la position approximative des zones d'hybridation les impliquant. Cinq ports échantillonnés dans l'étude sont indiqués par les points violets.

sud-eu. et *M. galloprovincialis* atl. et présente trois zones d'hybridation (Figures 8 et 2.1). Cinq ports ont été identifiés comme colonisés par les moules des docks (Figure 2.1).

2.1.3 Les ports : points chauds et ponts d'invasion

Les ports jouent un rôle important dans les invasions marines. En effet, ils constituent des « îles protégées pour les organismes marins introduits » (Bax et al. 2002). Les ports sont potentiellement les premiers récepteurs des espèces non natives puisqu'ils sont des hubs de la voie d'introduction la plus importante : le trafic maritime (Figure 6 ; Molnar et al. 2008). De ce fait, les ports commerciaux sont des points chauds d'introduction (J. M. Drake et Lodge 2004 ; Ferrario et al. 2017). Le transport d'espèces non indigènes a très certainement commencé avec les premiers voyages maritimes il y a des siècles (Carlton et Hodder 1995), cependant l'augmentation exponentielle des échanges en terme de volume et de distance depuis le début de l'ère industrielle a augmenté considérablement la pression de propagules.

Dans les invasions marines, les ports peuvent avoir un rôle de pont d'invasion vers des introductions secondaires régionales, à la fois vers d'autres ports ou vers le milieu naturel (Wasson et al. 2001 ; López-Legentil et al. 2015 ; Ferrario et al. 2017). Par exemple les marinas (ports de loisirs), souvent présentes à l'intérieur ou à proximité des ports de commerce, contribuent à la dispersion des espèces invasives (Clarke Murray et al. 2011). Les ports pourraient aussi servir de pas japonais temporaires pour des transferts sur de plus grandes distances (Apte et al. 2000). Des moules des docks ayant été récoltées dans plusieurs grands ports commerciaux et militaires, il était important d'évaluer l'ampleur de leur distribution dans les ports, à la fois commerciaux et de loisirs, afin de faire un état des lieux de l'introduction.

2.1.4 Résumé

Dans un premier temps, l'objectif de ce chapitre est de documenter l'étendue de l'introduction le long des côtes françaises (Manche et Atlantique) et la composition génétique de ces introductions. Nous concluons que le type initial observé à Cherbourg, que l'on appelle moule des docks (*dock mussel*), est admixé dans des proportions respectives $\sim 70/30\%$ entre *M. galloprovincialis* med. et *M. edulis* sud-eu. Bien qu'ayant envahi la majorité des ports analysés, la répartition des moules des docks est restreinte à l'intérieur des ports et elles forment des zones d'hybridation à faible échelle à l'entrée des ports avec les fonds génétiques locaux. Ces moules locales peuvent être soit *M. edulis* sud-eu. soit *M. galloprovincialis* atl. (Figure 2.1).

Les marqueurs informatifs entre *M. edulis* et *M. galloprovincialis* med. dans les moules des docks montrent une variance limitée de l'ascendance entre marqueurs et l'absence de fixations. Cette absence de fixations permet d'émettre l'hypothèse de l'absence d'incompatibilités post-zygotiques à très fort effet, qui étaient pourtant attendues (Bierne et al. 2006), et de balayages sélectifs importants. De plus, l'analyse des distorsions des fréquences alléliques par rapport à la moyenne génomique, à la fois entre ports mais aussi avec des admixtures indépendantes impliquant les mêmes espèces ou des espèces proches, montre des parallélismes importants dans certains cas. Ces parallélismes apparaissent lorsque les fonds génétiques sont similaires et peuvent avoir lieu à la fois à cause d'histoires évolutives partagées (pour les multiples introductions) mais aussi pour des raisons intrinsèques de sélection à l'échelle génomique (pour les admixtures indépendantes).

2.2 Article 2

Cet article est publié dans le journal *Evolutionary Applications*. Il est inclus dans l'édition spéciale intitulée « *Evolutionary Biology of Marine Invasions* ».

A. Simon, C. Arbiol, E. E. Nielsen, J. Couteau, R. Sussarellu, T. Burgeot, I. Bernard, J. W. P. Coolen, J.-B. Lamy, S. Robert, M. Skazina, P. Strelkov, H. Queiroga, I. Cancio, J. J. Welch, F. Viard et N. Bierne (2019b). « Replicated Anthropogenic Hybridisations Reveal Parallel Patterns of Admixture in Marine Mussels ». *Evolutionary Applications*. DOI : 10.1111/eva.12879. Prépubl.

Replicated anthropogenic hybridisations reveal parallel patterns of admixture in marine mussels.

Alexis Simon¹, Christine Arbiol¹, Einar Eg Nielsen², Jérôme Couteau³,
 Rossana Sussarellu⁴, Thierry Burgeot⁴, Ismaël Bernard⁵, Joop W.P.
 Coolen^{6,7}, Jean-Baptiste Lamy⁸, Stéphane Robert⁸, Petr Strelkov^{9,10},
 Henrique Queiroga¹¹, Ibon Cancio¹², John J. Welch¹³, Frédérique Viard¹⁴ and
 Nicolas Bierne¹

¹ISEM, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France.

²Section for Marine Living Resources, National Institute of Aquatic Resources, Technical University of Denmark, Silkeborg, Denmark.

³SARL TOXEM, Le Havre, France.

⁴Ifremer Unité Biogéochimie et Écotoxicologie, Centre Atlantique, Nantes.

⁵SAS Eurêka Mer, Lézardrieux, France.

⁶Wageningen Marine Research, P.O. Box 57, 1780 AB Den Helder, The Netherlands.

⁷Wageningen University, Aquatic Ecology and Water Quality Management Group, Droevendaalsesteeg 3a, 6708 PD Wageningen, The Netherlands.

⁸Ifremer, SG2M-LGPMM, Laboratoire de Génétique et Pathologie des Mollusques Marins, La Tremblade, France.

⁹St. Petersburg State University, Universitetskaya Emb. 7/9, St. Petersburg 199034, Russia.

¹⁰Laboratory of Monitoring and Conservation of Natural Arctic Ecosystems, Murmansk Arctic State University, Kapitana Egorova Str. 16, Murmansk 183038, Russia.

¹¹Department of Biology & CESAM, University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal.

¹²CBET Research Group, Dept. of Zoology and Animal Cell Biology, Fac. Science and Technology and Research Centre for Experimental Marine Biology and Biotechnology (PiE-UPV/EHU), University of the Basque Country (UPV/EHU), Bilbao, Spain.

¹³Department of Genetics, University of Cambridge, Downing St. Cambridge, CB23EH, UK.

¹⁴Sorbonne Universités, UPMC Univ Paris 06, CNRS, UMR 7144, Department AD2M, Station Biologique, Roscoff, France.

Abstract

Human-mediated transport creates secondary contacts between genetically differentiated lineages, bringing new opportunities for gene exchange. When similar introductions occur in different places, they provide informally replicated experiments for studying hybridisation. We here examined 4279 *Mytilus* mussels, sampled in Europe and genotyped with 77 ancestry informative markers. We identified a type of introduced mussels, called 'dock mussels', associated with port habitats and displaying a particular genetic signal of admixture between *M. edulis* and the Mediterranean lineage of *M. galloprovincialis*. These mussels exhibit similarities in their ancestry compositions, regardless of the local native genetic backgrounds and the distance separating colonised ports. We observed fine-scale genetic shifts at the port entrance, at scales below natural dispersal distance. Such sharp clines do not fit with migration-selection tension zone models, and instead suggest habitat choice and early stage adaptation to the port environment, possibly coupled with connectivity barriers. Variations in the spread and admixture patterns of dock mussels seem to be influenced by the local native genetic backgrounds encountered. We next examined departures from the average admixture rate at different loci, and compared human-mediated admixture events, to naturally admixed populations and experimental crosses. When the same *M. galloprovincialis* background was involved, positive correlations in the departures of loci across locations were found; but when different backgrounds were involved, no or negative correlations were observed. While some observed positive correlations might be best explained by a shared history and saltatory colonisation, others are likely produced by parallel selective events. Altogether, genome-wide effect of admixture seems repeatable, and more dependent on genetic background than environmental context. Our results pave the way towards further genomic analyses of admixture, and monitoring of the spread of dock mussels both at large and fine spatial scales.

Keywords: biological introductions, benthopelagic species, ports, secondary contact, clines, admixture.

1 Introduction

Biological introductions have evolutionary impacts on both native and introduced species, through ecological and genetic responses (Mooney & Cleland, 2001; Prentis, Wilson, Dormontt, Richardson, & Lowe, 2008; Strayer, Eviner, Jeschke, & Pace, 2006; Suarez & Tsutsui, 2008). This is especially so when ‘anthropogenic hybridisations’ lead to gene exchange (see McFarlane & Pemberton, 2019, for a recent review). Anthropogenic hybridisations have probably been underestimated, but have nevertheless been reported in diverse taxonomic groups, including plants, birds, fishes, mammals and invertebrates (Largiadèr, 2008, and references therein). For instance, in nineteen different fish families, half of the observed interspecific hybridisations have been attributed to human disturbances (Scribner, Page, & Bartron, 2000). The outcomes of these hybridisations could be similarly diverse. Hybridisation might favour the sustainable establishment of non-indigenous species (NIS) by facilitating adaptation to the local environment via the introgression of ‘ready-to-use’ alleles from native genomes. Immediate advantage could also be gained through heterosis at the initial stage of introduction (Ellstrand & Schierenbeck, 2000; Schierenbeck & Ellstrand, 2009; Suarez & Tsutsui, 2008). Conversely, hybridisation is often considered as ‘genetic pollution’ of the native species, raising concerns of ‘extinction by hybridization and introgression’ (Rhymer & Simberloff, 1996), although these concerns often neglect the possibility of genetic rescue (Fitzpatrick et al., 2019; Harris, Zhang, & Nielsen, 2019). Additionally, hybrid fitness depression might oppose introduction success, stopping the spread of the introduced lineage (Kovach et al., 2016), perhaps at a natural barrier (Barton, 1979b). Overall, the evolutionary consequences of anthropogenic hybridisation (i.e., gene flow, local introgression, reinforcement, or rescue) are likely to be strongly contingent on intrinsic and extrinsic factors, such as the accumulation of reproductive incompatibilities or local selection processes (Abbott et al., 2013).

Introductions with hybridisation can also shed light on the evolutionary process itself. Just like natural hybrid zones, human-induced hybrid zones can be seen as ‘natural laboratories for evolutionary studies’ (G. M. Hewitt, 1988, p. 158) (Abbott et al., 2013; Barton & Hewitt, 1989). Indeed, anthropogenic introductions have a special value, because they tend to be recent, informally replicated (taking place independently in different locations), and can often be compared to natural admixture events between the same lineages (Bouchemousse, Liautard-Haag, Bierne, & Viard, 2016). This is important because, even with genome-wide genetic data and powerful inferential methods, the traces of secondary contacts tend to erode over time, and can be confounded with other processes (Bertl, Ringbauer, & Blum, 2018; Bierne, Gagnaire, & David, 2013). Recent secondary contacts allow a unique window on the processes involved during the early phase of admixture, including the sorting of alleles in admixed populations (Schumer et al., 2018).

The blue mussel complex of species (*Mytilus edulis*) includes three species naturally distributed in temperate regions of the Northern hemisphere: *M. edulis* (Linnaeus 1758), *M. gal-*

loprovincialis (Lamarck 1819) and *M. trossulus* (Gould 1850). It constitutes a model for investigating the genetic and evolutionary consequences of marine invasions (Popovic, Matias, Bierne, & Riginos, 2019; Saarman & Pogson, 2015). Despite divergences estimated at 2.5 million years (MY) between *M. edulis* and *M. galloprovincialis* (Roux et al., 2014) and 3.5 MY between these and *M. trossulus* (Rawson & Hilbish, 1995), they are incompletely reproductively isolated and readily hybridise where they meet.

Where found in sympatry, the distribution of *M. edulis* and *M. galloprovincialis* are correlated with salinity, tidal height and wave exposure (Bierne, David, Langlade, & Bonhomme, 2002; Gardner, 1994). In certain cases, *M. edulis* occupies sheltered, deeper or estuarine environments, while *M. galloprovincialis* is found on more wave-exposed parts of the coast. In regions with a single species, however, individuals can occupy all niches. It should also be noted that independent contacts can show reversed associations with the environment, in agreement with the coupling hypothesis (Bierne, Welch, Loire, Bonhomme, & David, 2011). *M. galloprovincialis*, though known as the Mediterranean mussel, has a large natural distribution – from the Black Sea to the North of the British Isles – and is divided into two main lineages, Atlantic (Atl.) and Mediterranean (Med.). (Fraïsse, Belkhir, Welch, & Bierne, 2016; Popovic et al., 2019; Quesada, Zapata, & Alvarez, 1995; Roux et al., 2014; Zbawicka, Drywa, Śmietanka, & Wenne, 2012). These two lineages form hybrid zones in the Almeria-Oran front region (El Ayari, Trigui El Menif, Hamer, Cahill, & Bierne, 2019; Quesada, Beynon, & Skibinski, 1995; Quesada, Zapata, & Alvarez, 1995).

Mussels of the family *Mytilidae* have several traits making them prone to transportation by humans. As benthopelagic molluscs, their planktonic feeding larval stage allows long distance spread through both marine currents (Bayne, 1976; Branch & Steffani, 2004) and anthropogenic vectors, mostly via ballast water (Geller, Carlton, & Powers, 1994) or fouling (e.g. on hulls: Apte, Holland, Godwin, and Gardner, 2000; Casoli et al., 2016; or marine litter: Miller, Carlton, Chapman, Geller, and Ruiz, 2017; Miralles, Gomez-Agenjo, Rayon-Viña, Gyraitė, and Garcia-Vazquez, 2018; Węstawski and Kotwicki, 2018). Mussels are also heavily cultivated on a global scale (287,958 tonnes in 2016, FAO, 2018); they can therefore follow the two main introduction pathways of marine species: international shipping and aquaculture (Molnar, Gamboa, Revenga, & Spalding, 2008; Nunes, Katsanevakis, Zenetos, & Cardoso, 2014). While larval dispersal might allow a post-introduction range expansion, initial establishment also relies on avoiding demographic and genetic Allee effects. As such, successful establishment depends on either large propagule pressure (likely to occur in many marine NIS: Rius, Turon, Bernardi, Volckaert, and Viard, 2015; Viard, David, and Darling, 2016), or on hybridisation with a native species (Mesgaran et al., 2016). In *Mytilus* mussels, this is facilitated by both high fecundity and high density traits, and by their incomplete reproductive isolation.

Among *Mytilus* species, *M. galloprovincialis* has been introduced many times across the globe, in both the northern and southern hemispheres, and notably, along the Pacific coast of North America, in South America, South Africa, Asia, and Oceania (Branch & Steffani,

2004; Daguin & Borsa, 2000; Han, Mao, Shui, Yanagimoto, & Gao, 2016; Kartavtsev, Chichvarikhin, Kijima, Hanzawa, & Park, 2005; Larraín, Zbawicka, Araneda, Gardner, & Wenne, 2018; McDonald, Seed, & Koehn, 1991; Saarman & Pogson, 2015; Zbawicka, Trucco, & Wenne, 2018). By contrast, we only know of a few cases of *M. edulis* introductions – either transient or successful – into non-native areas (Casoli et al., 2016; Crego-Prieto et al., 2015; Fraïsse, Haguenaer, et al., 2018). Branch and Steffani (2004) reported that observed introductions of *M. galloprovincialis* happened close to large shipping ports, with a secondary range expansion from these points. For instance in South Africa, *M. galloprovincialis* spread rapidly and had varying impacts on local communities, modulated by wave action (Branch, Odendaal, & Robinson, 2008; Branch & Steffani, 2004). Wherever *Mytilus* species are native, *M. galloprovincialis* has been shown to be highly competitive and has often displaced local mussels (James T. Carlton, Geller, Reaka-Kudla, & Norse, 1999). *M. galloprovincialis* has also been reported in the subarctic and Arctic, notably in Norway (Brooks & Farmen, 2013; Mathiesen et al., 2016). Given the low divergence between Atl. and Med. *M. galloprovincialis*, and their assignment to the same species, introduced *M. galloprovincialis* has often been reported without further investigation of its origin, and when markers are insufficiently informative, the origin is necessarily unresolved. Nevertheless, it is clear that both lineages have been successfully introduced in multiple places worldwide (Atl. in South Africa and Australia; Med. in the Eastern and Western Pacific Ocean; see Daguin and Borsa, 2000; Han et al., 2016; Popovic et al., 2019; Zardi et al., 2018).

Just as mussels are model organisms for studying the processes underlying successful introduction of alien species, ports are model locations (Bax, Hayes, Marshall, Parry, & Thresher, 2002). Because they are hubs of maritime traffic, with high connectivity, they are bridgeheads towards expansion at regional scales (Drake & Lodge, 2004). Vessels have been shown to be a major introduction pathway, through various vectors, including ballast water, sea-chest and hull (Katsanevakis, Zenetos, Belchior, & Cardoso, 2013; Sylvester et al., 2011). In addition, ports are often distinct from nearby natural habitats, with particular environmental features (Chapman & Underwood, 2011, and references therein). These new niches can be colonised by opportunistic species, such as many NIS (Bishop et al., 2017, and references therein). Mussels are likely to be introduced and become established in ports due to their aforementioned life history traits, their robustness to environmental pollution (Mlouka et al., 2019; Roberts, 1976), and tolerance to a large range of environmental conditions in terms of temperature, salinity and wave action (both through individual plasticity and interspecific variability; Braby and Somero, 2006; Fly and Hilbish, 2013; Lockwood and Somero, 2011).

In this study, using a population genomic dataset comprising 4279 mussels genotyped at 77 ancestry informative SNPs, we examined mussel populations established in ports in North-West France (located along the Atlantic and the English Channel coastlines), and compared these to mussel populations established in the vicinity. This genetic survey allows us to report, for the first time, an unexpected and extensive introduction of a non-indigenous lineage of

M. galloprovincialis into five ports in our study area. We show that the introduced mussels have a distinctive genetic signature, originating from admixture between the Med. *M. galloprovincialis* and native *M. edulis*. We call these mussels, 'dock mussels', in recognition of their strong association with port environments. Dock mussel populations in ports appear to constitute stable admixed populations and form small-scale hybrid zones with native mussels at the port entrance, which can be either *M. edulis* or Atl. *M. galloprovincialis* depending on the region.

To place these populations in a wider context, we additionally analysed published and new samples of putative *M. galloprovincialis* in Norway (Mathiesen et al., 2016), and concluded that these are admixed mussels between Atl. *M. galloprovincialis* and the local North-European (North-Eu.) *M. edulis* lineage, resulting from an anthropogenic introduction. We also combined our data with multiple samples of admixed populations from natural hybrid zones, and laboratory crosses. This allowed us to compare multiple independent events of admixture, with a variety of ecological and genomic contexts.

The similarities and differences between these various admixed populations help to clarify the factors that determine the outcome of an introduction with hybridisation. In particular, we show that similar outcomes sometimes reflect shared colonisation history, but can also arise in genuinely independent colonisations. However, this predictability is highly background dependent, and replicated outcomes only appear when the same parental backgrounds are involved.

2 Methods

2.1 Sampling and genotyping

We aimed to examine mussel populations in ports, following the discovery of mussels with unexpected Med. *M. galloprovincialis* ancestry in the port of Cherbourg (France), as sampled in 2003 (Simon et al., 2019). Besides a new sampling in Cherbourg, we sampled seven additional ports and neighbouring natural populations. We also aimed to compare the admixture patterns observed in the ports to other admixed populations, involving different lineages of the same species. The sampling focused on populations where we had *a priori* expectations of admixture. Therefore, it should not be confused with a representative sample of the *M. edulis* complex, where populations are usually much closer to the reference parental populations. Most of the port sites were sampled between 2015 and 2017 and older samples were used as references or for temporal information. We either received samples from collaborators or directly sampled in the areas of interest (see Figure S1 and Table S1 for full details).

As part of our sampling process, we re-genotyped samples from several previous studies that reported the presence of *M. galloprovincialis* alleles, but had not assigned the samples to the Atl. or Med. *M. galloprovincialis* lineages. In particular, we used previously extracted

DNA from the following studies: (i) Mathiesen et al. (2016) who studied the genetics of *Mytilus* spp. in the sub-Arctic and Arctic using 81 randomly ascertained SNPs. They identified *M. galloprovincialis* and putative hybrids with *M. edulis* in the Lofoten islands, Svalbard and Greenland. Their parental reference samples included only the Atl. *M. galloprovincialis* lineage (Galicia, Spain). Our aim was to further assess the origin of the *M. galloprovincialis* ancestry. (ii) Coolen (2017) studied connectivity between offshore energy installations in the North Sea, characterising samples with 6 microsatellite markers and the locus Me15/16. He identified populations containing individuals with *M. galloprovincialis* ancestry, using an Atl. *M. galloprovincialis* reference as well (Lisbon, Portugal).

Samples originating from another oil platform from the Norwegian Sea (Murchison oil station, MCH) and one Norwegian sample (Gåseid, GAS) were also included. We note that the MCH oil rig was free of settled mussels at the time of deployment.

These natural samples were compared to laboratory crosses between *M. edulis* and Med. *M. galloprovincialis*, produced in Bierne, Bonhomme, Boudry, Szulkin, and David (2006), and genotyped in Simon, Bierne, and Welch (2018). Briefly, F1 hybrids were first produced by crossing five males and five females of *M. edulis* from the North Sea (Grand-Fort-Philippe, France) and *M. galloprovincialis* from the western Mediterranean Sea (Thau lagoon, France). F2s were produced by crossing one F1 female and five F1 males. Additionally, sex-reciprocal backcrosses to *M. galloprovincialis* were made, they are named BCG when the females were *M. galloprovincialis* and BCF1 when the female was F1 (Table 1). Production of crosses are described in full detail in Bierne, David, Boudry, and Bonhomme (2002), Bierne et al. (2006) and Simon et al. (2018).

We collected gill, mantle or hemolymph tissues from mussels either fixed in 96% ethanol or freshly collected for DNA extraction. We used the NucleoMag™ 96 Tissue kit (Macherey-Nagel) in combination with a Kingfisher Flex (serial number 711-920, ThermoFisher Scientific) extraction robot to extract DNA. We followed the kit protocol with modified volumes for the following reagents: 2× diluted magnetic beads, 200 µL of MB3 and MB4, 300 µL of MB5 and 100 µL of MB6. The extraction program is presented in Figure S2.

Genotyping was subcontracted to LGC genomics (Hoddesdon, UK) and performed with the KASP™ array method (Semagn, Babu, Hearne, & Olsen, 2014). We used a set of ancestry informative SNPs developed previously (Simon et al., 2018; Simon et al., 2019). For cost reduction, we used a subset of SNPs that were sufficient for species and population delineation. Multiple experiments of genotyping were performed. The results were pooled to obtain a dataset of 81 common markers.

2.2 Filtering

To obtain a clean starting dataset, we filtered loci and individuals for missing data. We then defined groups of individuals used as reference in downstream analyses and identified loci

Table 1: Groups used in the analyses of ancestry comparisons and correlations of distortion. The location and ancestry composition of sub-groups are indicated in Figure 2. The native genetic backgrounds possibly encountered is indicated for cases of introduction (n/a: not applicable).

Group	Native genetic background	Admixture pattern	Sub-group	Populations used
Dock mussels	South-Eu. <i>M. edulis</i> or Atl. <i>M. galloprovincialis</i>	Med. <i>M. galloprovincialis</i> / South-Eu. <i>M. edulis</i>	havre cher stmalo brest stnaz	Port of Le Havre Port of Cherbourg Port of Saint-Malo Bay of Brest Port of Saint-Nazaire
F2	n/a	<i>idem</i>	F2	F1 female × F1 males
Backcrosses (BCs)	n/a	<i>idem</i>	BCG BCF1	gallo_med females × F1 males F1 female × gallo_med males
Norway admixed	North-Eu. <i>M. galloprovincialis</i> (sometimes <i>M. trossulus</i>)	Atl. <i>M. galloprovincialis</i> / North-Eu. <i>M. edulis</i>	LOF GAS	Lofoten Islands, Norway Gåseid, Norway
Naturally admixed	n/a	Atl. <i>M. galloprovincialis</i> / South-Eu. <i>M. edulis</i>	ABD MCH JER HZSB	Aberdeen, Scotland Murchison oil station Jersey Island Mousterlin point (MOU) La Jument (JUM) Barres de Pen Bron (PEN) Chemoulin point (CHE) Groix Penestin (BIL_001) Le Pouliguen (POU_001) Houat Island (HOU_001)

deviating from Hardy-Weinberg expectations, to filter used markers for analyses depending on equilibrium hypotheses.

Analyses were carried out using R (v3.5.3, R Core Team, 2019) and custom Python 3 scripts for format conversions. Software packages and versions used are listed in Table S2. Decision thresholds for all analyses and dataset selections are summarised in Table S3.

First, control individuals duplicated between genotyping experiments were removed by keeping the one having the least missing data. Over 81 markers, the maximum number of mismatches observed between two duplicated individuals was 2 (without considering missing data), showing that the genotyping method is mostly accurate. A few individuals identified as affected by a *M. trossulus* transmissible cancer were removed from the dataset (Metzger et al., 2016; Riquet, Simon, & Bierne, 2017).

The dataset was filtered for missing data with a maximum threshold of 10% for markers over all individuals and 30% for individuals over all markers. This filtering yielded 4279 individuals genotyped at 77 loci (from the initial dataset composed of 4495 individuals genotyped over 81 loci). We separated nuclear (76 loci) and mitochondrial (1 locus) markers for downstream analyses. The mitochondrial marker (named 601) is located on the female mitochondria.

Most analyses required reference population samples. A list of reference individuals and groups was set *a priori* using the literature and our knowledge of the *M. edulis* species complex (Figure 1c and Table S4). We defined three levels of structure that we call L1, L2 and L3. L1 is the species level comprising *M. edulis* (edu), *M. galloprovincialis* (gallo) and *M. trossulus* (tros). L2 defines allopatric lineages in each species: (i) American (edu_am, East coast) and European (edu_eu) *M. edulis*; (ii) Atl. (gallo_atl) and Med. (gallo_med) *M. galloprovincialis*; (iii) Pacific (tros_pac), American (tros_am, East coast) and European (tros_eu, Baltic Sea) *M. trossulus*. Finally, L3 defines sub-populations where the differentiation is mainly due to local introgression following historic contacts between lineages (Fraïsse et al., 2016): (i) North-Eu. populations of *M. edulis* (edu_eu_north) were included (Simon et al., 2019). This lineage is present along the coast of Norway and meet with the South-Eu. lineage (edu_eu_south) along the Danish coast; (ii) Atl. *M. galloprovincialis* from the Iberian peninsula (gallo_atl_iber) and mussels from Brittany (gallo_atl_brit); (iii) West (gallo_med_west) and East (gallo_med_east) Med. *M. galloprovincialis*, the limit being set at the Siculo-Tunisian strait.

To improve this predefined set of reference samples, an initial genetic clustering was performed with the software *Admixture* (Alexander, Novembre, and Lange, 2009, full nuclear dataset, 3 clusters, 30 replicates, fig S4) and the results were combined with the *CLUMPAK* software (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015). All individuals with less than 85% ancestry from their putative cluster were removed from the reference set (this threshold was chosen to account for local introgression in some populations). This step ensures there are no migrants, either from introduction or from sympatric species, and no hybrids in the reference panel.

Once the reference dataset was established, Hardy-Weinberg equilibrium (HWE) was estimated in each L3 level for all markers. *edu_eu_south* was separated in two groups, corresponding to the bay of Biscay (int, as in Fraïsse et al., 2016) and the English Channel (ext), for this analysis only, as they do not mate randomly but do not show significant genetic differentiation (Table S6). We used the `hw.test` function of the R package `pegas` (Paradis, 2010) with 10^4 Monte Carlo permutations and a Benjamini-Yekutieli false discovery rate correction. Markers 604 and 190 were identified as significantly departing from HWE in at least one reference group (Figure S3).

2.3 Genetic map

Estimates of linkage between markers allow us to account for admixture linkage disequilibrium in ancestry estimation (see `Structure` analyses below), and to estimate time since admixture.

We used F2 crosses to produce a genetic map for a subset of markers analysed by Simon et al. (2018). This dataset comprises 97 markers genotyped for 110 reference *M. edulis* individuals, 24 reference *Med. M. galloprovincialis* individuals, 6 F1 parents (1 female, 5 males) and 132 F2 offspring. Markers that were not heterozygotic in all F1 parents, or with an allele frequency difference between species lower than 0.2 were removed to avoid spurious distortions and orientation. We also removed two markers with >10% missing data. This left a final dataset of 40 informative markers, and 114 F2 offspring. Alleles were oriented according to their frequencies in reference samples. We then used the R package `qt1` to produce a genetic map (Broman, Wu, Sen, & Churchill, 2003). Four additional markers were dropped by the internal checks in the package, for not passing the Mendelian segregation test in F2s (with Holm-Bonferroni correction). The final genetic map comprises 36 markers scattered among 16 linkage groups (Table S5). Only the first 8 linkage groups contain more than one marker.

An 'unlinked' set of markers was created by keeping the marker with the least missing data in each linkage group or physical contig. Markers not included in the linkage map analysis were considered to be unlinked. See Table S5 for a list of unlinked markers.

2.4 Population differentiation and genetic clustering

We aimed to identify known lineages of the *M. edulis* species complex to assign individual ancestry estimations and filter individuals based on their genetic compositions for downstream analyses.

Population differentiation analysis was used to assess the power of our set of ancestry-informative markers, and to test differences between admixed populations. Genetic clustering was then used to assign individuals to known lineages or to assess levels of admixture in the studied populations.

A principal component analysis (PCA) was performed in R, using the `ade4` package (Jombart, 2008). The genotype data were centred and scaled, with the replacement of missing

data by the mean allele frequencies. Any individuals identified as *M. trossulus* were removed from this analysis.

Hierarchical population differentiation tests were carried out with the R package `hierfstat` (Goudet, 2005). We used 10^4 permutations for all tests. The Weir and Cockerham F_{ST} estimator is reported when presenting population differentiation results. When calculating population differentiation between reference groups, markers with more than 30% missing data in *M. trossulus* populations were removed because of badly typed markers in this species (Table S3).

Ancestry estimation was performed with the Bayesian model implemented in the program `Structure` (Falush, Stephens, & Pritchard, 2003), which includes additional models of interest compared to the aforementioned `Admixture` software. Each result is composed of 25 replicates for each assessed number of genetic clusters, K , run for $8 \cdot 10^4$ steps after a $2 \cdot 10^4$ steps burn-in. The standard deviation for the α prior was set to 0.05 for better mixing of the chains. All analyses use uncorrelated allele frequencies (`FREQSCORR = 0`) and a separate and inferred α for each population (`POPALPHAS = 1`, `INFERALPHA = 1`, Wang, 2017). Replicates were merged with the program `CLUMPAK` (default parameters and MCL threshold set at 0.7) and the major clustering output of the most parsimonious K was used.

For `Structure` analyses, markers that departed from Hardy-Weinberg equilibrium in focal reference populations were removed to avoid departure from the algorithm model. The program was either run using the admixture model with linkage, using the F2 genetic map described above, or using a no-admixture model with the unlinked dataset (Table S5), as both models cannot be used simultaneously.

A first `Structure` analysis on the full dataset was used to remove all individuals with *M. trossulus* ancestry to focus on a 'reduced dataset' of *M. edulis* and *M. galloprovincialis*. Because, *M. trossulus* is present in sympatry in Norway and can hybridise with its congeners, a threshold of 10% ancestry was used to identify parental and most recent hybrid individuals (Table S3). From this reduced dataset, two analyses – with and without the admixture model – were performed (K in 3 to 6). Additionally, to allow a better classification of individuals at bay scales, `Structure` analyses were performed on a 'local dataset' with the ports and surrounding populations, with and without admixture, and without including the reference populations (K in 2 to 5). Finally, specific `Structure` runs with the linkage model were used to estimate the age of the admixture (cf. Supplementary information, section 1). Briefly, admixture linkage disequilibrium allows the estimation of the number of breakpoints per Morgan since the admixture event, r , which can be interpreted as an estimate of the number of generations since a single admixture event (Falush et al., 2003).

Mussels from the admixed populations with Atl. *M. galloprovincialis* (introduced and natural) were classified using the reduced dataset without admixture, using the yellow and grey clusters corresponding to pure Atl. *M. galloprovincialis* and admixed *M. galloprovincialis* respectively ($K = 5$, Figure S19). To obtain a finer classification in port areas, mussels

were assigned to *M. edulis*, Atl. *M. galloprovincialis* or dock mussel clusters using the local Structure analysis without admixture ($K = 3$, Figure S20). See Table S3 for details on the selection thresholds for each group and Figure S21 for independent plots of selected individuals.

The software *Newhybrids* (Anderson & Thompson, 2002) was used to evaluate the probability that individuals were first or second generation hybrids between the dock mussels and native lineages (Figures S26-S27).

2.5 Comparison of ancestry levels

To investigate the similarities and differences in the ancestry compositions of samples from different admixture events and localities (Table 1), we formally tested for variation in ancestry levels.

Independent comparisons were used for admixtures implicating Med. and Atl. *M. galloprovincialis*. For each population of interest, admixed individuals (identified as described in the previous section) were selected and native individuals were removed. The Structure ancestry estimates with admixture, identifying the four clusters *edu_eu_south*, *gallo_atl*, *gallo_med* and *edu_am*, were used ($K = 4$, Figure S21). This selection allowed a homogeneous comparison of ancestry levels between all admixed populations (Figure S23).

A non-parametric Kruskal-Wallis one-way ANOVA was used to test the statistical difference of the four ancestry values (Q) between populations of each admixture type. Additionally, a non-parametric post-hoc pairwise comparisons test was carried out, using the Dwass-Steel-Crichtlow-Fligner test (Crichtlow & Fligner, 1991; Hollander, Wolfe, & Chicken, 2015). We applied Benjamini-Yekutieli corrections for multiple testing.

To test the hypothesis of increased introgression of Med. *M. galloprovincialis* ancestry coming from dock mussels into Atl. *M. galloprovincialis* in the Bay of Brest, native Atl. *M. galloprovincialis* groups from Brittany were identified and their ancestries were compared: (i) mussels distant from the Bay of Brest, Northern Brittany population (*gallo_atl_brit*); (ii) individuals outside the Bay of Brest (the limit being the entrance straight), taken as reference local individuals; and (iii) individuals inside the Bay of Brest classified as Atl. *M. galloprovincialis* with the local Structure without admixture result (Figure S20).

2.6 Least cost distance analyses and Geographic cline fitting

To visualise transitions at the port entrance at the locus level, we fitted clines of allele frequencies along a spatial axis. The objective is to assess the concordance of transitions among markers and with the observed global ancestry.

As a proxy for connectivity between sampling sites, least cost path distance matrices were produced for each port and took into account obstacles such as land and human made barriers (e.g., breakwaters and seawalls). A raster of costs was built for each port from polygon shapefiles ('Trait de côte Histolitt Métropole et Corse V2', produced by SHOM and

IGN) modified to include small port structures that could stop larval dispersal or to exclude inaccessible parts. Locks inside ports were considered as opened for the purposes of distance calculation between isolated points. We used the program QGIS to handle polygons and raster creation. Land was coded as missing data and water was set to have a conductance of one. The R package `gdistance` was used to compute transition matrices based on those cost rasters and to compute least cost distances between points for each dataset (van Etten, 2017).

Geographic clines per SNP were fitted for each port (excluding Saint-Malo which only had one port sample) with the R package `hzar` (Derryberry, Derryberry, Maley, & Brumfield, 2014). The port of Le Havre was divided into two independent transects: North and South corresponding to the historic basins and the 'Port 2000' recent installations respectively. The least cost distance from the most inward site in each port (indicated by a triangle in Figure 3) was taken as a proxy for geographic distance and to project geographic relationships on a single axis. For the Bay of Brest, the starting site was taken as the right-most population in Figure 1g, up the Élorne estuary. The three points in the bottom-right corner of Figure 3e containing *M. galloprovincialis* ancestry were excluded from the fit, to account for discrepancies between least cost path distances and the presence of the dock mussels. Pure *M. edulis* individuals were removed for the analysis in the bay of Brest and Atl. *M. galloprovincialis* individuals for the ports of Le Havre, Saint-Nazaire and Cherbourg. Clines were fitted using a free scaling for minimum and maximum frequency values and independence of the two tails parameters. We used a burn-in of 10^4 and a chain length of 10^5 for the MCMC parameter fit. Only differentiated loci are presented in Figure 4 (panels a-d: allele frequency difference (AFD) > 0.5 , panel e: AFD > 0.3 ; see Figures S28-S32 for details).

2.7 Distortions from expected frequencies and correlations

Our data include multiple admixture events. To ask if outcomes were similar across events, we compared the deviations of marker allele frequencies from their expected values in each situation.

We denote the expected frequency of an allele in an admixed focal population as f_{exp} . This expected value is calculated from the observed allele frequencies in pure-lineage reference populations, and from the mean ancestry values across all markers for the focal population, as estimated from `Structure`.

Admixed population frequencies are calculated only with admixed individuals in each population (see section 2.5 for details and Figure S21 for selected individuals). We used the results of ancestry estimation from `Structure` with $K = 4$ clusters (`edu_eu_south`, `gallo_atl`, `gallo_med`, and `edu_am`) and summed ancestries from South-Eu. and American *M. edulis*,

giving the composite ancestry estimation Q_{edu} for each individual:

$$Q_{edu} = Q_{edu_eu_south} + Q_{edu_am} \quad (1)$$

In particular, with three reference populations, the expected allele frequency is:

$$f_{exp} = f_{local_edu} \cdot \bar{Q}_{edu} + f_{gallo_atl} \cdot \bar{Q}_{gallo_atl} + f_{gallo_med} \cdot \bar{Q}_{gallo_med} \quad (2)$$

Here, f values denote the allele frequencies in the reference population indicated by the subscript, and the Q -values denote the mean ancestry from the focal admixed population. gallo_med and gallo_atl correspond to the L2 level encompassing lower population classifications (fig 1c and Table S4) as the precise origin of the parental populations are not known below this level.

For lab crosses, the parental Med. *M. galloprovincialis* L3 level is known and corresponds to gallo_med_west. Therefore its frequency was used in place of f_{gallo_med} . For dock mussels the 'local edu' lineage is taken to be the South-Eu. *M. edulis* one (edu_eu_south). For LOF and GAS admixed populations, we used the North-Eu. *M. edulis* lineage (edu_eu_north) to estimate parental allele frequencies (f_{local_edu}) while using the usual \bar{Q}_{edu} estimation.

The deviation of the observe frequency f_{obs} from the expected frequency f_{exp} is defined as:

$$D = f_{obs} - f_{exp} \quad (3)$$

This computation allows us to estimate a distortion by locus from the average genomic expectation given the population ancestry and parental allele frequencies. The correlation of distortions by locus are then computed between admixed populations, corresponding to different admixture events (e.g. between one dock mussel and one Norway admixed population). For each correlation, we used Pearson's r to estimate the strength of the correlation and tested the significance with a permutation test ($5 \cdot 10^4$ permutations). The classic t-test was not used due to the distortions not following normality.

When multiple correlations pertained to the same null hypothesis (e.g. that distortions in lab backcrosses do not correlate with distortion in ports), and datasets contained possible non-independence (e.g., from migration of hybrids between ports), we used a modified Fisher's method to combine p values, developed by Poole, Gibbs, Shmulevich, Bernard, and Knijnenburg (2016) and implemented in the R package `EmpiricalBrownsMethod`.

3 Results

3.1 Differentiation between lineages and characterisation of admixed populations

We collected or reanalysed samples from several locations, with known or suspected admixture between different species or lineages of *Mytilus* mussels (Figure 1c, Table 1).

We first verified that our dataset could distinguish between species and focal lineages. Hierarchical genetic differentiation tests based on putative reference groups (Figure 1c, Table S4) showed significant F_{ST} distances until the grouping level L3. F_{ST} ranges between 0.72 and 0.81 at the species level (L1), between 0.38 and 0.48 for L2 levels within species and between 0.0024 and 0.31 for L3 levels within L2 (see Table S6 for details; note that our SNP panel is enriched for ancestry-informative SNPs and so these values should not be interpreted as genome-wide averages).

Initial PCA and Structure analyses identified the presence of all three *Mytilus* species. However, *M. trossulus* was present in only a few populations (i.e. Norway, North Sea), consistent with previous knowledge of its range (Figure S5). Because *M. trossulus* is not centrally relevant to the present work, individuals with more than 10% *M. trossulus* ancestry were removed from subsequent analyses.

After removing *M. trossulus* individuals, both the PCA (Figure 1a-b) and the Structure Bayesian clustering ($K = 4$, Figures S6-S15) show a clear differentiation between the parental lineages (edu_am, edu_eu_south, gallo_atl and gallo_med). Both methods also allow us to identify and further characterise three characteristic patterns of admixture in our data, which we called 'naturally admixed', 'Norway admixed' and 'dock mussels'. We describe each of these in detail below.

Each admixed pattern was further investigated by comparing ancestry estimations of populations to characterise the variation between locations (Structure Q -values, $K = 4$, Figure S23).

3.2 Natural hybridisation

Several samples are the result of natural admixture between Atl. *M. galloprovincialis* and South-Eu. *M. edulis* and are called 'naturally admixed' (Figure 1c, Table 1). This category includes geographically distant samples from Scotland (ABD), the English Channel island of Jersey (JER), the Murchison oil platform in the Norwegian Sea (MCH) and the natural hybrid zone in South Brittany (HZSB, Figure 2). As far as we know, these groups are free from human-mediated introductions.

Naturally admixed populations cover much of the range of admixture proportions observed between the two parental species (Figure S23). These four populations exhibit significant differences in their Atl. *M. galloprovincialis* ancestry, with the exception of the MCH/HZSB

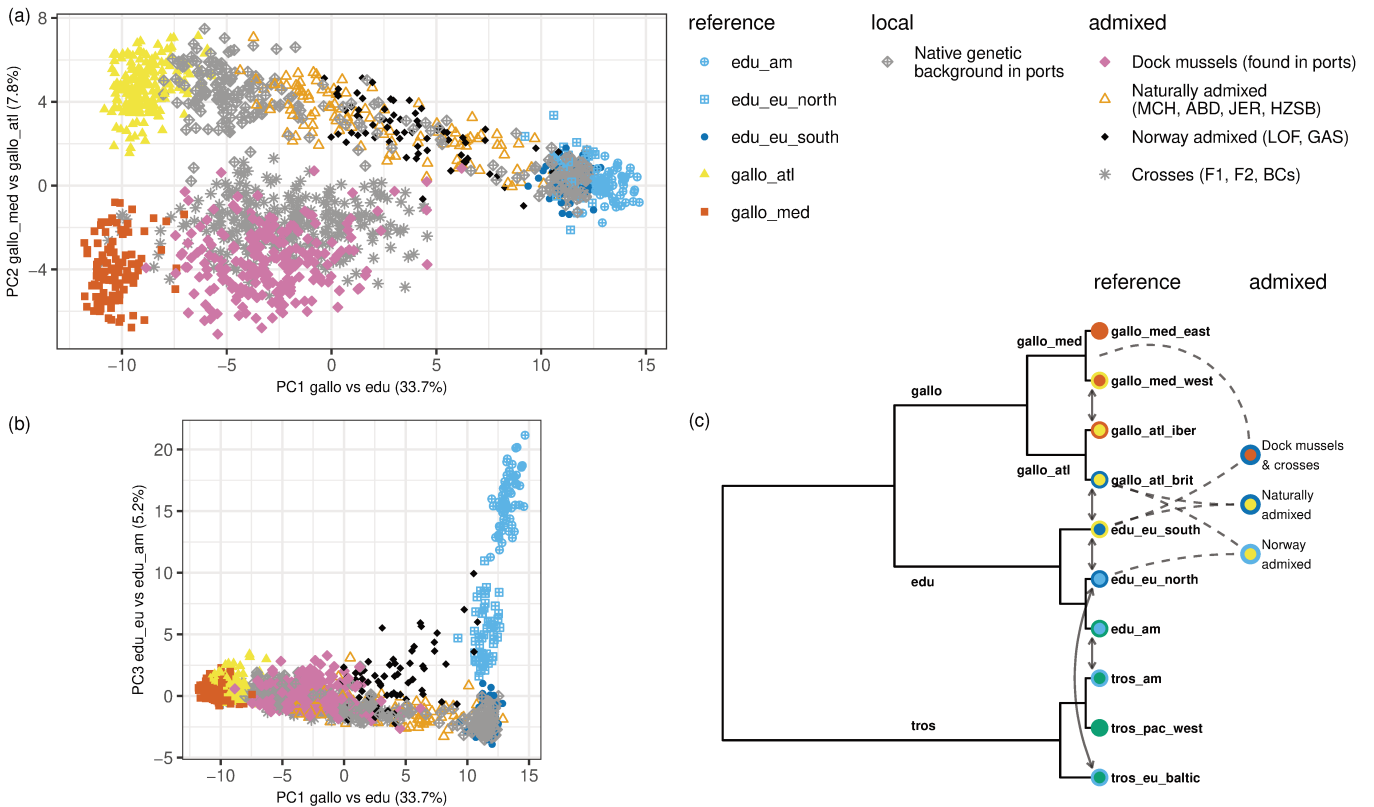


Figure 1: (a)-(b) Principal Component Analysis of reference samples and studied groups (*M. trossulus* samples were not considered). Locations in and around ports have been randomly sub-sampled for visual clarity (500 out of 1930 individuals retained) and individuals were classified as native genetic backgrounds (grey diamonds) or as dock mussels (pink diamonds) on the basis of a Structure analysis. The ports of interest are Le Havre, Cherbourg, Saint-Malo, Brest and Saint-Nazaire; see Figures 2 and 3 for details. (c) Schematic tree of lineage relationships presenting group names and colour schemes. External circle colours and arrows represent known local introgression between *Mytilus* spp. lineages. The three admixture types studied are presented in the right column.

comparison (Table S10). JER is the most *M. edulis*-like population while MCH and ABD are the most *M. galloprovincialis*-like, with HZSB being the most variable one. Interestingly, JER exhibit a homogeneous excess of South-Eu. *M. edulis* ancestry, contrasting with the Atl. *M. galloprovincialis* ancestry excess of the three other natural populations (Figures 2 and S23). Atl. *M. galloprovincialis* ancestry excess is usually observed in contact zones reflecting the asymmetric introgression with South-Eu. *M. edulis* (Fraisse et al., 2016).

3.3 Admixed populations in Norway

We named a second admixture pattern 'Norway admixed', because it includes two Norwegian populations (LOF, GAS). These admixed mussels involve Atl. *M. galloprovincialis* and North-Eu. *M. edulis* (Figure 1b), and are defined as non-indigenous (Mathiesen et al., 2016). LOF and GAS do not differ significantly at any of the four different ancestry estimates (Table S10).

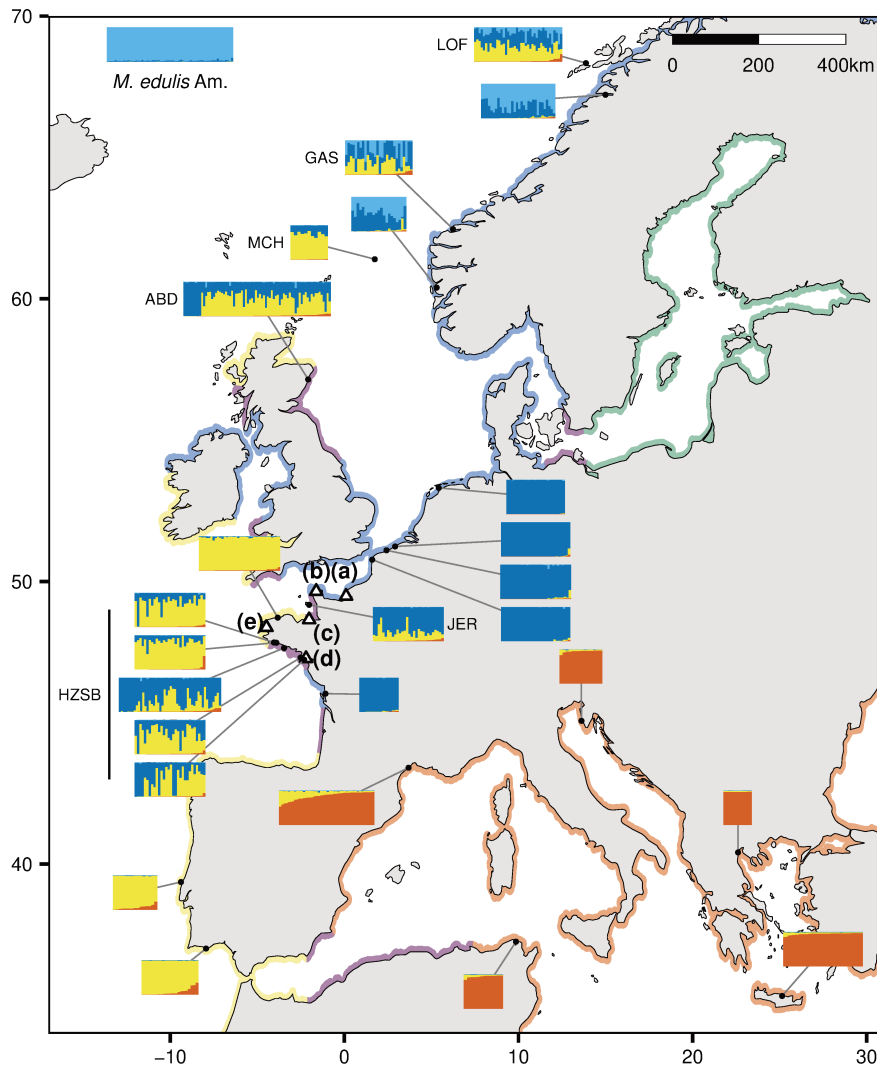


Figure 2: Location and ancestry composition of sites for reference and admixed populations. Barplots represent ancestries of individuals from the focal site, estimated by Structure with $K = 4$. In all barplots, individuals have been sorted from left to right by their level of Mediterranean *M. galloprovincialis* ancestry. Coloured coastlines indicate the approximate distribution of parental genetic background, with colour code as used in Figure 1. Hybrid zones are coloured in purple. Points (a)-(e) correspond to the ports of Le Havre, Cherbourg, Saint-Malo, Saint-Nazaire and Brest respectively, which are detailed in fig. 3.

These admixed mussels are on average composed of 40% Eu. *M. edulis* ($SD = 15.82$, $N = 63$), 16% American *M. edulis* ($SD = 15.35$), 41% Atl. *M. galloprovincialis* ($SD = 13.91$), and 3% Med. *M. galloprovincialis* ($SD = 3.83$) (Figures S21 and S23). The presence of individuals with some Atl. *M. galloprovincialis* ancestry was also confirmed in Svalbard (Figure S14; Mathiesen et al., 2016). On average, admixed mussels in Svalbard have lower proportions of Atl. *M. galloprovincialis* ancestry. These individuals were not used in downstream analyses, due to their small number.

Norway admixed populations were also compared to naturally admixed populations given they both involve the Atl. *M. galloprovincialis* lineage. Nearly all pairwise comparisons of

the Atl. *M. galloprovincialis* ancestry are significantly different, with the exception of the GAS/JER comparison (Table S10). GAS and LOF appear to be more similar to JER, with an excess of *M. edulis* ancestry, than they are to the other three naturally admixed populations.

3.4 Dock mussels

3.4.1 An admixture between geographically distant lineages

We identified a group that we labelled 'dock mussels', found in five French ports, and more rarely in their vicinity. They exhibit a characteristic admixture between Med. *M. galloprovincialis* and South-Eu. *M. edulis*, and are defined as the intermediate cluster between these two lineages (Figure 1, Table. 1). The selection of individuals defined as dock mussels is based on a Structure analysis without admixture (Figure. S20). Dock mussels are closer to Med. *M. galloprovincialis* than to *M. edulis* in the PCA, reflecting the estimated ancestries, and are not differentiated by other axes of the PCA (Figure 1a). Additionally, they show a large variance in all directions, presumably including inter-specific hybrids with *M. edulis* and inter-lineage hybrids with Atl. *M. galloprovincialis*. It is noteworthy that apart from the dock mussels, and the lab crosses between Med. *M. galloprovincialis* and South-Eu. *M. edulis*, no other population clusters in this region of the PCA (i.e. intermediate placement between Med. *M. galloprovincialis* and South-Eu. *M. edulis*). This implies that no natural hybridisation is observed between these two lineages in our dataset. This is in accordance with the distribution of the *Mytilus* lineages (Figure 2).

We analysed three other large ports to search for dock mussels, but none showed the presence of this class of mussels: La Rochelle (France, Figure S16), Bilbao (Spain, Figure S17) and New York city (USA, Figure S18).

In the five colonised ports, individuals of native parental genetic backgrounds are found in addition to dock mussels (Figures 1a-b and 3). These native mussels are (i) Pure South-Eu. *M. edulis* around Cherbourg, Le Havre and Saint-Nazaire, and (ii) Pure Atl. *M. galloprovincialis* from Brittany around Brest, Saint-Malo and Saint-Nazaire. We also observed intermediate individuals between Atl. *M. galloprovincialis* and *M. edulis* corresponding to admixed individuals or hybrids in the Bay of Brest area, Saint-Nazaire and Saint-Malo. All of these locations are in or close to natural hybrid zones between those two species, while the aquaculture of *M. edulis* in the Bay of Brest, imported from the Bay of Biscay, is an additional source of *M. edulis* in this area, especially since dispersing larvae from aquaculture sites are common (for details see Figure S11).

In term of estimated ancestries (Structure Q -values), dock mussels are on average composed of 25% Eu. *M. edulis* ($SD = 11.17$, $N = 879$), 69% Med. *M. galloprovincialis* ($SD = 11.85$), 4% Atl. *M. galloprovincialis* ($SD = 6.08$) and 2% American *M. edulis* ($SD = 3.04$) (Figure S21). Allele frequencies of dock mussels for markers differentiated between *M. edulis* and Med. *M. galloprovincialis* are also consistent with the observed levels

of admixture, and are strongly concordant between markers (Figure S22). All port populations are highly similar, both spatially and temporally, in their variance of allele frequencies regardless of their overall level of introgression (Figure S22).

When comparing ports, Cherbourg, Saint-Nazaire and Saint-Malo are the least introgressed populations (Figure S23, Table S11). Le Havre appear to be the most introgressed by South-Eu. *M. edulis*. Brest also have reduced levels of Med. *M. galloprovincialis* ancestry, equivalent to what is found in Le Havre, but due to an excess of Atl. *M. galloprovincialis* ancestry. Cherbourg, Saint-Malo and Saint-Nazaire do not differ significantly in South-Eu. *M. edulis*, Atl. and Med. *M. galloprovincialis* ancestries, despite the fact they are in different native species contexts.

For the port of Cherbourg, we were able to analyse several temporal samples between 2003 and 2017 (Figure 3b). These exhibit a small differentiation between the 2003 sample and later years (2015 and 2016; $F_{ST} = 0.0066$ and 0.0097 , Table S8) and this seems to be driven by a small increase in Med. *M. galloprovincialis* ancestry in 2015 and 2016 (significant only between 2003 and 2016, Table S12). The only other historical sample in our collection was a site in the Bay of Brest that showed the absence of dock mussels in 1997 (Pointe de L'Armorique, PtArm97, fig S11). However, this area also exhibited only one dock mussel genotype 20 years later (Brest-24).

3.4.2 Dating the admixture of dock mussels

To estimate the age of the admixture event which resulted in the dock mussels, we inferred levels of linkage disequilibria (Figure S24). Disequilibria were present, but at low levels indicating that there had been several generations of recombination since admixture. We computed a linkage map from the lab produced F2, and found that it was consistent with the disequilibria present in the dock mussels. Using this map, and the linkage option in the Structure package, we estimated the admixture time to be between 4 and 14 generations, depending on the port (Table S14 and supplementary methods).

As survival and lifetime are highly variable and environment dependent in mussels, it is difficult to translate these estimates into clock time. However, given that mussels reach maturity at ~1 year and have a high early life mortality rate, 1-2 years seems a reasonable estimate of the generation time, dating the admixtures at between 4 and 28 years ago. We note that our oldest sample from Cherbourg in 2003, provides one of the oldest estimates, and so could not be used to calibrate a 'recombination clock'.

3.4.3 Dock mussels are spatially restricted to ports

The individual ancestries were plotted spatially to assess their distribution in and around the five studied French ports (Figures 3).

The ports of interest are localised in regions characterised by different native species (Fig-

ure 2). The native species around Le Havre and Cherbourg is South-Eu. *M. edulis* while in the Bay of Brest, the native mussels are Atl. *M. galloprovincialis* (Figure 3). Saint-Malo and Saint-Nazaire lie on the limits of hybrid zones between *M. edulis* and *M. galloprovincialis*. However, surroundings of Saint-Malo are mostly inhabited by Atl. *M. galloprovincialis* (Figure 3c), and Saint-Nazaire is located in a zone mostly composed of *M. edulis* with the presence of Atl. *M. galloprovincialis* in sympatry (Figure 2 and 3d). Around the latter, local *M. galloprovincialis* are more introgressed by *M. edulis* than those found in Brittany as they lie at the far end of the South Brittany hybrid zone (Bierne et al., 2003).

Four of the five studied ports (all except Brest) have locked basins where the dock mussels were found. Importantly, dock mussels are nearly all localised inside port infrastructures, and we observed a sharp shift at the port entrance (Figure 3). For the ports of Saint-Nazaire, Saint-Malo, Cherbourg and Le Havre only four individuals with Med. *M. galloprovincialis* ancestry were detected in coastal wild populations (out of 341 individuals presented in Figure 3). Those individuals were observed at distances between a few hundred meters to 30 km from the entrance of the ports.

In the opposite direction (from the natural coast to the port), we mainly find native migrants close to the port entrance inside Le Havre, Cherbourg and Saint-Nazaire (Figure 3). Le Havre and Saint-Nazaire are the ports containing the largest number of *M. edulis* migrants, yet Le Havre is the only one where F1 hybrids between dock mussels and *M. edulis* have been observed (identified with Newhybrids, Figure S26).

The Bay of Brest is of particular interest for two reasons (Figure 3e1-e2). First, the local background is the Atl. *M. galloprovincialis* lineage, contrasting with the other ports where the native background is *M. edulis* (with the exception of Saint-Malo), and exhibiting higher sympatry inside port infrastructure than anywhere else. Second, mussels with a typical dock mussel admixed genetic background have been detected outside port infrastructures, which motivated an extensive sampling. Contrary to the other ports, dock mussels extensively colonised the local environment, mainly inside and close to estuarine areas.

Dock mussels are, however, restricted to the inside of the bay with no detectable influence on external *M. galloprovincialis* populations. We compared several groups of Atl. *M. galloprovincialis* from Brittany (away, close and inside the Bay of Brest) to assess the potential introgression from dock mussels to the local populations. While levels of *M. edulis* ancestry increased and levels of Atl. *M. galloprovincialis* decreased significantly from distant populations to inside the Bay of Brest, the levels of Med. *M. galloprovincialis* ancestry did not differ significantly (Table S13). Nonetheless, we note that the tail of the distribution of Med. *M. galloprovincialis* ancestry in the Bay of Brest is skewed towards higher values (Figure S23). This tail is due to the presence of hybrids between dock mussels and the local native Atl. *M. galloprovincialis* (Figure S27).

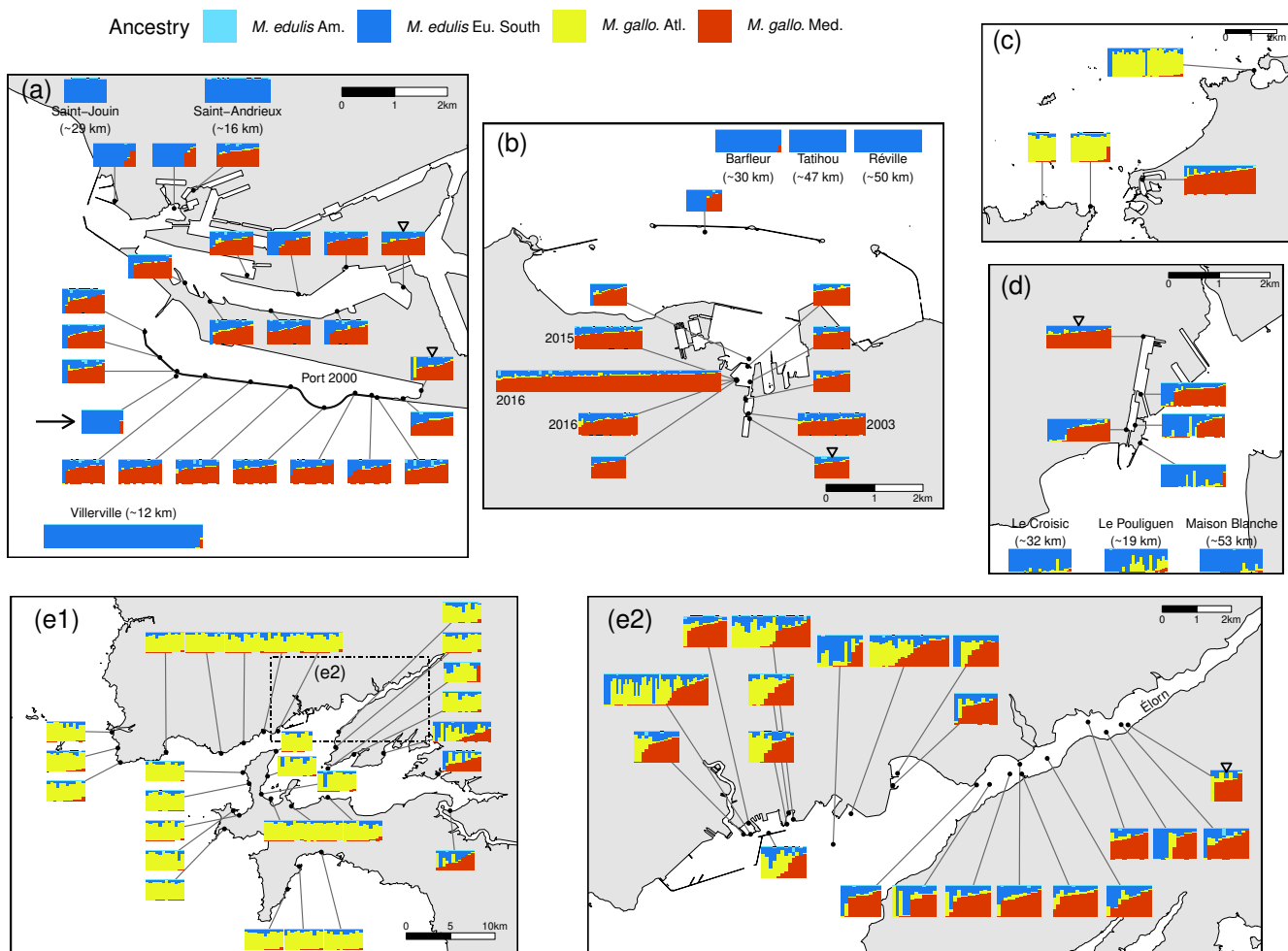


Figure 3: Ancestry composition of sites for each port. As in Figure 2, barplots represent the ancestry estimation for individuals at the indicated locations and are ordered from left to right by their Med. *M. galloprovincialis* ancestry. Barplots at the map edges correspond to distant populations with the least cost path distance from the port indicated in parentheses. The inner-most populations used to fit geographic clines are indicated by the reversed triangles. (a) Le Havre; note that the two distinct main basins (North and South-Port 2000) found in this port were separated for geographic cline analyses; the arrow indicates a site located on the estuary side of the dyke, characterised by a majority of *M. edulis* individuals. (b) Cherbourg; dates indicate collection year; all other samples were collected in 2017. (c) Saint-Malo. (d) Saint-Nazaire. (e1) Bay of Brest. (e2) Detailed map of the port of Brest and the Éloron estuary, which corresponds to the inset rectangle in panel (e1).

3.4.4 Geographic clines show sharp and concordant transitions at the port entrance

Allele frequencies shift sharply at the entrance of ports (Figure 4a-d) and clines are highly concordant both between markers and with the mean ancestry cline (red line). Compared to the reference Med. *M. galloprovincialis* frequencies, dock mussels show a global decrease of allele frequency due to a genome wide introgression from the local species.

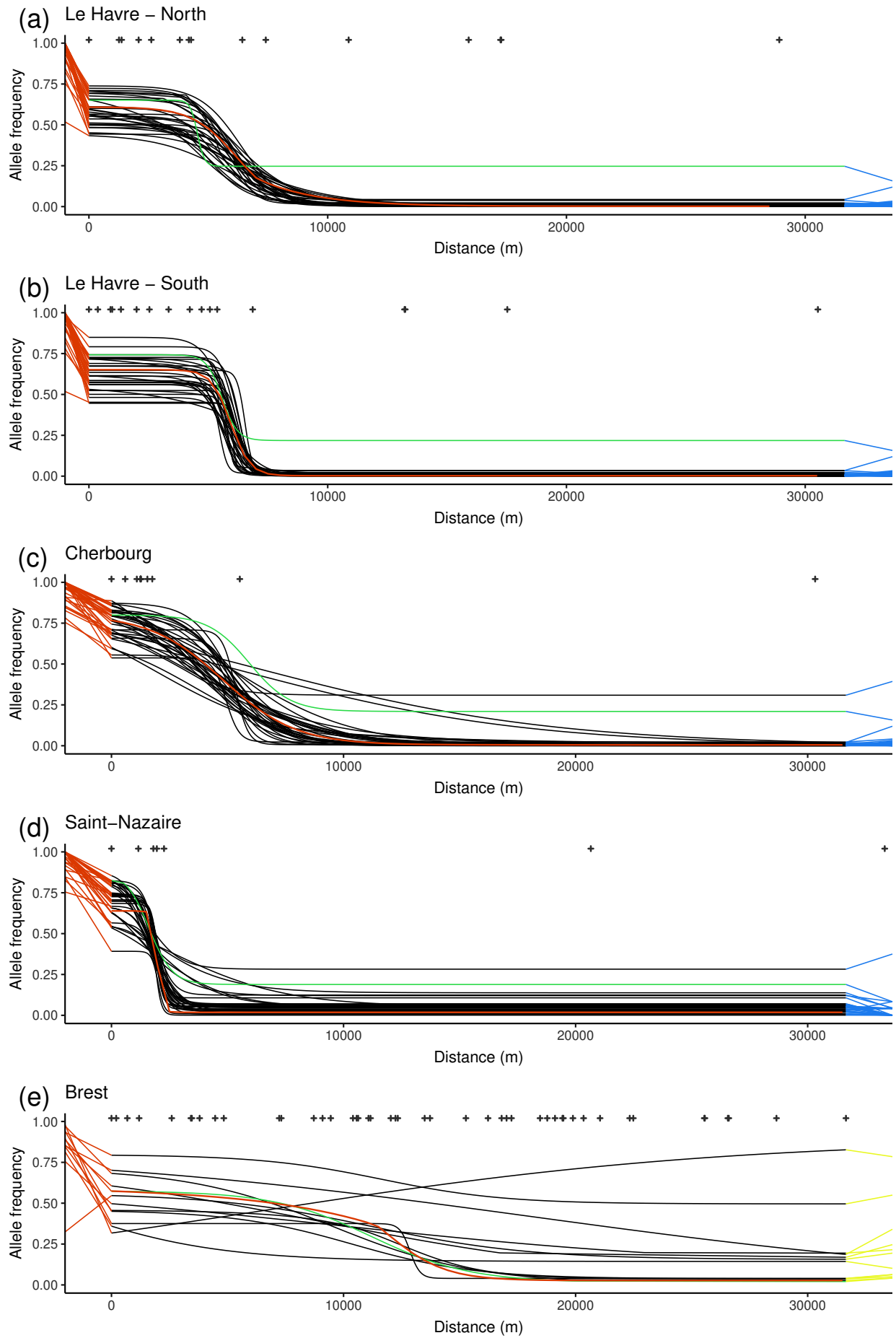


Figure 4: Geographic clines computed with the package *hzar* in each study ports (except St-Malo, see text). The x-axis is the distance from the most inward point (reversed triangles in figure 3) determined by a least-cost path analysis. Top crosses indicate the distance of each site considered. For representation purposes, some distant points are not displayed, but were used in the cline fit. Only alleles with a frequency difference of 0.5 between left-most port population and sea-side reference are presented (except for panel (e) where the threshold is 0.3), each with a distinct black line. For each marker, left and right segments join the frequency fitted at the end of the cline to the frequency observed in reference populations, with Med. *M. galloprovincialis* in orange and South -Eu. *M. edulis* in blue (or Atl. *M. galloprovincialis* in yellow). For (a)-(d), references are Mediterranean *M. galloprovincialis* on the left and *M. edulis* on the right. For (e), the right hand side reference is the local Atlantic *M. galloprovincialis*. The orange cline is the mean cline computed from the Mediterranean *M. galloprovincialis* Q-value from Structure, in mean proportion of ancestry. The cline of the female mitochondrial marker (601) is shown in green. (a) Le Havre, North transect (historic basin). (b) Le Havre, South transect (Port 2000). (c) Cherbourg. (d) Saint-Nazaire. (e) Bay of Brest.

Clines have narrow widths across all ports. Average widths are 3.99 km ($SD = 1.80$) and 1.30 km ($SD = 0.52$) for the North and South transects of Le Havre respectively (Figure 4a-d); 7.37 km ($SD = 5.38$) in Cherbourg (Figure 4c); 2.16 km ($SD = 2.15$) in Saint-Nazaire (Figure 4c), and 18.51 km ($SD = 14.03$) in the Bay of Brest (Figure 4e).

The difference between the North and South transects in Le Havre is best explained by the presence of more *M. edulis* or hybrid individuals at the entry of the North basin (Figure 3a). The interpretation in the Bay of Brest is more difficult due to two factors. First, the spread of dock mussels and sympatry with local ones in several populations make allele frequencies more variable between close populations (Figure 3e-f). Second, we had a reduced number of differentiated markers between Atl. and Med. *M. galloprovincialis* in our dataset with lower level of differentiation.

3.5 Repeatability of allele frequency deviations between admixture events

If admixture events are non-independent (e.g., due to migration between ports), or if admixture events are independent, but lead to repeatable patterns of natural selection, then we would expect to see the same alleles over- or under-represented in different locations.

We cannot compare allele frequencies directly, because different locations are characterised by different overall levels of ancestry. Therefore, for each marker, in each location, we calculated its deviation from expected values. These expected frequencies were calculated from the allele's frequencies in the reference parental populations, combined with the overall levels of ancestry in the sampled location (this is Barton's concordance analysis, eq. 1-3).

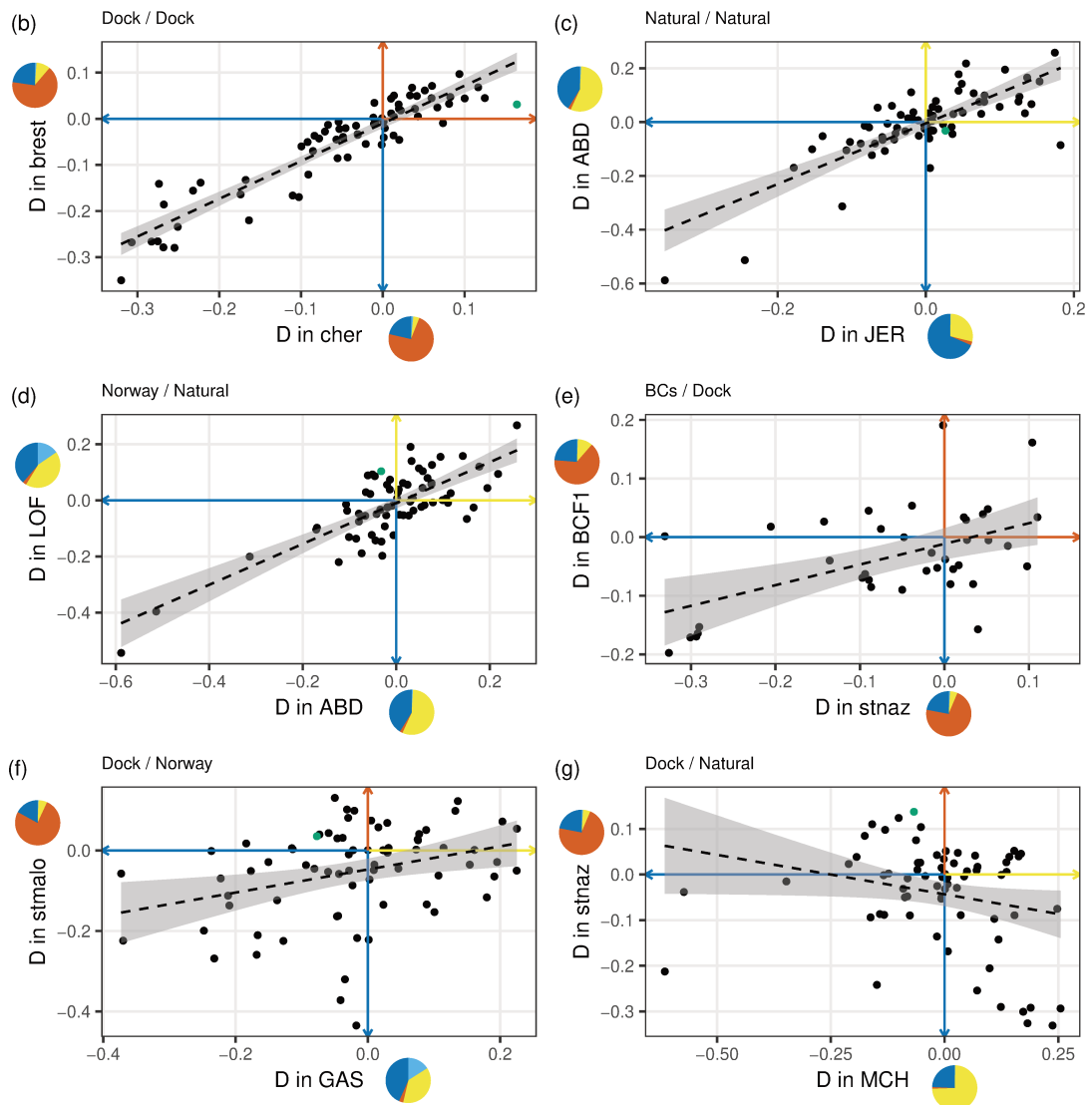
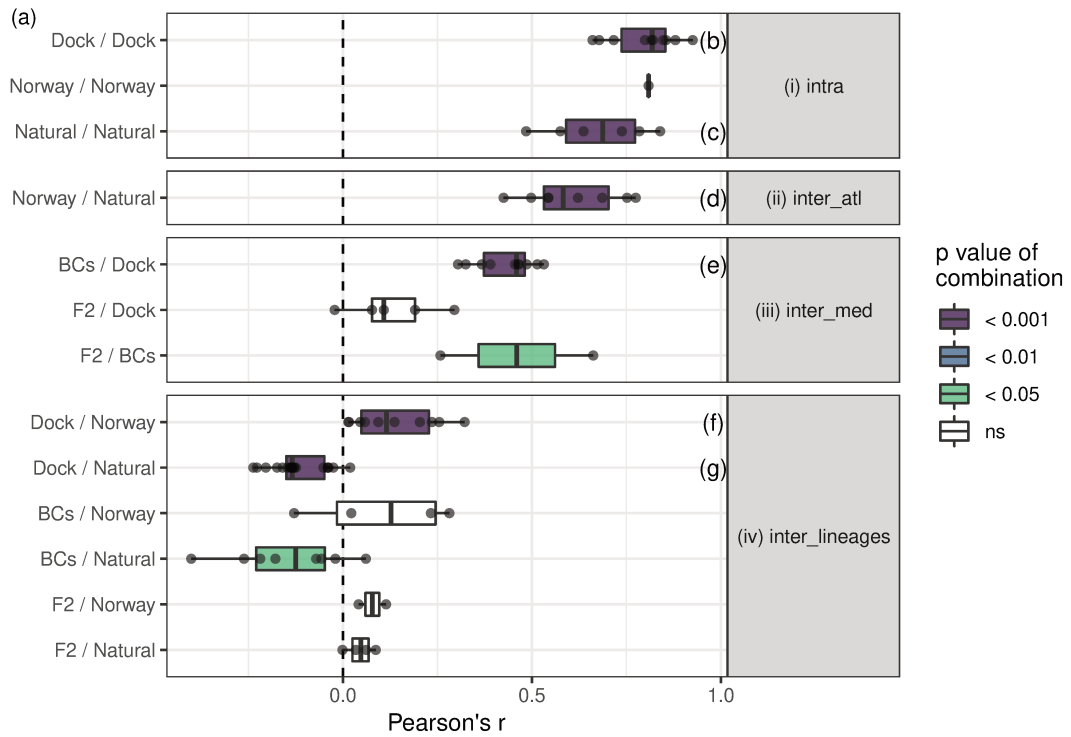


Figure 5: (a) Pearson's r correlation coefficients of distortions (D) between groups of admixture types. The admixture types are: dock mussels (Dock), Norway admixed (Norway), naturally admixed (Natural) or crosses (BCs and F2). Each grey dot is a correlation between two sites (e.g. havre vs. cher is one of the point shown in the Dock/Dock row or BCF1 vs. MCH in the BCs/Norway row). The significance level correspond to the combination of p values among comparisons (see methods). Four types of comparisons were tested: (i) intra - comparisons among the same types of admixture events; (ii) inter_atl - comparisons of the admixture events between Atl. *M. galloprovincialis* and *M. edulis*; (iii) inter_med - comparisons of the admixture events involving South-Eu. *M. edulis* and Med. *M. galloprovincialis*; (iv) inter_lineages - comparisons of admixture events between different backgrounds. Panels (b)-(g) at the bottom show examples of correlations between distortions computed in two locations, for the highest significance levels per type comparisons (purple colour in panel [a]). All correlations presented are significant and linear models with 95% confidence intervals are plotted. The colour of the axis shows the direction of the distortion in term of lineage, using the colour code shown in Figure 1. Pies show the mean ancestry composition of the population considered. Distortion corresponding to the mitochondrial marker (601) is highlighted in green in panels (b)-(g).

Examination of these allele frequency deviations showed some suggestive similarities between admixture events. For example, the mitochondrial marker (601) is differentiated between the Med. and the Atl. *M. galloprovincialis* lineages (Figure S38). This locus exhibits large distortions (D) towards the Med. *M. galloprovincialis* lineage in Le Havre, Cherbourg and Saint-Nazaire (0.11, 0.16, 0.13 respectively), while displaying smaller distortions in Brest and Saint-Malo (0.03 in both cases).

More formally, the repeatability of admixture events can be assessed by correlating the complete set of allele-frequency deviations between events. Four types of comparisons corresponding to differences in implicated lineages are presented in Figure 5.

We examined all pairwise comparisons involving the same parental backgrounds in similar conditions (Figure 5a-[i]): the five dock mussels populations from French ports ('Dock / Dock'), the two Norwegian introductions ('Norway / Norway'), and the natural hybrid zones ('Natural / Natural'). In each case, the allele frequency deviations are significantly and positively correlated between events, with large to medium effect sizes (Figures 5 and S33-S34). The same was also true when we compared the Norwegian introductions to the natural hybrid zones involving the same *M. galloprovincialis* genetic background ('Norway / Natural', Figure 5a-[ii]). Remarkably, strong correlations were also observed when we compared dock mussels to lab crosses involving the same lineages (Figure 5a-[iii]). The correlations were strongest for lab backcrosses (BCs), and much weaker and non-significant for the F2. This is consistent with the genetic makeup of the dock mussels, which have hybrid indexes closer to BC genotypes than to F2s (fig S23 and S25), albeit more recombined.

Globally, the level and consistency of correlations increases with the similarity between admixture events (from groups [iv] to [i] in Figure 5). Panels (i)-(iii) indicate that admixture events

of different kinds can lead to strongly repeatable results. But this is only true when the same genetic backgrounds are involved. To show this, Figure 5a-(iv) shows results from pairs of admixture events involving different backgrounds (e.g. Dock vs. Norway admixture). In this case, effect sizes are small to medium, and sometimes negative.

3.6 Additional putative anthropogenic introductions

While the overall genetic composition of many of our sampled populations was as expected, we also obtained some isolated but unexpected results which we report in the following section. First, the port of New York showed higher levels of South-Eu. *M. edulis* ancestry, up to 30%, compared to other populations from Long Island Sound (Figure S18). Therefore we cannot exclude the possibility that there has been an introduction of Eu. *M. edulis* in or close to the port of New York.

Second, outside of ports, multiple long distance migrants from different origins were identified. The reanalysis of the Coolen (2017) samples did not show any pure *M. galloprovincialis* individuals (Figure S13). However, one population contained six individuals composed of 10 to 30% Med. *M. galloprovincialis* ancestry (Q13A, Figure S13). This population is located offshore, at around 25 km from the entry of the port of Rotterdam, which is the largest commercial port of Europe. Given the greater proportion of migrants at this distance, as compared to results from other ports, the presence of dock mussels in Rotterdam is highly probable and will require further investigation.

Similarly, one population in the bay of Biscay, on the Atl. coast of Oléron Island (ROC_VER), contained an individual with pure Med. *M. galloprovincialis* ancestry and a few individuals with some levels of Med. *M. galloprovincialis* ancestry in an Atl. *M. galloprovincialis* background. Those latter individuals might plausibly be migrants Atl. *M. galloprovincialis* from the Basque Country. Indeed, unlike populations from Brittany, Iberian Atl. *M. galloprovincialis* populations south of the last hybrid zone with *M. edulis*, have low to medium levels of Med. *M. galloprovincialis* ancestry due to their contact with this lineage in the South (see Bilbao port samples, Figure S17 and classification as Atl. *M. galloprovincialis* in Figure S19).

Other unexpected ancestries were observed in other locations. For example, we found at least one Atl. *M. galloprovincialis* in the port of Le Havre (LeHa_P11, Figure S8). We also report here the presence of an F1 hybrid between *M. edulis* and Atl. *M. galloprovincialis* in the port of Sète (France, Mediterranean coast) despite the fact that neither of these lineages are found in this area. We also analysed two samples from a ferry hull collected in 2011 and 2013. The ferry crosses the English Channel between a *M. galloprovincialis* region in Brittany (Roscoff) and a hybrid zone in the UK (Plymouth) where *M. edulis* and *M. galloprovincialis* are found in sympatry (Hilbish, Carson, Plante, Weaver, & Gilg, 2002, and personal communication). Both samples showed a mixture of *M. edulis* and Atl. *M. galloprovincialis* individuals (Figure S15, Fer11 and Fer13), highlighting once again the role of ship traffic in the displacement of species

and their role as meeting points where hybridisation can occur.

We also detected a signature of Atl. *M. galloprovincialis* in the northern English Channel, and southern North Sea, indicating the presence or recurrent migration of Atl. *M. galloprovincialis* in those regions (Dieppe, Ostende, Ault, Dunkerque 'Dun', Figure S5).

Finally, one population from Korea (KOR, Figure S15) is completely composed of pure Med. *M. galloprovincialis*, corresponding to the known introduction in Asia (Han et al., 2016; McDonald et al., 1991). Another study showed that the introduction on the Pacific coast of the USA was similarly composed by pure Med. *M. galloprovincialis* (Simon et al., 2019). Those observations preclude the idea that previously observed Med. *M. galloprovincialis* introductions are related to dock mussels.

4 Discussion

We have uncovered a singular type of mussels in five ports in Western France. These dock mussel populations display a recent admixture pattern between non-native Med. *M. galloprovincialis* and South-Eu. *M. edulis*. While secondary admixture also occurred with genetic lineages encountered locally, dock mussels exhibit a high level of similarity between ports. In addition, our spatial sampling in ports allowed us to document the striking confinement and association of these genotypes to the interior of the ports, resulting in narrow shifts at port entrances. Some variation to this observation was, however, observed between ports, potentially due to their different layouts and conditions. Based on these results, we assume that dock mussels have been introduced.

By including and reanalysing *M. galloprovincialis* populations in Norway, experimental crosses, and newly identified admixed populations from several sites, we were able to compare admixture patterns between equivalent situations as well as between different genetic backgrounds and thus investigate the extent of parallelism in such secondary admixture processes.

4.1 The introduction of dock mussels and the timing of admixture

Dock mussels constitute homogeneous populations composed of around 70% Med. *M. galloprovincialis* ancestry, which may sometimes be called a 'hybrid swarm' due to a uni-modal distribution of hybrid indices and a complete mixing of ancestries along the genome (Allendorf, Leary, Spruell, & Wenburg, 2001; Beninde, Feldmeier, Veith, & Hochkirch, 2018; Jiggins & Mallet, 2000). We additionally show that there is ongoing secondary admixture between the dock mussel cluster and native genetic backgrounds, exemplified by the detection of F1 hybrids in Le Havre (Figure S26). While no F1 hybrids have been identified in the Bay of Brest by Newhybrids (Figure S27) – which most probably results from reduced power of identification between the two *M. galloprovincialis* lineages – the distribution of ancestries observed leaves little doubt that hybridisation is ongoing between dock mussels and Atl. *M. galloprovincialis*

(Figures 3 and S23). Given the possibilities of local admixture, the relative global homogeneity of dock mussels could be explained either by the recentness of the introduction, by the existence of extrinsic or intrinsic barriers to introgressions, or by both.

The evidence of limited natural dispersal outside ports, presented in this study, provides a strong case for a saltatory colonisation of ports through human-mediated 'jump dispersal'. In our view, the most parsimonious hypothesis of colonisation involves an initial admixture between pure Med. *M. galloprovincialis* and South-Eu. *M. edulis* in a yet unknown location, followed by secondary events of anthropogenically mediated dispersal. Both the genetic homogeneity of dock mussels and the absence of pure parental Med. *M. galloprovincialis* in all sampled ports provide arguments for this hypothesis. For instance in the Bay of Brest or in Saint-Malo, the presence of dock mussels with similar genetic compositions to the other ports (Figure S23), where the local native species is however different (i.e., predominantly Atl. *M. galloprovincialis* rather than *M. edulis*), suggests that the admixture with *M. edulis* happened before the introduction of dock mussels in these ports.

Ship traffic is thus likely to be the main source of these introductions to ports. The five studied infrastructures are large commercial and military ports that may have facilitated the primary introduction of mussels (C. L. Hewitt, Gollasch, & Minchin, 2009; Sylvester et al., 2011). Given the presence of marinas in the vicinity of the large studied ports and their colonisation by dock mussels, they constitute a possible way of secondary expansion at a regional scale. Indeed, marinas and associated activities, e.g. leisure boating, have been shown to contribute to regional NIS expansion (Clarke Murray, Pakhomov, & Therriault, 2011) and create chaotic genetic structure in both native and non-native species inhabiting these artificial habitats (Guzinski, Ballenghien, Daguin-Thiébaud, Lévêque, & Viard, 2018; Hudson, Viard, Roby, & Rius, 2016). For now, in the Bay of Brest, only the marinas close to the large port contained dock mussels. The other marinas outside of the bay (e.g. Camaret and Morgat, Figure S11 Brest-11 and 13 respectively) – potentially exchanging a lot of traffic with Brest marinas – did not, and this supports the absence of a secondary introduction. Colonisation seems therefore so far limited to large port infrastructure, and nearby marinas, with dispersal due to large vessel traffic. This situation might nonetheless change over time, and genetic monitoring should be pursued.

We have estimated an admixture time for dock mussels of 4 to 28 years ago. In addition to the inherent difficulty of this dating and the limitation of our dataset, we note that this estimate assumes neutrality, and no gene flow since admixture. We have evidence, at least in Le Havre, of a constant input of new chromosome tracts from the native *M. edulis*. In addition, we can suspect a continuing propagule pressure of Med. *M. galloprovincialis* from the maritime traffic. It is also likely that selection acts to maintain parental gene combinations against recombination (Bierne et al., 2006; Simon et al., 2018). Both effects, gene flow and selection, tend to bias the date estimates towards more recent times (Corbett-Detig & Nielsen, 2017). A precise estimation of the admixture event will require a recombination map in mussels and

the distribution of ancestry track lengths along the genome of admixed individuals.

Interestingly, in 1978, Prof. David Skibinski analysed hybrids from natural populations in the Swansea region (UK) with allozymes (Skibinski, Beardmore, & Ahmad, 1978) and noticed that the 'King's dock' populations (Swansea port) were unusual (Figure S39). Those populations showed linkage and Hardy-Weinberg equilibria, and intermediate allele frequencies between *M. edulis* and *M. galloprovincialis*. A closer look at the allele frequency shows that, at one particular allozyme subsequently shown to differentiate Atl. from Med. *M. galloprovincialis* (Ap, Quesada, Zapata, & Alvarez, 1995), King's dock populations had allele frequencies that were closer to those of Med. mussels than to local Atl. *M. galloprovincialis*. This evidence suggests that introduced dock mussels were already present, and already admixed with *M. edulis* at the same level in the Swansea port, 40 years ago. This provides further indication that our estimate of admixture time is potentially underestimated. The term 'dock mussels' was chosen in reference to this work. We do not know if dock mussels persisted in the Swansea port and this matter needs further investigation.

Both of the above considerations suggest that the admixture event leading to dock mussels is a few decades old. The mussel introductions therefore appear relatively recent, especially compared to the several centuries over which human maritime traffic could have been a vector of fouling NIS (J. T. Carlton & Hodder, 1995). However, as stated by Hulme (2009), 'the highest rates of introductions in Europe occurred in the last 25 years' (p. 11) due to an increase in the rate of global exchange. It is therefore possible that dock mussels were spread to multiple ports in this time-frame, especially if a large propagule size is a prerequisite for successful introduction under strong demographic and/or genetic Allee effect (Barton & Turelli, 2011).

Dock mussels are not isolated cases of anthropogenic hybridisation in the *M. edulis* species complex. Recently, Zbawicka et al. (2018) reported the presence of an admixed population between introduced Med. *M. galloprovincialis* and native *M. platensis* close to the city of Puerto Madryn in the middle of the Atlantic coast of Argentina. Their randomly ascertained SNPs did not allow a precise analysis of individual admixture proportions but the average admixture appeared well-balanced. In this issue, Popovic et al. (2019) reported two independent introductions of *M. galloprovincialis* in Australia, one by the Atl. *M. galloprovincialis* in Batemans Bay and the other by the Med. *M. galloprovincialis* in Sydney Harbour, both accompanied by admixture with the native genetic background (*M. planulatus*). In New-Zealand, Gardner, Zbawicka, Westfall, and Wenne (2016) found evidence suggesting possible admixture between introduced *M. galloprovincialis* and the native *Mytilus* species. Such observations are additional indications of the frequent occurrence of the admixture process where *M. galloprovincialis* has been introduced in an area already inhabited by a native lineage of *Mytilus*.

Conversely, there was little to no introgression during the introduction of Med. *M. galloprovincialis* in California (Saarman & Pogson, 2015) and Asia (Brannock, Wethey, & Hilbish, 2009, and Korean sample in this study) where the native species is *M. trossulus*. Those last

two cases may be the result of increased intrinsic and extrinsic reproductive isolation with *M. trossulus* that is much more divergent. Alternatively, the introduction and initial spread may have happened in a place devoid of native *M. trossulus* and with a more balanced demographic context than for dock mussels. Finally, events of admixture are not restricted to *M. galloprovincialis*. For instance, evidence of admixture has been found in the Kerguelen Islands (Fraïsse, Haguénauer, et al., 2018; Zbawicka, Gardner, & Wenne, 2019).

4.2 Confinement of the introduced mussels, local introgression and potential impacts

In all studied ports, the introduced dock mussels form sharp human-induced hybrid zones at the port entrance. By contrast, natural clines in mussels are usually on the order of tens to hundreds of kilometres (Lassen & Turano, 1978; Strelkov, Katolikova, & Väinölä, 2017; Väinölä & Hvilson, 1991). Saarman and Pogson (2015) also found differences in the sharpness of genomic clines between the anthropogenically driven contact in California and old natural secondary contacts. If the natural clines are due to post-zygotic selection in a tension zone model (Barton & Hewitt, 1985; Bierne, David, Boudry, & Bonhomme, 2002), then the narrow clines in ports imply additional processes. Those processes could include habitat choice during the larval settlement stage at a small spatial scale (Bierne et al., 2003; Comesaña & Sanjuan, 1997; Katolikova, Khaitov, Väinölä, Gantsevich, & Strelkov, 2016) or early stage larval or post-settlement ecological selection to the port environment. For instance, selection in mussels could act through attachment strength (Willis & Skibinski, 1992), pollution tolerance (Loria, Cristescu, and Gonzalez, 2019, for a review; and McKenzie, Brooks, and Johnston, 2011, for an example in a bryozoan), or competition for space linked to different growth rates (Branch & Steffani, 2004; Saarman & Pogson, 2015). Additionally, genetic differentiation in mussels has been shown to be associated with sewage treatment plants (Larsson, 2017; Larsson et al., 2016).

