



HAL
open science

How do species barriers decay? Concordance and local introgression in mosaic hybrid zones of mussels

Alexis Simon, Christelle Fraïsse, Tahani El Ayari, Cathy Liautard-haag, Petr Strelkov, John J Welch, Nicolas Bierne

► **To cite this version:**

Alexis Simon, Christelle Fraïsse, Tahani El Ayari, Cathy Liautard-haag, Petr Strelkov, et al.. How do species barriers decay? Concordance and local introgression in mosaic hybrid zones of mussels. *Journal of Evolutionary Biology*, 2021, 34, pp.208-223. <10.1111/jeb.13709>. <hal-03025054>

HAL Id: hal-03025054

<https://hal.science/hal-03025054v1>

Submitted on 26 Nov 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

How do species barriers decay? Concordance and local introgression 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

The *Mytilus* complex of marine mussel species forms a mosaic of hybrid zones, found across temperate regions of the globe. This allows us to study “replicated” instances of secondary contact between closely-related species. Previous work on this complex has shown that local introgression is both widespread and highly heterogeneous, and has identified SNPs that are outliers of differentiation between lineages. Here, we developed an ancestry-informative panel of such SNPs. We then compared their frequencies in newly-sampled populations, including samples from within the hybrid zones, and parental populations at different distances from the contact. Results show that close to the hybrid zones, some outlier loci are near to fixation for the heterospecific allele, suggesting enhanced local introgression, or the local sweep of a shared ancestral allele. Conversely, genomic cline analyses, treating local parental populations as the reference, reveal a globally high concordance among loci, albeit with a few signals of asymmetric introgression. Enhanced local introgression at specific loci is consistent with the early transfer of adaptive variants after contact, possibly including asymmetric bistable variants (Dobzhansky-Muller incompatibilities), or haplotypes loaded with fewer deleterious mutations. Having escaped one barrier, however, these variants can be trapped or delayed at the next barrier, confining the introgression locally. These results shed light on the decay of species barriers during phases of contact.

Keywords: Hybrid zones, local introgression, genomic clines, secondary contact, *Mytilus*

1 Introduction

Divergence between species is almost always accompanied by hybridisation during phases of contact (Abbott et al., 2013), and so the study of speciation is intertwined with the study of introgression (Aeschbacher, Selby, Willis, & Coop, 2017; Chaturvedi et al., 2020; Duranton et al., 2018; Gagnaire et al., 2018; Martin, Davey, Salazar, & Jiggins, 2019; Roesti, Moser, & Berner, 2013; Schumer et al., 2018). The outcome of any given contact will depend on intrinsic and extrinsic factors. When barriers are weak there can be massive introgression, leaving differentiation at only a few strongly selected loci. But even strong barriers can be rapidly crossed by adaptive alleles and flanking neutral variants (Barton, 1979a; Faure, David, Bonhomme, & Bierne, 2008).

If some variation is bi-stable (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 at the barrier, and thereby contributes to the barrier (Barton, 1979a; Barton & Bengtsson, 1986; Piálek & Barton, 1997). Stochastic processes, such as random drift or variable migration rates, can free the spreading wave of these variants one by one (Piálek & Barton, 1997). The overall result is a sequential decay of the species barrier. However, the process can be very slow, and so difficult to observe in action.