Although our sampling around ports was not exhaustive, dock mussels do appear to be restricted to the port interiors, with only a few introduced mussels detected in distant populations. While the presence of introduced migrants up to 30 km from ports may appear concerning, most distant individuals are hybrids between dock mussels and the local background (Figures 3 and S26-S27). Therefore, we can hypothesise that the propagule pressure from ports will be swamped by large native populations for most of the ports. Conversely, native mussels are relatively rare inside the ports (except for Brest). Were they more numerous, hybridisation might favour an increase in introgression by the possibility of backcrossing to the native mussels. The concern of genetic pollution seems increased in the Bay of Brest where the potential for dispersion and hybridisation appears greater. Additionally, populations of introduced mussels were found in basins closed by locks (Saint-Malo, Le Havre, Cherbourg, Saint-Nazaire). In such contexts, both the exit and entry of mussel larva from any species

may be limited and those populations may act as reservoirs of introduced backgrounds.

The introduction cases in ports and Norway agree well with the expectation of asymmetric introgression from the established taxon into the propagating one (Barton, 1979a; Currat, Ruedi, Petit, & Excoffier, 2008; Moran, 1981). Introgression levels can reach much higher levels in a moving hybrid zone than in stable ones (Currat et al., 2008). Genetic pollution by NIS is unlikely to be substantial during invasion, while the reverse is true although less concerning (Currat et al., 2008). However, when the invasion wave is halted and trapped at a natural barrier, density trough, or ecotone, introgression can start to proceed in both directions. Introgression of native mussel populations by dock mussel alleles could therefore become a concern. Nonetheless, the evolutionary future of Med. *M. galloprovincialis* alleles in native populations are hard to predict. They could for example be counter-selected like in the westslope cutthroat trout (*Oncorhynchus clarkii lewisi*), where introgression impacts the fitness of native populations and selection against introduced alleles in wild populations seems to be acting (Kovach et al., 2016; Muhlfeld et al., 2009). While this is an interesting outcome, some parts of the native genome may still be impacted. Indeed, in the brown trout (*Salmo trutta*), a haplotype-based method showed that residual introduced tracts are present in native populations and go undetected by classical ancestry estimation methods (Leitwein, Gagnaire, Desmarais, Berrebi, & Guinand, 2018).

The Bay of Brest is an interesting case study both in terms of implicated species – this is a crossroad between three lineages – and of introduction. In this area, unlike the other ports, introduced mussels were found beyond the major human-made structures. Yet, even in distant sites from ports, mussels were predominantly found on artificial structures (buoys, pillars, piers, etc.). However, this observation may be more related to space competition with oysters on natural sites than to habitat selection, as finding mussels of any type on natural rocky shores in the bay was difficult.

The spread of dock mussels in the Bay of Brest might be due to several interacting factors. First, the port – and notably the commercial area – has a more open layout compared to the other four ports (some of which, such as Saint-Nazaire, have locks at their entry). Second, compared to other ports, habitats suitable for dock mussels might have been available. Third, the closer genetic distance with Atl. *M. galloprovincialis* when compared to *M. edulis* might facilitate hybridisation by avoiding stronger reproductive incompatibilities (both pre- and post-zygotic). Therefore, the prediction of the invasion by dock mussels will require a thorough understanding of the reproductive incompatibilities between non-indigenous and native mussels (Blum, Walters, Burkhead, Freeman, & Porter, 2010; Hall, Hastings, & Ayres, 2006).

When interacting species have accumulated too many incompatibilities for hybridisation to lead to viable and fertile offspring, inter-specific mating represents lost reproductive effort (Allendorf et al., 2001). For less reproductively isolated species, hybridisation has been considered by Mesgaran et al. (2016) as a way to escape demographic Allee effects during colonisation. As small introduced mussel populations may suffer from a strong Allee effect, hybridisation has

potentially provided the initial demographic boost to the first introduction of Med. *M. galloprovincialis*. Conversely, hybrid breakdown would have impeded both the introduction of the hybrid background, which would then have required a tremendous propagule pressure from maritime traffic. The same applies to the subsequent spread of dock mussels, even if fitter (Barton & Turelli, 2011), and this could explain their confinement inside ports. Stochasticity (drift and variation in population density and dispersal) could free the introduced background after a lag time (Piálék & Barton, 1997). Although the delay is expected to be long, confined dock mussel populations could represent hidden bombshells able to escape and spread globally in the future.

The introduced dock mussels display an important component of *M. galloprovincialis* ancestry. Based on the worldwide spread and displacement of local species (Branch et al., 2008; Gardner et al., 2016; Saarman & Pogson, 2015), *M. galloprovincialis* is expected to have a competitive advantage in diverse conditions. It is thus tempting to predict that dock mussels should spread. However, the specific ecological characteristics of these dock mussels as well as the native mussels that first colonized the study ports are unknown, which strongly limits any attempts to predict the impact and the fate of the introduced populations. Their local impact will require further investigation. Nonetheless, we are left with the fact that in ports and in natural environments in the Bay of Brest, dock mussels have probably displaced the native lineages. Michalek, Ventura, and Sanders (2016) report impacts of hybridisation on *Mytilus* aquaculture in Scotland and Larraín et al. (2018) raise concerns of economic impacts in Chile. In Scotland, a recent demographic increase of *M. trossulus* has produced large losses to *M. edulis* aquaculture due to their colonisation of culture ropes and their shell being more fragile (Beaumont, Hawkins, Doig, Davies, & Snow, 2008; Dias et al., 2009). In Brittany and Normandy, most of the cultured mussels are imported spat from the bay of Biscay because *M. edulis* is easier to cultivate, with a shorter reproduction period, and has a higher commercial value for consumers than Atl. *M. galloprovincialis*. Therefore, spat collection of introduced mussels and involuntary culture of the wrong genetic background should impact the quality of cultured mussels and the growing cycle used in mussel farms, but only in case of a massive invasion.

While there may be few direct impacts of dock mussels on native and cultivated mussel populations, indirect effects via parasite hitch-hiking during introduction and their transmission to native species has been documented both in terrestrial and marine systems (Prenter, MacNeil, Dick, & Dunn, 2004; Torchin, Lafferty, & Kuris, 2002). We should therefore be concerned about the potential parasites NIS may have brought into natural and cultivated populations ('spillover' effect). Additionally, the 'spillback' effect, due to the NIS being a competent host for native parasites and constituting a new reservoir for local diseases, should not be neglected (Kelly, Paterson, Townsend, Poulin, & Tompkins, 2009). We can note that, at this time, we did not detect the *M. trossulus* transmissible cancer in dock mussels (Metzger et al., 2016; Riquet et al., 2017). On an evolutionary perspective, the introduction of Atl.

and Med. *M. galloprovincialis* into *M. edulis* ranges and the following gene flow may confer some parasitism adaptations to the native species. For example, it has been demonstrated that *M. galloprovincialis* is more resistant to Pea crab parasitism than *M. edulis* living in the same region (Seed, 1969).

If management is to be considered, multiple steps need to be taken. First, genetic detection methods such as the one used in this work need to be routinely used to assess the extent of the introduction in all large North-European ports. Second, the introduction is to be followed in time and space around the points of introduction, notably to determine the speed of the expansion front, if any, and thus ascertain if dock mussels are becoming invasive. Third, to understand the introduction process in the different ports, there needs to be an integration of genetics and ecology (Lawson Handley et al., 2011). However, we have a large gap in our ecological knowledge of the port environments and what influences mussel populations. A thorough study of the ecology of mussels in ports will be needed to untangle the roles of ecological variation in the distribution of dock mussels. Both habitat choice and post-settlement selection are likely to play a role. The final objective would be to produce a fine scale environmental niche model. Fourth, a vector risk assessment will be necessary to predict the possible human induced secondary displacements (e.g., Herborg, O'Hara, & Therriault, 2009). Finally, at a local scale, larval dispersal through oceanographic constraints will play a major role in the potential spread of dock mussels and dispersal models for NIS in ports will be needed (see David, Matthee, Loveday, & Simon, 2016, for an example at a large scale). While some studies of water flows, tide or wave physical constraints in ports of the English Channel exist (Guillou & Chapalain, 2011, 2012; Jouanneau, Sentchev, & Dumas, 2013), none include a biological module. A study of wave entrance in the southern basin of Le Havre would suggest the likely dispersal of *M. edulis* larvae within this basin (Guillou & Chapalain, 2012), while the whole basin proved populated by dock mussels, providing further evidence for habitat choice or early stage selection. Overall, a large effort will be needed to produce consistent models of larval dispersal at the scale of ports of interest. At a medium scale, in the Bay of Brest, the model of Bessin (2017) could help investigate the relative weights of dispersion, habitat selection and ecological constraints on the distribution of genetic backgrounds. At any rate, managing dock mussels will require the combination of vector risk assessment, network theory, and environmental niche and oceanographic models to build a complete risk assessment model (Frost et al., 2019; Herborg et al., 2009; Hulme, 2009).

In addition to allowing the study of introduction and evolutionary biology, the *Mytilus* model could be of interest for the recent field of urban ecology and evolution, investigating the impact of urbanisation on evolutionary trajectories and the feedbacks with the environment (Rivkin et al., 2018; Thompson, Rieseberg, & Schluter, 2018). The marine environment is not left untouched by urbanisation and human infrastructures have large impacts on coastal communities and their abiotic conditions (Critchley & Bishop, 2019; Mayer-Pinto et al., 2018). This is the 'Ocean Sprawl', in the words of Duarte et al. (2012), which has broad effects

encompassing connectivity modifications and environmental and toxicological changes (for a review see Firth et al., 2016).

4.3 Parallelism of distortions

The parallelism in allele frequency distortions that we observed between admixture events, suggests that patterns produced during such events can be highly repeatable. This is probably due to a combination of processes. As discussed above, port introductions are expected to partly share a pre-introduction history of admixture. The two introduced Atl. *M. galloprovincialis* populations we studied in Norway are also likely to share the same history of admixture. However, the composition in *M. edulis* ancestry of these populations is in accordance with an independent admixture event with the local *M. edulis* background. Naturally admixed Atl. *M. galloprovincialis* combine an old history of introgression during glacial oscillation periods (Fraïsse, Roux, et al., 2018; Roux et al., 2014) with ongoing local introgression from the native *M. edulis* populations in direct contact within the mosaic hybrid zone observed today (Fraïsse et al., 2016; Simon et al., 2019).

Shared colonisation history cannot be the whole story, however, because we also found repeatable patterns between admixture events that must be considered independent. This includes not only the comparisons of natural admixture to the introduced Atl. *M. galloprovincialis* (involving two different backgrounds of *M. edulis*: South- and North-Eu.), but also the comparison of port samples to experimental backcrosses.

Our comparison of Atl. *M. galloprovincialis* admixtures includes populations with a wide variety of contributions from the parental lineages. These range from a high *M. edulis* contribution in JER to a high Atl. *M. galloprovincialis* contribution in MCH. The high positive correlations of distortions observed between all Atl. *M. galloprovincialis* admixture, despite variable contributions of the two parental backgrounds, is particularly interesting. The calculation of D corrects for ancient introgression of parental backgrounds, and we are unlikely to have missed a hidden parental population given our broad geographic survey (this work and Simon et al., 2019) and the large-scale genetic panmixia usually observed in mussels outside hybrid zones (e.g. East vs. West Mediterranean Sea). Genomic regions do tend to deviate consistently toward an excess of *M. galloprovincialis* ancestry or an excess of *M. edulis* ancestry. This suggests selective processes and a shared architecture of the barrier to gene flow. A first possible explanation is that some loci are closer to barrier loci than others (Ravinet et al., 2017). Barrier loci can be local adaptation genes or genetic incompatibilities. Schumer et al. (2018) found that in several events of admixture between swordtail fish species contributing differently to the resulting population, local ancestry were nonetheless positively correlated. They showed that parallel correlations, despite opposite parental contributions, can be the result of selection in the same direction to resolve pairwise epistatic incompatibilities. In addition, an interesting interpretation of the parallelism observed in mussels would be that our loci belong

to genomic regions with different rates of recombination. *M. edulis* and *M. galloprovincialis* are close to the 2% net synonymous divergence limit (1.89%), where there is a high probability of strong reproductive isolation, either due to physical constraints or sufficient accumulation of incompatibilities (Roux et al., 2016). They are therefore expected to be incompatible at a high number of differentiated sites (Simon et al., 2018). With such a highly polygenic determinism of post-zygotic selection one expect a correlation between recombination rates and introgression (Barton & Bengtsson, 1986), which has recently been observed in multiple study systems (*Mimulus*, Aeschbacher, Selby, Willis, and Coop, 2017; sea bass, Duranton et al., 2018; oyster, Gagnaire et al., 2018; stickleback, Roesti, Moser, and Berner, 2013; swordtail fish, Schumer et al., 2018 or *Heliconius*, Martin, Davey, Salazar, and Jiggins, 2019).

While patterns of hybridisation are strongly repeatable when the same *M. galloprovincialis* lineages are involved, equally notable is the lack of repeatability with different lineages. A possible explanation is that different sets of incompatible loci may be implicated in the reproductive isolation between *M. edulis* and the two *M. galloprovincialis* lineages. However, the history of divergence between the two *M. galloprovincialis* lineages is much younger than the divergence with *M. edulis* and most of the fixed mutations are expected to be shared by the two lineages (Fraïsse et al., 2016). Additionally, Atl. *M. galloprovincialis* is in contact with *M. edulis* while Med. *M. galloprovincialis* is not. Atl. *M. galloprovincialis* has experienced a punctuated history of introgression possibly swamped by bi-stable incompatibilities with an asymmetric advantage to the *M. edulis* allele (Fraïsse et al., 2016; Gosset & Bierne, 2013; Simon et al., 2019). This differential introgression might have erased, or even reversed, the selective effects in the two *M. galloprovincialis* backgrounds. This hypothesis requires further theoretical and experimental investigation. Finally, given that karyotypic differences have been suggested between the two *M. galloprovincialis* lineages (Martínez-Lage, González-Tizón, & Méndez, 1996), they potentially exhibit different recombination landscapes impacting the outcome of distortions.

5 Conclusion

Mytilus mussels, with their introduction and hybridisation potential, are a particularly useful model for studying the parallelism of admixture events, and the range of outcomes of introductions with hybridisation. Our study shows that admixture between the same genetic backgrounds are highly repeatable. This repeatability can be explained both by a shared history of pre-introduction admixture and parallel genomic processes. One category of anthropogenic hybridisations, the 'dock mussels', exhibit homogeneous patterns of admixture among all studied populations, and appear to be restricted to environments of large commercial ports. Follow-up investigations will be needed to understand how selection, hybridisation, environmental conditions and dispersal are shaping the distribution and genomic architecture of these dock mussels and similar introductions.

Data availability: Raw data and scripts are available as a Zenodo archive [10.5281/zenodo.3375381](https://doi.org/10.5281/zenodo.3375381).

Acknowledgements: Data used in this work were partly produced through the genotyping and sequencing facilities of ISEM and LabEx CeMEB, an ANR "Investissements d'avenir" program (ANR-10-LABX-04-01). This project benefited from the Montpellier Bioinformatics Biodiversity platform supported by the LabEx CeMEB. We are grateful to the Marine Observation Department (SMO) at the Roscoff Biological Station for sampling in the bay of Brest. We thank Laure Paradis for providing maps of the study area and her assistance with map editing. We thank Prof. D. Skibinski for communications about the Swansea King's Dock mussels. We also thank people who contributed to the acquisition of samples or gave advice: Galice Hoarau, Alison Gallet, Filip Volckaert, Association Port Vivant (Le Havre), CRC Normandie, Julien Normand, Sophie Arnaud-Haond, Lea-Anne Henry, Thierry Comtet, Pieternella Luttikhuisen and Jørgen Berge (UiT The Arctic University of Norway). We acknowledge the help of staff and contractors at CNR International UK Ltd for helping to collect and preserve the samples from the Murchison oil station. We thank Richard Crooijmans and Bert Dibbits from Animal Breeding and Genomics Wageningen University & Research for supplying material. This work was supported by the ANR Project HySea (ANR-12-BSV7-0011). Maria Skazina was supported by the Russian Science Foundation project 19-74-20024.

References

- Abbott, R. J., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., ... Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229–246. doi:10.1111/j.1420-9101.2012.02599.x
- Aeschbacher, S., Selby, J. P., Willis, J. H., & Coop, G. (2017). Population-genomic inference of the strength and timing of selection against gene flow. *Proceedings of the National Academy of Sciences*, *114*(27), 7061–7066. doi:10.1073/pnas.1616755114
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. doi:10.1101/gr.094052.109
- Allendorf, F. W., Leary, R. F., Spruell, P., & Wenburg, J. K. (2001). The problems with hybrids: Setting conservation guidelines. *Trends in Ecology and Evolution*, *16*(11), 613–622. doi:10.1016/S0169-5347(01)02290-X
- Anderson, E. C., & Thompson, E. A. (2002). A Model-Based Method for Identifying Species Hybrids Using Multilocus Genetic Data. *Genetics*, *160*(3), 1217–1229.
- Apte, S., Holland, B. S., Godwin, L. S., & Gardner, J. P. A. (2000). Jumping ship: A stepping stone event mediating transfer of a non-indigenous species via a potentially unsuitable environment. *Biological Invasions*, *2*(1), 75–79. doi:10.1023/A:1010024818644
- Barton, N. H. (1979a). Gene flow past a cline. *Heredity*, *43*(3), 333–339. doi:10.1038/hdy.1979.86
- Barton, N. H. (1979b). The dynamics of hybrid zones. *Heredity*, *43*(3), 341–359. doi:10.1038/hdy.1979.87
- Barton, N. H., & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity*, *56*, 357–376. doi:10.1038/hdy.1986.135
- Barton, N. H., & Hewitt, G. M. (1985). Analysis of Hybrid Zones. *Annual Review of Ecology, Evolution, and Systematics*, *16*, 113–148. doi:10.1146/annurev.es.16.110185.000553
- Barton, N. H., & Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature*, *341*, 497–503. doi:10.1038/341497a0
- Barton, N. H., & Turelli, M. (2011). Spatial waves of advance with bistable dynamics: Cytoplasmic and genetic analogues of allee effects. *The American naturalist*, *178*(3), E48–E75. doi:10.1086/661246
- Bax, N., Hayes, K., Marshall, A., Parry, D., & Thresher, R. (2002). Man-made marinas as sheltered islands for alien marine organisms: Establishment and eradication of an alien invasive marine species. In C. R. Veitch & M. N. Clout (Eds.), *Turning the tide: The eradication of invasive species* (IUCN SSC Invasive Species Specialist Group, pp. 26–39). Gland, Switzerland and Cambridge, UK: IUCN.
- Bayne, B. L. (1976). 4. The biology of mussel larvae. In B. L. Bayne (Ed.), *Marine mussels: Their ecology and physiology* (Vol. 10, pp. 81–120). Cambridge University Press.
- Beaumont, A. R., Hawkins, M. P., Doig, F. L., Davies, I. M., & Snow, M. (2008). Three species of *Mytilus* and their hybrids identified in a Scottish Loch: Natives, relicts and invaders? *Journal of Experimental Marine Biology and Ecology*, *367*(2), 100–110. doi:10.1016/j.jembe.2008.08.021
- Beninde, J., Feldmeier, S., Veith, M., & Hochkirch, A. (2018). Admixture of hybrid swarms of native and introduced lizards in cities is determined by the cityscape structure and invasion history. *Proceedings of the Royal Society B: Biological Sciences*, *285*, 20180143. doi:10.1098/rspb.2018.0143

-
- Bertl, J., Ringbauer, H., & Blum, M. G. B. (2018). Can secondary contact following range expansion be distinguished from barriers to gene flow? *PeerJ*, 6, e5325. doi:10.7717/peerj.5325
- Bessin, C. (2017). *Dispersion larvaire de l'ophiure noire Ophiocomina nigra en mer d'Iroise : distribution des larves et modélisation du transport larvaire*. (Master Océanographie, Université de la Méditerranée).
- 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*, 273, 1253–1260. doi:10.1098/rspb.2005.3440
- 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(2), 447–461. doi:10.1046/j.1365-294X.2003.01730.x
- 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(2), 292–298. doi:10.1111/j.0014-3820.2002.tb01339.x
- 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. doi:10.3354/meps245157
- Bierne, N., Gagnaire, P.-A., & David, P. (2013). The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. *Current Zoology*, 59(1), 72–86. doi:10.1093/czoolo/59.1.72
- Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: Why genome scans may fail to map local adaptation genes. *Molecular Ecology*, 20, 2044–2072. doi:10.1111/j.1365-294X.2011.05080.x
- Bishop, M. J., Mayer-Pinto, M., Airoidi, L., Firth, L. B., Morris, R. L., Loke, L. H., ... Dafforn, K. A. (2017). Effects of ocean sprawl on ecological connectivity: Impacts and solutions. *Journal of Experimental Marine Biology and Ecology*, 492, 7–30. doi:10.1016/j.jembe.2017.01.021
- Blum, M. J., Walters, D. M., Burkhead, N. M., Freeman, B. J., & Porter, B. A. (2010). Reproductive isolation and the expansion of an invasive hybrid swarm. *Biological Invasions*, 12, 2825–2836. doi:10.1007/s10530-010-9688-9
- Bouchemousse, S., Liautard-Haag, C., Bierne, N., & Viard, F. (2016). Distinguishing contemporary hybridization from past introgression with postgenomic ancestry-informative SNPs in strongly differentiated *Ciona* species. *Molecular Ecology*, 25(21), 5527–5542. doi:10.1111/mec.13854
- Braby, C. E., & Somero, G. N. (2006). Ecological gradients and relative abundance of native (*Mytilus trossulus*) and invasive (*Mytilus galloprovincialis*) blue mussels in the California hybrid zone. *Marine Biology*, 148, 1249–1262. doi:10.1007/s00227-005-0177-0
- Branch, G. M., Odendaal, F., & Robinson, T. (2008). Long-term monitoring of the arrival, expansion and effects of the alien mussel *Mytilus galloprovincialis* relative to wave action. *Marine Ecology Progress Series*, 370, 171–183. doi:10.3354/meps07626
- Branch, G. M., & Steffani, C. N. (2004). Can we predict the effects of alien species? A case-history of the invasion of South Africa by *Mytilus galloprovincialis* (Lamarck). *Journal of Experimental Marine Biology and Ecology*, 300, 189–215. doi:10.1016/j.jembe.2003.12.007

- Brannock, P. M., Wethey, D. S., & Hilbish, T. J. (2009). Extensive hybridization with minimal introgression in *Mytilus galloprovincialis* and *M. trossulus* in Hokkaido, Japan. *Marine Ecology Progress Series*, 383, 161–171. doi:10.3354/meps07995
- Broman, K. W., Wu, H., Sen, S., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, 19, 889–890.
- Brooks, S. J., & Farnen, E. (2013). The Distribution of the Mussel *Mytilus* Species Along the Norwegian Coast. *Journal of Shellfish Research*, 32(2), 265–270. doi:10.2983/035.032.0203
- Carlton, J. T. [J. T.], & Hodder, J. (1995). Biogeography and dispersal of coastal marine organisms: Experimental studies on a replica of a 16th-century sailing vessel. *Marine Biology*, 121(4), 721–730. doi:10.1007/BF00349308
- Carlton, J. T. [James T.], Geller, J. B., Reaka-Kudla, M. L., & Norse, E. A. (1999). Historical Extinctions in the Sea. *Annual Review of Ecology and Systematics*, 30(1), 515–538. doi:10.1146/annurev.ecolsys.30.1.515
- Casoli, E., Ventura, D., Modica, M. V., Belluscio, A., Capello, M., Oliverio, M., & Ardizzone, G. D. (2016). A massive ingression of the alien species *Mytilus edulis* L. (Bivalvia: Mollusca) into the Mediterranean Sea following the Costa Concordia cruise-ship disaster. *Mediterranean Marine Science*, 17(2), 404–416. doi:10.12681/mms.1619
- Chapman, M., & Underwood, A. (2011). Evaluation of ecological engineering of “armoured” shorelines to improve their value as habitat. *Journal of Experimental Marine Biology and Ecology*, 400, 302–313. doi:10.1016/j.jembe.2011.02.025
- Clarke Murray, C., Pakhomov, E. A., & Therriault, T. W. (2011). Recreational boating: A large unregulated vector transporting marine invasive species: Transport of NIS by recreational boats. *Diversity and Distributions*, 17(6), 1161–1172. doi:10.1111/j.1472-4642.2011.00798.x
- Comesaña, A. S., & Sanjuan, A. (1997). Microgeographic allozyme differentiation in the hybrid zone of *Mytilus galloprovincialis* Lmk. and *M. edulis* L. on the continental European coast. *Helgoländer Meeresunters*, 51, 107–124. doi:10.1007/BF02908758
- Coolen, J. W. P. (2017). *North Sea reefs: Benthic biodiversity of artificial and rocky reefs in the southern North Sea* (PhD thesis, Wageningen University, Wageningen).
- Corbett-Detig, R., & Nielsen, R. (2017). A Hidden Markov Model Approach for Simultaneously Estimating Local Ancestry and Admixture Time Using Next Generation Sequence Data in Samples of Arbitrary Ploidy. *PLOS Genetics*, 13(1), e1006529. doi:10.1371/journal.pgen.1006529
- Crego-Prieto, V., Ardura, A., Juanes, F., Roca, A., Taylor, J. S., & Garcia-Vazquez, E. (2015). Aquaculture and the spread of introduced mussel genes in British Columbia. *Biological Invasions*, 17(7), 2011–2026. doi:10.1007/s10530-015-0853-z
- Critchley, L. P., & Bishop, M. J. (2019). Differences in soft-sediment infaunal communities between shorelines with and without seawalls. *Estuaries and Coasts*. doi:10.1007/s12237-019-00527-z
- Critchlow, D. E., & Fligner, M. A. (1991). On distribution-free multiple comparisons in the one-way analysis of variance. *Communications in Statistics - Theory and Methods*, 20(1), 127–139. doi:10.1080/03610929108830487
- Currat, M., Ruedi, M., Petit, R. J., & Excoffier, L. (2008). The Hidden Side of Invasions: Massive Introgression by Local Genes. *Evolution*, 62, 1908–1920. doi:10.1111/j.1558-5646.2008.00413.x

-
- Daguin, C., & Borsa, P. (2000). Genetic relationships of *Mytilus galloprovincialis* Lamarck populations worldwide: Evidence from nuclear-DNA markers. *Geological Society, London, Special Publications*, 177(1), 389–397. doi:10.1144/GSL.SP.2000.177.01.26
- David, A. A., Matthee, C. A., Loveday, B. R., & Simon, C. A. (2016). Predicting the Dispersal Potential of an Invasive Polychaete Pest along a Complex Coastal Biome. *Integrative and Comparative Biology*, 56(4), 600–610. doi:10.1093/icb/icw011
- Derryberry, E. P., Derryberry, G. E., Maley, J. M., & Brumfield, R. T. (2014). Hzar: Hybrid zone analysis using an R software package. *Molecular Ecology Resources*, 14(3), 652–663. doi:10.1111/1755-0998.12209
- Dias, P. J., Dordor, A., Tulett, D., Piertney, S., Davies, I. M., & Snow, M. (2009). Survey of mussel (*Mytilus*) species at Scottish shellfish farms. *Aquaculture Research*, 40, 1715–1722. doi:10.1111/j.1365-2109.2009.02274.x
- Drake, J. M., & Lodge, D. M. (2004). Global hot spots of biological invasions: Evaluating options for ballast–water management. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271, 575–580. doi:10.1098/rspb.2003.2629
- Duarte, C. M., Pitt, K. A., Lucas, C. H., Purcell, J. E., Uye, S.-i., Robinson, K., ... Condon, R. H. (2012). Is global ocean sprawl a cause of jellyfish blooms? *Frontiers in Ecology and the Environment*, 11(2), 91–97. doi:10.1890/110246
- Durantón, M., Allal, F., Fraïsse, C., Bierne, N., Bonhomme, F., & Gagnaire, P.-A. (2018). The origin and remolding of genomic islands of differentiation in the European sea bass. *Nature Communications*, 9, 2518. doi:10.1038/s41467-018-04963-6
- El Ayari, T., Trigui El Menif, N., Hamer, B., Cahill, A. E., & Bierne, N. (2019). The hidden side of a major marine biogeographic boundary: A wide mosaic hybrid zone at the Atlantic–Mediterranean divide reveals the complex interaction between natural and genetic barriers in mussels. *Heredity*, 122, 770–784. doi:10.1038/s41437-018-0174-y
- Ellstrand, N. C., & Schierenbeck, K. A. (2000). Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences*, 97(13), 7043–7050. doi:10.1073/pnas.97.13.7043
- Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics*, 164(4), 1567–1587.
- FAO. (2018). *FAO yearbook. Fishery and Aquaculture Statistics 2016*. Roma.
- Firth, L. B., Knights, A. M., Bridger, D., Evans, A. J., Mieszkowska, N., Moore, P. J., ... Hawkins, S. J. (2016). Ocean sprawl: Challenges and opportunities for biodiversity management in a changing world. In R. N. Hughes, D. J. Hughes, I. P. Smith, & A. C. Dale (Eds.), *Oceanography and Marine Biology* (Vol. 54, pp. 201–278). Boca Raton: CRC Press.
- Fitzpatrick, S. W., Bradburd, G. S., Kremer, C. T., Salerno, P. E., Angeloni, L. M., & Funk, W. C. (2019). Genetic rescue without genomic swamping in wild populations. *bioRxiv*. doi:10.1101/701706
- Fly, E. K., & Hilbish, T. J. [Thomas J.]. (2013). Physiological energetics and biogeographic range limits of three congeneric mussel species. *Oecologia*, 172(1), 35–46. doi:10.1007/s00442-012-2486-6
- Fraïsse, C., Belkhir, K., Welch, J. J., & Bierne, N. (2016). Local interspecies introgression is the main cause of extreme levels of intraspecific differentiation in mussels. *Molecular Ecology*, 25(1), 269–286. doi:10.1111/mec.13299
- Fraïsse, C., Haguénauer, A., Gerard, K., Weber, A. A.-T., Bierne, N., & Chenuil, A. (2018). Fine-grained habitat-associated genetic connectivity in an admixed population of mussels

- in the small isolated Kerguelen Islands. *bioRxiv*, 239244, ver. 4 recommended and peer-reviewed by *PCI Evol Biol*. doi:10.1101/239244
- Fraïsse, C., Roux, C., Gagnaire, P.-A., Romiguier, J., Faivre, N., Welch, J. J., & Bierne, N. (2018). The divergence history of European blue mussel species reconstructed from Approximate Bayesian Computation: The effects of sequencing techniques and sampling strategies. *PeerJ*, 6, e5198. doi:10.7717/peerj.5198
- Frost, C. M., Allen, W. J., Courchamp, F., Jeschke, J. M., Saul, W.-C., & Wardle, D. A. (2019). Using Network Theory to Understand and Predict Biological Invasions. *Trends in Ecology & Evolution*. doi:10.1016/j.tree.2019.04.012
- Gagnaire, P.-A., Lamy, J.-B., Cornette, F., Heurtebise, S., Dégremont, L., Flahauw, E., ... Lapègue, S. (2018). Analysis of Genome-Wide Differentiation between Native and Introduced Populations of the Cupped Oysters *Crassostrea gigas* and *Crassostrea angulata*. *Genome Biol. Evol.*, 10(9), 2518–2534. doi:10.1093/gbe/evy194
- Gardner, J. P. A. (1994). The structure and dynamics of naturally occurring hybrid *Mytilus edulis* (Linnaeus, 1758) and *Mytilus galloprovincialis* (Lamarck, 1819) populations: Review and interpretation. *Archiv Fur Hydrobiologie*, 1-2, 37–71.
- Gardner, J. P. A., Zbawicka, M., Westfall, K. M., & Wenne, R. (2016). Invasive blue mussels threaten regional scale genetic diversity in mainland and remote offshore locations: The need for baseline data and enhanced protection in the Southern Ocean. *Global Change Biology*. doi:10.1111/gcb.13332
- Geller, J. B., Carlton, J. T., & Powers, D. A. (1994). PCR-based detection of mtDNA haplotypes of native and invading mussels on the northeastern Pacific coast: Latitudinal pattern of invasion. *Marine Biology*, 119(2), 243–249. doi:10.1007/BF00349563
- Gosset, C. C., & Bierne, N. (2013). Differential introgression from a sister species explains high F_{ST} outlier loci within a mussel species. *Journal of Evolutionary Biology*, 26(1), 14–26. doi:10.1111/jeb.12046
- Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5, 184–186. doi:10.1111/j.1471-8286.2004.00828.x
- Guillou, N., & Chapalain, G. (2011). Effects of the coupling between TELEMAC2D and TOMAWAC on SISYPHE modelling in the outer Seine estuary. *Proceedings of the XVI-llth Telemac & Mascaret User Club*, 51–58.
- Guillou, N., & Chapalain, G. (2012). Modeling Penetration of Tide-Influenced Waves in Le Havre Harbor. *Journal of Coastal Research*, 28(4), 945–955. doi:10.2112/JCOASTRES-D-11-00192.1
- Guzinski, J., Ballenghien, M., Daguin-Thiébaud, C., Lévêque, L., & Viard, F. (2018). Population genomics of the introduced and cultivated Pacific kelp *Undaria pinnatifida*: Marinas – not farms – drive regional connectivity and establishment in natural rocky reefs. *Evolutionary Applications*, 11, 1582–1597. doi:10.1111/eva.12647
- Hall, R. J., Hastings, A., & Ayres, D. R. (2006). Explaining the explosion: Modelling hybrid invasions. *Proceedings of the Royal Society of London B: Biological Sciences*, 273, 1385–1389. doi:10.1098/rspb.2006.3473
- Han, Z., Mao, Y., Shui, B., Yanagimoto, T., & Gao, T. (2016). Genetic structure and unique origin of the introduced blue mussel *Mytilus galloprovincialis* in the north-western Pacific: Clues from mitochondrial cytochrome c oxidase I (COI) sequences. *Marine and Freshwater Research*, 68(2), 263–269. doi:10.1071/MF15186
- Harris, K., Zhang, Y., & Nielsen, R. (2019). Genetic rescue and the maintenance of native ancestry. *Conservation Genetics*, 20(1), 59–64. doi:10.1007/s10592-018-1132-1

-
- Herborg, L.-M., O'Hara, P., & Therriault, T. W. (2009). Forecasting the potential distribution of the invasive tunicate *Didemnum vexillum*. *Journal of Applied Ecology*, *46*(1), 64–72. doi:10.1111/j.1365-2664.2008.01568.x
- Hewitt, C. L., Gollasch, S., & Minchin, D. (2009). The Vessel as a Vector – Biofouling, Ballast Water and Sediments. In G. Rilov & J. A. Crooks (Eds.), *Biological Invasions in Marine Ecosystems* (Vol. 204, pp. 117–131). doi:10.1007/978-3-540-79236-9_6
- Hewitt, G. M. (1988). Hybrid zones-natural laboratories for evolutionary studies. *Trends in Ecology and Evolution*, *3*(7), 158–167. doi:10.1016/0169-5347(88)90033-X
- Hilbish, T. J. [T. J.], Carson, E. W., Plante, J. R., Weaver, L. A., & Gilg, M. R. (2002). Distribution of *Mytilus edulis*, *M. galloprovincialis*, and their hybrids in open-coast populations of mussels in southwestern England. *Marine Biology*, *140*, 137–142. doi:10.1007/s002270100631
- Hollander, M., Wolfe, D. A., & Chicken, E. (2015). The One-Way Layout. In *Nonparametric Statistical Methods* (pp. 202–288). doi:10.1002/9781119196037
- Hudson, J., Viard, F., Roby, C., & Rius, M. (2016). Anthropogenic transport of species across native ranges: Unpredictable genetic and evolutionary consequences. *Biology Letters*, *12*, 20160620. doi:10.1098/rsbl.2016.0620
- Hulme, P. E. (2009). Trade, transport and trouble: Managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, *46*, 10–18. doi:10.1111/j.1365-2664.2008.01600.x
- Jiggins, C. D., & Mallet, J. (2000). Bimodal hybrid zones and speciation. *Trends in Ecology & Evolution*, *15*(6), 250–255. doi:10.1016/S0169-5347(00)01873-5
- Jombart, T. (2008). ADEGENET: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405. doi:10.1093/bioinformatics/btn129
- Jouanneau, N., Sentchev, A., & Dumas, F. (2013). Numerical modelling of circulation and dispersion processes in Boulogne-sur-Mer harbour (Eastern English Channel): Sensitivity to physical forcing and harbour design. *Ocean Dynamics*, *63*, 1321–1340. doi:10.1007/s10236-013-0659-4
- Kartavtsev, Y. P., Chichvarkhin, A. Y., Kijima, A., Hanzawa, N., & Park, I.-S. (2005). Allozyme and Morphometric Analysis of Two Common Mussel Species of the Genus *Mytilus* (Mollusca, Mytilidae) in Korean, Japanese and Russian Waters. *Korean Journal of Genetics*, *24*(4), 289–306.
- Katolikova, M., Khaitov, V., Väinölä, R., Gantsevich, M., & Strelkov, P. (2016). Genetic, Ecological and Morphological Distinctness of the Blue Mussels *Mytilus trossulus* Gould and *M. edulis* L. in the White Sea. *PLOS ONE*, *11*(4), e0152963. doi:10.1371/journal.pone.0152963
- Katsanevakis, S., Zenetos, A., Belchior, C., & Cardoso, A. C. (2013). Invading European Seas: Assessing pathways of introduction of marine aliens. *Ocean & Coastal Management*, *76*, 64–74. doi:10.1016/j.ocecoaman.2013.02.024
- Kelly, D. W., Paterson, R. A., Townsend, C. R., Poulin, R., & Tompkins, D. M. (2009). Parasite spillback: A neglected concept in invasion ecology? *Ecology*, *90*(8), 2047–2056. doi:10.1890/08-1085.1
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179–1191. doi:10.1111/1755-0998.12387
- Kovach, R. P., Hand, B. K., Hohenlohe, P. A., Cosart, T. F., Boyer, M. C., Neville, H. H., ... Luikart, G. (2016). Vive la résistance: Genome-wide selection against introduced alleles

- in invasive hybrid zones. *Proceedings of the Royal Society B: Biological Sciences*, 283, 20161380. doi:10.1098/rspb.2016.1380
- Largiadèr, C. R. (2008). Hybridization and Introgression Between Native and Alien Species. In W. Nentwig (Ed.), *Biological Invasions* (Vol. 193, pp. 275–292). doi:10.1007/978-3-540-36920-2_16
- Larraín, M. A., Zbawicka, M., Araneda, C., Gardner, J. P. A., & Wenne, R. (2018). Native and invasive taxa on the Pacific coast of South America: Impacts on aquaculture, traceability and biodiversity of blue mussels (*Mytilus* spp.) *Evolutionary Applications*, 11, 298–311. doi:10.1111/eva.12553
- Larsson, J. (2017). *Genetic Aspects of Environmental Disturbances in Marine Ecosystems: Studies of the Blue Mussel in the Baltic Sea*. Södertörns högskola.
- Larsson, J., Lönn, M., Lind, E. E., Świeżak, J., Smolarz, K., & Grahn, M. (2016). Sewage treatment plant associated genetic differentiation in the blue mussel from the Baltic Sea and Swedish west coast. *PeerJ*, 4, e2628. doi:10.7717/peerj.2628
- Lassen, H. H., & Turano, F. J. (1978). Clinal variation and heterozygote deficit at the lap-locus in *Mytilus edulis*. *Marine Biology*, 49(3), 245–254. doi:10.1007/BF00391137
- Lawson Handley, L.-J., Estoup, A., Evans, D. M., Thomas, C. E., Lombaert, E., Facon, B., ... Roy, H. E. (2011). Ecological genetics of invasive alien species. *BioControl*, 56(4), 409–428. doi:10.1007/s10526-011-9386-2
- Leitwein, M., Gagnaire, P.-A., Desmarais, E., Berrebi, P., & Guinand, B. (2018). Genomic consequences of a recent three-way admixture in supplemented wild brown trout populations revealed by local ancestry tracts. *Molecular Ecology*, 27, 3466–3483. doi:10.1111/mec.14816
- Lockwood, B. L., & Somero, G. N. (2011). Invasive and native blue mussels (genus *Mytilus*) on the California coast: The role of physiology in a biological invasion. *Journal of Experimental Marine Biology and Ecology*, 400, 167–174. doi:10.1016/j.jembe.2011.02.022
- Loria, A., Cristescu, M. E., & Gonzalez, A. (2019). Mixed evidence for adaptation to environmental pollution. *Evolutionary Applications*. doi:10.1111/eva.12782
- Martin, S. H., Davey, J. W., Salazar, C., & Jiggins, C. D. (2019). Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLOS Biology*, 17(2), e2006288. doi:10.1371/journal.pbio.2006288
- Martínez-Lage, A., González-Tizón, A., & Méndez, J. (1996). Chromosome differences between European mussel populations (genus *Mytilus*). *Caryologia*, 49(3-4), 343–355. doi:10.1080/00087114.1996.10797379
- Mathiesen, S. S., Thyrring, J., Hemmer-Hansen, J., Berge, J., Sukhotin, A., Leopold, P., ... Nielsen, E. E. (2016). Genetic diversity and connectivity within *Mytilus* spp. in the subarctic and Arctic. *Evolutionary Applications*, 10, 39–55. doi:10.1111/eva.12415
- Mayer-Pinto, M., Cole, V. J., Johnston, E. L., Bugnot, A., Hurst, H., Airoidi, L., ... Dafforn, K. A. (2018). Functional and structural responses to marine urbanisation. *Environmental Research Letters*, 13, 014009. doi:10.1088/1748-9326/aa98a5
- McDonald, J. H., Seed, R., & Koehn, R. K. (1991). Allozymes and morphometric characters of three species of *Mytilus* in the Northern and Southern Hemispheres. *Marine Biology*, 111, 323–333. doi:10.1007/BF01319403
- McFarlane, S. E., & Pemberton, J. M. (2019). Detecting the True Extent of Introgression during Anthropogenic Hybridization. *Trends in Ecology & Evolution*, 34(4), 315–326. doi:10.1016/j.tree.2018.12.013

-
- McKenzie, L. A., Brooks, R., & Johnston, E. L. (2011). Heritable pollution tolerance in a marine invader. *Environmental Research*, *111*(7), 926–932. doi:10.1016/j.envres.2010.12.007
- Mesgaran, M. B., Lewis, M. A., Ades, P. K., Donohue, K., Ohadi, S., Li, C., & Cousens, R. D. (2016). Hybridization can facilitate species invasions, even without enhancing local adaptation. *Proceedings of the National Academy of Sciences*, *113*(36), 10210–10214. doi:10.1073/pnas.1605626113
- Metzger, M. J., Villalba, A., Carballal, M. J., Iglesias, D., Sherry, J., Reinisch, C., ... Goff, S. P. (2016). Widespread transmission of independent cancer lineages within multiple bivalve species. *Nature*, *534*, 705–709. doi:10.1038/nature18599
- Michalek, K., Ventura, A., & Sanders, T. (2016). Mytilus hybridization and impact on aquaculture: A Minireview. *Marine Genomics*, *27*, 3–7. doi:10.1016/j.margen.2016.04.008
- Miller, J. A., Carlton, J. T., Chapman, J. W., Geller, J. B., & Ruiz, G. M. (2017). Transoceanic dispersal of the mussel *Mytilus galloprovincialis* on Japanese tsunami marine debris: An approach for evaluating rafting of a coastal species at sea. *Marine Pollution Bulletin*, *132*, 60–69. doi:10.1016/j.marpolbul.2017.10.040
- Miralles, L., Gomez-Agenjo, M., Rayon-Viña, F., Gyraitė, G., & Garcia-Vazquez, E. (2018). Alert calling in port areas: Marine litter as possible secondary dispersal vector for hitchhiking invasive species. *Journal for Nature Conservation*, *42*, 12–18. doi:10.1016/j.jnc.2018.01.005
- Mlouka, R., Cachot, J., Boukadida, K., Clérandeau, C., Gourves, P.-Y., & Banni, M. (2019). Compared responses to copper and increased temperatures of hybrid and pure offspring of two mussel species. *Science of The Total Environment*, *685*, 795–805. doi:10.1016/j.scitotenv.2019.05.466
- Molnar, J. L., Gamboa, R. L., Revenga, C., & Spalding, M. D. (2008). Assessing the global threat of invasive species to marine biodiversity. *Frontiers in Ecology and the Environment*, *6*(9), 485–492. doi:10.1890/070064
- Mooney, H. A., & Cleland, E. E. (2001). The evolutionary impact of invasive species. *Proceedings of the National Academy of Sciences*, *98*(10), 5446–5451. doi:10.1073/pnas.091093398
- Moran, C. (1981). Genetic demarcation of geographical distribution by hybrid zones. In *Proceedings of the Ecological Society of Australia* (Vol. 11, pp. 67–73).
- Muhlfeld, C. C., Kalinowski, S. T., McMahon, T. E., Taper, M. L., Painter, S., Leary, R. F., & Allendorf, F. W. (2009). Hybridization rapidly reduces fitness of a native trout in the wild. *Biology Letters*, *5*(3), 328–331. doi:10.1098/rsbl.2009.0033
- Nunes, A. L., Katsanevakis, S., Zenetos, A., & Cardoso, A. C. (2014). Gateways to alien invasions in the European seas. *Aquatic Invasions*, *9*(2), 133–144. doi:10.3391/ai.2014.9.2.02
- Paradis, E. (2010). Pegas: An R package for population genetics with an integrated–modular approach. *Bioinformatics*, *26*, 419–420.
- Piálek, J., & Barton, N. H. (1997). The Spread of an Advantageous Allele Across a Barrier: The Effects of Random Drift and Selection Against Heterozygotes. *Genetics*, *145*(2), 493–504.
- Poole, W., Gibbs, D. L., Shmulevich, I., Bernard, B., & Knijnenburg, T. A. (2016). Combining dependent *P*-values with an empirical adaptation of Brown’s method. *Bioinformatics*, *32*(17), i430–i436. doi:10.1093/bioinformatics/btw438

- Popovic, I., Matias, A. M. A., Bierne, N., & Riginos, C. (2019). Twin introductions by independent invader mussel lineages are both associated with recent admixture with a native congener in Australia. *Evolutionary Applications*, eva.12857. doi:10.1111/eva.12857
- Prenter, J., MacNeil, C., Dick, J. T., & Dunn, A. M. (2004). Roles of parasites in animal invasions. *Trends in Ecology & Evolution*, 19(7), 385–390. doi:10.1016/j.tree.2004.05.002
- Prentis, P. J., Wilson, J. R. U., Dormontt, E. E., Richardson, D. M., & Lowe, A. J. (2008). Adaptive evolution in invasive species. *Trends in Plant Science*, 13(6), 288–294. doi:10.1016/j.tplants.2008.03.004
- Quesada, H., Beynon, C. M., & Skibinski, D. O. F. (1995). A mitochondrial DNA discontinuity in the mussel *Mytilus galloprovincialis* Lmk: Pleistocene vicariance biogeography and secondary intergradation. *Molecular Biology and Evolution*, 12(3), 521–524. doi:10.1093/oxfordjournals.molbev.a040227
- Quesada, H., Zapata, C., & Alvarez, G. (1995). A multilocus allozyme discontinuity in the mussel *Mytilus galloprovincialis*: The interaction of ecological and life-history factors. *Marine Ecology Progress Series*, 116, 99–115.
- R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., ... Westram, A. M. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, 30(8), 1450–1477. doi:10.1111/jeb.13047
- Rawson, P. D., & Hilbish, T. J. [Thomas J.]. (1995). Evolutionary relationships among the male and female mitochondrial DNA lineages in the *Mytilus edulis* species complex. *Molecular Biology and Evolution*. doi:10.1093/oxfordjournals.molbev.a040266
- Rhymer, J. M., & Simberloff, D. (1996). Extinction By Hybridization and Introgression. *Annual Review of Ecology and Systematics*, 27, 83–109. doi:10.1146/annurev.ecolsys.27.1.83
- Riquet, F., Simon, A., & Bierne, N. (2017). Weird genotypes? Don't discard them, transmissible cancer could be an explanation. *Evolutionary Applications*, 10, 140–145. doi:10.1111/eva.12439
- Rius, M., Turon, X., Bernardi, G., Volckaert, F. A. M., & Viard, F. (2015). Marine invasion genetics: From spatio-temporal patterns to evolutionary outcomes. *Biological Invasions*, 17(3), 869–885. doi:10.1007/s10530-014-0792-0
- Rivkin, L. R., Santangelo, J. S., Alberti, M., Aronson, M. F. J., de Keyser, C. W., Diamond, S. E., ... Johnson, M. T. J. (2018). A roadmap for urban evolutionary ecology. *Evolutionary Applications*, 12(3), 384–398. doi:10.1111/eva.12734
- Roberts, D. (1976). 3. Mussels and pollution. In B. L. Bayne (Ed.), *Marine mussels: Their ecology and physiology* (Bayne B. L., Vol. 10, pp. 81–120). Cambridge University Press.
- Roesti, M., Moser, D., & Berner, D. (2013). Recombination in the threespine stickleback genome-patterns and consequences. *Molecular Ecology*, 22(11), 3014–3027. doi:10.1111/mec.12322
- Roux, C., Fraïsse, C., Castric, V., Vekemans, X., Pogson, G. H., & Bierne, N. (2014). Can we continue to neglect genomic variation in introgression rates when inferring the history of speciation? A case study in a *Mytilus* hybrid zone. *Journal of Evolutionary Biology*, 27(8), 1662–1675. doi:10.1111/jeb.12425
- Roux, C., Fraïsse, C., Romiguier, J., Anciaux, Y., Galtier, N., & Bierne, N. (2016). Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence. *PLoS Biology*, 14(12), e2000234. doi:10.1371/journal.pbio.2000234

-
- Saarman, N. P., & Pogson, G. H. (2015). Introgression between invasive and native blue mussels (genus *Mytilus*) in the central California hybrid zone. *Molecular Ecology*, *24*(18), 4723–4738. doi:10.1111/mec.13340
- Schierenbeck, K. A., & Ellstrand, N. C. (2009). Hybridization and the evolution of invasiveness in plants and other organisms. *Biological Invasions*, *11*(5), 1093–1105. doi:10.1007/s10530-008-9388-x
- Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., ... Przeworski, M. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, *360*(6389), 656–660. doi:10.1126/science.aar3684
- Scribner, K. T., Page, K. S., & Bartron, M. L. (2000). Hybridization in freshwater fishes: A review of case studies and cytonuclear methods of biological inference. *Reviews in Fish Biology and Fisheries*, *10*, 293–323. doi:10.1023/A:1016642723238
- Seed, R. (1969). The incidence of the Pea crab, *Pinnotheres pisum* in the two types of *Mytilus* (Mollusca: Bivalvia) from Padstow, south-west England. *Journal of Zoology*, *158*(4), 413–420. doi:10.1111/j.1469-7998.1969.tb02158.x
- Semagn, K., Babu, R., Hearne, S., & Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement. *Molecular Breeding*, *33*, 1–14. doi:10.1007/s11032-013-9917-x
- Simon, A., Bierne, N., & Welch, J. J. (2018). Coadapted genomes and selection on hybrids: Fisher's geometric model explains a variety of empirical patterns. *Evolution Letters*, *2*(5), 472–498. doi:10.1002/evl3.66
- Simon, A., Fraïsse, C., El Ayari, T., Liutard-Haag, C., Strelkov, P., Welch, J. J., & Bierne, N. (2019). Local introgression at two spatial scales in mosaic hybrid zones of mussels. *bioRxiv*. doi:10.1101/818559
- Skibinski, D. O. F., Beardmore, J. A., & Ahmad, M. (1978). Genetic aids to the study of closely related taxa of the genus *Mytilus*. In B. Battaglia & J. A. Beardmore (Eds.), *Marine organisms: Genetics, ecology, and evolution* (pp. 469–486). Plenum Publishing Corporation.
- Strayer, D. L., Eviner, V. T., Jeschke, J. M., & Pace, M. L. (2006). Understanding the long-term effects of species invasions. *Trends in Ecology & Evolution*, *21*(11), 645–651. doi:10.1016/j.tree.2006.07.007
- Strelkov, P., Katolikova, M., & Väinölä, R. (2017). Temporal change of the Baltic Sea–North Sea blue mussel hybrid zone over two decades. *Marine Biology*, *164*(11). doi:10.1007/s00227-017-3249-z
- Suarez, A. V., & Tsutsui, N. D. (2008). The evolutionary consequences of biological invasions. *Molecular Ecology*, *17*(1), 351–360. doi:10.1111/j.1365-294X.2007.03456.x
- Sylvester, F., Kalaci, O., Leung, B., Lacoursière-Roussel, A., Murray, C. C., Choi, F. M., ... MacIsaac, H. J. (2011). Hull fouling as an invasion vector: Can simple models explain a complex problem? *Journal of Applied Ecology*, *48*(2), 415–423. doi:10.1111/j.1365-2664.2011.01957.x
- Thompson, K. A., Rieseberg, L. H., & Schluter, D. (2018). Speciation and the City. *Trends in Ecology & Evolution*, *33*(11), 815–826. doi:10.1016/j.tree.2018.08.007
- Torchin, M. E., Lafferty, K. D., & Kuris, A. M. (2002). Parasites and marine invasions. *Parasitology*, *124*(7), 137–151. doi:10.1017/S0031182002001506
- Väinölä, R., & Hvilson, M. M. (1991). Genetic divergence and a hybrid zone between Baltic and North Sea *Mytilus* populations (Mytilidae: Mollusca). *Biological Journal of the Linnean Society*, *43*(2), 127–148. doi:10.1111/j.1095-8312.1991.tb00589.x

- van Etten, J. (2017). R Package gdistance: Distances and Routes on Geographical Grids. *Journal of Statistical Software*, 76(13). doi:10.18637/jss.v076.i13
- Viard, F., David, P., & Darling, J. A. (2016). Marine invasions enter the genomic era: Three lessons from the past, and the way forward. *Current Zoology*, 62(6), 629–642. doi:10.1093/cz/zow053
- Wang, J. (2017). The computer program STRUCTURE for assigning individuals to populations: Easy to use but easier to misuse. *Molecular Ecology Resources*, 17(5), 981–990. doi:10.1111/1755-0998.12650
- Węśławski, J. M., & Kotwicki, L. (2018). Macro-plastic litter, a new vector for boreal species dispersal on Svalbard. *Polish Polar Research*, 39(1), 165–174. doi:10.24425/118743
- Willis, G. L., & Skibinski, D. O. F. (1992). Variation in strength of attachment to the substrate explains differential mortality in hybrid mussel (*Mytilus galloprovincialis* and *M. edulis*) populations. *Marine Biology*, 112(3), 403–408. doi:10.1007/BF00356285
- Zardi, G. I., McQuaid, C. D., Jacinto, R., Lourenço, C. R., Serrão, E. A., & Nicastro, K. R. (2018). Re-assessing the origins of the invasive mussel *Mytilus galloprovincialis* in southern Africa. *Marine and Freshwater Research*, 69(4), 607. doi:10.1071/MF17132
- Zbawicka, M., Drywa, A., Śmietanka, B., & Wenne, R. (2012). Identification and validation of novel SNP markers in European populations of marine *Mytilus* mussels. *Marine Biology*, 159(6), 1347–1362. doi:10.1007/s00227-012-1915-8
- Zbawicka, M., Gardner, J. P. A., & Wenne, R. (2019). Cryptic diversity in smooth-shelled mussels on Southern Ocean islands: Connectivity, hybridisation and a marine invasion. *Frontiers in Zoology*, 16, 32. doi:10.1186/s12983-019-0332-y
- Zbawicka, M., Trucco, M. I., & Wenne, R. (2018). Single nucleotide polymorphisms in native South American Atlantic coast populations of smooth shelled mussels: Hybridization with invasive European *Mytilus galloprovincialis*. *Genetics Selection Evolution*, 50, 5. doi:10.1186/s12711-018-0376-z

Supplementary information

A Admixture time

To estimate simply admixture time, *Structure* runs were used as indicated in Falush, Stephens, and Pritchard (2003) using the linkage model. For each port (or year for Cherbourg) only admixed dock mussels individuals were considered with additional reference populations being *edu_eu_south* and *gallo_med*. The *popflag* column was set to 1 for reference individuals to better estimate the parental allele frequencies. Only markers on the genetic map were used to avoid bias through undetected linkage with markers that could not be included in the genetic map analysis. Options used: *LINKAGE* = 1, *PFROMPOPFLAGONLY* = 1, *NOADMIX* = 0, *LOG10RSTART* = -2, *LOG10RMIN* = -3, *LOG10RMAX* = 2, *LOG10RPROPSD* = 0.1. See the parameter files for more details. Results of average r for 25 runs are presented in table S14, the number of generations is given by $r \times 100$ as genetic distance are given in cM.

B Supplementary Figures

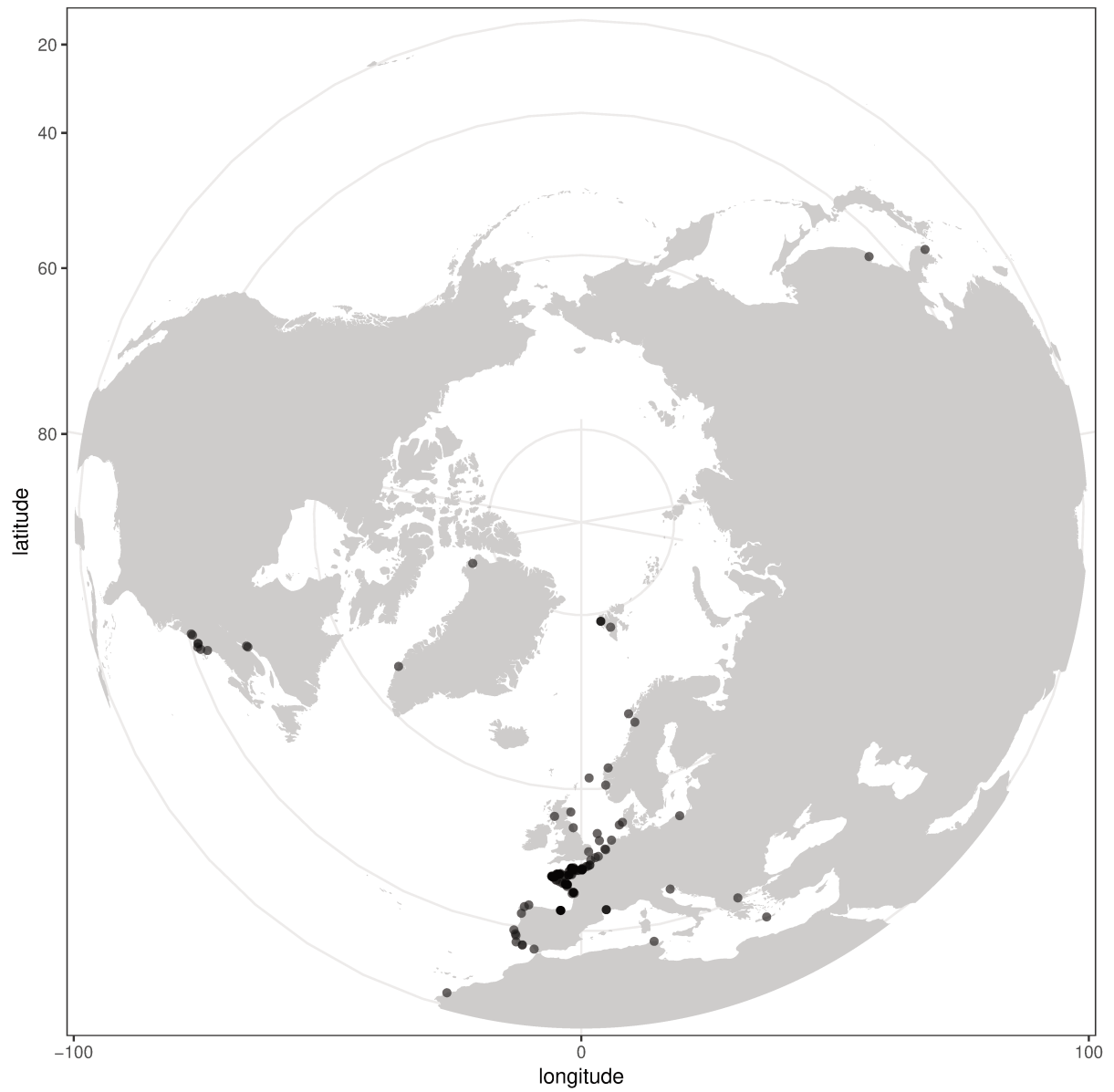


Figure S1: Location of all samples considered in the study.

Steps data

	Tip1	96 DW tip comb	
	Pick-Up	Tip Plate	
	Bind	Sample	
	Beginning of step	Precollect	No
		Release time, speed	00:00:30, Bottom mix
	Mixing / heating:	Mixing time, speed	00:04:30, Half mix
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	3
		Collect time [s]	1
	Wash 1	MB3	
	Beginning of step	Precollect	No
		Release time, speed	00:00:10, Bottom mix
	Mixing / heating:	Mixing time, speed	00:01:00, Half mix
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	3
		Collect time [s]	1
	Wash 2	MB4	
	Beginning of step	Precollect	No
		Release time, speed	00:00:10, Bottom mix
	Mixing / heating:	Mixing time, speed	00:01:00, Half mix
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	3
		Collect time [s]	1
	Wash 3	MB5	
	Beginning of step	Precollect	No
		Release beads	No
	Mixing / heating:	Mixing time, speed	00:00:30, Slow
		Heating during mixing	No
	End of step	Postmix	No
		Collect beads	No
	Elution	Elution	
	Beginning of step	Precollect	No
		Release time, speed	00:00:15, Bottom mix
	Mixing / heating:	Shake 1 time, speed	00:01:00, Medium
		Shake 2 time, speed	00:00:10, Bottom mix
		Loop count	8
		Heating temperature [°C]	72
		Preheat	Yes
	End of step	Postmix	No
		Collect count	3
		Collect time [s]	30
	ReleaseBeads1	MB5	
		Release time, speed	00:00:10, Half mix
	Leave	MB5	

Figure S2: Program for DNA extraction with the Kingfisher Flex robot.

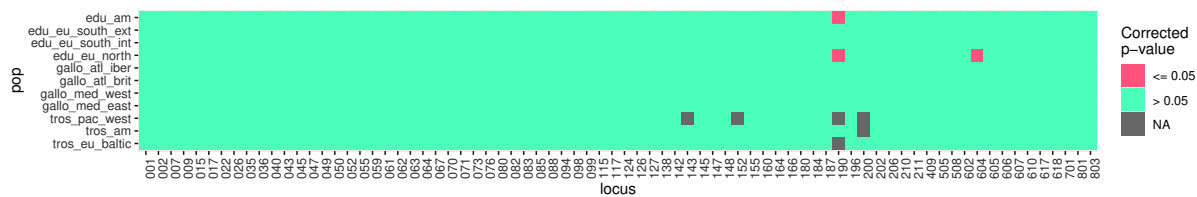


Figure S3: Hardy-Weinberg equilibrium test in each reference population. Benjamini-Yekutieli corrected p values are used.

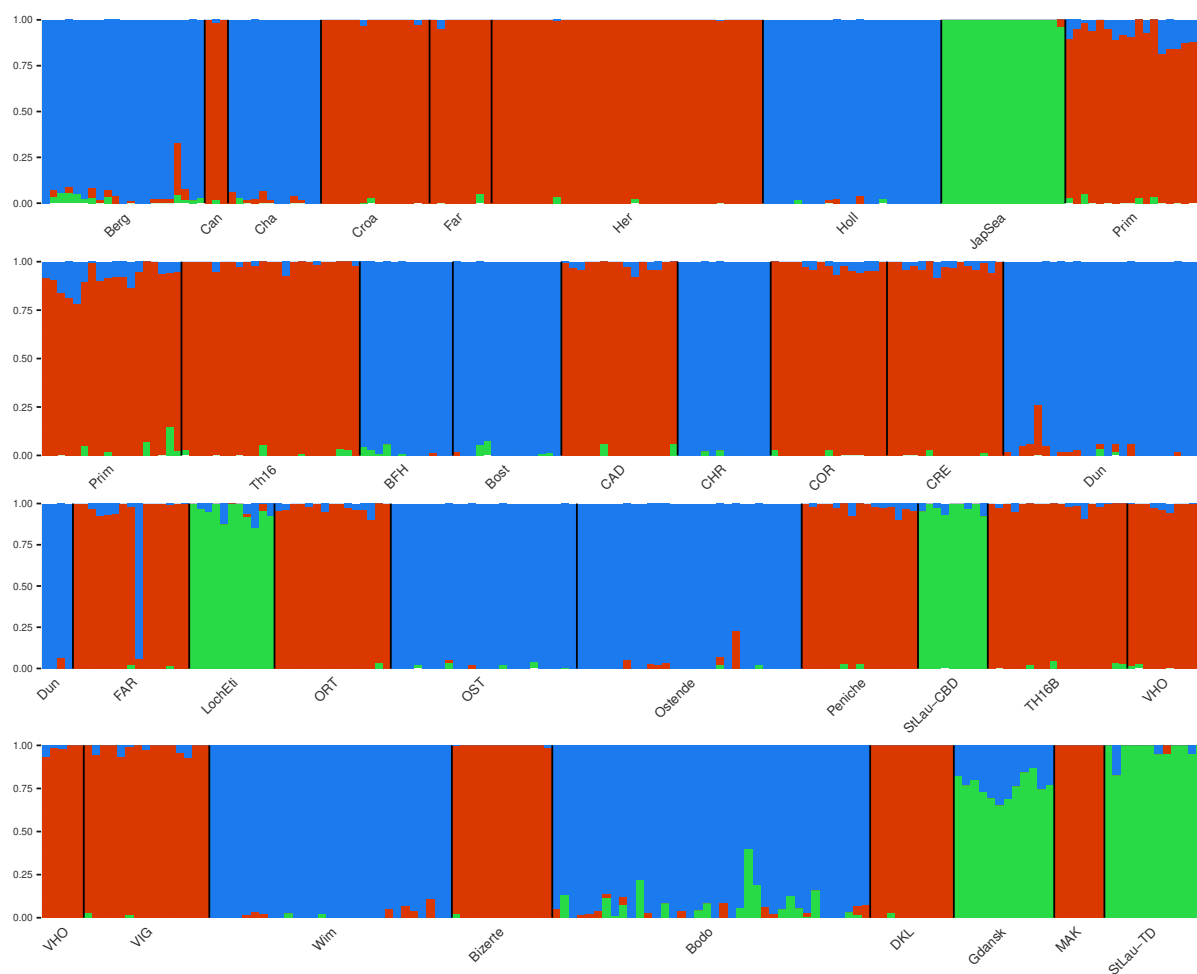


Figure S4: Admixture results for reference groups, $K = 3$, before filtration.

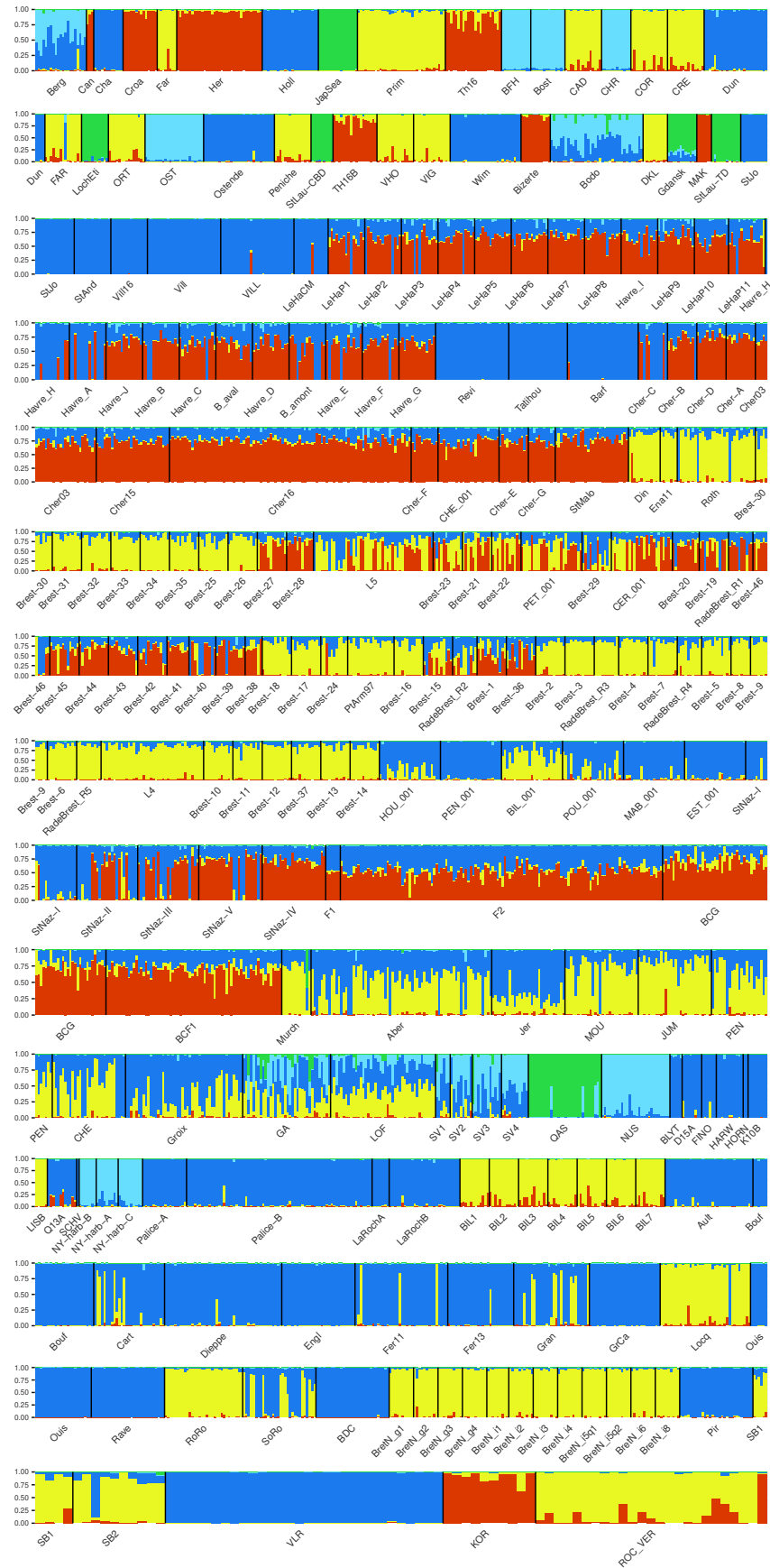


Figure S5: Structure results with all individuals, $K = 6$ major cluster.

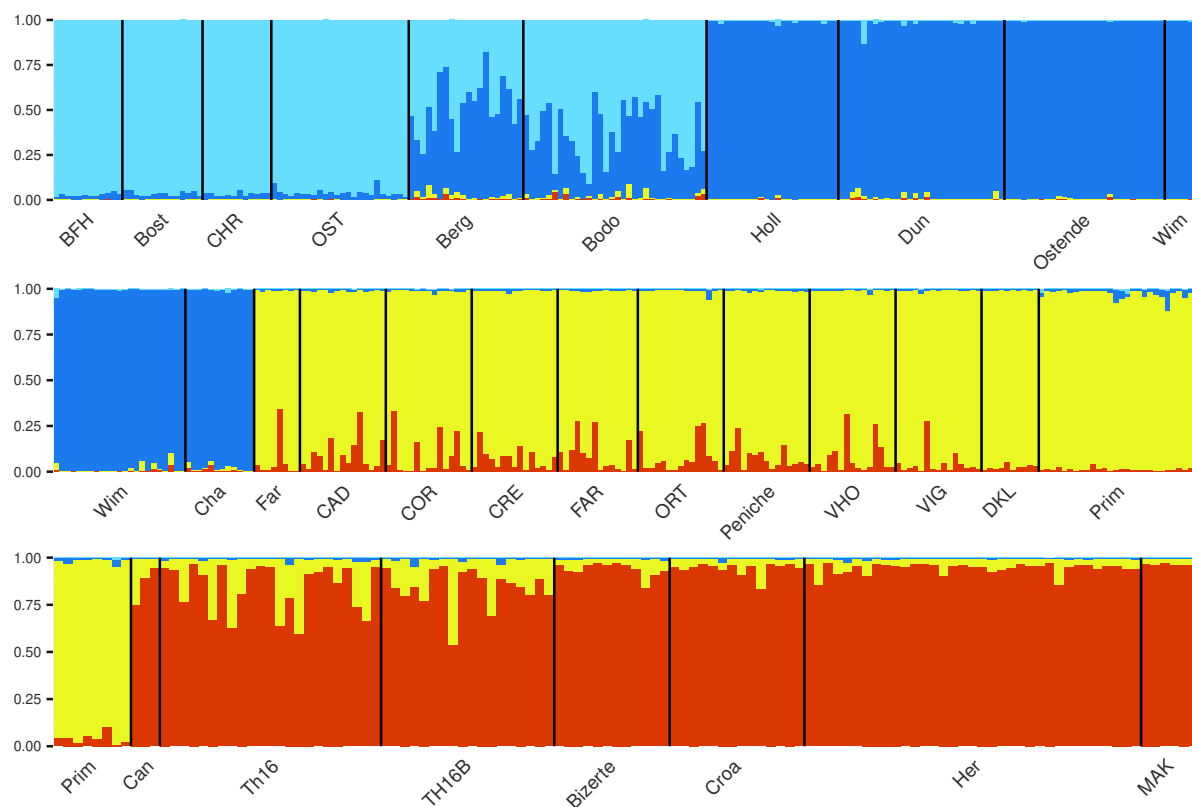


Figure S6: Structure results with all individuals excluding the ones with *M. trossulus* ancestry, $K = 4$ major cluster. Subset corresponding to reference populations only including individuals from the reference groups (table S4) after filtration.

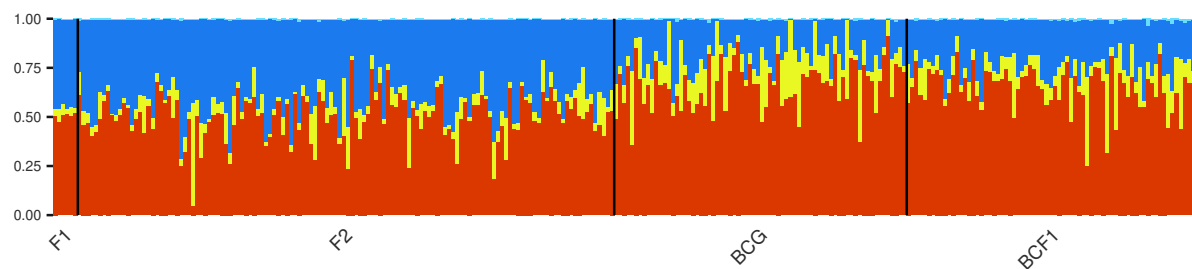


Figure S7: Subset of the lab crosses.

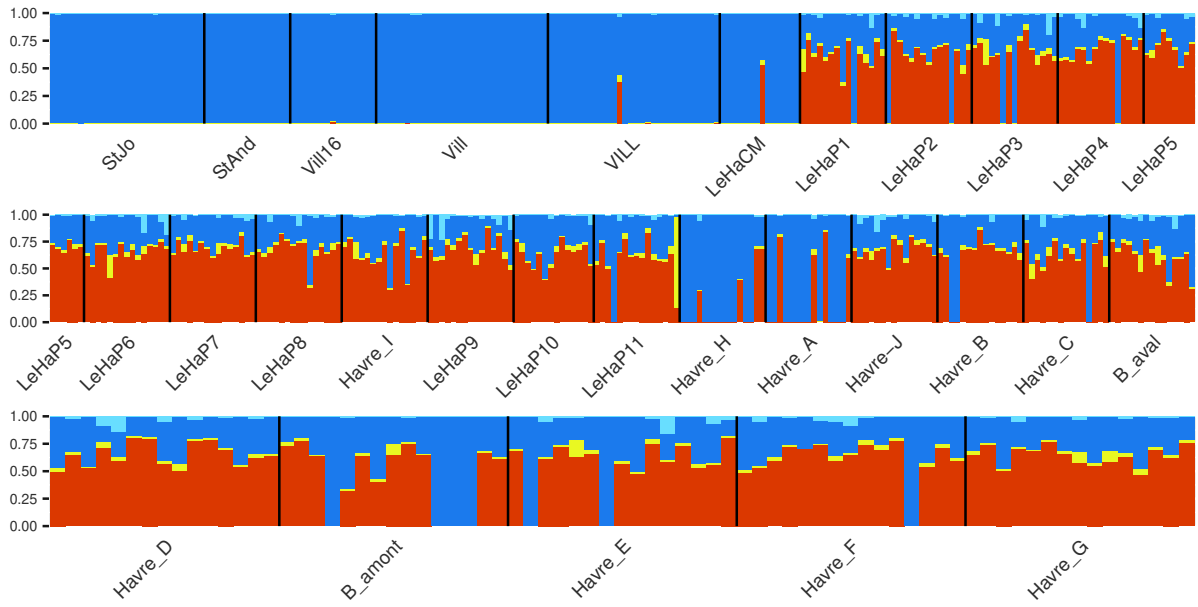


Figure S8: Subset of the port of Le Havre.

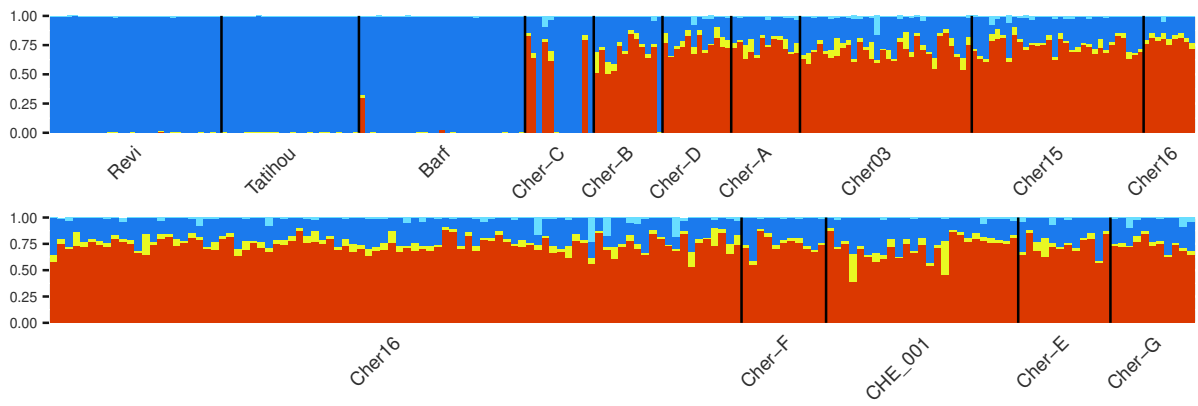


Figure S9: Subset of the port of Cherbourg.

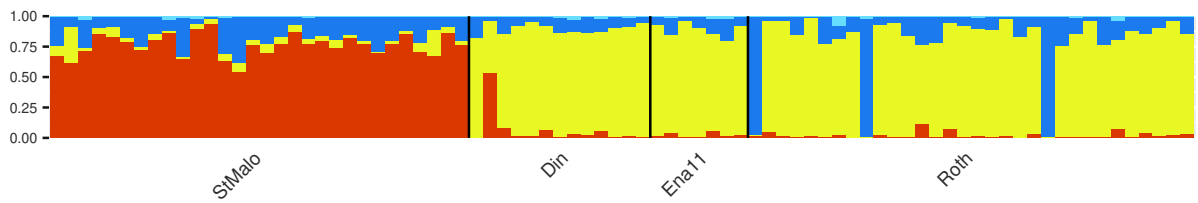


Figure S10: Subset of the port of Saint-Malo.

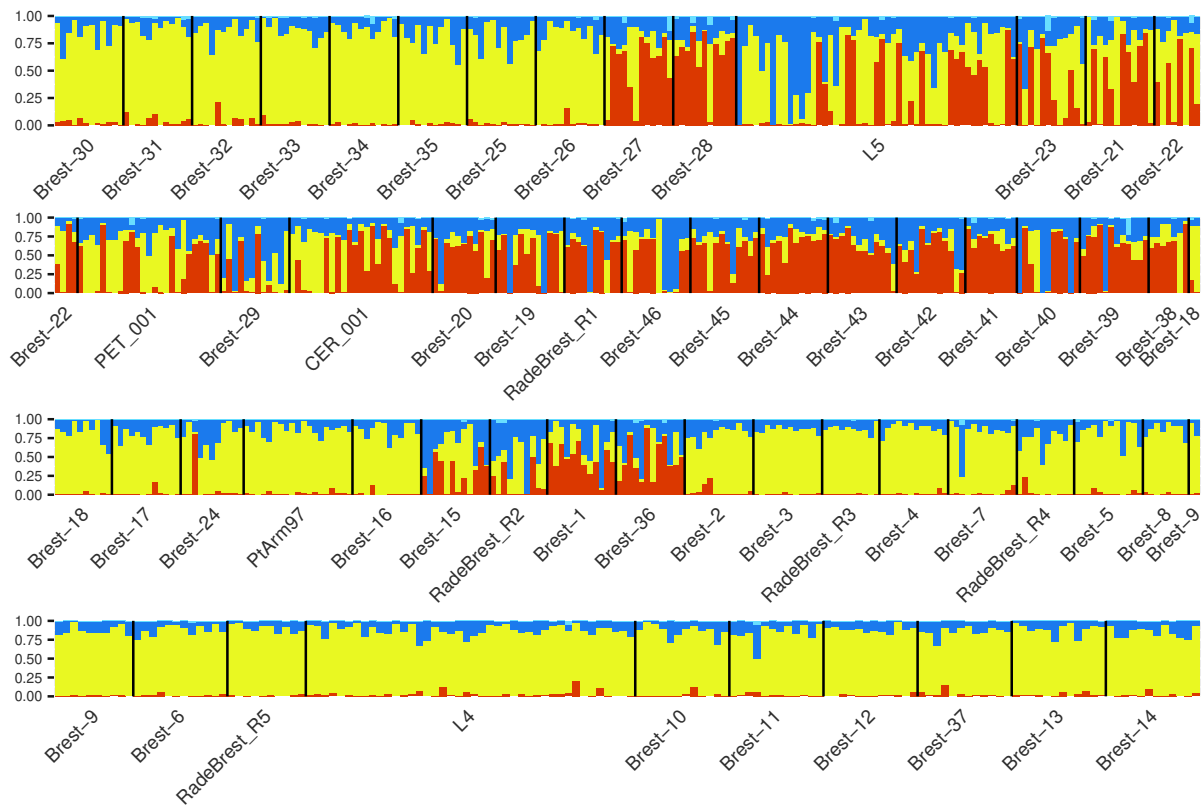


Figure S11: Subset of the bay of Brest area.

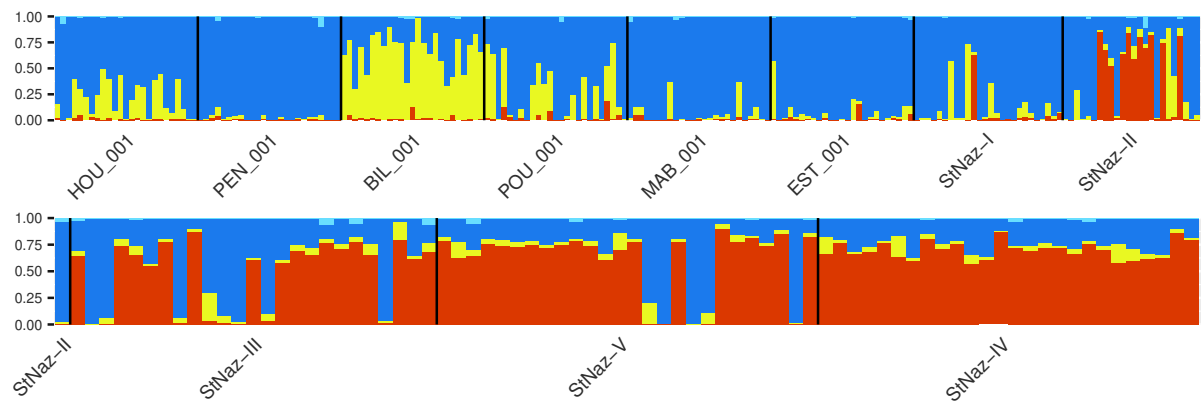


Figure S12: Subset of the port of Saint-Nazaire.

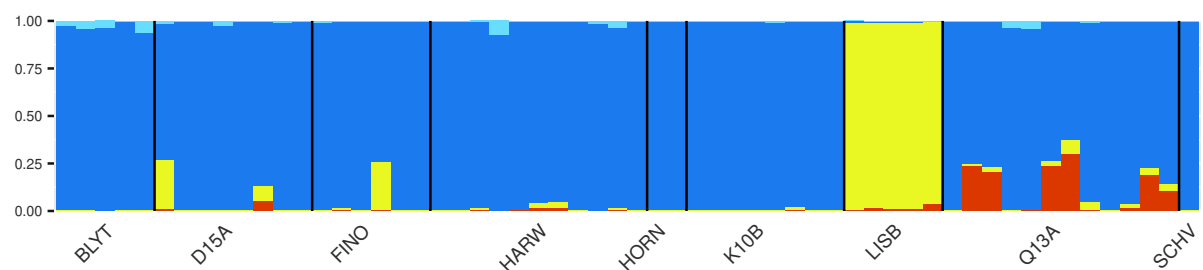


Figure S13: Subset of the populations studied in Coolen (2017).

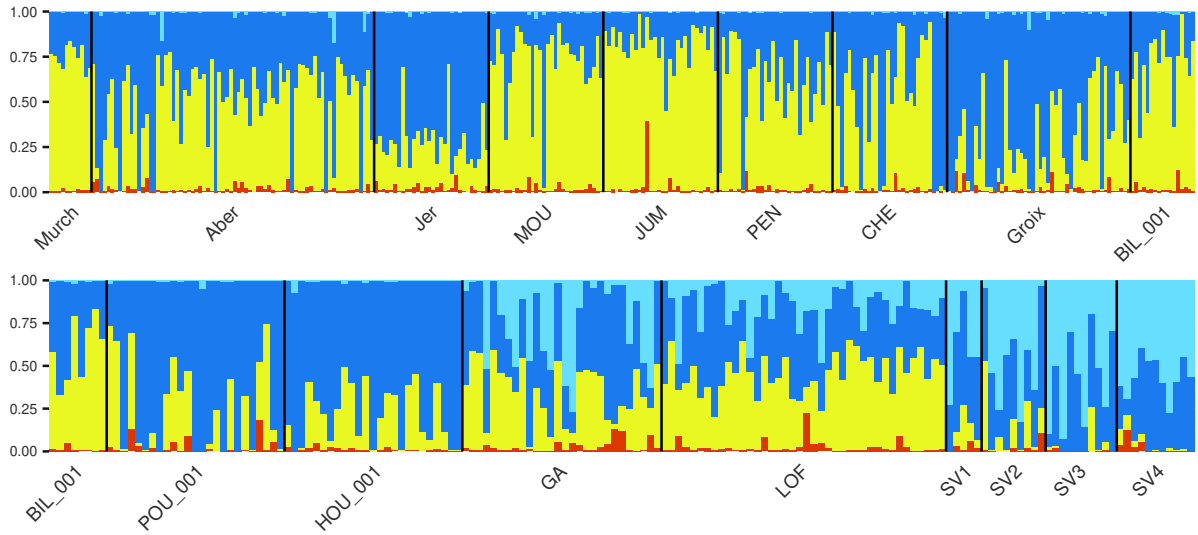


Figure S14: Subset of admixed populations with Atlantic *M. galloprovincialis* ancestries, including Natural admixed and Norway admixed. Populations LOF and SV1 to SV4 come from the Mathiesen et al. (2016) study.

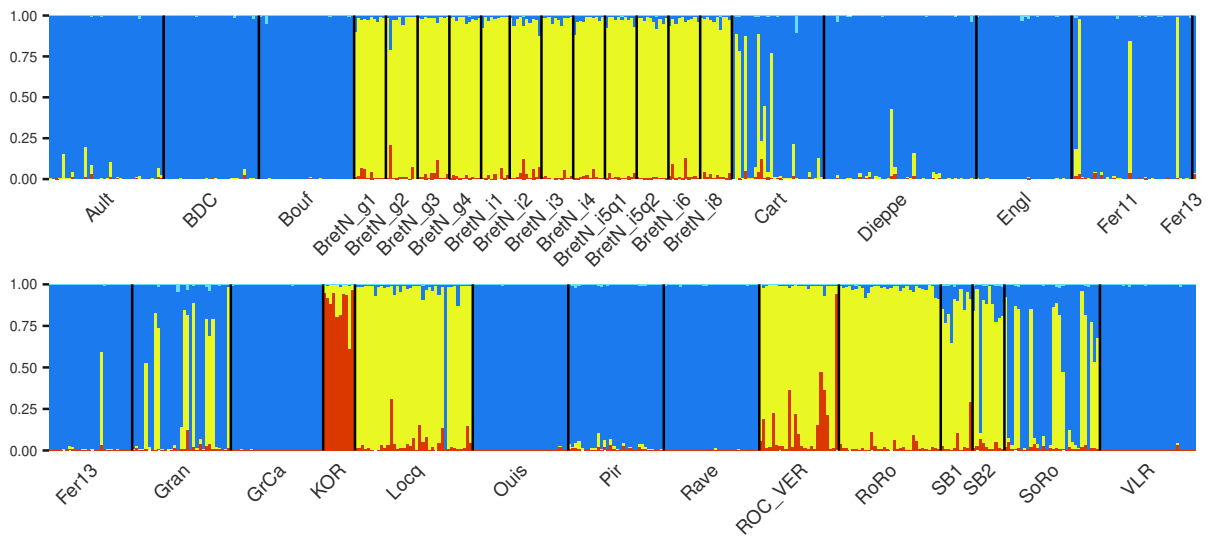


Figure S15: Subset of additional populations.

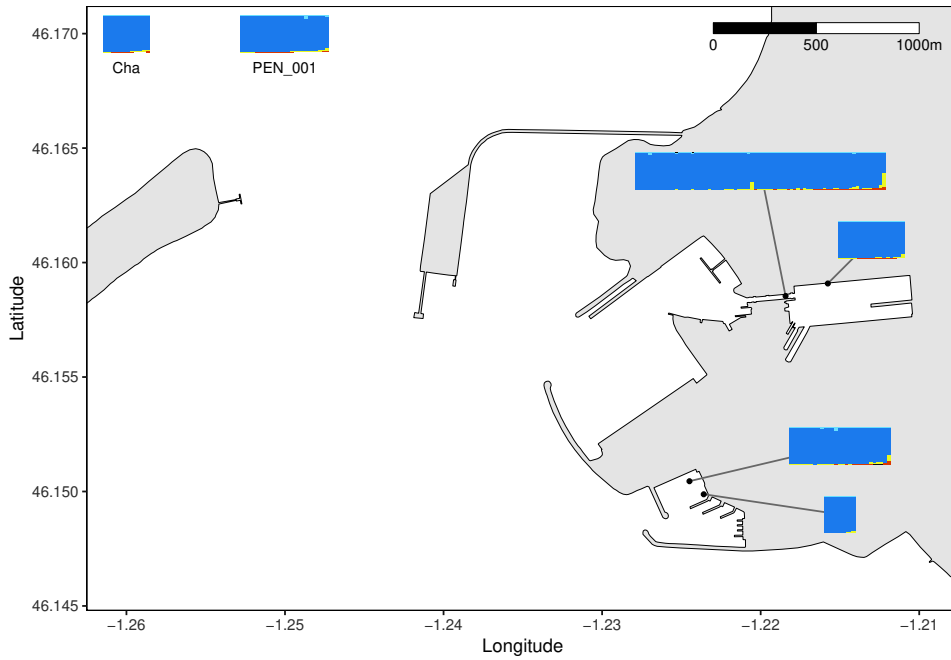


Figure S16: Map of the port of La Rochelle, bay of Biscay, France. Barplots represent individual ancestry proportions for each site. Barplots at the map edges correspond to distant populations. No Mediterranean *M. galloprovincialis* ancestry is detected in this port.

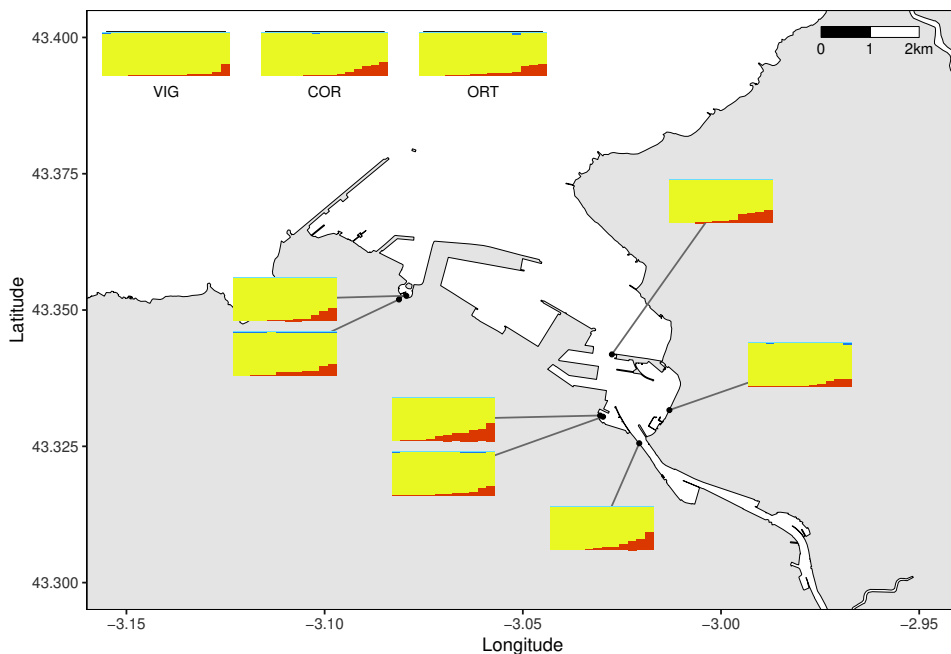


Figure S17: Map of the port of Bilbao, Basque Country, Spain. Sites inside the port do not exhibit different ancestry compositions than more distant populations

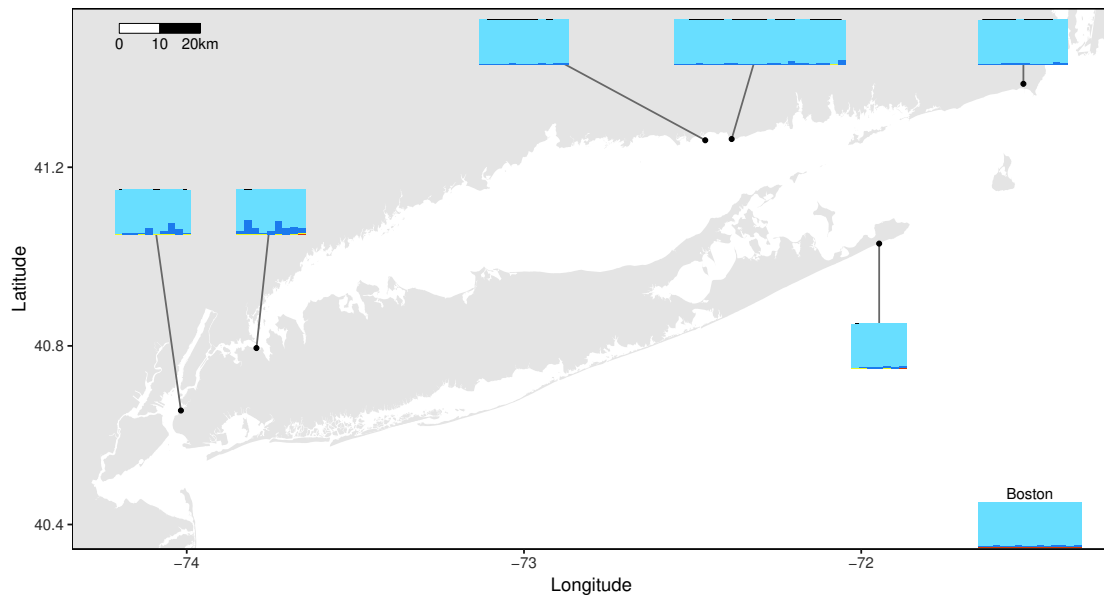


Figure S18: Map of the Long Island Sound, NY, USA. The two leftmost populations exhibit more European *M. edulis* ancestry than local American *M. edulis* populations.

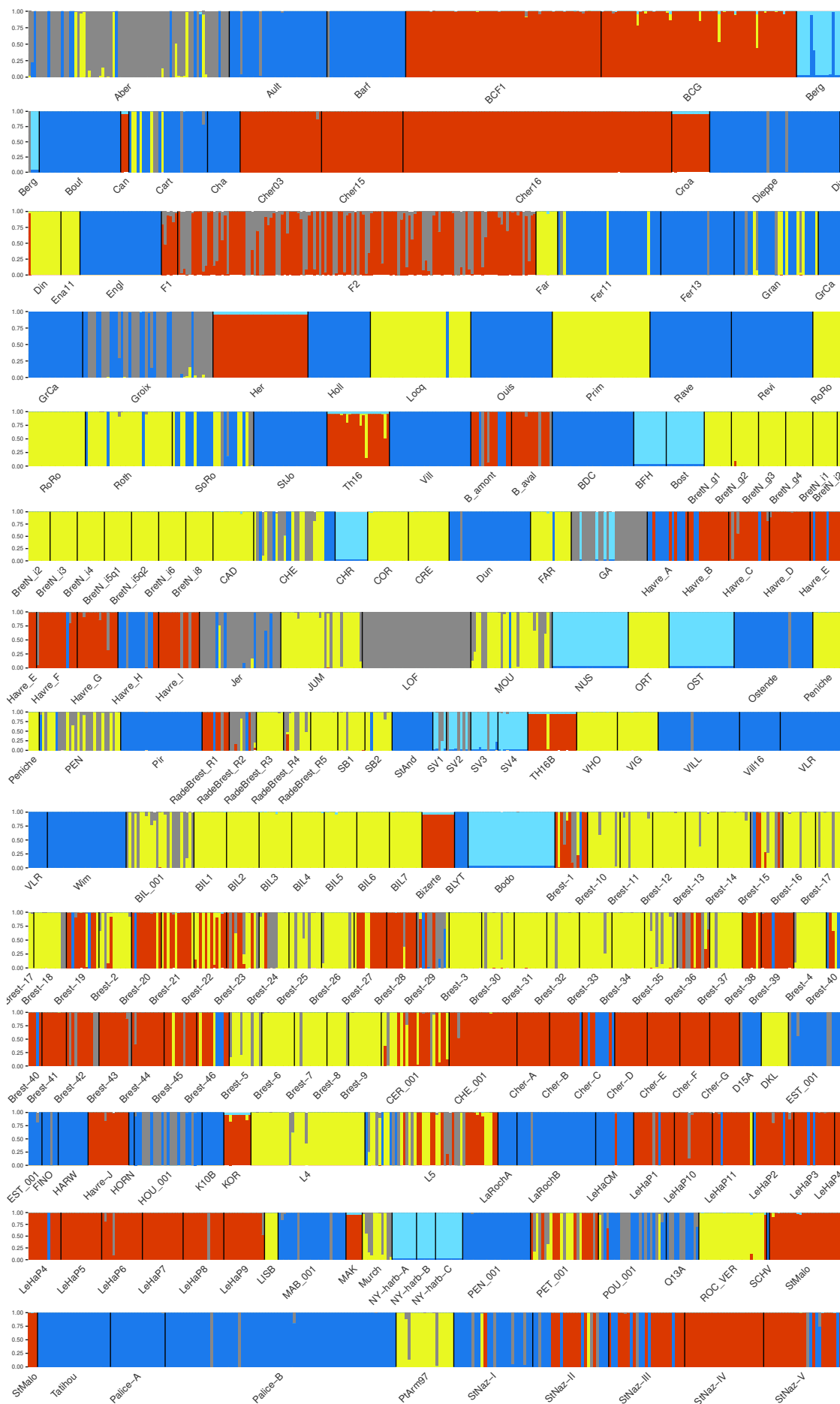


Figure S19: Structure without admixture model, $K = 5$

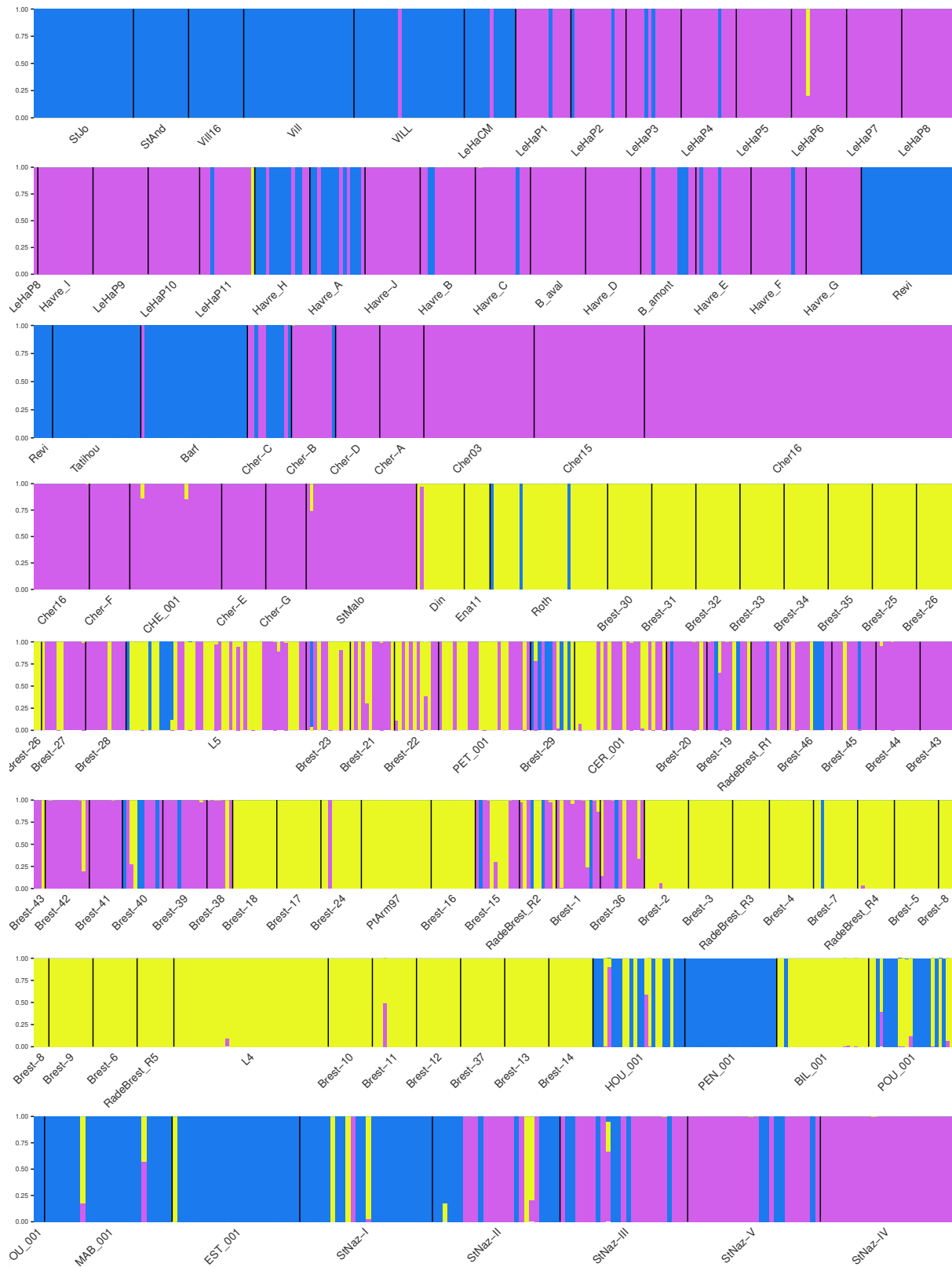


Figure S20: Structure without admixture model only with port and local populations, $K = 3$. This output is used to classify dock mussels for downstream analyses.

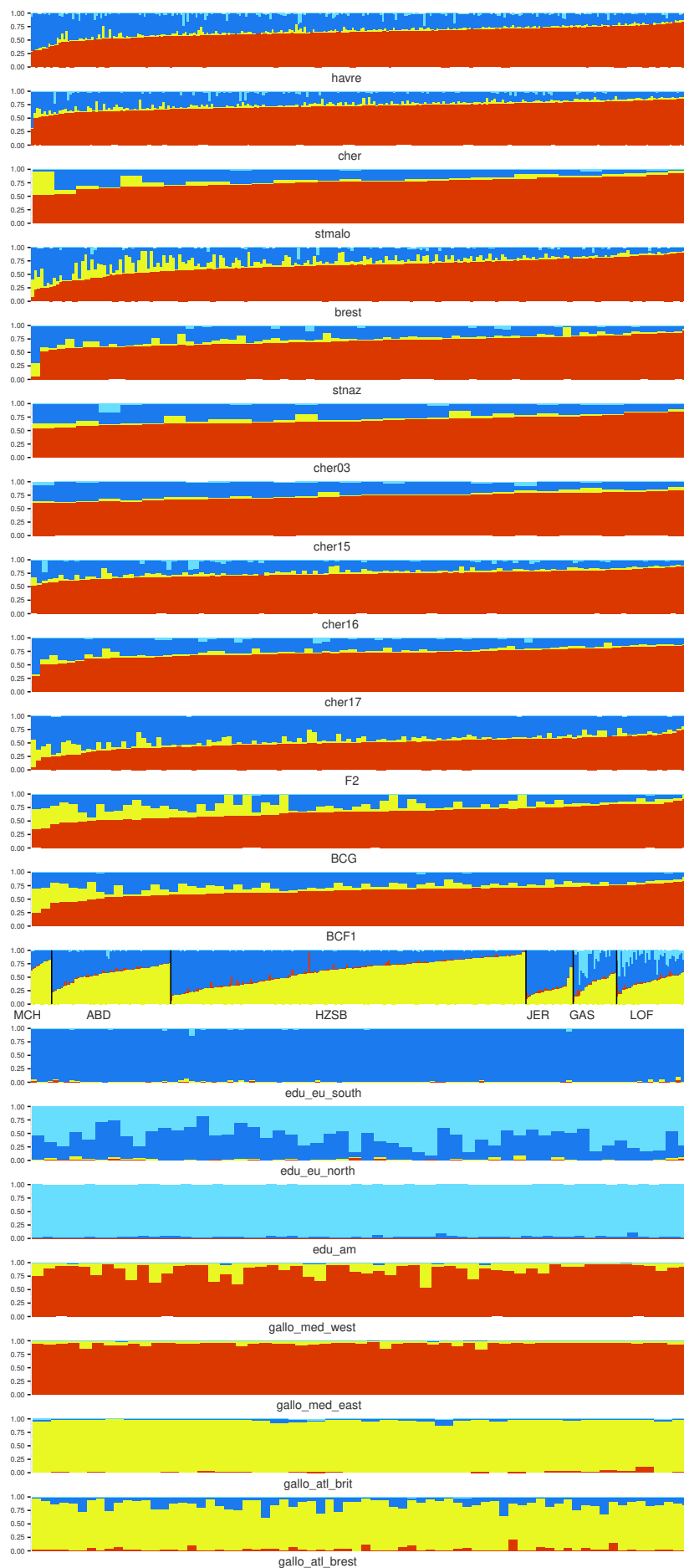


Figure S21: Structure plot of selected individuals for each group for downstream analyses

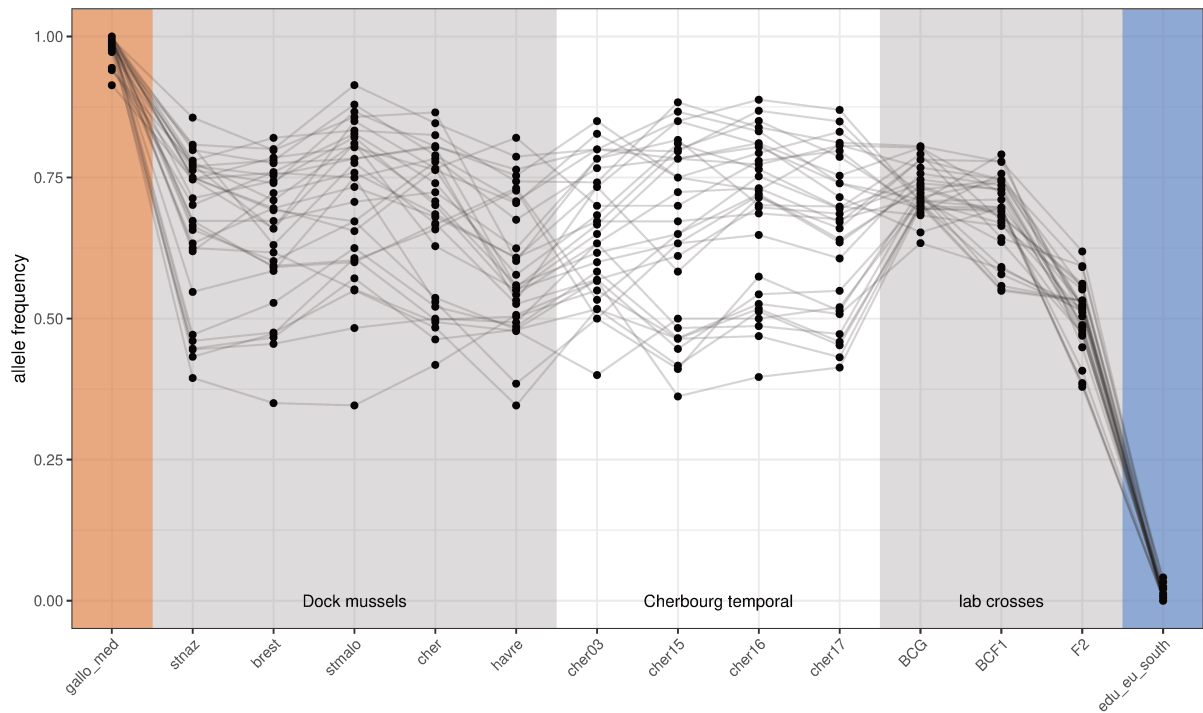


Figure S22: Allele frequencies of differentiated markers ($AFD > 0.9$) between Med. *M. galloprovincialis* and South-Eu. *M. edulis* in dock mussels, temporal samples of Cherbourg and laboratory crosses. Markers not heterozygous in the F1 cross were removed.

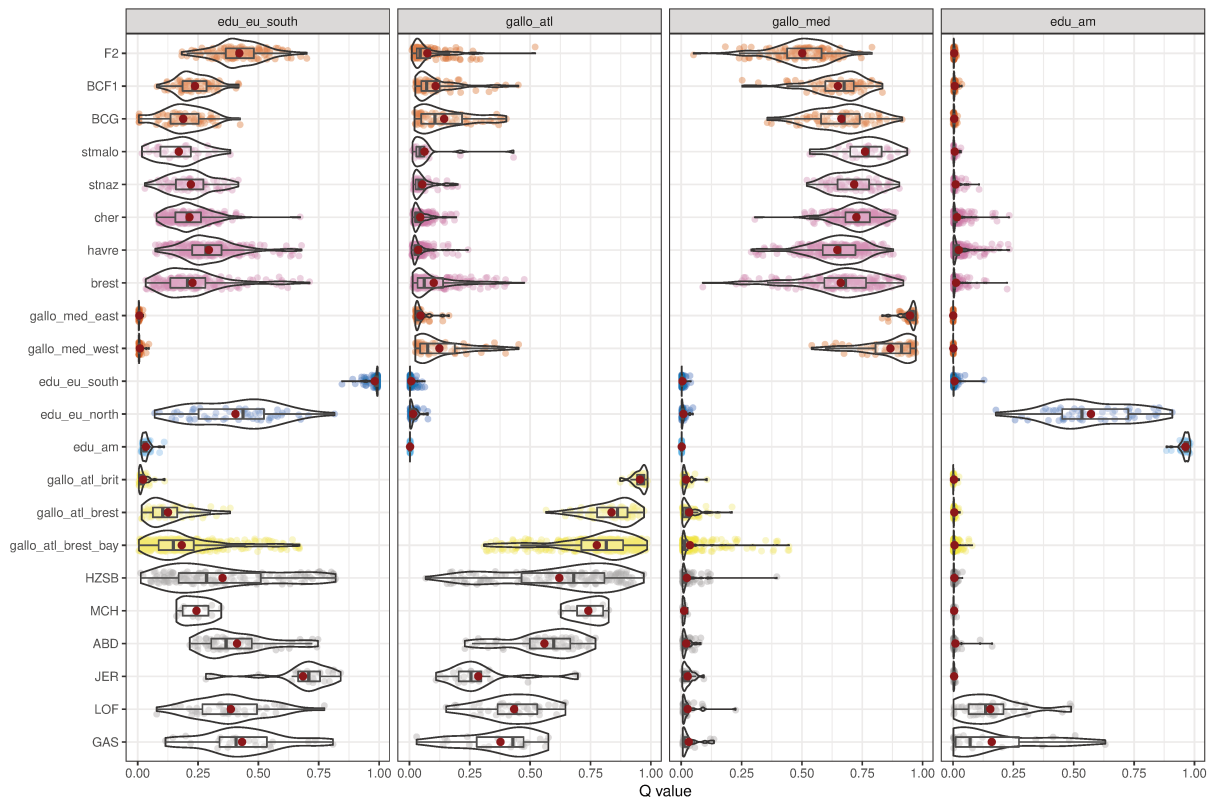


Figure S23: Ancestry levels for each Structure cluster in populations of admixed mussels and reference groups.

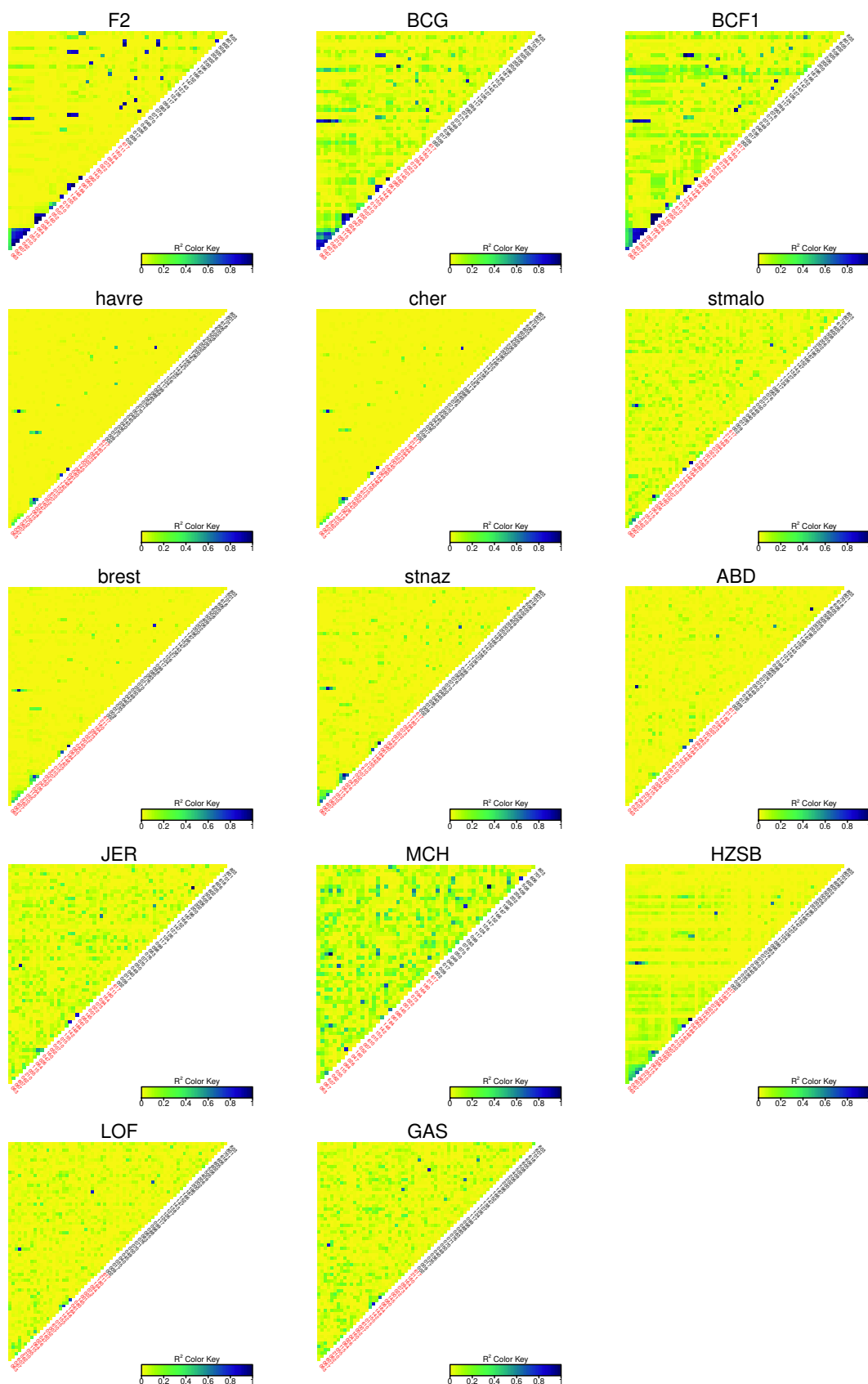


Figure S24: Linkage disequilibrium (R^2) in admixed populations. Markers are order following the genetic map (red labels) and then by name for loci that could be included in the genetic map construction.

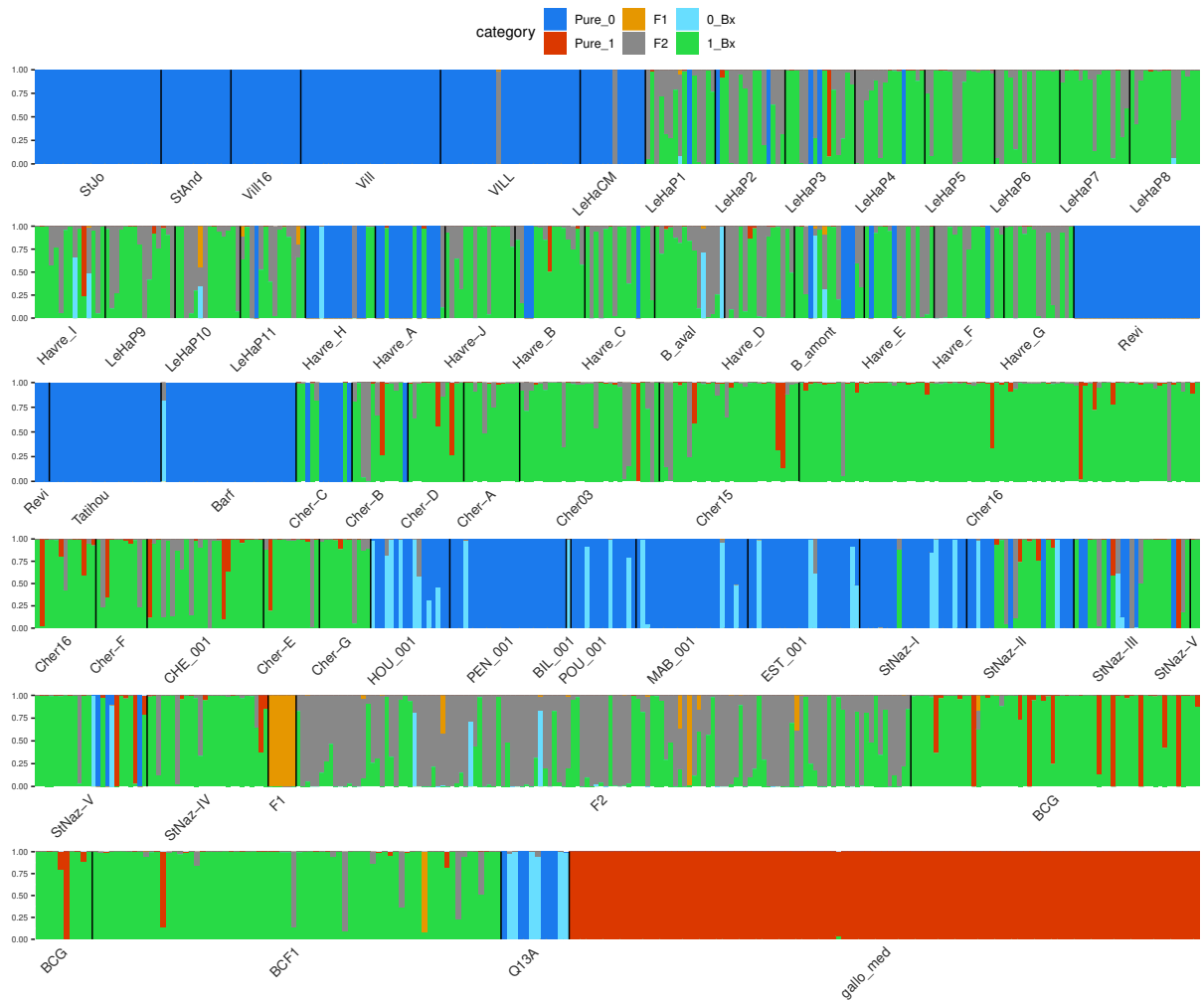


Figure S25: Newhybrids analysis including reference Mediterranean *M. galloprovincialis* (Pure_1) and local populations of all ports excluding Brest, containing South-Eu. *M. edulis* individuals (Pure_0). Lab crosses have been included and the results globally conform to the expected genotypes (F1, F2, BCG and BCF1). We hypothesise that pure Med. *M. galloprovincialis* assignments in ports are the result of marker sampling variance, as those assignments are also present in the lab BC samples.

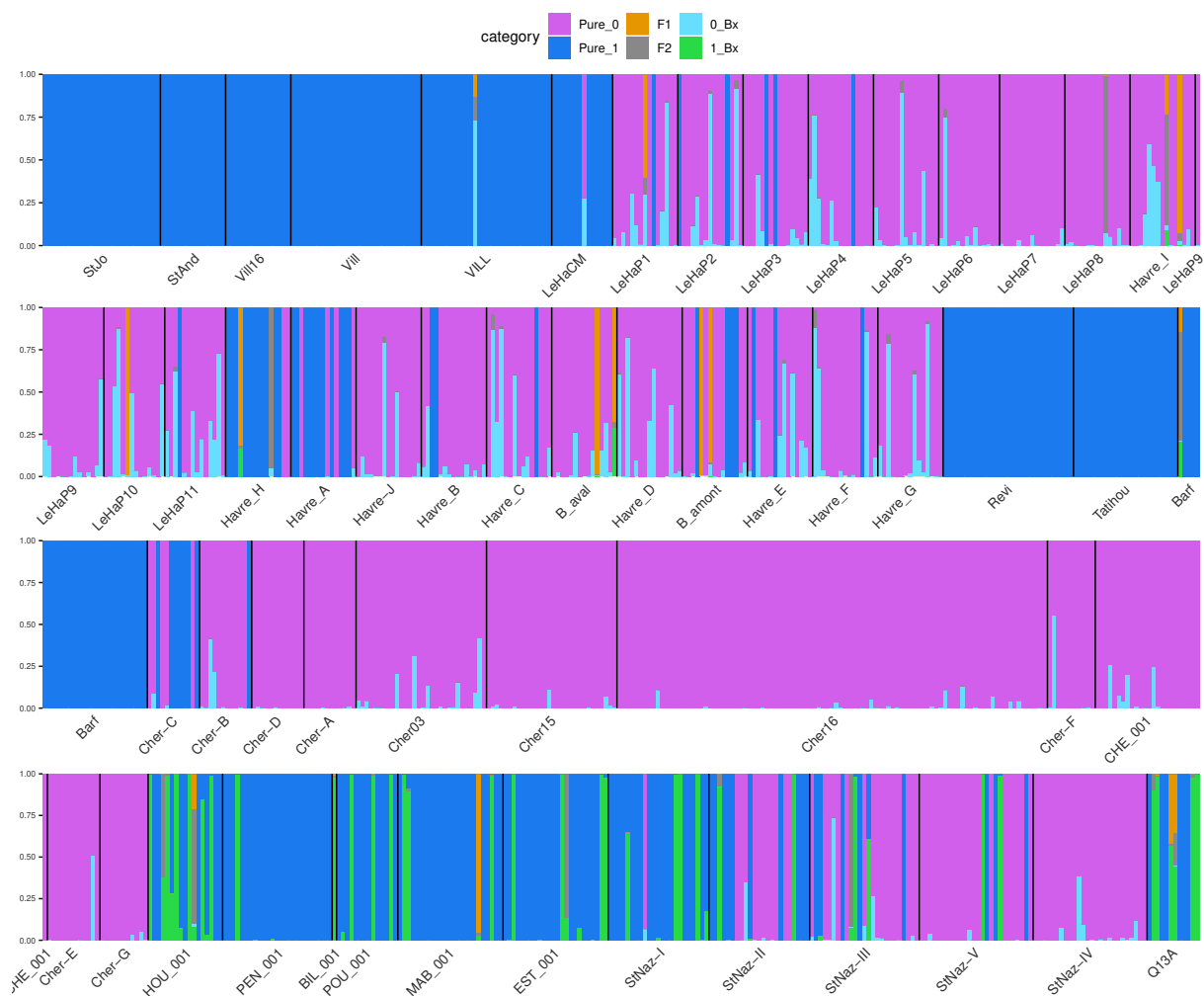


Figure S26: Newhybrids analysis comparing dock mussels (Pure_0) and the native Sout-Eu. *M. edulis* (Pure_1). In spite of the filtration to eliminate *Atl. M. galloprovincialis* genotypes, residual ancestry in individuals for populations around Saint-Nazaire (HOU_001 to StNaz-IV) produces spurious 1_Bx assignments.

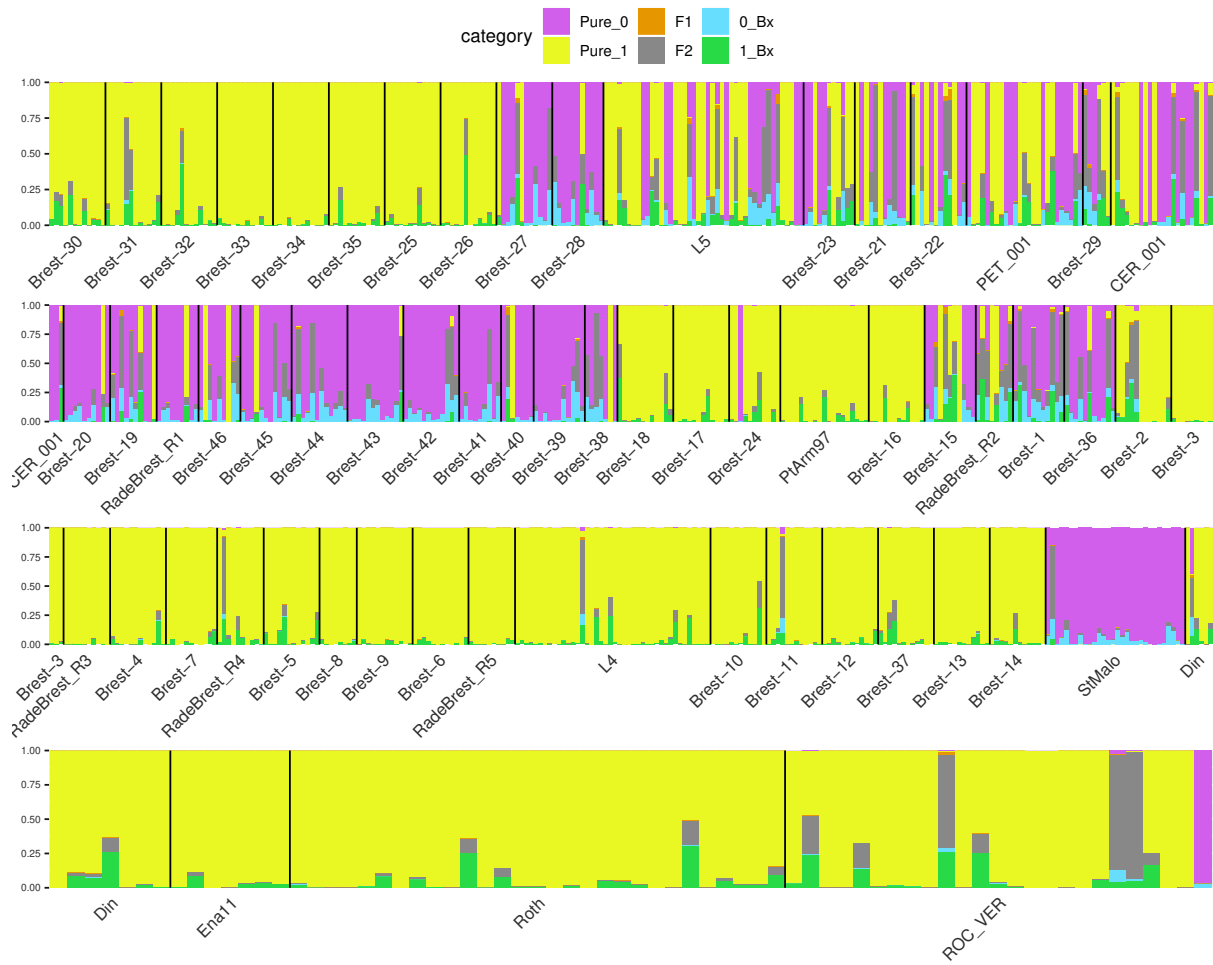


Figure S27: Newhybrids analysis comparing dock mussels (Pure_0) and native Atl. *M. galloprovincialis* (Pure_1).

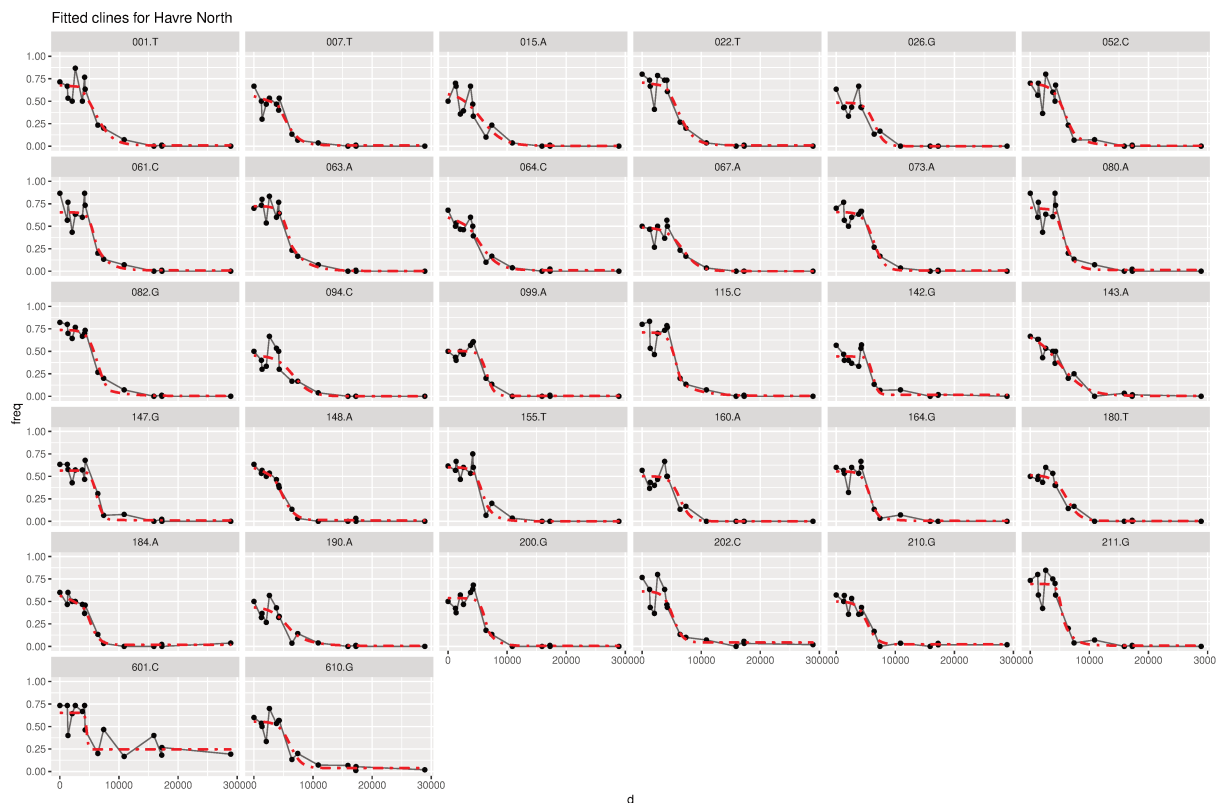


Figure S28: Details of geographic clines per marker fitted in Le Havre North transect. Each point is the allele frequency of the Med. *M. galloprovincialis* allele at one site.

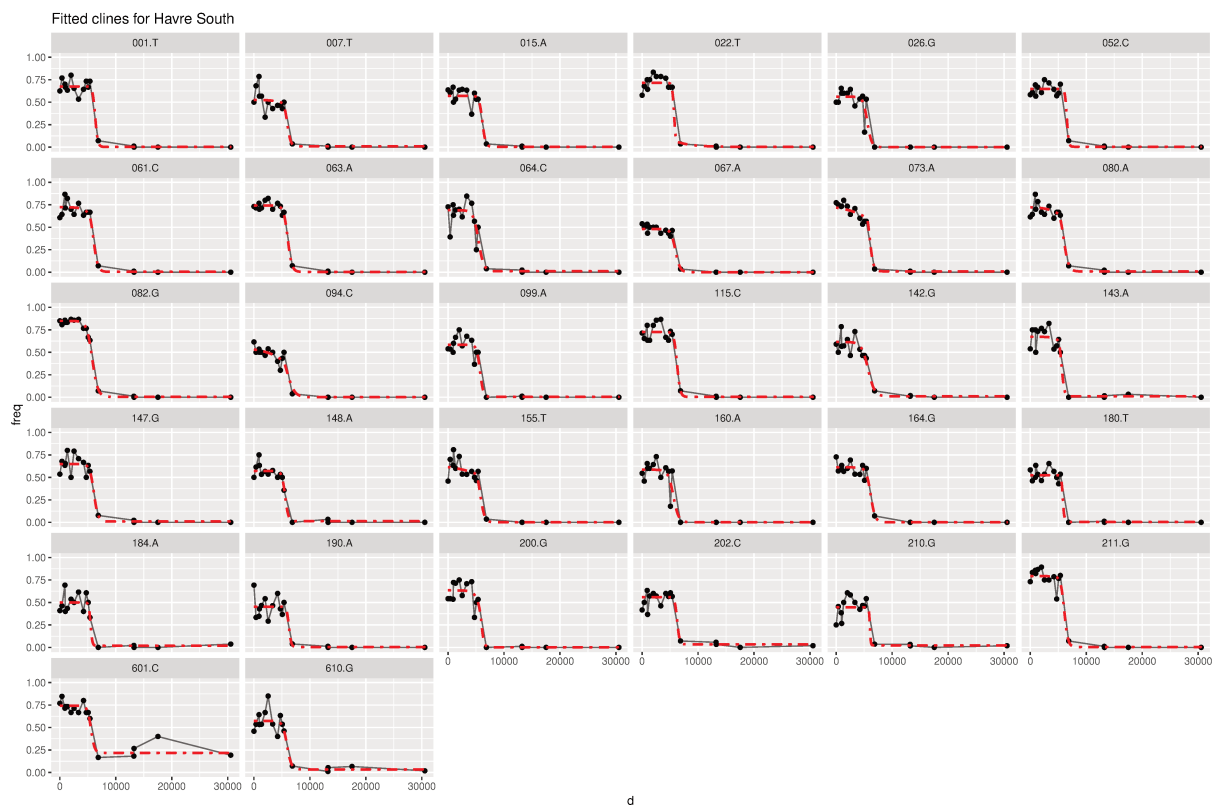


Figure S29: Details of geographic clines per marker fitted in Le Havre South transect.

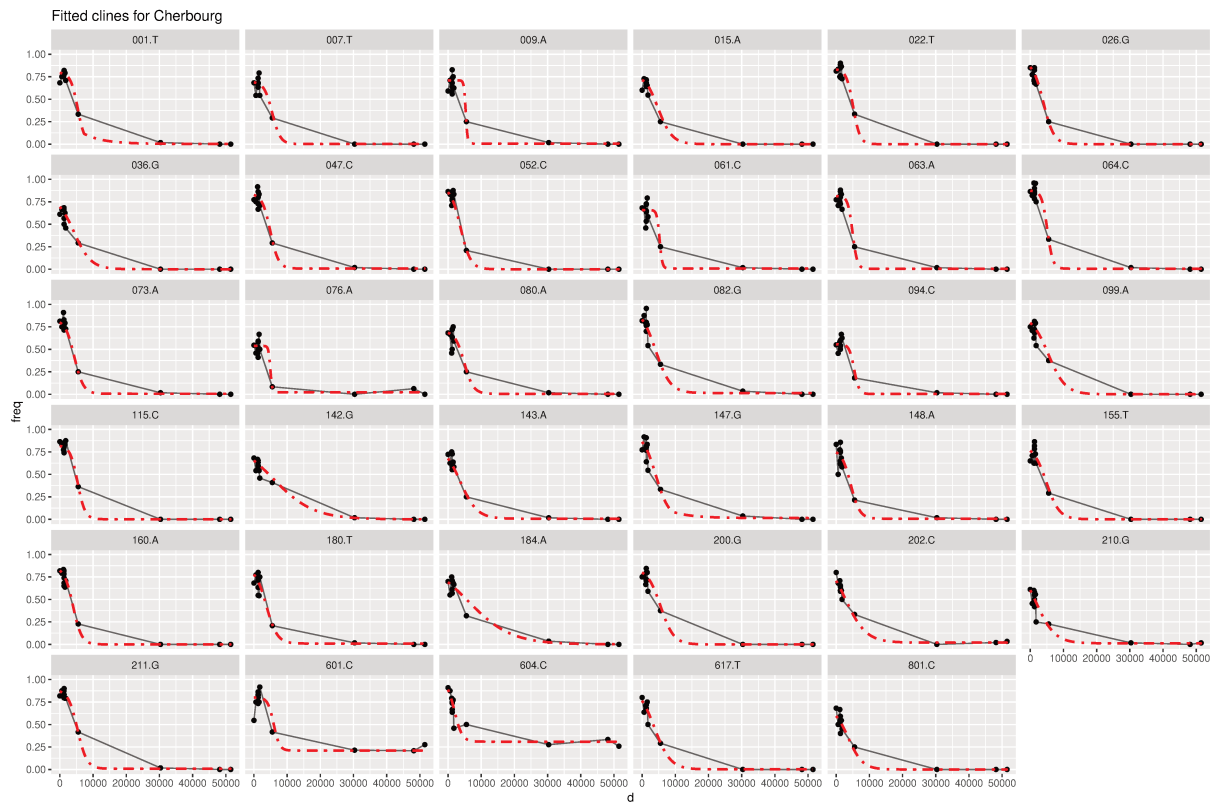


Figure S30: Details of geographic clines per marker fitted in Cherbourg.

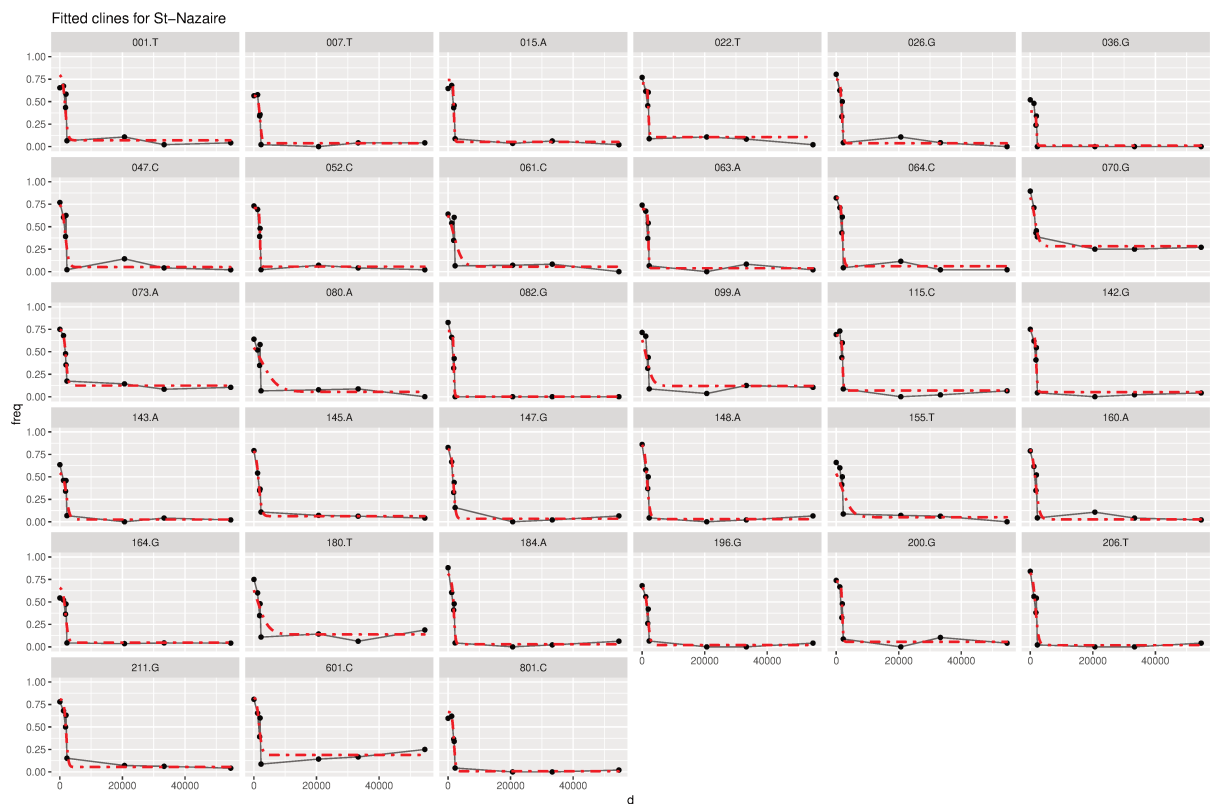


Figure S31: Details of geographic clines per marker fitted in Saint-Nazaire.

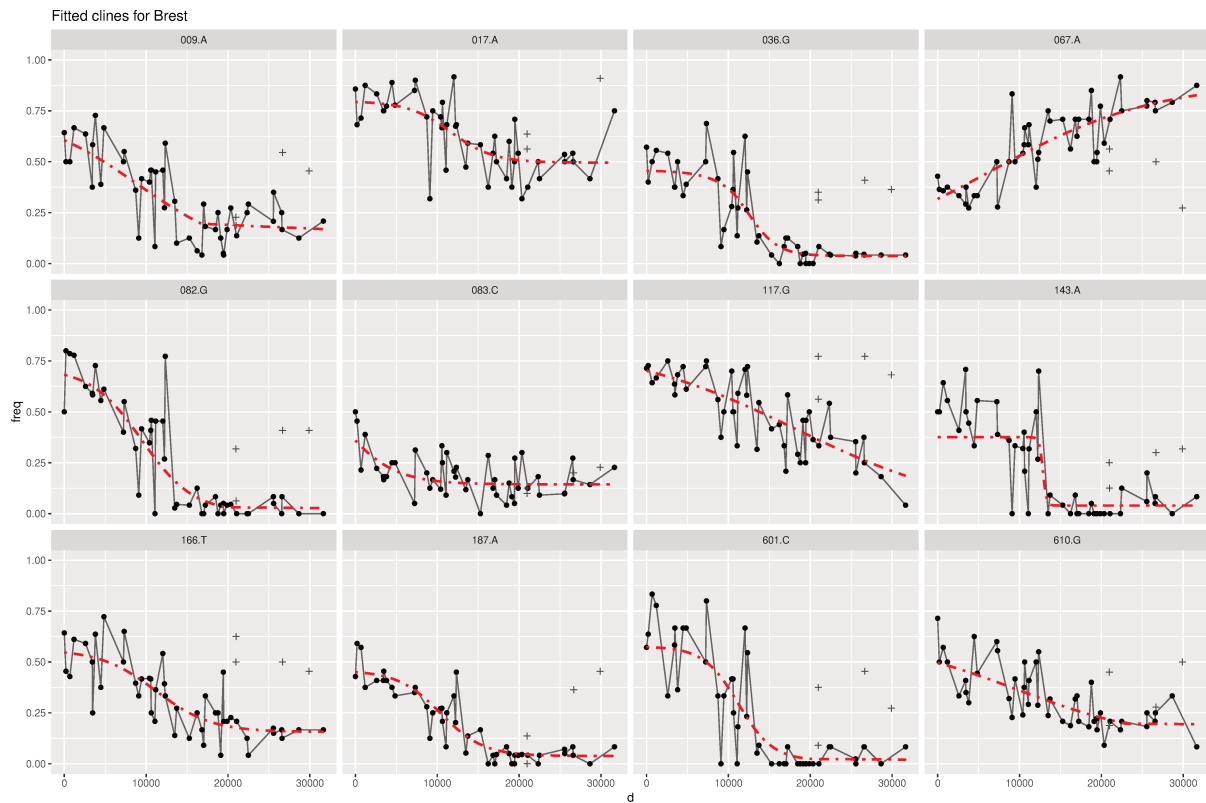


Figure S32: Details of geographic clines per marker fitted in the bay of Brest. Crosses are sites not considered for the fit.

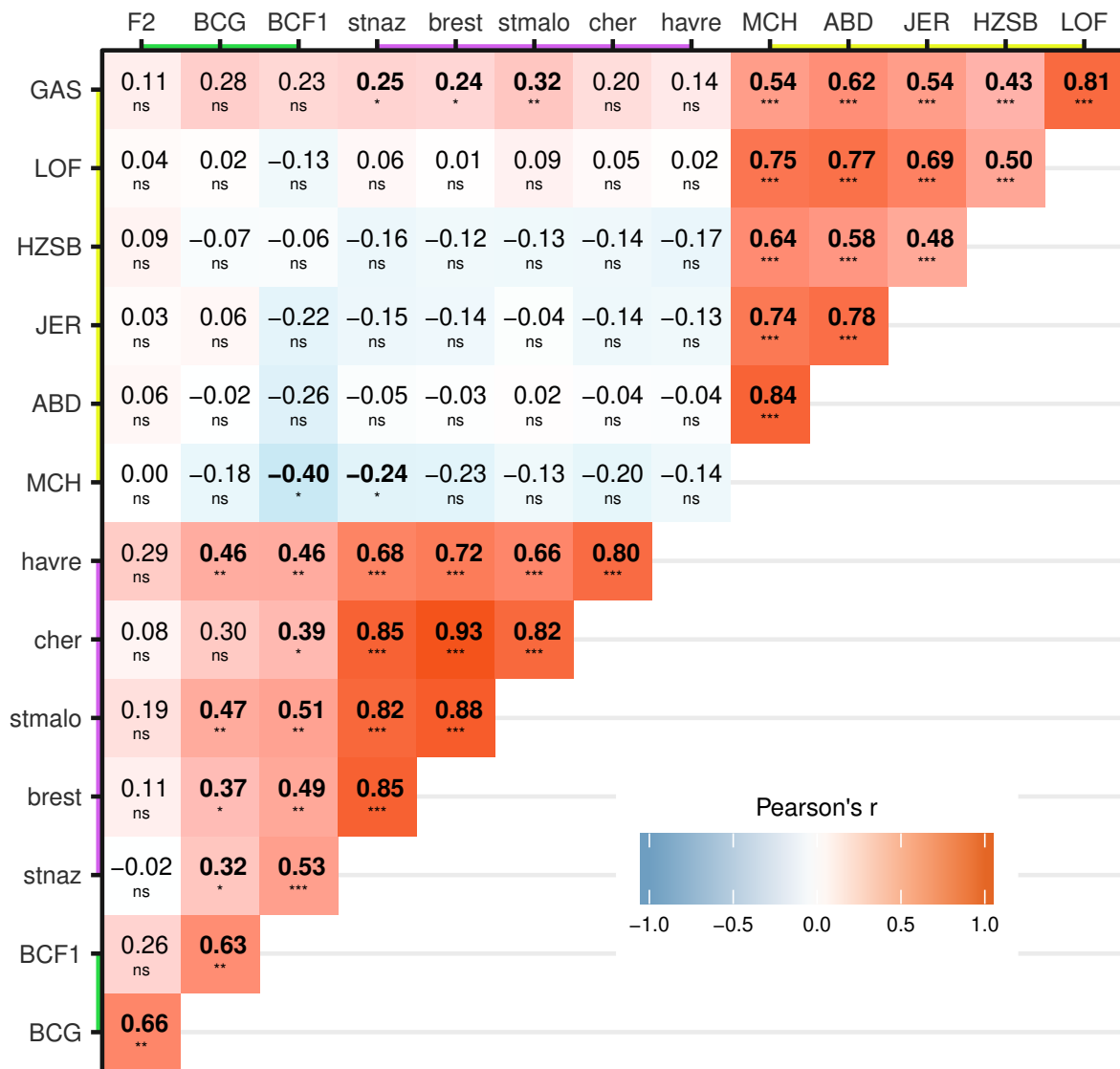


Figure S33: All correlations between populations.

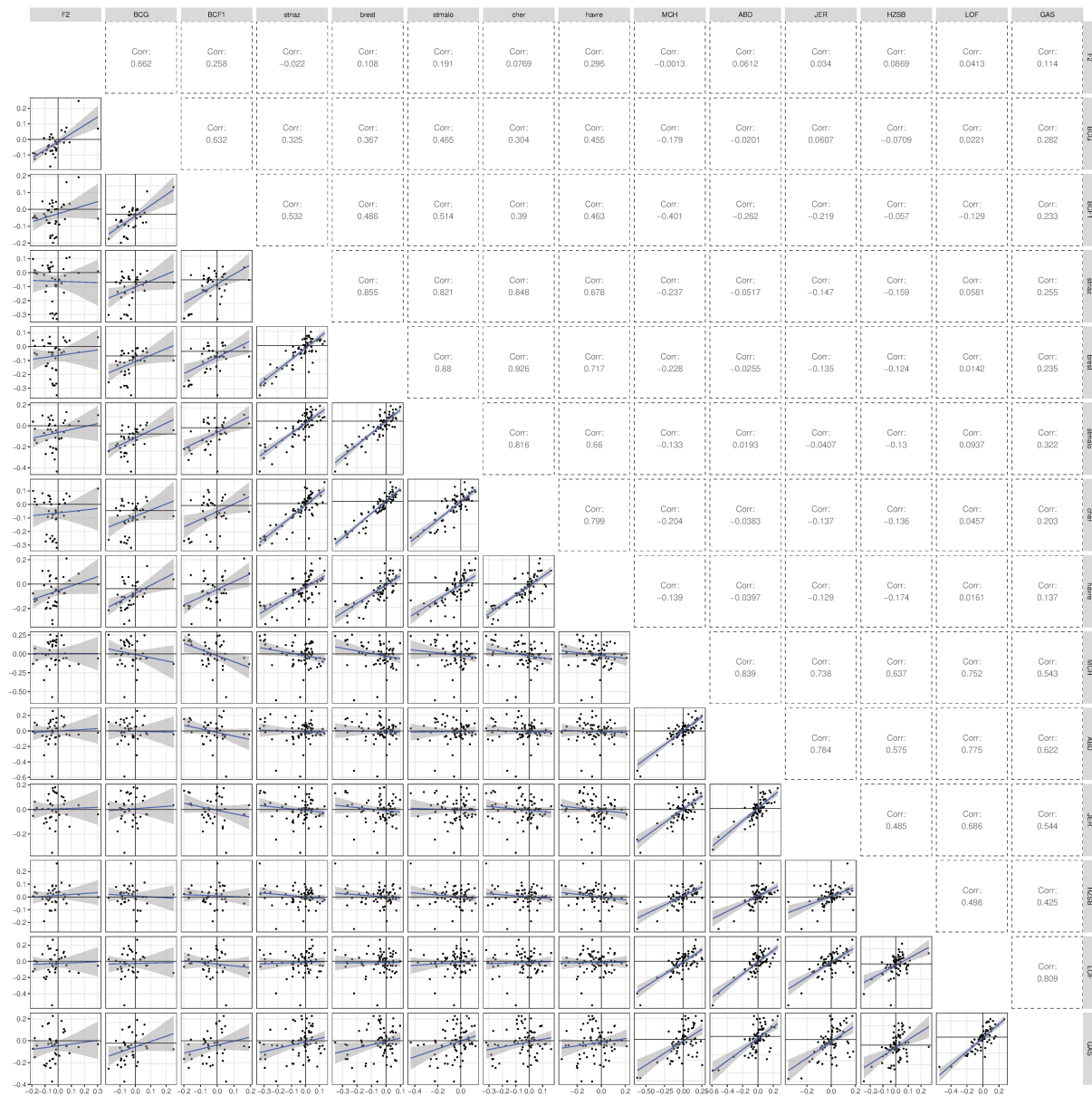


Figure S34: Scatter plots for all correlations between populations.

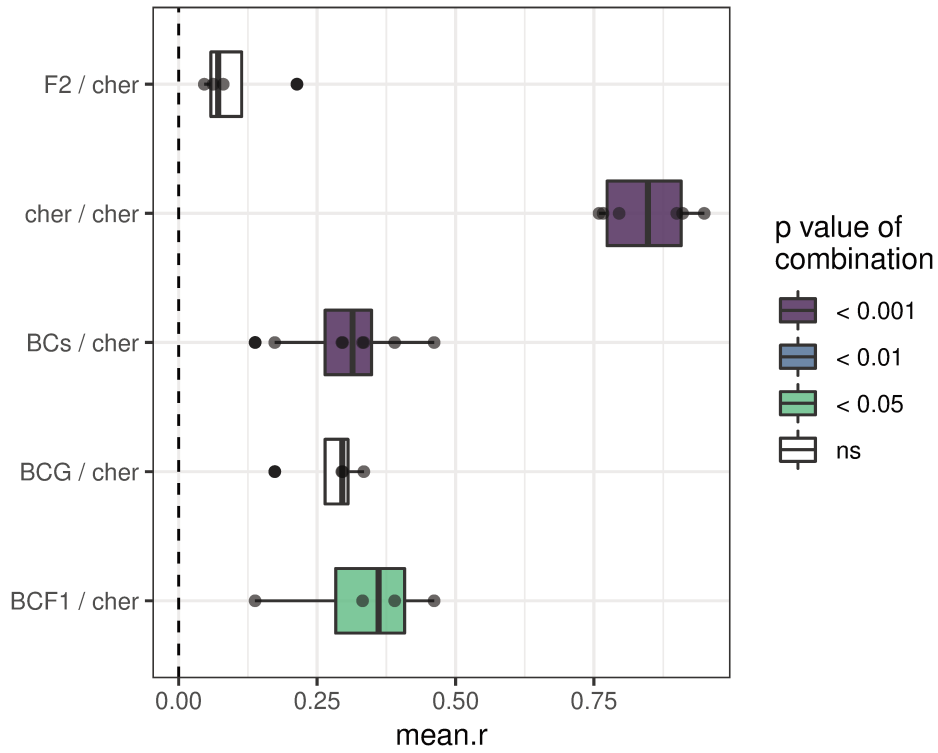


Figure S35: Correlation matrix for temporal variation in Cherbourg.

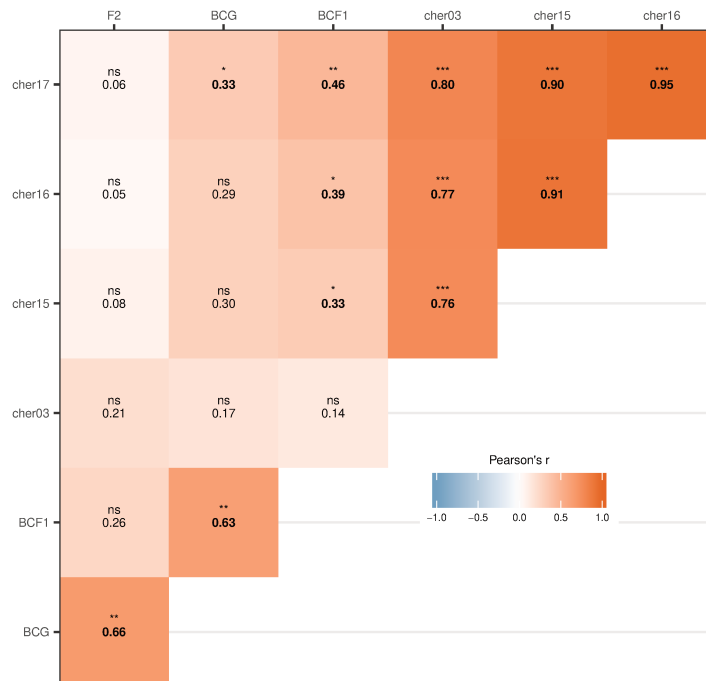


Figure S36: Correlation matrix for temporal variation in Cherbourg.

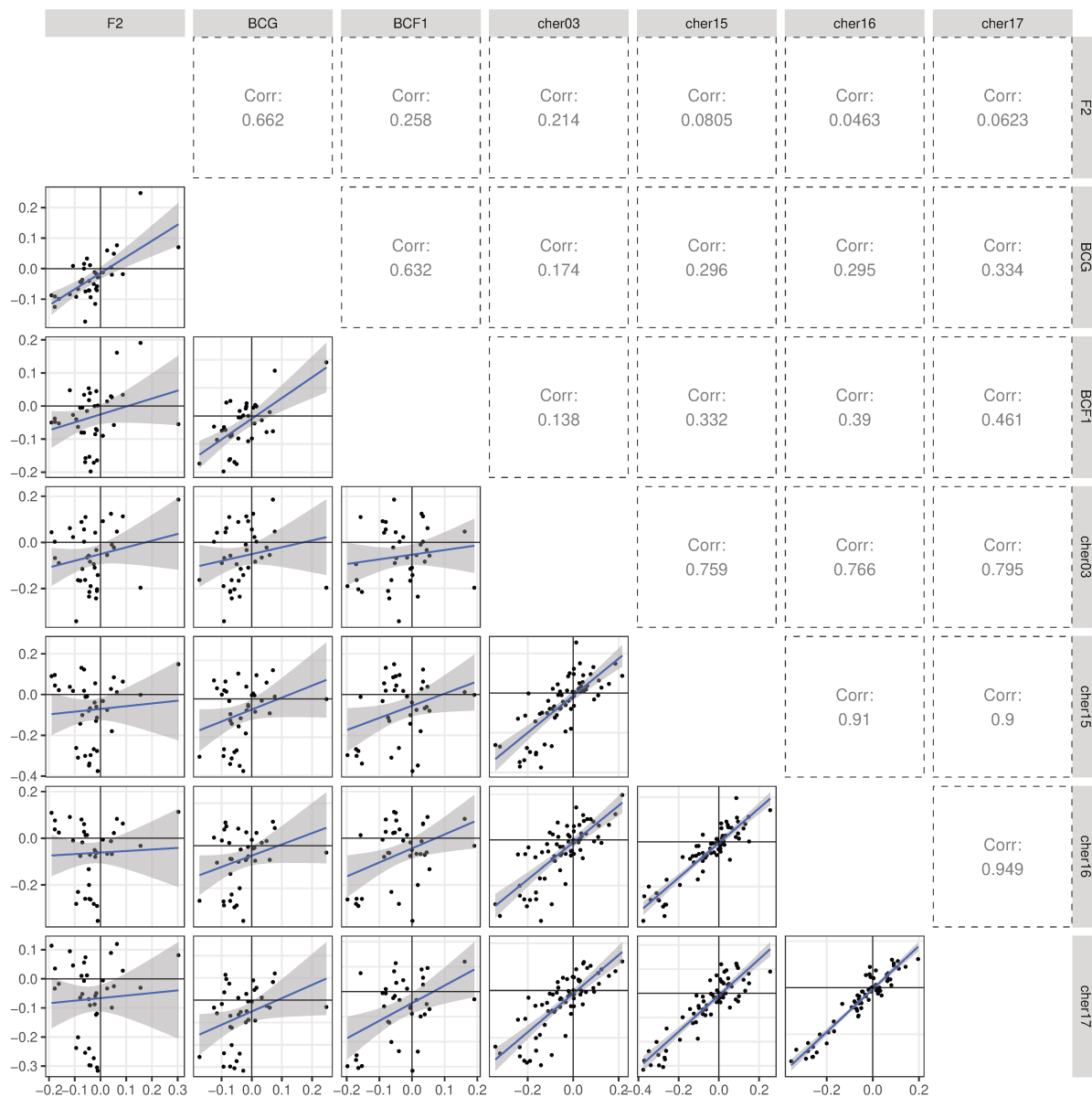


Figure S37: All correlations for Cherbourg.

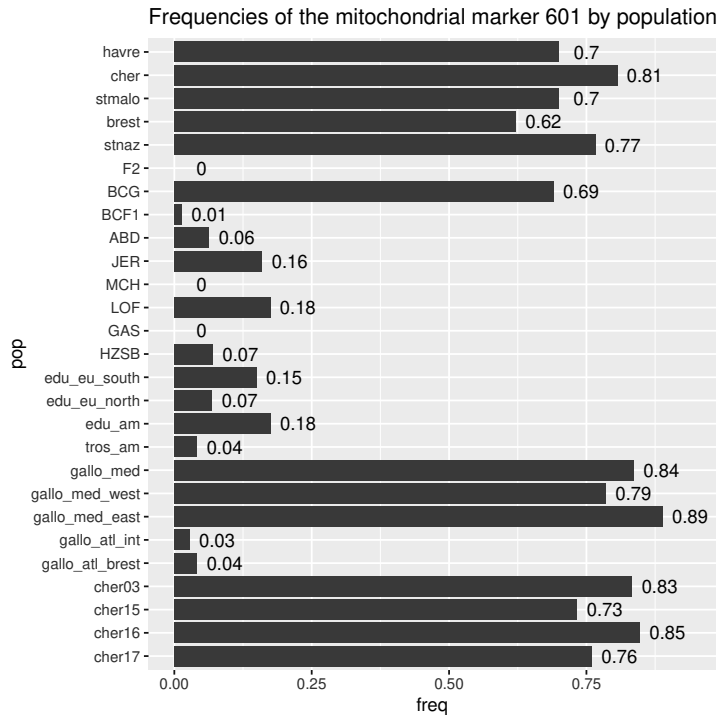


Figure S38: Allele frequencies of the female mitochondrial marker 601.



Figure S39: The mighty mussels: newspaper article about the Swansea King's dock mussels and the work of Prof. David Skibinski. South Wales Evening Post, 28th of June 1978.

C Supplementary Tables

Table S1: Population information with sample size after filtration and average missing data among population.

pop	longitude	latitude	locality	date	N	missing data
Aber	-2.07349	57.14758	ABD, Aberdeen, Scotland, UK		74	0.0254
Ault	1.42314	50.09058	Ault, France		36	0.0087
B_amont	0.14587	49.47124	Bassin Amont, Le Havre, France		15	0.0346
B_aval	0.13242	49.47447	Bassin Aval, Le Havre, France		15	0.0165
Barf	-1.25807	49.71015	Barfleur, France	2015	29	0.0152
BCF1			Backcross		72	0.0253
BCG			Backcross		72	0.0130
BDC	1.42282	50.09060	Bois-de-Cise, France	2016	30	0.0056
Berg	5.30573	60.39560	Bergen, Norway		21	0.0235
BFH	-72.46253	41.26056	Brandford Harbor, USA	2009	12	0.0162
BIL_001	-2.49015	47.44528	Le Bile - Penestin, France	11/01/17	25	0.0192
BIL1	-3.08124	43.35194	Zierbena, Bilbao harbor, Spain	17/07/17	12	0.0097
BIL2	-3.07939	43.35262	Zierbena, Bilbao harbor, Spain	17/07/17	12	0.0141
BIL3	-3.03050	43.33065	Santurtzi, Bilbao harbor, Spain	17/07/17	12	0.0097
BIL4	-3.02976	43.33044	Santurtzi, Bilbao harbor, Spain	17/07/17	12	0.0152
BIL5	-3.02063	43.32557	Portugalete, Bilbao harbor, Spain	17/07/17	12	0.0249
BIL6	-3.01302	43.33165	Getxo, Bilbao harbor, Spain	20/07/17	12	0.0119
BIL7	-3.02753	43.34187	Getxo, Bilbao harbor, Spain	20/07/17	12	0.0097
Bizerte	9.86000	37.26000	Bizerte, Tunisia		12	0.0357
BLYT	-1.49830	55.12580	Blythe, North Sea	14/11/14	5	0.0961
Bodo	14.99857	67.21838	Bodø, Norway	16/11/16	38	0.0308
Bost	-71.05000	42.36000	Boston, USA	2009	14	0.0427
Bouf	-0.72468	49.34802	Le Bouffay, France	2015	30	0.0156
Brest-1	-4.28628	48.31669	Lagonna Daoulas, France	08/07/17	12	0.0238
Brest-10	-4.58798	48.27521	Fishing dock, Camaret, France	09/07/17	12	0.0130
Brest-11	-4.59713	48.27972	Camaret marina, France	09/07/17	12	0.0162
Brest-12	-4.61044	48.26152	Veryac'h, France	09/07/17	12	0.0195
Brest-13	-4.49673	48.22357	Morgat marina, France	09/07/17	12	0.0087
Brest-14	-4.46343	48.23806	Postolonnec beach, France	09/07/17	12	0.0087
Brest-15	-4.36913	48.33790	Porz Tinduff, France	24/07/17	12	0.0173
Brest-16	-4.41397	48.33658	Lauberlac'h, France	24/07/17	12	0.0206
Brest-17	-4.44185	48.34370	Anse du Caro, France	24/07/17	12	0.0184
Brest-18	-4.43723	48.35935	Pointe Marloux, France	24/07/17	12	0.0314

Continued on next page

Table S1 – *Continued from previous page*

pop	longitude	latitude	locality	date	N	missing data
Brest-19	-4.43067	48.39313	Marina Moulin Blanc Nord, France	24/07/17	12	0.0130
Brest-2	-4.45367	48.29334	Lanvéoc, France	08/07/17	12	0.0216
Brest-20	-4.43248	48.39002	Marina Moulin Blanc Sud, France	24/07/17	11	0.0519
Brest-21	-4.47392	48.37994	Brest commerce A, France	25/07/17	12	0.0335
Brest-22	-4.47321	48.38280	Brest commerce B, France	25/07/17	12	0.0216
Brest-23	-4.48110	48.37762	Brest commerce C, France	25/07/17	12	0.0325
Brest-24	-4.45348	48.32235	Pointe de l'Armorique, France	25/07/17	11	0.0413
Brest-25	-4.55102	48.35945	Anse de Sainte-Anne, France	25/07/17	12	0.0152
Brest-26	-4.52923	48.36012	Pointe du Porzic Est, France	25/07/17	12	0.0130
Brest-27	-4.49093	48.37732	Marina du Chateau A, France	25/07/17	12	0.0152
Brest-28	-4.48892	48.38018	Marina du Chateau B, France	25/07/17	11	0.0519
Brest-29	-4.45585	48.37470	Brest commerce D, France	25/07/17	12	0.0206
Brest-3	-4.50891	48.28444	Le Fret, France	08/07/17	12	0.0076
Brest-30	-4.78104	48.35904	Le Conquet, France	27/07/17	12	0.0206
Brest-31	-4.77214	48.34305	Penzer, France	27/07/17	12	0.0087
Brest-32	-4.76820	48.32847	Pointe Saint-Mathieu, France	27/07/17	12	0.0108
Brest-33	-4.69895	48.33888	Plougonvelin, France	27/07/17	12	0.0292
Brest-34	-4.61538	48.33758	Petit Minou, France	27/07/17	12	0.0152
Brest-35	-4.58137	48.34809	Le Mengant, France	27/07/17	12	0.0119
Brest-36	-4.26814	48.27982	Térénez, France	05/08/17	12	0.0184
Brest-37	-4.51477	48.20070	île Vierge, France	06/08/17	12	0.0141
Brest-38	-4.34048	48.40575	Elorn 2, France	21/09/17	7	0.0148
Brest-39	-4.34317	48.40582	Elorn 2B, France	21/09/17	12	0.0357
Brest-4	-4.53997	48.29194	L'île du Renard, France	08/07/17	12	0.0065
Brest-40	-4.34898	48.40385	Elorn St Jean, France	21/09/17	11	0.0153
Brest-41	-4.35612	48.40648	Elorn 3, France	21/09/17	9	0.0216
Brest-42	-4.37208	48.39705	Elorn 6, France	21/09/17	12	0.0249
Brest-43	-4.38280	48.39553	Elorn 9, France	21/09/17	12	0.0216
Brest-44	-4.38205	48.39300	Elorn 11, France	21/09/17	12	0.0238
Brest-45	-4.38670	48.39298	Elorn 14, France	21/09/17	12	0.0173
Brest-46	-4.39465	48.39028	Elorn 15, France	21/09/17	12	0.0184
Brest-5	-4.54507	48.31460	Roscanvel, France	08/07/17	12	0.0249
Brest-6	-4.57300	48.30710	La Fraternité, France	08/07/17	12	0.0173
Brest-7	-4.55452	48.29658	Quelern, France	09/07/17	12	0.0141

Continued on next page

Table S1 – Continued from previous page

pop	longitude	latitude	locality	date	N	missing data
Brest-8	-4.53207	48.33974	Pointe des Espagnols, France	09/07/17	8	0.0568
Brest-9	-4.58254	48.31834	Îlot des Capucins, France	09/07/17	12	0.0141
BretN_g1	-3.71707	48.70014	venizella, Guimaec, France		10	0.0104
BretN_g2	-3.69819	48.69000	poul roudou, Locquirec, France		10	0.0078
BretN_g3	-3.67690	48.68796	lezingar, Locquirec, France		10	0.0104
BretN_g4	-3.64643	48.69481	pors ar villiec, Locquirec		10	0.0052
BretN_i1	-3.54638	48.72863	Léguer estuary, South point, France		9	0.0072
BretN_i2	-3.54230	48.72895	Banc du Guer sud, France		10	0.0117
BretN_i3	-3.58156	48.72139	Locquémo, France		10	0.0078
BretN_i4	-3.58371	48.72608	Locquemo, France		10	0.0065
BretN_i5q1	-3.58283	48.68886	Beg ar form, Saint-Michel en Grève, France		10	0.0091
BretN_i5q2	-3.58271	48.68914	Beg ar form, Saint-Michel en Grève, France		10	0.0039
BretN_i6	-3.61713	48.68507	Beg Douar, Plestin Les Grèves, France		10	0.0078
BretN_i8	-3.80425	48.72304	Roc'h Goalen, Plougasnou, France		10	0.0312
CAD	-6.29842	36.52719	Cadix, Spain	2016	15	0.0294
Can	3.69184	43.41226	Sète, canal, France		3	0.0909
Cart	-1.80842	49.37165	Carteret, France	2015	29	0.0331
CER_001	-4.44887	48.38265	Brest port, France	14/12/16	25	0.0135
Cha	-1.08059	46.03229	Charente, France		12	0.0119
CHE	-2.29965	47.23232	Pointe de Chemoulin, France	2016	30	0.0242
CHE_001	-1.62336	49.64574	Cherbourg, France	02/12/16	25	0.0156
Cher-A	-1.61953	49.64530	Cherbourg A, France	01/08/17	12	0.0227
Cher-B	-1.61981	49.64956	Cherbourg B, France	01/08/17	12	0.0249
Cher-C	-1.63244	49.67307	Cherbourg C, France	01/08/17	12	0.0390
Cher-D	-1.62020	49.64818	Cherbourg D, France	01/08/17	12	0.0487
Cher-E	-1.62080	49.64208	Cherbourg E, France	01/08/17	12	0.0433
Cher-F	-1.62342	49.64566	Cherbourg F, France	01/08/17	11	0.0342
Cher-G	-1.61964	49.63846	Cherbourg G, France	01/08/17	11	0.0614
Cher03	-1.61994	49.63942	Cherbourg, France		30	0.0108
Cher15	-1.62332	49.64557	Cherbourg, France	2015	30	0.0212
Cher16	-1.62332	49.64557	Cherbourg, France	2016	99	0.0365
CHR	-71.51917	41.38692	Charlestown, Snug Harbor Marina, USA	2009	12	0.0206
COR	-8.41393	43.37011	Coruña, Galicia, Spain	2016	15	0.0069
CRE	-8.97390	38.48047	Praia do Creio, Arrábida, Spain	2016	15	0.0078

Continued on next page

Table S1 – *Continued from previous page*

pop	longitude	latitude	locality	date	N	missing data
Croa	13.62091	45.07155	Croatia		14	0.0547
D15A	2.93460	54.32470	D15-A, North Sea	03/10/15	8	0.1445
Dieppe	1.09670	49.93556	Dieppe, France		48	0.0095
Din	-2.04937	48.63913	Dinard, France		13	0.0649
DKL	-15.93000	23.72000	Dahkla, Morocco		10	0.1013
Dun	2.42178	51.10313	Dunkerque, France	2016	30	0.0030
Ena11	-2.07507	48.64035	Saint-Enogat, Dinard, France		7	0.0816
Engl	-0.93377	49.39478	Englesqueville, France	2015	30	0.0177
EST_001	-2.53370	47.30370	Le Croisic, France		25	0.0088
F1			F1		6	0.0195
F2			F2		132	0.0364
Far	-7.93884	37.01568	Faro, Portugal		8	0.0633
FAR	-7.93884	37.01568	Faro, Portugal	2016	15	0.0190
Fer11	-3.96309	48.72045	ballast ferry	2011	38	0.0123
Fer13	-3.96309	48.72045	ballast ferry Armorique	16/05/13	27	0.0077
FINO	7.15830	55.19500	FINO 3, North Sea	23/09/15	6	0.2186
GA	6.24164	62.46399	GAS, Gaseid, Alesund, Norway	2015	36	0.0400
Gdansk	18.52794	54.60956	Gulf of Gdansk, Poland	23/08/12	12	0.1223
Gran	-1.59947	48.84003	Granville, France	2015	31	0.0201
GrCa	-1.03633	49.40803	Grancamp, France	2015	29	0.0233
Groix	-3.45911	47.64701	Groix, France		48	0.0249
HARW	1.28130	51.93480	Harwich, North Sea	15/11/15	11	0.1181
Havre_A	0.11327	49.48586	Le Havre, France	2016	15	0.0069
Havre_B	0.11612	49.47316	Le Havre, France	2016	15	0.0303
Havre_C	0.12268	49.47000	Le Havre, France	2016	15	0.0130
Havre_D	0.13867	49.46841	Le Havre, France	2016	15	0.0139
Havre_E	0.15604	49.46831	Le Havre, France	2016	15	0.0104
Havre_F	0.15841	49.47585	Le Havre, France	2016	15	0.0095
Havre_G	0.17341	49.47242	Le Havre, France	2016	15	0.0087
Havre_H	0.09767	49.48713	Le Havre, France	2016	15	0.0052
Havre_I	0.16510	49.45401	Le Havre, France	2016	15	0.0173
Havre-J	0.11832	49.48893	Le Havre, France	02/07/17	15	0.0502
Her	25.14420	35.33800	Heraklion, Greece		35	0.0171
Holl	5.42400	53.31000	Wadden Sea, The Netherlands		23	0.0056

Continued on next page

Table S1 – Continued from previous page

pop	longitude	latitude	locality	date	N	missing data
HORN	7.81100	55.47890	Horns Rev, North Sea	10/06/15	2	0.1299
HOU_001	-2.93723	47.41967	Ile de Houat, France	10/01/17	25	0.0192
JapSea	135.09238	38.73579	Katolikova, Japan Sea		16	0.1631
Jer	-2.02038	49.17308	JER, Jersey, France	2016	30	0.0143
JUM	-3.90044	47.83397	La Jument, France	2016	30	0.0104
K10B	3.25390	53.36260	K10-B, North Sea	01/10/14	8	0.1591
KOR	128.44707	34.83393	South Korea		10	0.1169
L4	-4.57230	48.28220	bay of Brest (Camaret), France	24/04/17	42	0.0247
L5	-4.48815	48.37733	bay of Brest (port), France	24/04/17	49	0.0358
LaRochA	-1.22359	46.14988	La Rochelle, France	24/06/17	7	0.1410
LaRochB	-1.22449	46.15045	La Rochelle, France	25/06/17	29	0.0636
LeHaCM	0.11380	49.45725	Le Havre, France	2017	14	0.0241
LeHaP1	0.10579	49.46466	Le Havre P1, France	2017	15	0.0277
LeHaP10	0.17332	49.45352	Le Havre P10, France	2017	14	0.0473
LeHaP11	0.17807	49.45456	Le Havre P11, France	2017	15	0.0857
LeHaP2	0.10996	49.46040	Le Havre P2, France	2017	15	0.0199
LeHaP3	0.11473	49.45825	Le Havre P3, France	2017	15	0.0156
LeHaP4	0.12137	49.45734	Le Havre P4, France	2017	15	0.0216
LeHaP5	0.13330	49.45634	Le Havre P5, France	2017	15	0.0727
LeHaP6	0.14413	49.45546	Le Havre P6, France	2017	15	0.0355
LeHaP7	0.15224	49.45185	Le Havre P7, France	2017	15	0.0268
LeHaP8	0.16015	49.45441	Le Havre P8, France	2017	15	0.0268
LeHaP9	0.16602	49.45376	Le Havre P9, France	2017	15	0.0502
LISB	-9.09260	38.76350	Lisbon, Portugal	14/02/15	5	0.1818
LochEti	-5.18317	56.45735	Loch Etive, Scotland, UK		11	0.1145
Locq	-3.59720	48.73691	Locquemeau, France		37	0.0190
LOF	13.87800	68.33800	Lofoten, Norway	2014	43	0.0417
MAB_001	-2.21610	46.99760	Maison Blanche, Noirmoutier, France		25	0.0088
MAK	22.61232	40.41618	Makrigiallos, Greece		6	0.1017
MOU	-4.04187	47.84282	Pointe de Mousterlin, France	2016	30	0.0242
Murch	1.75000	61.40000	MCH, Murchison oil station, Norwegian Sea		12	0.0465
NUS	-51.71040	64.19680	Nuuk, Greenland	2014	28	0.0756
NY-harb-A	-73.79327	40.79521	NYC Throgs Neck Bridge, USA	09/06/04	9	0.0534
NY-harb-B	-71.94682	41.02913	Montauk, USA	23/02/14	7	0.0686

Continued on next page

Table S1 – *Continued from previous page*

pop	longitude	latitude	locality	date	N	missing data
NY-harb-C	-74.01726	40.65532	Brooklyn piers, USA		10	0.0260
ORT	-7.80508	43.72246	Ortigueira, Galicia, Spain	2016	15	0.0139
OST	-72.38419	41.26328	Old Saybrook town, Hartland drive, USA	2009	24	0.0211
Ostende	2.91815	51.23815	Ostende, Belgium		29	0.0090
Ouis	-0.45892	49.33829	Quistreham, France	2015	30	0.0203
Palice-A	-1.21576	46.15909	La Palice, Bassin à flot, France	23/02/18	18	0.0065
Palice-B	-1.21843	46.15854	La Palice, lock, France	23/02/18	76	0.0106
PEN	-2.51383	47.30632	Barres de Pen Bron, France	2016	30	0.0212
PEN_001	-2.49363	47.49712	Camaret plage-Penestin, France	11/01/17	25	0.0223
Peniche	-9.37944	39.36844	Peniche, Portugal	2016	15	0.0121
PET_001	-4.47146	48.38119	Brest port, France	14/12/16	25	0.0151
Pir	-1.60083	49.16588	Pirou, France	2016	30	0.0074
POU_001	-2.41596	47.26029	Le Pouliguen - La Baule, France	12/01/17	25	0.0119
Prim	-3.82080	48.72122	Primel, France		36	0.0177
PtArm97	-4.45355	48.32563	Pointe de l'Armorique, plage des ducs d'albes, France		19	0.0444
Q13A	4.13610	52.19110	Q13-A, North Sea	28/05/14	12	0.0617
QAS	-69.24030	77.46500	Qaanaaq, Greenland	2014	30	0.1364
RadeBrest_R1	-4.39973	48.39014	Pont d'Iroise, France		10	0.0195
RadeBrest_R2	-4.36882	48.33810	Port du Tinduff, France		10	0.0169
RadeBrest_R3	-4.50680	48.28679	Le Fret, France		10	0.0091
RadeBrest_R4	-4.55083	48.30609	Quélern, Roscanvel, France		10	0.0104
RadeBrest_R5	-4.57194	48.28199	Pte Ste Barbe, Camaret, France		10	0.0104
Rave	-1.21023	49.47333	Ravenoville, France	2015	30	0.0130
Revi	-1.22933	49.57467	Réville, France	2015	30	0.0160
ROC_VER	-1.40556	45.98358	Rocher Vert Chaucre cote ouest Oléron, France	05/10/16	25	0.0151
RoRo	-3.38314	48.81866	Roc-Rouge, France		32	0.0114
Roth	-1.96213	48.68703	Rothéneuf, France		32	0.0288
SB1	-2.71614	48.55568	Pointe du Roselier, France		10	0.0039
SB2	-2.78885	48.58458	Pordic, plage du petit havre, France		10	0.0052
SCHV	4.25820	52.09870	Scheveningen, North Sea	08/07/14	1	0.0779
SoRo	-1.55248	48.73239	Sol-Roc, France		30	0.0472
StAnd	0.07980	49.54744	Saint-Andrieux, France		15	0.0087
StJo	0.15264	49.64395	Saint-Jouin, France	2015	27	0.0106
StLau-CBD	-69.46615	48.26941	Saint Lawrence, Cap de Bon Désir, Canada		9	0.1010

Continued on next page

Table S1 – *Continued from previous page*

pop	longitude	latitude	locality	date	N	missing data
StLau-TD	-69.69614	48.13441	Saint Lawrence, Tadoussac, Canada		12	0.1180
StMalo	-2.02165	48.64855	Saint-Malo port, France	16/07/17	30	0.0156
StNaz-I	-2.19894	47.27074	Saint-Nazaire, France	30/01/18	26	0.0080
StNaz-II	-2.20027	47.27391	Saint-Nazaire, France	30/01/18	25	0.0348
StNaz-III	-2.20259	47.27312	Saint-Nazaire, France	30/01/18	25	0.0358
StNaz-IV	-2.19813	47.28964	Saint-Nazaire, France	30/01/18	26	0.0330
StNaz-V	-2.19880	47.27941	Saint-Nazaire, France	30/01/18	26	0.0280
SV1	11.13620	79.11230	Svalbard – Kongsfjorden, Norway	2012	6	0.0455
SV2	11.13620	79.11230	Svalbard – Kongsfjorden, Norway	2013	9	0.0505
SV3	11.13620	79.11230	Svalbard – Kongsfjorden, Norway	2014	12	0.0649
SV4	15.60260	78.23810	Svalbard – Adventfjorden, Norway	2014	11	0.0106
Tatihou	-1.23627	49.58673	Tatihou, France	09/07/05	24	0.0076
Th16	3.69179	43.41227	Thau, France	2016	23	0.0169
TH16B	3.68782	43.41474	Thau, ponton station, France	2016	18	0.0043
VHO	-8.82312	37.38853	Vale dos Homens, Aljezur, Spain	2016	15	0.0078
VIG	-8.70214	42.26008	Vigo, Spain	2016	15	0.0104
Vill	0.11605	49.39935	Villerville, France	2015	30	0.0251
VILL	0.12383	49.40374	Villerville, France	2016	30	0.0108
Vill16	0.12351	49.40369	Villerville, France	2016	15	0.0035
VLR	0.79465	49.87812	Veules-les-Roses, France	2016	30	0.0052
Wim	1.60307	50.77263	Wimereux, France	2016	29	0.0049

Table S2: Softwares and R packages used in this study

software	version	reference	link
R	3.5.2	R Core Team, 2019	https://www.r-project.org/
Python	3.7.1	-	https://www.python.org/
Structure	2.3.4	Falush et al., 2003; Pritchard, Stephens, and Donnelly, 2000	https://web.stanford.edu/group/pritchardlab/structure.html
Admixture	1.3.0	Alexander and Lange, 2011	http://software.genetics.ucla.edu/admixture/
CLUMPAK	1.1	Kopelman, Mayzel, Jakobsson, Rosenberg, and Mayrose, 2015	http://clumpak.tau.ac.il/
QGIS	3.4.2	-	https://qgis.org/fr/site/
Newhybrids	1.1 beta 3	Anderson and Thompson, 2002	http://ib.berkeley.edu/labs/slatkin/eriq/software/software.htm

R package	version	reference
tidyverse	1.2.1	Wickham, 2017
reshape2	1.4.3	Wickham, 2007
adegenet	2.1.1	T. Jombart and Ahmed, 2011; Thibaut Jombart, 2008
pegas	0.11	Paradis, 2010
hierfstat	0.04-22	Goudet and Jombart, 2015
introgress	1.2.3	Gompert and Buerkle, 2010
qtl	1.44-9	Broman, Wu, Sen, and Churchill, 2003
LDheatmap	0.99-5	Shin, Blay, McNeney, and Graham, 2006
genetics	1.3.8.1	Warnes, Gorjanc, Leisch, and Man, 2013
hzar	0.2-5	Derryberry, Derryberry, Maley, and Brumfield, 2014
ggpubr	0.2	Kassambara, 2018
psych	1.8.12	Revelle, 2018
EmpiricalBrownsMethod	1.10.0	Poole, Gibbs, Shmulevich, Bernard, and Knijnenburg, 2016
ggstatsplots	0.0.9	Patil, 2019
foreach	1.4.4	Microsoft and Weston, 2017
doMC	1.3.5	Analytics and Weston, 2017
gdistance	1.2-2	van Etten, 2017

Table S3: Across all analyses multiple selection threshold have been used to select either individuals or markers for specific analyses. Decision thresholds are presented in table S3. For statistical tests we chose $\alpha = 0.05$ as the significant threshold.