A promising system to study these processes is the *Mytilus* complex of marine mussels, comprising the three species *M. trossulus*, *M. edulis* and *M. galloprovincialis*. Current transoceanic isolation (e.g. East and West Atlantic in *M. edulis* and *M. trossulus*, Riginos and Henzler, 2008; Varvio, Koehn, and Väinölä, 1988) and ancient allopatry (e.g. Atlantic and Mediterranean lineages in *M. galloprovincialis*, El Ayari, Trigui El Menif, Hamer, Cahill, and Bierne, 2019; Quesada, Zapata, and Alvarez, 1995) subdivides these taxa into infra-specific lineages (Fraïsse, Belkhir, Welch, & Bierne, 2016; Simon et al., 2020). However, they are broadcast spawners with a dispersive larval phase, and so true geographic isolation is relatively rare. For this reason, the complex is divided by numerous post-glacial hybrid zones, maintained by both intrinsic and extrinsic mechanisms of reproductive isolation (Bierne, Borsa, et al., 2003; El Ayari et al., 2019; Fraïsse et al., 2016; Hilbish et al., 2012; Riginos & Cunningham, 2005; Strelkov, Katolikova, & Väinölä, 2017). In particular, in the northern hemisphere all three species form a geographic mosaic delimited by many hybrid zones (Fraïsse et al., 2016).

This biogeographical history has two important consequences. First, contacts between similar lineages occur multiple times, and so we can observe the same species barrier in multiple demographic and ecological contexts (Abbott et al., 2013; Harrison & Larson, 2016; Simon et al., 2020). Second, with its narrow hybrid zones, combined with very long-range admixture (Fraïsse et al., 2016), this complex provides us with a continuum of divergence between the interacting taxa. Precisely because of this heterogeneity, however, it is challenging to identify diagnostic loci for the *Mytilus* complex.

Recently, Fraïsse et al. (2016) captured 1269 contigs from reference populations of the

Mytilus complex. Using these data, they identified regions that were divergent between species, and outliers of differentiation between populations of the same species. These intraspecific outliers are significant in two ways. First, they can be used to identify individual lineages. Second, phylogenetic analyses of gene trees suggested that the majority of the outliers were due to local introgression between species, rather than local intra-specific selective sweeps. As such, they provide information about introgression in the complex.

Here, we extend the analyses of Fraïsse et al. (2016). First, we built a SNP panel of a few hundred loci chosen from their outlier contigs, and therefore enriched in highly differentiated markers. Having included intra-specific outlier loci, we then examined their allele frequencies in a wider range of populations, allowing us to examine local introgression on a broader scale. Finally, we also sampled within hybrid zones, and carried out genomic cline / concordance analyses (Gompert & Buerkle, 2011; Macholán et al., 2011). Our aim here was to test the effect of the choice of parental populations (local *versus* distant) in the detection of discordant behaviour.

After verifying that our SNP panel is effective in identifying the lineages of interest, our results confirm that local introgression is both widespread in these *Mytilus* lineages, and heterogeneous across the genome. At some loci identified as intra-species outliers in Fraïsse et al. (2016), heterospecific alleles are sometimes fixed, or nearly so, in one parental patch close to a hybrid zone, while remaining nearly absent from other population patches farther from the zone. Genomic clines suggest high concordance among loci in the centre of hybrid zones, but only when local parental populations are used as reference. However, a few loci do depart from the genomic average, and demonstrate moderate 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 hybrid zones where lineages coexist.

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. We genotyped 441 individuals and extracted 72 genotypes by sequencing (GBS) from the published sequence dataset of Fraïsse et al. (2016).

2.2 Assay design

We aimed to genotype previously identified differentiated and outlier loci across a large number of samples in a cost-effective manner. For this purpose we used an Illumina BeadXpress® assay

Table 1: Definition of reference populations, classified by species level (L1), ocean basin or continent (L2), and regional groupings (L3). The ID column gives the population number referenced in Figure 1 and Table S1. Abbreviations used in the manuscript and figures are combinations of text in parentheses (e.g., edu_eu_north).