Variable used	for what?	with what?	kept if
Missing data of individuals	full dataset	genotypes	< 30%
	genetic map F2	genotypes	< 10%
Missing data of markers	full dataset	genotypes	< 10%
	F_{ST} between reference populations	genotypes of <i>M. trossulus</i> pop. only	< 30%
Q value or prob. of membership (C)	reference groups filtration	Admixture $K = 3$ (fig. S4)	$Q > 85\%$ from putative cluster
	<i>M. trossulus</i> individuals	full Structure $K = 5$ (fig. S5)	$Q_{tros} > 0.1$
	list of dock mussels	Structure local w/o admixture $K = 3$ (fig. S20)	$\in (\text{Port}) \wedge \in (C_{dock} > 0.9)$
	list of <i>M. edulis</i> in ports	Structure local w/o admixture $K = 3$	$\in (\text{Port}) \wedge \in (C_{edu} > 0.5)$
	list of gallo_atl in ports	Structure local w/o admixture $K = 3$	$\in (\text{Port}) \wedge \in (C_{gallo_atl} > 0.5)$
	list of admix gallo_atl	Structure w/o admixture $K = 5$ (fig. S19)	$\in (\text{focal pop.}) \wedge \in (C_{gallo_atl} + C_{admix_atl} > 0.9)$

Table S4: Reference panel groups and populations

L1	L2	L3	pop	N
edu	edu_am	edu_am	Bost	14
			CHR	12
			OST	24
	edu_eu	edu_eu_south	Dun (ext)	29
			Holl (ext)	23
			Ostende (ext)	28
			Wim (ext)	29
			Cha (int)	12
	edu_eu_north	Bodo	38	
		Berg	38	
gallo	gallo_atl	gallo_atl_iber	CAD	15
			COR	15
			CRE	15
			DKL	10
			Far	8
			FAR	14
			ORT	15
			Peniche	15
			VHO	15
			VIG	15
	gallo_atl_brit	Prim	36	
	gallo_med	gallo_med_east	Croa	14
			Her	35
			MAK	6
	gallo_med_west	Can	3	
		Th16	23	
TH16B		18		
Bizerte		12		
tros	tros_am	tros_am	LochEti	11
			StLau-CBD	9
			StLau-TD	11
	tros_eu	tros_eu_baltic	Gdansk	12
	tros_pac	tros_pac_west	JapSea	16

Table S5: Linkage map – (i) genetic map produced from lab F2 crosses, (ii) markers on the same contig and (iii) left markers with unknown relation.

	locus	linkage group	position (cM)	unlinked set
	063	1	0.00	•
	067	1	22.88	
	607	1	24.23	
	062	1	25.11	•
	210	1	28.33	
	055	1	28.77	
	034	1	31.99	
	115	1	32.44	
	064	2	0.00	
	026	2	1.36	
	047	2	1.36	
	160	2	1.36	•
	161	2	3.65	
	052	3	0.00	•
	007	3	9.70	
	610	3	27.41	
	015	4	0.00	
(i)	155	4	2.29	•
	900	4	14.32	
	142	5	0.00	
	148	5	3.22	
	184	5	3.22	•
	061	6	0.00	
	080	6	10.00	•
	082	7	0.00	•
	802	7	10.00	
	144	8	0.00	
	145	8	10.00	•
	001	9	0.00	•
	022	10	0.00	•
	073	11	0.00	•
	094	12	0.00	•
	164	13	0.00	•
	180	14	0.00	•
	211	15	0.00	•
	617	16	0.00	•
	202	R_L02_mira_Contig1	2420 (bp)	•
(ii)	602	R_L02_mira_Contig1	3684 (bp)	
	604	R_L02_mira_Contig1	6232 (bp)	
(iii)	All other markers not in the linkage map			•

General Note

<i>p</i> value	symbol
$p < 0.001$	***
$0.001 < p < 0.01$	**
$0.01 < p < 0.05$	*
$p > 0.05$	ns

When statistically significant, results are presented in bold.

Table S6: Fst values between reference groups.

Level	group 1	group 2	F_{ST}
L1	edu	gallo	0.8127***
	edu	tros	0.7226***
	gallo	tros	0.8086***
L2	gallo_med	gallo_atl	0.3823***
	edu_eu	edu_am	0.4845***
L3	gallo_med_west	gallo_med_east	0.0586***
	edu_eu_south_int	edu_eu_south_ext	0.0024 (ns)
	edu_eu_south_int	edu_eu_north	0.2086***
	edu_eu_south_ext	edu_eu_north	0.3129***
	gallo_atl_ext	gallo_atl_int	0.1555***

Table S7: Fst values and significance between ports.

	cher	havre	brest	stmalo
havre	0.0235***			
brest	0.0046***	0.0245***		
stmalo	0.0078 (ns)	0.0390***	0.0058 (ns)	
stnaz	0.0071 (ns)	0.0304***	0.0067**	0.0043 (ns)

Table S8: Fst values and significance between Cherbourg years.

	cher03	cher15	cher16
cher15	0.0066**		
cher16	0.0097*	0.0009 (ns)	
cher17	0.0032 (ns)	-0.0017 (ns)	-0.0007 (ns)

Table S9: Fst values and significance between naturally and Norway admixed.

	ABD	HZSB	GAS	JER	LOF
HZSB	0.0429***				
GAS	0.0736*	0.1021*			
JER	0.0974***	0.1439***	0.0335 (ns)		
LOF	0.0372***	0.0743***	0.0047 (ns)	0.0414 (ns)	
MCH	0.0346 (ns)	0.0462 (ns)	0.1795 (ns)	0.2330*	0.1163 (ns)

Table S10: Statistical tests of the differences in ancestry estimates between populations of implicating Atl. *M. galloprovincialis* and *M. edulis* (i.e. Norway admixed and naturally admixed). The Dwass-Steel-Crichtlow-Fligner test is used to compute post-hoc pairwise comparisons, the statistic (*W*) and corrected *p*-values for multiple tests are reported for each type of ancestry.

group1	group2	edu_eu_south		gallo_atl		gallo_med		edu_am	
		W	p.value	W	p.value	W	p.value	W	p.value
JER	ABD	-8.0014855	0.0000003	8.093009	0.0000003	-1.9301499	1.0000000	0.4648148	1.0000000
JER	MCH	-6.3869294	0.0000784	6.386518	0.0000392	-3.3998864	0.8073574	-1.2188032	1.0000000
JER	HZSB	-8.2758887	0.0000002	8.291331	0.0000002	-2.6247521	1.0000000	1.2462346	1.0000000
ABD	MCH	-5.7267576	0.0004261	5.920113	0.0001286	-2.5251363	1.0000000	-1.6665649	1.0000000
ABD	HZSB	-4.2334401	0.0152819	4.159408	0.0135876	-0.5622536	1.0000000	1.5813721	1.0000000
MCH	HZSB	1.5074295	1.0000000	-1.750562	0.7669834	2.3974014	1.0000000	2.9963772	0.1888831
GAS	LOF	-1.3324952	1.0000000	1.675714	0.7832573	-0.4038931	1.0000000	1.4839329	1.0000000
GAS	JER	5.8227527	0.0003818	-3.283418	0.0776347	0.9194070	1.0000000	-5.9576512	0.0001795
GAS	ABD	-1.2763083	1.0000000	6.491894	0.0000315	-0.3242880	1.0000000	-6.6249430	0.0000233
GAS	MCH	-4.5553253	0.0090960	6.585699	0.0000296	-2.0567182	1.0000000	-4.8952169	0.0033482
GAS	HZSB	-3.0020824	0.1682154	6.555299	0.0000296	-0.6873292	1.0000000	-7.1242722	0.0000059
LOF	JER	7.9519649	0.0000003	-5.949607	0.0001286	1.9547265	1.0000000	-8.4598018	0.0000000
LOF	ABD	0.7416394	1.0000000	6.047916	0.0001051	0.4258784	1.0000000	-10.4083896	0.0000000
LOF	MCH	-4.4370894	0.0106137	7.028091	0.0000090	-2.0570260	1.0000000	-6.6559353	0.0000233
LOF	HZSB	-2.5689731	0.3136611	7.008418	0.0000090	-0.0492258	1.0000000	-12.0687157	0.0000000

Table S11: Comparisons of ancestry estimates in dock mussel populations. See Table S10 for details.

group1	group2	edu_eu_south		gallo_atl		gallo_med		edu_am	
		W	p.value	W	p.value	W	p.value	W	p.value
bre	havre	10.7655111	0.0000000	-15.6453266	0.0000000	-3.341636	0.0714903	5.979967	0.0006899
bre	cher	0.5032010	1.0000000	-10.9995502	0.0000000	6.803901	0.0000110	2.996186	0.1429073
bre	stnaz	0.9072883	1.0000000	-5.6694271	0.0003583	3.482114	0.0674667	-1.141112	1.0000000
bre	stmalo	-2.9745824	0.1482498	-3.4396767	0.0627953	5.240644	0.0012351	-2.577176	0.2506348
havre	cher	-13.4998443	0.0000000	6.6257054	0.0000222	13.215188	0.0000000	-3.266336	0.1084834
havre	stnaz	-7.7970358	0.0000003	6.6026763	0.0000222	7.137681	0.0000044	-5.056126	0.0038002
havre	stmalo	-8.2353098	0.0000001	4.5053667	0.0070604	7.675113	0.0000008	-5.016890	0.0038002
cher	stnaz	0.8986060	1.0000000	2.4963584	0.2841525	-1.389627	0.9541097	-3.234122	0.1084834
cher	stmalo	-3.6887206	0.0511747	1.8065668	0.6554651	3.202155	0.0767377	-3.980020	0.0358256
stnaz	stmalo	-3.6202551	0.0511747	0.3570753	1.0000000	3.302863	0.0714903	-1.562515	0.8759137

Table S12: Comparisons of ancestry estimates in Cherbourg temporal sampling. See Table S10 for details.

group1	group2	edu_eu_south		gallo_atl		gallo_med		edu_am	
		W	p.value	W	p.value	W	p.value	W	p.value
cher17	cher16	-2.6860867	0.4230421	2.3094012	0.5026203	1.8889989	0.6677609	1.5834060	1
cher17	cher03	1.6852922	0.8582597	1.6552149	0.8892385	-2.6432869	0.3065542	0.3310927	1
cher17	cher15	0.4714781	1.0000000	-1.4395540	0.9074246	0.4614467	1.0000000	0.2006680	1
cher16	cher03	3.7937987	0.1074177	0.4877319	1.0000000	-4.2160701	0.0422489	-0.8576754	1
cher16	cher15	2.4189968	0.4273574	-3.1784465	0.3622642	-0.8510679	1.0000000	-0.7791183	1
cher03	cher15	-1.1499584	1.0000000	-2.4254679	0.5026203	2.6344868	0.3065542	-0.0209115	1

Table S13: Comparisons of ancestry estimates in Atl. *M. galloprovincialis* from Brittany. See Table S10 for details.

group1	group2	edu_eu_south		gallo_atl		gallo_med		edu_am	
		W	p.value	W	p.value	W	p.value	W	p.value
gallo_atl_brest_bay	gallo_atl_brest	-4.540429	0.0024335	4.074363	0.0072794	2.8645769	0.2354177	-1.909974	0.3243082
gallo_atl_brest_bay	gallo_atl_int	-12.819450	0.0000000	12.408811	0.0000000	-0.3787987	1.0000000	-8.974752	0.0000000
gallo_atl_brest	gallo_atl_int	-10.709842	0.0000000	10.484107	0.0000000	-2.4195419	0.2394843	-7.325735	0.0000006

Table S14: Estimation of the number of generations since admixture in ports.

port	average r	sd(r)	generations
brest	0.05873	0.00007	5.87
cher	0.07082	0.00006	7.08
cher03	0.12713	0.00061	12.71
cher15	0.07670	0.00031	7.67
cher16	0.06290	0.00008	6.29
cher17	0.06472	0.00013	6.47
havre	0.13902	0.00018	13.90
stmalo	0.08277	0.00066	8.28
stnaz	0.03472	0.00005	3.47

References

- Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, *12*(1), 246. doi:10.1186/1471-2105-12-246
- Analytics, R., & Weston, S. (2017). doMC: Foreach Parallel Adaptor for 'parallel' (Version R package version 1.3.5).
- Anderson, E. C., & Thompson, E. A. (2002). A Model-Based Method for Identifying Species Hybrids Using Multilocus Genetic Data. *Genetics*, *160*(3), 1217–1229.
- Broman, K. W., Wu, H., Sen, S., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, *19*, 889–890.
- Coolen, J. W. P. (2017). *North Sea reefs: Benthic biodiversity of artificial and rocky reefs in the southern North Sea* (PhD thesis, Wageningen University, Wageningen).
- Derryberry, E. P., Derryberry, G. E., Maley, J. M., & Brumfield, R. T. (2014). Hzar: Hybrid zone analysis using an R software package. *Molecular Ecology Resources*, *14*(3), 652–663. doi:10.1111/1755-0998.12209
- Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics*, *164*(4), 1567–1587.
- Gompert, Z., & Buerkle, C. A. (2010). Introgress: A software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, *10*(2), 378–384. doi:10.1111/j.1755-0998.2009.02733.x
- Goudet, J., & Jombart, T. [Thibaut]. (2015). Hierfstat: Estimation and Tests of Hierarchical F-Statistics.
- Jombart, T. [T.], & Ahmed, I. (2011). Adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. doi:10.1093/bioinformatics/btr521
- Jombart, T. [Thibaut]. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405. doi:10.1093/bioinformatics/btn129
- Kassambara, A. (2018). Ggpubr: 'ggplot2' Based Publication Ready Plots (Version R package version 0.2).
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure

-
- inferences across K. *Molecular Ecology Resources*, 15(5), 1179–1191. doi:10.1111/1755-0998.12387
- Mathiesen, S. S., Thyrring, J., Hemmer-Hansen, J., Berge, J., Sukhotin, A., Leopold, P., ... Nielsen, E. E. (2016). Genetic diversity and connectivity within *Mytilus* spp. in the subarctic and Arctic. *Evolutionary Applications*, 10, 39–55. doi:10.1111/eva.12415
- Microsoft, & Weston, S. (2017). Foreach: Provides Foreach Looping Construct for R (Version R package version 1.4.4).
- Paradis, E. (2010). Pegas: An R package for population genetics with an integrated–modular approach. *Bioinformatics*, 26, 419–420.
- Patil, I. (2019). Ggstatsplot: 'ggplot2' Based Plots with Statistical Details (Version R package version 0.0.8).
- Poole, W., Gibbs, D. L., Shmulevich, I., Bernard, B., & Knijnenburg, T. A. (2016). Combining dependent *P*-values with an empirical adaptation of Brown's method. *Bioinformatics*, 32(17), i430–i436. doi:10.1093/bioinformatics/btw438
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959. doi:10.1111/j.1471-8286.2007.01758.x
- R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Revelle, W. (2018). Psych: Procedures for Psychological, Psychometric, and Personality Research (Version R package version 1.8.12). Evanston, Illinois.
- Shin, J.-H., Blay, S., McNeney, B., & Graham, J. (2006). LDheatmap: An R Function for Graphical Display of Pairwise Linkage Disequilibria Between Single Nucleotide Polymorphisms. *J Stat Soft*, 16, Code Snippet 3.
- van Etten, J. (2017). R Package gdistance: Distances and Routes on Geographical Grids. *Journal of Statistical Software*, 76(13). doi:10.18637/jss.v076.i13
- Warnes, G., Gorjanc, w. c. f. G., Leisch, F., & Man, M. (2013). Genetics: Population Genetics (Version R package version 1.3.8.1).
- Wickham, H. (2007). Reshaping Data with the reshape Package. *Journal of Statistical Software*, 21(12), 1–20.
- Wickham, H. (2017). Tidyverse: Easily Install and Load the 'Tidyverse' (Version R package version 1.2.1).

Chapitre 3

Analyse des patrons d'admixture à l'échelle génomique

3.1 Contexte et résumé de l'étude

3.1.1 Des cartes de plus en plus précises de l'admixture

L'étude de la spéciation et de l'hybridation bénéficie depuis une dizaine d'années d'une augmentation de notre capacité à obtenir des quantités toujours plus grandes d'informations (en terme de nombre d'individus) à une résolution génomique toujours plus précise (Seehausen et al. 2014). Ces capacités se traduisent en terme de recherche par l'obtention de « cartes » de plus en plus précises d'évènements d'admixture, en s'intéressant à l'ascendance locale observée le long des génomes et à la fréquence des haplotypes (Schumer et al. 2014a ; Duranton et al. 2018 ; Leitwein et al. 2018 ; Schumer et al. 2018b). Avec cet accès à l'information d'ascendance locale, de nouvelles questions peuvent être explorées empiriquement comme l'effet des variations de recombinaison sur la rétention d'allèles interspécifiques lors d'évènements d'admixture (Schumer et al. 2018b), la recherche des régions barrières au sein du génome (Ravinet et al. 2017 ; Duranton et al. 2018), ou la recherche d'introgessions adaptatives (Martin et Jiggins 2017).

Dans le contexte de populations issues d'un évènement d'admixture, l'information génomique peut être utilisée pour rechercher des régions associées à des adaptations écologiques et pour étudier le parallélisme entre évènements similaires (Meier et al. 2018). La cartographie génomique des régions parentales peut aussi être utilisée dans l'étude des facteurs permettant la maintenance d'une espèce hybride (Elgvin et al. 2017).

3.1.2 Vision génomique des admixtures entre espèces du complexe *M. edulis*

Nous avons pu observer au chapitre précédent que l'admixture était assez homogène à l'échelle de 77 marqueurs informatifs sur l'ascendance, dans l'ensemble des évènements d'admixture étudiés dans le complexe d'espèces *M. edulis*. Cependant, étant donné que ces marqueurs avaient été sélectionnés pour leurs valeurs extrêmes de différenciation entre espèces dans le chapitre 1 (Fraïsse et al. 2016a), nous espérons avoir enrichi le jeu de données avec des marqueurs de régions barrières au flux de gènes. Ainsi, nous nous attendions à ce qu'ils contribuent à des distorsions importantes dans les croisements et les populations admixées.

De ce constat d'homogénéité du mélange des génomes *M. edulis* et *M. galloprovincialis*, se pose alors la question de savoir si certaines régions du génome résistent à l'introgession. Par exemple, y a-t-il des régions du génome des moules des docks qui restent fixées *M. galloprovincialis* de méditerranée? Cette information est importante si l'on souhaite confirmer une vision génique de l'espèce, à l'image du modèle *Heliconius* qui malgré des introgressions importantes dans le génome, garde ses caractéristiques taxonomiques d'espèce via la conservation de quelques régions génomiques liées notamment aux phénotypes de patron de couleur des ailes (Consortium 2012).

Les autres observations qui nous intéressent ici sont les corrélations de distorsions de fréquences alléliques entre admixtures indépendantes (chapitre 2). On peut se demander si les corrélations observées ne sont pas liées justement au caractère extrême des marqueurs utilisés, qui bien que ne créant pas de fixations, pourraient contribuer de manière démesurée aux patrons observés. La considération des régions barrières seules rendrait l'observation précédente peu représentative de l'ensemble du génome. L'objectif suivant était donc d'observer les corrélations de distorsions à une échelle génomique supérieure, si possible maximale, c'est à dire le génome entier.

3.1.3 Résumé

Afin de s'intéresser à la résolution de l'admixture, principalement dans les populations des moules des docks résultant d'un évènement d'admixture anthropogénique récent, nous avons séquencé à faible couverture (i.e. chaque position nucléotidique n'est séquencée que quelques fois) un ensemble de génomes de référence et de génomes admixés. L'objectif est de regarder comment l'admixture est résolue à l'échelle génomique et d'analyser le parallélisme des patrons observés dans des évènements d'admixture indépendants. Sous l'hypothèse d'un isolement reproductif déterminé par des locus barrières à effet fort (modèles classiques de sous-dominance ou BDMI), l'attendu était d'observer des fixations communes à un ensemble d'admixtures. Nous observons au contraire une homogénéité de l'admixture à l'échelle génomique et aucune fixation partagée entre populations portuaires. Malgré cela, les corrélations observées au chapitre 2 sont conservées entre les évènements d'admixture impliquant des lignées proches. Ces

résultats montrent que dans les populations admixées, de nombreuses combinaisons de régions génomiques parentales ont des valeurs sélectives élevées. Un modèle polygénique de l'isolement reproductif peut être invoqué pour expliquer ces observations (chapitre 4), qui puisse rendre compte des multiples possibilités de la résolution d'une admixture.

3.2 Article 3

Genomics of *Mytilus* admixtures confirms a polygenic architecture of species barriers

1 Introduction

Anthropogenic introductions change the distribution of species and can create recent secondary contacts between lineages that diverged in allopatry. One of the outcomes of those contacts can be admixture if the lineages are not completely reproductively isolated. Such admixture events are of particular interest for evolutionary biologists as it opens a window on the study of the resolution of admixture in natural conditions.

In the “speciation genes” and “genic” view of species (Wu 2001; Orr et al. 2004), speciation is seen as the build-up of co-adapted and cohesive sets of genes, defining a species. This concept is largely embraced in the literature on speciation research, and remains compatible with the recent realisation, during the last decades, that introgression is pervasive across a large range of divergences between related taxa. The transposition of this concept at a genomic scale are “islands of speciation” (Turner et al. 2005; Ellegren et al. 2012; Ravinet et al. 2017), that are thought to harbour barrier loci, i.e. loci implicated in a reduction of gene flow due to a role in reproductive isolation.

Under classic models of hybrid incompatibilities, incompatible alleles are expected to be purged from admixed populations. An underdominance model predicts the fixation of the fitter parental homozygous genotype while a pairwise Bateson-Dobzhansky-Muller incompatibility (BDMI) model predicts the fixation of an ancestral allele at one of the two interacting loci to resolve incompatible derived alleles interactions (Schumer et al. 2018). Both models should result in practice in genomic regions exhibiting large departures from the average admixture level or complete fixations. When looking for such patterns, evolutionary biologists often rely on either lab crosses using a few generations (Moyle and Nakazato 2008; Morán and Fontdevila 2014) or on natural hybrid zones (Janoušek et al. 2012). Lab crosses can give access to large deviations due to long genomic regions reshuffled during the first generation, while old natural hybrid zones give access to long term stable states after the resolution of admixtures. However,

on one hand lab crosses may lack enough genomic combinations to investigate the effects of smaller genomic regions and on the other hand natural hybrid zones may be too old to observe transient patterns of the resolution of admixture.

We recently identified two anthropogenic admixture events in the *M. edulis* species complex (Simon et al. 2019b). The first one is an admixture between the Mediterranean lineage of *M. galloprovincialis* and the South-European lineage of *M. edulis* present in the English Channel and the North Sea. Those admixed mussels form homogeneous populations in a few big commercial ports in France and have been called “dock mussels”. They are locally in contact with one of the two native species (or both depending on the port), the South-European *M. edulis* and the Atlantic lineage of *M. galloprovincialis*. The second anthropogenic admixture happened between Atlantic *M. galloprovincialis* and the North-European lineage of *M. edulis* along the Norwegian coast (Mathiesen et al. 2016; Simon et al. 2019b). Using a small set of ancestry informative markers, we showed that dock mussels exhibit homogeneous mixing of the studied SNPs in completely admixed populations. Nonetheless, we observed departures from average admixture proportions, both between ports and between ports and lab crosses, indicative of both common history of the admixture and parallel selective effects.

Here we used low-coverage sequencing to explore polymorphisms in reference populations of the *M. edulis* species complex and in populations of admixed ancestry. We were specifically interested in scanning admixed genomes for genomic regions either fixed for Mediterranean *M. galloprovincialis* ancestry, the so-called speciation genes / barrier loci of the genic view concept, or completely introgressed by local mussel lineage ancestry. We found limited evidence for both of these but clear parallelism in admixture patterns between independent admixture events. The genetic architecture of species barriers must, therefore, be better explained by a complex polygenic model than by a few strong effect speciation genes.

2 Material and methods

2.1 Biological material

Groups and populations are defined in Table 1, which also explains the sequencing strategy used to sequence the 144 individuals. We choose to use a low-coverage strategy for 48 individuals to reduce the cost of the experiment, while sequencing the most important admixed populations at mid-coverage.

Reference populations were designed to be from diverse places for a good representation of the lineages investigated in this study. The lineages are defined following the two studies Simon et al. (2019a) and Simon et al. (2019b). Individual information is available in Table S1. Briefly, the *M. trossulus* population contains individuals from the American Atlantic coast (Saint-Lawrence), Western Pacific (Japan Sea) and Northern Europe (Barents Sea). The Mediterranean *M. galloprovincialis* population contains individuals from its North-Western part

(Thau, France), its mid Southern part (Bizerte, Tunisia), the Adriatic Sea (Italy, Gerdol et al. 2019), and its Eastern part (Heraklion, Greece). The North-European *M. edulis* population contains individuals from the Norwegian coast (Bodø) and from the Barents Sea. The other populations are less dispersed in terms of geographic distance and correspond to reference populations used in the previous papers (Simon et al. 2019a,b).

From the previous study of Simon et al. (2019b), we selected admixed individuals to provide a large range of admixture compositions in each population, while excluding secondary hybrids (e.g. backcrosses between dock mussels and *M. edulis*). The only exception to this rule are the two populations of Brest secondary (for secondary admixture) and Q13A. Brest secondary represent secondary hybridisation events between dock mussels and local Atlantic *M. galloprovincialis* in the bay of Brest, identified in (Simon et al. 2019b). Q13A population is composed of 5 individuals from the study of Coolen (2017), reanalysed in Simon et al. (2019b) and identified as later generation hybrids between dock mussels with the local *M. edulis* the North Sea.

2.2 *M. galloprovincialis* reference genome and resequenced individuals

We use the reference genome recently produced by Gerdol et al. (2019), as well as 12 resequenced *M. galloprovincialis* at high-coverage ($\sim 20x$) from the same study. The reference genome is composed of 10,577 scaffolds, with a N50 of 208 Kb (contig N50 ~ 71 Kb). In the resequenced individuals, 6 are from a *M. galloprovincialis* Atl. population in Spain, and 6 are from a *M. galloprovincialis* Med. population in Italy. Those individuals correspond to the high-coverage category of the three sequencing experiments.

2.3 DNA preparation and sequencing

Two different sequencing strategies were applied to a set of 48 and 96 individuals respectively, with varying expected depths. The first batch was composed of 48 individuals with an expected depth of 2x sequenced on a HiSeq4000, and the second one was composed of 96 individuals with an expected depth of 8x sequenced on a Novaseq.

2.3.1 Low-coverage: HiSeq4000 libraries

Low Input, Transposase Enabled (LITE) Illumina compatible libraries were constructed at the Earlham Institute (Norwich, UK) using a bespoke protocol based on the Illumina Nextera kit for 48 individual at low-coverage ($\sim 2x$). A total of 1 ng of DNA was combined with 0.9 μ L of Nextera reaction buffer and 0.1 μ L Nextera enzyme in a reaction volume of 5 μ L and incubated for 10 minutes at 55 °C. Following the initial incubation, 2.5 μ L of 2 mM custom barcoded P5 and P7 compatible primers, 5 μ L 5x Kapa Robust 2G reaction buffer, 0.5 μ L 10 mM dNTPs,

Table 1: Sequencing strategy. Nine reference populations and ten admixed populations have been sequenced. Four types of admixture have been investigated (Simon et al. 2019b)): (A) admixture between Med. *M. galloprovincialis* and South-Eu. *M. edulis* corresponding to classic dock mussels; (B) secondary admixture between dock mussels and native Atl. *M. galloprovincialis* mussels in the bay of Brest; (C) secondary admixture between dock mussels and native South-Eu. *M. edulis* mussels in the North Sea (close to Rotterdam); (D) natural admixture between Atl. *M. galloprovincialis* and South-Eu. *M. edulis*; (E) admixture between Atl. *M. galloprovincialis* and North-Eu. *M. edulis*.

Group	Population	Coverage type			Total
		Low	Mid	High	
Outgroup	<i>M. californianus</i>		2		2
References	<i>M. trossulus</i>	4	2		6
	<i>M. galloprovincialis</i> Med	4	2	6	12
	<i>M. galloprovincialis</i> Atl South	2	2	6	10
	<i>M. galloprovincialis</i> Atl Britany	4	2		6
	<i>M. edulis</i> Europe South	4	2		6
	<i>M. edulis</i> bay of Brest	4			4
	<i>M. edulis</i> Europe North	4	2		6
	<i>M. edulis</i> America	4	2		6
Admixture A	Le Havre		12		12
	Cherbourg 2003		12		12
	Cherbourg 2017		12		12
	Saint-Malo		6		6
	Brest		12		12
Admixture B	Brest (secondary)		7		7
Admixture C	Q13A	5			5
Admixture D	Aberdeen		12		12
	Jersey	12			12
Admixture E	Norway	1	7		8
	Total	48	96	12	156

0.1 μL Kapa Robust 2G enzyme and 10.4 μL water were added and mixed. Viable library molecules were then amplified by incubating the sample for 72°C for 3 minutes, followed by 14 PCR cycles consisting of 95°C for 1 minute, 65°C for 20 seconds and 72°C for 3 minutes.

Post PCR, 20 μL of amplified DNA was added to, and mixed with, 20 μL of Kapa beads and incubated at room temperature for 5 minutes to precipitate DNA molecules > 200 bp onto the beads. The beads were then pelleted on a magnetic particle concentrator (MPC), the supernatant removed, and two 70% ethanol washes performed. After removal of the final 70% ethanol wash, the beads were left to dry for 5 minutes at room temperature before the beads were resuspended with 20 μL of 10 mM Tris-HCl, pH8. This was then incubated at room temperature for 5 minutes to elute the DNA molecules. The tube was then placed back on the MPC, the beads allowed to pellet, and the aqueous phase containing the size selected DNA molecules transferred to a new tube.

The size distribution of each purified library was then determined by diluting 3 μL in 18 μL 10 mM Tris-HCl, pH8 and running this on a Perking Elmer GX. Using the proprietary software, the smear analysis function was used to determine the amount of material in the 400 to 600 bp size range and this information used to equimolar pool purified libraries. Once pooled the samples were then subjected to size selection on a Sage Science 1.5% BluePippin cassette recovering molecules between 400 and 600 bp. QC of the size selected pool was performed by running 1 μL aliquots on a Life Technologies Qubit high sensitivity assay and an Agilent DNA High Sense BioAnalyser chip and the concentration of viable library molecules measured using qPCR. For sequencing, library pools were loaded at 10pM based on an average of the qubit and qPCR concentrations, both calculated using an average molecule size of 425 bp, and sequenced on the Illumina HiSeq4000. Each sample was sequenced two times, on two different lanes of the same flow cell.

2.3.2 Mid-coverage: Novaseq libraries

Libraries for the 96 mid-coverage ($\sim 8\times$) individuals were prepared with the Truseq Nano DNA HT Sample Preparation Kit, following protocols and guidelines from the kit, at the GeT PlaGe - Genotoul sequencing platform (Toulouse, France). DNA samples were first quality checked with nanodrop and picogreen dosages for all samples. An additional fragment analyser run was used for 12 random samples.

All libraries were dosed with qPCR and checked with a fragment analyser. Equimolar pools were produced and first sequenced on an Illumina Miseq for further readjustments. The final equimolar pool of libraries was sequenced two times, on two different flow cells of an Illumina Novaseq.

2.4 Data preprocessing

Two sequencing runs for the same individuals were treated as different read-groups, and only merged at the MarkDuplicates step. All paired-end files were first quality filtered and trimmed using the `fastp` software (Chen et al. 2018). We used the low complexity filter with a threshold of 20, the base correction for paired-end data, and the adapter sequences specific for each experiment.

Each filtered paired-end run was then mapped to the reference genome using `bwa mem` (Li 2013). To account for divergence between the species of the *Mytilus* complex, we decreased the stringency of the mapping, by using the following `bwa mem` parameters `-k 10 -L 3 -B 2 -O 3`, as was done in Fraïsse et al. (2016). We also added the read-group information during this step.

We used the GATK framework to remove duplicates and perform an indel-realignment step (McKenna et al. 2010). `MarkDuplicates` from GATK4 was used to remove duplicates and merge the two read-groups for each individual, with an optical pixel distance of 2500. We finally used `RealignerTargetCreator` and `IndelRealigner` from GATK3 (as unavailable in the 4th version) to realign indels.

We used `mosdepth` (Pedersen and Quinlan 2018) to compute the distribution of coverage for each individuals. We used the 95th upper quantile of the distribution to define the upper coverage threshold for each individual.

2.5 SNP calling and filtering

We used the ANGSD framework to call single nucleotide polymorphisms (SNPs) which provides means of using genotype likelihoods for population genetics analyses, and is especially designed for low-coverage data (Korneliussen et al. 2014)).

In all the `angsd` calls, we used the following common parameters: use only the reads with a unique best hit (`-uniqueonly=1`); use only reads with both mates mapped (`-only_proper_pairs=1`); remove reads that are not primary mapped, failed to map or are marked as duplicate (`-remove_bad=1`); skip triallelic sites (`-skiptrialelic=1`); use bases with a minimum mapping quality of 20 (`-minmapq=20`); use bases with a minimum quality score of 25 (`-minq=25`); and only work with sites with a p -value less than $1e-6$ (`-snp_pval=1e-6`). Finally, we always use the samtools model of genotype likelihood (`-gl=2`) when computing them (`-doglf=1`).

A first pass with `angsd` was carried out to obtain polymorphic positions using all individuals. The additional options were a minimum major allele frequency (MAF) of 0.05 (`-minmaf=0.05`), a total maximum depth over all individuals of 15,600 ($100 \times$ number of individuals, `-setmaxdepth=15600`) and maximum of 25% missing data (`-minind=117`). This first step yielded 63,641,442 polymorphic sites. It provided counts of allelic coverage for each individual at each site.

For those polymorphic sites, we computed statistics of Hardy-Weinberg equilibrium (HWE)

in each of the reference populations (`-dohwe=1`). A site was considered to depart from HWE if the Fisher's method combined p -values for all reference populations was < 0.01 . An additional filter was applied on the per-individual coverage to remove high-coverage regions. A site was considered without excess coverage if at least half of the individuals had a coverage lower than their computed upper coverage threshold (cf. section 2.4). Sites were filtered out if they did not satisfy both criteria. `angsd` was run a final time using those sites to produce a filtered dataset, where genotype likelihood and called genotypes were outputted. We used a posterior threshold of 0.9 for genotype calling. This filtering yielded 29,493,381 polymorphic sites.

2.6 F_{ST} computation

We computed F_{ST} between all comparisons of reference populations (Table 1) using the ANGSD framework. F_{ST} computations requires the estimation of pairwise site frequency spectrums (SFS) from site allele frequencies for each reference population (using the command `realSFS`). The joint SFS serves as a prior in the following computation. Then, pairwise weighted Weir and Cockerham F_{ST} s were computed for all common sites between the two focal populations (`realSFS fst index` and `realSFS fst [print|stats|stats2]`).

2.7 Principal components analysis

A principal components analysis (PCA) was performed using a pseudo-haploid method to alleviate batch effects due to coverage differences between experiments. The method uses a random sampling at each base to reduce each individual in a haploid state (`angsd` option `-doibs=1`). Additional options to the default ones are `-minMaf 0.05 -minInd 117 -doMajorMinor 3 -doMaf 1 -doCounts 1 -GL 2 -doIBS 1 -doCov 1 -makeMatrix 1`. This step produces a covariance matrix between individuals, on which eigenvalues and eigenvectors could be computed in R. When removing individuals from the PCA, the eigen decomposition was recomputed from the reduced covariance matrix.

2.8 Allele frequencies in admixed populations of fixed sites between parental populations, and concordance analysis

The called genotype posterior probabilities (section 2.5, VCF format) were first filtered to keep sites with more than 25% missing data and 5% MAF. We then filtered sites that presented a $F_{ST} > 0.9$ in at least one comparison involving any reference populations of *M. edulis* and *M. galloprovincialis* (retrieving 53,400 sites).

To investigate the variation of allele frequencies between admixed populations, we looked at fixed sites between the two reference populations having contributed to the admixture (as a simplifying assumption). In R, we first removed sites with more than 50% missing data

in any reference population. Then we kept only fixed sites between the two focal reference populations.

As we work with fixed sites, the distortion of allele frequency in one population is taken to be $D = f - \mu$, where f is the frequency of the *M. galloprovincialis* allele in the focal population and μ the average frequency of *M. galloprovincialis* alleles on all fixed sites, which is taken as the genomic average of admixture. As such, D is oriented and positive values represent distortions toward *M. galloprovincialis* while negative ones are distortions toward *M. edulis*.

When correlating the distortions of two populations, we used Pearson's correlation coefficient and a permutation test with $1e^4$ replicates. In one comparison between two populations, as some markers were present on the same genomic scaffold, we randomly selected one SNP per scaffold and bootstrapped this operation 100 times to obtain a distribution of correlation coefficients. This bootstrapping was used to avoid spurious inflation of the correlation by linkage correlation inside over represented scaffolds.

2.9 Local ancestry estimation and time of admixture

We used the program `ancestry_hmm` from Corbett-Detig and Nielsen (2017) to estimate the timing of a pulse of *M. edulis* having produced the observed level of admixture in dock mussels. The method uses a hidden markov model to jointly estimate the posterior probability of genotypes at each position and a timing of admixture pulses. The panel was prepared using the filtered VCF file. We used the custom script of the method to prepare the panel (`vcf2ahmm.pl`, https://github.com/russcd/Ancestry_HMM). We used a minimum 10% allele frequency difference between reference panels, a recombination rate of $2e^{-8}$ bp/Morgans (Bierne 2010), a minimum of 10 alleles called in each reference panel and a simple pruning of linkage disequilibrium (LD) with a minimum distance of 5000 bp between consecutive markers. Reference panels were the Mediterranean *M. galloprovincialis* and South-European *M. edulis* (including *M. edulis* from the bay of Brest).

Seven runs were performed: five runs per dock mussel population (le Havre, two years of Cherbourg, Saint-Malo and Brest), one run with all dock mussels populations combined and one run combining le Havre and the two temporal samples of Cherbourg. The proportions of *M. edulis* ancestry during the imputed pulse was taken to be the mean allele frequency computed with fixed sites in the previous section.

3 Results and discussion

3.1 Challenges of the sequencing results

Our sequencing strategy includes variable coverages (Figure 1), requiring to take into account those effects in the analyses. In each experiment, most individuals were close to the targeted

coverage. Only two individuals, corresponding to the two *M. californianus* mussels, show reduced median coverage and proportion of genome covered at $1\times$ (Figure 1C). After investigation, this discrepancy is due to a high rate of duplicated reads, as the raw reads output were equivalent to the other individuals and the mapping proportion of reads were higher than 99%.

The low coverage experiment drastically reduces the proportion of the genome explored (Figure 1), yet most individuals have a more than 50% genome representation in the dataset. Variance in coverage between experiments are additionally causing differences in the imputation of genotypes at each position, which could bias some population genetics analyses. For this reason, we used for most analyses the genotype likelihood framework of ANGSD (Korneliussen et al. 2014).

We observed the variances in mapping coverage along the genome. Regions with no coverage appear to be localised to the same regions in multiple individuals, and this seems to be more frequent than a random distribution. In light of the recent pre-publication of the reference genome (Gerdol et al. 2019), arguing for a pan-genome structure in *Mytilus* mussels, those regions could be structural (e.g. presence-absence) variations. This possibility needs to be further investigated.

3.2 Differentiation between lineages and characteristics of admixed populations

Global pairwise F_{ST} between reference populations are distributed between 0 and 0.39 for all comparisons inside the *Mytilus* species complex (Table 2). The F_{ST} fit with expectations of the differentiation between populations (Simon et al. 2019a,b). Inside each species, F_{ST} are around one order of magnitude lower than between species. We can note that *M. edulis* from the bay of Brest can be considered equivalent to the South-Eu. *M. edulis*. While those mussels most probably come from the bay of Biscay (due to spat transplant for aquaculture), the global genomic differentiation is not high enough to split it from the South-European population in downstream analyses (although we will see below that they are slightly more introgressed by *M. edulis* alleles in the PCA, as previously reported by Fraïsse et al. 2016 and Simon et al. 2019a).

We used a pseudo-haploid method to compute a principal components analysis, which provided an analysis of the full genome without batch effects (PCA, Figure 2). *M. californianus* and *M. trossulus* were excluded from the eigen decomposition to focus on the reference populations of interest – as in Simon et al. (2019b) no influence of *M. trossulus* was detected in the studied admixed populations.

Admixed populations between Atlantic *M. galloprovincialis* and *M. edulis* shows a large variance in ancestry proportions when the populations of Aberdeen and Jersey are combined (Figure 2). This fact is firstly due to a difference of mean admixture between Aberdeen and

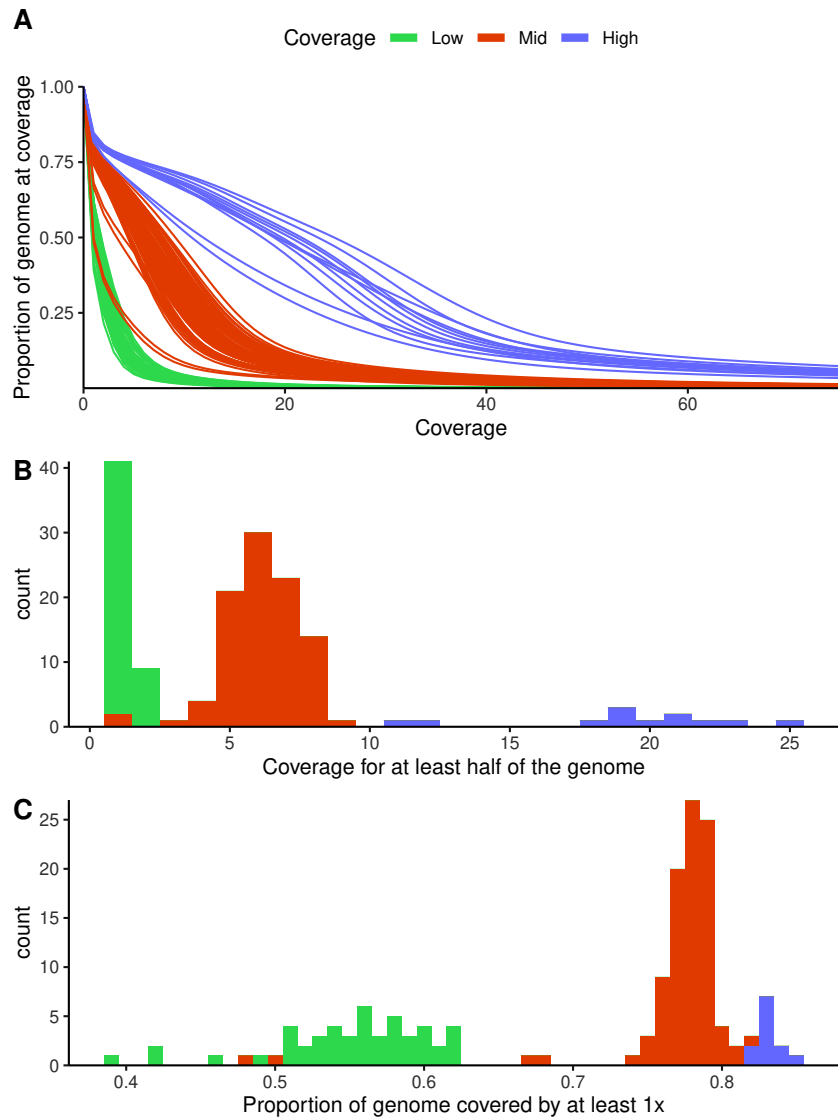


Figure 1: (A) Cumulative distribution indicating the proportion of total bases covered for a least the given coverage value. (B) Histogram of the median coverage (closer coverage level corresponding to half of the genome). (C) Histogram of the proportion of the genome with a least one read mapped.

Table 2: Global pairwise F_{ST} . The computation is described in section 2.6

	M edu Brest bay	M edu Europe North	M edu Europe South	M gallo Atl Brittany	M gallo Atl South	M gallo Med	M trossulus
M edu America	0.06	0.02	0.06	0.25	0.29	0.33	0.36
M edu Brest bay		0.00	-0.01	0.21	0.26	0.30	0.37
M edu Europe North			0.00	0.21	0.26	0.30	0.34
M edu Europe South				0.23	0.28	0.32	0.38
M gallo Atl Brittany					0.02	0.08	0.36
M gallo Atl South						0.04	0.37
M gallo Med							0.39

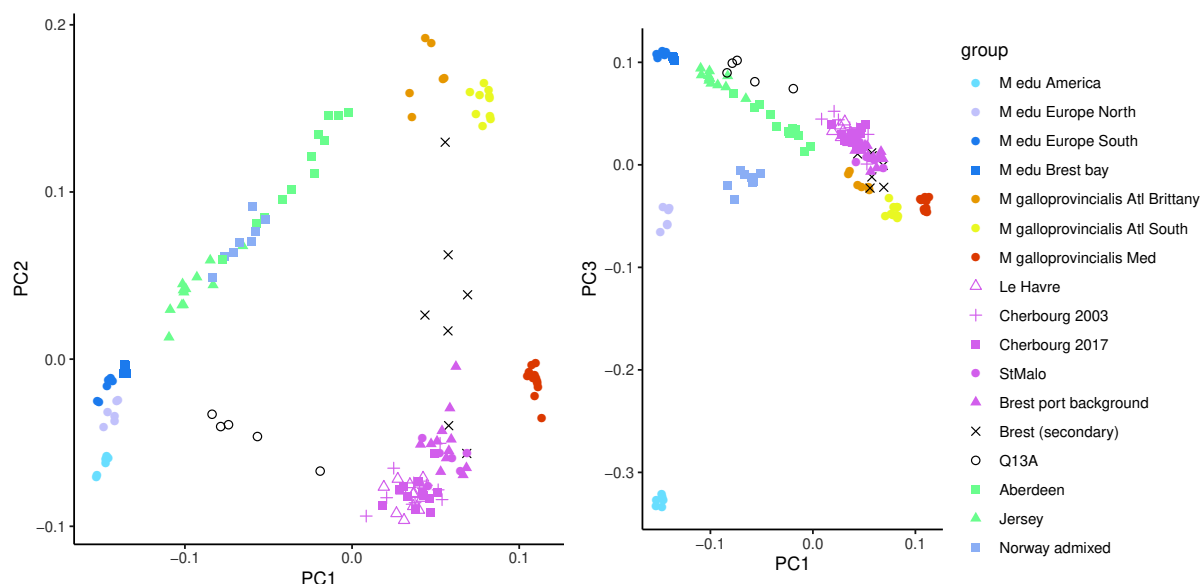


Figure 2: PCA using a pseudo-haploidisation method.

Jersey, representing two positions on the admixture spectrum. In addition, a high variance in ancestry proportions is observed in Aberdeen, maybe because this population is close to local *M. edulis* populations that continuously hybridise with them. Although Aberdeen and Jersey are our two natural admixture populations the average and variance of individual ancestry proportions are different in these two populations as desired during the design of the project.

This study includes the two anthropogenic admixture events identified in Simon et al. (2019b): (i) admixture A which are dock mussels, admixed between Mediterranean *M. galloprovincialis* and South-European *M. edulis*, established in French ports; and (ii) Admixture E which are an admixed population between Atlantic *M. galloprovincialis* and North-European *M. edulis*, present along the Norwegian coast (Table 1, Figure 2). While the same strategy was

used for those anthropogenic admixture events, the variance in individual ancestry proportions is reduced, both in Norway and dock mussels, most probably because the inflow of *M. edulis* tracts is lower in these populations.

In each species, lineages are well dissociated by the first three axes of the PCA (Figure 2). The position of the admixed clusters is globally intermediate in the PC plans between their putative parental groups. The Norwegian population is intermediate between the North-European *M. edulis* and Atlantic *M. galloprovincialis*, and dock mussels intermediate between South-European *M. edulis* and Mediterranean *M. galloprovincialis* (Figure 2). However, admixed clusters are slightly shifted on the second axis of the PCA, suggesting either that we did not sample the correct source population or that drift or selection happened in those admixed populations and somehow moved admixed populations away from the straight line between putative parental groups. This is well visible in the first PC plan where dock mussels seem more “Mediterranean” and *M. galloprovincialis* from Brittany more “Atlantic” than their respective *M. galloprovincialis* parents on PC2. Given the large sampling of the Mediterranean and Atlantic *M. galloprovincialis* populations used here and the low variance presented by this reference population in the PCA, it seems unlikely that we did not sample the diversity present in the Mediterranean Sea and the Atlantic Ocean. More probably, random drift or selection could draw variants segregating in the parental populations to higher frequencies than in the introduced populations. Although differentiated, Atlantic and Mediterranean *M. galloprovincialis* are sharing most of the variation through introgression across the contact zone between those lineages (El Ayari et al. 2019; Simon et al. 2019a). Selection in admixed genomes can promote a kind of counter-introgression from alleles of the alternative *M. galloprovincialis* genomes. Drift can also drive fixations of variants that are in low and high frequency in the parental population and these variants can be enriched in variants differentiated between Atlantic and Mediterranean *M. galloprovincialis*.

The secondary hybridized individuals identified in Simon et al. (2019b) are from two types. First the five individuals from Q13A, close to the port of Rotterdam (the Netherlands; Coolen 2017), are at mid distance between dock mussels and South-European *M. edulis*. (Figure 2 PC1-PC2), although one individual presents a closer position to the dock mussels cluster. Genetic intergradation seems more advanced in this area than in French ports and has resulted in these hybrid genomes with a higher ancestry of *M. edulis* than Mediterranean *M. galloprovincialis*.

Second, seven individuals from the bay of Brest were initially identified as containing more ancestry from Atlantic *M. galloprovincialis* than the average local dock mussels (Figure 2 PC2). In those, four appear to correspond to hybrids between dock mussels and the local *M. galloprovincialis*. One is either a local *M. galloprovincialis* or late backcross to the native mussels. Two are clustering with the dock mussels. Interestingly, two mussels classified as dock mussels show an increase in Atlantic *M. galloprovincialis* ancestry. The results obtained with the genomes do not completely overlap results obtained with the 77 ancestry-informative

SNPs regarding Atlantic *M. galloprovincialis* ancestry estimations, mainly because the initial panel was composed of only few SNPs differentiated between Atlantic and Mediterranean *M. galloprovincialis* (Simon et al. 2019b).

In the dock mussels cluster, two groups are visible (Figure 2 PC1-PC2). The first group contains mainly Brest and Saint-Malo individuals while the second is mainly composed of Le Havre and Cherbourg. In terms of location of those individuals, the main difference resides in the local *Mytilus* species they are in contact with. Le Havre and Cherbourg are in a *M. edulis* region while Saint-Malo and Brest are in a *M. galloprovincialis* region.

We also analysed four *M. edulis* from Brest that clusters with South-European *M. edulis*, although with a slight displacement towards Atlantic *M. galloprovincialis* in accordance with the Bay of Biscay *M. edulis* being slightly more introgressed (Fraïsse et al. 2016; Simon et al. 2019a).

Finally we observe on the fourth axis of the PCA a pattern that could be symptomatic of the presence of inversions (Figure S1). Indeed, the two species are grouped in three levels (mainly visible for *M. edulis*) and the clustering is not related to geography. The investigation of the presence of inversions will require the use of local PCA along the genome (Li and Ralph 2019).

3.3 Admixture patterns in replicated admixture events

We consider here only four reference populations: the Atlantic and Mediterranean *M. galloprovincialis* and the South and North European *M. edulis*. Three admixture types were investigated, A-C, D and E (Table 1).

First we looked at fixed sites between *M. galloprovincialis* and *M. edulis* overall (2196 SNPs on 600 scaffolds; Figure 3A). The main observation is that we confirm a rather homogeneous distribution of admixture proportions in the genome, with very few fixations. Some fixations of the major parent allele (the parent that contributes most to the admixed population) are observed but could be expected by the sampling variance. When populations with the same admixture history can be pooled to increase the sample size, such as dock mussel populations, no more fixation is observed. Fixations of the minor parent allele happen sometimes and are less well explained by the sampling variance. Still in the pooled sample of dock mussels, we observed no fixation of the minor parent allele. When we consider the overall distribution, the distribution is always asymmetrical, with a larger tail towards the minor ancestry parent. Asymmetries of the distributions could result from variations in recombination rates along the genome, which have been shown to influence introgression rates. For example Schumer et al. (2018) showed that ancestry from the minor parent of the admixture is more frequent in regions with high recombination rates and distant from targets of selection.

When looking exclusively at dock mussels (2196 SNPs on 600 scaffolds; Figure 3B), we observe that the fixations or large deviations are not shared by all populations and may be

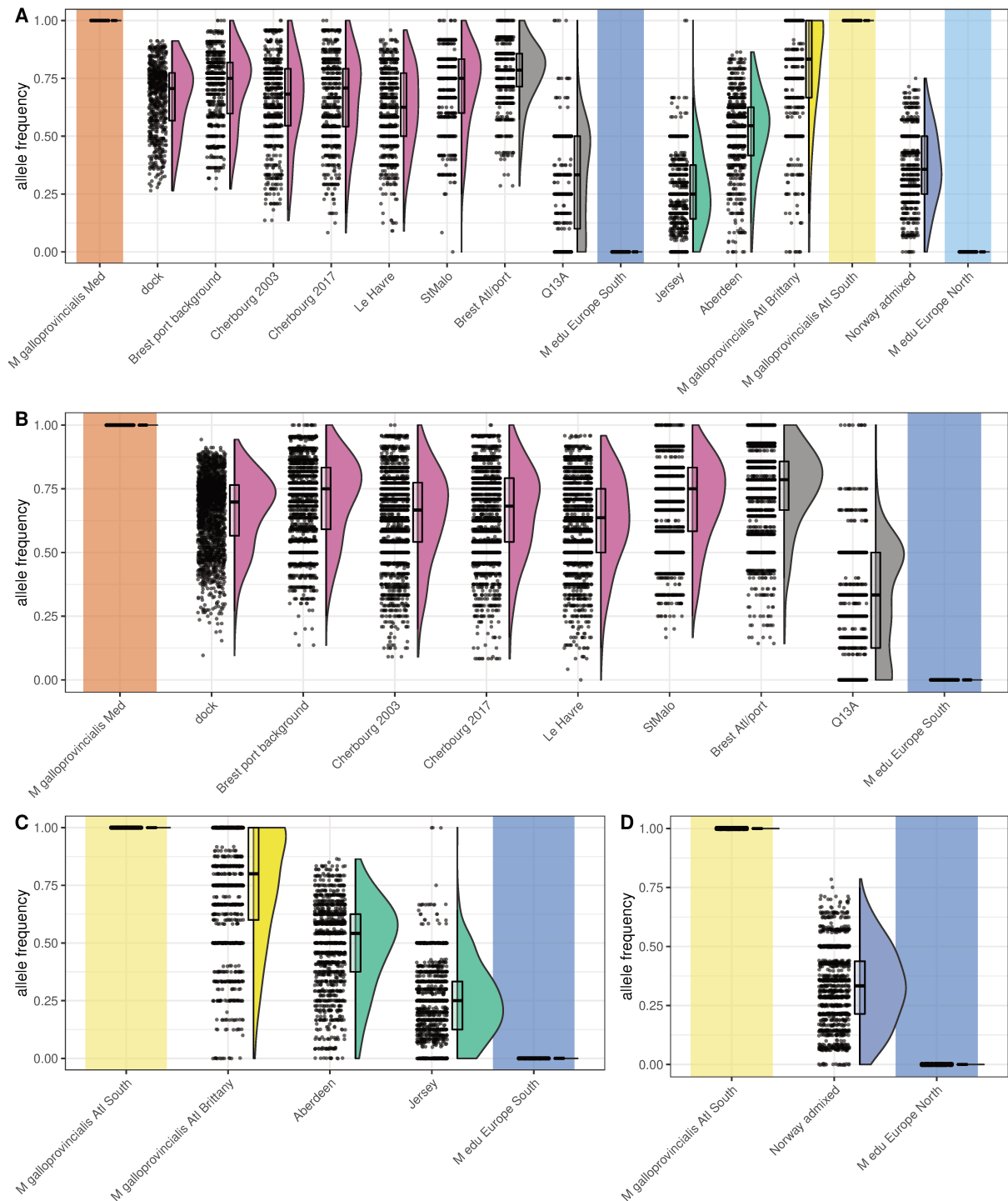


Figure 3: Frequency distributions of admixed populations for fixed SNPs. The “dock” group is the combination of populations from all admixture A type (Table 1). (A) Fixed SNPs between all *M. galloprovincialis* (Atlantic and Mediterranean) and *M. edulis* (North and South Eu.) populations (780 SNPs on 229 scaffolds). (B) Fixed SNPs between Mediterranean *M. galloprovincialis* and South-Eu. *M. edulis* (2196 SNPs on 600 scaffolds). (C) Fixed SNPs between Atlantic *M. galloprovincialis* and South-Eu. *M. edulis* (1348 SNPs on 357 scaffolds). (D) Fixed SNPs between Atlantic *M. galloprovincialis* and North-Eu. *M. edulis* (1363 SNPs on 378 scaffolds).

reflecting local drift or local differences in admixture resolution. The absence of shared fixations or large distortions is more prone on the *M. galloprovincialis* side, however the global asymmetry is conserved and large distortions are present in the *M. edulis* direction.

In the admixture D type, natural admixtures between Atlantic *M. galloprovincialis* and *M. edulis* (Figure 3C), we observe increased fixations of *M. edulis* alleles that seem to reflect patterns seen in the introgressed population of *M. galloprovincialis* in Brittany. When considering only fixed sites between *M. galloprovincialis* from Brittany and *M. edulis* (Figure S2), fixations in Aberdeen disappear. This observation suggests older comets of introgression in the parental population. Note that comets of introgression are interesting but were known and the genome project aims to understand admixture resolution in recent anthropogenic admixture events. In Jersey the major parent is *M. edulis*. The asymmetry of the distribution is reversed and we observe a few fixations of *M. galloprovincialis* alleles.

Finally, the Norwegian admixture type E (Figure 3D) shows the most balanced distribution of allele frequencies, with fixation pertaining more to important *M. edulis* introgression than to large distortions. In addition, no *M. galloprovincialis* allele fixation is observed, in contrary to the observation in Jersey.

With observations from multiple events of admixture, both anthropogenic and natural, we conclude that admixture resolution in mussels does not fit well with models of strong effect species genes / barrier loci (involved in reproductive isolation). Under these models, we would have expected fixations of species gene alleles in parallel in replicated admixture events, while we observed either no fixation or private fixations. Particularly clear in dock mussels overall, we observe a strong homogeneity of the admixture along the genome that must produce a large number of fit combinations. A polygenic model could better explain this result and explains that different solutions exist to resolve the admixture (Simon et al. 2018).

The quest of selected loci under a genome scan framework is here rather to search in the side of enhanced introgression of the minor parent allele. The 5% lower quantile of the frequencies in all dock mussels contains 110 SNPs spread on 40 scaffolds. We identified six SNPs with missense amino-acid changes including four radical changes (vep output using the annotation produced by Gerdol et al. 2019). Four are within a Cadherin protein, one is impacting a Tumor Necrosis Factor Receptor (TNFR) with cysteine rich domain and one touches an OTU domain cysteine protease possibly related to TNF. When investigating outliers in populations implicating Atlantic *M. galloprovincialis* (the Brittany population, Jersey and Aberdeen), three scaffolds appear to be in common in the *M. edulis* allele distortions. On those scaffolds, we identified three common missense amino-acid variants with radical changes with two in the Period protein (Per, implicated in circadian rhythms) and one in an Ankyrin repeat-containing domain. Further investigations of those candidate genes will be necessary to draw any conclusion on their role in species incompatibilities or local adaptation. Interestingly a TNF factor had been identified as an introgression outlier in Fraisse et al. (2016). We posit that finding proteins linked to this function in this scan (TNFR and OTU domain) might not

be simple chance.

3.4 Correlated distortions between common admixture types

Using 780 fixed SNPs on 229 scaffolds, we confirm the results obtained in Simon et al. (2019b). Although deviations to the average admixture proportion is not consistent across replicated admixture events, we observed an overall correlation of the deviations. Dock mussels exhibit the strongest correlations between the four ports studied and between the two temporal samples of Cherbourg. As argued in Simon et al. (2019b), this observation pertains to the probable common history of admixture and colonisation shared by those populations.

Correlations between Atlantic and Mediterranean *M. galloprovincialis* type admixtures are low and non-significant (Figure 4). Nonetheless, the Atlantic Brittany *M. galloprovincialis* reference population shows increase correlations with dock mussels compared to other populations. This correlation is increased when comparing to the secondary Brest hybridisation, as it involves the local Brittany *M. galloprovincialis* as one parent.

The other group of high correlations is between admixtures implicating the Atlantic *M. galloprovincialis* (Figure 4). As previously observed, while not implicating the same *M. edulis* lineage, the natural admixtures are correlated with the Norwegian admixture. The correlation is particularly high between Norway and Aberdeen, which could be linked to a shared parent in British Isles *M. galloprovincialis*, different from the Brittany *M. galloprovincialis*.

3.5 New estimation of an admixture time

We used a hidden markov model (Corbett-Detig and Nielsen 2017) to estimate the time of admixture between the Mediterranean *M. galloprovincialis* and South-European *M. edulis*. Estimation was performed both for each population of dock mussels and to two combined datasets, an overall dataset dock mussels and a dataset only including le Havre and Cherbourg populations (Table 3).

Table 3: Ancestry HMM estimates of admixture time. The estimate for Saint-Malo and Brest reached the upper limit of generations (500) and are not reported. The timing of the pulse of ancestry resulting in the observed level of *M. edulis* ancestry is estimated in number of generations.

Population	<i>M. edulis</i> pulse proportion	Time estimate (gen.)
All	0.34	82
Havre + Cher	0.30	28
Le Havre	0.38	64
Cherbourg 2003	0.36	75
Cherbourg 2017	0.34	25
Saint-Malo	0.30	900
Brest	0.30	300



Figure 4: Mean bootstrap Pearson's correlation coefficient r for all comparisons (100 replicates) using one fixed SNP per scaffold between *M. galloprovincialis* and *M. edulis* (229 SNPs, bootstrapping draws randomly in each scaffold). The 95% confidence interval is indicated in brackets. Bold numbers indicate that the worst p -value of the replicates is inferior to $5e^{-4}$ for Bonferroni correction (which is a particularly stringent condition).

Credible time estimates (excluding Saint-Malo and Brest) range between 28 and 82 generations. As observed in Simon et al. (2019b) with a more rudimentary method, the admixture is relatively recent and confirms the role of recent human introduction in this process. A time estimate of more than 28 generations is also in line with the observation of dock mussels in the port of Swansea 40 years ago if we take a generation time of ~ 2 years (Skibinski et al. 1978; Simon et al. 2019b).

The result of 82 generations when using all dock mussels may pertain to biases introduced by the Brest and Saint-Malo populations. Indeed, inference on those populations gives large times of admixture (Table 3). We hypothesise that this inflation of the admixture time is due to the influence of Atlantic Brittany *M. galloprovincialis* ancestry in those populations (see discussion of Figure 2), that may introduce in dock mussels genomes short tracts of *M. edulis* (including comets of introgression, Fraïsse et al. 2016; Simon et al. 2019a) resulting from much older introgressions in Atlantic Brittany *M. galloprovincialis*.

Results of 64 and 75 generations estimates in le Havre and Cherbourg 2003 are counter-intuitive. First ongoing hybridisation in le Havre is likely to introduce longer tracts of *M. edulis* in the population which should bias the estimation of the time to admixture downward rather than upward (Racimo et al. 2015). Second the population of Cherbourg 2003 should have a younger time estimate than the 2017 sample (or at least equivalent). This paradoxical result was already obtained with few markers but with known recombination map positions in Simon et al. (2019b). A new pulse of *M. edulis* introgression in between the two sampling times might explain the results but the ancestry proportions are very similar in the two sampling dates and the estimated *M. edulis* pulse proportion is higher in 2003 than 2017.

Overall, admixture timing estimations could be improved by correctly accounting for ancestral LD (shown to bias estimates, Corbett-Detig and Nielsen 2017) and by using multiple pulse models or multiple contributing populations. For example, the estimation in Brest and Saint-Malo could be improved by taking into account an Atlantic *M. galloprovincialis* component. Improving the LD pruning would require to obtain a genetic map in one of the reference population.

Conclusion

The analysis of genome sequences in replicated admixture events confirmed that there exists a diversity of sufficiently fit hybrid genotypes bridging the various genetic lineages of the *M. edulis* species complex. A genic view of species would have predicted that, with an access to whole genome sequences, we should have found speciation genes / barrier loci consistently fixed despite admixture in the rest of the genome. This was not what we found. Although departures from average admixture proportions are correlated between admixed populations, suggesting shared selective effects affecting a particular genomic region, this observation is not totally consistent and varies with the lineages implicated in the admixture. We did not

always find fixations and, if present, they were often not shared between different populations. In addition, parallelism is also driven by genome regions in which the introgression of the minor parent allele is enhanced. These regions create an asymmetric distribution of deviations with a tail toward excess of minor parent alleles. This distribution is not easily explained by underdominance or pairwise BDMIs with equal fitness of parental genotypes. Such models would have predicted the fixation of the major parent genotypes at an underdominant locus and at one of the two partner loci of a BDMI (Schumer et al. 2018). Asymmetric fitness of parental genotypes is one explanation as the direction of fixation depends of the equilibrium frequency, \hat{p} , of the bi-stable variant (underdominant or BDMI, Barton and Turelli 2011). Symmetric bi-stable variants have $\hat{p} = 0.5$ and we predict fixation of the major parent allele after admixture, but some asymmetric bi-stable variant might have \hat{p} above the admixture proportion and the minor parent allele is expected to fix. Asymmetric fitness of parental genotypes is therefore one ingredient to explore. However it does not explain the absence of fixation. We posit that polygenic selection should be invoked in addition to explain the paucity of fixations in our data. Under a polygenic model, multiple combinations of parental alleles can produce the same fitness and recombination uncouples the associations such that small effect mutations that contribute to departure initially tend to recover a nearly neutral behavior after some generations of admixture and do not necessarily reach fixation. The next chapter explores the utility of Fisher's geometric model to include polygenic selection and asymmetric parental fitness effects in the prediction of hybrid fitness.

References

- Barton, N. H. and M. Turelli (2011). "Spatial Waves of Advance with Bistable Dynamics: Cytoplasmic and Genetic Analogues of Allee Effects." *The American naturalist* 178.3, E48–E75. DOI: 10.1086/661246.
- Bierne, N. (2010). "The Distinctive Footprints of Local Hitchhiking in a Varied Environment and Global Hitchhiking in a Subdivided Population: The Distinctive Footprints of Local and Global Hitchhiking". *Evolution* 64.11, pp. 3254–3272. DOI: 10.1111/j.1558-5646.2010.01050.x.
- Chen, S., Y. Zhou, Y. Chen, and J. Gu (2018). "Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor". *Bioinformatics* 34, pp. i884–i890. DOI: 10.1101/274100.
- Coolen, J. W. P. (2017). "North Sea Reefs: Benthic Biodiversity of Artificial and Rocky Reefs in the Southern North Sea". PhD thesis. Wageningen: Wageningen University.
- Corbett-Detig, R. and R. Nielsen (2017). "A Hidden Markov Model Approach for Simultaneously Estimating Local Ancestry and Admixture Time Using Next Generation Sequence Data in Samples of Arbitrary Ploidy". *PLOS Genetics* 13.1. Ed. by H. M. Kang, e1006529. DOI: 10.1371/journal.pgen.1006529.
- El Ayari, T., N. Trigui El Menif, B. Hamer, A. E. Cahill, and N. Bierne (2019). "The Hidden Side of a Major Marine Biogeographic Boundary: A Wide Mosaic Hybrid Zone at the Atlantic–Mediterranean Divide Reveals the Complex Interaction between Natural and Genetic Barriers in Mussels". *Heredity* 122, pp. 770–784. DOI: 10.1038/s41437-018-0174-y.

-
- Ellegren, H. et al. (2012). “The Genomic Landscape of Species Divergence in *Ficedula* Flycatchers”. *Nature* 491.7426, pp. 756–760. DOI: 10.1038/nature11584.
- Fraïsse, C., K. Belkhir, J. J. Welch, and N. Bierne (2016). “Local Interspecies Introgression Is the Main Cause of Extreme Levels of Intraspecific Differentiation in Mussels”. *Molecular Ecology* 25.1, pp. 269–286. DOI: 10.1111/mec.13299.
- Gerdol, M. et al. (2019). “Massive Gene Presence/Absence Variation in the Mussel Genome as an Adaptive Strategy: First Evidence of a Pan-Genome in Metazoa”. *bioRxiv*. DOI: 10.1101/781377.
- Janoušek, V. et al. (2012). “Genome-Wide Architecture of Reproductive Isolation in a Naturally Occurring Hybrid Zone between *Mus Musculus Musculus* and *M. m. Domesticus*: REPRODUCTIVE ISOLATION IN HOUSE MOUSE”. *Molecular Ecology* 21.12, pp. 3032–3047. DOI: 10.1111/j.1365-294X.2012.05583.x.
- Korneliusson, T. S., A. Albrechtsen, and R. Nielsen (2014). “ANGSD: Analysis of Next Generation Sequencing Data”. *BMC Bioinformatics* 15.1. DOI: 10.1186/s12859-014-0356-4.
- Li, H. and P. Ralph (2019). “Local PCA Shows How the Effect of Population Structure Differs Along the Genome”. *Genetics* 211, pp. 289–304. DOI: 10.1534/genetics.118.301747.
- Li, H. (2013). “Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM”.
- Mathiesen, S. S. et al. (2016). “Genetic Diversity and Connectivity within *Mytilus* Spp. in the Subarctic and Arctic”. *Evolutionary Applications* 10, pp. 39–55. DOI: 10.1111/eva.12415.
- McKenna, A. et al. (2010). “The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data”. *Genome Research* 20.9, pp. 1297–1303. DOI: 10.1101/gr.107524.110.
- Morán, T. and A. Fontdevila (2014). “Genome-Wide Dissection of Hybrid Sterility in *Drosophila* Confirms a Polygenic Threshold Architecture”. *Journal of Heredity* 105.3, pp. 381–396. DOI: 10.1093/jhered/esu003.
- Moyle, L. C. and T. Nakazato (2008). “Complex Epistasis for Dobzhansky-Muller Hybrid Incompatibility in *Solanum*”. *Genetics* 181.1, pp. 347–351. DOI: 10.1534/genetics.108.095679.
- Orr, H. A., J. P. Masly, and D. C. Presgraves (2004). “Speciation Genes”. *Current Opinion in Genetics & Development* 14.6, pp. 675–679. DOI: 10.1016/j.gde.2004.08.009.
- Pedersen, B. S. and A. R. Quinlan (2018). “Mosdepth: Quick Coverage Calculation for Genomes and Exomes”. *Bioinformatics* 34.5, pp. 867–868. DOI: 10.1093/bioinformatics/btx699.
- Racimo, F., S. Sankararaman, R. Nielsen, and E. Huerta-Sánchez (2015). “Evidence for Archaic Adaptive Introgression in Humans.” *Nature reviews. Genetics* 16.6, pp. 359–371. DOI: 10.1038/nrg3936.
- Ravinet, M. et al. (2017). “Interpreting the Genomic Landscape of Speciation: A Road Map for Finding Barriers to Gene Flow”. *Journal of Evolutionary Biology* 30.8, pp. 1450–1477. DOI: 10.1111/jeb.13047.
- Schumer, M. et al. (2018). “Natural Selection Interacts with Recombination to Shape the Evolution of Hybrid Genomes”. *Science* 360.6389, pp. 656–660. DOI: 10.1126/science.aar3684.
- Simon, A., N. Bierne, and J. J. Welch (2018). “Coadapted Genomes and Selection on Hybrids: Fisher’s Geometric Model Explains a Variety of Empirical Patterns”. *Evolution Letters* 2.5, pp. 472–498. DOI: 10.1002/evl3.66.
- Simon, A., C. Fraïsse, T. El Ayari, C. Liautard-Haag, P. Strelkov, J. J. Welch, and N. Bierne (2019a). “Local Introgression at Two Spatial Scales in Mosaic Hybrid Zones of Mussels”. *bioRxiv*. DOI: 10.1101/818559. Submitted.

- Simon, A. et al. (2019b). “Replicated Anthropogenic Hybridisations Reveal Parallel Patterns of Admixture in Marine Mussels”. *Evolutionary Applications*. DOI: 10.1111/eva.12879. Pre-published.
- Skibinski, D. O. F., J. A. Beardmore, and M. Ahmad (1978). “Genetic Aids to the Study of Closely Related Taxa of the Genus *Mytilus*”. *Marine Organisms: Genetics, Ecology, and Evolution*. Ed. by B. Battaglia and J. A. Beardmore. Proceedings of a NATO Advanced Study Research Inst. on the Genetics, Evolution, and Ecology of Marine Organisms Held in the Fondazione Giorgio Cini, Venice, 1977. Plenum Publishing Corporation, pp. 469–486.
- Turner, T. L., M. W. Hahn, and S. V. Nuzhdin (2005). “Genomic Islands of Speciation in *Anopheles Gambiae*”. *PLoS Biology* 3.9, pp. 1572–1578. DOI: 10.1371/journal.pbio.0030285.
- Wu, C.-I. (2001). “The Genic View of the Process of Speciation: Genic View of the Process of Speciation”. *Journal of Evolutionary Biology* 14.6, pp. 851–865. DOI: 10.1046/j.1420-9101.2001.00335.x.

Supplementary information

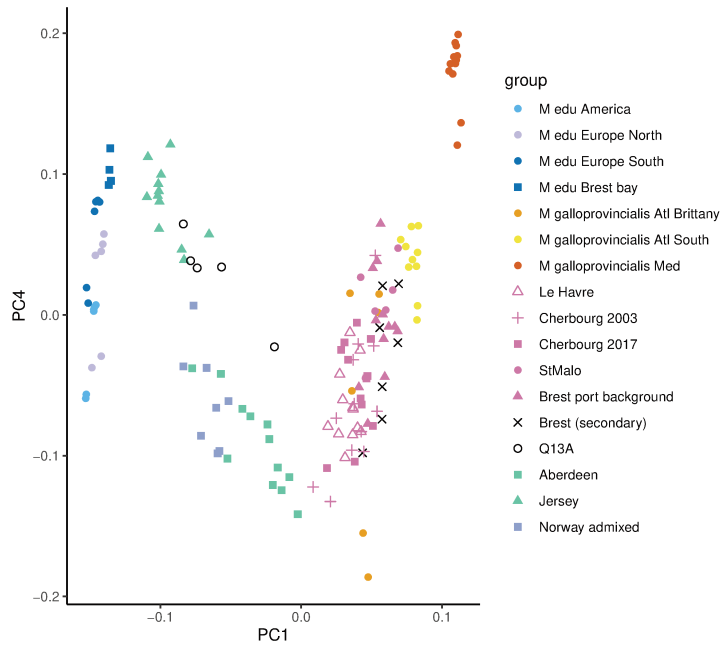


Figure S1: Axis 1 and 4 for the PCA using the pseudo-haploid method.

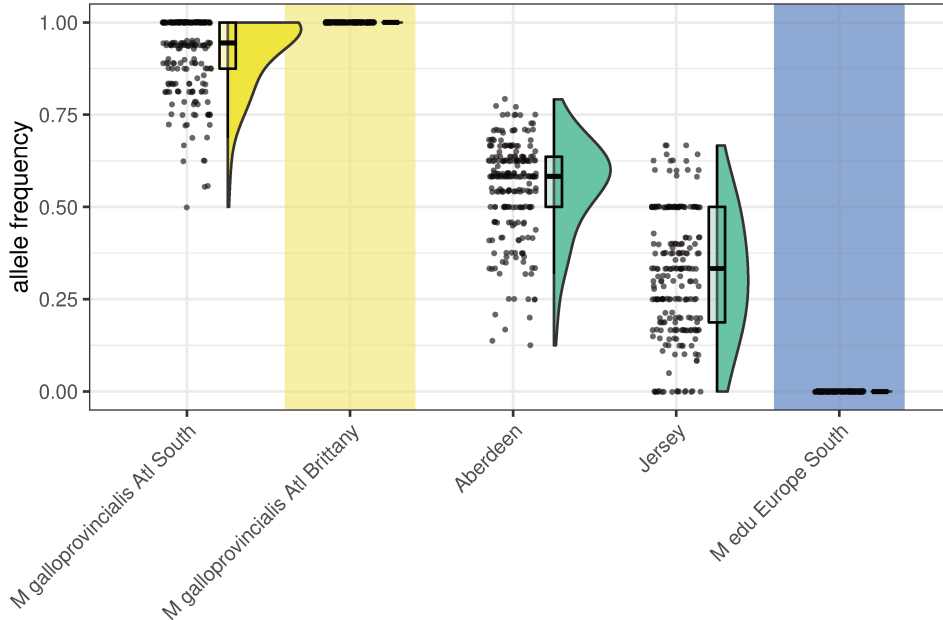


Figure S2: Allele frequencies distributions for the natural admixtures of Aberdeen and Jersey, keeping fixed SNPs between the Atlantic *M. galloprovincialis* from Brittany and South-European *M. edulis* (231 SNPs on 107 scaffolds).

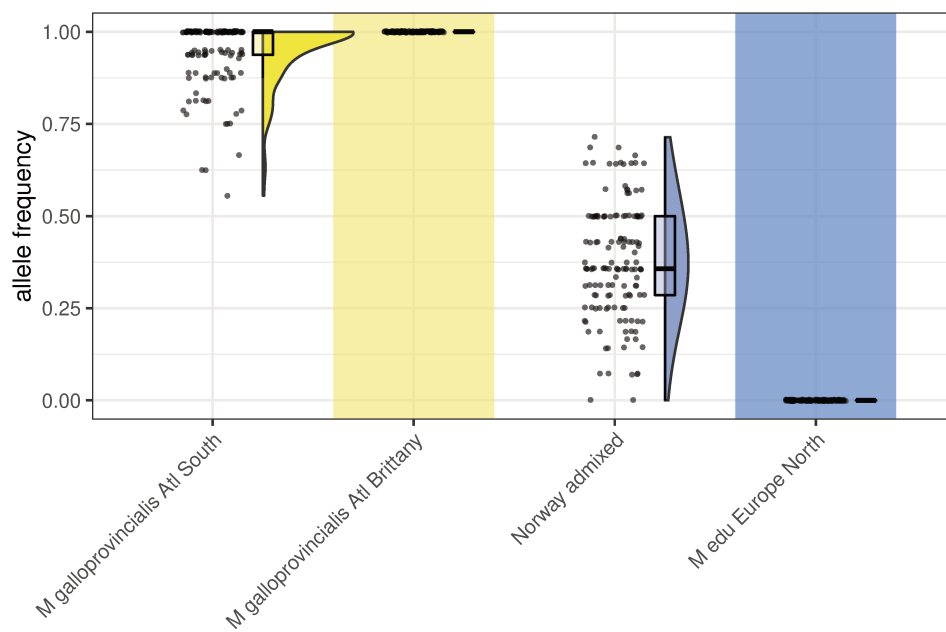


Figure S3: Allele frequencies distributions for the Norwegian admixture, keeping fixed SNPs between the Atlantic *M. galloprovincialis* from Brittany and North-European *M. edulis* (142 SNPs on 83 scaffolds).

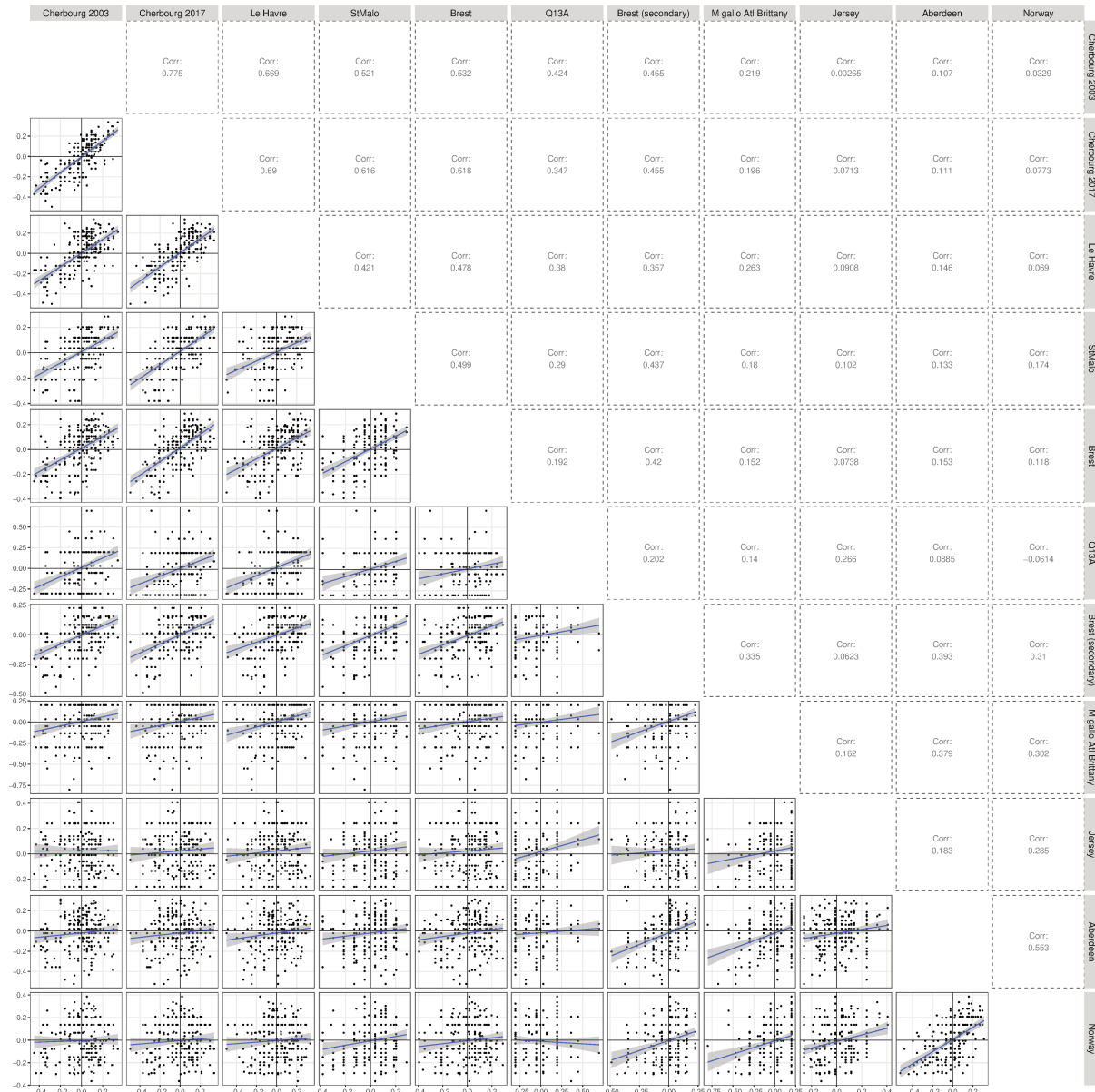


Figure S4: Each pair of correlation from Figure 4



Figure S5: Replicates of the correlation coefficient for random draws of markers on each scaffold.

Table S1: Individual information.

ind	group	experiment	pop	locality	lat	long
Aber-50	Aberdeen	Novaseq	Aber	ABD, Aberdeen, Scotland, UK	57.15	-2.07
Aber-52	Aberdeen	Novaseq	Aber	Aberdeen, Scotland, UK	57.15	-2.07
Aber-53	Aberdeen	Novaseq	Aber	Aberdeen, Scotland, UK	57.15	-2.07
Aber-54	Aberdeen	Novaseq	Aber	Aberdeen, Scotland, UK	57.15	-2.07
Aber-59	Aberdeen	Novaseq	Aber	Aberdeen, Scotland, UK	57.15	-2.07
Aber-61	Aberdeen	Novaseq	Aber	Aberdeen, Scotland, UK	57.15	-2.07
Aber-62	Aberdeen	Novaseq	Aber	Aberdeen, Scotland, UK	57.15	-2.07
Aber-64	Aberdeen	Novaseq	Aber	Aberdeen, Scotland, UK	57.15	-2.07
Aber-72	Aberdeen	Novaseq	Aber	Aberdeen, Scotland, UK	57.15	-2.07
Aber-75	Aberdeen	Novaseq	Aber	Aberdeen, Scotland, UK	57.15	-2.07
Aber-78	Aberdeen	Novaseq	Aber	Aberdeen, Scotland, UK	57.15	-2.07
Aber-79	Aberdeen	Novaseq	Aber	Aberdeen, Scotland, UK	57.15	-2.07
Brest-1-02	Brest Atl/port	Novaseq	Brest-1	Lagonna Daoulas, France	48.32	-4.29
Brest-15-08	Brest Atl/port	Novaseq	Brest-15	Porz Tinduff, France	48.34	-4.37
Brest-16-04	Brest Atl/port	Novaseq	Brest-16	Lauberlac'h, France	48.34	-4.41
Brest-19-06	Brest Atl/port	Novaseq	Brest-19	Marina Moulin Blanc Nord, France	48.39	-4.43
Brest-21-07	Brest Atl/port	Novaseq	Brest-21	Brest commerce A, France	48.38	-4.47
Brest-22-01	Brest Atl/port	Novaseq	Brest-22	Brest commerce B, France	48.38	-4.47
Brest-38-03	Brest Atl/port	Novaseq	Brest-38	Elorn 2, France	48.41	-4.34
Brest-19-10	Brest port background	Novaseq	Brest-19	Marina Moulin Blanc Nord, France	48.39	-4.43
Brest-20-04	Brest port background	Novaseq	Brest-20	Marina Moulin Blanc Sud, France	48.39	-4.43
Brest-22-11	Brest port background	Novaseq	Brest-22	Brest commerce B, France	48.38	-4.47
Brest-23-05	Brest port background	Novaseq	Brest-23	Brest commerce C, France	48.38	-4.48
Brest-27-11	Brest port background	Novaseq	Brest-27	Marina du Chateau A, France	48.38	-4.49
Brest-36-03	Brest port background	Novaseq	Brest-36	Térénez, France	48.28	-4.27
Brest-36-06	Brest port background	Novaseq	Brest-36	Térénez, France	48.28	-4.27
Brest-38-07	Brest port background	Novaseq	Brest-38	Elorn 2, France	48.41	-4.34
Brest-39-04	Brest port background	Novaseq	Brest-39	Elorn 2B, France	48.41	-4.34

Continued on next page

Table S1 – *Continued from previous page*

ind	group	experiment	pop	locality	lat	long
Brest-42-06	Brest port background	Novaseq	Brest-42	Elorn 6, France	48.40	-4.37
Brest-45-10	Brest port background	Novaseq	Brest-45	Elorn 14, France	48.39	-4.39
Brest-46-01	Brest port background	Novaseq	Brest-46	Elorn 15, France	48.39	-4.39
Cher03-04	Cherbourg 2003	Novaseq	Cher03	Cherbourg, France	49.64	-1.62
Cher03-05	Cherbourg 2003	Novaseq	Cher03	Cherbourg, France	49.64	-1.62
Cher03-06	Cherbourg 2003	Novaseq	Cher03	Cherbourg, France	49.64	-1.62
Cher03-07	Cherbourg 2003	Novaseq	Cher03	Cherbourg, France	49.64	-1.62
Cher03-10	Cherbourg 2003	Novaseq	Cher03	Cherbourg, France	49.64	-1.62
Cher03-11	Cherbourg 2003	Novaseq	Cher03	Cherbourg, France	49.64	-1.62
Cher03-18	Cherbourg 2003	Novaseq	Cher03	Cherbourg, France	49.64	-1.62
Cher03-19	Cherbourg 2003	Novaseq	Cher03	Cherbourg, France	49.64	-1.62
Cher03-21	Cherbourg 2003	Novaseq	Cher03	Cherbourg, France	49.64	-1.62
Cher03-22	Cherbourg 2003	Novaseq	Cher03	Cherbourg, France	49.64	-1.62
Cher03-25	Cherbourg 2003	Novaseq	Cher03	Cherbourg, France	49.64	-1.62
Cher03-29	Cherbourg 2003	Novaseq	Cher03	Cherbourg, France	49.64	-1.62
Cher-A-01	Cherbourg 2017	Novaseq	Cher-A	Cherbourg A, France	49.65	-1.62
Cher-A-02	Cherbourg 2017	Novaseq	Cher-A	Cherbourg A, France	49.65	-1.62
Cher-B-01	Cherbourg 2017	Novaseq	Cher-B	Cherbourg B, France	49.65	-1.62
Cher-B-09	Cherbourg 2017	Novaseq	Cher-B	Cherbourg B, France	49.65	-1.62
Cher-C-02	Cherbourg 2017	Novaseq	Cher-C	Cherbourg C, France	49.67	-1.63
Cher-D-10	Cherbourg 2017	Novaseq	Cher-D	Cherbourg D, France	49.65	-1.62
Cher-E-11	Cherbourg 2017	Novaseq	Cher-E	Cherbourg E, France	49.64	-1.62
Cher-E-12	Cherbourg 2017	Novaseq	Cher-E	Cherbourg E, France	49.64	-1.62
Cher-F-02	Cherbourg 2017	Novaseq	Cher-F	Cherbourg F, France	49.65	-1.62
Cher-F-08	Cherbourg 2017	Novaseq	Cher-F	Cherbourg F, France	49.65	-1.62
Cher-G-01	Cherbourg 2017	Novaseq	Cher-G	Cherbourg G, France	49.64	-1.62
Cher-G-07	Cherbourg 2017	Novaseq	Cher-G	Cherbourg G, France	49.64	-1.62
Jer-02	Jersey	Hiseq	Jer	Jersey, France	49.17	-2.02

Continued on next page

Table S1 – *Continued from previous page*

ind	group	experiment	pop	locality	lat	long
Jer-03	Jersey	Hiseq	Jer	Jersey, France	49.17	-2.02
Jer-04	Jersey	Hiseq	Jer	Jersey, France	49.17	-2.02
Jer-09	Jersey	Hiseq	Jer	Jersey, France	49.17	-2.02
Jer-10	Jersey	Hiseq	Jer	Jersey, France	49.17	-2.02
Jer-11	Jersey	Hiseq	Jer	Jersey, France	49.17	-2.02
Jer-12	Jersey	Hiseq	Jer	Jersey, France	49.17	-2.02
Jer-19	Jersey	Hiseq	Jer	Jersey, France	49.17	-2.02
Jer-22	Jersey	Hiseq	Jer	Jersey, France	49.17	-2.02
Jer-24	Jersey	Hiseq	Jer	Jersey, France	49.17	-2.02
Jer-26	Jersey	Hiseq	Jer	Jersey, France	49.17	-2.02
Jer-28	Jersey	Hiseq	Jer	Jersey, France	49.17	-2.02
B-amont-02	Le Havre	Novaseq	B_amont	Bassin Amont, Le Havre, France	49.47	0.15
Havre-B-08	Le Havre	Novaseq	Havre_B	Le Havre, France	49.47	0.12
Havre-C-06	Le Havre	Novaseq	Havre_C	Le Havre, France	49.47	0.12
Havre-E-05	Le Havre	Novaseq	Havre_E	Le Havre, France	49.47	0.16
Havre-F-11	Le Havre	Novaseq	Havre_F	Le Havre, France	49.48	0.16
Havre-G-13	Le Havre	Novaseq	Havre_G	Le Havre, France	49.47	0.17
Havre-J-08	Le Havre	Novaseq	Havre-J	Le Havre, France	49.49	0.12
LeHaP11-13	Le Havre	Novaseq	LeHaP11	Le Havre P11, France	49.45	0.18
LeHaP2-03	Le Havre	Novaseq	LeHaP2	Le Havre P2, France	49.46	0.11
LeHaP3-10	Le Havre	Novaseq	LeHaP3	Le Havre P3, France	49.46	0.11
LeHaP4-08	Le Havre	Novaseq	LeHaP4	Le Havre P4, France	49.46	0.12
LeHaP9-13	Le Havre	Novaseq	LeHaP9	Le Havre P9, France	49.45	0.17
Mcalif-AK-08	M californianus	Novaseq	Alaska			
Mcalif-PB-41	M californianus	Novaseq	Punta Baja			
SHM-01	M edu America	Hiseq	10-LgIs-SHM	Snug Harbor Marina, Charlestown, Long Island, USA	41.39	-71.52
SHM-02	M edu America	Novaseq	10-LgIs-SHM	Snug Harbor Marina, Charlestown, Long Island, USA	41.39	-71.52
QNT-38	M edu America	Hiseq	11-LgIs-QNT	Quonochontaug, Long Island, USA	41.33	-71.71

Continued on next page

Table S1 – Continued from previous page

ind	group	experiment	pop	locality	lat	long
QNT-03	M edu America	Novaseq	11-Lgls-QNT	Quonochontaug, Long Island, USA	41.33	-71.71
OST-02	M edu America	Hiseq	OST	Old Saybrook town, Hartland drive, USA	41.26	-72.38
OST-20	M edu America	Hiseq	OST	Old Saybrook town, Hartland drive, USA	41.26	-72.38
Brest-15-02	M edu Brest bay	Hiseq	Brest-15	Porz Tinduff, France	48.34	-4.37
Brest-19-09	M edu Brest bay	Hiseq	Brest-19	Marina Moulin Blanc Nord, France	48.39	-4.43
Brest-29-03	M edu Brest bay	Hiseq	Brest-29	Brest commerce D, France	48.37	-4.46
Brest-39-05	M edu Brest bay	Hiseq	Brest-39	Elorn 2B, France	48.41	-4.34
RET-13	M edu Europe North	Hiseq	19-Retinskoye	Retinskoye, Russia	69.11	33.39
RET-01	M edu Europe North	Novaseq	19-Retinskoye	Retinskoye, Russia	69.11	33.39
Bodo-01	M edu Europe North	Hiseq	Bodo	Bodø, Norway	67.22	15.00
Bodo-04	M edu Europe North	Hiseq	Bodo	Bodø, Norway	67.22	15.00
Bodo-08	M edu Europe North	Hiseq	Bodo	Bodø, Norway	67.22	15.00
Bodo-13	M edu Europe North	Novaseq	Bodo	Bodø, Norway	67.22	15.00
Holl-01	M edu Europe South	Hiseq	Holl	Wadden Sea, The Netherlands	53.31	5.42
Holl-02	M edu Europe South	Hiseq	Holl	Wadden Sea, The Netherlands	53.31	5.42
Ouis-02	M edu Europe South	Hiseq	Ouis	Ouistreham, France	49.34	-0.46
Ouis-01	M edu Europe South	Novaseq	Ouis	Ouistreham, France	49.34	-0.46
Revi-02	M edu Europe South	Hiseq	Revi	Réville, France	49.57	-1.23
Revi-01	M edu Europe South	Novaseq	Revi	Réville, France	49.57	-1.23
Brest-18-10	M galloprovincialis Atl Brittany	Novaseq	Brest-18	Pointe Marloux, France	48.36	-4.44
Brest-21-06	M galloprovincialis Atl Brittany	Novaseq	Brest-21	Brest commerce A, France	48.38	-4.47
Brest-31-02	M galloprovincialis Atl Brittany	Hiseq	Brest-31	Penzer, France	48.34	-4.77
Brest-34-11	M galloprovincialis Atl Brittany	Hiseq	Brest-34	Petit Minou, France	48.34	-4.62
Prim-02	M galloprovincialis Atl Brittany	Hiseq	Prim	Primel, France	48.72	-3.82
Prim-06	M galloprovincialis Atl Brittany	Hiseq	Prim	Primel, France	48.72	-3.82
COR-04	M galloprovincialis Atl South	Hiseq	COR	Coruña, Galicia, Spain	43.37	-8.41
COR-05	M galloprovincialis Atl South	Novaseq	COR	Coruña, Galicia, Spain	43.37	-8.41
1642-AF-101	M galloprovincialis Atl South	Mgallo	Gerdol2019_Atl			

Continued on next page

Table S1 – *Continued from previous page*

ind	group	experiment	pop	locality	lat	long
1642-AF-102	M galloprovincialis Atl South	Mgallo	Gerdol2019_Atl			
1642-AF-103	M galloprovincialis Atl South	Mgallo	Gerdol2019_Atl			
1642-AF-104	M galloprovincialis Atl South	Mgallo	Gerdol2019_Atl			
1642-AF-105	M galloprovincialis Atl South	Mgallo	Gerdol2019_Atl			
1642-AF-106	M galloprovincialis Atl South	Mgallo	Gerdol2019_Atl			
PEN-08	M galloprovincialis Atl South	Hiseq	Peniche	Peniche, Portugal	39.37	-9.38
PEN-01	M galloprovincialis Atl South	Novaseq	Peniche	Peniche, Portugal	39.37	-9.38
Bizerte-19	M galloprovincialis Med	Hiseq	Bizerte	Bizerte, Tunisia	37.26	9.86
Bizerte-23	M galloprovincialis Med	Hiseq	Bizerte	Bizerte, Tunisia	37.26	9.86
1642-AF-107	M galloprovincialis Med	Mgallo	Gerdol2019_Med			
1642-AF-108	M galloprovincialis Med	Mgallo	Gerdol2019_Med			
1642-AF-109	M galloprovincialis Med	Mgallo	Gerdol2019_Med			
1642-AF-110	M galloprovincialis Med	Mgallo	Gerdol2019_Med			
1642-AF-111	M galloprovincialis Med	Mgallo	Gerdol2019_Med			
1642-AF-112	M galloprovincialis Med	Mgallo	Gerdol2019_Med			
Her-17	M galloprovincialis Med	Hiseq	Her	Heraklion, Greece	35.34	25.14
Her-19	M galloprovincialis Med	Novaseq	Her	Heraklion, Greece	35.34	25.14
Th16-07	M galloprovincialis Med	Hiseq	Th16	Thau, France	43.41	3.69
Th16-04	M galloprovincialis Med	Novaseq	Th16	Thau, France	43.41	3.69
RET-11	M trossulus	Hiseq	19-Retinskoye	Retinskoye, Russia	69.11	33.39
RET-14	M trossulus	Novaseq	19-Retinskoye	Retinskoye, Russia	69.11	33.39
JapSea-07	M trossulus	Hiseq	JapSea	Vostok Biological Station, Vostok Bay	42.89	132.73
JapSea-01	M trossulus	Hiseq	JapSea	Vostok Biological Station, Vostok Bay	42.89	132.73
StLau-TD-23	M trossulus	Hiseq	StLau-TD	Saint Lawrence, Tadoussac, Canada	48.13	-69.70
StLau-TD-27	M trossulus	Novaseq	StLau-TD	Saint Lawrence, Tadoussac, Canada	48.13	-69.70
BergAR-88	Norway admixed	Hiseq	Berg	Bergen, Norway	60.40	5.31
GA-05	Norway admixed	Novaseq	GA	Gaseid, Alesund, Norway	62.46	6.24
GA-12	Norway admixed	Novaseq	GA	Gaseid, Alesund, Norway	62.46	6.24

Continued on next page

Table S1 – *Continued from previous page*

ind	group	experiment	pop	locality	lat	long
GA-19	Norway admixed	Novaseq	GA	Gaseid, Alesund, Norway	62.46	6.24
GA-29	Norway admixed	Novaseq	GA	Gaseid, Alesund, Norway	62.46	6.24
LOF-01	Norway admixed	Novaseq	LOF	Lofoten, Norway	68.34	13.88
LOF-03	Norway admixed	Novaseq	LOF	Lofoten, Norway	68.34	13.88
LOF-04	Norway admixed	Novaseq	LOF	Lofoten, Norway	68.34	13.88
Q13A-11	Q13A	Hiseq	Q13A	Q13-A, North Sea	52.19	4.14
Q13A-02	Q13A	Hiseq	Q13A	Q13-A, North Sea	52.19	4.14
Q13A-03	Q13A	Hiseq	Q13A	Q13-A, North Sea	52.19	4.14
Q13A-06	Q13A	Hiseq	Q13A	Q13-A, North Sea	52.19	4.14
Q13A-07	Q13A	Hiseq	Q13A	Q13-A, North Sea	52.19	4.14
StMalo-01	StMalo	Novaseq	StMalo	Saint-Malo port, France	48.65	-2.02
StMalo-04	StMalo	Novaseq	StMalo	Saint-Malo port, France	48.65	-2.02
StMalo-07	StMalo	Novaseq	StMalo	Saint-Malo port, France	48.65	-2.02
StMalo-08	StMalo	Novaseq	StMalo	Saint-Malo port, France	48.65	-2.02
StMalo-09	StMalo	Novaseq	StMalo	Saint-Malo port, France	48.65	-2.02
StMalo-10	StMalo	Novaseq	StMalo	Saint-Malo port, France	48.65	-2.02

Chapitre 4

Modèle polygénique et prédiction de la valeur sélective des hybrides

4.1 Contexte et résumé de l'étude

Nous avons pu observer dans les trois chapitres précédents des patrons de résolution d'admixture sur des temps différents qui peuvent paraître à première vue contradictoires. J'entends ici par résolution de l'admixture, le processus selon lequel, dans une population admixée ou introgressée, les variants sont triés par l'action de la sélection naturelle (intrinsèque et extrinsèque), de la stochasticité et de la recombinaison. Le paradoxe apparent réside dans le fait d'observer à l'échelle des espèces du complexe *M. edulis* des comètes d'introgression à la fois locales et globales (chapitre 1), alors que les populations admixées et les croisements présentent peu de distorsions extrêmes, que ce soit sur ~ 80 marqueurs effectivement différenciés (chapitre 2) ou à l'échelle du génome (chapitre 3).