L1	L2	L3	comment	ID
<i>M. trossulus</i> (tros)	America (am)	-	Saint Lawrence	08, 09
	Europe (eu)	North (north)	Russia and Scandinavia	18, 19
		Baltic (baltic)	Baltic sea	22
	Pacific (pac)	East (east)	From Californian coast	01, 02, 05
<i>M. edulis</i> (edu)	America (am)	-	Long Island region	10-17
	Europe (eu)	external (ext)	-	26-28
		local (int)	-	35-37
		North (north)	Russia and Scandinavia	18, 19
<i>M. galloprovincialis</i> (gallo)	Atlantic (atl)	external (ext)	-	40-41
		local (int)	-	31-32
	Mediterranean (med)	West (west)	-	51-54
		East (east)	-	55-57
		Black Sea (bs)	-	58

with Veracode™ technology (GoldenGate Genotyping Assay). This method is based on the hybridisation and PCR amplification of SNP- and locus-specific oligos differentially bound to two different dyes. The genotype of an individual is provided by the relative fluorescence. We designed an assay of 384 SNPs (being the multiplexing limit of the technology).

Loci were selected, prior to genotyping, based on their differentiation either between species or between populations of the same species, using the published results of Fraïsse 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 et al., 2016, <http://www.scbi.uma.es/mytilus/index.php>). It contains 1269 contigs sequenced for 72 individuals from eleven different locations. Contig length ranges from 97 to 25,242 bp with a mean size of 3,389 bp and a cumulative length of 4.3 Mbp.

From the called genotypes of Fraïsse et al. (2016), markers with a minimum allele frequency of 0.05 and a maximum missing data percentage of 50% computed on all individuals were retained. Coverage was estimated as a mean computed on three populations (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 target 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

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). “all” corresponds to the other populations.

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	17
gallo / edu	100	65
gallo_med / all	6	1
tros_am / tros_eu	34	12
tros / all	96	68
Total	384	212

probable design success (ADT score > 0.4). F_{ST} was computed between pairs of populations and species such as those shown in Table 2. A first filtration step kept the 5 highest F_{ST} markers per contig for each comparison. We ensured a $F_{ST} > 0.5$ for all comparisons, except within edu_am and between gallo_atl and gallo_med where we used a $F_{ST} > 0.4$. Then, the six SNPs with the highest ADT scores per contig were retained. Finally, one SNP per contig and per category with the highest F_{ST} was selected (Table 2). Only for the *M. trossulus* comparison, the 70 SNPs with the highest ADT scores were retained. This SNP selection provides a final dataset of 384 SNPs, balanced across comparisons (Table 2).

2.3 Genotyping and filtering

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 and because of our choice to retain only good quality SNPs. Forty 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* only (overall 12% missing data). The genotypes for the 77 GBS individuals were retrieved from the calling in Fraïsse 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.

Three samples were removed from the analyses prior to filtering as they have already proven to be cancerous individuals (Riquet, Simon, & Bierne, 2017) (Por_40 from population 34; Arsud_05, Arsud_07 from population 38). 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., 2017), two additional samples from the Mediterranean sea were removed (MTP_05 and Collo_10 from populations 48 and 52 respectively). Filtering yielded a total of 514 individuals genotyped at 212 markers.

Reference populations were defined based on previous knowledge of the *Mytilus* species complex (Fraïsse et al., 2016; Simon et al., 2020). They are used first as a classification reference for newly genotyped individuals, to compute allele frequencies in analyses, and to serve as parental populations. 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 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 geographically close to it. We used a preliminary *Admixture* analysis (v1.3.0; Alexander, Novembre, and Lange, 2009) at the species level (L1) using $K = 3$ and default settings (30 independent runs merged with CLUMPAK; v1.1; Kopelman, Mayzel, Jakobsson, Rosenberg, and Mayrose, 2015) to correct the reference groups by removing inter-specific migrants or hybrids, as sympatry and hybridisation is a common phenomenon in the *Mytilus* species complex. Individuals assigned to those populations, and not filtered out, constitute the “reference dataset”.

Hardy-Weinberg equilibrium was tested for the remaining markers within putative panmictic clusters outside of known hybrid zones (pegas 0.11, Paradis, 2010). 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. (2020) 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 *ade4* R package (v2.1.3; Jombart, 2008), with replacement of missing data by the mean allele frequencies. Following the PCA, a dimensional reduction method, UMAP (Diaz-Papkovich, Anderson-Trocme, & Gravel, 2019; 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., 2019). This method was performed using the python package `umap-learn` (v0.3.9; McInnes and Healy, 2018) and the R wrapper package `umap` (v0.2.5.0; Konopka, 2019).