4.1.1 Modèles d'isolement génétique classiques face aux observations

Le modèle le plus simple de sous-dominance (i.e. une valeur sélective réduite des hétérozygotes) dans une population admixée mène inéluctablement à la fixation de l'un des allèles parentaux ou la séparation dans l'espace de ces deux allèles (i.e. la formation d'une zone de tension). Le temps de fixation et la force de la barrière vont dépendre des paramètres démographiques, du coefficient de sélection du génotype hétérozygote et de la migration. Ce comportement bi-stable est souvent interprété de cette façon dans les zones d'hybridation (chapitre 1), bien que parfois confondu avec de l'adaptation locale (Bierne et al. 2011).

De façon proche, un modèle BDMI (Encadré 2) suppose la fixation d'au moins un des allèles parentaux, par la contre sélection des combinaisons d'allèles dérivé/dérivé. L'allèle dérivé qui sera éliminé dépendra de la contribution initiale des parents à l'admixture. Ce principe est illustré dans Schumer et al. (2018b) (Fig. S6), d'où la Figure 4.1 est reprise, et permet dans

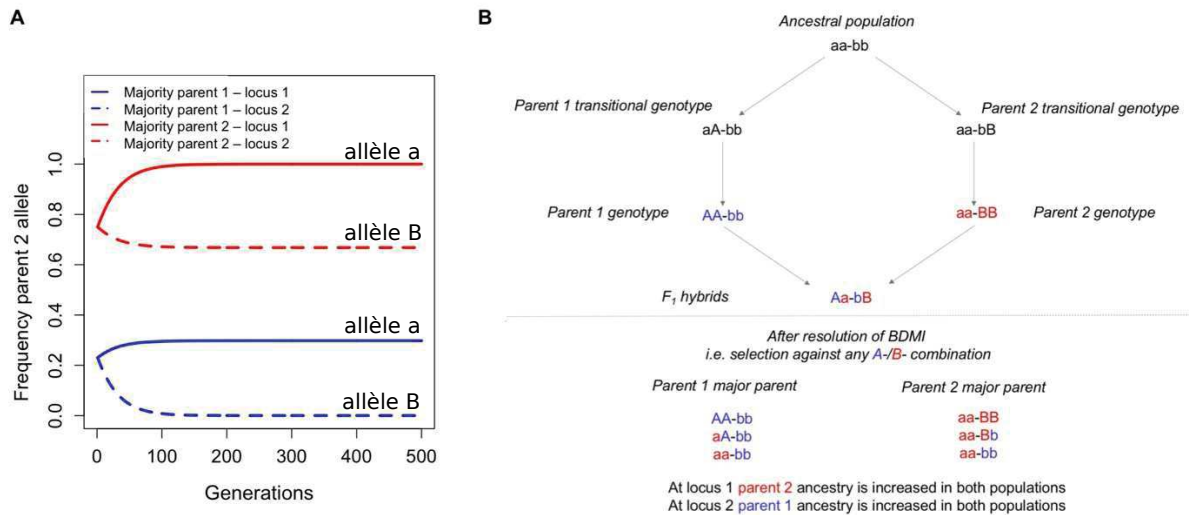


Figure 4.1 : Tirée de Schumer et al. (2018b). (A) Modélisation du changement de fréquence de l'allèle du parent 2 à chaque locus (1 et 2, traits pleins et pointillés respectivement), dans deux situations différentes : bleu - lorsque le parent 1 est majoritaire ; rouge - lorsque le parent 2 est majoritaire. Dans les deux cas de proportion d'admixture, les variations de fréquences sont corrélées à chaque locus. Le modèle utilisé considère ici que la BDMI est neutre, c'est-à-dire que les allèles dérivés n'ont pas d'avantage sélectif par rapport aux ancestraux. (B) Mécanisme menant à cette dynamique. Le modèle BDMI est similaire à l'Encadré 2. Après la purge de l'allèle dérivé provenant du parent minoritaire, l'allèle dérivé du parent majoritaire, bien qu'ayant diminué en fréquence, est toujours présent dans la population admixée.

cet article d'expliquer la corrélation d'ascendance dans une région du génome malgré des proportions d'admixture initiales inversées.

Ces deux types de modèles supposent des effets sélectifs assez forts sur un nombre limité de marqueurs afin de maintenir une barrière aux flux de gènes suffisantes au maintien des espèces. Les variations de fréquences devraient alors être relativement rapides et produire des balayages sélectifs importants. Cependant, les observations empiriques chez les moules, à la fois sur les croisements expérimentaux de F₂s et BCs – chez qui relativement peu de BDMIs classiques ont été retrouvées –, et dans de multiples événements d'admixture laissent à penser que peu de régions génomiques constituent une barrière forte aux flux de gènes et que l'admixture produit une multitude de combinaisons viables sans dépression hybride. Enfin, dans des conditions naturelles avec de la recombinaison et de la migration, les modèles classiques de BDMI ne sont pas effectifs pour le maintien des barrières d'espèces (Lindtke et Buerkle 2015).

4.1.2 Passage à une vision polygénique

Afin de réconcilier les observations des patrons d'admixture dans les croisements et les populations admixées, il est nécessaire de considérer des modèles polygéniques permettant de prendre en compte des relations épistatiques à large échelle. De manière générale, les études d'associations génomiques (GWAS) ont permis relativement récemment de réaliser que beau-

coup de traits étaient hautement polygéniques (Boyle et al. 2017 ; Sachdeva et Barton 2018a), via notamment l'augmentation des puissances de détection. S'ajoute aussi à la réalisation que des modèles polygéniques peuvent mieux prédire des données classiques d'incompatibilité (e.g. un seuil polygénique de la stérilité hybride chez les drosophiles ; Morán et Fontdevila 2014), également la considération que des polymorphismes peuvent contribuer à l'isolement reproductif (Cutter 2012).

La prédiction du maintien d'un isolement reproductif dans des contacts secondaires et d'une hétérogénéité du flux de gènes est liée à la présence d'interactions intra-génomiques (Lindtke et Buerkle 2015). Ces interactions intra-génomiques peuvent être modélisées de manière directe par des réseaux de gènes co-adaptés ou des voies de régulation génétique décrits à partir d'observations empiriques (Lindtke et Buerkle 2015), mais aussi via des propriétés émergentes de co-adaptation et de forme de paysage de fitness dans le modèle géométrique de Fisher (Tenaillon et al. 2007 ; Chevin et al. 2014 ; Fraïsse et al. 2016b).

4.1.3 Résumé


Dans cette étude, nous utilisons le cadre théorique du modèle géométrique de Fisher afin d'explorer la valeur sélective des premières générations d'hybridation. En utilisant une approximation permettant de réduire la dimensionnalité du modèle et d'augmenter sa flexibilité, nous montrons qu'il permet d'obtenir des prédictions classiques de l'étude de la spéciation. De plus, la flexibilité du modèle permet une interpolation entre différentes approches théoriques initialement développées pour prédire des cas extrêmes observés dans des croisements. Ce modèle permet une première exploration à l'échelle génomique des patrons d'incompatibilité dans des croisements, sans avoir recours à des « gènes de spéciation » à effets forts, et la prédiction de nouvelles observations chez les moules : une distorsion vers une augmentation globale de l'hétérozygotie dans un croisement F2.

4.2 Article 4

Cet article a été publié dans le journal *Evolution Letters*.

A. Simon, N. Bierne et J. J. Welch (2018). « Coadapted Genomes and Selection on Hybrids : Fisher's Geometric Model Explains a Variety of Empirical Patterns ». *Evolution Letters* 2.5, p. 472-498. DOI : 10.1002/evl3.66

Coadapted genomes and selection on hybrids: Fisher's geometric model explains a variety of empirical patterns

Alexis Simon,^{1,2,3}  Nicolas Bierne,^{1,2} and John J. Welch²

¹Institut des Sciences de l'Évolution UMR5554, Université de Montpellier, CNRS-IRD-EPHE-UM, France

²Department of Genetics, University of Cambridge, Downing St. Cambridge, CB23EH, United Kingdom

³E-mail: alexis.simon@umontpellier.fr

Received March 5, 2018

Accepted June 6, 2018

Natural selection plays a variety of roles in hybridization, speciation, and admixture. Most research has focused on two extreme cases: crosses between closely related inbred lines, where hybrids are fitter than their parents, or crosses between effectively isolated species, where hybrids suffer severe breakdown. But many natural populations must fall into intermediate regimes, with multiple types of gene interaction, and these are more difficult to study. Here, we develop a simple fitness landscape model, and show that it naturally interpolates between previous modeling approaches, which were designed for the extreme cases, and invoke either mildly deleterious recessives, or discrete hybrid incompatibilities. Our model yields several new predictions, which we test with genomic data from *Mytilus* mussels, and published data from plants (*Zea*, *Populus*, and *Senecio*) and animals (*Mus*, *Teleogryllus*, and *Drosophila*). The predictions are generally supported, and the model explains a number of surprising empirical patterns. Our approach enables novel and complementary uses of genome-wide datasets, which do not depend on identifying outlier loci, or "speciation genes" with anomalous effects. Given its simplicity and flexibility, and its predictive successes with a wide range of data, the approach should be readily extendable to other outstanding questions in the study of hybridization.

KEY WORDS: Dobzhansky–Muller incompatibilities, Haldane's Rule, heterozygosity, inbreeding, speciation genetics, sterility.

Impact summary

When individuals of different populations mate, the offspring will carry new combinations of alleles. Sometimes the new combinations bring fitness benefits (heterosis). This is often true, for example, when the parental lines are closely related and highly inbred: a fact that can be exploited in artificial breeding programs. Sometimes, the hybrids are much less fit than their parents (hybrid breakdown), suggesting that the populations may be distinct species. These different outcomes depend on the ways in which the alleles interact, and so comparing the outcomes of different types of hybridization can tell us a lot about gene interactions. We developed a general mathematical multigenic model that makes simple predictions for the fitness of hybrids of any type. We show that our model can account for a large number of empirical patterns, including some that were not well explained by alternative theories,

developed for specific cases. We tested our predictions with new data from mussels, and published data from plants and animals, and obtained a good fit. Our framework suggests a new and complementary approach to analyzing genomic data from hybrids, which does not rely on identifying small regions of the genome with anomalous effects.

Hybridization and admixture lead to natural selection acting on alleles in different genetic backgrounds. Most classical studies of hybridization can be placed into one of two classes. The first, involves crosses between closely related inbred lines, where there is no coadaptation between the deleterious alleles that differentiate the lines, such that most hybrids are fitter than their parents. Wright's single-locus theory of inbreeding was developed to interpret data of this kind (Wright 1922, 1977; Crow 1952; Hallauer et al. 2010). The second, involves crosses between

effectively isolated species, where viable and fertile hybrids are very rare. Data of this kind are often analyzed by focusing on a small number of “speciation genes,” and interpreted using models of genetic incompatibilities (Dobzhansky 1937; Coyne and Orr 1989; Orr 1995; Gavrillets 2004; Welch 2004; Kalirad and Azevedo 2017).

The differences between these types of hybridization are clear, but it is equally clear that they are extremes of a continuum. Furthermore, the intermediate stages of this continuum are of particular interest, because they include phenomena such as incipient speciation, and occasional introgression between partially isolated populations (Waser 1993; Rosas et al. 2010; Mendez et al. 2012; Fraïsse et al. 2016a; Duranton et al. 2017). However, it can be difficult to model natural selection in this intermediate regime, not least because models require a large number of parameters when they include epistatic effects between many loci. The empirical study of hybrid genotypes in this regime is also difficult. The analysis of lab crosses often focuses on segregation distortions of large effect, and pairwise incompatibilities (Coyne and Orr 2004; Abbott et al. 2013). This QTL-mapping framework can miss small effect mutations (Noor et al. 2001; Rockman 2012), which are difficult to identify individually, but whose cumulative effect can be substantial (Boyle et al. 2017).

One promising approach is to use Fisher’s geometric model, which assigns fitness values to genotypes using a model of optimizing selection on quantitative traits (Fisher 1930; Orr 1998; Welch and Waxman 2003; Martin and Lenormand 2006). The tools of quantitative genetics have often been used to study hybridization (e.g., Melchinger 1987; Lynch 1991; Demuth and Wade 2005; Fitzpatrick 2008), but Fisher’s model is fully additive at the level of phenotype, and the “traits” need not correspond in any simple way to standard quantitative traits (Rosas et al. 2010; Martin 2014; Schiffman and Ralph 2017). Instead, the goal is to generate a rugged fitness landscape, which includes a wide variety of mutational effect sizes and epistatic interactions on fitness, with a minimum of free parameters (Orr 1998; Blanquart et al. 2014; Barton 2017; Hwang et al. 2017).

Here, we build on previous studies (Mani and Clarke 1990; Barton 2001; Rosas et al. 2010; Chevin et al. 2014; Fraïsse et al. 2016b; Schiffman and Ralph 2017), and use Fisher’s geometric model to study hybridization. We develop a simple Brownian bridge approximation, and show that it can naturally interpolate between previous modeling approaches (Wright 1922, 1977; Orr 1995; Gavrillets 2004), which are appropriate for the two extreme types of hybridization. We then show how the model can account for surprising empirical patterns that have been observed in both regimes (Wright 1977; Moehring 2011; Moran et al. 2017). Finally, we show that the model yields novel predictions for selection on heterozygosity, and test these predictions with a wide range of new and existing datasets (Table 1).

Models and Results

THE MODELS

Notation and basics

We will consider hybrids between two diploid populations, labeled P1 and P2, each of which is genetically uniform, but which differ from each other by d substitutions. Considering all possible combinations of the two homozygotes and the heterozygote, the populations could generate 3^d distinct hybrid genotypes, and each might differ in their overall fitness, or in some component of fitness, such as fertility or viability. Below, we will focus on rank order differences between different classes of hybrid (e.g., high vs. low heterozygosity, males vs. females, F1 vs. F2 etc.). As such, following Turelli and Orr (2000), we describe hybrids using a “breakdown score,” S , which is larger for hybrids that have lower values of the fitness component of interest. Breakdown might relate to fitness via,

$$\ln w \propto -S^{\alpha/2} \quad (1)$$

in which case, the parameter α adjusts the rate at which fitness declines with breakdown. This is related to the overall levels of fitness dominance and epistasis (Hinze and Lamkey 2003; Tenaillon et al. 2007; Fraïsse et al. 2016b), and so these can vary independently of other results. We now define the key quantity f , as the expected value of S for a given class of hybrid, scaled by the expected value for an unfit reference class.

$$f \equiv \frac{E(S)}{E(S_{\dagger})} \quad (2)$$

Here, $E(S_{\dagger})$, is the expected breakdown score for the class of hybrid with the lowest fitness, on the condition that the parental genotypes are optimally fit. In this case, f can vary between zero, for the best possible class of hybrid, and one, for the worst possible class. We will also consider the case where the parental types are themselves suboptimal, with their own level of “breakdown,” denoted f_{P1} and f_{P2} . In this case, when $f_{P1}, f_{P2} > 0$, then the value of f for the hybrids can be greater than one (see below).

To define classes of hybrid, we also follow Turelli and Orr (2000). We pay particular attention to interpopulation heterozygosity, and define p_{12} as the proportion of the divergent sites that carry one allele from each of the parental types. We also define p_1 and p_2 as the proportion of divergent sites that carry only alleles originating from P1 or P2, respectively. Since $p_1 + p_2 + p_{12} = 1$, it is convenient to introduce the hybrid index, h , which we define as the total proportion of divergent sites originating from P2 (Barton and Gale 1993).

$$h \equiv p_2 + \frac{1}{2}p_{12} \quad (3)$$

This quantity is closely related to measures of ancestry (e.g., Falush et al. 2003; Fitzpatrick 2012; Christe et al. 2016),

Table 1. Datasets.

Hybridization	<i>N</i>	Sex	#Markers	Cross	<i>g_X</i>	Fitness measure	Reference
<i>Zea mays</i> inbred lines	–	♀	–	Various	–	Excess yield	See Fig. 3
<i>Mytilus edulis/galloprovincialis</i>	132	♀/♂	43	F2	–	–	This study
	144	♀/♂	43	BC1	–	–	
<i>Drosophila pseudoobscura/persimilis</i>	1141/1036	♀	2 X; 11 A	BC1	0.37	Motile sperm: present/absent	Noor et al. (2001)
<i>Drosophila sechellia/simulans</i>	200/200	♀	8 X; 31 A	BC1	0.17	Sperm quantity: 3-pt. scale	Macdonald and Goldstein (1999)
<i>Drosophila santomea/yakuba</i>	550/549	♀	10 X; 22 A	BC1	0.17	Motile sperm: 9-pt. scale	Moehring et al. (2006a, 2006b)
<i>Teleogryllus oceanicus/commodus</i>	248	♂	–	Various	0.30	Egg and offspring number	Moran et al. (2017)
<i>Populus alba/tremula</i>	137	♀♂	~12,000	WH	–	Survival after 4 years	Christe et al. (2016)
<i>Senecio aethnensis/chrysanthemifolius</i>	64	♀	966	F2	–	Necrotic/Healthy	Chapman et al. (2016)
<i>Mus musculus musculus/domesticus</i>	185	♀	14,220	WH	0.039	Testes weight	Turner and Harr (2014)
	305	♀	202 (16 X; 182 A)	F2	0.039	Prop. abnormal sperm	White et al. (2011)

N, The number of individual hybrids, divided by backcross direction where appropriate; #Markers, The number of genetic markers used to estimate p_{12} and h , sometimes divided into X-linked and Autosomal. BC1, First backcross; WH, Wild hybrids. g_X , weight given to X-linked markers in the calculation of overall genome composition, calculated from the length of annotated coding sequence.

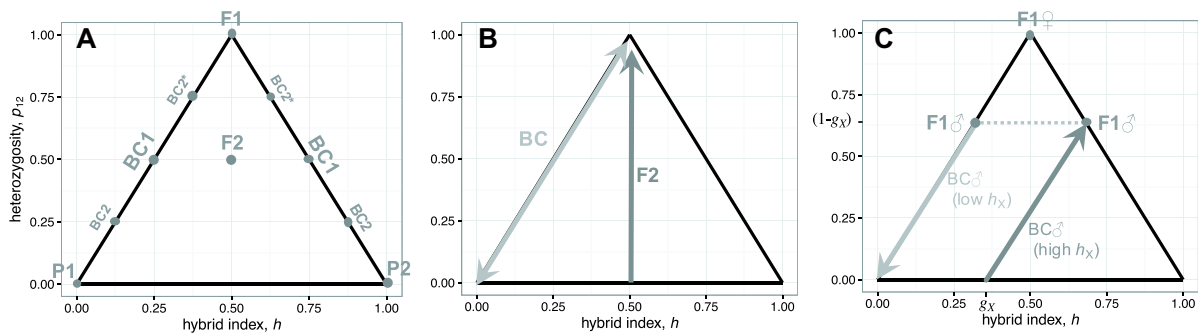


Figure 1. Diploid hybrid genomes can be represented as points in a triangular space, indicating their hybrid index, h (the total proportion of divergent alleles that originate from parental line P2), and their interpopulation heterozygosity, p_{12} (the proportion of divergent sites that carry one allele from each parental line). Panel (A) shows how the standard crosses are placed in this space, with biparental inheritance. The two parental lines P1 and P2 are at the lower corners. The initial F1 cross (P1 × P2) will be heterozygous at all divergent sites, and so be found at the top corner. Individuals from later crosses will vary due to segregation and recombination, but the F2 (F1 × F1) will tend to be found toward the center, while backcrosses (such as BC1 = F1 × P1 or F1 × P2) will be found on the upper edges (see Fig. 3 for more details of these backcrosses). Panel (B) illustrates the selection on heterozygosity predicted by Fisher’s geometric model, when the parental lines are well adapted (eq. 9). In backcrosses, heterozygosity will be under diversifying selection, favoring both extreme values. By contrast, in the F2, we predict directional selection toward higher heterozygosity. Panel (C) illustrates some complications introduced by heteromorphic sex chromosomes (see eq. 16 and Appendix 4). With XO sex determination, male F1 carry no heterozygosity on the X, which will tend to reduce their fitness, consistent with Haldane’s Rule. For male backcrosses (F1 × P1), the selection acting on (autosomal) heterozygosity, will depend on the alleles carried on the X. When the X carries mostly P2 alleles, fitter individuals will be more heterozygous (darker gray arrow). When the X carries mostly P1 alleles, the fittest individuals will carry no heterozygosity (lighter gray arrow). g_x is the proportion of the divergent sites found on the X, and is set at 37%, as we have estimated for *Drosophila pseudoobscura*.

although it considers only divergent sites. We can now describe each individual genotype via its heterozygosity, p_{12} , and hybrid index, h . This means that all hybrids can be represented as points in a triangular space, as shown in Figure 1 A (Gompert and Buerkle 2010; Fitzpatrick 2012). Our goal in this article is to represent this space as a kind of fitness landscape, and show how f can vary with p_{12} and h .

Fisher’s geometric model

Fisher’s model is defined by n quantitative traits under selection toward an intermediate optimum (Fisher 1930). We will assume that fitness is always measured in a fixed environment, but we make no assumptions about how this optimum might have moved during the parental divergence (see Appendix 1). If the selection function is multivariate normal, including correlated selection, then we can rotate the axes and scale the trait values, to specify n new traits that are under independent selection of different strengths (Waxman and Welch 2005; Martin and Lenormand 2006; Martin 2014). Examples with $n = 2$ traits are shown in Figure 2. We now define the breakdown score of a phenotype as

$$S \equiv \sum_{i=1}^n \lambda_i z_i^2 \tag{4}$$

where, for trait i , z_i is its deviation from the optimum and λ_i is the strength of selection. By assumption, all mutational changes

act additively on each trait, but their effects on fitness can vary with the phenotype of the individual in which they appear, and this yields fitness dominance and epistasis (Appendix 1).

The d substitutions that differentiate P1 and P2 can be considered as a chain of vectors, which connect the two parental phenotypes (Fig. S1). While the sizes and directions of these vectors will generally be unknown, in Appendix 1, we show that the chain can be approximated as a constrained random walk, or Brownian bridge (Revuz and Yor 1999, Ch. 1). This approximation relies on the fact that hybrid genomes contain the fixed alleles in effectively random combinations, and it gives accurate results for a wide range of assumptions about the divergence process (Figs. S2–S3), including adaptation of the parental lines to different environments (Appendix 1; Fig. S4).

Under the Brownian bridge approximation, the quantity $E(S_{\dagger})$, that appears in equation (2), is given by

$$E(S_{\dagger}) = d \sum_{i=1}^n \lambda_i v_i \tag{5}$$

where v_i is the expected variance contributed to trait i by a fixed mutation in heterozygous state and we have assumed that any maladaptation has accrued independently in the two parental lines (see Appendix 1 and Table S1). The key quantity f is given by

$$f = f_{P1} + \beta_1 h(1 - \beta_2 h) - p_{12} \tag{6}$$

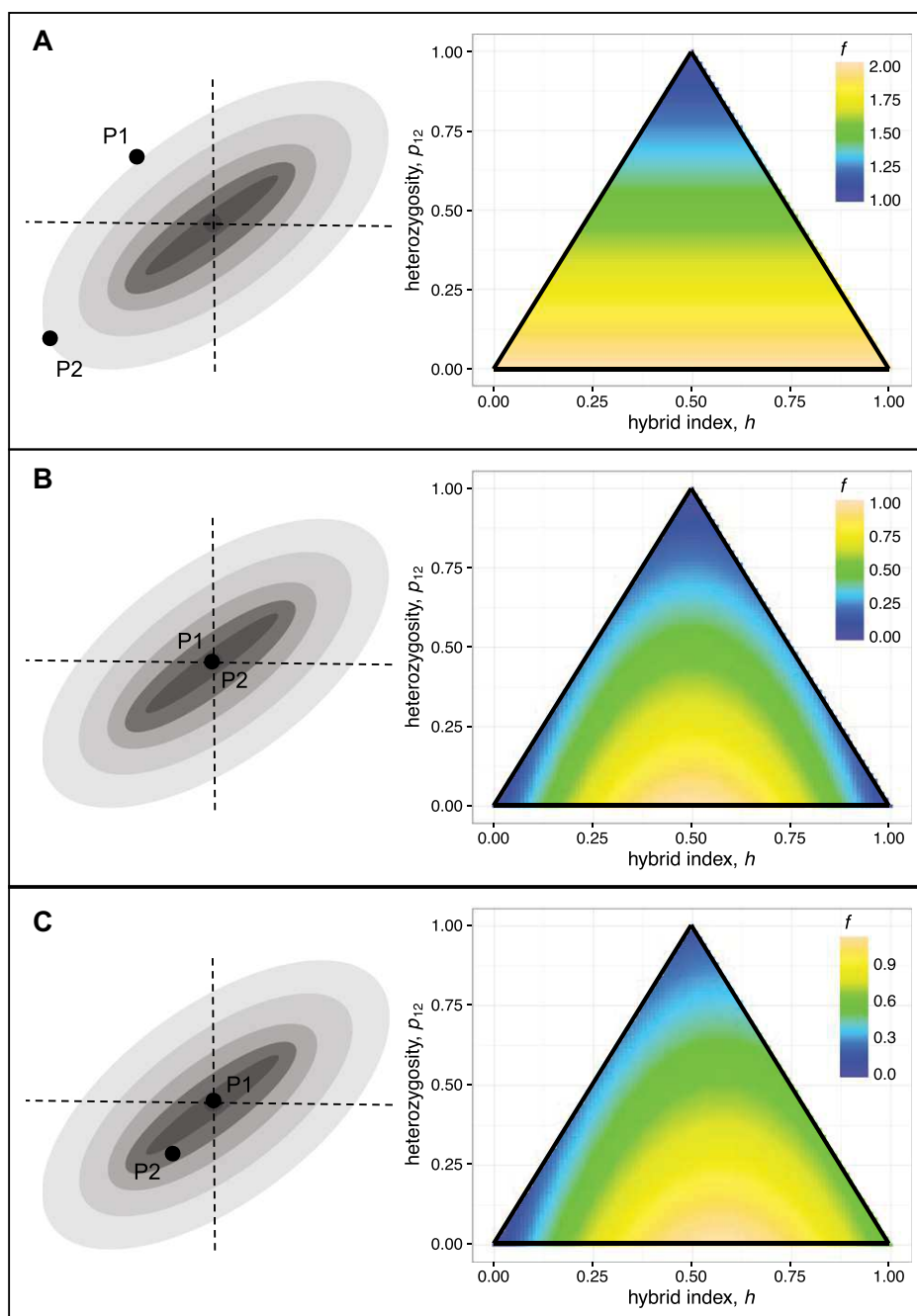


Figure 2. Fisher's geometric model generates a flexible series of fitness landscapes for hybrid genotypes. The left-hand panels show a schematic representation of the model, with $n = 2$ traits, each under optimizing selection of differing strengths. (The orientation of the axes shows that the model allows for correlated selection, although this is ignored in the text, by rotating the axes). The right-hand panels show how the expected breakdown score (eqs. 2 and 6), varies for hybrids of different types. Panels (A)–(C) show different assumptions about levels of parental maladaptation. Panel (A) shows a scenario where all of the parental divergence is maladaptive, with no tendency for their alleles to be coadapted. In this case, hybrid fitness increases with their heterozygosity, as predicted by Wright's single-locus theory of inbreeding (eq. 8; Wright 1922, 1977). Panel (B) shows a scenario where the parental lines are optimal (or, at least, very well adapted compared to the worst class of hybrid that can be formed between them). In this case, the hybrid index is under symmetrical diversifying selection, and the form of selection on heterozygosity will vary for different cross types (eqs. 9, 13, and 15). This landscape can also be generated by a general model of genetic incompatibilities (see Appendix 2). Panel (C) shows a situation where only one of the parental lines is maladapted (eq. 10 with $f_{p_2} = 0.5$).

where

$$\begin{aligned}\beta_1 &\equiv 4 - 2f_{P1} \\ \beta_2 &\equiv \frac{4 - f_{P1} - f_{P2}}{4 - 2f_{P1}}\end{aligned}\quad (7)$$

(see Appendix 1 for full details). Two features of equation (6) are immediately notable. First, it does not depend on any of the model parameters. For example, the number of traits, n , could affect the accuracy of our approximation (since S will tend to approach normality as n increases); but it does not appear in equation (6), which depends solely on p_{12} , h and the levels of parental maladaptation. Second, for a given value of h , the fitness of hybrids will tend to increase with their heterozygosity, p_{12} . This prediction agrees with the widespread finding of heterosis (Crow 1952; Table S1 of Fraïsse et al. 2016b). Indeed, by rearranging equation (6) it is clear that hybrids will tend to be fitter than parental line P1 (such that $f < f_{P1}$) on the condition that $p_{12} > \beta_1 h(1 - \beta_2 h)$.

It is also useful to consider equation (6) in a few special cases. First, let us assume that all of the divergence between the parents is maladaptive, without any tendency for coadaptation between their alleles. In this case, the parental phenotypes can be treated as unconstrained random walks away from the optimum. Each fixed mutation, in homozygous state, will contribute an expected $4v_i$ to the variance on each trait, and so, from equations (4) and (5), we have $E(f_{P1}) = E(f_{P2}) = 2$ (see also Appendix 1). With this high level of parental maladaptation, the expected breakdown score of the hybrids is

$$f = 2 - p_{12}, \quad f_{P1}, f_{P2} = 2 \quad (8)$$

The fitness landscape implied by equation (8) is illustrated in Figure 2 A. With these extreme levels of parental maladaptation, hybrid breakdown depends only on the heterozygosity. Indeed, this result is closely related to Wright's (1922) single-locus theory of inbreeding, which was developed to analyze crosses between closely related inbred lines, where all divergence between the parental lines comprises deleterious alleles. Wright's theory therefore appears as a special case of Fisher's geometric model, when parental alleles show no coadaptation (see below).

Now let us consider the other extreme case, where the parental alleles are completely coadapted, such that P1 and P2 both have optimal fitness, but realized by different combinations of alleles. In this case, we find

$$f = 4h(1 - h) - p_{12}, \quad f_{P1}, f_{P2} = 0 \quad (9)$$

The fitness landscape implied by equation (9) is illustrated in Figure 2 B. With well-adapted parents, the hybrid index is under symmetrical diversifying selection.

As with equation (8), equation (9) can also be derived via an alternative route, using a model of speciation genetics. In particular, we show in Appendix 2 that equation (9) can be obtained from a general model of "Dobzhansky–Muller incompatibilities," each involving a small number of loci (Orr 1995; Turelli and Orr 2000; Gavrillets 2004; Welch 2004; Fraïsse et al. 2016b). The agreement between the two models depends on a particular set of assumptions about the dominance of incompatibilities, namely (i) partial recessivity on average, and (ii) greater reduction in fitness when they contain homozygous alleles from both parental lines. However, we show in Appendix 2 that these assumptions are biologically realistic, and implied by a number of well-established empirical patterns (Turelli and Orr 2000). The result is that equation (9) can be interpreted as a model of genetic incompatibilities, but without the large number of free parameters that these models can require.

Equation (9) is expected to apply to many cases of wild hybridization, because it should provide a good approximation even if the parental populations are not truly optimal. The only requirement is that they be much better adapted than the worst possible class of hybrid, such that $f_{P1}, f_{P2} \ll 1$.

Nevertheless, the general form of equation (6) also allows us to explore intermediate cases. For example, if only P2 is maladapted, then we find

$$f = 4h \left(1 - \left(1 - \frac{1}{4} f_{P2} \right) h \right) - p_{12}, \quad f_{P1} = 0 \quad (10)$$

In this case, as illustrated in Figure 2 C, selection on hybrid index is skewed toward the fitter parent. Below, we will show how all three of these special cases (Fig. 2 A–C) can be used in data analysis.

TESTING THE PREDICTIONS WITH BIPARENTAL INHERITANCE

Fitness differences between crosses

The simplest predictions from equation (6) assume standard biparental inheritance at all loci. In this case, the standard cross types can be easily located on our fitness landscapes. This is shown in Figure 1 A.

With well-adapted parental lines (Fig. 2 B) hybrids of high fitness are expected only for the initial F1 cross (P1 × P2) and breakdown is predicted for later crosses, such as the first backcross (BC1: F1 × P1 or F1 × P2) or the F2 (F1 × F1). This pattern of F1 hybrid vigor (high F1 fitness) followed by breakdown in later crosses, has widespread empirical support (see references in Table S1 of Fraïsse et al. 2016b and Rosas et al. 2010).

With highly maladapted parents, by contrast, hybrids of all types can be fitter than their parents (see Fig. 2 A). Plentiful data of this kind come from highly inbred lines of *Zea mays* (Neal 1935; Wright 1977; Melchinger 1987; Hinze and Lamkey

2003; Hallauer et al. 2010). To analyze these data, a widely used proxy for fitness is the excess yield of a cross (i.e., its increase in yield relative to the parental lines), scaled by the excess yield of the F1. From equations (1)–(2), and using a Taylor expansion, we find

$$\frac{w - w_P}{w_{F1} - w_P} \approx \frac{f_P^{\alpha/2} - f^{\alpha/2}}{f_P^{\alpha/2} - f_{F1}^{\alpha/2}} \quad (11)$$

$$= p_{12}, \quad f_P = 2, \alpha = 2 \quad (12)$$

where the subscript P refers to both parental lines, which are assumed to have similar yields. For later crosses, these values will vary between individuals within a cross, due to segregation and recombination, but in this section we ignore this variation, and assume that p_{12} and h take their expected values for a given cross type (Fig. 1 A). A fuller treatment is outlined in Appendix 3.

Equation (12) confirms that Fisher’s model can produce Wright’s (1922) single-locus predictions for inbreeding, but only when all divergence between the parental lines is deleterious ($f_P = 2$), and when increases in breakdown score act independently on log fitness ($\alpha = 2$). These single-locus predictions have strong support in *Zea mays* (Neal 1935; Wright 1977; Melchinger 1987; Hinze and Lamkey 2003; Hallauer et al. 2010). For example, as shown in Figure 3 A, the excess yield of the F2 is often around 50%, which is equal to its expected heterozygosity (Wright 1977; Hallauer et al. 2010). It is also notable that the two outlying points (from Shehata and Dhawan 1975), are variety crosses, and not inbred lines in the strict sense.

Despite this predictive success, Wright (1977) noted a pattern that single-locus theory could not explain. In Wright’s words: “the most consistent deviation from expectation [...] is the low yield of F2 in comparison with the first backcrosses” (Wright 1977, p. 39). Because $E(p_{12}) = \frac{1}{2}$ for both crosses, this cannot be explained without epistasis, or coadaptation between the alleles in the parental lines. In fact, the pattern is predicted by Fisher’s model, when there is a small amount of coadaptation, such that $1 < f_P < 2$. This yields a fitness surface with a small amount of curvature, which is intermediate between those shown in Figure 2 A and 2 B.

Figure 3 B plots the four relevant datasets collated by Wright, and compares the results to predictions from equations (6) and (11), with $f_P = 1.5$ and $\alpha = 3$. The model predicts the roughly linear increase in yield with mean heterozygosity, as with single locus theory, but also predicts the consistent difference between BC1 and the F2.

Selection on heterozygosity within crosses

In the previous section, we ignored between-individual variation in heterozygosity within a given cross type. In this section, we

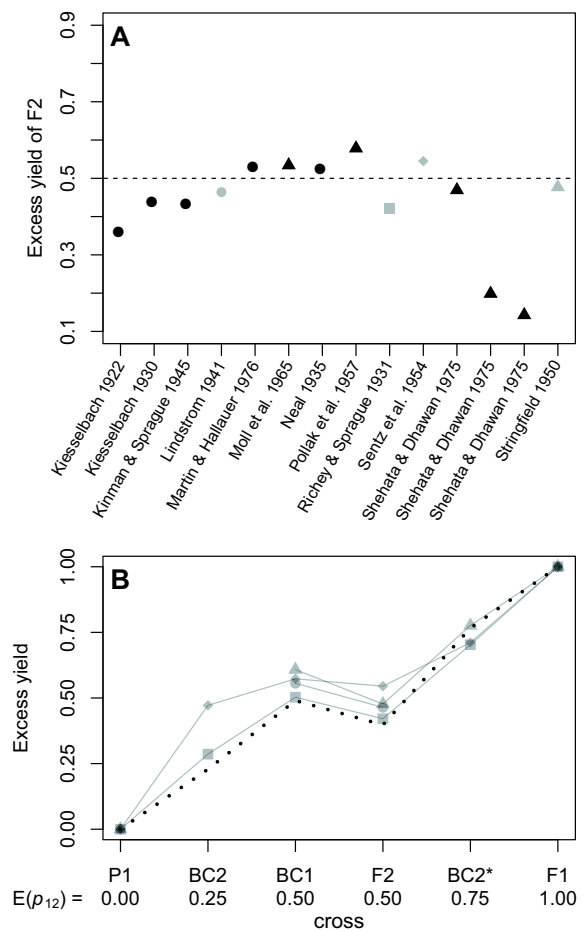


Figure 3. Data on hybrid vigor (i.e., increased yield), from crosses of inbred *Zea mays*. The original data were collated by Wright (1977; see his Table 2.23), and Hallauer et al. (2010; see their Table 9.13), including only data from single crosses, where there was hybrid vigor in the F2, and yield measured in quintals per hectare. Panel (A) plots the excess yield of the F2 (eq. 11). Results are shown for variety crosses (black triangles), as well as crosses of inbred lines in the strict sense (all other points). The dashed line shows the prediction of 0.5 from single-locus theory (eq. 12). Panel (B) shows the four datasets collated by Wright (1977), which allow us to compare the F2 and various backcrosses. These crosses, chosen to yield different levels of heterozygosity, are the parental type (P1), the second backcross (BC2 = (F1 × P1) × P1); the first backcross (BC1 = F1 × P1), the F2 (F1 × F1), second backcross to the other parent (BC2* = (F1 × P1) × P2), and the F1 (P1 × P2) (The data of Stringfield (1950), shown as gray triangles, replace BC2* with an F2 between two distinct F1, involving three distinct strains, but the predictions are unchanged). The gray symbols for the four datasets correspond to those used in panel (A). The dotted line in panel (B) shows predictions from Fisher’s model, assuming that the between-strain divergence contains limited coadaptation. The prediction uses equations (13), (14), and (11), with $f_P = 1.5$, and $\alpha = 3$. The model predicts both the roughly linear increase in vigor with heterozygosity, and the systematic difference between BC1 and F2.

show how natural selection is predicted to act on this heterozygosity.

First, let us consider the F2. In this case, we have $4h(1 - h) \approx 1$ with relatively little variation between individuals (see Appendix 3 for details). Therefore, if both parents have similar levels of maladaptation, equation (6) is well approximated by

$$f_{F2} \approx 1 + \frac{1}{2}f_P - p_{12} \quad (13)$$

The prediction is that p_{12} will be under directional selection in the F2, favoring individuals with higher heterozygosity. This is illustrated in Figure 1 B.

Now let us consider a backcross: F1 \times P1. In this case, all of the homozygous alleles must come from P1, such that $4h(1 - h) = p_{12}(2 - p_{12})$, and so equation (6) becomes

$$f_{BC} = f_P + (1 - f_P)p_{12} - \left(1 - \frac{1}{2}f_P\right)p_{12}^2 \quad (14)$$

$$= p_{12}(1 - p_{12}), \quad f_P = 0 \quad (15)$$

So backcrosses will tend to be least fit when they have intermediate levels of heterozygosity. When parents are well adapted, heterozygosity is under symmetrical disruptive selection, favoring heterozygosities that are either higher or lower than $p_{12} = 0.5$ (eq. 15). This is illustrated in Figure 1 B.

To test these predictions, we used a new set of genetic data from hybrids of the mussel species: *Mytilus edulis* and *Mytilus galloprovincialis* (Bierne et al. 2002, 2006). These species fall at the high end of the continuum of divergence during which introgression persists among incipient species (Roux et al. 2016). We used experimentally bred F2 and BC1, with selection imposed implicitly, by the method of fertilization, and by our genotyping only individuals who survived to reproductive age (Bierne et al. 2002, 2006; see Methods and Fig. S6 for full details).

To estimate heterozygosity in each hybrid individual, we used the 43 markers that were heterozygous in all of the F1 hybrids used to make the subsequent crosses (see Fig. S6). We then asked whether the distribution of p_{12} values in recombinant hybrids was symmetrically distributed around its Mendelian expectation of $p_{12} = 0.5$, or whether it was upwardly biased, as would be expected if directional selection were acting on heterozygosity. As shown in the first column of Table 2, Wilcoxon tests found that heterozygosities in surviving hybrids were significantly higher than expected, in both the F2 and backcross. These results may have been biased by the inclusion of individuals with missing data, because they showed higher heterozygosity (see Table S2). We therefore repeated the test with these individuals excluded. As shown in the second column of Table 2, results were little changed, although the bias toward high heterozygosities was now weaker in the backcross.

Interpreting these results is complicated by the ongoing gene flow between *M. edulis* and *M. galloprovincialis* in nature (Fraïsse et al. 2016a). To test for this, we genotyped 129 pure-species individuals, and repeated our analyses with a subset of 33 markers that were strongly differentiated between the pure species (see Methods, Fig. S6 and Table S3 for details). With these markers, there was evidence of elevated heterozygosity in the F2, but not the backcross (Table 2 third column). We also noticed that many of our backcross hybrids, though backcrossed to *M. galloprovincialis*, carried homozygous alleles that were typical of *M. edulis*. We therefore repeated our analysis after excluding these “F2-like” backcrosses. Results, shown in the fourth column of Table 2, showed that the reduced BC dataset showed no tendency for elevated heterozygosity. However, the bias toward higher heterozygosities remained in the F2, even when we subsampled to equalize the sample sizes.

Despite the problems of interpretation due to introgression and shared variants, the results support the prediction of equations (13)–(15): that directional selection on heterozygosity should act in the F2, but weakly or not at all in the backcross.

PREDICTIONS OF FISHER'S GEOMETRIC MODEL WITH SEX-SPECIFIC INHERITANCE

Results above assumed exclusively biparental inheritance. But the predictions of Fisher's model are easily extended to include heteromorphic sex chromosomes, or loci with strictly uniparental inheritance, such as organelles (Coyne and Orr 1989; Turelli and Orr 2000; Turelli and Moyle 2007; Fraïsse et al. 2016b). In this section, we introduce the approach, and demonstrate its flexibility, while the full derivations are collected in Appendix 4. The basic method of introducing sex-specific inheritance is to write p_{12} and h as weighted sums of contributions from different types of locus. For example, to focus on the contribution of the X chromosome versus the autosomes, we can write

$$p_{12} = g_X p_{12,X} + g_A p_{12,A} \\ h = g_X h_X + g_A h_A \quad (16)$$

where the subscripts denote the chromosome type (so that $p_{12,A}$ is the proportion of divergent sites on the autosomes that are heterozygous), and g_X and g_A are weightings, which sum to one (Turelli and Orr 2000).

We can now predict differences in hybrids of different sexes. For example, previous authors have shown that Fisher's model predicts Haldane's Rule: the tendency of sex-specific breakdown to appear in the heterogametic sex (Haldane 1922; Turelli and Orr 2000; Barton 2001; Fraïsse et al. 2016b; Schiffman and Ralph 2017; see Fig. 1 C). Using equations (16) and (6), this basic insight can easily be extended to give formal conditions for

Table 2. Tests for selection on heterozygosity in F2 and Backcrosses of *Mytilus* mussels.

Markers:	43	43	33	33
Dataset:	All	No missing data	No missing data	Subsampled
Cross	\hat{p}_{12} (N) P -value	\hat{p}_{12} (N) P -value	\hat{p}_{12} (N) P -value	\hat{p}_{12} (N) P -value
F2	0.57 (132) $1.5 \times 10^{-6***}$	0.56 (88) $6.4 \times 10^{-4***}$	0.55 (91) 0.0033**	0.56 (56) 0.0020**
BC	0.57 (144) $1.3 \times 10^{-5***}$	0.53 (94) 0.0282*	0.53 (105) 0.0569	0.52 (56) 0.5815

\hat{p}_{12} , the estimated median heterozygosity; N , the number of hybrid individuals sampled; P -value, result of a Wilcoxon test of the null hypothesis median $p_{12} = 0.5$ (* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$). F2, random mating of F1 between *M. galloprovincialis* and *M. edulis*; BC, Backcross of the F1 to *M. galloprovincialis*. No missing data, all individuals with missing data for any of the markers were excluded; Subsampled, for the BC, any individual carrying a marker that was homozygous for the major allele carried by wild *M. edulis* populations was excluded; for the F2, we downsampled by sequencing order to equalize sample sizes.

Haldane's Rule, with different levels of parental maladaptation, and uniparental inheritance (see Appendix 4).

Sex chromosomes also complicate predictions about selection on heterozygosity within crosses. Consider, for example, male backcrosses in an XY species, where the X chromosome is large, the Y chromosome is small or degenerate, and the parental lines are reasonably fit. These backcross males will vary in their autosomal heterozygosity ($p_{12,A}$), but the selection on this heterozygosity will vary with the alleles carried on the X. This is illustrated in Figure 1 C. If the backcross is made to P1, but h_X is large (i.e. divergent X-linked alleles come mainly from P2), then heterozygosity will tend to be under positive selection (see darker arrow in Fig. 1 C); but if h_X is small, then heterozygosity will tend to be under negative selection (see lighter arrow in Fig. 1 C, Appendix 4 and Fig. S8 for details).

To test this prediction, we used data from Noor et al. (2001) (Fig. S7). These authors generated male backcrosses between *Drosophila pseudoobscura* and *Drosophila persimilis*. These species are ideal for our test, because the X carries ~37% of the coding sequence, the Y is degenerate, and the hybrids can be fully sterile (see Table 1, Methods, and Noor et al. 2001). Each backcross male was scored for sterility, and multiple genetic markers, including two markers on the X (see Table 1). We characterized the X as low- h_X if both markers carried the P1 allele, and as high- h_X if both markers carried the P2 allele; we excluded intermediate individuals, for which we have no clear prediction. We then regressed sterility on heterozygosity, $p_{12,A}$, but allowed the intercept and the slope to vary with h_X . Results, shown in Table 3, show that both h_X and $p_{12,A}$ are important predictors of sterility in these backcrosses, and that their effects interact. With low h_X , sterility is generally rarer, but positively correlated with $p_{12,A}$. With high h_X , by contrast, sterility is more common, and negatively correlated with $p_{12,A}$. The predictions of Fisher's model are therefore supported, in both cross directions.

The interaction between h_X and $p_{12,A}$ may also help to explain a broader set of patterns observed by Moehring (2011)

in multiple datasets of *Drosophila* backcrosses (Macdonald and Goldstein 1999; Noor et al. 2001; Moehring et al. 2006a, 2006b and see Table 1). Moehring found that male fertility correlated strongly and negatively with h_X , while correlations with $p_{12,A}$ were weak and inconsistent. This pattern was not consistent with predictions of a simple model of X-autosome incompatibilities (Moehring 2011), but it is consistent with Fisher's geometric model (see Fig. 1 C, Fig. 2 B, and Appendix 4).

Female backcrosses in *Teleogryllus*

Fisher's model can help to explain more complex patterns. To illustrate this, we will consider the data of Moran et al. (2017) from the field crickets *Teleogryllus oceanicus* and *T. commodus*. These species have XO sex determination, and a large X chromosome ($g_X \approx 0.3$, from Moran et al. 2017). They are also a rare exception to Haldane's Rule, with F1 sterility appearing solely in XX females (Hogan and Fontana 1973). Moran et al. (2017) hypothesized that female sterility might be caused by negative interactions between heterospecific alleles on the X, which appear together in F1♀, but not in F1♂.

To test this hypothesis, they generated backcross females with identical recombinant autosomes, but different levels of heterozygosity on the X, $p_{12,X}$. The hypothesis of X-X incompatibilities predicts that fertility will decrease with $p_{12,X}$, but this was not observed. This is shown in Figure 4 A.

To try and explain the patterns that were observed, let us note two clear asymmetries in the data. First, there are strong differences between the fitness levels of the two parental species in lab conditions, with *T. commodus* (labeled P2 in Fig. 4) producing around half the eggs and offspring of *T. oceanicus* (P1 in Fig. 4). This suggests that equation 10 might apply to these data. Figure 4 shows a second asymmetry in the data: the reciprocal F1 (female-male vs. male-female) have very different fitness (Turelli and Moyle 2007; Fraïsse et al. 2016b). Because the X and autosomes will be identical for both cross directions, this implies some sort of parent-of-origin effect on the phenotype. This could be either uniparental inheritance, or selective silencing

Table 3. Regressions of male sterility on autosomal heterospecificity in *Drosophila* backcrosses.

Backcross to	N	Model	Best-fit coefficients			AIC
			intercept (low h_X)	intercept (high h_X)	for $p_{12,A}$	
<i>D. persimilis</i>	577	two intercepts	-1.071	1.797	-	596.79
		two intercepts + one slope	-2.165	0.764	2.147	580.41
		two intercepts + two slopes	-3.033	2.871	3.746 (low- h_X); -1.973 (high- h_X)	558.91
<i>D. pseudoobscura</i>	610	two intercepts	-0.197	2.962	-	603.53
		two intercepts + one slope	-2.031	1.219	3.620	558.18
		two intercepts + two slopes	-2.485	4.034	4.505 (low- h_X); -1.876 (high- h_X)	545.87

AIC, Akaike Information Criterion; preferred model shown in bold.

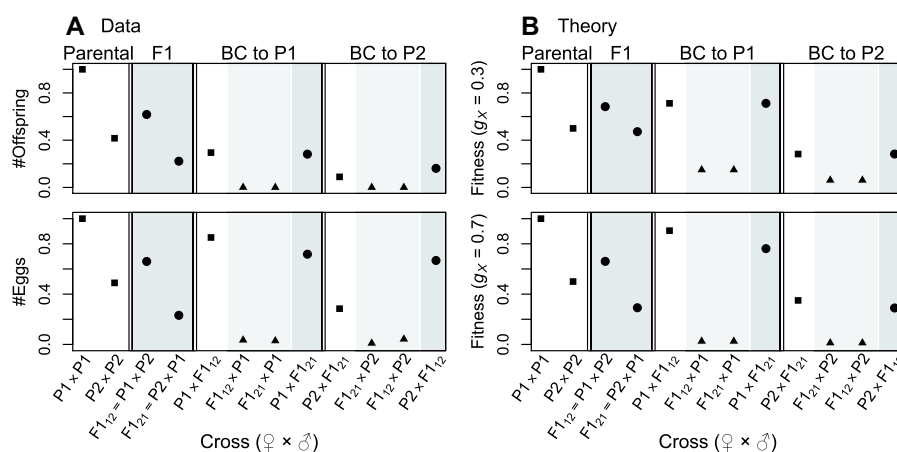


Figure 4. Data on female fertility, from crosses of the field crickets *Teleogryllus oceanicus* (P1), and *T. commodus* (P2), from Moran et al. (2017), compared to theoretical predictions from Fisher's geometric model. Plotting styles denote the level of X-linked heterozygosity: high (circles and darker shading); intermediate (triangles and lighter shading) or low (squares and no shading). To plot the data (column A), we used the mean number of offspring per pair (upper panel), or mean number of eggs per pair (lower panel), each normalized by the value for P1 (see the supplementary information of Moran et al. 2017 for full details). The theoretical predictions (column B) are listed in Table S4. In the upper panel, we assume $g_X = 0.3$, as estimated from the chromosome sizes, and complete silencing of the paternal X (such that $\pi = 1$ in Table S4). In the lower panel, we assume $g_X = 0.7$, and incomplete silencing of the paternal X ($\pi = 0.8$), to improve fit to the egg data. While predictions apply to the rank order of fitnesses, to aid visualization, we plot $w = e^{-6f^2}$ (see eq. 1), and set the parameter f_{P2} via $0.5 = e^{-6f_{P2}}$, to reflect the lower fertility of this species under lab conditions.

(Turelli and Moyle 2007; Fraïsse et al. 2016b). One possibility is the speculation of Hoy and collaborators (see e.g., Hoy et al. 1977; Butlin and Ritchie 1989; Dr. Peter Moran pers. comm.), that dosage compensation in *Teleogryllus* involves silencing of the paternal X.

We can now use the foregoing assumptions to predict the levels of breakdown for each of the crosses produced by Moran et al. (2017). The predictions for each cross are listed in Table S4, and plotted in the upper panel of Figure 4 B. This simple model explains several striking aspects of the *Teleogryllus* data.

Adjusting the parameter values can further improve the fit. For example, if we increase g_X (as would be case if divergent sites affecting female fecundity were clustered on the X), and assume that paternal X silencing is incomplete (affecting 80% of the divergent sites), then the results, shown in the lower panel of Figure 4 B, agree well with the data on *Teleogryllus* egg number, as shown in the lower-left-hand panel (only the high fitness of P2 \times F1₁₂ is poorly predicted).

Further adjustments are possible, but these soon become *ad hoc*, at least without further knowledge of the true nature of parent-of-origin effects in *Teleogryllus*. The important point here is that

Fisher's geometric model explains several key features of these hybrid data, while using only a single parameter derived from the data themselves; and even this parameter (f_{P2}), was estimated from the parental control lines.

ESTIMATING THE FITNESS SURFACE

Across a diverse collection of hybrids, equation (6) predicts that the hybrid index will be under disruptive selection, and heterozygosity under directional selection. This prediction can be tested with datasets containing estimates of fitness, h and p_{12} for many hybrid individuals. Exactly such an analysis was presented by Christe et al. (2016), for families of wild hybrids from the forest trees, *Populus alba* and *P. tremula* (Lindtke et al. 2012, 2014; Christe et al. 2016). These authors scored survival over four years in a common-garden environment, and fit a generalized linear model to their binary data (binary logistic regression, with "family" as a random effect), and predictors including linear and quadratic terms in p_{12} and h . Model selection favored a four-term model, with terms in p_{12} , h , and h^2 (see Table S6, and Supplementary information of Christe et al. 2016 for full details). For comparison with our theoretical predictions, we can write their best fit model in the following form:

$$y = const + \beta_0 (\beta_1 h (1 - \beta_2 h) - p_{12}) \quad (17)$$

where y is the fitted value for hybrid breakdown. In its general form, equation (6) should hold and Fisher's model predicts that $\beta_0 > 0$, $0 \leq \beta_1 \leq 4$, and $\frac{1}{2} \leq \beta_2 \leq 2$. Additionally if parental maladaptation is not highly asymmetrical then it predicts $\beta_2 \approx 1$.

The best-fit model of Christe et al. (2016) corresponds to $\hat{\beta}_0 = 2.963$, $\hat{\beta}_1 = 2.777$ and $\hat{\beta}_2 = 0.934$, which supports the predictions of directional selection toward higher heterozygosity, and near-symmetrical diversifying selection on the hybrid index.

To obtain confidence intervals on these parameters, we fit equation (17) to the raw data of Christe et al. (2016). We also searched for other data sets, from which we could estimate the hybrid fitness surface. After applying some quality controls (see Methods and Table S2), we identified one other dataset of wild hybrids, from the mouse subspecies *Mus musculus musculus/domesticus*, where male testes size was the proxy for fertility (Turner and Harr 2014). We also found four datasets of controlled crosses: F2 from the same mouse subspecies (White et al. 2011), and the ragworts *Senecio aethnensis* and *S. chrysanthemifolius* (Chapman et al. 2016); and the *Drosophila* backcrosses discussed above (Macdonald and Goldstein 1999; Moehring et al. 2006a, 2006b). Unlike the data from wild hybrids, these single-cross datasets leave large regions of the fitness surface unsampled (see Fig. 1); nevertheless, they each contain enough variation in h and p_{12} for a meaningful estimation. Details of

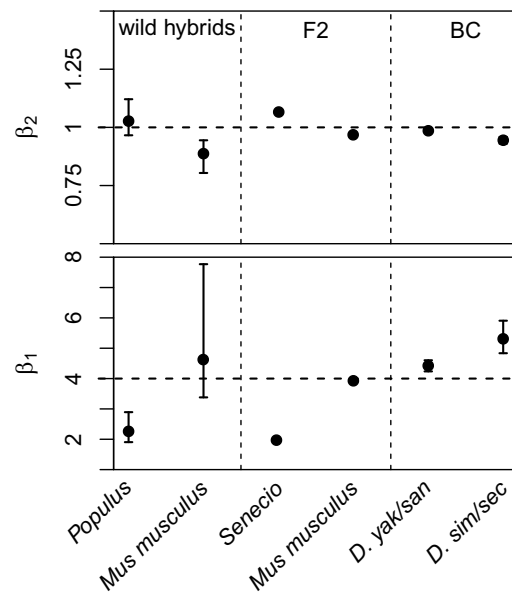


Figure 5. Best fit parameter estimates for the GLM of equation (17), with fitness and genomic data from six sets of hybrids (see Table 1 for details). The upper panel shows estimates of the coefficient β_2 that determines the form of selection acting on the hybrid index, h . Estimates of $\beta_2 = 1$ are consistent with symmetrical diversifying selection. The lower panel show estimates of the coefficient β_1 that determine the relative strength of selection acting on the hybrid index. Estimates of $\beta_1 = 4$ are predicted when the parental types are well adapted (eq. 9), while estimates $0 < \beta_1 < 4$ are predicted when the parental types are maladapted (eq. 6). Confidence intervals are defined as values that reduce the AIC by 2 units. These measures of uncertainty were not obtained for the F2 data, where variation in the hybrid index contributed little to the model fit, as predicted by equation (13). Full details of the model fitting are found in the Methods and Tables S5 and S6.

all six datasets are shown in Table 1, and they are plotted in Figures S9–S11.

Figure 5 shows a summary of the estimated parameters, and full results are reported in Tables S5 and S6, and Figures S9–S11. Taken together, the results show good support for the predictions of equation (6). For all six datasets there was evidence of significant positive selection on heterozygosity ($\hat{\beta}_0 > 0$ was preferred in all cases). Furthermore, for all six datasets, we inferred diversifying selection acting on the hybrid index. Estimates of β_2 , shown in the upper panel of Figure 5, show that this selection was near-symmetrical in all cases, such that $\hat{\beta}_2 \approx 1$. The poorest fit to the predictions was found for the *Drosophila* backcrosses, where estimates of β_1 were significantly greater than the predicted upper bound of $\beta_1 = 4$ (Fig. 5 lower panel). But these datasets were least suited to our purpose, because estimates of h and p_{12} depend strongly on our rough estimate of the relative contributions of the X and autosomes (see Methods), and because they lack

F2-like genotypes, from the center of the fitness surface (Fig. 1 A; Fig. S11). By contrast, results for the *Mus musculus* F2 (White et al. 2011), are remarkably close to the predictions of equation (9) (Fig. 5; Fig. S10).

Two other features of the results deserve comment. First, for the two F2 datasets, it was not possible to provide meaningful confidence intervals for β_1 and β_2 . This is because, for these two datasets, the terms in h and h^2 did not make a significant contribution to model fit, and so the preferred model contained only selection on p_{12} (see Table S6). This is consistent with our earlier prediction of equation (13), and stems from the low variation in $4h(1-h)$ among F2 hybrids (see Appendix 3 and Figs. S9 and S10).

Second, for two of the datasets, *Populus* and *Senecio*, the estimates of β_1 are substantially lower than 4 (Fig. 5; Fig. S9). This is suggestive of parental maladaptation, creating true heterosis in the hybrids (see eq. 6). Consistent with this inference, there is independent evidence of parental load and F1 hybrid vigor in both species pairs (*Populus*: Caseys et al. 2015; Christe et al. 2017; *Senecio*: Abbott and Brennan 2014).

Discussion

In this article, we have used Fisher's geometric model to develop predictions for the relative fitness of any class of hybrid. The modeling approach is simple, with few free parameters, and it generates a wide range of testable predictions. We have tested some of these predictions with new and published datasets (Table 1), and the major predictions of the model are well supported.

We emphasize that our approach is designed for coarse-grained patterns in the data, and typical outcomes of hybridization, without considering the particular set of substitutions that differentiate the parental lines, or the particular combination of alleles in an individual hybrid. The limitations of such an approach are seen in the low r^2 values associated with our model fitting (Table S5). Nevertheless, our approach should enable novel and complementary uses of genomic datasets, which do not depend on identifying individual loci with anomalous effects. In this approach, all alleles, and not just a handful of "speciation genes" might contribute to hybrid fitness.

A second goal of the present work was to show how Fisher's model can interpolate between previous modeling approaches, namely the classical theory of inbreeding (Wright 1922; Crow 1952), and models of genetic incompatibilities, each involving a small number of loci (Dobzhansky 1937; Orr 1995; Gavrilits 2004; Welch 2004). We have also shown that Fisher's model can account for empirical patterns that each approach has struggled to explain (Wright 1977; Moehring 2011; Moran et al. 2017).

Though we have stressed their similarities, we should also stress that Fisher's model and the model of incompatibilities remain different in two ways. First, as shown in Appendix 2, the two models agree only when the dominance relations of incompatibilities take a particular set of values (eq. A36)—albeit a biologically realistic set (Appendix 2; Turelli and Orr 2000). Second, even when predictions are identical for the quantity f (eq. 2), the two approaches still make different predictions for other kinds of data, which were not considered in the present work. The most important difference is the dependency of log fitness on d , the genomic divergence between the species. Under Fisher's geometric model, the log fitness of hybrids declines with $-d^{\alpha/2}$ (eqs. 1–2 and 5–6). By contrast, with the simplest models of incompatibilities, log fitness declines with $-d^{\ell\alpha/2}$ where ℓ is the number of loci involved in each incompatibility (Orr 1995; Welch 2004; Appendix 2, eq. A22). This is due to a snowball effect, where the number of incompatibilities grows explosively with d^ℓ . This is a genuine difference between the modeling approaches, although truly discriminatory tests may be difficult. For example, it may not always be possible to distinguish between a snowballing model with a low value of α (equivalent to strong positive epistasis between incompatibilities), or a model where α is higher, but where the number of "incompatibilities" does not snowball, because they appear and disappear as the genetic background changes (Welch 2004; Fraïsse et al. 2016b; Guerrero et al. 2017; Kalirad and Azevedo 2017). Furthermore, Fisher's model can generate an "apparent snowball effect," when introgressed segments contain multiple divergent sites (Fraïsse et al. 2016b), and this is true of the strongest empirical demonstrations of the effect (Matute et al. 2010; Moyle and Nakazato 2010; Wang et al. 2015).

Finally, given the simplicity and flexibility of the modeling approach explored here, and its predictive successes with a range of data, it should be readily extendable to address other outstanding questions in the study of hybridization. These include the putative role of hybridization in adaptive evolution (e.g., Mendez et al. 2012; Fraïsse et al. 2016b, 2016a; Duranton et al. 2017), the effects of recombination in shaping patterns of divergence (Aeschbacher et al. 2017; Schumer et al. 2018), the analysis of clines (Barton and Gale 1993; Macholán et al. 2011; Baird et al. 2012), and the roles of intrinsic versus extrinsic isolation (Chevin et al. 2014, Fig. S4). Given its ability to interpolate between models of different and extreme kinds, it should also be particularly useful for understanding hybridization in intermediate regimes, where parental genomes are characterized by both maladaptation and allelic coadaptation, or where the architecture of isolation involves many genes of small or moderate effect. Data—including those analyzed here—suggest that such architectures might be quite common (Davis and Wu 1996; Maside and Naveira 1996; Rosas et al. 2010;

Morán and Fontdevila 2014; Baird 2017; Buerkle 2017; Boyle et al. 2017).

Methods

Mytilus DATA

Conserved tissues from the mussel species, *Mytilus edulis* and *Mytilus galloprovincialis*, and their hybrids, were retained from the work of Bierne et al. (2002, 2006). As reported in those studies, *M. edulis* from the North of France were crossed with *M. galloprovincialis* from the French Mediterranean coast to produce F1 hybrids (five males and one female; Bierne et al. 2002). The F1 were then used to produce an F2, and sex-reciprocal backcrosses to *M. galloprovincialis* (which we denote here as BC₁₂ and BC₂₁; Bierne et al. 2006). In particular, oocytes from the F1 female were fertilized by the pooled sperm of the five F1 males producing F2 individuals, from which 132 individuals were sampled; oocytes from the F1 female were fertilized by pooled sperm of five *M. galloprovincialis* males to produce BC₁₂, from which 72 individuals were sampled; and five *M. galloprovincialis* females were fertilized by pooled sperm from the five F1 males, producing BC₂₁, from which 72 individuals were sampled. In addition to these hybrids, we also genotyped 129 individuals from “reference” populations of the two species, found in regions with relatively little contemporary introgression. In particular, we sampled *M. galloprovincialis* from Thau in the Mediterranean Sea; and sampled *M. edulis* from four locations in the North Sea and English Channel (The Netherlands, Saint-Jouin, Villerville and Réville). Full details of these reference populations are found in Table S3.

In each case, gill tissues were conserved in ethanol at -20°C . DNA was extracted using a NucleaMag 96 Tissue kit (Macherey-Nagel) and a KingFisherTM Flex (ThermoFisher Scientific). We then genotyped all samples for 98 *Mytilus* markers that were designed from the data of Fraïsse et al. (2016a). The flanking sequences of the 98 SNPs are provided in Table S7. Genotyping was subcontracted to LGC-genomics and performed with the KASPTM chemistry (Kompetitive Allele Specific PCR, Semagn et al. 2014). Results are shown in Figure S6. Many of the 98 markers are not diagnostic for *M. edulis* and *M. galloprovincialis*, and so we retained only the 43 that were scored as heterozygous in all six of the F1 hybrids. To obtain a reduced set of strongly diagnostic markers, we measured sample allele frequencies in our pure species *M. edulis* and *M. galloprovincialis* samples (pooling *M. edulis* individuals across the four sampling locations; Table S3), and retained only markers for which the absolute difference in allele frequencies between species was $> 90\%$. This yielded the set of 33 markers used for the right-hand columns in Table 2. The “sub-sampled” data shown in the fourth column of Table 2, excluded any BC hybrid who carried the major allele typical of *M. edulis*

in homozygous form. This yielded 56 BC hybrids. Because sequencing order was haphazard, to equalize the sample sizes, we retained the first 56 F2 to be sequenced.

COLLATION OF PUBLISHED DATA

We searched the literature for published datasets combining measures of individual hybrid fitness, with genomic data that could be used to estimate p_{12} and h . We rejected any datasets where the measure of fecundity or fertility took an extreme low value for one of the pure species, suggesting that it is not a good proxy for fitness (e.g., Orgogozo et al. 2006), data where the fitness proxy correlated strongly with a measure of genetic abnormality such as aneuploidy (Xu and He 2011), or data where the states of many markers could not be unambiguously assigned, for example, due to shared variation. Before estimating the fitness surface, we also excluded any dataset where there was a highly significant rank correlation between the proportion of missing data in an individual, and either their heterozygosity, or fitness. For this reason, we did not proceed with reanalyses of the excellent datasets of Li et al. (2011), or Routtu et al. (2014) (see Table S2 for full details). For our reanalysis of the *Teleogryllus* data of Moran et al. (2017), we did not consider data from the second backcross, to avoid complications due to selection on heterozygosity during the first backcross, and because of an anomalous hatching rate in the *T. oceanicus* controls. For our reanalysis of the *Mus musculus* F2 (White et al. 2011), we used a conservative subset of these data; we excluded any individual where any X-linked marker was scored as heterozygous (indicative of sequencing errors in heterogametic males; White et al. 2011), and controlled for variation in the uniparentally inherited markers, by retaining only individuals carrying *M. m. domesticus* mitochondria, and the *M. m. musculus* Y. However, results were little changed when we used all 304 individuals with fertility data (Table S6). Results were also unaffected when we used alternative proxies for fitness (Table S6). For the wild *Mus musculus* hybrids data (Turner and Harr 2014), fixed markers and their orientation between species were determined using previously published data of Staubach et al. (2012) and Yang et al. (2009). See details in the Dryad data package.

ESTIMATION OF g_X FROM ANNOTATED GENOMES

To fit the generalized linear models (GLM) to data from species with XY sex determination (Table 1), we needed to account for the different marker densities on the X and autosomes. Accordingly, we estimated the overall values of p_{12} and h from equation (16). We estimated the weightings, g_X and g_A , from the total length of coding sequences associated with each chromosome type, ignoring the small contributions from the Y and mitochondria. In each case, we obtained the longest protein product for each unique gene, and then summed their lengths, using a custom R script. The

g_X values, shown in Table 1, were calculated as the total length of X-linked CDS divided by the total CDS length. For *Mus musculus*, we used the reference genome assembly “GRCm38.p5.” For *Drosophila simulans*, we used the assembly “GCA_000754195.3 ASM75419v2,” and for *Drosophila yakuba* “GCA_000005975.1 dyak_caf1.” For *Drosophila pseudoobscura*, the current annotation was downloaded from FlyBase release 3.04 (Gramates et al. 2017). The .gtf file was then sorted and merged (combining overlapping coding sequences on each chromosome) using BEDTools (Quinlan and Hall 2010). Coding sequence lengths were calculated and summed over each chromosome, using custom awk commands.

GLM METHODS

The linear model results shown in Table 3, Figure 5, Tables S5 and S6, and Figures S9–S11, were all fit in R v. 3.3.2 (R Core Team 2016). For datasets with quantitative fitness measures (White et al. 2011; Turner and Harr 2014; Fig. S10) we used the standard general linear models, with Gaussian errors, and chose data transformations to reduce heteroscedasticity. For binary fitness data (Noor et al. 2001; Christe et al. 2016; Chapman et al. 2016; Table 3; Fig. S9), we used binomial regression with a logit link implemented in the *glm* function; and with ordinal fitness data (Macdonald and Goldstein 1999; Moehring et al. 2006b; Fig. S11) we used proportional odds logistic regression (Agresti 2003), implemented in the *polr* function. In these cases, the *P*-values shown in Table S6 were calculated by comparing the *t*-value to the upper tail of normal distribution, as in a Wald test. For the non-Gaussian models, we also report McFadden’s pseudo- r^2 , defined as one minus the ratio of log likelihoods for the fitted and null models (McFadden 1974).

DATA ARCHIVING

All data and analysis scripts have been deposited in the Dryad repository (DOI: <https://doi.org/10.5061/dryad.2hb2br7>). Source code used for the simulations is available here https://github.com/alxsimon/Hybrid_fitness_Fisher_geometric_model.

AUTHOR CONTRIBUTIONS

A.S. and J.W. worked on the mathematical models and derivations. All authors designed the study, collected the data and wrote the manuscript.

ACKNOWLEDGMENTS

We are very grateful to all of the authors of the original data reanalyzed here and especially to the following, who provided clarifications or reformatting: Luisa Bresadola, Charlie Brummer, Mark A. Chapman, Camille Christe, Eric Hagg, Xionglei He, Christian Lexer, Xuehui Li, Amanda Moehring, Bret Payseur, Michael White, and Gavin Woodruff. We are also grateful to Andrea Betancourt for advice on processing the genome annotations, and to Peter Moran, who spotted a serious mistake in an earlier draft. Stuart Baird, Jon Slate, David Waxman, and anonymous reviewers, provided helpful comments that led to substantial improvements. This work was partly funded by the Agence Nationale de la Recherche

(HYSEA project, ANR-12-BSV7-0011). NB acknowledges a fellowship of the French Embassy in the United Kingdom, with Churchill College Cambridge. A.S. acknowledges a fellowship of the Labex CeMEB and the doctoral school GAIA. J.W. acknowledges Joanna Weinert.

The authors declare no conflicts of interests.

LITERATURE CITED

- Abbott, R. J., and A. C. Brennan. 2014. Altitudinal gradients, plant hybrid zones and evolutionary novelty. *Philos. Trans. R. Soc. B Biol. Sci.* 369: 20130346.
- Abbott, R. J., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, et al. 2013. Hybridization and speciation. *J. Evol. Biol.* 26: 229–246.
- Aeschbacher, S., J. P. Selby, J. H. Willis, and G. Coop. 2017. Population-genomic inference of the strength and timing of selection against gene flow. *Proc. Natl. Acad. Sci.* 114: 7061–7066.
- Agresti, A. 2003. *Categorical data analysis*. 2nd ed. John Wiley & Sons, Inc., Hoboken, New Jersey.
- Baird, S. J. E. 2017. The impact of high-throughput sequencing technology on speciation research: maintaining perspective. *J. Evol. Biol.* 30: 1482–1487.
- Baird, S. J. E., A. Ribas, M. Macholán, T. Albrecht, J. Piálek, and J. Göü de Bellocq. 2012. Where are the wormy mice? A reexamination of hybrid parasitism in the European house mouse hybrid zone: helminth parasites in the house mouse hybrid zone. *Evolution* 66: 2757–2772.
- Barton, N. H. 2001. The role of hybridization in evolution. *Mol. Ecol.* 10: 551–568.
- . 2017. How does epistasis influence the response to selection? *Heredity* 118:96–109.
- Barton, N. H., and K. S. Gale. 1993. Genetic analysis of hybrid zones. Pp. 13–45 *in* Hybrid zones and the evolutionary process, R. G. Harrison ed. Oxford Univ. Press, Oxford.
- Bierne, N., F. Bonhomme, P. Boudry, M. Szulkin, and P. David. 2006. Fitness landscapes support the dominance theory of post-zygotic isolation in the mussels *Mytilus edulis* and *M. galloprovincialis*. *Proc. R. Soc. B* 273: 1253–1260.
- Bierne, N., P. David, P. Boudry, and F. Bonhomme. 2002. Assortative fertilization and selection at larval stage in the mussels *Mytilusedulis* and *M. galloprovincialis*. *Evolution* 56: 292–298.
- Blanquart, F., G. Achaz, T. Bataillon, and O. Tenailon. 2014. Properties of selected mutations and genotypic landscapes under Fisher’s geometric model. *Evolution* 68: 3537–3554.
- Bouchemousse, S., L. Lévêque, G. Dubois, and F. Viard. 2016. Co-Occurrence and Reproductive Synchrony Do Not Ensure Hybridization between an Alien Tunicate and Its Interfertile Native Congener. *Evol. Ecol.* 30: 69–87.
- Boyle, E. A., Y. I. Li, and J. K. Pritchard. 2017. An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169: 1177–1186.
- Buerkle, C. A. 2017. Inconvenient truths in population and speciation genetics point towards a future beyond allele frequencies. *J. Evol. Biol.* 30: 1498–1500.
- Butlin, R. K., and M. G. Ritchie. 1989. Genetic coupling in mate recognition systems: what is the evidence? *Biol. J. Linn. Soc.* 37: 237–246.
- Caseys, C., C. Stritt, G. Glauser, T. Blanchard, and C. Lexer. 2015. Effects of hybridization and evolutionary constraints on secondary metabolites: the genetic architecture of phenylpropanoids in European populus species. *PLoS ONE* 10: e0128200.
- Chapman, M. A., S. J. Hiscock, and D. A. Filatov. 2016. The genomic bases of morphological divergence and reproductive isolation driven

- by ecological speciation in *Senecio* (Asteraceae). *J. Evol. Biol.* 29: 98–113.
- Chevin, L.-M., G. Decorzent, and T. Lenormand. 2014. Niche dimensionality and the genetics of ecological speciation. *Evolution* 68: 1244–1256.
- Christe, C., K. N. Stölting, M. Paris, C. Frasse, N. Bierne, and C. Lexer. 2017. Adaptive evolution and segregating load contribute to the genomic landscape of divergence in two tree species connected by episodic gene flow. *Mol. Ecol.* 26: 59–76.
- Christe, C., K. N. Stölting, L. Bresadola, B. Fussi, B. Heinze, D. Wegmann, and C. Lexer. 2016. Selection against recombinant hybrids maintains reproductive isolation in hybridizing *Populus* species despite F₁ fertility and recurrent gene flow. *Mol. Ecol.* 25: 2482–2498.
- Connallon, T. and A. G. Clark. 2014. Balancing Selection in Species with Separate Sexes: Insights from Fisher's Geometric Model. *Genetics* 197: 991–1006.
- Coyne, J. A., and H. A. Orr. 1989. Two rules of speciation. Pp. 180–207 in D. Otte and J. A. Endler, eds. *Speciation and its consequences*. Sinauer Associates, Sunderland, MA.
- . 2004. *Speciation*. Vol. 37. Sinauer Associates, Sunderland, MA.
- Crow, J. F. 1952. Dominance and overdominance. In *Heterosis*, J. W. Gowen ed. Iowa State College Press, Ames, Iowa.
- Davis, A. W., and C.-I. Wu. 1996. The broom of the Sorcerer's apprentice: the fine structure of a chromosomal region causing reproductive isolation between two sibling species of *Drosophila*. *Genetics* 143: 1287–1298.
- Demuth, J. P., and M. J. Wade. 2005. On the theoretical and empirical framework for studying genetic interactions within and among species. *Am. Nat.* 165: 524–536.
- Dobzhansky, T. G. 1937. *Genetics and the origin of species*. Columbia Univ. Press, New York.
- Duranton, M., F. Allal, C. Fraïsse, N. Bierne, F. Bonhomme, and P.-A. Gagnaire. 2018. The origin and remodeling of genomic islands of differentiation in the European Sea Bass *Nature Communications* 9, no. 1. <https://doi.org/10.1038/s41467-018-04963-6>.
- Duranton, M., F. Allal, C. Fraïsse, N. Bierne, F. Bonhomme, and P.-A. Gagnaire. 2017. The origin and remodeling of genomic islands of differentiation in the European sea bass. *bioRxiv*. <https://doi.org/10.1101/223750>.
- Edmunds, S. 2002. Does Parental Divergence Predict Reproductive Compatibility? *Trends Ecol. Evol.* 17: 520–527.
- Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567–1587.
- Fisher, R. A. 1930. *The genetical theory of natural selection*. Clarendon Press, Oxford.
- Fitzpatrick, B. M. 2008. Hybrid dysfunction: population genetic and quantitative genetic perspectives. *Am. Nat.* 171: 491–498.
- . 2012. Estimating ancestry and heterozygosity of hybrids using molecular markers. *BMC Evol. Biol.* 12: 131.
- Fraïsse, C., J. A. D. Elderfield, and J. J. Welch. 2014. The Genetics of Speciation: Are Complex Incompatibilities Easier to Evolve? *J. Evol. Biol.* 27: 688–699.
- Fraïsse, C., K. Belkhir, J. J. Welch, and N. Bierne. 2016a. Local interspecies introgression is the main cause of extreme levels of intraspecific differentiation in mussels. *Mol. Ecol.* 25:269–286.
- Fraïsse, C., P. A. Gunnarsson, D. Roze, N. Bierne, and J. J. Welch. 2016b. The genetics of speciation: insights from Fisher's geometric model. *Evolution* 70:1450–1464.
- Gavrilets, S. 2004. *Fitness landscapes and the origin of species*. Princeton Univ. Press, Princeton.
- Gompert, Z., and C. A. Buerkle. 2010. Introgress: a software package for mapping components of isolation in hybrids. *Mol. Ecol. Resour.* 10: 378–384.
- Gramates, L. S., S. J. Marygold, G. dos Santos, J.-M. Urbano, G. Antonazzo, B. B. Matthews, A. J. Rey, C. J. Tabone, M. A. Crosby, D. B. Emmert, et al. 2017. FlyBase at 25: looking to the future. *Nucleic Acids Res.* 45: D663–D671.
- Guerrero, R. F., C. D. Muir, S. Josway, and L. C. Moyle. 2017. Pervasive antagonistic interactions among hybrid incompatibility loci. *PLoS Genet.* 13: e1006817.
- Haldane, J. B. S. 1922. Sex ratio and unisexual sterility in hybrid animals. *J. Genet.* 12: 101–109.
- Hallauer, A. R., M. J. Carena, and J. B. Mirandad Filho. 2010. Quantitative genetics in maize breeding, *Handbook of plant breeding*. Vol. 6. Springer, Berlin.
- Hinze, L. L., and K. R. Lamkey. 2003. Absence of epistasis for grain yield in elite maize hybrids. *Crop Sci.* 43: 46–56.
- Hogan, T. W., and P. G. Fontana. 1973. Restoration of meiotic stability following artificial hybridisation and selection in *Teleogryllus* (Orth., Gryllidae). *Bull. Entomol. Res.* 62: 557–563.
- Hoy, R., J. Hahn, and R. Paul. 1977. Hybrid cricket auditory behavior: evidence for genetic coupling in animal communication. *Science* 195: 82.
- Hwang, S., S.-C. Park, and J. Krug. 2017. Genotypic complexity of Fisher's geometric model. *Genetics* 206: 1049–1079.
- Kalirad, A., and R. B. R. Azevedo. 2017. Spiraling complexity: a test of the snowball effect in a computational model of RNA folding. *Genetics* 206: 377–388.
- Kiesselbach, T. (1922). *Corn investigations*. Neb. Agric. Exp. Stn. Bull. 20:5–151.
- Kiesselbach, T. 1930. The use of advanced generation hybrids as parents of double-cross seed corn. *J. Am. Soc. Agron.* 22: 614–625.
- Kinman, M. L., and G. F. Sprague. 1945. Relation between number of parental lines and theoretical performance of synthetic varieties of corn. *J. Am. Soc. Agron.* 37: 341–351.
- Li, X., X. Wang, Y. Wei, and E. C. Brummer. 2011. Prevalence of segregation distortion in diploid alfalfa and its implications for genetics and breeding applications. *Theor. Appl. Genet.* 123: 667–679.
- Lindstrom, E. W. 1941. Analysis of modern maize breeding principles and methods. Pp. 667–679 in *Proceedings of the 7th International Genetics Congress*, Edinburgh, Scotland.
- Lindtke, D., Z. Gompert, C. Lexer, and C. A. Buerkle. 2014. Unexpected ancestry of *Populus* seedlings from a hybrid zone implies a large role for postzygotic selection in the maintenance of species. *Mol. Ecol.* 23: 4316–4330.
- Lindtke, D., C. A. Buerkle, T. Barbará, B. Heinze, S. Castiglione, D. Bartha, and C. Lexer. 2012. Recombinant hybrids retain heterozygosity at many loci: new insights into the genomics of reproductive isolation in *Populus*. *Mol. Ecol.* 21: 5042–5058.
- Lynch, M. 1991. The genetic interpretation of inbreeding depression and outbreeding depression. *Evolution* 45: 622–629.
- Macdonald, S. J., and D. B. Goldstein. 1999. A quantitative genetic analysis of male sexual traits distinguishing the sibling species *Drosophila simulans* and *D. sechellia*. *Genetics* 153: 1683–1699.
- Macholán, M., S. J. E. Baird, P. Duřková, P. Munclinger, B. V. Bímová, and J. Piálek. 2011. Assessing multilocus introgression patterns: a case study on the mouse X chromosome in central Europe: heterogeneity of introgression on the mouse X chromosome. *Evolution* 65: 1428–1446.
- Mank, J. E., D. J. Hosken, and N. Wedell. 2011. Some Inconvenient Truths about Sex Chromosome Dosage Compensation and the Potential Role of Sexual Conflict. *Evolution* 65: 2133–2144.

- Manna, F., G. Martin, and T. Lenormand. 2011. Fitness Landscapes: An Alternative Theory for the Dominance of Mutation. *Genetics* 189: 923–937.
- Mani, G. S., and B. C. Clarke. 1990. Mutational order: a major stochastic process in evolution. *Proc. R. Soc. Lond. B Biol. Sci.* 240: 29–37.
- Martin, G. 2014. Fisher's geometrical model emerges as a property of complex integrated phenotypic networks. *Genetics* 197: 237–255.
- Martin, G., and T. Lenormand. 2006. A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution* 60: 893–907.
- Martin, J. M., and A. R. Hallauer. 1976. Relation between heterozygosity and yield for four types of maize inbred lines. *Egypt. J. Genet. Cytol.* 5: 119–135.
- Maside, X. R., and H. F. Naveira. 1996. On the difficulties of discriminating between major and minor hybrid male sterility factors in *Drosophila* by examining the segregation ratio of sterile and fertile sons in backcrossing experiments. *Heredity* 77: 433–438.
- Matute, D. R., I. A. Butler, D. A. Turissini, and J. A. Coyne. 2010. A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science* 329: 1518–1521.
- McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. Pp. 105–142 in P. Zarembka ed. *Frontiers in econometrics*. Academic Press, Cambridge, Massachusetts.
- Melchinger, A. E. 1987. Expectation of means and variances of testcrosses produced from F2 and backcross individuals and their selfed progenies. *Heredity* 59: 105–115.
- Mendez, F. L., J. C. Watkins, and M. F. Hammer. 2012. A haplotype at STAT2 introgressed from neanderthals and serves as a candidate of positive selection in papua New Guinea. *Am. J. Hum. Genet.* 91: 265–274.
- Moehring, A. J., A. Llopart, S. Elwyn, J. A. Coyne, and T. F. C. Mackay. 2006a. The genetic basis of postzygotic reproductive isolation between *Drosophila santomea* and *D. yakuba* due to hybrid male sterility. *Genetics* 173: 225–233.
- . 2006b. The genetic basis of prezygotic reproductive isolation between *Drosophila santomea* and *D. yakuba* due to mating preference. *Genetics* 173: 215–223.
- Moehring, A. J. 2011. Heterozygosity and its unexpected correlations with hybrid sterility. *Evolution* 65: 2621–2630.
- Moll, R. H., J. H. Lonquist, J. V. Fortuno, and E. C. Johnson. 1965. The relationship of heterosis and genetic divergence in maize. *Genetics* 52: 139–144.
- Moran, P. A., M. G. Ritchie, and N. W. Bailey. 2017. A rare exception to Haldane's rule: are X chromosomes key to hybrid incompatibilities? *Heredity* 118:554–562.
- Morán, T., and A. Fontdevila. 2014. Genome-wide dissection of hybrid sterility in *Drosophila* confirms a polygenic threshold architecture. *J. Hered.* 105: 381–396.
- Moyle, L. C., and T. Nakazato. 2010. Hybrid incompatibility “Snowballs” between solanum species. *Science* 329: 1521–1523.
- Neal, N. P. 1935. Decrease in yielding capacity in advanced generations of hybrid corn. *J. Am. Soc. Agron.* 51: 666–670.
- Newcomer, J. T., N. K. Neerchal, and J. G. Morel. 2008. Computation of Higher Order Moments from Two Multinomial Overdispersion Likelihood Models. *Dep. Math. Stat. Univ. Md. Baltim.* USA.
- Noor, M. A. F., K. L. Grams, L. A. Bertucci, Y. Almendarez, J. Reiland, and K. R. Smith. 2001. The genetics of reproductive isolation and the potential for gene exchange between *Drosophila pseudoobscura* and *D. persimilis* via backcross hybrid males. *Evolution* 55: 512–521.
- Orgogozo, V., K. W. Broman, and D. L. Stern. 2006. High-resolution quantitative trait locus mapping reveals sign epistasis controlling ovule number between two *Drosophila* species. *Genetics* 173: 197–205.
- Orr, H. A. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* 139: 1805–1813.
- . 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52: 935–949.
- Pollak, E., H. F. Robinson, and R. E. Comstock. 1957. Inter-population hybrids in open-pollinated varieties of maize. *Am. Nat.* 91: 387–391.
- Quinlan, A. R., and I. M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842.
- R Core Team. 2016. R: a language and environment for statistical computing. R found. *Stat. Comput.* Vienna Austria <https://www.R-project.org>.
- Revuz, D., and M. Yor. 1999. *Continuous martingales and Brownian motion*. 3rd ed. Springer-Verlag, New York.
- Richey, F. D., and G. F. Sprague. 1931. Experiments on hybrid vigor and convergent improvement in corn. 267. U.S. Department of Agriculture, Washington, D.C., pp. 1–22.
- Rockman, M. V. 2012. The QTN program and the alleles that matter for evolution: All that's gold does not glitter. *Evolution* 66: 1–17.
- Rosas, U., N. H. Barton, L. Copsey, P. Barbier de Reuille, and E. Coen. 2010. Cryptic variation between species and the basis of hybrid performance. *PLoS Biol.* 8: e1000429.
- Routtu, J., M. D. Hall, B. Albere, C. Beisel, R. Bergeron, A. Chaturvedi, J.-H. Choi, J. Colbourne, L. De Meester, M. T. Stephens, et al. 2014. An SNP-based second-generation genetic map of *Daphnia magna* and its application to QTL analysis of phenotypic traits. *BMC Genomics* 15: 1033.
- Roux, C., C. Fraïsse, J. Romiguier, Y. Anciaux, N. Galtier, and N. Bierne. 2016. Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biol.* 14: e2000234.
- Schiffman, J. S., and P. L. Ralph. 2017. System drift and speciation. *bioRxiv*. <https://doi.org/10.1101/231209>.
- Schumer, M., C. Xu, D. L. Powell, A. Durvasula, L. Skov, C. Holland, J. C. Blazier, S. Sankaraman, P. Andolfatto, G. G. Rosenthal, et al. 2018. Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* 360: 656–660.
- Semagn, K., R. Babu, S. Hearne, and M. Olsen. 2014. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Mol. Breed.* 33: 1–14.
- Sentz, J. C., H. F. Robinson, and R. E. Comstock. 1954. Relation between heterozygosity and performance in maize. *Agron. J.* 46: 514–520.
- Shehata, A. H., and N. L. Dhawan. 1975. Genetic analysis of grain yield of maize as manifested in diverse varietal populations and their crosses. *Egypt. J. Genet. Cytol.* 4: 90–116.
- Staubach, F., A. Lorenc, P. W. Messer, K. Tang, D. A. Petrov, and D. Tautz. 2012. Genome patterns of selection and introgression of haplotypes in natural populations of the house mouse (*Mus Musculus*). *PLoS Genetics* 8: e1002891. <https://doi.org/10.1371/journal.pgen.1002891>.
- Stringfield, G. 1950. Heterozygosity and hybrid vigor in maize. *Agron. J.* 42: 45–112.
- Tenaillon, O., O. K. Silander, J.-P. Uzan, and L. Chao. 2007. Quantifying organismal complexity using a population genetic approach. *PLoS ONE* 2: e217.
- Turelli, M., and L. C. Moyle. 2007. Asymmetric postmating isolation: Darwin's corollary to haldane's rule. *Genetics* 176: 1059–1088.
- Turelli, M., and H. A. Orr. 2000. Dominance, epistasis and the genetics of postzygotic isolation. *Genetics* 154: 1663.
- Turner, L. M., and B. Harr. 2014. Genome-wide mapping in a house mouse hybrid zone reveals hybrid sterility loci and Dobzhansky-Muller interactions. *Elife* 3: e02504.

- Wang, R. J., M. A. White, and B. A. Payseur. 2015. The pace of hybrid incompatibility evolution in house mice. *Genetics* 201: 229–242.
- Waser, N. M. 1993. Population structure, optimal outbreeding and assortative mating in angiosperms. Pp. 173–199 in N. W. Thornhill, ed. *The natural history of inbreeding and outbreeding: Theoretical and empirical perspectives*. Chicago Univ. Press, Chicago.
- Waxman, D., and J. J. Welch. 2005. Fisher’s microscope and haldane’s ellipse. *Am. Nat.* 166: 447–457.
- Welch, J. J. 2004. Accumulating dobzhansky-muller incompatibilities: reconciling theory and data. *Evolution* 58: 1145–1156.
- Welch, J. J., and D. Waxman. 2003. Modularity and the cost of complexity. *Evolution* 57: 1723–1734.
- White, M. A., B. Steffy, T. Wiltshire, and B. A. Payseur. 2011. Genetic dissection of a key reproductive barrier between nascent species of house mice. *Genetics* 189: 289–304.
- Wright, S. 1922. The effects of inbreeding and crossbreeding on guinea pigs : III. crosses between highly inbred families. *U. S. Dep. Agric. Bull.* 1121.
- . 1977. Inbreeding depression and heterosis: plants. In *Evolution and the genetics of populations, volume 3: Experimental results and evolutionary deductions, vol. 3*. Chicago Univ. Press, Chicago.
- Wu, C.-I. and A. W. Davis. 1993. Evolution of Postmating Reproductive Isolation: The Composite Nature of Haldane’s Rule and Its Genetic Bases. *Am. Nat.* 142: 187–212.
- Xu, M., and X. He. 2011. Genetic incompatibility dampens hybrid fertility more than hybrid viability: yeast as a case study. *PLoS ONE* 6:e18341.
- Yang, H., Y. Ding, L. N. Hutchins, J. Szatkiewicz, T. A. Bell, B. J. Paigen, J. H. Graber, F. P.-M. de Villena, and G. A. Churchill. 2009. A customized and versatile high-density genotyping array for the mouse. *Nature Methods* 6: 663–666.

Associate Editor: Z. Gompert

Appendix 1: The Brownian Bridge Approximation

In this Appendix, we derive the Brownian bridge approximation for the breakdown score of hybrid genotypes, under Fisher’s geometric model. The Appendix is in three parts. First, we state our approximation without much justification. We will also assume haploid genomics, which makes the approximation simpler to explain, and also helps to draw connections with a previous model of epistasis introduced by Barton and Gale (1993). In the second section, we justify the approximation, by showing how it relates to explicit models of genetic and phenotypic divergence. We also show that the approximation is quite robust to varying assumptions about the divergence process. Third, and finally, we introduce the full, diploid version of the approximation, which is used for the results in the main text.

Results with haploid genetics

In this section, and the following section, we will assume that the parental species P1 and P2 are haploid. In this case, of course,

there is no heterozygosity ($p_{12} = 0$) and the hybrid index is $h = p_2 = 1 - p_1$.

We can now consider the d substitutions that differentiate P1 and P2 as a chain of vectors, in the n -dimensional phenotypic space, which connect one of the parental phenotypes to the other. Our approximation is to treat this path, on each of the n traits, as describing a Brownian bridge, that is, a random walk constrained at its start and end points (Revuz and Yor 1999, Ch. 1 for details). This approximation will be justified by appealing to the random combinations of alleles that appear together in hybrid genotypes (see below).

To derive our approximation, let $B(t)$ denote a Brownian bridge, taking place over a single unit of time, and with a rate σ_B . $B(t)$ is normally distributed, with the following mean:

$$E(B(t)) = (1 - t)B(0) + tB(1), \quad 0 \leq t \leq 1 \quad (A1)$$

where $B(0)$ and $B(1)$ are the fixed points at the beginning and end of the walk. The covariance at two time points is given by

$$\text{Cov}(B(t_1), B(t_2)) = \sigma_B^2(1 - t_2)t_1, \quad 0 \leq t_1 \leq t_2 \leq 1 \quad (A2)$$

(Revuz and Yor 1999, Ch. 1). For our purposes, the start and end points of the walk are given by the trait values of the parental lines, so that, for trait i , $B(0) = z_{P1,i}$ and $B(1) = z_{P2,i}$. The total length of the walk represents the complete path of d substitutions, each with typical effect v_i , so that $\sigma_B^2 = dv_i$. Finally, to determine the trait value for a hybrid carrying dh alleles that come from P2, we just “stop” the random walk at the intermediate “time” h . As such, using equations (A1) and (A2), the trait value z_i will be a normal random variable with the following properties:

$$z_i \sim N(\mu_i, \sigma_i^2)$$

$$\mu_i \equiv E(z_i) = (1 - h)z_{P1,i} + hz_{P2,i} \quad (A3)$$

$$\sigma_i^2 \equiv \text{Var}(z_i) = dv_i h(1 - h) \quad (A4)$$

The expectations and variances in these expressions involve two sorts of averaging. First, there is averaging over different realizations of evolutionary process, that is, the different sets of substitutions that might differentiate P1 and P2. Second, there is averaging over the subset of substitutions that might appear in a hybrid with a given value of h . Only this second averaging process will apply to real datasets, and it helps to explain the connection between our approximation and explicit models of divergence (see below).

To complete the results for haploidy, we note that the breakdown score, S , depends on the squared trait value (see eq. 4 of the main text). From normal theory, we then have:

$$E(z_i^2) = \sigma_i^2 + \mu_i^2 \tag{A5}$$

$$\text{Var}(z_i^2) = 2\sigma_i^2(\sigma_i^2 + 2\mu_i^2) \tag{A6}$$

and so S will be approximately gamma distributed, with a mean and variance given by the weighted sum of these quantities (see also Rosas et al. 2010). The key quantity f , as defined in equation (2), is then found to be

$$f \equiv \frac{E(S)}{E(S^{\dagger})} = \frac{d \sum_i^n \lambda_i v_i h(1-h) + \sum_i^n \lambda_i ((1-h)z_{P1,i} + hz_{P2,i})^2}{d \sum_i^n \lambda_i v_i} = h(1-h) + f_{mal} \tag{A7}$$

where

$$f_{mal} = \frac{\sum_i \lambda_i ((1-h)z_{P1,i} + hz_{P2,i})^2}{d \sum_i \lambda_i v_i} \tag{A8}$$

The term f_{mal} comes from the μ_i^2 in equation (A5), and describes the contribution to hybrid breakdown from parental maladaptation. The value of f_{mal} depends on the relative positions in phenotypic space of the two parental types. It is useful to consider a few simple cases, which are summarized in Table S1, and which we also use in the main text. First, let us consider the case where maladaptation accrues independently in the two parental lines. In this case, shown in Table S1b, we expect no covariance between their phenotypic deviations, and so f_{mal} is just a weighted sum of the relative breakdown scores for the two parental lines

$$f_{mal} = (1-h)^2 f_{P1} + h^2 f_{P2}, \quad \overline{z_{P1,i} z_{P2,i}} = 0 \tag{A9}$$

where $f_{P1} = \sum_i \lambda_i z_{P1,i}^2 / \sum_i \lambda_i d v_i$. The assumption of independent maladaptation should apply to many real datasets, and so equation (A9) is used in the main text, as the basis of equations (6)–(9) (see below for details). Independent maladaptation need not apply in every case, however, and it might be the case that P1 and P2 are identically maladapted, perhaps because of a shared unsuitability to a lab environment. This case, however, turns out to be comparable to the case with no parental adaptation. As shown in Table S1c, when both parental phenotypes are sub-optimal, but identical, all genotypes suffer an additional constant contribution to their maladaptation. Finally, we can imagine a case where the two populations tend to become maladapted in opposite directions, such that the midparental phenotype is optimal. Results for this case are shown in Table S1d, but it is probably the least realistic of the three scenarios.

When parental lines are well adapted, such that $f_{mal} \approx 0$, then our haploid model is closely connected to a simple model of epistasis introduced by Barton and Gale (1993). They provide results for cline shape and linkage disequilibrium under this model (Barton and Gale 1993; p. 18).

Comparison to simple models of divergence

The approximation introduced above includes only two fixed points—the parental phenotypes—and ignores their common ancestor, and all of the intermediate steps in their divergence. Nevertheless, we can be fairly confident that all of these intermediates had reasonably high fitness. The reason it is possible to ignore these intermediate steps is that mutations fix in a particular order during divergence, but in hybrids, they appear in random combinations, due to recombination and segregation. This makes the Brownian bridge approximation robust to a wide range of different assumptions about population divergence.

In this section of the Appendix, we illustrate this point, and explore the robustness of the approximation. For simplicity, we continue to assume haploid genetics, and consider only one phenotypic trait (hence, we write $z_i = z$ throughout). We will also limit ourselves to models with a low mutation rate, such that fixations occur one at a time. Given this last assumption, we can order the d fixations that differentiate P1 and P2 in a single chain, starting at P1, and then going back in time to the most recent common ancestor, and then forwards in time again toward P2. This is illustrated in Figure S1. We will denote as $z(j)$, the trait value in the j th step of this chain, such that $z(0)$ is the trait value for P1, and $z(d)$ is the trait value for P2. Each genotype will contain a different number of alleles from P2, and the hybrid index of phenotype $z(j)$ is simply $h = j/d$. Finally, the phenotypic effect of the j th substitution in the chain is defined by

$$m_j \equiv z(j) - z(j-1) \tag{A10}$$

Figure S1 illustrates these quantities.

Now, let us first consider the case where both populations remain well adapted, or close to the optimum, throughout the divergence process. We can then model the phenotype at step j as a normal random variable.

$$z(j) \sim N(0, v/2) \tag{A11}$$

From equations (A10) and (A11), it follows that the effects of neighboring substitutions in the chain will negatively covary, implying that maladapted populations will tend to fix mutations that return them closer to the optimum:

$$\text{Cov}(m_i, m_j) = \begin{cases} v, & i = j \\ -v/2, & i = j \pm 1 \\ 0, & \text{otherwise} \end{cases} \tag{A12}$$

Now let us consider the phenotype that contains dh alleles from P2, namely $z(dh)$. This phenotype can be written as the sum of substitutional changes, starting from the initial P1 phenotype:

$$z(dh) = z(0) + \sum_{i=1}^{dh} m_i \quad (\text{A13})$$

We define the relative breakdown of this genotype as $f_h \equiv z^2(dh)/(dv)$, and the expected value of this quantity is

$$\begin{aligned} E(f_h) &\equiv \frac{E(z^2(dh))}{dv} = \frac{1}{dv} \left[\text{Var}(z(0)) + 2\text{Cov}\left(z(0), \sum_{i=1}^{dh} m_i\right) \right. \\ &\quad \left. + \text{Var}\left(\sum_{i=1}^{dh} m_i\right) \right] \\ &= \frac{1}{dv} \left[\text{Var}(z(0)) + 2\text{Cov}\left(z(0), \sum_{i=1}^{dh} m_i\right) \right. \\ &\quad \left. + dh\text{Var}(m_i) + 2 \sum_{i < j \leq dh} \text{Cov}(m_i, m_j) \right] \\ &= \frac{1}{2d} + h \left(\sqrt{2}\rho_{z(0),m_i} + 1 \right. \\ &\quad \left. + (dh - 1)\rho_{m_i,m_j} \right) \quad (\text{A14}) \end{aligned}$$

where $\rho_{z(0),m_i}$ is the average correlation between the initial P1 phenotype, and the effect of each substitution; and ρ_{m_i,m_j} is the average correlation between each distinct pair of substitutions. Let us first evaluate equation (A14) when the dh mutations are added to the hybrid in the same fixed order that they appear in the chain of fit intermediates (Fig. S1). In this case, from equation (A12), only the first of the dh substitutions will covary with the initial phenotype, and so we have $\rho_{z(0),m_i} = -1/(\sqrt{2}dh)$. Similarly, each substitution will only covary with its preceding and subsequent substitutions, such that $\rho_{m_i,m_j} = -\frac{1}{2} \frac{2(dh-1)}{dh(dh-1)} = -1/(dh)$. As such, when mutations are added in this fixed order, equation (A14) yields:

$$E(f_h) = \frac{1}{2d} = E(f_P) \quad (\text{A15})$$

So the expected breakdown of all genotypes, regardless of their hybrid index, is equal to that of the parental lines. Equation (A15) reflects the fact that each of these “hybrid” genotypes is one of the fit intermediates, and all will be very fit compared to the worst class of hybrid, unless the divergence, d , is very small. This result can be contrasted with the result for real hybrids, where segregation and recombination will bring together collections of alleles that fixed at different times. If the dh substitutions in equation (A13) are chosen at random, without replacement, from the complete set of d substitutions, then the correlation

terms in equation (A14) are both reduced by a factor h , reflecting the fact that consecutive substitutions in the chain will only appear together with probability h . With $\rho_{z(0),m_i} = -1/(\sqrt{2}d)$, and $\rho_{m_i,m_j} = -1/d$, equation (A14) becomes

$$E(f_h) = h(1 - h) + E(f_P) \quad (\text{A16})$$

Even without conditioning on the parental phenotypes, equation (A16) is now very close to the Brownian bridge approximation of equation (A7). This is illustrated in Figure S2a. In this simulation run, the diverging populations remained close to the optimum throughout (as indicated by the jagged black line). By contrast, hybrids, which contain a randomized collection of alleles, can be very unfit on average (as shown by the smoother black curves). Their expected deviation from the optimum is well described by the Brownian bridge approximation (see colored dotted lines, which show eq. A7). Because the maladaptation of the parental lines is small, the analytical predictions are little affected if we explicitly account for the observed maladaptation, by calculating f_{mal} from equation (A8) (see red dotted lines in Fig. S2a), or if we ignore parental maladaptation altogether, by setting f_{mal} to zero (see blue dotted lines in Fig. S2a).

The model discussed above assumed that the effects of fixed mutations were normally distributed (see eqs. A10 and A11 and Fig. S2a), but the effect sizes of real substitutions might show substantial deviations from normality. To investigate the effects of nonnormal effect sizes, Figure S2b shows results when divergence included a few substitutions of very large effect (simulated by drawing the trait values, $z(i)$, from a Cauchy distribution). The results show that the Brownian bridge continues to give a very good fit. This is because the approximation follows from robust, central-limit-type behavior, and so it does not depend strongly on the normality of the allelic effects.

The approximation is also robust if we assume more complicated patterns of divergence. For example, Figure S2c shows results where we have assumed that both populations adapted independently to a shifting optimum. This was modeled by generating the $z(i)$ from an Ornstein–Uhlenbeck process, where the ancestral population (characterized by an intermediate value of h), was strongly maladapted.

All of the examples above assume well-adapted parental lines. By contrast, under strong inbreeding, or mutation accumulation, divergence might be entirely maladaptive, with no tendency for populations to revert to the optimum. To model this, we can generate the allelic effects, m_i , as uncorrelated normal variables, starting from an optimal ancestral state. In this case, if we assume that the parental lines each contain $d/2$ substitutions, then their phenotypes are distributed as $z(0), z(d) \sim N(0, vd/2)$, such that $E(f_P) = 1/2$. As shown in Figure S2d, the Brownian bridge approximation remains accurate in this case. For a single trait, the

accuracy does depend strongly on knowing the true parental trait values (dotted red lines), and the approximation that uses their expected squared values (dotted blue lines) can be quite inaccurate. Nevertheless, over multiple traits, we can expect the errors of this approximation to cancel out. A clear feature of this approximation, is that the expected levels of breakdown are constant for all haploid hybrids (the blue line in Fig. S2d is flat). This reflects the fact that, without natural selection, both the parental lines, and all possible hybrid genotypes are simply random assemblies of alleles, without any coadaptation. This echoes the result for diploids with maximal parental maladaptation, where breakdown levels depend solely on heterozygosity, and not on the hybrid index (see eq. 8 and Fig. 2A).

Examples illustrated in Figure S2 use simple heuristic simulations, where the phenotypes $z(j)$ were drawn from a known distribution. In Figures S3 and S4 we show results when these phenotypes were generated by explicit population genetic simulations, including multiple phenotypic traits, and the appearance, fixation, or loss of new mutations. The full simulation procedure is described in Fraïsse et al. (2016b), but all simulations assumed allopatric divergence between P1 and P2, haploid genomics, and a low mutation rate, such that fixation or loss events could be tracked one mutation at a time. In the simulations shown in Figure S3, we began with a core set of parameters (Fig. S3a), and then varied these one at a time. Specifically, we varied the curvature of the fitness function, α , with a core set of parameters (eq. 1; Fig. S3b), the size of the parental populations (Fig. S3c), the total number of traits, n (eq. 4; Fig. S3d), and the shape of the distribution of mutant effects (Fig. S3e). These different parameter regimes led to very different levels of genetic divergence between the parents (i.e., different values of d), and different distributions of fixed phenotypic effects. Nevertheless, the Brownian bridge approximation remained very accurate in all cases. Finally, we moved the phenotypic optima during the simulation, so that much of the divergence took place by positive selection (see Fraïsse et al. 2016b; Fig. S3f). This sometimes led to an increase in the variance in breakdown score (Fig. S3f), but otherwise, the Brownian bridge approximation remained accurate.

Figure S4 shows further simulations, where the optima can move in discrete jumps (as in Orr 1998; Barton 2001). For $n = 2$ traits, we compare divergence with a stable, fixed optimum (Fig. S4a), optima that vary identically for the two diverging populations (Fig. S4b), and optima that move in different directions in the environments experienced by P1 and P2 (Fig. S4c). These scenarios correspond, roughly, to mutation-order speciation via drift (Fig. S4a) or selection (Fig. S4b), and to ecological speciation (Fig. S4c), where reproductive isolation evolves as a byproduct of adaptation to different environments (Mani and Clarke 1990; Coyne and Orr 2004; Gavrillets 2004; Chevin et al. 2014).

When the optima can vary, we can also distinguish between intrinsic reproductive isolation, which involves maladaptation in traits that are under identical selection in all environments, and extrinsic isolation, which involves maladaptation in traits whose optima vary (Chevin et al. 2014). To illustrate this, the right-hand columns in Figure S4 compare the breakdown caused by selection on both traits, measured in the environment to which P1 has adapted, and the breakdown due to one trait only, corresponding to the vertical axis on the fitness landscapes, for which the optimum never varies. This shows how the approach can be used to model breakdown that is caused by the correlated effects of adaptation to distinct environments (due to pleiotropy or linkage), even when hybrid fitness is measured in a benign environment, where these divergent selection pressures do not apply.

Extension to diploidy

In this final section of the Appendix, we extend the results above to diploid genetics, as described in the main text. In this case, we must explicitly consider heterozygosity, p_{12} , and the hybrid index must include contributions from both homozygous and heterozygous alleles: $h = p_2 + p_{12}/2$.

To apply the Brownian bridge approximation to diploidy, we will assume that the chain of random vectors applies to alleles in their heterozygous state. Because all mutations are assumed to be additive on the phenotype, this means that the random walk, which begins at the phenotype of P1, will now end at the midparental phenotype, instead of the phenotype of P2. It also means that, to deal with alleles that appear in homozygous state, some sections of the walk must be counted twice. In particular, we need to study the following quantity:

$$B_{dip}(t_1, t_2) \equiv B(t_1 + t_2) + B(t_1) - B(0) \quad (A17)$$

where $B(\cdot)$ is a Brownian bridge, as described in equations (A1) and (A2) above. From these equations, we find:

$$E(B_{dip}(t_1, t_2)) = (1 - 2t_1 - t_2)B(0) + (2t_1 - t_2)B(1) \quad (A18)$$

$$\begin{aligned} \text{Var}(B_{dip}(t_1, t_2)) &= \text{Var}(B(t_1 + t_2)) + \text{Var}(B(t_1)) \\ &\quad + 2\text{Cov}(B(t_1), B(t_1 + t_2)) \\ &= \sigma_B^2 \{(t_1 + t_2)(1 - t_1 - t_2) \\ &\quad + t_1(1 - t_1) + 2(1 - t_1 - t_2)t_1\} \\ &= \sigma_B^2 \{t_2(1 - t_2) + 4t_1(1 - t_1 - t_2)\} \quad (A19) \end{aligned}$$

For our diploid problem, we start the walk at the P1 phenotype: $B(0) = z_{P1,i}$, but end it at the midparent $B(1) = (z_{P1,i} + z_{P2,i})/2$. The total walk length is $\sigma_B^2 = d v_i$, just as with

haploidy, but now v_i is understood as the typical effect of an allele in heterozygous state. Finally, we take the intermediate “times” as $t_1 = p_2$ (this section of the walk is counted twice, to account for homozygous alleles), and $t_2 = p_{12}$ (this section is counted once, to account for heterozygous alleles). This results in

$$z_i \sim N(\mu_i, \sigma_i^2)$$

$$\mu_i \equiv E(z_i) = (1 - 2p_2 - p_{12})z_{P1,i} + (2p_2 + p_{12})\frac{z_{P1,i} + z_{P2,i}}{2}$$

$$= (1 - h)z_{P1,i} + hz_{P2,i} \quad (A20)$$

$$\sigma_i^2 \equiv \text{Var}(z_i) = dv_i(p_{12}(1 - p_{12}) + 4p_1p_2)$$

$$= dv_i(4h(1 - h) - p_{12}) \quad (A21)$$

We note that equation (A20) is identical to the haploid result of equation (A3). This shows that, once the hybrid index is correctly defined, all of the results concerning parental maladaptation (as shown in eqs. A8 and A9 and Table S1) apply equally to diploids and haploids. Combining equations (A20)–(A21) with equations (A5)–(A6), and equation (A9), leads directly to the diploid results shown in the main text.

The only other difference to note is that maladaptive diploid substitutions now contribute a possible $2^2v_i = 4v_i$ to the trait variance. This implies that highly maladapted parents, who have fixed $\sim d/2$ such mutations, will have a breakdown score of $E(S_P) \sim \frac{d}{2}4 \sum_i \lambda_i v_i$, such that $E(f_P) = 2$ under diploidy (as opposed to $E(f_P) = 1/2$ under haploidy). The diploid result is used in the main text, to explore data from highly inbred lines of *Zea mays*.

Appendix 2: Fisher’s Model and Dobzhansky-Muller Incompatibilities

In this Appendix, we show that a major prediction of Fisher’s geometric model, assuming well-adapted parents (eq. 9), can also be derived from a model of genetic incompatibilities, each involving a small number of loci. Such models stem from the classic work of Dobzhansky 1937 and have been important in the study of speciation genetics (Fraïsse et al. 2014; Orr 1995; Turelli and Orr 2000; Gavrillets 2004; Welch 2004; Demuth and Wade 2005; Fraïsse et al. 2016b). We begin by introducing the model in general form, and then explore different ways of assigning the parameters, which determine the contribution of individual incompatibilities to the overall level of breakdown.

A general model of incompatibilities

Following Orr (1995), let us assume that certain combinations of alleles, at $\ell \leq d$ of the divergent loci, can be intrinsically

incompatible, while all other combinations confer high fitness. By assumption, the pure species genotypes, and their ancestral states, must be fit, but all other combinations have a fixed probability ε_ℓ of being incompatible.

The worst class of hybrid will contain all of the possible incompatibilities, and so its expected breakdown score will be proportional to the expected number of incompatibilities. This was calculated by Welch (2004, eqs. 1–2):

$$E(S_{i,I}) \propto \varepsilon_\ell \binom{d}{\ell} (2^\ell - \ell - 1) \quad (A22)$$

Here, and below, we use the subscript I to indicate a model of incompatibilities. To derive f_I (eq. 2), we note that hybrids will have higher fitness when some of the incompatibilities are absent from their genomes (Turelli and Orr 2000). The probability that an incompatibility is present depends on how many of the ℓ loci are heterozygous. For a genotype comprising i loci that are homozygous for the P1 allele, j loci homozygous for the P2 allele, and k loci that are heterozygous, the probability required is:

$$\pi_{ijk} = \frac{2^k - u(i) - u(j)}{2^\ell - 2}, \quad \text{with } u(x) = \begin{cases} 0 & \text{if } x > 0 \\ 1 & \text{if } x = 0 \end{cases} \quad (A23)$$

which is the proportion of the possible combinations of heterospecific alleles that are present in an “ ijk ” genotype.

Incompatibilities may also have reduced effects due to recessivity, when their negative effects are masked by the presence of alternative, compatible alleles (Turelli and Orr 2000). To model this, we introduce the free parameter s_{ijk} , which is the expected increase in breakdown when an incompatibility appears in an ijk genotype. Finally, in a hybrid genome characterized by p_1 , p_2 , and p_{12} , the trinomial expansion of $(p_1 + p_2 + p_{12})^\ell$, tells us how many ℓ -locus genotypes of each type it is expected to contain. Putting these together, we have

$$f_I = \sum_{i+j+k=\ell} \binom{\ell}{i, j, k} p_1^i p_2^j p_{12}^k \pi_{ijk} s_{ijk} \quad (A24)$$

where

$$\binom{\ell}{i, j, k} \equiv \frac{\ell!}{i!j!k!}$$

Equations (A23) and (A24) extend results with $\ell = 2$ and $\ell = 3$ from Turelli and Orr (2000), and represent a general model of breakdown caused by incompatibilities. A notable feature of these equations is their large number of free parameters. Even with symmetry between P1 and P2 (such that $s_{ijk} = s_{jik}$), we will still require a total of $\lfloor \ell(1 + \ell/4) \rfloor$ different s_{ijk} values to specify the model (i.e., three extra parameters for two-locus incompatibilities, five parameters for three-locus incompatibilities etc.). There is good empirical evidence for, at least, two-, three-, and four-locus incompatibilities (Fraïsse et al. 2014), and so with the s_{ijk} and the

ε_ℓ , the full model would depend on at least 17 free parameters. By contrast, equation (9), from Fisher's geometric model, has no free parameters. The incompatibility-based model is therefore much more flexible, but also much more difficult to explore.

In the following section, we examine different ways of assigning the s_{ijk} . We will show that a particular set of values make the model identical to Fisher's geometric model (eq. 9). We will also follow Turelli and Orr (2000), and show that these values are biologically plausible, given well-established empirical patterns, and especially Haldane's Rule (see Haldane 1922 and Appendix 4).

The dominance relations of incompatibilities

Let us begin by assigning the following functional form for the s_{ijk} :

$$s_{ijk} \propto \left(\frac{1}{2}\right)^{\delta k} \quad (\text{A25})$$

where below, we will use a constant of proportionality such that $s_{ijk} = (1/2)^{\delta k} \cdot 2(2^\ell - 2)$, to simplify the final results. In equation (A25), the parameter δ allows us to tune the dominance of incompatibilities, measured in terms of breakdown score, rather than fitness.

When $\delta = 1$, then each heterozygous locus halves the effects of incompatibility. This is equivalent to assuming that incompatibilities act multiplicatively, since each heterozygous locus halves the number of times that the incompatible combination of alleles is present in the genome. The s_{ijk} under multiplicative selection ($\delta = 1$) are illustrated by the green points in Figure S5.

To determine the predictions of this model, let us substitute equation (A25) into equation (A24), and set $\delta = 1$. After some algebra, we find:

$$f_I = 2 \left[1 - (p_2 + \frac{1}{2}p_{12})^\ell - (p_1 + \frac{1}{2}p_{12})^\ell \right], \quad \delta = 1 \quad (\text{A26})$$

$$\equiv 2 \left[1 - h^\ell - (1-h)^\ell \right], \quad \delta = 1 \quad (\text{A27})$$

$$= 4h(1-h), \quad \ell = 2, \delta = 1 \quad (\text{A28})$$

where h is the hybrid index (eq. 3). As such, when incompatibilities act multiplicatively, breakdown will depend solely on the hybrid index, and not on the heterozygosity. When pairwise incompatibilities are considered (eq. A28), the model becomes equivalent to the haploid version of Fisher's geometric model (see eq. A7), and also to the model of Barton and Gale (1993).

More importantly, equation (A28) predicts that breakdown should not change between the F1 and F2 crosses (both of which have $h = \frac{1}{2}$), and that homogametic F1 will have the highest possible breakdown score. As such, this multiplicative model cannot predict hybrid breakdown between the F1 and F2, nor Haldane's

Rule, and both patterns have widespread empirical support (see Table A1 of Fraïsse et al. 2016b).

Now let us consider another extreme assumption. We assume that incompatibilities are fully recessive, such that no breakdown appears unless all incompatible alleles appear in homozygous or hemizygous form. We model this by making δ very large, such that $s_{ijk} = 0$ unless $k = 0$. These values are illustrated by the red points in Figure S5. With the assumption of complete recessivity, we find:

$$f_I = 2 \left[(p_1 + p_2)^\ell - p_1^\ell - p_2^\ell \right], \quad \delta \rightarrow \infty \quad (\text{A29})$$

$$= 4h(1-h) - 2p_{12} + p_{12}^2, \quad \ell = 2, \delta \rightarrow \infty \quad (\text{A30})$$

Equation (A29) also fails to predict Haldane's Rule, unless there is substantial uniparental inheritance from both the male and female parents (see Appendix 4). This is because $f_I = 0$ if $p_1 p_2 = 0$, and so both male and female F1 will have identical and optimal fitness. For similar reasons, equation (A29) predicts that the fitness of heterogametic backcrosses will decrease with $p_{12,A}$, and this prediction is not supported by the relevant data (Moehring 2011).

We have shown that both extreme regimes (no recessivity, and complete recessivity) yield unsupported predictions. But what values of δ are biologically plausible? To answer this question, let us consider Haldane's Rule under an incompatibility-based model, and ignoring uniparental inheritance. Assuming that males are heterogametic, Haldane's Rule will hold when

$$f_{I,F1\sigma} > f_{I,F1\varphi} \quad (\text{A31})$$

(see Appendix 4). If we use g_X and $1 - g_X$ to denote the proportions of the parental divergence that is found on the X and autosomes (as in eq. 16 of the main text), then using equations (A24) and (A25), equation (A31) is found to be equivalent to:

$$\left(2(1 - g_X) + 2^\delta g_X \right)^\ell - (1 - g_X)^\ell - \left(1 - g_X + 2^\delta g_X \right)^\ell > 2^\ell - 2 \quad (\text{A32})$$

This condition is most difficult to satisfy when incompatibilities involve two loci ($\ell = 2$), and in this case, we find the solution:

$$\delta > \ln \left(\frac{2 - g_X}{1 - g_X} \right) / \ln(2). \quad (\text{A33})$$

The value of δ that is required to yield Haldane's Rule will therefore increase with g_X . Toward the limit of the biologically plausible range, when two-thirds of the between-species divergence is X-linked ($g_X = 2/3$) Haldane's Rule will hold only if $\delta > 2$. As such, setting $\delta = 2$, such that each heterozygous locus reduces the breakdown score by a factor of four, will yield Haldane's Rule in most cases. The s_{ijk} values from equation (A25) with $\delta = 2$ are shown as yellow points in Figure S5. Another

feature of the model with $\delta = 2$ is that it produces parameter dependencies that are very close to those predicted by Fisher's geometrical model (see also Manna et al. 2011). The similarity is clearest with two-locus incompatibilities, where we find

$$f_I = \left(\frac{1}{2}\right)^{\delta-2} p_{12} \left(1 - p_{12} \left[1 - \left(\frac{1}{2}\right)^\delta\right]\right) + 4p_1 p_2, \quad \ell = 2 \quad (\text{A34})$$

$$= 4h(1 - h) - p_{12} + \frac{1}{4}p_{12}^2, \quad \ell = 2, \delta = 2 \quad (\text{A35})$$

Comparing equation (A35) to equation (9), shows that $f_I \approx f$ when we use equation (A25) with $\delta = 2$.

We can then go further, and find a set of s_{ijk} values that yield exactly the same dependencies as Fisher's model. To do this, we set $f_I = f$, using equations (9) and (A24), and then solve for the s_{ijk} . After some algebra, we find

$$s_{ijk} = \frac{(\ell - k)\ell - (i - j)^2}{\ell(\ell - 1)\pi_{ijk}} \quad (\text{A36})$$

These values are shown as blue points in Figure S5. Equation (A36) looks unwieldy, and it was derived solely to make the models agree. Nevertheless, it embodies biologically plausible assumptions about incompatibilities. First, the similarities between the blue and yellow points in Figure S5 show that Fisher's model corresponds to the assumption of partial recessivity, at levels sufficient to produce Haldane's Rule. Second, equation (A36) states that incompatibilities will have stronger effects when alleles from both parental species appear in homozygous state. For example, if the three alleles ABc form an incompatibility (with upper and lower case letters distinguishing alleles from P1 and P2), then equation (A36) predicts that the genotype Aa/BB/cc (with $ijk = 111$) will tend to have lower fitness than the genotype AA/BB/Cc (with $ijk = 201$) even though both genotypes contain the incompatibility, and both comprise two homozygous loci and one heterozygous locus.

This section has shown that Fisher's geometric model implies assumptions about "incompatibilities," that yield well supported predictions. However, it remains true that the full incompatibility-based model (eqs. A22–A24) is much more flexible. While that model is very parameter rich, the results above also suggest a family of models that might combine the flexibility and simplicity. When we assume only pairwise incompatibilities ($\ell = 2$), then equations (A28), (A30), (A35), and the result from Fisher's model (eq. 9), are all special cases of

$$f = 4h(1 - h) + (2f_{\text{het}} - r - 2)p_{12} + (1 + r - f_{\text{het}})p_{12}^2 \quad (\text{A37})$$

where f_{het} is the breakdown experienced by the global heterozygote (the genotype with $p_{12} = 1$), and r is the rate of change in breakdown with p_{12} .

Appendix 3: Segregation and Recombination

In this Appendix, we consider the effects of segregation and recombination on the expected levels of breakdown. For a recombinant cross, such as the F2 or backcross, p_{12} and h might vary between individuals, and so we must treat f as a random variable. To see this, let us write equation (9) as

$$\begin{aligned} f &= 4h(1 - h) - p_{12} \\ &= p_{12}(1 - p_{12}) + 4p_1 p_2 \\ &= p_{12} - p_{12}^2 + 4p_1 - 4p_1^2 - 4p_1 p_{12} \end{aligned} \quad (\text{A38})$$

and so its expected value is:

$$\begin{aligned} E(f) &= \bar{p}_{12}(1 - \bar{p}_{12}) + 4\bar{p}_1 \bar{p}_2 - \text{Var}(p_{12}) - 4(\text{Var}(p_1) \\ &\quad + \text{Cov}(p_{12}, p_1)) \end{aligned} \quad (\text{A39})$$

where overbars represent expected values. The variances and covariances in equation (A39) will depend on the distribution of the divergence across the genome, and on patterns of segregation and recombination. However, we can derive simple and useful predictions if we assume that the divergence is equally distributed among m freely recombining regions. In this case, we can use results from a multinomial distribution.

$$\text{Var}(p_{12}) = \frac{\bar{p}_{12}(1 - \bar{p}_{12})}{m} \quad (\text{A40})$$

$$\text{Var}(p_1) = \frac{\bar{p}_1(1 - \bar{p}_1)}{m} \quad (\text{A41})$$

$$\text{Cov}(p_{12}, p_1) = -\frac{\bar{p}_1 \bar{p}_{12}}{m} \quad (\text{A42})$$

These results also apply to estimators from m independently segregating markers. As an example, let us compare the first backcross and the F2, in a population with strictly biparental inheritance. For the first backcross, we have $\bar{p}_{12} = \frac{1}{2}$ and $p_1 p_2 = 0$, and so

$$E(f_{\text{BC1}}) = \frac{1}{4} \left(1 - \frac{1}{m}\right) \quad (\text{A43})$$

For the F2, we have $\bar{p}_{12} = \frac{1}{2}$ and $\bar{p}_1 = \bar{p}_2 = \frac{1}{4}$, and so

$$\begin{aligned} E(f_{\text{F2}}) &= \frac{1}{2} \left(1 - \frac{1}{m}\right) \\ &= 2E(f_{\text{BC1}}) \end{aligned} \quad (\text{A44})$$

and so the predicted breakdown in the F2 is always double that of the first backcross. By the same method, we can also calculate the variance of f , but this requires higher order moments of the multinomial distribution (Newcomer et al. 2008), and this can yield lengthy expressions. However, the following results for the

F2 are required to justify the approximation of equation (13) from the main text.

$$E(4h(1-h)) = 1 - \frac{1}{2m} \approx 1, \quad \bar{p}_{12} = \frac{1}{2}, \quad \bar{p}_1 = \frac{1}{4} \quad (\text{A45})$$

$$\text{Var}(4h(1-h)) = \frac{2m-1}{4m^3} \approx \frac{1}{2m^2}, \quad \bar{p}_{12} = \frac{1}{2}, \quad \bar{p}_1 = \frac{1}{4} \quad (\text{A46})$$

Comparing equation (A46) to equation (A40) shows that most of the variance in F2 breakdown will be due to variation in heterozygosity.

Appendix 4: Predictions of Fisher's Geometric Model with Sex-Specific Inheritance

In this Appendix, we will develop some predictions of Fisher's geometric model, after relaxing the assumption of strictly biparental inheritance. These results justify claims in the main text, and also provide analytical support for some assertions in Fraïsse et al. (2016b).

Additional notation and basics

As described in the main text, the basic strategy for incorporating sex-specific inheritance is to write the genome-wide measures as weighted sums of contributions from loci of different types. For example, if we have a system with an X chromosome and autosomes, and a subset of loci that are strictly maternally or paternally inherited, then we could write:

$$\begin{aligned} p_{12} &= g_X p_{12,X} + g_A p_{12,A} + g_{\text{♀}} p_{12,\text{♀}} + g_{\text{♂}} p_{12,\text{♂}} \\ p_1 &= g_X p_{1,X} + g_A p_{1,A} + g_{\text{♀}} p_{1,\text{♀}} + g_{\text{♂}} p_{1,\text{♂}} \\ p_2 &= g_X p_{2,X} + g_A p_{2,A} + g_{\text{♀}} p_{2,\text{♀}} + g_{\text{♂}} p_{2,\text{♂}} \\ h &= g_X h_X + g_A h_A + g_{\text{♀}} h_{\text{♀}} + g_{\text{♂}} h_{\text{♂}} \end{aligned} \quad (\text{A47})$$

where $g_X + g_A + g_{\text{♀}} + g_{\text{♂}} = 1$. The weightings will vary with the proportion of the divergence found at loci of each type, and with the typical effect sizes of the fixed differences. Following equation (3), the hybrid index for the X chromosome, can also be written as

$$h_X = p_{2,X} + \frac{p_{12,X}}{2} \quad (\text{A48})$$

and similarly for the other quantities. Here, as in the main text, we have assumed an XO system, such that females are homogametic, and males heterogametic. However, the equations are flexible enough to model a range of biological situations. For example, haplodiploidy might be modeled by setting $g_X = 1$, and the uniparentally inherited loci (included via $g_{\text{♀}}$ and $g_{\text{♂}}$) might include

the Y (or W) chromosome, organelles, selectively silenced regions of the X (or Z), or other imprinted loci (Turelli and Moyle 2007).

With sex-specific inheritance, we also require assumptions about sex-specific selection, and this can be incorporated in several ways (Connallon and Clark 2014; Fraïsse et al. 2016b). For example, sexual conflict can be modeled by assuming that there are differences in the optimal trait values for each sex (Connallon and Clark 2014). Alternatively, we could assume that some subset of the traits is under selection in only one sex, for example traits involved in spermatogenesis or oogenesis (Wu and Davis 1993; Coyne and Orr 2004). Finally, we could assume that sexes are under identical selection, which will usually require assumptions about dosage compensation (see below and Fraïsse et al. 2016b).

Patterns in the F1: Haldane's rule and asymmetry

This section presents results for the initial F1 cross ($P_1 \times P_2$), and particularly the findings of Haldane's Rule, and parental sex asymmetry (Haldane 1922; Turelli and Orr 2000; Turelli and Moyle 2007; Fraïsse et al. 2016b).

Haldane's Rule applies to offspring of different sexes, in species with chromosomal sex determination. It states that, when F1 breakdown is stronger in offspring of one sex, it will tend to be the heterogametic sex (Haldane 1922; Turelli and Orr 2000). Previous authors have noted that Fisher's model predicts this pattern (Barton 2001; Fraïsse et al. 2016b; Schiffman and Ralph 2017), and here, we extend this insight to give formal conditions for Haldane's Rule. We will assume identical selection in both sexes, and that pure-species individuals of both sexes have the same fitness. This implies a form of dosage compensation, such that X-linked alleles have identical effects in homozygous or hemizygous state (Mank et al. 2011; Fraïsse et al. 2016b). Assuming an XO or XY system, we expect greater breakdown in the heterogametic (male) sex on the condition that:

$$f_{F1\text{♀}} < f_{F1\text{♂}}. \quad (\text{A49})$$

In homogametic females, all divergent sites on the X and autosomes will be heterozygous, such that $p_{12} = g_A + g_X$, while by definition, uniparentally inherited loci will come from one parent alone, such that $p_1 p_2 = g_{\text{♀}} g_{\text{♂}}$. In heterogametic males, by contrast, only autosomal sites will be heterozygous, such that $p_{12} = g_A$, and X-linked sites will be hemizygous and maternally inherited, such that $p_1 p_2 = (g_{\text{♀}} + g_X) g_{\text{♂}}$. Putting these results together with equation (6), and using

$$\Delta \equiv g_{\text{♀}} - g_{\text{♂}} \quad (\text{A50})$$

to denote the difference in the contributions of exclusively maternally and paternally inherited sites, we find:

$$f_{F1\varnothing} = g_{\varnothing} + g_{\sigma} - \Delta^2 + f_{P\varnothing} \frac{(1 + \Delta)^2}{4} + f_{P\sigma} \frac{(1 - \Delta)^2}{4} \quad (\text{A51})$$

$$= \frac{f_{P\varnothing} + f_{P\sigma}}{4}, \quad g_{\varnothing}, g_{\sigma} = 0 \quad (\text{A52})$$

$$f_{F1\sigma} = g_{\varnothing} + g_{\sigma} + g_X - (\Delta + g_X)^2 + f_{P\varnothing} \frac{(1 + \Delta + g_X)^2}{4} + f_{P\sigma} \frac{(1 - \Delta - g_X)^2}{4} \quad (\text{A53})$$

$$= g_X(1 - g_X) + (1 + g_X^2) \frac{f_{P\varnothing} + f_{P\sigma}}{4} + g_X \frac{f_{P\varnothing} - f_{P\sigma}}{2}, \quad g_{\varnothing}, g_{\sigma} = 0 \quad (\text{A54})$$

where we have used $f_{P\varnothing}$ and $f_{P\sigma}$ to denote the initial maladaptation of the maternal and paternal lines used to make the F1. We can now combine equations (A49)–(A54) to give formal conditions for Haldane’s Rule. We find that the heterogametic sex will tend to show more breakdown if

$$0 < g_X < g_X^* \quad (\text{A55})$$

where

$$g_X^* \equiv 2 \frac{2 - \Delta(4 - (f_{P\varnothing} + f_{P\sigma})) + f_{P\varnothing} - f_{P\sigma}}{4 - (f_{P\varnothing} + f_{P\sigma})} \quad (\text{A56})$$

$$= 1 - 2\Delta, \quad f_{P\varnothing}, f_{P\sigma} = 0 \quad (\text{A57})$$

$$= 2 \frac{1 - \Delta(2 - f_P)}{2 - f_P}, \quad f_{P\varnothing} = f_{P\sigma} \equiv f_P \quad (\text{A58})$$

$$= \frac{4 - 2f_{P\sigma}}{4 - f_{P\sigma}}, \quad \Delta, f_{P\varnothing} = 0 \quad (\text{A59})$$

Several observations follow from the results above. First, the conditions for Haldane’s Rule are always harder to satisfy as Δ increases, and so, from equation (A50), the presence of loci with exclusively maternal inheritance makes Haldane’s Rule less likely, while paternal inheritance makes it more likely. Second, Haldane’s Rule is usually more likely when the parental lines are maladapted. This is an effect of heterosis, that is an intrinsic advantage to heterozygosity, because the homogametic sex will tend to be more heterozygous. Third, the sole exception to this pattern is where the paternal line is much less fit than the maternal (i.e., if $f_{P\sigma} \gg f_{P\varnothing}$ as in eq. A59). In this case, Haldane’s Rule is less likely, because male F1 carry less of the unfit paternal genome.

The patterns above also clarify the broader role of uniparental effects in determining F1 fitness (Fraïsse et al. 2016b). For example, when there is no strictly uniparental inheritance, such that $g_{\varnothing}, g_{\sigma} = 0$, then the homogametic F1 will always be fitter than the average of the parental lines (eq. A52). By contrast, when uniparental inheritance is present (eq. A51), then the relative breakdown remains roughly constant. This implies that the absolute breakdown score will decline steadily with d , the genetic divergence between the parental lines (see eqs. 2 and 5). As such, Fisher’s model requires uniparental inheritance to explain the observation of an “F1 speciation clock” (Edmunds 2002; Fraïsse et al. 2016b). This also helps to determine the conditions for rare exceptions to Haldane’s Rule (as observed, e.g., in *Teleogryllus* Moran et al. 2017). We predict such exceptions only when uniparentally inherited loci act on traits that are subject to selection only in the homogametic sex (Fraïsse et al. 2016b).

Uniparental inheritance is also required, by definition, to explain another widespread phenomenon: the strong asymmetry in fitness between the reciprocal F1, that is male–female versus female–male crosses of the same species pair (Turelli and Moyle 2007; Fraïsse et al. 2016b). Such asymmetry is found even in species without sex chromosomes, or even separate sexes, and so it cannot be connected to Haldane’s Rule in any rigid sense (Bouchemousse et al. 2016; Fraïsse et al. 2016b). The results above (eqs. A51–A54) predict such asymmetry only in special circumstances: when there are uniparental effects of different sizes ($\Delta \neq 0$), and parental lines with different levels of maladaptation ($f_{P\sigma} \neq f_{P\varnothing}$). These are the assumptions that we used to analyze the *Teleogryllus* data of Moran et al. (2017) (see below). However, the observations of F1 asymmetry are very widespread, and apply even to species pairs that are both well adapted (Turelli and Moyle 2007; Fraïsse et al. 2016b). The reason for this apparent discrepancy is that the present work considers the expected breakdown, conditioned only on the levels of parental maladaptation ($f_{P\varnothing}$ and $f_{P\sigma}$), and so expectations for the reciprocal F1 must be identical when $f_{P\sigma} = f_{P\varnothing}$. If, by contrast, we condition on the phenotypic effects of uniparentally inherited alleles, then Fisher’s geometric model can account for the asymmetries observed; this was demonstrated by Fraïsse et al. (2016b).

Patterns in backcrosses

In this section, we provide further details of the analyses of backcross data. These include the male backcrosses from *Drosophila* (Noor et al. 2001; Macdonald and Goldstein 1999; Moehring et al. 2006a, 2006b), that were reanalyzed by Moehring (2011), and the female backcrosses from *Teleogryllus* presented by Moran et al. (2017). In both cases, the data come from hybrids of a single sex, and the fitness traits are components of fertility that are plausibly sex-specific (i.e., spermatogenesis and oogenesis; see Table 1).

As such, we treat the predictions of Fisher's geometric model as applying to a single sex.

The *Drosophila* data (Macdonald and Goldstein 1999; Noor et al. 2001; Moehring et al. 2006a, 2006b) come from three pairs of reciprocal backcrosses. The species pairs are *D. simulans/sechellia* (Macdonald and Goldstein 1999); *D. santomea/yakuba* (Moehring et al. 2006a, 2006b); and *D. pseudoobscura/persimilis* (Noor et al. 2001). In all three cases, the crosses can produce sterile hybrids, and so it is safe to assume that the parental lines are well adapted, compared to the worst possible class of hybrid, so that we can derive predictions from equation (9). Our analyses will also neglect loci with exclusively uniparental inheritance, because they do not qualitatively alter the predictions in this case. With these assumptions, as noted by Moehring (2011), hybrids can be characterized by two measures of "heterospecificity," namely, the autosomal heterozygosity, $p_{12,A}$, and the hybrid index of the X, h_X . All datasets scored fertility in each hybrid fly as a binary, or ordinal trait (see Table 1), and so Moehring (2011) asked whether fertility problems varied systematically with $p_{12,A}$ and h_X . She found, in all six crosses, that backcross fertility problems correlated strongly and positively with h_X , but correlations with $p_{12,A}$ were weak and inconsistent. This is shown in Figure S7, and Table 3 of Moehring (2011). To show how Fisher's geometric model might account for these patterns, we can write equation (9) as

$$\begin{aligned} f_{BC\sigma} &= 4h(1-h) - p_{12} \\ &= p_{12}(1-p_{12}) + 4p_2p_1 \\ &= a(1-a) + 4x(1-x-a) \\ &= a - a^2 + 4x - 4x^2 - 4ax \end{aligned} \quad (\text{A60})$$

where

$$\begin{aligned} a &\equiv (1-g_X)p_{12,A} \\ x &\equiv g_X h_X \end{aligned} \quad (\text{A61})$$

Figure S8a depicts the fitness surface of equation (A60) as a function of x and a . Each dataset of hybrid males could occupy a rectangular region of this surface, as determined by the its value of g_X , and this is how the data are plotted in Figure S7. The regions also correspond to a region of Figure 1 C, comprising the parallelogram delimited by the dotted horizontal line and the two arrows. From annotated *Drosophila* genomes, we estimated that $g_X = 0.17$ might characterize the *simulans/sechellia* and *yakuba/santomea* pairs, and that $g_X = 0.37$ might characterize the *pseudoobscura/persimilis* pair (Table 1; see Methods for details). The regions of parameter space for these values of g_X are marked on Figure S8a, while panels b-e show slices through the fitness surface for these values. In both cases, breakdown increases steadily with h_X , except in the improbable case that the recombinant autosomes were completely heterozygous (Fig. S8b-c). This is consistent with the positive correlations observed by Moehring (2011). By contrast, the dependencies on $p_{12,A}$ (Fig. S8d-e) vary in sign. This is consistent with the lack of consistent correlations with $p_{12,A}$ observed by Moehring (2011) (Fig. S7). Figure S8e also shows how, when g_X is large, the correlations with $p_{12,A}$ reverse in sign, for extreme values of h_X . Figure S7e-f shows how the data of Noor et al. (2001) divide naturally into individuals with high, medium and low values of h_X , and together, this explains the rationale of the test presented in Table 3.

To analyze the female backcross data of Moran et al. (2017) for *Teleogryllus*, we used equation (10), which assumes that P1 is well adapted, while P2 is maladapted. We also assume that a proportion π of the paternal X chromosome is silenced. In this case, the quantities p_{12} and h are calculated without considering silenced alleles, because these make no contribution to the phenotype. This rich dataset contains a wide variety of cross types, and so the full predictions for all of the relevant hybrids are listed in Table S4.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1: A schematic representation of a phenotype evolving over time in two populations, labeled P1 and P2, starting from their most recent common ancestor (MRCA).

Figure S2: The breakdown associated with haploid hybrid genotypes, under simple models of phenotypic divergence.

Figure S3: The breakdown associated with haploid hybrid genotypes, under explicit population genetic simulations of phenotypic divergence.

Figure S4: The breakdown associated with haploid hybrid genotypes, under explicit population genetic simulations of phenotypic divergence, in scenarios involving discrete jumps in the optimal value for one of $n = 2$ traits.

Figure S5: The effects of an incompatibility on hybrid breakdown score, s_{ijk} (eq. 41), when the incompatibility appears in a genotype comprising i loci that are homozygous for alleles from one parental species, j loci that are homozygous for alleles from the other parental species, and k loci that are heterozygous.

Figure S6: Genotype plot for the raw *Mytilus* data.

Figure S7: Plots of the *Drosophila* male backcross data reanalyzed here (see Table 1 and Moehring 2011).

Figure S8: Predictions of Fisher's geometric model for heterogametic male hybrids.

Figure S9: Estimation of the fitness surface for interspecific hybrids from plants (Table 1), namely wild hybrids of the genus *Populus* (row (a); Christe et al. 2016), and an F2 cross of the genus *Senecio* (row (b); Chapman et al. 2016).

Figure S10: Estimation of the fitness surface for subspecific hybrids from *Mus musculus* (Table 1, White et al. 2011; Turner and Harr 2014).

Figure S11: Estimation of the fitness surface for backcross male hybrids from *Drosophila* species pairs (Table 1, Macdonald and Goldstein 1999; Moehring et al. 2006a, b).

Table S1: The contribution of parental maladaptation to hybrid breakdown.

Table S2: Checks for appropriateness for genomic data sets.

Table S3: The reference populations for *Mytilus* crosses.

Table S4: Expected breakdown scores for homogametic female hybrids with paternal X silencing.

Table S5: Inferring the hybrid fitness surface from genomic data sets.

Table S6: The significance of individual regression coefficients.

Table S7: Information on the 98 markers used for the *Mytilus* genotyping.

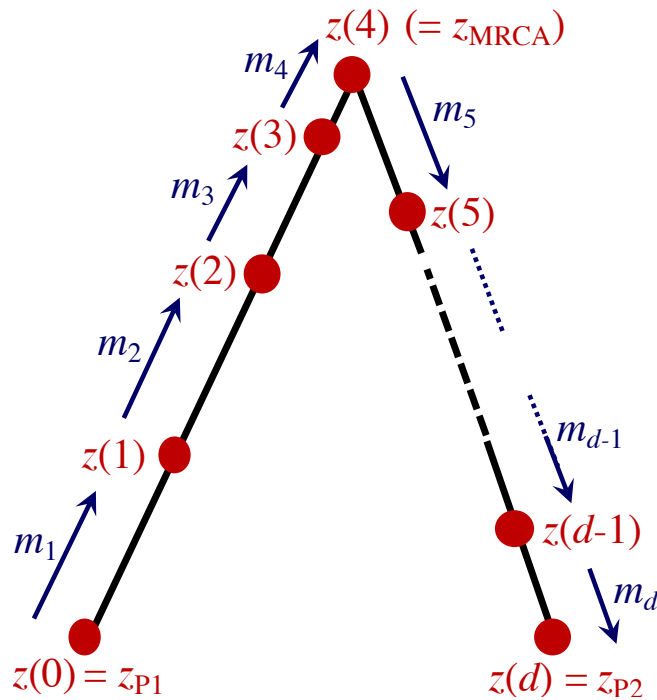


Figure S1: A schematic representation of a phenotype evolving over time in two populations, labeled P1 and P2, starting from their most recent common ancestor (MRCA). The populations differ by d substitutions. Each red dot represents the value of the phenotypic trait after each substitution, $z(i)$. The changes in phenotype, due to allelic substitutions, are denoted m_i (eq. 27), and are shown as blue arrows. These changes can be ordered in time, starting from the extant population P1, and going back in time to their common ancestor (here arbitrarily placed at $z(4)$), and then forward again in time to population P2. In haploids, the value of the hybrid index at each point in the chain, is the proportion of P2 alleles that the genotype contains, i.e., $h = i/d$ for phenotype $z(i)$. In real hybrids between P1 and P2, the substituted alleles may appear in random combinations due to segregation and recombination. The expected phenotype of such a hybrid can be found by randomizing the order of the m_i , and then locating the point in the chain with the required hybrid index, h . When their order is randomized, the chain of phenotypic changes can approximate a Brownian bridge between the parental phenotypes.

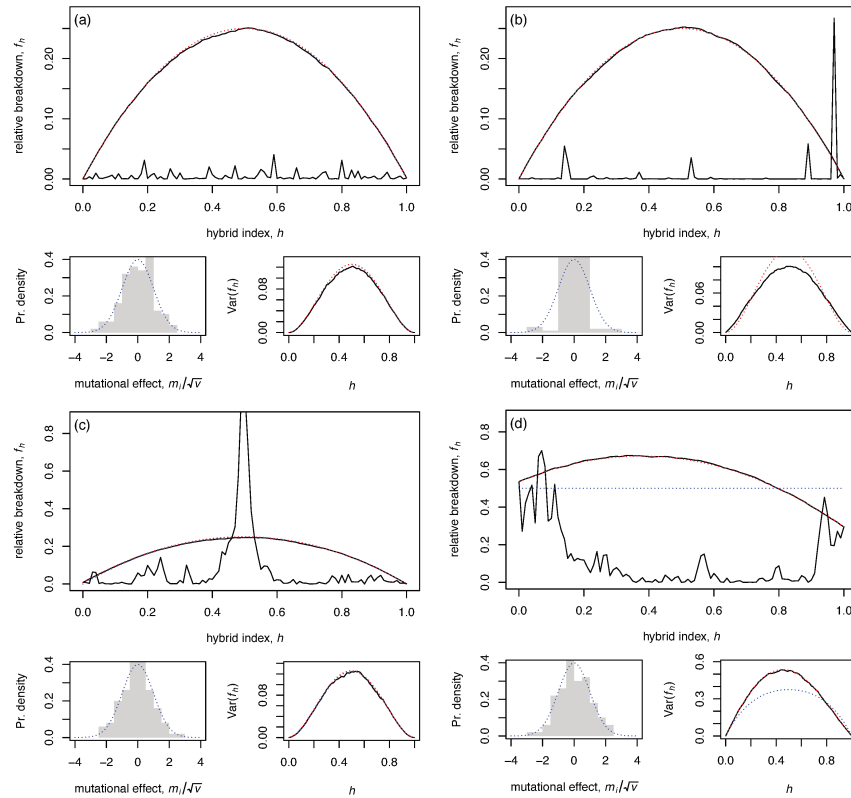


Figure S2: The breakdown associated with haploid hybrid genotypes, under simple models of phenotypic divergence. Each genotype is a hybrid between parental lines P1 and P2, which differ by d substitutions, and so a genotype with hybrid index h , carries $d(1 - h)$ alleles from P1, and dh alleles from P2. Each group of plots represents a single simulation of $d = 100$ substitutions, for which $d + 1$ trait values were generated at random. In the larger plots of each group, the black lines show the standardized breakdown score, which depends on the squared deviation of the trait from its optimal value: $f_h \equiv z^2(dh)/vd$. The jagged lines show the values when the hd P2 alleles were added in a fixed order, reflecting the chain of substitutions that connects P1 to P2 via their common ancestor (see Fig. S1). (This common ancestor carries the ancestral alleles from both lines, and will therefore be characterized by an intermediate value of h .) The smoother, black curves show results when the hd alleles were chosen at random, to simulate the generation of a hybrid by random segregation and recombination (curves show the average of 10,000 such randomizations). These results are very close to expectations for a Brownian bridge (see colored dotted lines). The analytical predictions are compared when we account correctly for the trait values of the parents (red dotted lines, which show eqs. 24-25), and when we use the approximation of eq. 26 with parental maladaptation set at its expected value (blue dotted lines). Smaller subplots show the effects of the $d = 100$ substitutions (eq. 27; left-hand smaller plots), and the variance in breakdown across the 10,000 randomizations (right-hand smaller plots; see eq. 23). Each group of plots shows a different set of assumptions about the original divergence of P1 and P2. In panel (a), we assumed that the ancestral populations always remained close to the optimum, by drawing the intermediate phenotypes from a normal distribution (see eq. 28). In panel (b) we violated the assumption of normality, by drawing the phenotypes from a Cauchy distribution with an undefined variance. This led to some mutations having very large effects, but made little difference to the accuracy of the Brownian bridge. In panel (c), we assumed that the ancestral population was distant from the optimum, and that the populations adapted in parallel. This was achieved by generating the phenotypes according to an Ornstein-Uhlenbeck process. Lastly, in panel (d) we assumed that populations diverged without any effective natural selection, by drawing the mutational effects from uncorrelated normal distributions. This situation would lead to $E(f_p) = 1/2$ and might apply to divergence under mutation accumulation, or strong inbreeding.

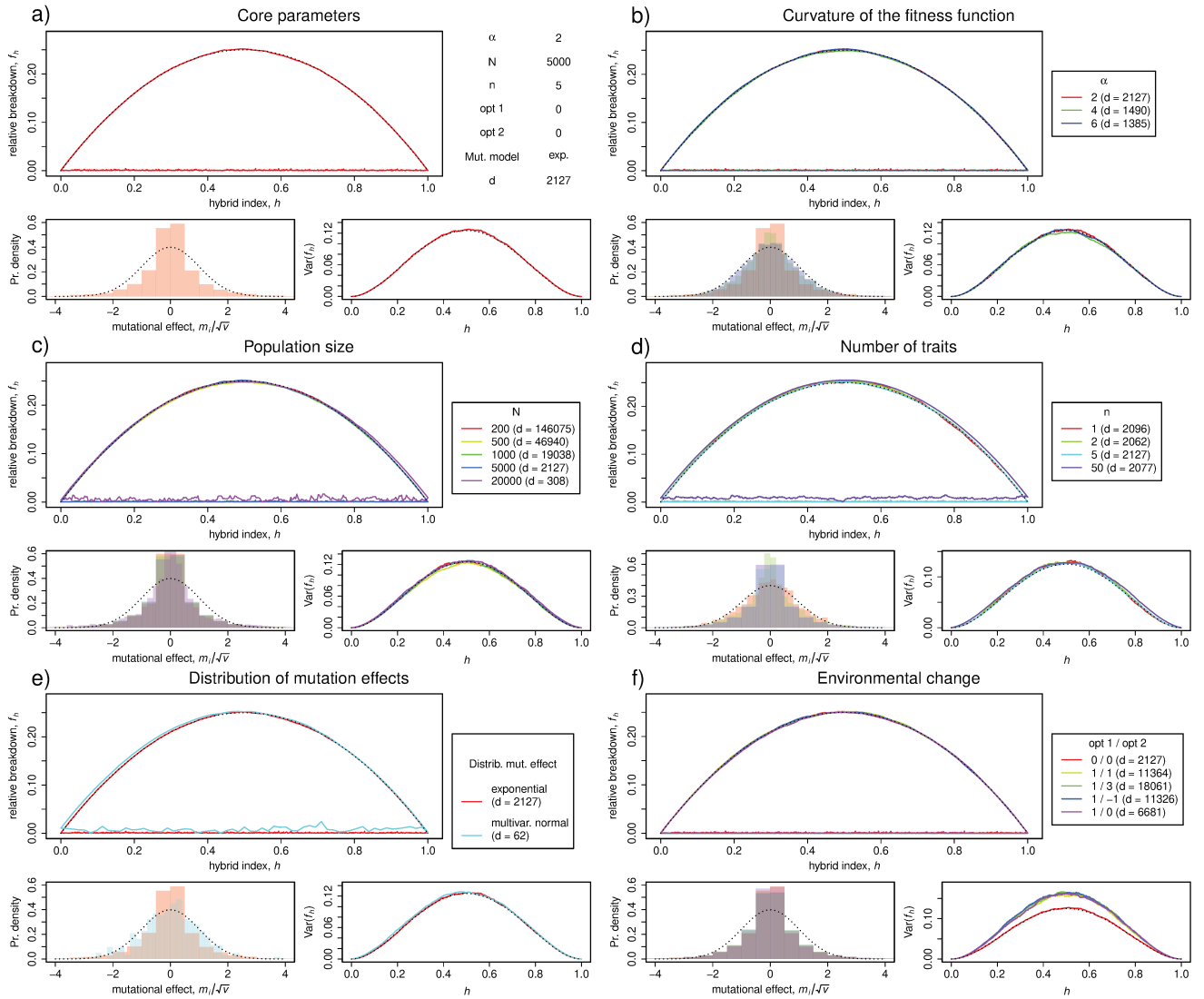


Figure S3: The breakdown associated with haploid hybrid genotypes, under explicit population genetic simulations of phenotypic divergence. The full simulation methods are described by Fraïsse et al. (2016b), but in each case, we assumed haploid genomics, and tracked the appearance, fixation or loss of mutations one at a time. Panel (a) shows results for a core parameter set. In subsequent panels, we varied the curvature of fitness function (α ; panel b; eq. 1); the size of each of the parental populations (N ; panel c), the number of phenotypic traits (n ; panel d; eq. 4); the shape of the distribution of mutational effects (exponential versus multivariate normal; panel e), and the model of environmental change (panel f). In particular, we moved the phenotypic optima for one of the n traits, according to a smooth sin wave function, which returned to its starting point 100 times over the course of a simulation. The optima tracked by the two populations sometimes differed, with the optimum tracked by P1 moving between 0 and opt1, and the optimum tracked by P2 between 0 and opt2. Each simulation comprised an expected 10^7 new mutations. We assumed that selection on all traits was isotropic (all $\lambda_i = 1$; eq. 4); the width of the mutational distribution was set such that the mean selection coefficient of a mutation appearing in an optimal genotype was -0.01, and the constant of proportionality in eq. 1 was set at 0.5. The number of mutations that fixed in each simulation (d) is indicated in parentheses in the legends. All other details match Figure S2.

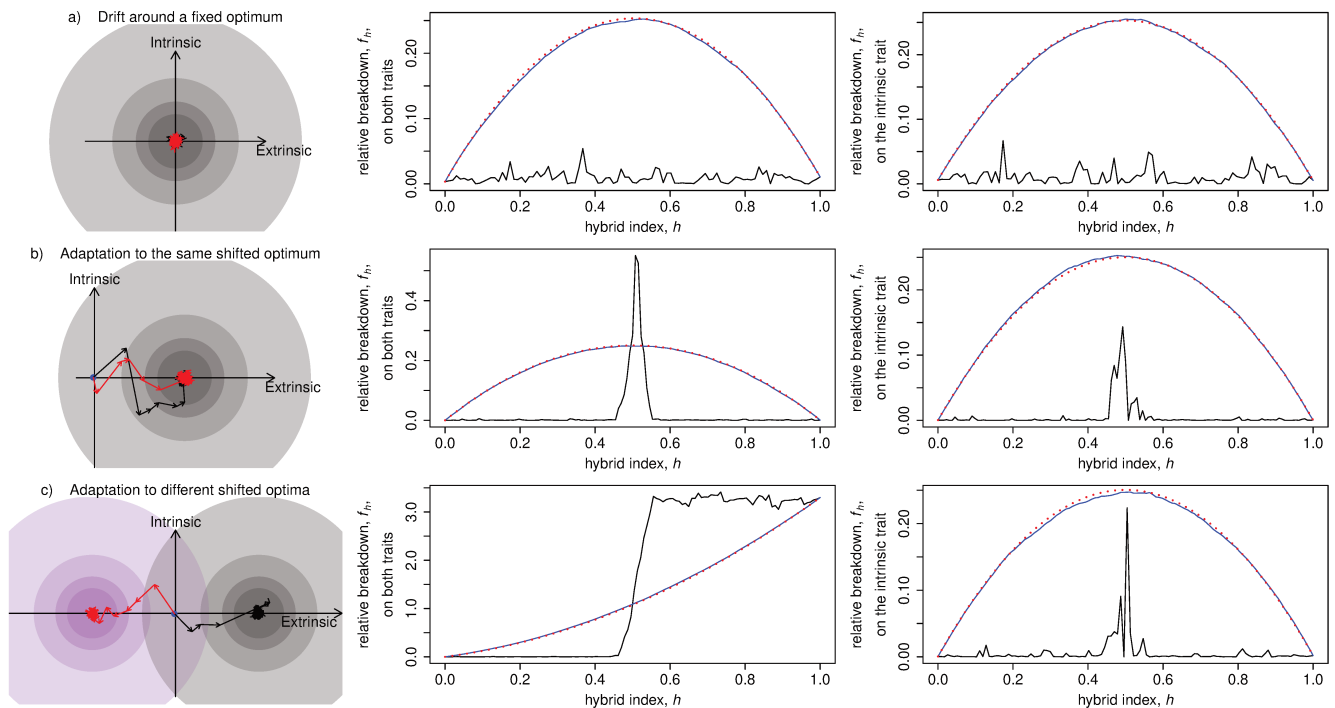


Figure S4: The breakdown associated with haploid hybrid genotypes, under explicit population genetic simulations of phenotypic divergence, in scenarios involving discrete jumps in the optimal value for one of $n = 2$ traits. This allows us to distinguish between intrinsic isolation (involving the maladaptation of traits whose optimum never moves), and extrinsic isolation (involving traits whose optimal value can vary between environments). Left-hand panels represent the fitness landscapes experienced by P1 and P2, with the initial optima found at the origin, and the concentric circles showing its final position, to which the parental populations adapt. Arrows show the phenotypic effects of each mutation that fixed during the divergence (red for P1 and black for P2). The central column shows the breakdown due to maladaptation in both traits, measured in the environment experienced by P1. The right-hand column shows breakdown due to maladaptation in one of the two traits. This is equivalent to measuring a single component of fitness, or to measuring total fitness in an environment where selection does not act on one of the traits. Row (a) corresponds to drift around a common fixed optimum for the two populations. Row (b) shows a scenario in which the optimum moves in the same way for both diverging populations. Row (c) shows a scenario in which the two populations adapt to different shifted optima, such that $f_{P2} \simeq 3$. Cases (a) and (b) correspond to mutation order speciation by drift or adaptation respectively. Case (c) represent an example of ecological speciation. We arbitrarily chose parameters so that the ancestral population suffered an approximately 12% drop in fitness after the environmental shift. All other simulation methods and details are as in Figure S3.

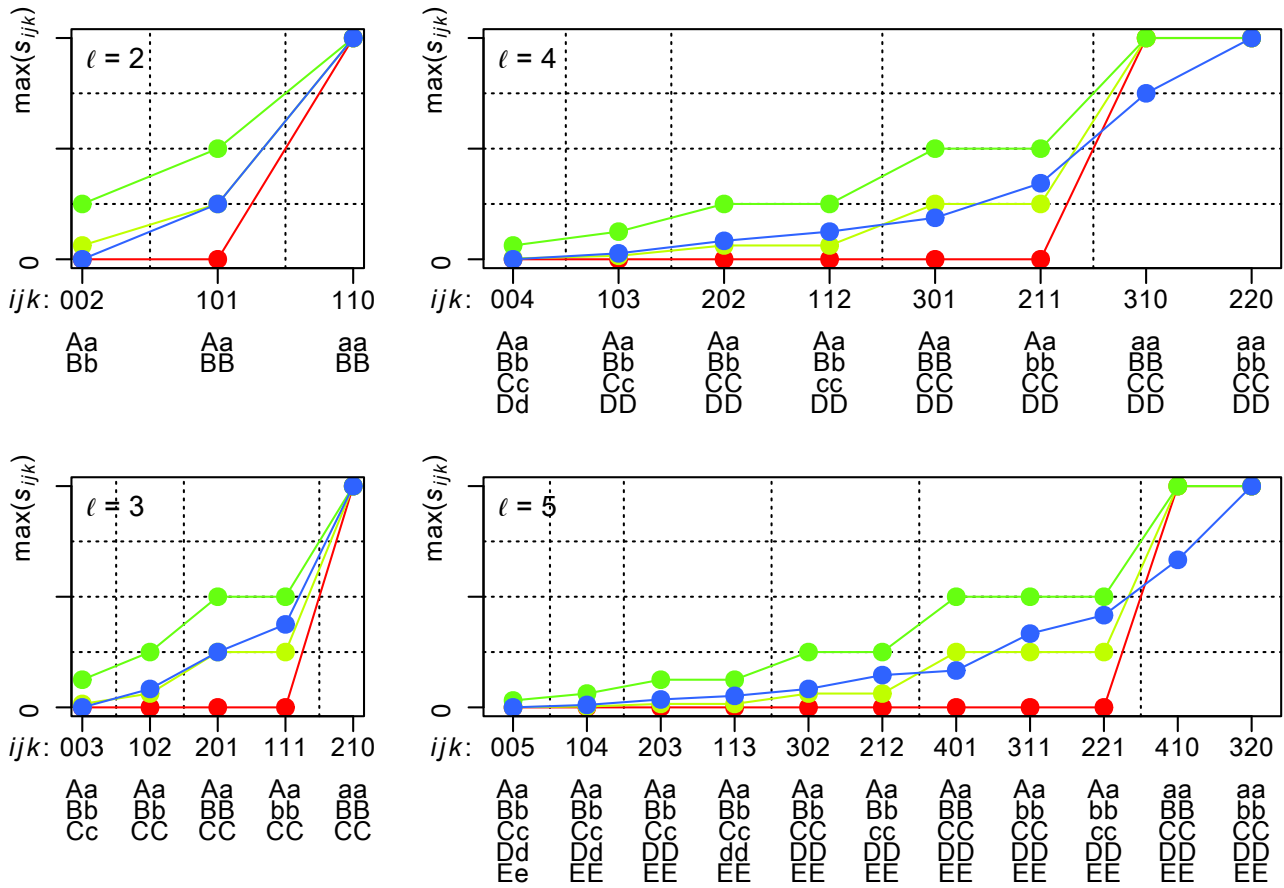


Figure S5: The effects of an incompatibility on hybrid breakdown score, s_{ijk} (eq. 41), when the incompatibility appears in a genotype comprising i loci that are homozygous for alleles from one parental species, j loci that are homozygous for alleles from the other parental species, and k loci that are heterozygous. The four panels show values for incompatibilities involving $\ell = 2, 3, 4$ and 5 loci. Values are compared when incompatible combinations have multiplicative effects (green points; eq. 42 with $\delta = 0$), when incompatibilities are wholly recessive (red points; eq. 42 with $\delta \rightarrow \infty$), and when they show partial recessivity at a level sufficient to generate Haldane's Rule (eq. 42 with $\delta = 2$; and treating hemizygous loci as equivalent to homozygous loci). Also shown are values assigned according to eq. 53 (blue points), which exactly reproduce the dependencies predicted by Fisher's geometric model with well adapted parents (eq. 9). Shown on the horizontal axis are the ijk values, and below, are example diploid genotypes, with upper and lower case letters used to distinguish alleles from the two parental lines, P1 and P2. Vertical dotted lines group genotypes with the same level of heterozygosity (i.e., the same value of k).

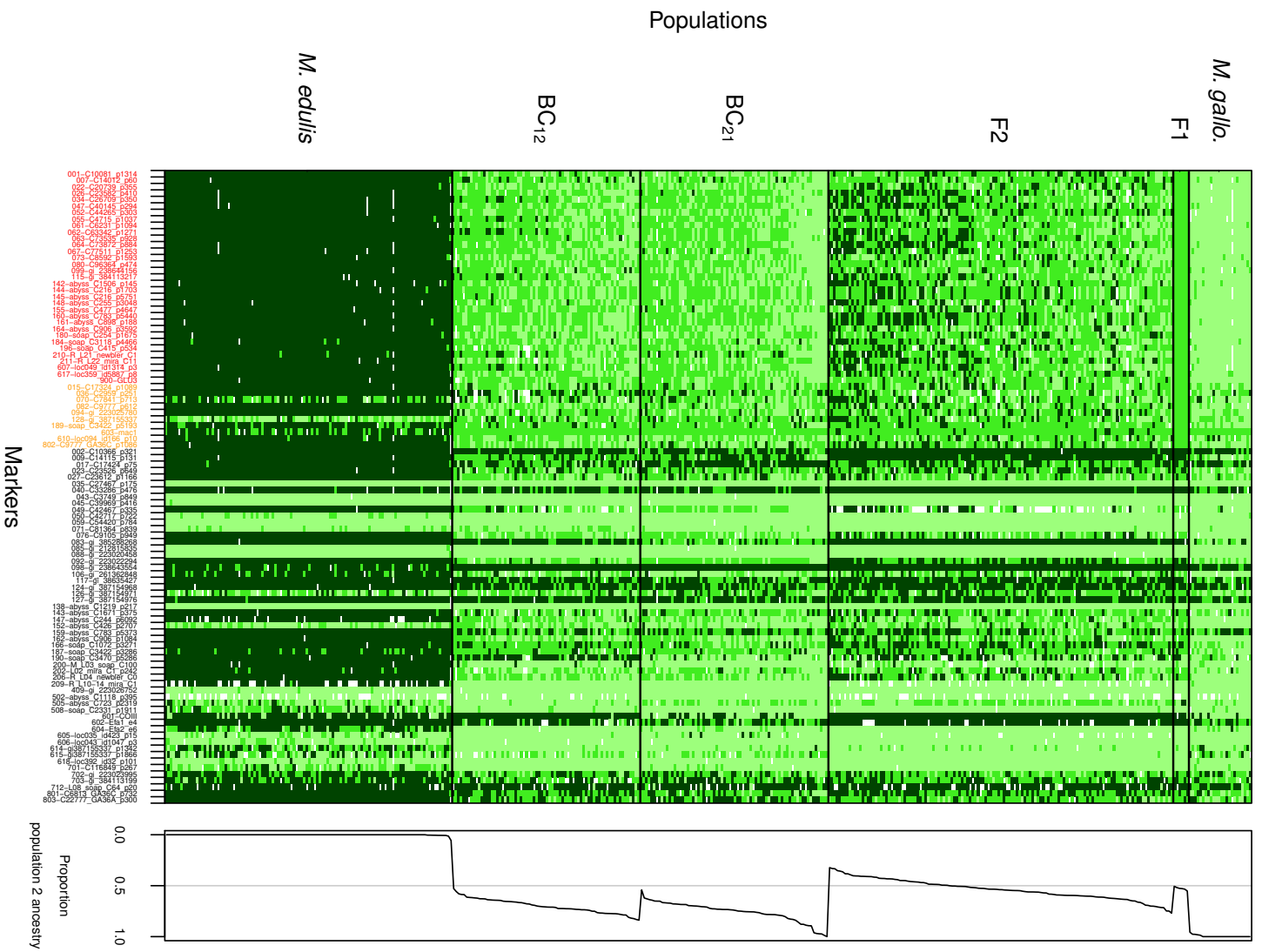


Figure S6: Genotype plot for the raw *Mytilus* data. This plot was generated with the R package *introgress* (v1.2.3; Gompert and Buerkle 2010). Columns represent markers, and rows represent individuals, labeled by cross type. Note that both backcrosses were made to *M. galloprovincialis*, and the subscripts differ according to whether the F1 was the female parent (BC₁₂), or the male parent (BC₂₁). Each colored rectangle represents the state of the genotype as homozygous for the *M. edulis* allele (dark green), homozygous for the *M. galloprovincialis* allele (light green), heterozygous (mid green), or missing data (white). The 33 markers that were strongly differentiated between the species are labeled in red, and the additional 10 that were scored as homozygous in all 6 F1 are labeled in orange. The right-hand-side plot represents the estimated proportion of *M. galloprovincialis* ancestry (population 2).

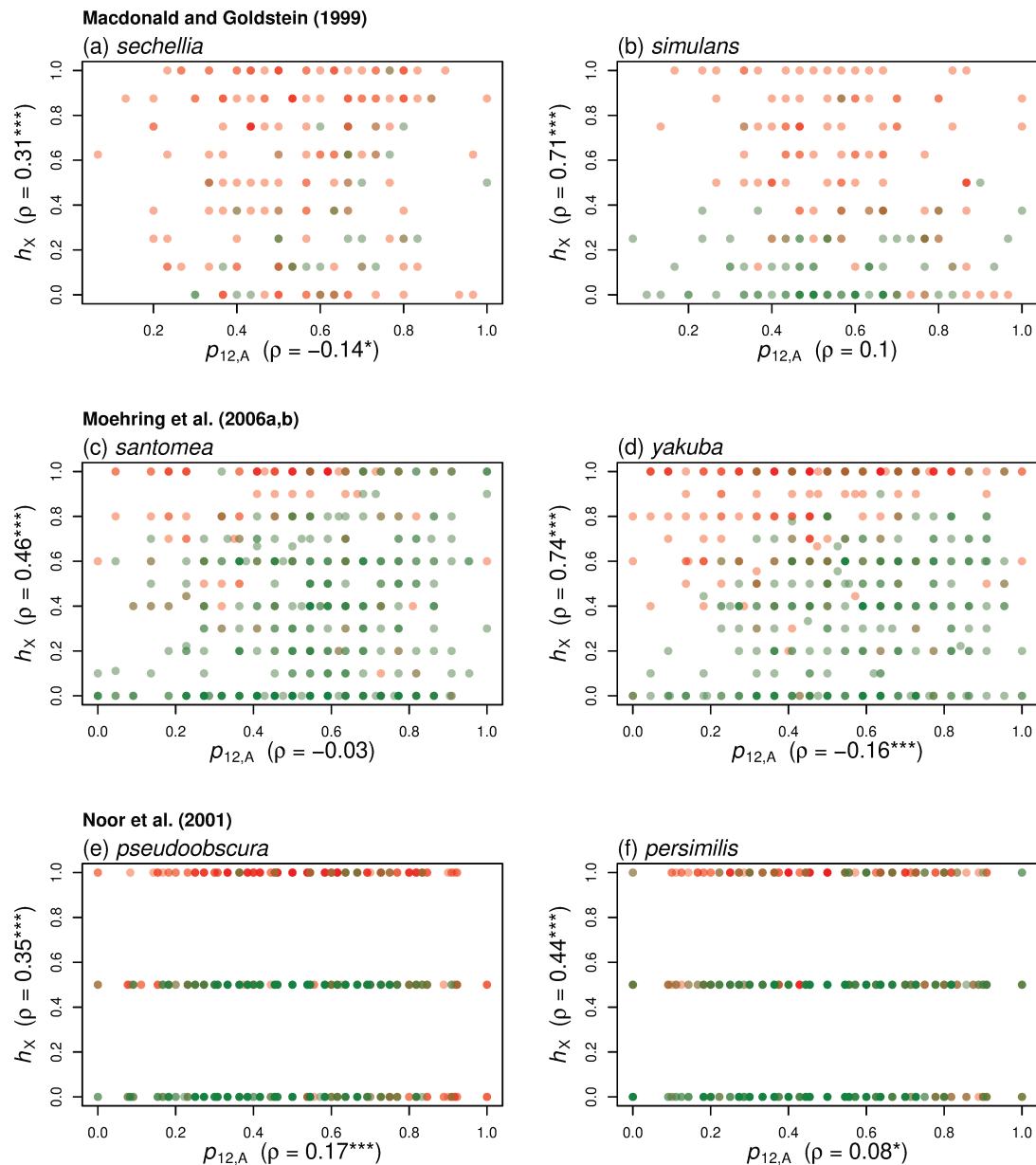


Figure S7: Plots of the *Drosophila* male backcross data reanalyzed here (see Table 1 and Moehring 2011). For all species pairs F1 σ had low fertility, consistent with Haldane’s Rule, and so male hybrids were derived from the backcross F1 φ \times P1 σ . Each point represents an individual hybrid, and is plotted as a function of its interspecies heterozygosity on the autosomes ($p_{12,A}$), and its hybrid index on the X (h_X), as estimated from a variable number of markers. Each row contains data from a reciprocal backcross, with the left-hand panel showing the backcross to the maternal species from the F1. Each individual was also scored for “sterility” on an ordinal scale, and red points show individuals with the highest possible sterility score (Panels a, b, e, and f), or falling within one of the two highest classes (Panels c-d). In parentheses, each axis also contains the results of correlating $p_{12,A}$ or h_X with sterility scores, replicating results of Moehring (2011). Shown are the Pearson’s correlation coefficients, ρ , and a measure of significance of a Spearman’s rank correlation (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$). In all cases, there was a strong tendency for h_X to correlate positively with sterility scores, while the correlations with $p_{12,A}$ show no consistent pattern. These findings are consistent with predictions from Fisher’s geometric model (see Fig. S8). Panels (e) and (f) show that h_X was estimated from two markers in this study, providing a natural division of the data into high- h_X and low- h_X individuals see Table 3.

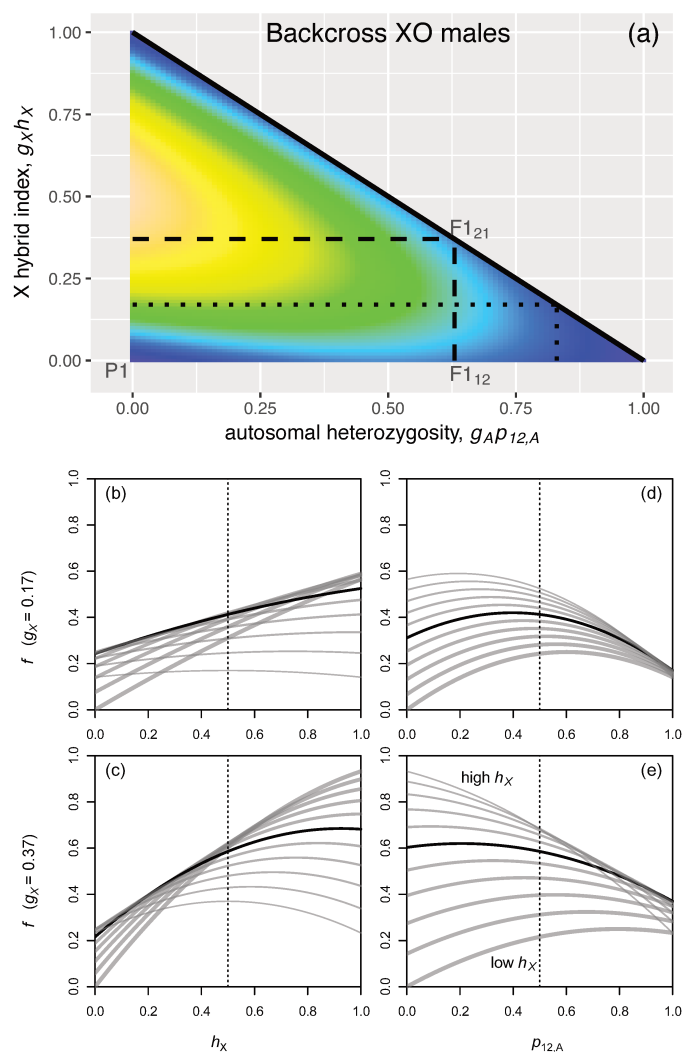


Figure S8: Predictions of Fisher's geometric model for heterogametic male hybrids. For simplicity, the predictions neglect any contributions from uniparentally-inherited loci other than the X, and assume that the parental types are well adapted. For concreteness, we assume XO sex determination, so that hybrids differ in their autosomal heterozygosity, $p_{12,A}$ and hybrid index of the X, h_X (i.e., for a backcross to P1, the proportion of divergent alleles on the X that come from P2). Panel (a) shows the fitness surface as a function of these two quantities. The dashed lines delimit the region that would apply to a species pair with $g_X = 0.17$ (as we have estimated for *Drosophila simulans/sechellia* and *D. santomea/yakuba*), and the dotted lines delimit the region that would apply to a species pair with $g_X = 0.37$ (as we have estimated for *Drosophila persimilis/pseudoobscura*). This region matches the parallelogram of Figure 1c. Panels (b)-(e) show slices through this fitness surface, with vertical dotted lines showing the Mendelian expectations for a first backcross. Panels (b) and (c) show the dependency on h_X . Results are shown when the autosomal heterozygosity is equal to its Mendelian expectation of $E(p_{12,A}) = \frac{1}{2}$ (black line), and over a range of values from $p_{12,A} = 0$ (thickest gray line) to $p_{12,A} = 1$ (thinnest gray line). Similarly, panels (d) and (e) show the dependency on autosomal heterozygosity, when h_X is at its expected value of $E(h_X) = \frac{1}{2}$ (black line), or over a range of values from $h_X = 0$ (thickest gray line) to $h_X = 1$ (thinnest gray line). Together, the plots show that Fisher's geometric model can account for the surprising results of Moehring (2011), and generate a new supported prediction (Table 3).

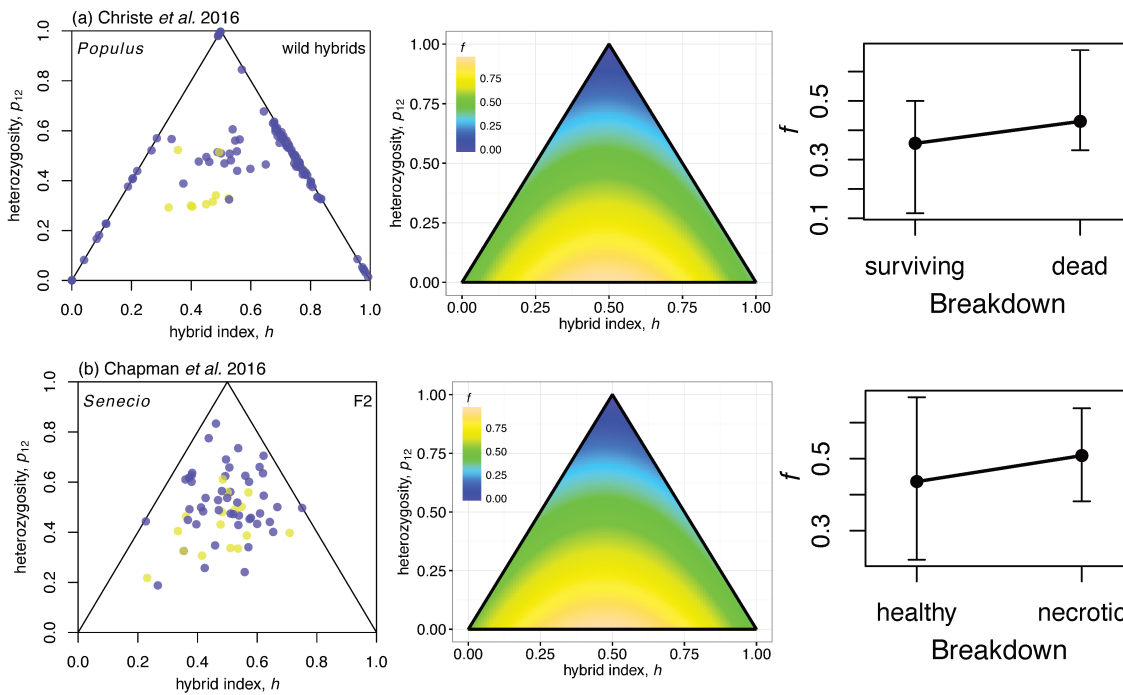


Figure S9: Estimation of the fitness surface for interspecific hybrids from plants (Table 1), namely wild hybrids of the genus *Populus* (row (a); Christe et al. 2016), and an F2 cross of the genus *Senecio* (row (b); Chapman et al. 2016). The left-hand panels show the data. Each point represents an individual hybrid, plotted as a function of their hybrid index, h and inter-species heterozygosity, p_{12} . Hybrid breakdown was scored as a binary scale for both data sets (Table 1), and the points are colored as low breakdown (blue) or high breakdown (yellow). The central panels show the fitness surfaces estimated from these data, using binary logistic regression, with predictors h , h^2 and p_{12} (see eq. 17, and Tables S5 and S6). These estimated surfaces resemble theoretical predictions with intermediate levels of parental maladaptation (compare Fig. 2a-b), which is consistent with evidence of F1 hybrid vigor in these taxa (*Populus*: Caseys et al. 2015; *Senecio*: Abbott and Brennan 2014). The right-hand panels show the fitted values of the predictor, for the high- and low-breakdown categories. Points and confidence intervals show the median and 90% quantiles.