The `Structure` software (v2.3.4; Falush, Stephens, and Pritchard, 2003) 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 12. The standard deviation for the α prior was set to 0.05 for better mixing of the chains. The `Clumpak` software (v1.1; Kopelman et al., 2015) was used to investigate and aggregate `Structure` outputs with an Markov clustering (MCL) threshold of 0.9. Only major clusters are presented in the results.

2.5 Hybrid zone 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 “distant”, according to observed levels of introgression or geography. A local parental population is defined as a population close to the focal hybrid zone and being potentially subject to recent gene flow from the hybrid zone. A distant parental population is defined as a population distant from the hybrid zone, which has been less influenced by recent gene flow from it but can still have a substantial rate of established introgression due to past contacts. In a few cases a local parental population was not available in our sampling design and in these cases, a distant population was used instead (Table 3). Excluding the special cases of the Øresund and Scotland hybrid zones, “local” references correspond to populations less than 250 km away from the central populations (in great circle distance). Similarly, “distant” references are composed of populations more than 450 km away from the central populations.

For each hybrid zone, three analyses were carried out: (i) estimation of hybrid indexes with the R package `introgress` (v1.2.3; Gompert and Buerkle, 2010), (ii) a local Bayesian clustering analysis using `Structure`, and (iii) a Bayesian 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, in the Øresund hybrid zone we compare *M. edulis* (P1) and *M. trossulus*

Table 3: Grouping of populations for each hybrid zone considered in this study. We consider parental populations (P1/P2) from either the distant or the local zone. See Figure 1-2 and Table S1 for population information. In the populations listed for parental groups, only the individuals representative of the focal group are retained (e.g. only *M. edulis* individuals in populations 18 and 19 in the Øresund analysis).

Hybrid zone	Lineages	P1 distant	P1 local	Central	P2 local	P2 distant
Øresund	edu_eu/tros_eu_baltic	18, 19	–	20, 21	22	08, 09
Scotland	edu_eu/gallo_atl	26-27	–	24	31, 32	40, 41
Brittany	edu_eu/gallo_atl	26-27	35-37	33, 34	31, 32	40, 41
Aquitaine	edu_eu/gallo_atl	26-27	35-37	38	39	40, 41
Algeria	gallo_atl/gallo_med	40, 41	–	42-50	51, 52	55-57
North European <i>M. edulis</i>	edu_eu/edu_am	26-27	–	18, 19	–	10-14, 17

(P2). Therefore, *M. trossulus* individuals living in sympatry in the P1 populations were removed from this reference. 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. The tros_am population was used as a distant reference for the Øresund hybrid zone, as the tros_eu_north population had a lower number of individuals and a higher level of missing data. In the North European *M. edulis* hybrid zone (Table 3), populations 15 and 16 of the edu_am reference population exhibited an increased introgression from European *M. edulis* and were removed (see section 3.1).

Genomic clines are used to detect markers deviating from the average genomic expectation given the distribution of hybrid indexes. The program `bgc` (v1.03) 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 `rhd5` (v2.32.0; Fischer, Pau, and 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 (2020).

For each hybrid zone, we performed two analyses called “local” and “distant” in which the local or distant parental populations are taken as references, respectively (Table 3). Both analyses consider the “central” population as admixed.

We estimated locus-specific posterior distributions for the cline centre (α) and rate of transition in ancestry along the admixture gradient (β) (Gompert & Buerkle, 2011). Non-zero values of α shift the cline from the centre (negative values to parent 1 and positive ones to parent 2). β controls the steepness of the transition, with positive values inducing a steep cline

and negative ones inducing a flattened cline. Loci were classified as having “excess ancestry” for α or β if the 95% quantiles of the posterior distributions did not include zero (as in Gompert and Buerkle, 2012).