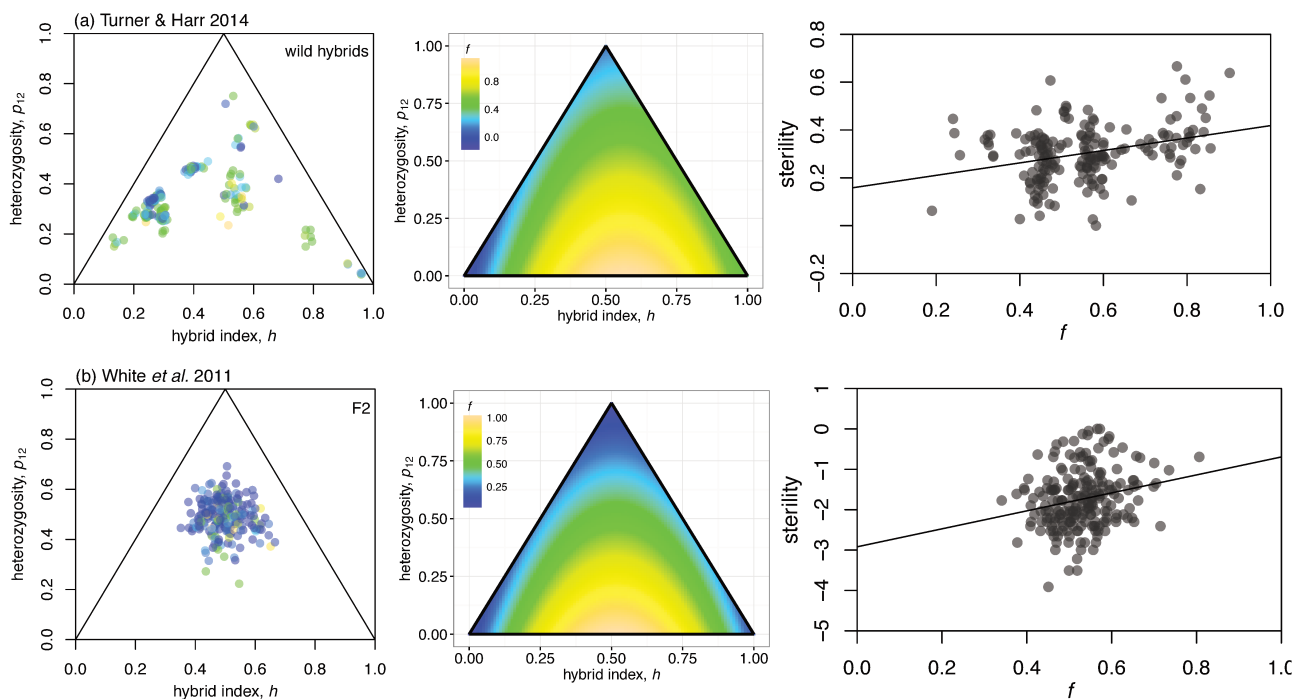


Figure S10: Estimation of the fitness surface for subspecific hybrids from *Mus musculus* (Table 1, White et al. 2011; Turner and Harr 2014). Because these males are heterogametic, to calculate the hybrid index, hemizygous alleles were treated as equivalent to homozygous alleles. Row (a) shows wild hybrids, with testes size as a proxy for fertility (Turner and Harr 2014), while row (b) shows an F2 cross, using the proportion of abnormal sperm (White et al. 2011). Since breakdown was scored on a continuous scale, the left-hand panels color each point with the color scheme of the fitness surfaces. The fitness surfaces were fit by standard least-squares regression, as shown in the right-hand panels. All other details match Supplementary Figure S9.

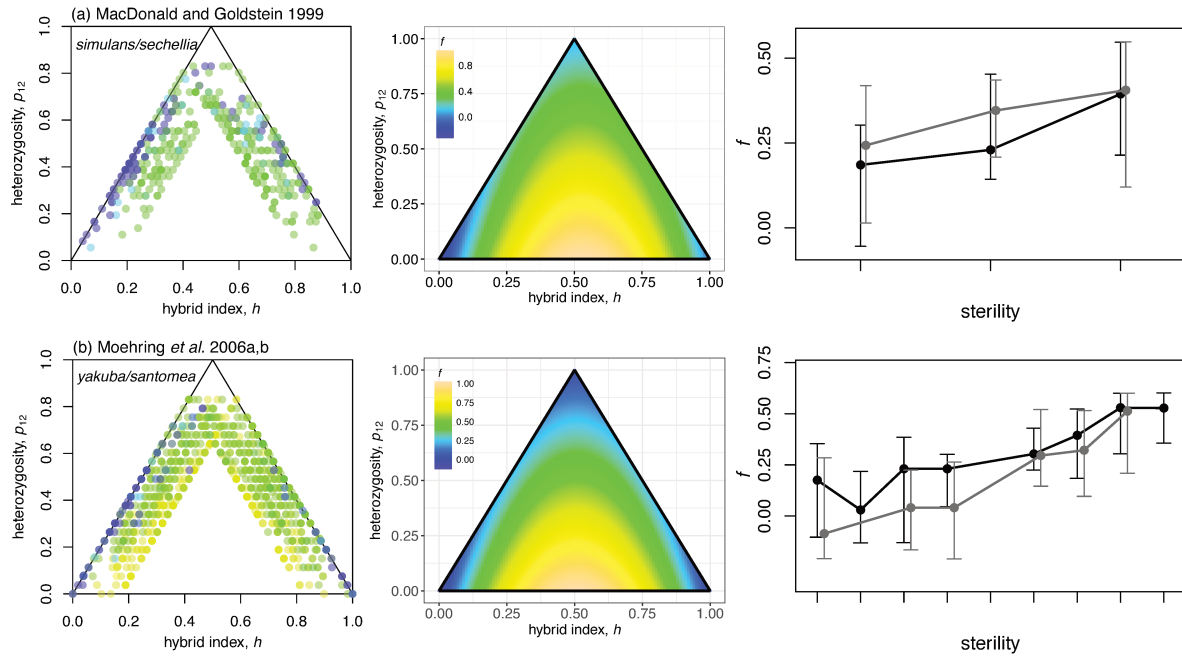


Figure S11: Estimation of the fitness surface for backcross male hybrids from *Drosophila* species pairs (Table 1, Macdonald and Goldstein 1999; Moehring et al. 2006a,b). For these data, the proxy for breakdown, male sterility, was scored on an ordinal scale, and so the models were fit with proportional odds logistic regression. The right-hand panels group the fitted values of the linear model by sterility category, and do so separately for the two backcross directions (black lines showing results for backcrosses to the maternal species from the F1, and gray lines show the backcross to the paternal species). This shows that the same patterns appear in both halves of the data. All other details match Supplementary Figures S9 and S10.

Table S1: The contribution of parental maladaptation to hybrid breakdown

Description	Condition	f_{mal}
(a) Exact result	-	$\frac{\sum_i \lambda_i ((1-h)z_{P1,i} + hz_{P2,i})^2}{d \sum_i \lambda_i v_i}$
(b) Independent maladaptation	$\overline{z_{P1,i} z_{P2,i}} = 0$	$(1-h)^2 f_{P1} + h^2 f_{P2}$
(c) Identical parental phenotypes	$z_{P1,i} = z_{P2,i}$	f_P
(d) Optimal midparent	$z_{P1,i} + z_{P2,i} = 0$	$(1-2h)^2 f_P$

Table S2: Checks for appropriateness for genomic data sets

	Hybridization	N	#markers	Cross	Fitness measure, w	p_0 : mean (sd)	ρ_{w,p_0} (p -val)	ρ_{p_{12},p_0} (p -val)	Reference
<i>Mytilus</i>	<i>edulis/galloprovincialis</i>	132	43	F2	-	0.034 (0.074)	-	0.244 (0.005**)	This study
		144	43	BC1	-	0.019 (0.045)	-	0.302 (0.0002***)	This study
<i>Senecio</i>	<i>aethnensis/chrysanthemifolius</i>	64	966	F2	Necrotic/Healthy	0.013 (0.014)	0.145 (0.253)	0.273 (0.029*)§	Chapman et al. 2016
<i>Mus musculus</i>	<i>musculus/domesticus</i>	185	14,220	WH	Testes weight	0.012 (0.006)	-0.099 (0.181)	0.039 (0.595)	Turner and Harr 2014
<i>Mus musculus</i>	<i>musculus/domesticus</i>	310	182	F2	Prop. abnormal sperm	0.030 (0.046)	0.056 (0.329)	-0.071 (0.215)	White et al. 2011
<i>Daphnia magna</i>	Germany/Finland	353	1324	F2	Fecundity ($N = 193$)	0.124 (0.092)	0.138 (0.056)	-0.206 (<10 ⁻⁴ **)	Routtu et al. 2014
<i>Medicago sativa</i>	<i>falcata/caerulea</i>	152	103	F2	Binary survival	0.158 (0.146)	-0.511 (<10 ⁻⁹ **)	-0.125 (0.124)	Li et al. 2011
<i>Saccharomyces</i>	<i>cerevisiae/paradoxus</i>	89	93	F2	Growth rate	0.080 (0.067)	-0.571 (<10 ⁻⁸ **)	-	Xu and He 2011

Notes: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. p_0 is the proportion of markers with missing data for each individual. For the data of White et al. (2011) this was calculated for autosomal markers only, while for the data of Xu and He (2011) it was defined as the proportion of loci scored as heterozygous, because for these haploid spores, heterozygotes must represent sequencing error, or aneuploidy. ρ and p -val are from Spearman's rank correlation tests. ρ_{w,p_0} is the correlation between fitness and the proportion of missing data; ρ_{p_{12},p_0} is the correlation between the proportion of sites with missing data, and the proportion scored as heterozygous. § this correlation is attributable to two outlying points. All other details match Table 1.

Table S3: The reference populations for *Mytilus* crosses

Species	Location	N	Latitude	Longitude
<i>Mytilus galloprovincialis</i>	Thau, France	22	43.412268	3.691788
<i>Mytilus edulis</i>	The Netherlands	22	53.31	5.424
	Saint-Jouin, Normandy, France	27	49.643953	0.152643
	Villerville, Normandy, France	29	49.399352	0.116054
	Réville, Normandy, France	29	49.5746667	-1.2293333333

Table S4: Expected breakdown scores for homogametic female hybrids with paternal X silencing

Hybrid	Cross ($\text{♀} \times \text{♂}$)	$E(p_{12})$	$E(p_1)$	$E(p_2)$	$E(h)$	$f; p_{12} = E(p_{12}), h = E(h)$
Parental	P1 \times P1	0	1	0	0	0
	P2 \times P2	0	0	1	1	f_{P2}
F1	F1 ₁₂ = P1 \times P2	$1 - \pi g_X$	πg_X	0	$\frac{1 - \pi g_X}{2}$	$\pi g_X - \pi^2 g_X^2 + \frac{(1 - \pi g_X)^2}{4} f_{P2}$
	F1 ₂₁ = P2 \times P1	$1 - \pi g_X$	0	πg_X	$\frac{1 + \pi g_X}{2}$	$\pi g_X - \pi^2 g_X^2 + \frac{(1 + \pi g_X)^2}{4} f_{P2}$
BC to P1	P1 \times F1 ₁₂	$\frac{1 - g_X}{2}$	$\frac{1 + g_X}{2}$	0	$\frac{1 - g_X}{4}$	$\frac{1 - g_X^2}{4} + \frac{(1 - g_X)^2}{16} f_{P2}$
	F1 ₁₂ \times P1	$\frac{1 - \pi g_X}{2}$	$\frac{1}{2}$	$\frac{\pi g_X}{2}$	$\frac{1 + \pi g_X}{4}$	$\frac{1 - \pi^2 g_X^2}{4} + \pi g_X + \frac{(1 + \pi g_X)^2}{16} f_{P2}$
	F1 ₂₁ \times P1	$\frac{1 - \pi g_X}{2}$	$\frac{1}{2}$	$\frac{\pi g_X}{2}$	$\frac{1 + \pi g_X}{4}$	$\frac{1 - \pi^2 g_X^2}{4} + \pi g_X + \frac{(1 + \pi g_X)^2}{16} f_{P2}$
	P1 \times F1 ₂₁	$\frac{1 - (2\pi - 1)g_X}{2}$	$\frac{1 + (2\pi - 1)g_X}{2}$	0	$\frac{1 - (2\pi - 1)g_X}{4}$	$\frac{1 - (2\pi - 1)^2 g_X^2}{4} + \frac{(1 - (2\pi - 1)g_X)^2}{16} f_{P2}$
BC to P2	P2 \times F1 ₂₁	$\frac{1 - g_X}{2}$	0	$\frac{1 + g_X}{2}$	$\frac{3 + g_X}{4}$	$\frac{1 - g_X^2}{4} + \frac{(3 + g_X)^2}{16} f_{P2}$
	F1 ₂₁ \times P2	$\frac{1 - \pi g_X}{2}$	$\frac{\pi g_X}{2}$	$\frac{1}{2}$	$\frac{3 - \pi g_X}{4}$	$\frac{1 - \pi^2 g_X^2}{4} + \pi g_X + \frac{(3 - \pi g_X)^2}{16} f_{P2}$
	F1 ₁₂ \times P2	$\frac{1 - \pi g_X}{2}$	$\frac{\pi g_X}{2}$	$\frac{1}{2}$	$\frac{3 - \pi g_X}{4}$	$\frac{1 - \pi^2 g_X^2}{4} + \pi g_X + \frac{(3 - \pi g_X)^2}{16} f_{P2}$
	P2 \times F1 ₁₂	$\frac{1 - (2\pi - 1)g_X}{2}$	0	$\frac{1 + (2\pi - 1)g_X}{2}$	$\frac{3 + 3(2\pi - 1)g_X}{4}$	$\frac{1 - (2\pi - 1)^2 g_X^2}{4} + \frac{(3 + (2\pi - 1)g_X)^2}{16} f_{P2}$

Notes: All expected breakdown scores were calculated from eq. 10, which assumes that parental species P1 is optimal, while P2 may be maladapted with breakdown $f_{P2} \geq 0$. $0 \leq g_X \leq 1$ denotes the proportion of the divergence that is located on the X chromosome; and $0 \leq \pi \leq 1$ denotes the proportion of the parental X chromosome that is silenced. The quantities p_{12} , p_1 , p_2 and h are calculated without considering silenced alleles on the paternal X, because these make no contribution to the phenotype.

Table S5: Inferring the hybrid fitness surface from genomic data sets

Hybridization	<i>N</i>	Response variable	β_0	β_1	β_2	r^2	AIC
<i>Populus alba/tremula</i>	137	Binary: Survival	[0]	-	-	[0]	104.75
			5.33	[4]	[1]	0.129	93.485
			12.44	2.229	[1]	0.225	85.647
			11.66	2.279	1.028	0.228	87.293
<i>Senecio F2 aethnensis/chrysanthemifolius</i>	64	Binary: Necrotic/Healthy	[0]	-	-	[0]	73.979
			6.13	[4]	[1]	0.0857	69.807
			6.54	1.797	[1]	0.0940	71.210
			6.42	1.963	1.065	0.0951	73.136
<i>Mus musculus ♂ musculus/domesticus</i>	185	-log(testes weight)	[0]	-	-	[0]	140.39
			0.82	[4]	[1]	0.0638	130.19
			0.75	4.425	[1]	0.0645	132.06
			0.83	4.622	0.888	0.1090	125.04
<i>Mus musculus F2♂ musculus/domesticus</i>	213	log(prop. abnormal sperm)	[0]	-	-	[0]	465.09
			2.22	[4]	[1]	0.0480	456.62
			2.23	3.430	[1]	0.0481	458.60
			2.23	3.933	0.968	0.0486	460.48
<i>Drosophila BC1♂ yakuba/santomea</i>	1099	Ordinal: Motile sperm	[0]	-	-	[0]	2702.84
			12.95	[4]	[1]	0.1928	2186.39
			12.24	4.347	[1]	0.1960	2179.90
			12.51	4.413	0.986	0.1986	2174.87
<i>Drosophila BC1♂ simulans/sechellia</i>	400	Ordinal: Sperm quantity	[0]	-	-	[0]	679.57
			13.24	[4]	[1]	0.1816	558.89
			11.68	5.006	[1]	0.2022	546.97
			11.45	5.314	0.945	0.2541	513.90

Notes: r^2 is the proportion of the variance explained, or McFadden's pseudo- r^2 for ordinal or binary data. AIC: Akaike Information Criterion, with favored model in bold. Parameter values that were not estimated, but fixed at their null or theoretically predicted values are shown in square brackets.

Table S6: The significance of individual regression coefficients

Hybridization	<i>N</i>	Response variable	intercept	<i>h</i>	<i>h</i> ²	<i>p</i> ₁₂	<i>r</i> ²	AIC
<i>Populus</i>	137	Binary: Survival	-2.63†	8.229**	-7.685**	-2.963**	-	-
			-1.831	26.564**	-27.301***	-11.600***	0.228	-
<i>Senecio</i> F2	64	Binary: Necrotic/Healthy	-0.863	12.606	-13.430	-6.422*	0.095	73.14
			1.714	-	-	-5.959*	0.088	69.62
<i>Mus musculus</i>	185	-log(testes weight)	-2.732	3.816***	-3.388***	-0.8256**	0.109	-
		1-(testes weight)/max(testes weight)	0.106	1.612***	-1.400***	-0.339*	0.106	-
<i>Mus musculus</i> F2	213	log(prop. abnormal sperm)	-2.881	8.749	-8.466	-2.225**	0.049	460.48
			-0.691	-	-	-2.150**	0.045	457.20
	212	-sqrt(density of normal sperm)	-3.992	10.159	-9.378	-3.404**	0.048	649.41
			-1.320	-	-	-3.348**	0.045	645.99
	305	log(prop. abnormal sperm)	-2.498	6.811	-6.533	-3.340***	0.037	655.00
			-0.773	-	-	-1.929**	0.035	652.76
<i>Drosophila yak/sant</i>	1099	Ordinal: Motile sperm	-	55.646***	-54.478***	-12.514***	0.199	-
<i>Drosophila sim/sech</i>	400	Ordinal: Sperm quantity	-	60.899***	-57.551***	-11.461***	0.254	-

Notes: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. r^2 is the proportion of the variance explained by the regression, or McFadden's pseudo- r^2 for binary or ordinal regressions. AIC: Akaike Information Criterion, used for comparing the fit of different sized models to the same data.

† Results reproduced from Christe et al. (2016), including "family" as a factor.

Table S7: Information on the 98 markers used for the *Mytilus* genotyping. Flanking sequences are provided in an annex file `Mytilus_markers_flanking_sequences.csv`. Contigs listed here can be obtained on the MytilusDB website (<http://www.scbi.uma.es/mytilus/index.php>, Fraïsse et al. 2016a), with the exception of TranscriptomeGalloConcat.

Marker name	SNP	Contig	Position
001-C10081_p1314	[A/T]	Contig10081_GA36A	1314
002-C10366_p321	[A/G]	Contig10366_GA36G	321
007-C14012_p60	[A/T]	Contig14012_GA36H	60
009-C14115_p131	[A/T]	Contig14115_GA36B	131
015-C17324_p1089	[A/G]	Contig17324_GA36A	1089
017-C17424_p75	[A/T]	Contig17424_GA36H	75
022-C20739_p355	[A/T]	Contig20739_GA36B	355
023-C23526_p649	[A/T]	Contig23526_GA36A	649
026-C23582_p410	[A/G]	Contig23582_GA36B	410
027-C23612_p1166	[C/T]	Contig23612_GA36D	1166
034-C26709_p350	[A/T]	Contig26709_GA36B	350
035-C27467_p175	[C/T]	Contig27467_GA36B	175
036-C2959_p251	[A/G]	Contig2959_GA36D	251
040-C33286_p476	[A/G]	Contig33286_GA36C	476
043-C3749_p849	[A/T]	Contig3749_GA36A	849
045-C39969_p416	[A/G]	Contig39969_GA36B	416
047-C40145_p294	[A/C]	Contig40145_GA36A	294
049-C42467_p335	[C/T]	Contig42467_GA36A	335
050-C42717_p722	[C/T]	Contig42717_GA36B	722
052-C44265_p303	[C/T]	Contig44265_GA36C	303
055-C4715_p1037	[A/C]	Contig4715_GA36B	1037
059-C54420_p784	[A/G]	Contig54420_GA36A	784
061-C6231_p1094	[A/C]	Contig6231_GA36B	1094
062-C63342_p1271	[A/T]	Contig63342_GA36B	1271
063-C73535_p928	[A/C]	Contig73535_GA36A	928
064-C73872_p884	[C/T]	Contig73872_GA36D	884
067-C77511_p1253	[A/C]	Contig77511_GA36A	1253
070-C7841_p713	[A/G]	Contig7841_GA36D	713
071-C81364_p839	[A/G]	Contig81364_GA36B	839
073-C8592_p1593	[A/T]	Contig8592_GA36D	1593
076-C9105_p949	[A/G]	Contig9105_GA36D	949
080-C96364_p474	[A/G]	Contig96364_GA36A	474
082-C9777_p612	[C/G]	Contig9777_GA36C	612
083-gi_385288268	[A/C]	gi_385288268_emb_Contig56466	1072
085-gi_212815835	[A/C]	gi_212815835_gb_GE749068	643
088-gi_223020458	[A/C]	gi_223020458_gb_FL493225	306
092-gi_223022294	[C/T]	gi_223022294_gb_FL494235	300
094-gi_223025780	[C/T]	gi_223025780_gb_FL490738	366
098-gi_238643554	[G/T]	gi_238643554_gb_Contig12266	389
099-gi_238644156	[A/C]	gi_238644156_gb_Contig35188	123
106-gi_261362848	[A/G]	gi_261362848_gb_M12A_cDNA4	139
115-gi_384113217	[C/T]	gi_384113217_emb_HE609130	794
117-gi_38635427	[C/G]	gi_38635427_emb_AJ609591	80
124-gi_387154968	[A/C]	gi_387154968_emb_HE609112	131
126-gi_387154971	[A/G]	gi_387154971_emb_Contig4866	1908
127-gi_387154976	[C/T]	gi_387154976_emb_HE609129	229
128-gi_387155337	[A/G]	gi_387155337_emb_HE610079	1120
138-abyss_C1219_p217	[A/T]	H_L1_abyss_Contig1219	2176
142-abyss_C1506_p145	[A/G]	H_L1_abyss_Contig1506	1455

143-abyss_C1671_p375	[A/G]	H_L1_abyss_Contig1671	3759
144-abyss_C216_p1703	[A/C]	H_L1_abyss_Contig216	1703
145-abyss_C216_p5751	[A/T]	H_L1_abyss_Contig216	5751
147-abyss_C244_p6092	[A/G]	H_L1_abyss_Contig244	6092
148-abyss_C255_p3048	[A/G]	H_L1_abyss_Contig255	3048
152-abyss_C426_p2707	[C/T]	H_L1_abyss_Contig426	2707
155-abyss_C477_p4647	[A/T]	H_L1_abyss_Contig477	4647
159-abyss_C783_p5373	[A/G]	H_L1_abyss_Contig783	5373
160-abyss_C783_p5440	[A/G]	H_L1_abyss_Contig783	5440
161-abyss_C898_p188	[C/T]	H_L1_abyss_Contig898	188
162-abyss_C906_p1084	[G/T]	H_L1_abyss_Contig906	1084
164-abyss_C906_p3592	[G/T]	H_L1_abyss_Contig906	3592
166-soap_C1072_p3271	[A/T]	H_L1_soap_Contig1072	3271
180-soap_C254_p1675	[G/T]	H_L1_soap_Contig254	1675
184-soap_C3118_p4466	[A/G]	H_L1_soap_Contig3118	4466
187-soap_C3422_p3286	[A/T]	H_L1_soap_Contig3422	3286
189-soap_C3422_p5193	[A/G]	H_L1_soap_Contig3422	5193
190-soap_C3470_p5286	[A/G]	H_L1_soap_Contig3470	5286
196-soap_C415_p534	[A/G]	H_L1_soap_Contig415	534
200-M_L03_soap_C100	[A/G]	M_L03_soap_Contig100	6168
202-L02_mira_C1_p242	[C/T]	R_L02_mira_Contig1	2420
206-R_L04_newbler_C0	[C/T]	R_L04_newbler_contig00002	5658
209-R_L10-14_mira_C1	[A/G]	R_L10-14_mira_Contig1	16260
210-R_L21_newbler_C1	[G/T]	R_L21_newbler_Contig1	18611
211-R_L22_mira_C11	[G/T]	R_L22_mira_Contig11	5294
409-gi_223026752	[A/G]	gi_223026752_gb_Contig46241	515
502-abyss_C1118_p395	[T/C]	H_L1_abyss_Contig1118	3954
505-abyss_C723_p2319	[T/C]	H_L1_abyss_Contig723	2319
508-soap_C2331_p1911	[A/G]	H_L1_soap_Contig2331	1911
601-COIII	[C/T]	mitochondria	-
602-Efa1_e4	[A/T]	R_L02_mira_Contig1	6232
603-mac1	[C/T]	gi_5114427_gb_Port34a	2372
604-Efa2_e6	[A/C]	R_L02_mira_Contig1	6232
605-loc035_id423_p15	[T/A]	TranscriptomeGalloConcat	1593146
606-loc043_id1047_p3	[G/A]	TranscriptomeGalloConcat	3639171
607-loc049_id1314_p3	[T/C]	TranscriptomeGalloConcat	3833043
610-loc094_id166_p10	[A/G]	TranscriptomeGalloConcat	10850806
614-gi387155337_p1342	[GGAA/TGAT]	gi_387155337_emb_HE610079	1342
615-gi387155337_p1866	[A/G]	gi_387155337_emb_HE610079	1866
617-loc359_id5887_p8	[T/C]	TranscriptomeGalloConcat	89998577
618-loc392_id32_p101	[C/T]	TranscriptomeGalloConcat	101064357
701-C116849_p267	[A/T]	Contig116849_GA36C	267
702-gi_223023995	[A/T]	gi_223023995_gb_FL493336	177
703-gi_384113199	[A/C]	gi_384113199_emb_Contig12194	4203
712-L08_soap_C64_p20	[A/G]	M_L08_soap_Contig64	20
801-C6813_GA36C_p732	[C/T]	Contig6813_GA36C	732
802-C9777_GA36C_p1086	[C/T]	Contig9777_GA36C	1086
803-C22777_GA36A_p300	[C/T]	Contig22777_GA36A	300
900-GLU3	[CAATATTAAGTG/ TCATATAAAGCT]	gi_961463_dbj_D63778	2261

References

- Macdonald, S. J. and D. B. Goldstein (1999). "A Quantitative Genetic Analysis of Male Sexual Traits Distinguishing the Sibling Species *Drosophila Simulans* and *D. Sechellia*". *Genetics* 153.4, pp. 1683–1699.
- Moehring, A. J., A. Llopart, S. Elwyn, J. A. Coyne, and T. F. C. Mackay (Mar. 17, 2006a). "The Genetic Basis of Postzygotic Reproductive Isolation Between *Drosophila Santomea* and *D. Yakuba* Due to Hybrid Male Sterility". *Genetics* 173.1, pp. 225–233.
- (Mar. 17, 2006b). "The Genetic Basis of Prezygotic Reproductive Isolation Between *Drosophila Santomea* and *D. Yakuba* Due to Mating Preference". *Genetics* 173.1, pp. 215–223.
- Gompert, Z. and C. A. Buerkle (2010). "Introgress: A Software Package for Mapping Components of Isolation in Hybrids". *Mol. Ecol. Resour.* 10.2, pp. 378–384.
- Li, X., X. Wang, Y. Wei, and E. C. Brummer (Aug. 2011). "Prevalence of Segregation Distortion in Diploid Alfalfa and Its Implications for Genetics and Breeding Applications". *Theor. Appl. Genet.* 123.4, pp. 667–679.
- Moehring, A. J. (2011). "Heterozygosity and Its Unexpected Correlations with Hybrid Sterility". *Evolution* 65.9, pp. 2621–2630.
- White, M. A., B. Steffy, T. Wiltshire, and B. A. Payseur (Sept. 1, 2011). "Genetic Dissection of a Key Reproductive Barrier Between Nascent Species of House Mice". *Genetics* 189.1, pp. 289–304.
- Xu, M. and X. He (2011). "Genetic Incompatibility Dampens Hybrid Fertility More Than Hybrid Viability: Yeast as a Case Study". *PLoS ONE* 6.4, e18341.
- Abbott, R. J. and A. C. Brennan (June 23, 2014). "Altitudinal Gradients, Plant Hybrid Zones and Evolutionary Novelty". *Philos. Trans. R. Soc. B Biol. Sci.* 369.1648, p. 20130346.
- Routtu, J. et al. (2014). "An SNP-Based Second-Generation Genetic Map of *Daphnia Magna* and Its Application to QTL Analysis of Phenotypic Traits". *BMC Genomics* 15, p. 1033.
- Turner, L. M. and B. Harr (2014). "Genome-Wide Mapping in a House Mouse Hybrid Zone Reveals Hybrid Sterility Loci and Dobzhansky-Muller Interactions". *Elife* 3, e02504.
- Casays, C., C. Stritt, G. Glauser, T. Blanchard, and C. Lexer (May 26, 2015). "Effects of Hybridization and Evolutionary Constraints on Secondary Metabolites: The Genetic Architecture of Phenylpropanoids in European *Populus* Species". *PLoS ONE* 10.5, e0128200.
- Chapman, M. A., S. J. Hiscock, and D. A. Filatov (Jan. 2016). "The Genomic Bases of Morphological Divergence and Reproductive Isolation Driven by Ecological Speciation in *Senecio* (Asteraceae)". *J. Evol. Biol.* 29.1, pp. 98–113.
- Christe, C. et al. (2016). "Selection against Recombinant Hybrids Maintains Reproductive Isolation in Hybridizing *Populus* Species despite F₁ Fertility and Recurrent Gene Flow". *Mol. Ecol.* 25.11, pp. 2482–2498.
- Fraïsse, C., K. Belkhir, J. J. Welch, and N. Bierne (2016a). "Local Interspecies Introgression Is the Main Cause of Extreme Levels of Intraspecific Differentiation in Mussels". *Mol. Ecol.* 25.1, pp. 269–286.
- Fraïsse, C., P. A. Gunnarsson, D. Roze, N. Bierne, and J. J. Welch (2016b). "The Genetics of Speciation: Insights from Fisher's Geometric Model". *Evolution* 70.7, pp. 1450–1464.

Chapitre 5

Discussion générale et perspectives

5.1 La continuité du complexe d'espèces *Mytilus*

Cette thèse a permis de produire un jeu de données sans précédent sur le complexe d'espèces *M. edulis* en hémisphère nord. L'ensemble des chapitres, ainsi que quelques projets annexes, ont permis de génotyper 12 364 individus sur des SNPs informatifs. Parmi ces individus, 144 ont été sélectionnés pour être séquencés (12 génomes publiés ont pu y être ajoutés Gerdol et al. 2019).

Ce gros jeu de données, en terme de nombre d'individus analysés et de répartition géographique, donne une image nouvelle à ce complexe d'espèces marines, que l'on comprend seulement en regardant l'ensemble des données (Figure 5.1). L'augmentation de l'effort d'échantillonnage, surtout pour les espèces *M. galloprovincialis* et *M. edulis*, nous fait réaliser qu'il existe une continuité entre les différentes lignées du complexe. Bien que l'échantillonnage soit limité entre certaines populations, surtout pour *M. trossulus* qui n'était pas l'espèce centrale de cette thèse, nous pouvons penser que les transitions n'y sont pas moins continues. Le complexe d'espèces *M. edulis* représente un ensemble de points dans le processus de spéciation, puisqu'il comprend plusieurs étapes, plusieurs niveaux de différenciation et d'isolement reproductif : le niveau de l'espèce, les différenciations entre bassins océaniques et les différenciations populationnelles au sein de chaque lignée.

Cette conclusion semble présenter un paradoxe entre l'observation de zones d'hybridation qui paraissent stables et peu perméables entre les différentes espèces de *Mytilus* spp., et une continuité extrêmement importante entre les clusters de génotypes. Qu'est-ce qui maintient alors des entités spécifiques au sein du complexe ? Peut-être que la résolution de ce paradoxe réside dans le fait qu'il existe une multitude de scénarios possibles à la rencontre entre deux lignées divergentes, et que la continuité observée est le fruit de la contingence des contacts, remplissant le spectre des résultats d'admixture en différents endroits.

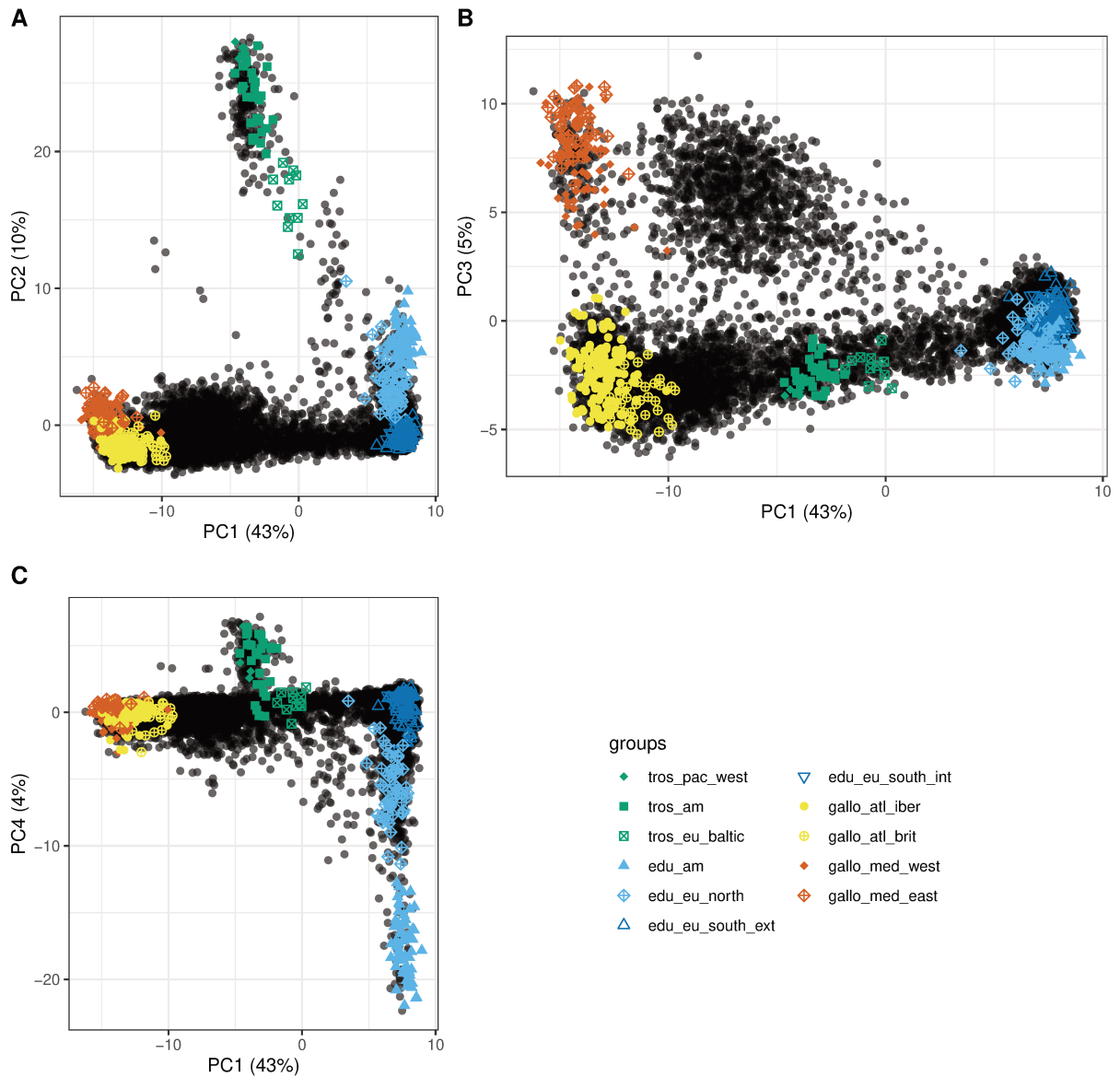


Figure 5.1 : Analyse en composantes principales de la base de données des génotypes produits durant la thèse. Correspondance des groupes – tros_pac_west : *M. trossulus* du Pacifique ouest ; tros_am : *M. trossulus* de l'Atlantique nord-ouest ; tros_eu_baltic : *M. trossulus* de la mer Baltique ; edu_am : *M. edulis* de l'Atlantique nord-ouest ; edu_eu_north : *M. edulis* nord européenne ; edu_eu_south : *M. edulis* sud européenne (int et ext pour interne et externe à la zone d'hybridation mosaïque) ; gallo_atl_iber : *M. galloprovincialis* atlantique de la côte ibérique ; gallo_atl_brit : *M. galloprovincialis* atlantique de la côte bretonne ; gallo_med_west : *M. galloprovincialis* méditerranéenne ouest ; gallo_med_east : *M. galloprovincialis* méditerranéenne est.

5.2 Les moules des docks

5.2.1 Une invasion en devenir ou est-ce déjà le cas ?

Cette thèse se veut être un point de référence dans l'étude de l'introduction des moules des docks. De nombreuses connaissances restent encore à recueillir sur ce modèle, puisque notre travail n'a exploré que les volets de la génétique et des distributions spatiales. Malgré cela, s'il fallait classer les moules des docks dans la typologie des espèces invasives avec nos connaissances actuelles (Figure 4 ; Blackburn et al. 2011), le niveau minimal serait C3, i.e. que « les individus survivent dans la nature au point d'introduction, se reproduisent et forment une population auto-suffisante ». De plus, si l'on considère les quelques individus portuaires retrouvés à l'extérieur des ports, l'établissement d'une population dans l'estuaire de l'Élorn dans la rade de Brest, ainsi que les différents ports constituant sûrement déjà des dispersions secondaires, le statut pourrait être davantage supérieur et pourrait aller jusqu'au niveau maximum E, i.e. « une espèce complètement invasive, avec des individus dispersant, survivant et se reproduisant dans des sites multiples avec un spectre plus ou moins important d'habitats et aires d'observation ».

Le point concernant la survie dans la nature peut ici être débattu. Blackburn et al. (2011) utilisent la dispersion et la reproduction dans la nature (terme *wild* utilisé) dans plusieurs niveaux de leur classification. Concernant les moules des docks, la grande majorité des individus ont été échantillonnés dans des milieux anthropisés ou sur des structures anthropiques en dehors des ports. Il convient donc de se demander à la fois si la classification inclue les milieux anthropisés comme les ports dans la catégorie "naturels" (i.e. non artificiels) et si les moules des docks peuvent occuper des niches du milieu naturel. Le caractère admixé des moules des docks nous renvoie à E. Anderson (1948) qui notait que les perturbations environnementales ouvraient un ensemble de niches qui pourraient être plus adaptées aux hybrides qu'à leurs parents. Dans certains cas, Ellstrand et Schierenbeck (2000) trouvent que les hybrides sont presque exclusivement présents dans des milieux ayant subi des perturbations anthropiques.

Les moules des docks ont-elles un avantage dans un environnement anthropisé ? Sont-elles adaptées à cet environnement ? Difficile de répondre sans des expériences de transplantation à la fois de moules des docks en dehors des ports (qui poseraient des problèmes éthiques) mais aussi de moules natives dans les ports. Les populations qui pourraient être utilisées pour cette question se trouvent dans la rade de Brest, puisque l'on trouve à la fois des moules natives et des moules des docks en sympatrie dans les milieux anthropisés, mais également des moules des docks en dehors du port commercial de Brest et des marinas. L'avant port du Havre est aussi une zone de co-existence des moules des docks avec les natives. D'après J. L. Lockwood et al. (2013, chapitre 5), contrairement à ce qui est souvent admis en biologie des invasions, les perturbations anthropiques ne font pas forcément pencher la balance en faveur des espèces non natives. Cependant, ils postulent que dans un tout nouvel environnement anthropisé jamais rencontré par les espèces natives, des espèces introduites peuvent avoir un

avantage sélectif. Dans le cas des moules, il est évident que la colonisation des ports n'est pas une caractéristique unique aux moules des docks puisqu'on y retrouve des individus natifs, et que certains ports non impactés (e.g. La Rochelle) sont colonisés par *M. edulis*. Malgré cela, il reste difficile de savoir si les ports envahis comme Cherbourg ou Le Havre abritaient des moules natives avant l'arrivée des moules des docks.

Ainsi, nous réalisons qu'il nous manque un volet écologique sur les moules des docks, mais également sur les environnements portuaires en général, afin de tirer plus de conclusions. Par exemple, il n'est pas pertinent d'utiliser un modèle de niche qui serait produit sur des données écologiques de l'une des deux espèces parentales (*M. edulis* ou *M. galloprovincialis*), puisque les traits liés à l'écologie pourraient avoir été largement modifiés dans les populations admixées des ports. La cartographie précise des environnements colonisés par les moules introduites sera donc nécessaire, en parallèle d'une comparaison des traits physiologiques entre les espèces parentales et les nouvelles populations portuaires.

Afin de comprendre le processus d'invasion, il faudra étendre l'échantillonnage à la fois à des ports plus petits (e.g. Saint-Malo) mais aussi à d'autres grands ports de commerce pouvant représenter de nouvelles sources (e.g. Rotterdam, puisque des migrants hybrides ont été retrouvés non loin). L'augmentation de l'effort d'échantillonnage devra passer par l'emploi de nouvelles méthodes impliquant à la fois l'usage de marqueurs caractéristiques, mais aussi de techniques utilisant l'ADN environnemental. Les techniques d'ADN environnemental sont de plus en plus développées dans le domaine de la biologie des invasions et semblent prometteuses. Elles peuvent être soit ciblées en utilisant par exemple des PCR quantitatives (Wood et al. 2019), soit beaucoup plus générales en utilisant des techniques de metabarcoding (Comtet et al. 2015; Holman et al. 2019). Dans la partie 2.1.2, je conclus qu'il est particulièrement difficile de différencier les lignées du complexe *M. edulis* sur la base de metabarcoding, la méthode la plus prometteuse est donc l'utilisation de marqueurs multiples diagnostiques et ciblés, permettant d'avoir des seuils de détection bas et une précision accrue.

D'où viennent les moules des docks? Tout d'abord l'admixture à l'origine de ces populations semble assez récente à l'échelle temporelle du transport maritime humain. Dans le chapitre 2 nous estimons l'admixture de quelques années à une vingtaine, mais la présence d'une population de moules des docks il y a 40 ans à Swansea (UK) indique un âge un peu plus important. L'absence d'individus parentaux *M. galloprovincialis* méditerranéens dans l'ensemble des ports échantillonnés indique soit que l'admixture initiale ne s'est pas faite dans ces ports, soit que les parents ont disparus depuis, ou encore que l'effort d'échantillonnage n'était pas assez grand. Les moules des docks partagent très certainement une histoire commune et le scénario le plus parcimonieux repose sur une admixture initiale dans un endroit encore non identifié, puis une dispersion entre ports via le trafic maritime.

Une question particulièrement intéressante à explorer est de savoir si l'hybridation avec les moules natives, en formant des zones de tension à l'entrée des ports, est un mécanisme qui retarde ou empêche l'invasion du milieu extérieur. En effet, si l'on considère que l'espèce

M. galloprovincialis a une valeur sélective intrinsèque supérieur à *M. edulis* (Bierne et al. 2006), l'introduction de la lignée méditerranéenne aurait très bien pu déboucher sur une invasion importante du milieu naturel comme il a été observé dans de multiples régions du monde (cf. partie i.3.5). Il serait donc intéressant de pouvoir explorer les différents scénarios de manière théorique.

5.2.2 Celle dont il ne faut pas prononcer le nom : l'espèce hybride ?

En dehors de leur spécificité écologique d'être restreinte dans les environnements portuaires ou estuariens, les moules des docks présentent aussi la caractéristique d'être une population (à l'échelle de l'ensemble des ports) homogène génétiquement. Cette homogénéité d'admixture, le maintien d'un cluster propre aux populations portuaires et la formation de clines à l'entrée des ports nous font nous demander si les moules des docks seraient un exemple d'espèce hybride homoploïde.

La spéciation hybride homoploïde (ou « recombinoire ») est définie comme l'émergence d'une nouvelle lignée produite par la reproduction sans polyploïdisation entre deux lignées parentales divergentes. Cette nouvelle lignée, possédant la même ploïdie que les parents, est reproductivement isolée des lignées parentales (Mallet 2007 ; Abbott et al. 2013 ; Elgvin et al. 2017).

Au sens de Mallet (2007), les espèces sont des clusters génotypiques multilocus. Une espèce hybride représente alors un cluster intermédiaire entre deux clusters parentaux, qui est stable et reste distinct malgré un contact avec les deux formes parentales. Selon Abbott et al. (2013), la combinaison des allèles parentaux doit contribuer à l'établissement d'une nouvelle population et au maintien de son originalité malgré le flux de gènes provenant des parents. On peut facilement imaginer que les espèces hybrides pourraient représenter toute une continuité dans les proportions génomiques provenant des deux parents, avec d'un côté l'introggression de quelques gènes et de l'autre une contribution symétrique. Cependant qu'est-ce qui distingue ce processus de l'introggression ? Le point le plus important ici est que la combinaison des deux génomes parentaux doit contribuer à l'isolement reproductif afin d'obtenir une lignée stable et isolée des deux parents (Schumer et al. 2014b, 2018a).

Un exemple commun dans la littérature de la spéciation hybride liée à l'introggression de quelques gènes est le cas des papillons *Heliconius*. Par exemple, *H. elevatus* a un génome provenant majoritairement d'un parent alors que certaines régions génomiques contrôlant les patrons de couleurs des ailes, impliqués dans l'isolement pré-zygotique, proviennent d'une autre espèce (Consortium 2012). De l'autre côté du spectre, un exemple qui semble à première vue proche du cas des moules des docks est celui du moineau italien (*Passer italiae*, Elgvin et al. 2017). Cette espèce, issue de l'hybridation entre le moineau domestique (*P. domesticus*) et le moineau espagnol (*P. hispaniolensis*), a une composition génomique intermédiaire dans des proportions respectives ~60/40. De plus, le moineau italien présente un isolement reproductif,

possiblement dû à des facteurs pré- et post-zygotiques, et une absence de flux de gènes malgré un contact actuel avec les deux espèces parentales (Hermansen et al. 2014 ; Trier et Hermansen 2014).

D'un point de vue théorique, Blanckaert et Bank (2018) s'intéressent à l'hypothèse que le tri réciproque des incompatibilités génétiques au sein d'une population hybride puisse produire un isolement post-zygotique avec les deux espèces parentales. En utilisant des modèles de type BDMI à deux ou quatre locus, ils concluent que la zone des paramètres permettant la spéciation hybride est assez restreinte et dépendante du degré de liaison entre locus incompatibles, de la taille de la population et des contributions relatives des deux parents à la population hybride. L'étude de Blanckaert et Bank (2018) démontre que la spéciation hybride par tri des incompatibilités devrait être rare dans la nature et qu'elle requiert des circonstances bien précises pour avoir lieu. Ce qui est en accord avec le très petit nombre de cas réellement avérés (Schumer et al. 2014b).

Le premier stade de la spéciation hybride dans des cas de contribution importante des deux parents, implique potentiellement la création de ce que beaucoup d'auteurs appellent un essaim hybride (traduction littérale de *hybrid swarm* ; Abbott et al. 2013). Ce terme correspond à un groupe admixé possédant une distribution unimodale d'indices hybrides et un mélange important des ascendances parentales le long du génome (Jiggins et Mallet 2000 ; Beninde et al. 2018). L'étude de ce type d'admixture est particulièrement importante pour la compréhension des mécanismes impliqués dans le processus de spéciation hybride (Abbott et al. 2013). Elle peut permettre la compréhension de la mise en place de nouvelles barrières aux flux de gènes, ou du réarrangement d'incompatibilités existantes (Hermansen et al. 2014).

Quels sont les éléments qui pourraient contribuer à un isolement reproductif entre les moules des docks et les autres lignées ? Tout d'abord, l'isolement reproductif partiel entre *M. edulis* et *M. galloprovincialis* semble démontré, à la fois par des expériences de croisement (Bierne et al. 2002a, 2006), et par le maintien de zones de tension entre ces entités en de multiples points (Bierne et al. 2002b, 2003b). De plus, il existe des différences de caryotypes entre les deux lignées de *M. galloprovincialis*, qui pourraient hypothétiquement permettre un isolement reproductif entre les moules des docks et les *M. galloprovincialis* atlantiques (Martínez-Lage et al. 1996). Enfin, quelques mécanismes d'isolement pré-zygotique existant entre espèces de moules pourraient aussi entrer en jeu (Bierne et al. 2002a ; Springer et Crespi 2007 ; voir la partie i.3.2 pour plus de détails sur l'isolement reproductif).

Dans le cas des moules des docks, il me semble impossible de conclure qu'elles constituent une espèce hybride avec les informations dont on dispose. Bien que cette population puisse être qualifiée d'essaim hybride, plusieurs arguments tendent à invalider une hypothèse de spéciation hybride. Tout d'abord, malgré un nombre de contacts secondaires important dans le complexe d'espèces *M. edulis*, aucune espèce hybride n'y est répertoriée à ce jour. On peut cependant noter que la population *M. trossulus* de la mer Baltique, fortement introgressée par *M. edulis* (Fraïsse et al. 2018a), pourrait être considérée pour une étude plus poussée des mécanismes

d'isolement avec les populations parentales. Ce complexe d'espèces ne semble pas prôner à ce type d'évènements, mais plutôt à des évènements d'introgression importants (Fraisse et al. 2016a ; Simon et al. 2019a). Malgré une composition génomique intermédiaire, il existe des différences d'introgression entre ports qui pourraient être liées à du flux de gènes local, de *M. edulis* vers les moules des docks (e.g. au Havre), et entre les *M. galloprovincialis* et les moules des docks (e.g. en rade de Brest) (Simon et al. 2019b). Ce flux de gènes doit passer par des individus de première génération (F1s identifiés dans Simon et al. 2019b), mais aussi par des rétro-croisements vers l'une ou l'autre des populations (plus difficiles à identifier). Enfin, l'isolement reproductif des moules des docks avec la lignée parentale *M. galloprovincialis* méditerranéenne n'est pas testée naturellement dans ce cas, puisque qu'elle n'est pas présente dans la zone d'introduction.

Le test de l'isolement reproductif nécessiterait des croisements en laboratoire entre les moules des docks et l'ensemble des lignées actrices dans ce scénario, i.e. les lignées parentales *M. edulis* et *M. galloprovincialis* méditerranéenne, ainsi que la lignée *M. galloprovincialis* atlantique. De plus, il faudra connaître de façon plus précise la séparation entre niches écologiques qui existe entre les différentes populations et si cela a un impact sur l'isolement pré-zygotique.

5.3 Un impact génétique ? Pollution génétique, conservation et gestion

En terme écologique, les espèces invasives sont considérées comme une cause d'extinction majeure (Bellard et al. 2016). Cependant, l'impact génétique est plus difficile à évaluer, de manière technique mais aussi conceptuelle.

Blackburn et al. 2014 classifient l'hybridation comme un des impacts environnementaux des espèces invasives. Leur catégorie « impact massif » (cf. Figure 5) de la classe hybridation correspond au cas où les hybrides sont fertiles et sains, conduisant à la perte d'une population native « pure », et ce même après éradication de l'espèce invasive. Les auteurs qualifient ce phénomène d'« extinction génomique », qui fait aussi référence au concept d'« extinction par hybridation » (Rhymer et Simberloff 1996 ; D. E. Wolf et al. 2001). De la même façon, le concept de « pollution génétique » (lancé par Butler 1994) repose sur une vision subjective de l'introgression, centrée sur une vision fixiste et phénotypique du concept d'espèce. La plupart des termes utilisés dans le domaine de la biologie des invasions et de la conservation sont péjoratifs (Rhymer et Simberloff 1996). Ils sous-entendent soit que les hybrides ont une valeur sélective inférieure, ce qui n'est pas forcément le cas, soit qu'une population « pure » a une valeur supérieure à nos yeux. Le concept de pureté d'une population pose de nombreux problèmes conceptuels, allant bien au-delà du domaine des invasions biologiques et de la conservation. Considéré avec trop de zèle, ce concept peut mener à des politiques de conservation possiblement aberrantes ou contre-productives, comme dans le cas du management des tortues des

Galápagos (Loire et Galtier 2017 ; J. M. Miller et al. 2018).

Le consensus dans la littérature semble être qu'une intervention est nécessaire lorsque l'hybridation et le mélange génétique sont clairement d'origine anthropique, reposant sur un principe de sauvegarde d'une « nature pré-anthropisée » (Allendorf et al. 2001). De plus, une lignée hybride ne devrait être protégée que si elle contient la seule information génétique restante d'un taxon qui a disparu (Allendorf et al. 2001). Enfin, comme concluent Rhymer et Simberloff (1996) : « Les conservationnistes n'ont pas besoin de sonner l'alarme à chaque fois que des populations échangent des gènes ».

Pour le moment, l'impact génétique des moules des docks reste limité. Dans les ports de Cherbourg, Le Havre et Saint-Nazaire, les moules des docks sont confinées à l'intérieur des ports. Nous pouvons penser que les quelques migrants portuaires ne constituent pas une pression démographique assez importante pour entamer une vague d'avance. De plus, si quelques marqueurs adaptatifs échappaient à la zone de tension présente à l'entrée des ports, il faudrait évaluer l'opinion de la communauté scientifique sur la situation. Les moules locales ne seraient-elles plus de l'espèce *M. edulis* avec des allèles méditerranéens ? Sachant que l'introggression s'est produite à de nombreuses reprises lors de l'histoire du complexe d'espèces, et que les génomes sont déjà très fortement admixés naturellement, il est, de mon point de vue, bien difficile de prédire si cette introgression, favorisée par les activités humaines, a un effet positif ou négatif.

La revue de Rhymer et Simberloff (1996) rappelle aussi que l'introggression est un processus évolutif courant, et qu'il faut considérer l'impact de l'hybridation avec parcimonie et choisir ses combats de conservation de manière pratique et pragmatique. Alors faut-il éradiquer les moules des docks ? D'un premier abord, ma réponse serait plutôt non, puisque les impacts négatifs d'une éradication seraient certainement supérieurs à ses impacts positifs. Tout d'abord, les espèces du complexe d'espèces *M. edulis* ne sont pas menacées d'extinction, et le potentiel d'extinction d'une population qui serait unique semble limité. Les moules ont de larges tailles de population, un cycle de vie rapide et souvent de larges aires de distribution. Comme nous l'avons montrée au chapitre 1, une possible introgression peut rester locale et rester couplée à la zone de tension suivante. Dans un second temps, à l'instar des moules naturelles qui sont reconnues comme des organismes clés de voûte aillant un rôle central dans l'écosystème (*multiplex hub*, Kéfi et al. 2016) et un rôle d'ingénieur des écosystèmes (Borthagaray et Carranza 2007), les bancs de moules des docks constituent un habitat et une ressource pour d'autres espèces, et leur éradication impacterait certainement d'autres taxons indigènes. De plus, une telle éradication devrait recourir à des moyens importants, comme l'utilisation d'un traitement au chlore (Bax et al. 2002), qui aurait des impacts environnementaux désastreux. Enfin, l'éradication devrait être effectuée dans l'ensemble des ports colonisés par les moules des docks, afin d'éviter une recolonisation depuis une autre population. La recolonisation pourrait également être limitée par le contrôle des vecteurs (e.g. traitement des eaux de ballast). Décider d'une éradication est donc un sujet complexe, impliquant des questions écologiques mais aussi

sociétales (Myers et al. 2000). De manière moins brutale, la densité de population portuaire ou le remplacement d'une partie des populations par des moules locales pourraient peut-être suffire à envahir (*swamp*) le fond génétique méditerranéen. Cependant cette considération requiert le développement d'un volet théorique en amont.

5.4 Considérations théoriques

5.4.1 Résolution de l'admixture

Les développements récents de la génomique des populations ont ouvert la porte à l'étude des patrons d'admixture au sein des zones de contact (L. M. Turner et Harr 2014 ; Elgvin et al. 2017 ; Duranton et al. 2018 ; Schumer et al. 2018b). Ceci à la fois sur des échelles de temps courts, c'est-à-dire la résolution des incompatibilités lors des premières générations d'hybridation (Schumer et al. 2018b), mais également sur des temps plus importants, correspondant par exemple au devenir évolutif d'un bloc génomique dans un fond génomique non spécifique (Duranton et al. 2018).

Comme nous avons pu le voir au chapitre 4, les modèles classiques d'incompatibilités souffrent d'un manque de flexibilité lors d'une mise à l'échelle génomique et parfois d'un manque de réalisme si l'on considère que la plupart des traits phénotypiques sont hautement polygéniques (Boyle et al. 2017). Je ne milite pas ici pour l'abandon des modèles simples à effets forts, qui représentent très clairement certains cas d'incompatibilités documentées, mais pour l'incorporation de l'ensemble des échelles afin de pouvoir représenter la diversité des patrons génomiques observés.

Bien que l'on se soit limité aux hybrides de premières générations dans le chapitre 4 (Simon et al. 2018), la prochaine étape est d'étudier la résolution de l'admixture sur différentes échelles de temps. L'objectif n'est pas seulement de produire des simulations individus centrées, mais également d'obtenir des prédictions analytiques sur les probabilités d'introgession des variants (à la manière de Sachdeva et Barton 2018b). Un pas dans cette direction a déjà été entrepris par Schneemann et al. (in prep.). Dans cette étude, le modèle a été étendu pour inclure de la dominance phénotypique variable, permettant une plus grande flexibilité des prédictions. En lien avec le parallélisme observé chez les moules (chapitres 2 et 3), le modèle permet d'obtenir des corrélations des valeurs sélectives marginales des variants entre les BCs, mais aussi des corrélations réduites lorsque l'on compare les BCs aux F2s. Ces résultats traduisent une dépendance de la trajectoire des variants à l'environnement génomique, différents entre types de croisement ou entre types d'admixture si on étend le raisonnement.

Dans une optique similaire, de nouveaux modèles ont été développés afin de suivre l'introgession de blocs génomiques. Sachdeva et Barton (2018a) s'intéressent à l'effet de la sélection polygénique sur l'introgession en utilisant un modèle infinitésimal affectant un trait sous sélection. L'effet d'un bloc de génome d'une longueur donnée sur la valeur du trait va alors dépendre

de l'effet additif de tous ses locus. Au cours des générations, la taille du bloc va varier par recombinaison et ainsi changer son effet sur la valeur du trait. Par exemple, un bloc initialement neutre peut être décomposé par recombinaison en de multiples blocs, chacun pouvant avoir des valeurs de traits positifs ou négatifs. L'une de leur conclusion sur la différence entre la propagation initiale et à long terme de l'introgession est particulièrement intéressante : la recombinaison joue un rôle différent durant ces deux phases. Dans la première phase, la recombinaison dilue l'avantage sélectif d'un bloc génomique positivement sélectionné, pouvant agir dans le sens inverse à la sélection. Au cours de la seconde phase, la recombinaison permet de remettre ensemble des blocs plus petits, tous positivement sélectionnés, pouvant finir par former des clusters de liaison extrêmement favorables. En d'autres termes, des longs fragments vont être d'abord sélectionnés et être découpés par la recombinaison. Les fragments restants à effets positifs ayant survécu à cette étape et ayant augmenté en fréquence, peuvent être remis ensemble, en évitant les interférences de Hill-Robertson (Sachdeva et Barton 2018b).

Une notion supplémentaire accompagne intrinsèquement la mise à l'échelle polygénique, celle de la contingence et de la répétabilité. Si chaque locus a un effet faible et que de nombreux locus sont disponibles pour l'action de la sélection, il semble à première vue que la dérive génétique (i.e. la stochasticité) aura une part importante à jouer dans la répétabilité du processus de tri des variants introgressés. Sachdeva et Barton (2018b), en utilisant le même modèle que précédemment, concluent que le taux de croissance initial d'un fragment dans la population receveuse est corrélé avec sa probabilité d'introgession à long terme. Ainsi, les variants d'un fragment avec une forte probabilité d'introgession sont susceptibles d'introgesser de manière répétée entre répliquats. L'étude empirique de la contingence passera à la fois par la cartographie des résolutions d'admixture dans des événements d'hybridation parallèles récents (e.g. chez les moules), et par l'obtention de probabilité d'introgession à long terme des variants, en utilisant des contacts répliqués plus anciens (e.g. chez les moules).

5.4.2 Devenir d'un allèle introduit

La littérature abonde de modèles théoriques sur l'invasion d'une espèce avec des considérations purement écologiques et éco-évolutives (i.e. sans hybridation ; Williamson 1996). Beaucoup moins d'études ont exploré la modélisation de l'invasion avec hybridation. La revue de Hall et al. (2006) s'intéresse à ce type de modélisation – même si le terme « invasion hybride » est particulièrement mal choisi. Je renvoie le lecteur vers cette revue qui dresse un portrait des différents modèles utilisés en écologie et génétique des invasions. Nous allons nous concentrer ici sur les paramètres pouvant impacter le devenir d'un allèle introduit et les possibles liens entre modèles génétiques liés à l'hybridation et modèles de la biologie des invasions.

Il peut être intéressant d'envisager la modélisation à l'échelle d'un locus, afin de pouvoir faire des généralisations et extensions. Le sort d'un allèle non natif peut théoriquement être

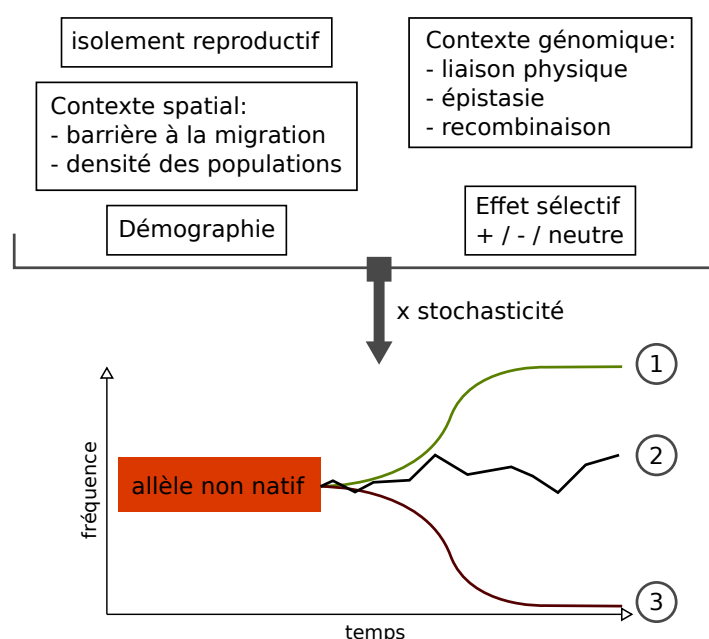


Figure 5.2 : Les multiples facteurs contrôlant le devenir d'un allèle non natif. On considère ici la fréquence d'un allèle introduit, à un locus donné, dans l'ensemble des populations où une introduction avec hybridation a eu lieu. L'allèle peut (1) envahir l'ensemble des populations, (3) disparaître et être remplacé par l'allèle natif ou (2) ségréger en fréquence intermédiaire, soit parce que les populations natives et introduites sont séparées spatialement et conservent des fréquences différentes, soit parce que la fréquence est homogène à un niveau intermédiaire dans toutes les populations.

très différent de celui de l'espèce introduite si du flux de gènes est possible avec une espèce native. Un allèle non natif peut par exemple disparaître même si l'espèce introduite remplace la native (Figure 5.2 (3)). À l'inverse, il peut introgresser l'espèce native même si l'espèce introduite finit par disparaître (Figure 5.2 (1)). Par exemple, cela pourrait être le cas dans les deux scénarios d'introgessions adaptatives. À la lumière de la théorie des clines, chaque locus va avoir un « comportement » dépendant du contexte démographique, écologique et spatial. S'ajoute à cela le contexte génomique dans lequel il se trouve (épistasie) et la corrélation avec les locus proches suivant l'architecture génomique (liaison). Pris ensemble, tous ces paramètres vont déterminer le résultat d'un évènement d'admixture (Figure 5.2).

À la vue de l'ensemble des paramètres qui contrôlent le sort d'un allèle non natif (Figure 5.2), il paraît vain d'espérer discuter de modèles verbaux de l'ensemble des scénarios. Ainsi, je souhaite simplement souligner quelques cas possibles et quelques liens avec la théorie actuelle, afin d'envisager de nouvelles directions de recherche. Les paramètres influençant la variation de la fréquence allélique sont non seulement nombreux mais ils interagissent aussi les uns avec les autres pour créer des contextes particuliers. De plus, ces contextes sont non déterministes et soumis à la stochasticité biologique (Figure 5.2).

Les modèles ayant exploré les contacts secondaires et la formation des zones de tension sont nombreux et ont explorés les effets de beaucoup de paramètres (Bazykin 1969; Slatkin et Maruyama 1975; Barton 1979, 1983; Piálek et Barton 1997; Barton et Shpak 2000; Barton et Turelli 2011). L'ensemble de la théorie liée à l'hybridation et aux zones d'hybridation s'applique au contexte des invasions avec hybridation. Cependant, les introductions présentent des paramètres particuliers. À la différence d'un contact secondaire entre fronts d'expansion, où des individus parentaux migrent de manière continue des zones externes vers la zone de contact, une introduction peut être coupée de la population parentale initiale ou ne recevoir

des propagules que de manière intermittente. De plus, d'un point de vue démographique, les cas d'introduction présentent des variabilités importantes en terme de taille et de fréquence de propagule (Simberloff 2009). Contrairement à un contact secondaire classique, l'asymétrie démographique peut être extrême. Ce paramètre peut influencer de façon importante le devenir d'allèles introduits, à la fois via l'action de la dérive au sein de la population introduite, par des effets de type densité-dépendance (effet Allee génétique), ou des phénomènes d'envahissement (*swamping*) génétique par l'espèce native. La population introduite peut être initialement petite, et un allèle, même sélectionné positivement, aura du mal à monter un gradient de densité. La pression de propagule apparaît dans ce contexte importante, non seulement dans la probabilité d'établissement d'une espèce introduite (aspect écologique, Simberloff 2009), mais aussi dans la balance contrôlant le niveau d'introgession qui peut avoir lieu entre l'espèce native et l'espèce introduite, et ceci dans les deux sens.

Le travail de Currat et al. (2008) est particulièrement intéressant à cet égard. Les auteurs s'intéressent à l'effet de la démographie et du niveau de reproduction interspécifique sur l'introgession de marqueurs neutres dans un contexte d'invasion, à l'aide d'un modèle spatial explicite. L'une de leurs conclusions importantes est que l'introgession ne se produit presque qu'exclusivement de l'espèce native vers l'espèce introduite dans le processus d'invasion, et ce indépendamment de la taille de population relative. Cependant, ce type de modèle ne prend en compte que les aspects démographique et stochastique.

Pour l'aspect empirique, le suivi des populations portuaires et la caractérisation des paramètres qui changent entre les milieux biotiques et abiotiques rencontrés seront particulièrement importants. Par exemple, on peut se demander quels paramètres sont liés aux différences de niveau d'introgession entre Le Havre et d'autres ports comme Cherbourg (Simon et al. 2019b). Est-ce simplement une variation aléatoire ou sont-elles liées à un contexte démographique différent ? L'observation de terrain permet de dire que les populations naturelles *M. edulis* ne semblent pas denses autour de Cherbourg, à la différence du port du Havre. La pression de migration des *M. edulis* natives est alors peut-être beaucoup plus importante au Havre qu'à d'autres endroits.

L'une des directions de recherche possible serait l'adaptation des modèles de réaction-diffusion aux contextes des invasions avec hybridation. En effet, ce type de modèle est déjà au cœur des théories des variations alléliques, des zones de tension et des vagues d'avance de la biologie des invasions (Fisher 1937 ; Haldane 1948 ; Roques 2013). Un ensemble de complexifications des modèles, provenant des différents milieux de la génétique et de l'écologie, pourrait être appliqué ici. Par exemple, des considérations de densité-dépendance (effets Allee) ont déjà été proposées par Barton et Turelli (2011). La migration génotype-dépendante a aussi été explorée par Nagylaki et Moody (1980) et Novak et Kollár (2017). Dans le domaine des vagues d'avance écologiques, les environnements hétérogènes (Berestycki et al. 2005a,b ; Berestycki et Hamel 2012) et les événements de migration longue distance (en utilisant des noyaux à queues lourdes ; e.g. Barton et Turelli 2011) ont été développés. Enfin, la grande

majorité des modèles de réaction-diffusion considère une seule dimension spatiale qui, bien que permettant une approximation correcte de nombreux cas, ne permet pas d'explorer toutes les conditions. Par exemple, on peut imaginer vouloir explorer les conséquences d'hétérogénéités spatiales en deux dimensions (chapitre IV de Roques 2013).

Une fois la question des phénomènes mono-locus (ou à un faible nombre de locus) traitée, il pourrait y avoir un nouveau changement d'échelle, en proposant une intégration génomique de ces concepts reprenant l'ensemble des théories polygéniques proposées dans l'étude de l'admixture (cf. partie 5.4.1). Les blocs d'introgession peuvent varier (i) en taille suivant la recombinaison, (ii) en valeur sélective intrinsèque suivant les variants qui le composent, et (iii) en fréquence à un point de l'espace donné suivant leur effet sélectif intrinsèque, leur environnement génomique (à la fois de liaison et d'épistasie), l'effet de la dérive, et enfin la migration depuis les dèmes voisins.

5.5 Moule des docks, vecteur d'un parasite d'un nouveau genre ?

En addition à leurs impacts écologiques et évolutifs directs, les espèces introduites peuvent aussi transmettre des parasites auto-stoppeurs aux espèces natives (*spillover* ; Torchin et al. 2002 ; Prenter et al. 2004). S'ajoute à cette menace le potentiel des espèces invasives à devenir des réservoirs pour des parasites natifs (*spillback* ; Kelly et al. 2009).

Un nouveau type de parasite, un cancer transmissible, a récemment été découvert chez les moules *Mytilus* spp. (Metzger et al. 2016). Un cancer transmissible est une lignée cancéreuse étant passée d'un parasitisme à l'échelle de l'individu à celle de l'espèce (voir de plusieurs espèces). Un cancer classique se trouve dans un cul de sac évolutif à la mort de son hôte. Le cancer transmissible a pu échapper à ce problème par acquisition de la capacité à se transmettre d'un individu à un autre, survivant ainsi à son extinction. Ce phénomène se traduit par la prolifération, dans un individu donné, de cellules d'un autre individu (voire d'une autre espèce).

Jusqu'à récemment, nous pensions les cancers transmissibles des événements rares nécessitant des conditions très particulières. Les deux exemples connus étaient le cancer transmissible du chien (CTVT) et le cancer du diable de Tasmanie (DFT1 et 2). Le cancer du chien est une lignée cancéreuse sexuellement transmissible qui se propage dans les populations canines depuis 4000 à 8500 ans environ (Baez-Ortega 2019). Le cancer du diable de Tasmanie est, quant à lui, beaucoup plus récent avec une première description en 1996 (Stammnitz et al. 2018). Cependant, cinq cancers transmissibles supplémentaires ont été découverts chez des bivalves marins (Metzger et Goff 2016 ; Metzger et al. 2016), comprenant une espèce du complexe *M. edulis*, remettant en question le caractère improbable de ce phénomène.

L'observation de cette leucémie transmissible chez *M. trossulus* en Colombie Britannique

(Metzger et al. 2016) nous a tout d'abord fait réaliser que certaines observations d'allèles *M. trossulus* dans nos données européennes pouvaient être interprétées à la lumière de cette découverte (Riquet et al. 2017). Nous avons pu montrer récemment qu'un cancer transmissible d'origine *M. trossulus* avait pu passer la barrière d'espèces et était aussi présent en Europe (chez *M. edulis* et *M. galloprovincialis*) et au Chili (chez *M. chilensis*; (Yonemitsu et al. 2019)). L'étude de quatre locus des lignées cancéreuses a permis de mettre en évidence que (i) les cancers transmissibles du Chili et d'Europe ont des allèles quasiment identiques, montrant une origine commune récente, et que (ii) cette nouvelle lignée est différente de celle déjà identifiée en Colombie Britannique (Yonemitsu et al. 2019).

Une des hypothèses sur la transmission du cancer transmissible entre individus *Mytilus* est que la lignée clonale ne se propage pas verticalement. Ainsi, il faudrait un transfert de cellules cancéreuses d'un individu à l'autre, probablement via le milieu extérieur par filtration. Ce mode de transmission nécessite probablement une certaine proximité des individus. Ainsi, la présence d'une même lignée cancéreuse dans des populations très distantes (Chili et Europe), n'ayant pas échangé de migrants naturels depuis au moins 10 000 ans, requiert l'hypothèse que des moules introduites aient joué le rôle de vecteur. Cette hypothèse requiert également la proximité des moules émettrices aux vectrices (e.g. en Europe) puis la proximité des moules vectrices aux receveuses (e.g. au Chili). Ainsi, les ports apparaissent comme un environnement particulièrement propice à ce type de transfert.

Cette hypothèse soulève de nouvelles questions sur l'impact des moules des docks et un potentiel rôle dans la dispersion du cancer transmissible : Les moules des docks peuvent-elles être des vecteurs du cancer transmissible ? Sont-elles plus ou moins susceptibles au cancer que d'autres lignées ? Cette susceptibilité en fait-elle des réservoirs locaux ? Toutes ces questions ne pourront être traitées qu'après une compréhension à la fois de l'écologie et de la dispersion des moules des docks, de la susceptibilité des différentes espèces de moules au cancer, mais également de la compréhension des mécanismes d'infection et de dispersion du cancer transmissible dans le complexe d'espèces *M. edulis*.

5.6 Conclusion

En conclusion, ce travail de thèse renforce l'idée que le complexe d'espèces des moules bleues (*M. edulis*) est un modèle marin très important pour l'étude de la spéciation et de l'hybridation. Non content de comporter de nombreux niveaux de différenciation illustrant la continuité de la spéciation, ce modèle présente une diversité importante de contacts secondaires en terme d'âges et de contextes, entre l'ensemble des espèces le constituant. L'étude des zones d'hybridation du complexe d'espèces *M. edulis*, ayant déjà largement contribué par le passé à la compréhension des mécanismes d'isolement reproductif, continue de nourrir de nouveaux questionnements liés aux introgressions locales et au maintien des zones de tension en milieu marin. En effet, malgré la description d'un ensemble presque exhaustif des populations de l'hémisphère nord, nous continuons à découvrir de nouvelles lignées, de nouvelles zones d'hybridation et de nouvelles populations fortement introgressées.

Des plus, le modèle des moules est source de nombreuses introductions et invasions à l'échelle mondiale, notamment par l'espèce *M. galloprovincialis*. Ces introductions produisent des contacts secondaires récents, particulièrement intéressants pour explorer la phase initiale de ceux-ci. Les introductions ayant produit les moules des docks, caractérisées dans cette thèse, sont particulièrement fascinantes. Elles nous permettent d'explorer des questions évolutives en regardant les patrons de résolution de l'admixture et en les comparant à des événements similaires. D'un autre côté, elles pourront aussi apporter d'importantes réponses au domaine de la biologie des invasions sur des questions récurrentes : Quel est le rôle de la diversité génétique dans les potentiels d'établissement et d'invasion ? Quel est le rôle des introgressions adaptatives dans l'invasion ? Enfin, l'étude de l'écologie des moules des docks sera une étape clé dans la compréhension des facteurs leur permettant de prospérer au cœur de milieux anthropisés ou ceux les empêchant de coloniser en grand nombre le milieu naturel. L'étude des populations externes aux ports en rade de Brest sera essentielle afin de répondre à une question centrale : Quelles sont les parts intrinsèques et extrinsèques contrôlant l'avance des moules des docks ?

Si il y a bien un message important à retenir de cette thèse, lorsque l'on regarde les patrons d'admixture et les mécanismes d'isolement reproductifs, c'est que la vision « gène de spéciation » ne suffit plus pour expliquer l'accumulation de preuves du caractère extrêmement polygénique de ces processus. Il faut donc accompagner la révolution génomique, produisant une résolution toujours plus grande, d'une révolution de la théorie afin d'englober cette nouvelle vision omnigénique de la spéciation.

Annexe : *Evolution digest*

Digest du journal *Evolution* sur la prise en compte de la sélection aux sites liés dans les inférences démographiques.

A. Simon et M. Duranton (2018). « Digest : Demographic Inferences Accounting for Selection at Linked Sites ». *Evolution*. DOI : 10.1111/evo.13504

Digest: Demographic inferences accounting for selection at linked sites[†]

Alexis Simon^{1,*}  and Maud Duranton^{1,2,*}

¹Institut des Sciences de l'Evolution-Montpellier, Université de Montpellier, CNRS-IRD-EPHE-UM, France

²E-mail: maud.duranton@umontpellier.fr

Received April 18, 2018

Accepted May 7, 2018

Complex demography and selection at linked sites can generate spurious signatures of divergent selection. Unfortunately, many attempts at demographic inference consider overly simple models and neglect the effect of selection at linked sites. In this issue, Rougemont and Bernatchez (2018) applied an approximate Bayesian computation (ABC) framework that accounts for indirect selection to reveal a complex history of secondary contacts in Atlantic salmon (*Salmo salar*) that might explain a high rate of latitudinal clines in this species.

Identifying signatures of selection within genomes is a long-standing objective in evolutionary biology. This quest is complicated by factors such as species' demographic history, which has long been recognized for its potential to produce footprints that mimic selection. Secondary contacts, for instance, can generate clines in allele frequencies correlated with spatial variation in ecological factors that may be interpreted incorrectly as representing local adaptation to an environmental gradient (Fig. 1). More recently, selection at linked sites has received particular attention as another confounding mechanism that generates genomic regions of increased divergence between populations as expected in the presence of divergent selection (Fig. 1B). Furthermore, selection at linked sites seems to affect a sufficiently large fraction of the genome of many species and challenges the basic assumption that most loci evolve neutrally. Despite those issues, few studies have attempted to include the effects of selection at linked sites in demographic divergence models (Roux et al. 2016; Rougeux et al. 2017).

In this issue, Rougemont and Bernatchez (2018) present an inference of the demographic history of Atlantic salmon (*Salmo salar*) populations using a dataset of 5034 SNPs genotyped in

77 populations covering the whole species' range. They first reassessed the population structure and confirmed the existence of genetic subdivision among regions. They then used an approximate Bayesian computation (ABC) method to compare demographic scenarios with various histories of gene flow, including divergence with migration, ancient migration, and secondary contact.

ABC avoids the explicit computation of likelihood, allowing users to choose the most appropriate summary statistics to be combined for model selection. This approach allows users to evaluate more complex and flexible models compared to other inference methods. Furthermore, ABC can be coupled with machine learning, which is well adapted to high-dimensional data. Although selection was not explicitly simulated in this study, the implemented models build on the theoretical predictions that selection against introgression locally reduces the effective migration rate between populations (Barton and Bengtsson 1986), whereas selection at linked sites enhances the effect of drift by locally reducing the effective population size (Charlesworth et al. 1993). Even if selection at linked sites can be either positive (through selective sweeps) or negative (through background selection), these two mechanisms both reduce neutral nucleotide diversity over time.

Using this modeling framework, Rougemont and Bernatchez (2018) provide a new perspective on the evolutionary history of a widely studied system in which a complex demographic history

*These authors are cofirst authors.

[†]This article corresponds to Rougemont Q., L. Bernatchez. 2018 The demographic history of Atlantic Salmon (*Salmo salar*) across its distribution range reconstructed from Approximate Bayesian Computations. *Evolution*. <https://doi.org/10.1111/evo.13486>.

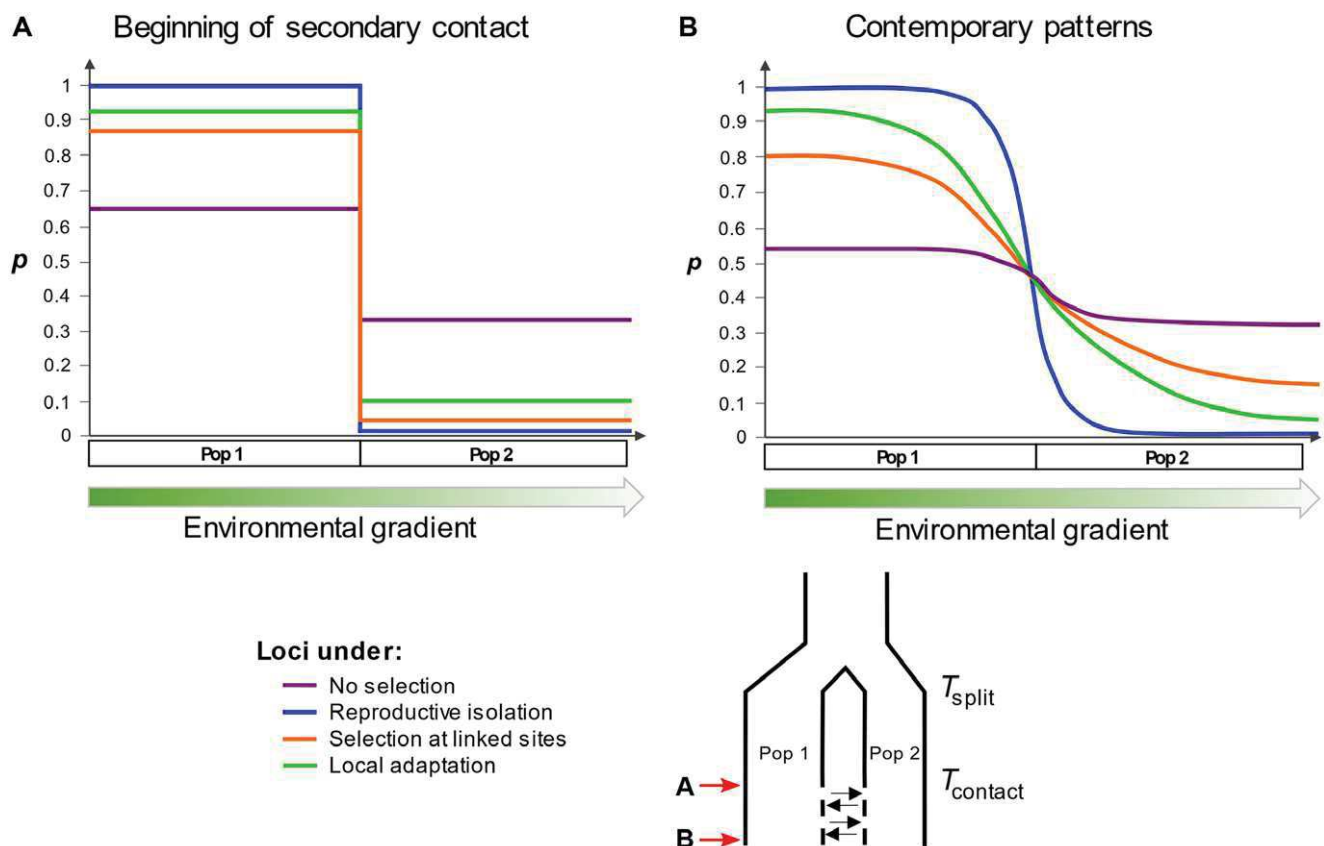


Figure 1. The fate of polymorphisms affected by different processes before and after admixture. Each color represents a different locus affected by a specific process during divergence (i.e., between T_{split} and T_{contact}) and secondary contact. Plots represent allele frequency clines across space, where populations 1 and 2 occupy different parts of an environmental gradient. The red arrows point to the time when cline patterns are observed. Since they are evolving under different processes, loci reach various levels of divergence and experience different levels of gene flow. (A) During divergence, differentiation builds up by either selection or genetic drift. Local adaptation (green) and reproductive isolation (blue) alleles rapidly increase in frequency in one population and decrease in the other. Loci under selection at linked sites (orange) diverge more rapidly than neutral ones (violet) because linked selection locally reduces the effective population size, which increases genetic drift. (B) During the contact, gene flow rapidly erodes the differentiation of the neutral and linked selected loci, but the cline of the second is steeper because it was more differentiated. Local adaptation loci experience reduced gene flow, and their clines are fixed by the environment. The reproductive isolation cline rapidly converged with the local adaptation cline. Clines of reproductive isolation and selection at linked sites can therefore be mistaken for local adaptation if these processes have not been taken into account when performing demographic inferences.

may confound the detection of local adaptation. A scenario of secondary contacts was favored both between and within American and European Atlantic salmon populations, providing strong evidence for a reticulated demographic history. This work emphasizes the importance of identifying a null model of demographic history for a given species that accounts for both population structure and selection at linked sites. Future studies seeking to identify signatures of divergent selection in natural populations will need to take into account variation in effective population size both in time and along the genome. Finally, this study illustrates the flexibility of ABC demographic inferences to account for dataset particularities such as the ascertainment bias in SNP chip arrays,

confirming that this approach could be widely applied to many nonmodel species.

Future studies would benefit from going a step further toward modeling the whole picture of the demographic history by considering more populations simultaneously. This goal could be achieved within an ABC framework, because population structure and phylogeography of Atlantic salmon are now well known. However, for species for which there is no a priori knowledge about population history, considering a large number of populations (meaning an exponentially large number of alternative scenarios) can quickly become intractable in terms of computational time. One way forward could be to use non-ABC methods with

minimal assumptions (e.g., Minimal-Assumption Genomic Inference of Coalescence, Weissman and Hallatschek 2017; or Approximate Blockwise Likelihood Estimation, Reddy et al. 2017). Such methods would capture the key aspects of populations' history and provide the starting point for further refinements. Those methods also have the potential to incorporate the aforementioned confounding factors such as selection at linked sites. In conclusion, an important prerequisite before starting to uncover signatures of divergent selection is to identify a relevant null model of demographic history accounting for indirect selective effects.

LITERATURE CITED

- Barton, N., and B. O. Bengtsson. 1986. The barrier to genetic exchange between hybridising populations. *Heredity* 57:357–376.
- Charlesworth, B., M. T. Morgan, D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134:1289–1303.
- Reddy, C. B., M. J. Hickerson, L. A. Frantz, and K. Lohse. 2017. Blockwise site frequency spectra for inferring complex population histories and recombination. *bioRxiv*, 077958. <https://doi.org/10.1101/077958>.
- Rougemont, Q., and L. Bernatchez. 2018. The demographic history of Atlantic Salmon (*Salmo salar*) across its distribution range reconstructed from approximate Bayesian computations. *Evolution*. <https://doi.org/10.1111/evo.13486>.
- Rougeux, C., L. Bernatchez, and P. A. Gagnaire. 2017. Modeling the multiple facets of speciation-with-gene-flow toward inferring the divergence history of lake whitefish species pairs (*Coregonus clupeaformis*). *Genome Biol. Evol.* 9:2057–2074.
- Roux, C., C. Fraïsse, J. Romiguier, Y. Anciaux, N. Galtier, and N. Bierne. 2016. Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLOS Biol.* 14. <https://doi.org/10.1371/journal.pbio.2000234>.
- Weissman, D. B., and O. Hallatschek. 2017. Minimal-assumption inference from population-genomic data. *Elife* 6:e24836.

Associate Editor: K. Moore
Handling Editor: M. Noor

SUBMIT A DIGEST

Digests are short (~500 word), news articles about selected original research included in the journal, written by students or postdocs. These digests are published online and linked to their corresponding original research articles. For instructions on Digests preparation and submission, please visit the following link: <https://sites.duke.edu/evodigests/>.

Liste des figures

1	Processus de spéciation et barrières aux flux de gènes	3
2	Scénarios démographiques d'isolement	4
3	Modèle BDMI	8
4	Typologie du processus d'invasion biologique	13
5	Classification de l'impact des invasions	13
6	Voies d'introduction en milieu marin	14
7	Photographies des trois espèces d'intérêt	18
8	Distribution du complexe d'espèce <i>M. edulis</i>	20
9	Relations phylogénétiques schématisques du complexe <i>M. edulis</i>	21
10	Zone grise de la spéciation	24
2.1	Distribution naturelle des moules dans le nord de la France	79
4.1	Trajectoire de résolution d'une BDMI	208
5.1	ACP de l'ensemble des données collectées durant la thèse	258
5.2	Les facteurs contrôlant le devenir d'un allèle non natif	267

Liste des tableaux

1	Les introductions de <i>M. galloprovincialis</i>	28
---	------------------------------------------------------------	----

Liste des encadrés

1	Encadré 1 : Bases moléculaires de la différenciation génétique	6
2	Encadré 2 : Incompatibilité de Bateson-Dobzhansky-Muller (BDMI)	8
3	Encadré 3 : la diaspora <i>Mytilus</i>	29

Glossaire

Allopatrie

L'allopatrie correspond à la situation où deux taxons ont une aire de distribution non chevauchante. 4

BC

Abbréviation de *Backcross*. Désigne un rétro-croisement, c'est-à-dire le croisement entre un hybride (généralement un F1) et un individu d'une des deux populations parentales. On désigne souvent le nombre de générations de rétrocroisement par un nombre, par exemple un BC1 sera la première génération de rétro-croisement entre un F1 et l'une des populations parentales. 2, 7, 30, 31, 176, 231

Cryptique

Qualifie une espèce morphologiquement très proche d'une ou plusieurs autres espèces et n'étant pas reconnue taxonomiquement. Elle satisfait cependant un certain nombre de critères de à l'une des définitions de l'espèce, e.g. une divergence génétique élevée, une absence de flux de gènes ou un isolement reproductif pré- ou post-zygotique. 14, 78

Cryptogénique

Qualifie une espèce pour laquelle le caractère natif ou introduit ne peut pas être déterminé de manière fiable, voir Carlton (1996) et Carlton (2009). 14

Dème

En modélisation, désigne une population panmictique. Les dèmes sont souvent utilisés afin de discrétiser l'espace en des unités homogènes dans lesquelles on peut utiliser un certain nombre de modèles classiques de la génétique des populations. 235

F1

Désigne le croisement de première génération entre deux individus parentaux. Dans une espèce diploïde, un individu F1 ne contient pas de recombinaison entre les deux génomes parentaux et est hétérozygotes sur tous les locus fixés entre les deux populations. 4, 7, 8, 16, 30, 31, 229

F2

Désigne le croisement de seconde génération, entre deux F1. 7, 30, 31, 176, 177, 231

GWAS

Abbréviation de *Genome Wide Association Study*. Désigne l'étude de la corrélation entre des variants génomiques et des traits phénotypiques. 9, 176

Introgression

Augmentation en fréquence de l'allèle d'une autre espèce (ou lignée) dans une espèce focale. Processus produit par l'hybridation entre les deux espèces, puis par le rétro-croisement avec l'espèce focale. Le terme peut aussi bien servir pour désigner une augmentation globale sur l'ensemble du génome, que pour une petite région génomique en particulier.. iii, 8, 10, 11, 17–19, 21–23, 27, 31, 33–35, 173–175, 227, 229–235, 237

Phasé

Se dit d'un génome dont on sait sur quel chromosome homologue se trouve les allèles (soit paternel, soit maternel). Ceci correspond donc à la reconstruction des haplotypes paternels et maternels. La phase peut être obtenue par des méthodes statistiques ou par séquençage des génomes parentaux.. 10

Propagule

En biologie des invasions, désigne un groupe d'individus étant introduits ensemble en un même point du temps et de l'espace. La pression de propagule est le nombre total d'individus introduits en un point donné (taille de propagule) et leur fréquence (nombre de propagules). 11

Sympatrie

La sympatrie correspond à la situation où deux taxons ont une aire de distribution commune. 23, 25, 225

Vague d'avance

Terme utilisé notamment en biologie des invasions pour désigner la propagation d'une espèce ou d'un génotype dans l'espace, dont la distribution de densité ressemble à une vague durant la propagation. 17, 230, 234

Valeur sélective

Espérance de la contribution relative aux générations ultérieures d'une classe (allèle, génotype, classe phénotypique). Elle est estimée, suivant les situations, par une ou plusieurs composantes (survie, fertilité, nombre de descendants à la génération suivante, etc.) associées à la classe considérée (*fitness* en anglais). Tiré de Thomas et al. 2016. iv, 1–3, 6–9, 16, 31, 175, 177, 227, 229, 235

Vigueur hybride

Désigne l'augmentation de la valeur sélective d'un hybride comparé à l'attendu sachant la valeur sélective des deux parents. 7

Épistatique

Désigne la modification de l'effet (expression) d'un locus par un (ou plusieurs) autres locus. L'épistasie peut être compris comme l'effet du fond génétique sur un locus. 2, 8, 9, 30, 31, 176

Bibliographie

- Abbott, R. J. (1992). « Plant Invasions, Interspecific Hybridization and the Evolution of New Plant Taxa ». *Trends in Ecology & Evolution* 7.12, p. 401-405. DOI : [https://doi.org/10.1016/0169-5347\(92\)90020-C](https://doi.org/10.1016/0169-5347(92)90020-C).
- Abbott, R. J. et al. (2013). « Hybridization and Speciation ». *Journal of Evolutionary Biology* 26.2, p. 229-246. DOI : [10.1111/j.1420-9101.2012.02599.x](https://doi.org/10.1111/j.1420-9101.2012.02599.x).
- Ahmad, M. et J. A. Beardmore (1976). « Genetic Evidence That the "Padstow Mussel" Is *Mytilus Galloprovincialis* ». *Marine Biology* 35.2, p. 139-147. DOI : [10.1007/BF00390935](https://doi.org/10.1007/BF00390935).
- Ahyong, S. et al. (2018). « World Register of Introduced Marine Species (WRiMS) ». DOI : [10.14284/347](https://doi.org/10.14284/347).
- Allendorf, F. W., R. F. Leary, P. Spruell et J. K. Wenburg (2001). « The Problems with Hybrids : Setting Conservation Guidelines ». *Trends in Ecology and Evolution* 16.11, p. 613-622. DOI : [10.1016/S0169-5347\(01\)02290-X](https://doi.org/10.1016/S0169-5347(01)02290-X).
- Anderson, A. S., A. L. Bilodeau, M. R. Gilg et T. J. Hilbish (2002). « Routes of Introduction of the Mediterranean Mussel (*Mytilus Galloprovincialis*) to Puget Sound and Hood Canal ». *Journal of Shellfish Research* 21.1, p. 75-79.
- Anderson, E. (1948). « Hybridization of the Habitat ». *Evolution* 2.1, p. 1-9. DOI : [10.2307/2405610](https://doi.org/10.2307/2405610).
- Anderson, E. et G. L. Stebbins (1954). « Hybridization as an Evolutionary Stimulus ». *Evolution* 8.4, p. 378-388. DOI : [10.2307/2405784](https://doi.org/10.2307/2405784).
- Anttila, C., R. King, C. Ferris, D. Ayres et D. Strong (2000). « Reciprocal Hybrid Formation of *Spartina* in San Francisco Bay ». *Molecular Ecology* 9.6, p. 765-770.
- Appeltans, W. et al. (2012). « The Magnitude of Global Marine Species Diversity ». *Current Biology* 22.23, p. 2189-2202. DOI : [10.1016/j.cub.2012.09.036](https://doi.org/10.1016/j.cub.2012.09.036).
- Apte, S., B. S. Holland, L. S. Godwin et J. P. A. Gardner (2000). « Jumping Ship : A Stepping Stone Event Mediating Transfer of a Non-Indigenous Species via a Potentially Unsuitable Environment ». *Biological Invasions* 2.1, p. 75-79. DOI : [10.1023/A:1010024818644](https://doi.org/10.1023/A:1010024818644).
- Araneda, C., M. A. Larraín, B. Hecht et S. Narum (2016). « Adaptive Genetic Variation Distinguishes Chilean Blue Mussels (*Mytilus Chilensis*) from Different Marine Environments ». *Ecology and Evolution* 6.11, p. 3632-3644. DOI : [10.1002/ece3.2110](https://doi.org/10.1002/ece3.2110).
- Baez-Ortega, A. (2019). « Somatic Evolution and Global Expansion of an Ancient Transmissible Cancer Lineage ». *Science* 365, eaau9923.
- Bailey, S. A. (2015). « An Overview of Thirty Years of Research on Ballast Water as a Vector for Aquatic Invasive Species to Freshwater and Marine Environments ». *Aquatic Ecosystem Health & Management* 18.3, p. 261-268. DOI : [10.1080/14634988.2015.1027129](https://doi.org/10.1080/14634988.2015.1027129).