3 Results

We use a set of 212 differentiated and outlier 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 ancestry-informative panel of SNPs. 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 (Fraisse et al., 2016), and so we first tested if the markers were still informative in our broader 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. each side of the Atlantic ocean or the 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 et al. (2016), showing that our markers could differentiate the lineages of interest (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. While our dataset contained many differentiated loci as expected, only two markers showed fixed differences between species. Indeed, when considering all reference individuals of one species, only marker 147 was fixed between *M. edulis* and *M. galloprovincialis* (Contig H_L1_abysse_Contig244 position 6092) and only marker 015 was fixed between *M. edulis* and *M. trossulus* (Contig Contig17324_GA36A position 1089).

Using our SNP panel, five genetic clusters could be defined without ambiguity in a Principal Component Analysis (Figure S2; and noting that the order of the PCs is not meaningful, due to our method of selecting SNPs). As shown in Figure 2 using a dimension reduction algorithm (UMAP), 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*. (Note that the placement of the clusters is not meaningful due to the UMAP algorithm). We observed that the Baltic population of *M. trossulus* (22) is introgressed by *M. edulis*, as previously described

(Fraïsse et al., 2016), and this cluster can be identified with Structure at $K = 8$ (Figure S3).

The remaining individuals in Figure 2 form clouds labelled A to E and represent the various admixed populations we sampled (see also Figure 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 corresponds to group D (Figure 2).

Group E contains a small set of individuals sampled in the leisure marina of Cherbourg in 2003, on the French coast of the English Channel (population 29). These individuals exhibit 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 expected local genetic lineage.

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 lineage of *M. edulis* (edu_am; Figures 2, S4, S5). Additionally, they do not differentiate on an independent axis on a classic PCA (Figure S2). The group defined as North-European *M. edulis* does not exhibit visible introgression from the sympatric *M. trossulus*. While two F1 individuals were found between the North-European *M. edulis* and *M. trossulus* in population 19, the levels of introgression between these sympatric populations were not higher than those in the American populations. 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 B corresponds to admixed individuals of the Øresund hybrid zone (populations 20-21) between the newly identified North-European *M. edulis* and the Baltic *M. trossulus* (Figure 2).

The 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 infra-specific ancestry coming from Europe (already noticed by Simon et al., 2020). An analysis considering only *M. edulis* samples, Figure S4, shows more clearly that individuals from the most southern populations of the Long Island Sound (15, 16) sometimes have higher European ancestry than other American populations (e.g. Boston, population 17).

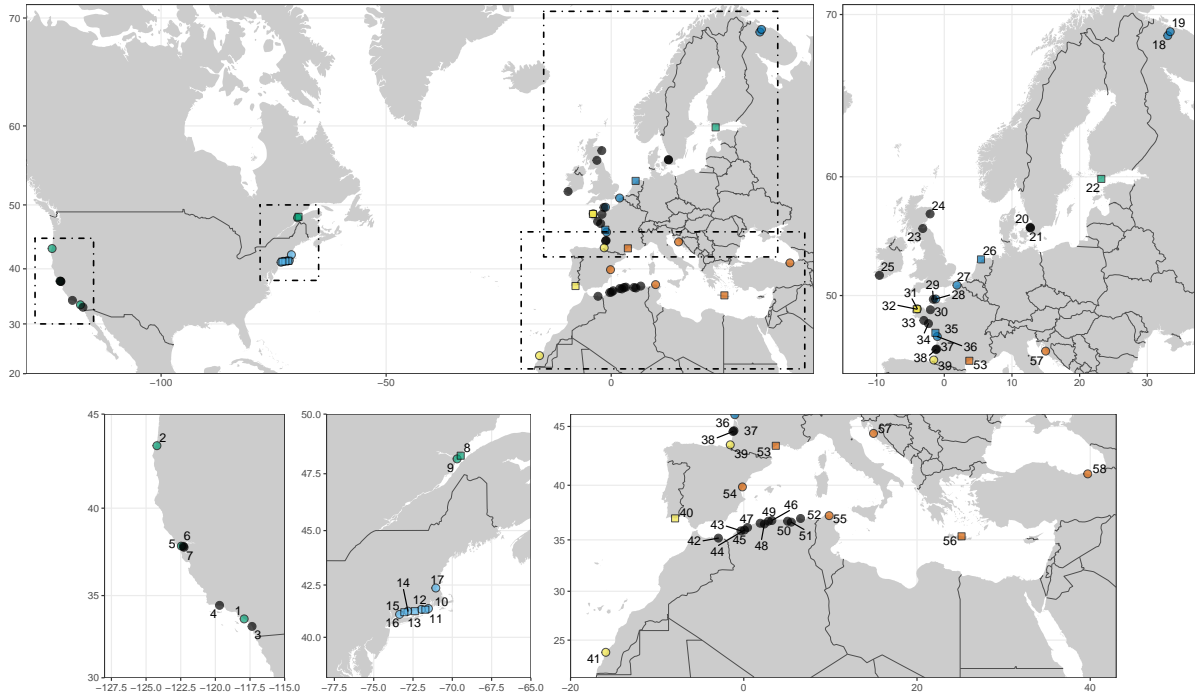


Figure 1: Location of the studied populations. Squares: genotype by sequencing (GBS, data from Fraïsse, Belkhir, Welch, and Bierne, 2016). Circles: BeadXpress genotypes. The same colour code is used in Figure 2. More detailed population information is available in Table S1.

3.2 Genetic differences between patches of parental populations close to or distant from hybrid zones

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). This representation illustrates that outlier loci (identified by Fraïsse et al., 2016) between populations within parental species most often correspond to a high frequency of the heterospecific allelic state (defined on the basis of allele frequency differences between species) in the population close to hybrid zones: blue smudges (*edulis* allelic state) in *M. galloprovincialis*, red smudges (*galloprovincialis* allelic state) in *M. edulis* (Figure 3), blue smudges (*edulis* allelic state) in *M. trossulus* and green smudges (*trossulus* allelic state) in *M. edulis* (Figure 4). The discordance between populations close to and distant from hybrid zones is indicative of differential introgression (similar reasoning to the now well established Patterson's D or Reich's f_3), but at a given SNP the allelic state is not informative of the ancestry, and so this observation can also be explained by an increased frequency of a shared ancestral allele. However, Fraïsse et al. (2016) used gene phylogenies to identify that roughly half of these outliers were the consequence of enhanced local introgression while the other half was the consequence of a local selective sweep independent of introgression. Because the present study only uses strongly differentiated and outlier SNPs, we lack the neutral baseline required to re-conduct outlier tests done by Fraïsse et al. (2016). Nevertheless, we confirm a strong

signal that does not seem consistent with a fully neutral interpretation knowing the very low level of genetic differentiation between populations of the same species at the genome scale (Fraïsse et al., 2016).

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. *M. galloprovincialis* populations are known to be more introgressed by *M. edulis* alleles in the Atlantic than the Mediterranean and our results confirm and illustrate this trend (see mean hybrid index, h , 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 (34-35 and 26-28, Figure 1), 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 alleles with an *M. edulis* allelic state, contrasting with the rest of the Mediterranean basin (Figure 3A, B6). The South-European *M. edulis*, both external and internal to the mosaic, have nearly as many smudges of *M. galloprovincialis* allelic state (Figure 3A, B4); and around half of these extend to Northern Europe.