- Barker, B. S., J. E. Cocio, S. R. Anderson, J. E. Braasch, F. A. Cang, H. D. Gillette et K. M. Dlugosch (2018). « Potential Limits to the Benefits of Admixture during Biological Invasion ». *Molecular Ecology*. DOI : 10.1111/mec.14958.
- Barsotti, G. et C. Meluzzi (1968). « Osservazioni Su *Mytilus Edulis* L. e *Mytilus Galloprovincialis* Lamarck ». *Conchiglie* 4, p. 50-58.
- Barton, N. H. et G. M. Hewitt (1981). « The Genetic Basis of Hybrid Inviability in the Grasshopper *Podisma Pedestris* ». *Heredity* 47.3, p. 367-383. DOI : 10.1038/hdy.1981.98.
- Barton, N. H. (1979). « The Dynamics of Hybrid Zones ». *Heredity* 43.3, p. 341-359. DOI : 10.1038/hdy.1979.87.
- (1983). « Multilocus Clines ». *Evolution* 37.3, p. 454-471. DOI : 10.2307/2408260.
- Barton, N. H. et B. O. Bengtsson (1986). « The Barrier to Genetic Exchange between Hybridising Populations ». *Heredity* 56, p. 357-376. DOI : 10.1038/hdy.1986.135.
- Barton, N. H. et G. M. Hewitt (1985). « Analysis of Hybrid Zones ». *Annual Review of Ecology, Evolution, and Systematics* 16, p. 113-148. DOI : 10.1146/annurev.es.16.110185.000553.
- (1989). « Adaptation, Speciation and Hybrid Zones. » *Nature* 341, p. 497-503. DOI : 10.1038/341497a0.
- Barton, N. H. et M. Shpak (2000). « The Effect of Epistasis on the Structure of Hybrid Zones ». *Genetical Research* 75.2, p. 179-198. DOI : 10.1017/S0016672399004334.
- Barton, N. H. et M. Turelli (2011). « Spatial Waves of Advance with Bistable Dynamics : Cytoplasmic and Genetic Analogues of Allee Effects. » *The American naturalist* 178.3, E48-E75. DOI : 10.1086/661246.
- Bateson, W. (1909). « Heredity and Variation in Modern Lights ». *Darwin and modern science*. Sous la dir. d'A. C. Seward, p. 85-101.
- Bax, N., K. Hayes, A. Marshall, D. Parry et R. Thresher (2002). « Man-Made Marinas as Sheltered Islands for Alien Marine Organisms : Establishment and Eradication of an Alien Invasive Marine Species ». *Turning the Tide : The Eradication of Invasive Species*. Sous la dir. de C. R. Veitch et M. N. Clout. IUCN SSC Invasive Species Specialist Group. Gland, Switzerland and Cambridge, UK : IUCN, p. 26-39.
- Bazykin, A. D. (1969). « Hypothetical Mechanism of Speciation ». *Evolution* 23.4, p. 685. DOI : 10.2307/2406862.
- Beaumont, A. R., M. P. Hawkins, F. L. Doig, I. M. Davies et M. Snow (2008). « Three Species of *Mytilus* and Their Hybrids Identified in a Scottish Loch : Natives, Relicts and Invaders ? » *Journal of Experimental Marine Biology and Ecology* 367.2, p. 100-110. DOI : 10.1016/j.jembe.2008.08.021.
- Bellard, C., P. Cassey et T. M. Blackburn (2016). « Alien Species as a Driver of Recent Extinctions ». *Biology letters*. DOI : 10.1098/rsbl.2015.0623.
- Beninde, J., S. Feldmeier, M. Veith et A. Hochkirch (2018). « Admixture of Hybrid Swarms of Native and Introduced Lizards in Cities Is Determined by the Cityscape Structure and Invasion History ». *Proceedings of the Royal Society B : Biological Sciences* 285, p. 20180143. DOI : 10.1098/rspb.2018.0143.
- Berestycki, H. et F. Hamel (2012). « Generalized Transition Waves and Their Properties ». *Communications on Pure and Applied Mathematics* LXV, p. 592-648.
- Berestycki, H., F. Hamel et L. Roques (2005a). « Analysis of the Periodically Fragmented Environment Model : I – Species Persistence ». *Journal of Mathematical Biology* 51.1, p. 75-113. DOI : 10.1007/s00285-004-0313-3.

- (2005b). « Analysis of the Periodically Fragmented Environment Model : II—Biological Invasions and Pulsating Travelling Fronts ». *Journal de Mathématiques Pures et Appliquées* 84.8, p. 1101-1146. DOI : 10.1016/j.matpur.2004.10.006.
- Bertl, J., H. Ringbauer et M. G. B. Blum (2018). « Can Secondary Contact Following Range Expansion Be Distinguished from Barriers to Gene Flow ? » *PeerJ* 6, e5325. DOI : 10.7717/peerj.5325.
- Beyer, J. et al. (2017). « Blue Mussels (*Mytilus Edulis* Spp.) as Sentinel Organisms in Coastal Pollution Monitoring : A Review ». *Marine Environmental Research* 130, p. 338-365. DOI : 10.1016/j.marenvres.2017.07.024.
- Bierne, N., F. Bonhomme, P. Boudry, M. Szulkin et P. David (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* 273, p. 1253-1260. DOI : 10.1098/rspb.2005.3440.
- Bierne, N., F. Bonhomme et P. David (2003a). « Habitat Preference and the Marine-Speciation Paradox ». *Proceedings of the Royal Society of London. Series B* 270.1522, p. 1399-1406. DOI : 10.1098/rspb.2003.2404.
- Bierne, N., P. Borsa, C. Daguin, D. Jollivet, F. Viard, F. Bonhomme et P. David (2003b). « Introgression Patterns in the Mosaic Hybrid Zone between *Mytilus Edulis* and *M. Galloprovincialis*. » *Molecular Ecology* 12.2, p. 447-461. DOI : 10.1046/j.1365-294X.2003.01730.x.
- Bierne, N., P. David, P. Boudry et F. Bonhomme (2002a). « Assortative Fertilization and Selection at Larval Stage in the Mussels *Mytilus Edulis* and *M. Galloprovincialis*. » *Evolution* 56.2, p. 292-298. DOI : 10.1111/j.0014-3820.2002.tb01339.x.
- Bierne, N., P. David, A. Langlade et F. Bonhomme (2002b). « Can Habitat Specialisation Maintain a Mosaic Hybrid Zone in Marine Bivalves ? » *Marine Ecology Progress Series* 245, p. 157-170. DOI : 10.3354/meps245157.
- Bierne, N., P.-A. Gagnaire et P. David (2013). « The Geography of Introgression in a Patchy Environment and the Thorn in the Side of Ecological Speciation ». *Current Zoology* 59.1, p. 72-86. DOI : 10.1093/czoolo/59.1.72.
- Bierne, N., J. Welch, E. Loire, F. Bonhomme et P. David (2011). « The Coupling Hypothesis : Why Genome Scans May Fail to Map Local Adaptation Genes ». *Molecular Ecology* 20, p. 2044-2072. DOI : 10.1111/j.1365-294X.2011.05080.x.
- Bishop, M. J. et al. (2017). « Effects of Ocean Sprawl on Ecological Connectivity : Impacts and Solutions ». *Journal of Experimental Marine Biology and Ecology* 492, p. 7-30. DOI : 10.1016/j.jembe.2017.01.021.
- Blackburn, T. M. et al. (2011). « A Proposed Unified Framework for Biological Invasions ». *Trends in Ecology & Evolution* 26.7, p. 333-339. DOI : 10.1016/j.tree.2011.03.023.
- Blackburn, T. M. et al. (2014). « A Unified Classification of Alien Species Based on the Magnitude of Their Environmental Impacts ». *PLoS Biology* 12.5, e1001850. DOI : 10.1371/journal.pbio.1001850.
- Blanckaert, A. et C. Bank (2018). « In Search of the Goldilocks Zone for Hybrid Speciation ». *PLOS Genetics* 14.9. Sous la dir. de J. Zhang, e1007613. DOI : 10.1371/journal.pgen.1007613.
- Bock, D. G. et al. (2015). « What We Still Don't Know about Invasion Genetics ». *Molecular Ecology* 24.9, p. 2277-2297. DOI : 10.1111/mec.13032.
- Bonham, V., J. L. Shields et C. Riginos (2019). « CABI : *Mytilus Galloprovincialis*. » *Invasive Species Compendium*. Wallingford, UK : CAB International. URL : www.cabi.org/isc.

- Borthagaray, A. I. et A. Carranza (2007). « Mussels as Ecosystem Engineers : Their Contribution to Species Richness in a Rocky Littoral Community ». *Acta Oecologica* 31.3, p. 243-250. DOI : 10.1016/j.actao.2006.10.008.
- Bouchemousse, S., C. Liautard-Haag, N. Bierne et F. Viard (2016). « Distinguishing Contemporary Hybridization from Past Introgression with Postgenomic Ancestry-Informative SNPs in Strongly Differentiated *Ciona* Species ». *Molecular Ecology* 25.21, p. 5527-5542. DOI : 10.1111/mec.13854.
- Boyle, E. A., Y. I. Li et J. K. Pritchard (2017). « An Expanded View of Complex Traits : From Polygenic to Omnigenic ». *Cell* 169.7, p. 1177-1186. DOI : 10.1016/j.cell.2017.05.038.
- Branch, G. M., F. Odendaal et T. Robinson (2008). « Long-Term Monitoring of the Arrival, Expansion and Effects of the Alien Mussel *Mytilus Galloprovincialis* Relative to Wave Action ». *Marine Ecology Progress Series* 370, p. 171-183. DOI : 10.3354/meps07626.
- Branch, G. M. et C. N. Steffani (2004). « Can We Predict the Effects of Alien Species? A Case-History of the Invasion of South Africa by *Mytilus Galloprovincialis* (Lamarck) ». *Journal of Experimental Marine Biology and Ecology* 300, p. 189-215. DOI : 10.1016/j.jembe.2003.12.007.
- Brannock, P. M., D. S. Wetthey et T. J. Hilbish (2009). « Extensive Hybridization with Minimal Introgression in *Mytilus Galloprovincialis* and *M. Trossulus* in Hokkaido, Japan ». *Marine Ecology Progress Series* 383, p. 161-171. DOI : 10.3354/meps07995.
- Brannock, P. et T. Hilbish (2010). « Hybridization Results in High Levels of Sterility and Restricted Introgression between Invasive and Endemic Marine Blue Mussels ». *Marine Ecology Progress Series* 406, p. 161-171. DOI : 10.3354/meps08522.
- Britton-Davidian, J., F. Fel-Clair, J. Lopez, P. Alibert et P. Boursot (2005). « Postzygotic Isolation between the Two European Subspecies of the House Mouse : Estimates from Fertility Patterns in Wild and Laboratory-Bred Hybrids ». *Biological Journal of the Linnean Society* 84.3, p. 379-393. DOI : 10.1111/j.1095-8312.2005.00441.x.
- Burri, R. (2017). « Dissecting Differentiation Landscapes : A Linked Selection's Perspective ». *Journal of Evolutionary Biology* 30.8, p. 1501-1505. DOI : 10.1111/jeb.13108.
- Butler, D. (1994). « Bid to Protect Wolves from Genetic Pollution ». *Nature* 370.6490, p. 497-497. DOI : 10.1038/370497a0.
- Cáceres-Martínez, J. (1997). « Mussel Fishery and Culture in Baja California, Mexico : History, Present Status, and Future ». *The History, Present Condition, and Future of the Molluscan Fisheries of North and Central America and Europe, Volume 2, Pacific Coast*. Sous la dir. de C. L. J. MacKenzie, V. G. J. Burrell, A. Rosenfield et W. L. Hobart. T. 2. NOAA Technical Report NMFS 128.
- Campbell, C. R., J. W. Poelstra et A. D. Yoder (2018). « What Is Speciation Genomics? The Roles of Ecology, Gene Flow, and Genomic Architecture in the Formation of Species ». *Biological Journal of the Linnean Society* 124.4, p. 561-583. DOI : 10.1093/biolinnea/bly063.
- Carlton, J. T. et J. Hodder (1995). « Biogeography and Dispersal of Coastal Marine Organisms : Experimental Studies on a Replica of a 16th-Century Sailing Vessel ». *Marine Biology* 121.4, p. 721-730. DOI : 10.1007/BF00349308.
- Carlton, J. T. (1996). « Biological Invasions and Cryptogenic Species ». *Ecology* 77.6, p. 1653-1655. DOI : 10.2307/2265767.
- (2009). « Deep Invasion Ecology and the Assembly of Communities in Historical Time ». *Biological Invasions in Marine Ecosystems*. Sous la dir. de G. Rilov et J. A. Crooks. Réd.

- par M. M. Caldwell, G. Heldmaier, R. B. Jackson, O. L. Lange, H. A. Mooney, E. .-.-D. Schulze et U. Sommer. T. 204. Berlin, Heidelberg : Springer Berlin Heidelberg, p. 13-56. DOI : 10.1007/978-3-540-79236-9_2.
- Carlton, J. T., J. B. Geller, M. L. Reaka-Kudla et E. A. Norse (1999). « Historical Extinctions in the Sea ». *Annual Review of Ecology and Systematics* 30.1, p. 515-538. DOI : 10.1146/annurev.ecolsys.30.1.515.
- Case, A. L. et J. H. Willis (2008). « Hybrid Male Sterility in *Mimulus* (Phrymaceae) Is Associated with a Geographically Restricted Mitochondrial Rearrangement ». *Evolution* 62.5, p. 1026-1039. DOI : 10.1111/j.1558-5646.2008.00360.x.
- Casoli, E., D. Ventura, M. V. Modica, A. Belluscio, M. Capello, M. Oliverio et G. D. Ardizzone (2016). « A Massive Ingression of the Alien Species *Mytilus Edulis* L. (Bivalvia : Mollusca) into the Mediterranean Sea Following the Costa Concordia Cruise-Ship Disaster ». *Mediterranean Marine Science* 17.2, p. 404-416. DOI : 10.12681/mms.1619.
- Chapple, D. G., K. A. Miller, F. Kraus et M. B. Thompson (2013). « Divergent Introduction Histories among Invasive Populations of the Delicate Skink (*Lampropholis Delicata*) : Has the Importance of Genetic Admixture in the Success of Biological Invasions Been Overemphasized ? » *Diversity and Distributions* 19.2, p. 134-146. DOI : 10.1111/j.1472-4642.2012.00919.x.
- Charlesworth, B., M. T. Morgan et D. Charlesworth (1993). « The Effect of Deleterious Mutations on Neutral Molecular Variation ». 134, p. 1289-1303.
- Charlesworth, B., D. Charlesworth, J. A. Coyne et C. H. Langley (2016). « Hubby and Lewontin on Protein Variation in Natural Populations : When Molecular Genetics Came to the Rescue of Population Genetics ». *Genetics* 203.4, p. 1497-1503. DOI : 10.1534/genetics.115.185975.
- Chenuil, A., A. E. Cahill, N. Délémontey, E. Du Salliant du Luc et H. Fanton (2019). « Problems and Questions Posed by Cryptic Species. A Framework to Guide Future Studies ». *From Assessing to Conserving Biodiversity : Conceptual and Practical Challenges*. Sous la dir. d'E. Casetta, J. Marques da Silva et D. Vecchi. Cham : Springer International Publishing, p. 77-106. DOI : 10.1007/978-3-030-10991-2_4.
- Chevin, L.-M., G. Decorzent et T. Lenormand (2014). « Niche Dimensionality and the Genetics of Ecological Speciation ». *Evolution* 68.5, p. 1244-1256. DOI : 10.1111/evo.12346.
- Christe, C., K. N. Stölting, M. Paris, C. Fra sse, N. Bierne et C. Lexer (2017). « Adaptive Evolution and Segregating Load Contribute to the Genomic Landscape of Divergence in Two Tree Species Connected by Episodic Gene Flow ». *Molecular Ecology* 26.1, p. 59-76. DOI : 10.1111/mec.13765.
- Clarke Murray, C., E. A. Pakhomov et T. W. Therriault (2011). « Recreational Boating : A Large Unregulated Vector Transporting Marine Invasive Species : Transport of NIS by Recreational Boats ». *Diversity and Distributions* 17.6, p. 1161-1172. DOI : 10.1111/j.1472-4642.2011.00798.x.
- Colautti, R. I., J. M. Alexander, K. M. Dlugosch, S. R. Keller et S. E. Sultan (2017). « Invasions and Extinctions through the Looking Glass of Evolutionary Ecology ». *Philosophical Transactions of the Royal Society B : Biological Sciences* 372.1712, p. 20160031. DOI : 10.1098/rstb.2016.0031.
- Comesaña, A. S. et A. Sanjuan (1997). « Microgeographic Allozyme Differentiation in the Hybrid Zone of *Mytilus Galloprovincialis* Lmk. and *M. Edulis* L. on the Continental European Coast ». *Helgoländer Meeresunters* 51, p. 107-124. DOI : 10.1007/BF02908758.

- Comtet, T., A. Sandionigi, F. Viard et M. Casiraghi (2015). « DNA (Meta)Barcoding of Biological Invasions : A Powerful Tool to Elucidate Invasion Processes and Help Managing Aliens ». *Biological Invasions* 17.3, p. 905-922. DOI : 10.1007/s10530-015-0854-y.
- Consortium, H. G. (2012). « Butterfly Genome Reveals Promiscuous Exchange of Mimicry Adaptations among Species ». *Nature* 487, p. 5.
- Coolen, J. W. P. (2017). « North Sea Reefs : Benthic Biodiversity of Artificial and Rocky Reefs in the Southern North Sea ». PhD thesis. Wageningen : Wageningen University. URL : <http://edepot.wur.nl/404837>.
- Corbett-Detig, R. B., J. Zhou, A. G. Clark, D. L. Hartl et J. F. Ayroles (2013). « Genetic Incompatibilities Are Widespread within Species. » *Nature* 504.7478, p. 135-7. DOI : 10.1038/nature12678.
- Coustau, C., F. Renaud et B. Delay (1991). « Genetic Characterization of the Hybridization between *Mytilus Edulis* and *M. Galloprovincialis* on the Atlantic Coast of France ». *Marine Biology* 111.1, p. 87-93. DOI : 10.1007/BF01986350.
- Coutts, A. D. et T. J. Dodgshun (2007). « The Nature and Extent of Organisms in Vessel Sea-Chests : A Protected Mechanism for Marine Bioinvasions ». *Marine Pollution Bulletin* 54.7, p. 875-886. DOI : 10.1016/j.marpolbul.2007.03.011.
- Coyne, J. A. et H. A. Orr (2004). *Speciation*. T. 37. Sinauer Associates Sunderland, MA.
- Crego-Prieto, V., A. Ardura, F. Juanes, A. Roca, J. S. Taylor et E. Garcia-Vazquez (2015). « Aquaculture and the Spread of Introduced Mussel Genes in British Columbia ». *Biological Invasions* 17.7, p. 2011-2026. DOI : 10.1007/s10530-015-0853-z.
- Cruickshank, T. E. et M. W. Hahn (2014). « Reanalysis Suggests That Genomic Islands of Speciation Are Due to Reduced Diversity, Not Reduced Gene Flow ». *Molecular Ecology* 23.13, p. 3133-3157. DOI : 10.1111/mec.12796.
- Currat, M., M. Ruedi, R. J. Petit et L. Excoffier (2008). « The Hidden Side of Invasions : Massive Introgression by Local Genes ». *Evolution* 62, p. 1908-1920. DOI : 10.1111/j.1558-5646.2008.00413.x.
- Cutter, A. D. (2012). « The Polymorphic Prelude to Bateson–Dobzhansky–Muller Incompatibilities ». *Trends in Ecology & Evolution* 27.4, p. 209-218. DOI : 10.1016/j.tree.2011.11.004.
- Daguin, C. (2000). « Phylogéographie Des Moules Du Complexe d'espèces *Mytilus Edulis* ». Université de Montpellier.
- Daguin, C. et P. Borsa (2000). « Genetic Relationships of *Mytilus Galloprovincialis* Lamarck Populations Worldwide : Evidence from Nuclear-DNA Markers ». *Geological Society, London, Special Publications* 177.1, p. 389-397. DOI : 10.1144/GSL.SP.2000.177.01.26.
- Davis, A. W. et C.-I. Wu (1996). « The Broom of the Sorcerer's Apprentice : The Fine Structure of a Chromosomal Region Causing Reproductive Isolation Between Two Sibling Species of *Drosophila* ». *Genetics* 143.3, p. 1287-1298.
- Davis, M. A. et al. (2011). « Don't Judge Species on Their Origins ». *Nature* 474, p. 153-154. DOI : 10.1038/474153a.
- De Queiroz, K. (2007). « Species Concepts and Species Delimitation ». *Systematic Biology* 56.6, p. 879-886. DOI : 10.1080/10635150701701083.
- Dlugosch, K. M. et I. M. Parker (2008). « Founding Events in Species Invasions : Genetic Variation, Adaptive Evolution, and the Role of Multiple Introductions ». *Molecular Ecology* 17.1, p. 431-449. DOI : 10.1111/j.1365-294X.2007.03538.x.

- Dobzhansky, T. (1936). « Studies on Hybrid Sterility. II. Localization of Sterility Factors in *Drosophila Pseudoobscura* Hybrids ». *Genetics* 21.2, p. 113-135. URL : <https://www.ncbi.nlm.nih.gov/pubmed/17246786>.
- Drake, J. (2007). « Hull Fouling Is a Risk Factor for Intercontinental Species Exchange in Aquatic Ecosystems ». *Aquatic Invasions* 2.2, p. 121-131. DOI : 10.3391/ai.2007.2.2.7.
- Drake, J. M. et D. M. Lodge (2004). « Global Hot Spots of Biological Invasions : Evaluating Options for Ballast-Water Management ». *Proceedings of the Royal Society of London. Series B : Biological Sciences* 271, p. 575-580. DOI : 10.1098/rspb.2003.2629.
- Duranton, M., F. Allal, C. Fraïsse, N. Bierne, F. Bonhomme et P.-A. Gagnaire (2018). « The Origin and Remolding of Genomic Islands of Differentiation in the European Sea Bass ». *Nature Communications* 9, p. 2518. DOI : 10.1038/s41467-018-04963-6.
- Duvaux, L., K. Belkhir, M. Boulesteix et P. Boursot (2011). « Isolation and Gene Flow : Inferring the Speciation History of European House Mice ». *Molecular Ecology* 20.24, p. 5248-5264. DOI : 10.1111/j.1365-294X.2011.05343.x.
- Ehrenfeld, J. G. (2010). « Ecosystem Consequences of Biological Invasions ». *Annual Review of Ecology, Evolution, and Systematics* 41.1, p. 59-80. DOI : 10.1146/annurev-ecolsys-102209-144650.
- El Ayari, T., N. Trigui El Menif, B. Hamer, A. E. Cahill et N. Bierne (2019). « The Hidden Side of a Major Marine Biogeographic Boundary : A Wide Mosaic Hybrid Zone at the Atlantic-Mediterranean Divide Reveals the Complex Interaction between Natural and Genetic Barriers in Mussels ». *Heredity* 122, p. 770-784. DOI : 10.1038/s41437-018-0174-y.
- Elgvin, T. O. et al. (2017). « The Genomic Mosaicism of Hybrid Speciation ». *Science Advances* 3.6. DOI : 10.1126/sciadv.1602996.
- Ellegren, H. et al. (2012). « The Genomic Landscape of Species Divergence in *Ficedula* Flycatchers ». *Nature* 491.7426, p. 756-760. DOI : 10.1038/nature11584.
- Ellstrand, N. C. et K. A. Schierenbeck (2000). « Hybridization as a Stimulus for the Evolution of Invasiveness in Plants? ». *Proceedings of the National Academy of Sciences* 97.13, p. 7043-7050. DOI : 10.1073/pnas.97.13.7043.
- Elton, C. S. (1958). *The Ecology of Invasions by Animals and Plants*. University of Chicago Press.
- Excoffier, L., I. Dupanloup, E. Huerta-Sánchez, V. C. Sousa et M. Foll (2013). « Robust Demographic Inference from Genomic and SNP Data ». *PLoS Genetics* 9.10. Sous la dir. de J. M. Akey, e1003905. DOI : 10.1371/journal.pgen.1003905.
- Facon, B., B. Genton, J. Shykoff, P. Jarne, A. Estoup et P. David (2006). « A General Eco-Evolutionary Framework for Understanding Bioinvasions ». *Trends in Ecology & Evolution* 21.3, p. 130-135. DOI : 10.1016/j.tree.2005.10.012.
- Facon, B., J.-P. Pointier, P. Jarne, V. Sarda et P. David (2008). « High Genetic Variance in Life-History Strategies within Invasive Populations by Way of Multiple Introductions ». *Current Biology* 18.5, p. 363-367. DOI : 10.1016/j.cub.2008.01.063.
- FAO (2018). *FAO Yearbook. Fishery and Aquaculture Statistics 2016*. Roma. 104 p.
- Ferrario, J., S. Caronni, A. Occhipinti-Ambrogi et A. Marchini (2017). « Role of Commercial Harbours and Recreational Marinas in the Spread of Non-Indigenous Fouling Species ». *Biofouling* 33.8, p. 651-660. DOI : 10.1080/08927014.2017.1351958.

- Ferris, C., R. King et A. Gray (1997). « Molecular Evidence for the Maternal Parentage in the Hybrid Origin of *Spartina Anglica* CE Hubbard ». *Molecular Ecology* 6.2, p. 185-187.
- Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford, UK : Clarendon Press.
- (1937). « The Wave of Advance of Advantageous Genes ». *Annals of eugenics* 7.4, p. 355-369. URL : <http://onlinelibrary.wiley.com/doi/10.1111/j.1469-1809.1937.tb02153.x/full> (visité le 07/03/2017).
- Fishman, L. et J. H. Willis (2001). « Evidence for Dobzhansky-Muller Incompatibilities Contributing to the Sterility of Hybrids between *Mimulus Guttatus* and *M. Nasutus* ». *Evolution* 55.10, p. 1932-1942. DOI : 10.1111/j.0014-3820.2001.tb01311.x.
- Fitzpatrick, B. M., J. R. Johnson, D. K. Kump, H. B. Shaffer, J. J. Smith et S. R. Voss (2009). « Rapid Fixation of Non-Native Alleles Revealed by Genome-Wide SNP Analysis of Hybrid Tiger Salamanders ». *BMC Evolutionary Biology* 9.1, p. 176.
- Flagella, M. M., M. Verlaque, A. Soria et M. C. Buia (2007). « Macroalgal Survival in Ballast Water Tanks ». *Marine Pollution Bulletin* 54.9, p. 1395-1401. DOI : 10.1016/j.marpolbul.2007.05.015.
- Fraïsse, C., K. Belkhir, J. J. Welch et N. Bierne (2016a). « Local Interspecies Introgression Is the Main Cause of Extreme Levels of Intraspecific Differentiation in Mussels ». *Molecular Ecology* 25.1, p. 269-286. DOI : 10.1111/mec.13299.
- Fraïsse, C., P. A. Gunnarsson, D. Roze, N. Bierne et J. J. Welch (2016b). « The Genetics of Speciation : Insights from Fisher's Geometric Model ». *Evolution* 70.7, p. 1450-1464. DOI : 10.1111/evo.12968.
- Fraïsse, C., A. Haguénauer, K. Gerard, A. A.-T. Weber, N. Bierne et A. Chenuil (2018a). « Fine-Grained Habitat-Associated Genetic Connectivity in an Admixed Population of Mussels in the Small Isolated Kerguelen Islands ». *bioRxiv* 239244, ver. 4 recommended and peer-reviewed by PCI Evol Biol. DOI : 10.1101/239244.
- Fraïsse, C., C. Roux, P.-A. Gagnaire, J. Romiguier, N. Faivre, J. J. Welch et N. Bierne (2018b). « The Divergence History of European Blue Mussel Species Reconstructed from Approximate Bayesian Computation : The Effects of Sequencing Techniques and Sampling Strategies ». *PeerJ* 6, e5198. DOI : 10.7717/peerj.5198.
- Gardner, J. P. A. (1994). « The Structure and Dynamics of Naturally Occurring Hybrid *Mytilus Edulis* (Linnaeus, 1758) and *Mytilus Galloprovincialis* (Lamarck, 1819) Populations : Review and Interpretation. » *Archiv Fur Hydrobiologie* 1-2, p. 37-71.
- Gardner, J. P. A. et D. O. F. Skibinski (1990). « Genotype-Dependent Fecundity and Temporal Variation of Spawning in Hybrid Mussel (*Mytilus*) Populations ». *Marine Biology* 105.1, p. 153-162. DOI : 10.1007/BF01344281.
- Gardner, J. P. A., D. O. F. Skibinski et C. D. Bajdik (1993). « Shell Growth and Viability Differences Between the Marine Mussels *Mytilus Edulis* (L.), *Mytilus Galloprovincialis* (Lmk.), and Their Hybrids From Two Sympatric Populations in S.W. England ». *The Biological Bulletin* 185.3, p. 405-416. DOI : 10.2307/1542481.
- Gardner, J. P. A., M. Zbawicka, K. M. Westfall et R. Wenne (2016). « Invasive Blue Mussels Threaten Regional Scale Genetic Diversity in Mainland and Remote Offshore Locations : The Need for Baseline Data and Enhanced Protection in the Southern Ocean ». *Global Change Biology*. DOI : 10.1111/gcb.13332.
- Gavrilets, S. et A. Hastings (1996). « Founder Effect Speciation : A Theoretical Reassessment ». *The American Naturalist* 147.3, p. 466-491. DOI : 10.1086/285861.

- Geller, J. B., J. T. Carlton et D. A. Powers (1994). « PCR-Based Detection of mtDNA Haplotypes of Native and Invading Mussels on the Northeastern Pacific Coast : Latitudinal Pattern of Invasion ». *Marine Biology* 119.2, p. 243-249. DOI : 10.1007/BF00349563.
- Gérard, K., N. Bierne, P. Borsa, A. Chenuil et J.-P. Féral (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* 49.1, p. 84-91. DOI : 10.1016/j.ympev.2008.07.006.
- Gerdol, M. et al. (2019). « Massive Gene Presence/Absence Variation in the Mussel Genome as an Adaptive Strategy : First Evidence of a Pan-Genome in Metazoa ». *bioRxiv*. DOI : 10.1101/781377.
- Gosling, E., S. Doherty et N. Howley (2008). « Genetic Characterization of Hybrid Mussel (*Mytilus*) Populations on Irish Coasts ». *Journal of the Marine Biological Association of the United Kingdom* 88.2, p. 341-346. DOI : 10.1017/S0025315408000957.
- Gosset, C. C. et N. Bierne (2013). « Differential Introgression from a Sister Species Explains High F_{ST} Outlier Loci within a Mussel Species ». *Journal of Evolutionary Biology* 26.1, p. 14-26. DOI : 10.1111/jeb.12046.
- Grant, W. S. et M. I. Cherry (1985). « *Mytilus Galloprovincialis* Lmk. in Southern Africa ». *J. Exp. Mar. Biol. Ecol.* 90, p. 179-191. DOI : 10.1016/0022-0981(85)90119-4.
- Guerrero, R. F., C. D. Muir, S. Josway et L. C. Moyle (2017). « Pervasive Antagonistic Interactions among Hybrid Incompatibility Loci ». *PLoS Genetics* 13.6, e1006817. DOI : 10.1371/journal.pgen.1006817.
- Gutenkunst, R. N., R. D. Hernandez, S. H. Williamson et C. D. Bustamante (2009). « Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data ». *PLoS Genetics* 5.10. Sous la dir. de G. McVean, e1000695. DOI : 10.1371/journal.pgen.1000695.
- Haldane, J. B. (1948). « The Theory of a Cline ». *Journal of genetics* 48.3, p. 277-284. URL : <http://www.springerlink.com/index/F37743140767K133.pdf> (visité le 20/04/2017).
- Hall, R. J., A. Hastings et D. R. Ayres (2006). « Explaining the Explosion : Modelling Hybrid Invasions ». *Proceedings of the Royal Society of London B : Biological Sciences* 273, p. 1385-1389. DOI : 10.1098/rspb.2006.3473.
- Harr, B. (2006). « Genomic Islands of Differentiation between House Mouse Subspecies ». *Genome Research* 16.6, p. 730-737. DOI : 10.1101/gr.5045006.
- Harris, H. (1966). « Enzyme Polymorphisms in Man ». *Proc. R. Soc. Lond. B* 164.995, p. 298-310. DOI : 10.1098/rspb.1966.0032.
- Harrison, R. G. (1986). « Pattern and Process in a Narrow Hybrid Zone ». *Heredity* 56.3, p. 337-349. DOI : 10.1038/hdy.1986.55.
- Harrison, R. G. et E. L. Larson (2016). « Heterogeneous Genome Divergence, Differential Introgression, and the Origin and Structure of Hybrid Zones ». *Molecular Ecology* 25.11, p. 2454-2466. DOI : 10.1111/mec.13582.
- Heath, D. D., P. D. Rawson et T. J. Hilbish (1995). « PCR-Based Nuclear Markers Identify Alien Blue Mussel (*Mytilus* Spp.) Genotypes on the West Coast of Canada ». *Canadian Journal of Fisheries and Aquatic Sciences* 52.12, p. 2621-2627. DOI : 10.1139/f95-851.
- Hermansen, J. S., F. Haas, C. N. Trier et R. I. Bailey (2014). « Hybrid Speciation through Sorting of Parental Incompatibilities in Italian Sparrows ». *Molecular Ecology* 23, p. 5831-5842.

- Hewitt, G. M. (1988). « Hybrid Zones-Natural Laboratories for Evolutionary Studies ». *Trends in Ecology and Evolution* 3.7, p. 158-167. DOI : 10.1016/0169-5347(88)90033-X.
- (2000). « The Genetic Legacy of the Quaternary Ice Ages. » *Nature* 405.6789, p. 907-913. DOI : 10.1038/35016000.
- (2011). « Quaternary Phylogeography : The Roots of Hybrid Zones ». *Genetica* 139.5, p. 617-638. DOI : 10.1007/s10709-011-9547-3.
- Hilbish, T. J., A. Mullinax, S. I. Dolven, A. Meyer, R. K. Koehn et P. D. Rawson (2000). « Origin of the Antitropical Distribution Pattern in Marine Mussels (*Mytilus* Spp.) : Routes and Timing of Transequatorial Migration ». *Marine Biology* 136.1, p. 69-77. DOI : 10.1007/s002270050010.
- Hilbish, T., J. Timmons, V. Agrawal, K. Schneider et M. Gilg (2003). « Estuarine Habitats Protect Hybrid Mussels from Selection ». *Journal of Experimental Marine Biology and Ecology* 292.2, p. 177-186. DOI : 10.1016/S0022-0981(03)00161-8.
- Hill, W. G. et A. Robertson (1966). « The Effect of Linkage on Limits to Artificial Selection ». *Genetical Research* 8, p. 269-294. DOI : 10.1017/S001667230800949X.
- Holman, L. E., M. de Bruyn, S. Creer, G. Carvalho, J. Robidart et M. Rius (2019). « Detection of Introduced and Resident Marine Species Using Environmental DNA Metabarcoding of Sediment and Water ». *Scientific Reports* 9.1, p. 11559. DOI : 10.1038/s41598-019-47899-7.
- Hovick, S. M. et K. D. Whitney (2014). « Hybridisation Is Associated with Increased Fecundity and Size in Invasive Taxa : Meta-Analytic Support for the Hybridisation-Invasion Hypothesis ». *Ecology Letters* 17.11. Sous la dir. de J. Gurevitch, p. 1464-1477. DOI : 10.1111/ele.12355.
- Hubby, J. L. et R. C. Lewontin (1966). « A Molecular Approach to the Study of Genic Heterozygosity in Natural Populations. I. the Number of Alleles at Different Loci in *Drosophila Pseudoobscura* ». *Genetics* 54.2, p. 577-594.
- Janoušek, V. et al. (2012). « Genome-Wide Architecture of Reproductive Isolation in a Naturally Occurring Hybrid Zone between *Mus Musculus Musculus* and *M. m. Domesticus* : REPRODUCTIVE ISOLATION IN HOUSE MOUSE ». *Molecular Ecology* 21.12, p. 3032-3047. DOI : 10.1111/j.1365-294X.2012.05583.x.
- Jiggins, C. D. et J. Mallet (2000). « Bimodal Hybrid Zones and Speciation ». *Trends in Ecology & Evolution* 15.6, p. 250-255. DOI : 10.1016/S0169-5347(00)01873-5.
- Kartavtsev, Y. P., A. Y. Chichvarkhin, A. Kijima, N. Hanzawa et I.-S. Park (2005). « Allozyme and Morphometric Analysis of Two Common Mussel Species of the Genus *Mytilus* (Mollusca, Mytilidae) in Korean, Japanese and Russian Waters ». *Korean Journal of Genetics* 24.4, p. 289-306.
- Katolikova, M., V. Khaitov, R. Väinölä, M. Gantsevich et P. Strelkov (2016). « Genetic, Ecological and Morphological Distinctness of the Blue Mussels *Mytilus Trossulus* Gould and *M. Edulis* L. in the White Sea ». *PLOS ONE* 11.4. Sous la dir. de D. J. Colgan, e0152963. DOI : 10.1371/journal.pone.0152963.
- Kautsky, N. et S. Evans (1987). « Role of Biodeposition by *Mytilus Edulis* in the Circulation of Matter and Nutrients in a Baltic Coastal Ecosystem ». *Marine Ecology Progress Series* 38, p. 201-212. DOI : 10.3354/meps038201.
- Kéfi, S., V. Miele, E. A. Wieters, S. A. Navarrete et E. L. Berlow (2016). « How Structured Is the Entangled Bank? The Surprisingly Simple Organization of Multiplex Ecological

- Networks Leads to Increased Persistence and Resilience ». *PLOS Biology* 14.8. Sous la dir. de D. Simberloff, e1002527. DOI : 10.1371/journal.pbio.1002527.
- Kelly, D. W., R. A. Paterson, C. R. Townsend, R. Poulin et D. M. Tompkins (2009). « Parasite Spillover : A Neglected Concept in Invasion Ecology ? » *Ecology* 90.8, p. 2047-2056. DOI : 10.1890/08-1085.1.
- Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press.
- Kirkpatrick, M. et N. H. Barton (2006). « Chromosome Inversions, Local Adaptation and Speciation ». *Genetics* 173.1, p. 419-434. DOI : 10.1534/genetics.105.047985.
- Kirkpatrick, M. et V. Ravné (2002). « Speciation by Natural and Sexual Selection : Models and Experiments ». *The American Naturalist* 159.S3, S22-S35. DOI : 10.1086/338370.
- Klibansky, L. K. J. et M. A. McCartney (2014). « Conspecific Sperm Precedence Is a Reproductive Barrier between Free-Spawning Marine Mussels in the Northwest Atlantic Mytilus Hybrid Zone ». *PLoS ONE* 9.9. Sous la dir. de M. Hart, e108433. DOI : 10.1371/journal.pone.0108433.
- Knowlton, N. (1993). « Sibling Species in the Sea ». *Annual Review of Ecology and Systematics* 24, p. 189-216. DOI : 10.1146/annurev.es.24.110193.001201.
- Koehn, R. K. (1991). « The Genetics and Taxonomy of Species in the Genus Mytilus ». *Aquaculture* 94.2-3, p. 125-145. DOI : 10.1016/0044-8486(91)90114-M.
- Koehn, R. K., R. Milkman et J. B. Mitton (1976). « Population Genetics of Marine Pelecypods. IV. Selection, Migration and Genetic Differentiation in the Blue Mussel *Mytilus Edulis* ». *Evolution* 30.1, p. 2-32. DOI : 10.1111/j.1558-5646.1976.tb00878.x.
- Koehn, R. K. et J. B. Mitton (1972). « Population Genetics of Marine Pelecypods. I. Ecological Heterogeneity and Evolutionary Strategy at an Enzyme Locus ». *The American Naturalist* 106.947, p. 47-56. DOI : 10.1086/282750.
- Koehn, R. K., R. I. E. Newell et F. Immermann (1980). « Maintenance of an Aminopeptidase Allele Frequency Cline by Natural Selection ». *Proc. Natl. Acad. Sci. USA* 77.9, p. 5385-5389.
- Kolbe, J. J., R. E. Glor, L. Rodríguez Schettino, A. C. Lara, A. Larson et J. B. Losos (2004). « Genetic Variation Increases during Biological Invasion by a Cuban Lizard ». *Nature* 431.7005, p. 177-181. DOI : 10.1038/nature02807.
- Kruuk, L. E. B., S. J. E. Baird, K. S. Gale et N. H. Barton (1999). « A Comparison of Multilocus Clines Maintained by Environmental Adaptation or by Selection against Hybrids ». *Genetics* 153.4, p. 1959-1971. URL : <http://www.genetics.org/content/153/4/1959.short> (visité le 20/04/2017).
- Larraín, M. A., M. Zbawicka, C. Araneda, J. P. A. Gardner et R. Wenne (2018). « Native and Invasive Taxa on the Pacific Coast of South America : Impacts on Aquaculture, Traceability and Biodiversity of Blue Mussels (*Mytilus* Spp.) » *Evolutionary Applications* 11, p. 298-311. DOI : 10.1111/eva.12553.
- Larson, E. L. et al. (2018). « The Evolution of Polymorphic Hybrid Incompatibilities in House Mice ». *Genetics*, genetics.300840.2018. DOI : 10.1534/genetics.118.300840.
- Lee, S. et B. Morton (1985). « The Introduction of the Mediterranean Mussel *Mytilus Galloprovincialis* into Hong Kong ». *Malacological Review* 18.1-2, p. 107-109.
- Leitwein, M., P.-A. Gagnaire, E. Desmarais, P. Berrebi et B. Guinand (2018). « Genomic Consequences of a Recent Three-Way Admixture in Supplemented Wild Brown Trout Populations Revealed by Local Ancestry Tracts ». *Molecular Ecology* 27, p. 3466-3483. DOI : 10.1111/mec.14816.

- Lessios, H. A. (2009). « Speciation in Sea Urchins ». *Echinoderms : Durham : Proceedings of the 12th International Echinoderm Conference, 7-11 August 2006, Durham, New Hampshire, U.S.A.* Sous la dir. de L. Harris, S. Böttger, C. Walker et M. Lesser. CRC Press, p. 91-101. DOI : 10.1201/9780203869543.
- Levings, C. D., J. R. Cordell, S. Ong et G. E. Piercey (2004). « The Origin and Identity of Invertebrate Organisms Being Transported to Canada's Pacific Coast by Ballast Water ». *Canadian Journal of Fisheries and Aquatic Sciences* 61.1, p. 1-11. DOI : 10.1139/f03-135.
- Lindtke, D. et C. A. Buerkle (2015). « The Genetic Architecture of Hybrid Incompatibilities and Their Effect on Barriers to Introgression in Secondary Contact ». *Evolution* 69.8, p. 1987-2004. DOI : 10.1111/evo.12725.
- Lockwood, B. L. et G. N. Somero (2011). « Invasive and Native Blue Mussels (Genus *Mytilus*) on the California Coast : The Role of Physiology in a Biological Invasion ». *Journal of Experimental Marine Biology and Ecology* 400, p. 167-174. DOI : 10.1016/j.jembe.2011.02.022.
- Lockwood, J. L., M. F. Hoopes et M. P. Marchetti (2013). *Invasion Ecology*. John Wiley & Sons.
- Loire, E. et N. Galtier (2017). « Lacking Conservation Genomics in the Giant Galápagos Tortoise ». *bioRxiv* 101980, ver. 4 peer-reviewed and recommended by Peer Community in Evolutionary Biology. DOI : 10.1101/101980.
- López-Legentil, S., M. L. Legentil, P. M. Erwin et X. Turon (2015). « Harbor Networks as Introduction Gateways : Contrasting Distribution Patterns of Native and Introduced Ascidians ». *Biological Invasions* 17.6, p. 1623-1638. DOI : 10.1007/s10530-014-0821-z.
- Macholán, M., S. J. E. Baird, P. Dufková, P. Munclinger, B. V. Bímová et J. Piálek (2011). « Assessing Multilocus Introgression Patterns : A Case Study on the Mouse X Chromosome in Central Europe : Heterogeneity of Introgression on the Mouse X Chromosome ». *Evolution* 65.5, p. 1428-1446. DOI : 10.1111/j.1558-5646.2011.01228.x.
- Mallet, J. (2007). « Hybrid Speciation ». *Nature* 446.7133, p. 279-283. DOI : 10.1038/nature05706.
- Mani, G. S. et B. C. Clarke (1990). « Mutational Order : A Major Stochastic Process in Evolution ». *Proceedings of the Royal Society of London B : Biological Sciences* 240.1297, p. 29-37. DOI : 10.1098/rspb.1979.0086.
- Martin, S. H. et C. D. Jiggins (2017). « Interpreting the Genomic Landscape of Introgression ». *Current Opinion in Genetics & Development* 47, p. 69-74. DOI : 10.1016/j.gde.2017.08.007.
- Martínez-Lage, A., A. González-Tizón et J. Méndez (1996). « Chromosome Differences between European Mussel Populations (Genus *Mytilus*) ». *Caryologia* 49.3-4, p. 343-355. DOI : 10.1080/00087114.1996.10797379.
- Mathiesen, S. S. et al. (2016). « Genetic Diversity and Connectivity within *Mytilus* Spp. in the Subarctic and Arctic ». *Evolutionary Applications* 10, p. 39-55. DOI : 10.1111/eva.12415.
- Maynard-Smith, J. et J. Haigh (1974). « The Hitch-Hiking Effect of a Favourable Gene ». *Genetical Research* 23.1, p. 23-35. DOI : 10.1017/S0016672300014634.
- McDonald, J. H., R. Seed et R. K. Koehn (1991). « Allozymes and Morphometric Characters of Three Species of *Mytilus* in the Northern and Southern Hemispheres ». *Marine Biology* 111, p. 323-333. DOI : 10.1007/BF01319403.
- McDonald, J. H. et R. K. Koehn (1988). « The Mussels *Mytilus Galloprovincialis* and *M. Trossulus* on the Pacific Coast of North America ». *Marine Biology* 99.1, p. 111-118. DOI : 10.1007/BF00644984.

- McDonald, J. H., R. K. Koehn, E. S. Balakirev, G. P. Manchenko, A. I. Pudovkin, S. O. Sergievskii et K. V. Krutovskii (1990). « Species Identity of the "Common Mussel" Inhabiting the Asiatic Coasts of the Pacific Ocean. » *Soviet Journal of Marine Biology* 16.1, p. 10-18.
- McGrath, C. L., J.-F. Gout, P. Johri, T. G. Doak et M. Lynch (2014). « Differential Retention and Divergent Resolution of Duplicate Genes Following Whole-Genome Duplication ». *Genome Research* 24.10, p. 1665-1675. DOI : 10.1101/gr.173740.114.
- McKenzie, J. L., R. S. Dhillon et P. M. Schulte (2016). « Steep, Coincident, and Concordant Clines in Mitochondrial and Nuclear-Encoded Genes in a Hybrid Zone between Subspecies of Atlantic Killifish, *Fundulus Heteroclitus* ». *Ecology and Evolution* 6.16, p. 5771-5787. DOI : 10.1002/ece3.2324.
- Meier, J. I., D. A. Marques, C. E. Wagner, L. Excoffier et O. Seehausen (2018). « Genomics of Parallel Ecological Speciation in Lake Victoria Cichlids ». *Molecular Biology and Evolution* 35.6. Sous la dir. de D. Agashe, p. 1489-1506. DOI : 10.1093/molbev/msy051.
- Mesgaran, M. B., M. A. Lewis, P. K. Ades, K. Donohue, S. Ohadi, C. Li et R. D. Cousens (2016). « Hybridization Can Facilitate Species Invasions, Even without Enhancing Local Adaptation ». *Proceedings of the National Academy of Sciences* 113.36, p. 10210-10214. DOI : 10.1073/pnas.1605626113.
- Metzger, M. J. et S. P. Goff (2016). « A Sixth Modality of Infectious Disease : Contagious Cancer from Devils to Clams and Beyond ». *PLOS Pathogens* 12.10. Sous la dir. de D. C. Sheppard, e1005904. DOI : 10.1371/journal.ppat.1005904.
- Metzger, M. J. et al. (2016). « Widespread Transmission of Independent Cancer Lineages within Multiple Bivalve Species ». *Nature* 534, p. 705-709. DOI : 10.1038/nature18599.
- Milkman, R. et L. Beaty (1970). « Large-Scale Electrophoretic Studies of Allelic Variation in *Mytilus Edulis* ». *Biological Bulletin*. T. 139, p. 430.
- Miller, J. A., J. T. Carlton, J. W. Chapman, J. B. Geller et G. M. Ruiz (2017). « Transoceanic Dispersal of the Mussel *Mytilus Galloprovincialis* on Japanese Tsunami Marine Debris : An Approach for Evaluating Rafting of a Coastal Species at Sea ». *Marine Pollution Bulletin* 132, p. 60-69. DOI : 10.1016/j.marpolbul.2017.10.040.
- Miller, J. M. et al. (2018). « Genome-Wide Assessment of Diversity and Divergence Among Extant Galapagos Giant Tortoise Species ». *Journal of Heredity* 109.6, p. 611-619. DOI : 10.1093/jhered/esy031.
- Mineur, F., T. Belsher, M. P. Johnson, C. A. Maggs et M. Verlaque (2007). « Experimental Assessment of Oyster Transfers as a Vector for Macroalgal Introductions ». *Biological Conservation* 137.2, p. 237-247. DOI : 10.1016/j.biocon.2007.02.001.
- Mineur, F., M. P. Johnson et C. A. Maggs (2008). « Macroalgal Introductions by Hull Fouling on Recreational Vessels : Seaweeds and Sailors ». *Environmental Management* 42.4, p. 667-676. DOI : 10.1007/s00267-008-9185-4.
- Miralles, L., M. Gomez-Agenjo, F. Rayon-Viña, G. Gyraitė et E. Garcia-Vazquez (2018). « Alert Calling in Port Areas : Marine Litter as Possible Secondary Dispersal Vector for Hitchhiking Invasive Species ». *Journal for Nature Conservation* 42, p. 12-18. DOI : 10.1016/j.jnc.2018.01.005.
- Mlouka, R., J. Cachot, K. Boukadida, C. Clérandeau, P.-Y. Gourves et M. Banni (2019). « Compared Responses to Copper and Increased Temperatures of Hybrid and Pure Offspring of Two Mussel Species ». *Science of The Total Environment* 685, p. 795-805. DOI : 10.1016/j.scitotenv.2019.05.466.
- Molnar, J. L., R. L. Gamboa, C. Revenga et M. D. Spalding (2008). « Assessing the Global Threat of Invasive Species to Marine Biodiversity ». *Frontiers in Ecology and the Environment* 6.9, p. 485-492. DOI : 10.1890/070064.

- Mooney, H. A. et E. E. Cleland (2001). « The Evolutionary Impact of Invasive Species ». *Proceedings of the National Academy of Sciences* 98.10, p. 5446-5451. DOI : 10.1073/pnas.091093398.
- Morais, P. et M. Reichard (2018). « Cryptic Invasions : A Review ». *Science of The Total Environment* 613-614, p. 1438-1448. DOI : 10.1016/j.scitotenv.2017.06.133.
- Morán, T. et A. Fontdevila (2014). « Genome-Wide Dissection of Hybrid Sterility in *Drosophila* Confirms a Polygenic Threshold Architecture ». *Journal of Heredity* 105.3, p. 381-396. DOI : 10.1093/jhered/esu003.
- Moyle, L. C. et T. Nakazato (2008). « Complex Epistasis for Dobzhansky-Muller Hybrid Incompatibility in *Solanum* ». *Genetics* 181.1, p. 347-351. DOI : 10.1534/genetics.108.095679.
- Muller, H. J. (1942). « Isolating Mechanisms, Evolution, and Temperature ». *Biol. Symp.* 6, p. 71-125. URL : <https://ci.nii.ac.jp/naid/10024791088/en/>.
- Murgarella, M., D. Puiu, B. Novoa, A. Figueras, D. Posada et C. Canchaya (2016). « A First Insight into the Genome of the Filter-Feeder Mussel *Mytilus Galloprovincialis* ». *Plos One* 11.3, e0151561-e0151561. DOI : 10.1371/journal.pone.0151561.
- Myers, J. H., D. Simberloff, A. M. Kuris et J. R. Carey (2000). « Eradication Revisited : Dealing with Exotic Species ». *Trends in Ecology & Evolution* 15.8, p. 316-320. DOI : 10.1016/S0169-5347(00)01914-5.
- Nagylaki, T. et M. Moody (1980). « Diffusion Model for Genotype-Dependent Migration ». *Proceedings of the National Academy of Sciences* 77.8, p. 4842-4846. URL : <http://www.pnas.org/content/77/8/4842.short> (visité le 07/03/2017).
- Nguyen, T. T., B. J. Hayes et B. A. Ingram (2014). « Genetic Parameters and Response to Selection in Blue Mussel (*Mytilus Galloprovincialis*) Using a SNP-Based Pedigree ». *Aquaculture* 420-421, p. 295-301. DOI : 10.1016/j.aquaculture.2013.11.021.
- Noor, M. A. F., K. L. Grams, L. A. Bertucci, Y. Almendarez, J. Reiland et K. R. Smith (2001). « The Genetics of Reproductive Isolation and the Potential for Gene Exchange between *Drosophila Pseudoobscura* and *D. Persimilis* via Backcross Hybrid Males ». *Evolution* 55.3, p. 512-521. DOI : 10.1554/0014-3820(2001)055[0512:TGORIA]2.0.CO;2.
- Nosil, P. (2012). *Ecological Speciation*. OUP Oxford.
- Novak, S. et R. Kollár (2017). « Spatial Gene Frequency Waves Under Genotype-Dependent Dispersal ». *Genetics* 205.1, p. 367-374. DOI : 10.1534/genetics.116.193946.
- Nunes, A. L., S. Katsanevakis, A. Zenetos et A. C. Cardoso (2014). « Gateways to Alien Invasions in the European Seas ». *Aquatic Invasions* 9.2, p. 133-144. DOI : 10.3391/ai.2014.9.2.02.
- Ohta, T. (1973). « Slightly Deleterious Mutant Substitutions in Evolution ». *Nature* 246, p. 96-98.
- Orr, H. A., J. P. Masly et D. C. Presgraves (2004). « Speciation Genes ». *Current Opinion in Genetics & Development* 14.6, p. 675-679. DOI : 10.1016/j.gde.2004.08.009.
- Orr, H. A. (1990). « "Why Polyploidy Is Rarer in Animals Than in Plants" Revisited ». *The American Naturalist* 136.6, p. 759-770. DOI : 10.1086/285130.
- Orr, H. A. et S. Irving (2001). « Complex Epistasis and the Genetic Basis of Hybrid Sterility in the *Drosophila Pseudoobscura* Bogota-USA Hybridization ». *Genetics* 158.3, p. 1089-1100. URL : <http://www.genetics.org/content/158/3/1089>.

- Oyarzún, P. A., J. E. Toro, J. I. Cañete et J. P. A. Gardner (2016). « Bioinvasion Threatens the Genetic Integrity of Native Diversity and a Natural Hybrid Zone : Smooth-Shelled Blue Mussels (*Mytilus* Spp.) in the Strait of Magellan ». *Biological Journal of the Linnean Society* 117.3, p. 574-585. DOI : 10.1111/bij.12687.
- Piálek, J. et N. H. Barton (1997). « The Spread of an Advantageous Allele Across a Barrier : The Effects of Random Drift and Selection Against Heterozygotes ». *Genetics* 145.2, p. 493-504. URL : <http://www.genetics.org/content/145/2/493>.
- Piálek, J., H. C. Hauffe et J. B. Searle (2005). « Chromosomal Variation in the House Mouse ». *Biological Journal of the Linnean Society* 84.3, p. 535-563. DOI : 10.1111/j.1095-8312.2005.00454.x.
- Popovic, I., A. M. A. Matias, N. Bierne et C. Riginos (2019). « Twin Introductions by Independent Invader Mussel Lineages Are Both Associated with Recent Admixture with a Native Congener in Australia ». *Evolutionary Applications*, evo.12857. DOI : 10.1111/eva.12857.
- Prenter, J., C. MacNeil, J. T. Dick et A. M. Dunn (2004). « Roles of Parasites in Animal Invasions ». *Trends in Ecology & Evolution* 19.7, p. 385-390. DOI : 10.1016/j.tree.2004.05.002.
- Ptacek, M. B., H. C. Gerhardt et R. D. Sage (1994). « Speciation by Polyploidy in Treefrogs : Multiple Origins of the Tetraploid, *Hyla Versicolor* ». *Evolution* 48.3, p. 898-908. DOI : 10.1111/j.1558-5646.1994.tb01370.x.
- Quesada, H., R. Wenne et D. O. F. Skibinski (1995a). « Differential Introgression of Mitochondrial DNA across Species Boundaries within the Marine Mussel Genus *Mytilus* ». *Proceedings of the Royal Society B : Biological Sciences* 262.1363, p. 51-56. DOI : 10.1098/rspb.1995.0175.
- Quesada, H., C. M. Beynon et D. O. F. Skibinski (1995b). « A Mitochondrial DNA Discontinuity in the Mussel *Mytilus Galloprovincialis* Lmk : Pleistocene Vicariance Biogeography and Secondary Intergradation. ». *Molecular Biology and Evolution* 12.3, p. 521-524. DOI : 10.1093/oxfordjournals.molbev.a040227.
- Quesada, H., C. Zapata et G. Alvarez (1995c). « A Multilocus Allozyme Discontinuity in the Mussel *Mytilus Galloprovincialis* : The Interaction of Ecological and Life-History Factors ». *Marine Ecology Progress Series* 116, p. 99-115. URL : <http://www.int-res.com/articles/meps/116/m116p099.pdf>.
- Ramírez, S. C. et J. Cáceres-Martínez (1999). « Settlement of the Blue Mussel *Mytilus Galloprovincialis* Lamarck on Artificial Substrates in Bahía de Todos Santos BC, Mexico ». *Journal of shellfish research* 18.1, p. 33-39.
- Ravinet, M. et al. (2017). « Interpreting the Genomic Landscape of Speciation : A Road Map for Finding Barriers to Gene Flow ». *Journal of Evolutionary Biology* 30.8, p. 1450-1477. DOI : 10.1111/jeb.13047.
- Rawson, P. D., P. O. Yund et C. Slaughter (2003). « Patterns of Gamete Incompatibility between the Blue Mussels *Mytilus Edulis* and *M. Trossulus* ». *Marine Biology* 143.2, p. 317-325. DOI : 10.1007/s00227-003-1084-x.
- Rawson, P. D. et T. J. Hilbish (1995). « Evolutionary Relationships among the Male and Female Mitochondrial DNA Lineages in the *Mytilus Edulis* Species Complex. ». *Molecular Biology and Evolution* 12.5, p. 893-901. DOI : 10.1093/oxfordjournals.molbev.a040266.

- Rhymer, J. M. et D. Simberloff (1996). « Extinction By Hybridization and Introgression ». *Annual Review of Ecology and Systematics* 27, p. 83-109. DOI : 10.1146/annurev.ecolsys.27.1.83.
- Ribeiro, Â. M., C. A. Canchaya, F. Penalzoa, J. Galindo et R. R. da Fonseca (2018). « Population Genomic Footprints of Environmental Pollution Pressure in Natural Populations of the Mediterranean Mussel ». *Marine Genomics*. DOI : 10.1016/j.margen.2018.10.009.
- Rieseberg, L. H., S.-C. Kim, R. A. Randell, K. D. Whitney, B. L. Gross, C. Lexer et K. Clay (2007). « Hybridization and the Colonization of Novel Habitats by Annual Sunflowers ». *Genetica* 129.2, p. 149-165. DOI : 10.1007/s10709-006-9011-y.
- Riquet, F., A. Simon et N. Bierne (2017). « Weird Genotypes ? Don't Discard Them, Transmissible Cancer Could Be an Explanation ». *Evolutionary Applications* 10, p. 140-145. DOI : 10.1111/eva.12439.
- Rius, M. et J. A. Darling (2014). « How Important Is Intraspecific Genetic Admixture to the Success of Colonising Populations ? ». *Trends in Ecology & Evolution* 29.4, p. 233-242. DOI : 10.1016/j.tree.2014.02.003.
- Roberts, D. (1976). « 3. Mussels and Pollution ». *Marine Mussels : Their Ecology and Physiology*. Sous la dir. de B. L. Bayne. Bayne B. L. T. 10. International Biological Programme. Cambridge University Press, p. 81-120.
- Romiguier, J. et al. (2014). « Comparative Population Genomics in Animals Uncovers the Determinants of Genetic Diversity ». *Nature* 515.7526, p. 261-263. DOI : 10.1038/nature13685.
- Roques, L. (2013). *Modèles de Réaction-Diffusion Pour l'écologie Spatiale*. Editions Quae.
- Rougeux, C., L. Bernatchez et P.-A. Gagnaire (2017). « Modeling the Multiple Facets of Speciation-with-Gene-Flow toward Inferring the Divergence History of Lake Whitefish Species Pairs (*Coregonus Clupeaformis*) ». *Genome Biology and Evolution* 9.8, p. 2057-2074. DOI : 10.1093/gbe/evx150.
- Roux, C., C. Fraïsse, V. Castric, X. Vekemans, G. H. Pogson et N. Bierne (2014). « Can We Continue to Neglect Genomic Variation in Introgression Rates When Inferring the History of Speciation ? A Case Study in a *Mytilus* Hybrid Zone ». *Journal of Evolutionary Biology* 27.8, p. 1662-1675. DOI : 10.1111/jeb.12425.
- Roux, C., C. Fraïsse, J. Romiguier, Y. Anciaux, N. Galtier et N. Bierne (2016). « Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence ». *PLoS Biology* 14.12. Sous la dir. de C. Moritz, e2000234. DOI : 10.1371/journal.pbio.2000234.
- Ruesink, J. L., H. S. Lenihan, A. C. Trimble, K. W. Heiman, F. Micheli, J. E. Byers et M. C. Kay (2005). « Introduction of Non-Native Oysters : Ecosystem Effects and Restoration Implications ». *Annual Review of Ecology, Evolution, and Systematics* 36.1, p. 643-689. DOI : 10.1146/annurev.ecolsys.36.102003.152638.
- Ruiz, G. M., P. W. Fofonoff, J. T. Carlton, M. J. Wonham et A. H. Hines (2000). « Invasion of Coastal Marine Communities in North America : Apparent Patterns, Processes, and Biases ». *Annual Review of Ecology and Systematics* 31.1, p. 481-531. DOI : 10.1146/annurev.ecolsys.31.1.481.
- Saarman, N. P. et G. H. Pogson (2015). « Introgression between Invasive and Native Blue Mussels (Genus *Mytilus*) in the Central California Hybrid Zone ». *Molecular Ecology* 24.18, p. 4723-4738. DOI : 10.1111/mec.13340.
- Saavedra, C. et E. Bachère (2006). « Bivalve Genomics ». *Aquaculture* 256.1-4, p. 1-14. DOI : 10.1016/j.aquaculture.2006.02.023.

- Sachdeva, H. et N. H. Barton (2018a). « Introgression of a Block of Genome under Infinitesimal Selection ». *Genetics* 209.4, p. 1279-1303. DOI : 10.1534/genetics.118.301018.
- (2018b). « Replicability of Introgression under Linked, Polygenic Selection ». *Genetics*. DOI : 10.1534/genetics.118.301429.
- Sakai, A. K. et al. (2001). « The Population Biology of Invasive Species ». *Annual review of ecology and systematics* 32.1, p. 305-332. URL : <http://www.annualreviews.org/doi/abs/10.1146/annurev.ecolsys.32.081501.114037> (visité le 05/07/2017).
- Sanjuan, A., C. Zapata et G. Alvarez (1994). « *Mytilus Galloprovincialis* and *M. Edulis* on the Coasts of the Iberian Peninsula ». *Marine Ecology Progress Series* 113, p. 131-146. DOI : 10.3354/meps113131.
- Sax, D. F. et J. H. Brown (2000). « The Paradox of Invasion ». *Global Ecology and Biogeography* 9.5, p. 363-371. URL : <http://onlinelibrary.wiley.com/doi/10.1046/j.1365-2699.2000.00217.x/full> (visité le 05/07/2017).
- Schierenbeck, K. A. et N. C. Ellstrand (2009). « Hybridization and the Evolution of Invasiveness in Plants and Other Organisms ». *Biological Invasions* 11.5, p. 1093-1105. DOI : 10.1007/s10530-008-9388-x.
- Schluter, D. (2009). « Evidence for Ecological Speciation and Its Alternative ». *Science* 323.5915, p. 737-741. DOI : 10.1126/science.1160006.
- Schneemann, H., B. De Sanctis et J. J. Welch (in prep.). « The Genetics and Geometry of Hybridization ». En prép.
- Schumer, M., R. Cui, D. L. Powell, R. Dresner, G. G. Rosenthal et P. Andolfatto (2014a). « High-Resolution Mapping Reveals Hundreds of Genetic Incompatibilities in Hybridizing Fish Species ». *Elife* 3, e02535. DOI : 10.7554/eLife.02535.001.
- Schumer, M., G. G. Rosenthal et P. Andolfatto (2014b). « How Common Is Homoploid Hybrid Speciation ? » *Evolution* 68.6, p. 1553-1560. DOI : 10.1111/evo.12399.
- (2018a). « What Do We Mean When We Talk about Hybrid Speciation ? » *Heredity* 120.4, p. 379-382. DOI : 10.1038/s41437-017-0036-z.
- Schumer, M. et al. (2018b). « Natural Selection Interacts with Recombination to Shape the Evolution of Hybrid Genomes ». *Science* 360.6389, p. 656-660. DOI : 10.1126/science.aar3684.
- Seed, R. (1976). « 1. Ecology ». *Marine Mussels : Their Ecology and Physiology*. Sous la dir. de B. L. Bayne. International Biological Programme. Cambridge University Press, p. 13-65.
- Seehausen, O. et al. (2014). « Genomics and the Origin of Species. » *Nature reviews. Genetics* 15.3, p. 176-92. DOI : 10.1038/nrg3644.
- Simberloff, D. (2009). « The Role of Propagule Pressure in Biological Invasions ». *Annual Review of Ecology, Evolution, and Systematics* 40.1, p. 81-102. DOI : 10.1146/annurev.ecolsys.110308.120304.
- Simon, A., N. Bierne et J. J. Welch (2018). « Coadapted Genomes and Selection on Hybrids : Fisher's Geometric Model Explains a Variety of Empirical Patterns ». *Evolution Letters* 2.5, p. 472-498. DOI : 10.1002/evl3.66.
- Simon, A. et M. Duranton (2018). « Digest : Demographic Inferences Accounting for Selection at Linked Sites ». *Evolution*. DOI : 10.1111/evo.13504.
- Simon, A., C. Fraïsse, T. El Ayari, C. Liutard-Haag, P. Strelkov, J. J. Welch et N. Bierne (2019a). « Local Introgression at Two Spatial Scales in Mosaic Hybrid Zones of Mussels ». *bioRxiv*. DOI : 10.1101/818559. Soumis.
- Simon, A. et al. (2019b). « Replicated Anthropogenic Hybridisations Reveal Parallel Patterns of Admixture in Marine Mussels ». *Evolutionary Applications*. DOI : 10.1111/eva.12879. Prépubl.

- Skibinski, D. O. F., M. Ahmad et J. A. Beardmore (1978a). « Genetic Evidence of Naturally Occurring Hybrids between *Mytilus Edulis* and *M. Galloprovincialis* ». 32.2, p. 354-364. DOI : 10.1038/nature01240.
- Skibinski, D. O. F. et J. A. Beardmore (1979). « A Genetic Study of Intergradation between *Mytilus Edulis* and *Mytilus Galloprovincialis* ». *Experientia* 35.11, p. 1442-1444. DOI : 10.1007/BF01962773.
- Skibinski, D. O. F., J. A. Beardmore et M. Ahmad (1978b). « Genetic Aids to the Study of Closely Related Taxa of the Genus *Mytilus* ». *Marine Organisms : Genetics, Ecology, and Evolution*. Sous la dir. de B. Battaglia et J. A. Beardmore. Proceedings of a NATO Advanced Study Research Inst. on the Genetics, Evolution, and Ecology of Marine Organisms Held in the Fondazione Giorgio Cini, Venice, 1977. Plenum Publishing Corporation, p. 469-486.
- Skurikhina, L. A., Y. F. Kartavtsev, A. Y. Chichvarkhin et M. V. Pan'kova (2001). « Study of Two Species of Mussels, *Mytilus Trossulus* and *Mytilus Galloprovincialis* (Bivalvia, Mytilidae), and Their Hybrids in Peter the Great Bay of the Sea of Japan with the Use of PCR Markers ». 37.12, p. 4.
- Slatkin, M. et T. Maruyama (1975). « Genetic Drift in a Cline ». *Genetics* 81.1, p. 209-222. URL : <http://www.genetics.org/content/81/1/209.short> (visité le 10/04/2017).
- Slaughter, C., M. A. McCartney et P. O. Yund (2008). « Comparison of Gamete Compatibility Between Two Blue Mussel Species in Sympatry and in Allopatry ». *The Biological Bulletin* 214.1, p. 57-66. DOI : 10.2307/25066660.
- Smadja, C. et G. Ganem (2005). « Asymmetrical Reproductive Character Displacement in the House Mouse ». *Journal of Evolutionary Biology* 18.6, p. 1485-1493. DOI : 10.1111/j.1420-9101.2005.00944.x.
- Springer, S. A. et B. J. Crespi (2007). « Adaptive Gamete-Recognition Divergence in a Hybridizing *Mytilus* Population ». *Evolution* 61.4, p. 772-783. DOI : 10.1111/j.1558-5646.2007.00073.x.
- Stammnitz, M. R. et al. (2018). « The Origins and Vulnerabilities of Two Transmissible Cancers in Tasmanian Devils ». *Cancer Cell* 33.4, 607-619.e15. DOI : 10.1016/j.ccell.2018.03.013.
- Storchová, R., S. Gregorová, D. Buckiová, V. Kyselová, P. Divina et J. Forejt (2004). « Genetic Analysis of X-Linked Hybrid Sterility in the House Mouse ». *Mammalian Genome* 15.7. DOI : 10.1007/s00335-004-2386-0.
- Strelkov, P., M. Katolikova et R. Väinölä (2017). « Temporal Change of the Baltic Sea-North Sea Blue Mussel Hybrid Zone over Two Decades ». *Marine Biology* 164.11, p. 214. DOI : 10.1007/s00227-017-3249-z.
- Stuckas, H., K. Stoof, H. Quesada et R. Tiedemann (2009). « Evolutionary Implications of Discordant Clines across the Baltic *Mytilus* Hybrid Zone (*Mytilus Edulis* and *Mytilus Trossulus*) ». *Heredity* 103.2, p. 146-156. DOI : 10.1038/hdy.2009.37.
- Sylvester, F. et al. (2011). « Hull Fouling as an Invasion Vector : Can Simple Models Explain a Complex Problem ? » *Journal of Applied Ecology* 48.2, p. 415-423. DOI : 10.1111/j.1365-2664.2011.01957.x.
- Szymura, J. M. et N. H. Barton (1986). « Genetic Analysis of a Hybrid Zone Between the Fire-Bellied Toads, *Bombina Bombina* and *B. Variegata*, Near Cracow in Southern Poland ». *Evolution* 40.6, p. 1141-1159.
- Templeton, A. R. (2008). « The Reality and Importance of Founder Speciation in Evolution ». *BioEssays* 30.5, p. 470-479. DOI : 10.1002/bies.20745.

- Tenaillon, O., O. K. Silander, J.-P. Uzan et L. Chao (2007). « Quantifying Organismal Complexity Using a Population Genetic Approach ». *PLoS ONE* 2.2, e217. DOI : 10.1371/journal.pone.0000217.
- Thomas, F., T. Lefevre et M. Raymond (2016). *Biologie Évolutive*. Biologie. De Boeck supérieur. URL : <https://books.google.fr/books?id=OMFEDQAAQBAJ>.
- Thompson, K. (2014). *Where Do Camels Belong?* Greystone books.
- Tincu, J. A. et S. W. Taylor (2004). « Antimicrobial Peptides from Marine Invertebrates ». *Antimicrobial Agents and Chemotherapy* 48.10, p. 3645-3654. DOI : 10.1128/AAC.48.10.3645-3654.2004.
- Todesco, M. et al. (2019). « Massive Haplotypes Underlie Ecotypic Differentiation in Sunflowers ». *bioRxiv*. DOI : 10.1101/790279.
- Torchin, M. E., K. D. Lafferty et A. M. Kuris (2002). « Parasites and Marine Invasions ». *Parasitology* 124.7, p. 137-151. DOI : 10.1017/S0031182002001506.
- Tóth, E. G., Á. Bede-Fazekas, G. G. Vendramin, F. Bagnoli et M. Höhn (2017). « Mid-Pleistocene and Holocene Demographic Fluctuation of Scots Pine (*Pinus Sylvestris* L.) in the Carpathian Mountains and the Pannonian Basin : Signs of Historical Expansions and Contractions ». *Quaternary International*. DOI : 10.1016/j.quaint.2017.11.024.
- Trier, C. N. et J. S. Hermansen (2014). « Evidence for Mito-Nuclear and Sex-Linked Reproductive Barriers between the Hybrid Italian Sparrow and Its Parent Species ». *PLOS Genetics* 10.1, p. 10.
- Turelli, M. et H. A. Orr (1995). « The Dominance Theory of Haldane's Rule. » *Genetics* 140.1, p. 389-402. URL : <http://www.genetics.org/content/140/1/389.short> (visité le 19/10/2016).
- (2000). « Dominance, Epistasis and the Genetics of Postzygotic Isolation ». *Genetics* 154.4, p. 1663. URL : <http://www.genetics.org/content/154/4/1663.abstract>.
- Turner, L. M. et B. Harr (2014). « Genome-Wide Mapping in a House Mouse Hybrid Zone Reveals Hybrid Sterility Loci and Dobzhansky-Muller Interactions ». *Elife* 3, e02504. DOI : 10.7554/eLife.02504.
- Turner, T. L., M. W. Hahn et S. V. Nuzhdin (2005). « Genomic Islands of Speciation in *Anopheles Gambiae* ». *PLoS Biology* 3.9, p. 1572-1578. DOI : 10.1371/journal.pbio.0030285.
- Uliano-Silva, M. et al. (2018). « A Hybrid-Hierarchical Genome Assembly Strategy to Sequence the Invasive Golden Mussel, *Limnoperna Fortunei* ». *GigaScience* 7.2. DOI : 10.1093/gigascience/gix128.
- Väinölä, R. et M. M. Hvilsom (1991). « Genetic Divergence and a Hybrid Zone between Baltic and North Sea *Mytilus* Populations (Mytilidae : Mollusca) ». *Biological Journal of the Linnean Society* 43.2, p. 127-148. DOI : 10.1111/j.1095-8312.1991.tb00589.x.
- Van Belleghem, S. M., K. De Wolf et F. Hendrickx (2016). « Behavioral Adaptations Imply a Direct Link between Ecological Specialization and Reproductive Isolation in a Sympatrically Diverging Ground Beetle : BRIEF COMMUNICATION ». *Evolution* 70.8, p. 1904-1912. DOI : 10.1111/evo.12998.
- Van Belleghem, S. M. et al. (2018). « Patterns of Z Chromosome Divergence among *Heliconius* Species Highlight the Importance of Historical Demography ». *Molecular Ecology*. DOI : 10.1111/mec.14560.

- Varvio, S.-L., R. K. Koehn et R. Väinölä (1988). « Evolutionary Genetics of the *Mytilus Edulis* Complex in the North Atlantic Region ». *Marine Biology* 98.1, p. 51-60. DOI : 10.1007/BF00392658.
- Vermeij, G. J. (1992). « Trans-Equatorial Connections between Biotas in the Temperate Eastern Atlantic ». *Marine Biology* 112.2, p. 343-348. DOI : 10.1007/BF00702481.
- Viard, F., P. David et J. A. Darling (2016). « Marine Invasions Enter the Genomic Era : Three Lessons from the Past, and the Way Forward ». *Current Zoology* 62.6, p. 629-642. DOI : 10.1093/cz/zow053.
- Wang, S. et al. (2017). « Scallop Genome Provides Insights into Evolution of Bilaterian Karyotype and Development ». *Nature Ecology & Evolution* 1.5, p. 0120. DOI : 10.1038/s41559-017-0120.
- Wasson, K., C. J. Zabin, L. Bedinger, M. Cristina Diaz et J. S. Pearse (2001). « Biological Invasions of Estuaries without International Shipping : The Importance of Intraregional Transport ». *Biological Conservation* 102.2, p. 143-153. DOI : 10.1016/S0006-3207(01)00098-2.
- Węśławski, J. M. et L. Kotwicki (2018). « Macro-Plastic Litter, a New Vector for Boreal Species Dispersal on Svalbard ». *Polish Polar Research* 39.1, p. 165-174. DOI : 10.24425/118743.
- Westfall, K. M. et J. P. A. Gardner (2013). « Interlineage *Mytilus Galloprovincialis* Lmk. 1819 Hybridization Yields Inconsistent Genetic Outcomes in the Southern Hemisphere ». *Biological Invasions* 15.7, p. 1493-1506. DOI : 10.1007/s10530-012-0385-8.
- (2010). « Genetic Diversity of Southern Hemisphere Blue Mussels (*Bivalvia* : *Mytilidae*) and the Identification of Non-Indigenous Taxa ». *Biological Journal of the Linnean Society* 101.4, p. 898-909.
- Wilkins, N. P., K. Fujino et E. M. Gosling (1983). « The Mediterranean Mussel *Mytilus Galloprovincialis* Lmk. in Japan ». *Biological Journal of the Linnean Society* 20.4, p. 365-374. DOI : 10.1111/j.1095-8312.1983.tb01597.x.
- Williamson, M. (1996). *Biological Invasions*. Springer Science & Business Media.
- Wilson, J., I. Matejusova, R. E. McIntosh, S. Carboni et M. Bekaert (2018). « New Diagnostic SNP Molecular Markers for the *Mytilus* Species Complex ». *PLOS ONE* 13.7. Sous la dir. de T.-Y. Chiang, e0200654. DOI : 10.1371/journal.pone.0200654.
- Wolf, D. E., N. Takebayashi et L. H. Riesberg (2001). « Predicting the Risk of Extinction through Hybridization ». *Conservation Biology* 15.4, p. 1039-1053.
- Wolf, J. B. W. et H. Ellegren (2017). « Making Sense of Genomic Islands of Differentiation in Light of Speciation ». *Nature Reviews Genetics* 18.2, p. 87-100. DOI : 10.1038/nrg.2016.133.
- Wood, S. A. et al. (2019). « Considerations for Incorporating Real-Time PCR Assays into Routine Marine Biosecurity Surveillance Programmes : A Case Study Targeting the Mediterranean Fanworm (*Sabella Spallanzanii*) and Club Tunicate (*Styela Clava*) ». *Genome* 62.3, p. 137-146. DOI : 10.1139/gen-2018-0021.
- Yonemitsu, M. A. et al. (2019). « A Single Clonal Lineage of Transmissible Cancer Identified in Two Marine Mussel Species in South America and Europe ». *eLife* 8, e47788. DOI : 10.7554/eLife.47788.
- Zane, L. (2019). « A Genome-Wide Approach to the Phylogeography of the Mussel *Mytilus Galloprovincialis* in the Adriatic and the Black Seas ». *Frontiers in Marine Science* 6, p. 16.

- Zbawicka, M., A. Burzyński, D. O. F. Skibinski et R. Wenne (2010). « Scottish *Mytilus Trossulus* Mussels Retain Ancestral Mitochondrial DNA : Complete Sequences of Male and Female mtDNA Genomes ». *Gene* 456.1-2, p. 45-53. DOI : 10.1016/j.gene.2010.02.009.
- Zbawicka, M., J. P. A. Gardner et R. Wenne (2019). « Cryptic Diversity in Smooth-Shelled Mussels on Southern Ocean Islands : Connectivity, Hybridisation and a Marine Invasion ». *Frontiers in Zoology* 16, p. 32. DOI : 10.1186/s12983-019-0332-y.
- Zhang, G. et al. (2012). « The Oyster Genome Reveals Stress Adaptation and Complexity of Shell Formation ». *Nature* 490.7418, p. 49-54. DOI : 10.1038/nature11413.
- Zigler, K. S., M. A. McCartney, D. R. Levitan et H. A. Lessios (2005). « Sea Urchin Binding Divergence Predicts Gamete Compatibility ». *Evolution* 59.11, p. 2399-2404. DOI : 10.1111/j.0014-3820.2005.tb00949.x.

Remerciements

Je tiens tout d'abord à remercier tous les membres du jury, qui ont accepté d'évaluer mon travail. Je souhaite aussi remercier les membres de mon comité de thèse, John, Patrice, Frédérique, Éric, Denis et Benoît, pour avoir pris le temps de me conseiller dans la construction de ce projet de thèse.

La majeure partie de ma gratitude va à Nicolas (oui j'en garde un peu pour les autres). Je ne pense pas qu'un doctorant puisse espérer avoir un meilleur encadrant que toi, aussi bien humainement que scientifiquement. Merci de ton investissement hors norme pour transmettre tes connaissances avec passion. Ton investissement en tes étudiants va bien au delà d'une simple relation prof/élève et des heures régulières de travail, et pour cela je t'en suis très reconnaissant. J'espère pouvoir collaborer avec toi pendant encore très longtemps. Merci aussi à toi John, qui malgré l'absence de titre officiel, a été un réel deuxième encadrant. À l'image de Nico, tu combines qualités humaines et scientifiques. J'ai beaucoup apprécié passer du temps à Cambridge, à la fois pour l'ambiance très british de la ville mais surtout pour avoir pu collaborer avec toi. Tu as su amélioré la qualité de mon anglais et toujours réussi à faire ressortir clairement le message dans des ébauches de papiers désorganisés.

Merci bien sûr à l'ensemble des membres de l'équipe SEA (prochainement BEM), actuels ou passés, Pierre-Alexandre, François, Maud, Pierre, Adrien, Maurine, Christine, Florentine, Christelle et Ahmed, pour contribuer à la bonne ambiance et toujours être là pour filer un coup de main. Merci aussi à Sophie et Cathy, en souvenir de la bonne vieille époque de la station de Sète. C'est particulièrement agréable de pouvoir faire de la recherche dans une ambiance familiale et sans prise de tête. Merci pour les nombreuses discussions scientifiques ou non, à midi, autour d'un café ou sur la plage. Maud! 3 ans et demis déjà, en fait tu es un peu ma jumelle de thèse. Merci d'avoir partagé avec moi hauts et bas, et d'avoir toujours été là pour pouvoir s'énerver de quelque chose ensemble. Je te souhaite beaucoup de chances dans tes nouvelles aventures américaines. Merci aussi à toutes les personnes du bâtiment 24. Merci au reste de l'ISEM et aux doctorants et post-docs pour l'ambiance agréable de recherche, les journal clubs et bien sûr les légendaires week-ends d'intégration.

Petite partie séjour à Roscoff, où embruns et galettes bretonnes étaient au rendez-vous. Merci à Denis, Benoît et l'équipe UMI Biologie évolutive et écologie des algues qui m'ont accueilli pendant plusieurs mois à la station biologique de Roscoff. Bien que ce séjour roscovite n'ait pas encore porté ses fruits en matière de modélisation, il a permis la mise en place de beaucoup de questionnements qui je pense me suivront encore pendant longtemps. Merci beaucoup à toi Frédérique, pour les discussions sur les invasions et cette collaboration fructueuse sur les moules des docks. L'équipe BEM est particulièrement chanceuse d'accueillir bientôt une chercheuse comme toi. Merci à Éric, avec qui j'ai pu discuter des modèles océanographiques. Merci bien sûr à tous les doctorants de la station, avec qui on ne s'ennuie pas. Merci Camille de m'avoir laissé squatter chez toi et de m'avoir converti au mode de vie roscovite. De nature casanière, cet animal ne serais pas souvent sorti de sa tanière sans toi. Ce séjour à Roscoff m'aura aussi permis plusieurs escapades en rade de Brest, qui se sont avérées très fructueuses en terme de pêche aux moules (surtout que les gens de la ville ne m'ont pas pris mon panier).

Merci Hugo pour les discussions du soir et pour toujours chopper mes références, même si elles sont parfois fatigantes. On a eu de la chance de pouvoir se suivre depuis l'ENS jusqu'à Montpellier. Merci aux autres anciens potes, aux groupes de la "fine équipe" et "aplysies et politique", d'avoir notamment supporté mes partages de liens toujours plus débiles les uns que les autres, vive les bractées et les gros poissons.

Merci aux kayakistes de Montpellier, pour m'avoir intégré dans leur équipe et m'avoir

permis de reprendre ma passion d'enfance. Je n'avais jamais espéré pouvoir intégrer un jour la meilleure équipe de France (voir d'Europe mais il faudra que l'on prouve ça l'année prochaine). Merci pour l'ambiance, aussi bien sportive qu'amicale, et les médailles.

Merci à ma famille de faire le déplacement depuis la Normandie pour ma soutenance et de toujours avoir poussé ma curiosité.

Enfin merci à la personne la plus importante dans ma vie, Alison, qui a supporté la distance pendant toute cette thèse. Merci d'être là pour moi et dans ma vie.

Résumé

Les activités anthropiques sont en train de créer de nouveaux contacts entre des lignées génétiques différenciées, nous donnant accès à la phase initiale et à des réplicats de contacts secondaires. Le complexe *M. edulis* est composé d'un ensemble d'espèces et de lignées partiellement isolées reproductivement, présentant une grande diversité de contacts secondaires et de flux de gènes, constituant un modèle de choix pour l'étude de l'hybridation et de la spéciation. Nous avons profité de plusieurs événements de contact secondaires, à la fois naturels, anthropiques, et expérimentaux, pour étudier leurs effets sur les patrons génomiques d'admixture. En effet, plusieurs événements d'introduction de l'espèce *M. galloprovincialis* dans plusieurs ports français mais aussi en Norvège, provenant de lignées différentes et s'étant admixées avec les espèces locales, ont été identifiés d'abord à l'aide d'une centaine de marqueurs génétiques informatifs. Les populations portuaires forment des zones de contact à très fine échelle entre les populations introduites à l'intérieur des ports et les populations locales à l'extérieur. Ces populations portuaires présentent une introgression importante de gènes des lignées locales avec lesquelles elles sont en contact. La comparaison de ces événements d'admixture anthropiques avec des cas impliquant d'autres lignées mais aussi avec des croisements expérimentaux ont permis d'identifier des parallélismes dans la distorsion des fréquences alléliques lorsque les mêmes lignées sont impliquées. Afin d'explorer les patrons génomiques produits par ces événements, nous analysons 156 génomes de populations de référence et d'individus admixés. Cette analyse permet de montrer que les corrélations de distorsions entre événements d'admixture sont conservés à l'échelle génomique. Enfin, un modèle théorique permettant de prédire la valeur sélective des hybrides a été développé en utilisant le cadre du modèle géométrique de Fisher. Ce modèle considère les interactions épistatiques à l'échelle du génome, émergeant de la co-adaptation des mutations produites lors de la divergence. Il permet entre autres de s'affranchir des limitations des modèles classiques d'incompatibilités génétiques tout en conservant des prédictions en accord avec la théorie et les données de l'isolement post-zygotique. Ce modèle, simple et flexible, permet de prédire de nombreuses observations empiriques de l'étude de l'hybridation, dont les distorsions de ségrégation chez les moules. Il permet enfin d'expliquer que les populations admixées retrouvent une valeur sélective élevée bien qu'aillant une combinaison d'allèles provenant des deux génomes parentaux.

Abstract

Anthropogenic activities are creating new contacts between genetically differentiated lineages, providing access to the initial phase and replicates of secondary contacts. The complex *M. edulis* contains several species and lineages partially reproductively isolated, exhibiting a large diversity of secondary contacts and levels of gene flow. It constitutes therefore a good model for the study of hybridisation and speciation. We took advantage of several secondary contact events, both natural, anthropogenic and experimental, to investigate their effects on genomic admixture patterns. Indeed, events of *M. galloprovincialis* introduction from two different lineages, implicating admixture between the introduced species and the native species, have been identified in several French ports and in Norway. The introduction has first been identified with a hundred ancestry informative markers. Mussel port populations are forming fine scale contact zones between the inside of the ports and local populations located outside. Those port populations exhibit an important rate of introgression from local lineages with which they are in contact. The comparison of those anthropogenic admixture events with cases implicating other lineages and experimental crosses allowed us to identify parallelisms in allele frequency distortions when similar lineages are hybridising. In the aim of investigating the genomic patterns produced by such events, we analysed 156 genomes of reference and admixed populations, showing the correlation of distortions between events is conserved at the genomic scale. Finally, we developed a theoretical model for the prediction of hybrid fitness using the framework of Fisher's geometric model. This model takes into account genome scale epistatic interactions, emerging from the co-adaptation of mutations produced during divergence between the two lineages. The model is simple and flexible and predicts many empirical observations of hybridization research including segregation distortions in mussels. Finally, it explains that admixed populations retrieve high fitness while being a combination of alleles from the two parental genomes.