For the *M. edulis*-*M. trossulus* comparison, we investigated two intergradations, in America and Europe (Figure 4). American and North European *M. edulis* populations show introgression from *M. trossulus*, with shared patterns of heterospecific alleles. Those two populations share potential local introgressions (Figure 4A), and almost none are private to either population (though see marker 183 in Figure 4). Regarding introgression from *M. edulis* to *M. trossulus*, whereas the Baltic *M. trossulus* has been highly introgressed (exhibiting a mean hybrid index of $h = 0.15$), American *M. trossulus* shows lower levels of global introgression (mean hybrid index of $h = 0.05$).

3.3 Concordance in 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 centre of the hybrid zone where two species or lineages coexist; (ii) local parental populations (P1 and P2 local) on each side of the hybrid zone and impacted by recent hybridisation; and (iii) distant parental populations (P1 and P2 distant) which are not in direct contact with the focal hybrid zone and so potentially less affected by local introgression.

A genomic cline analysis is designed to identify loci that shift from the genome wide expectation on the basis of the multi-locus hybrid index. Departure from the expectation is usually interpreted as evidence of selection (adaptive introgression, incompatibility or balancing selection). We posit that such analysis can strongly depend on the reference populations used in the analysis. Our outlier SNPs, for instance, should not appear as outlier in the genomic clines analysis if the local parental populations are used as references, but should if distant populations are used instead. We therefore contrast analyses done with different references.

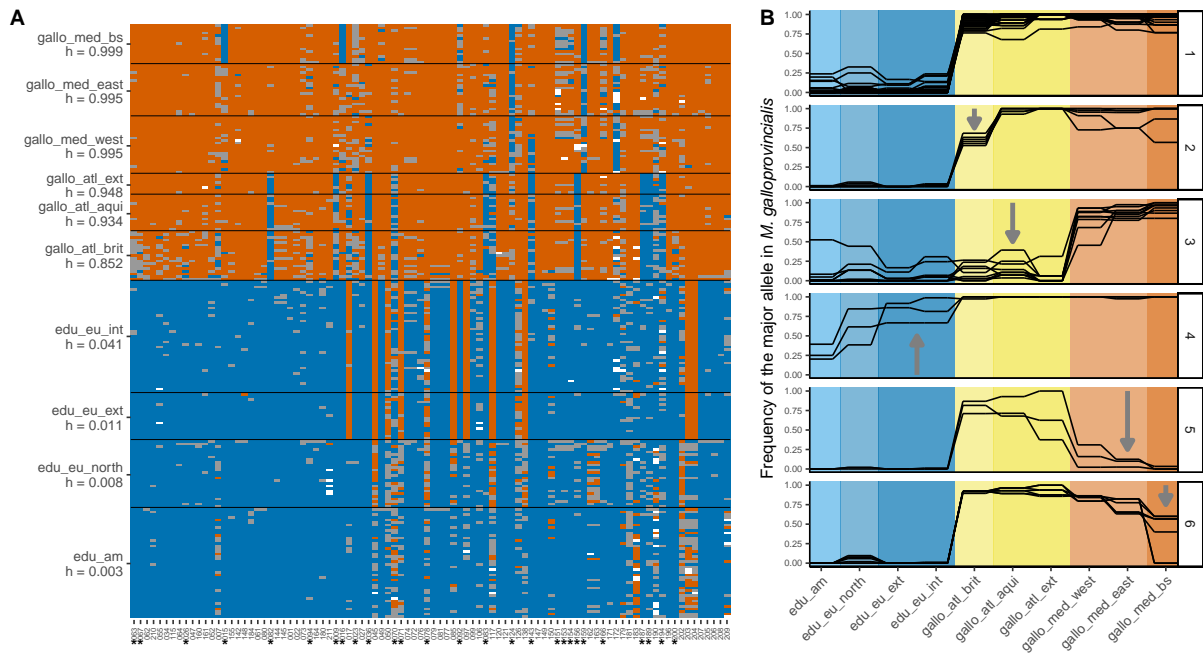


Figure 3: Visualisation of introgression between *M. galloprovincialis* and *M. edulis* using the reference dataset. **(A)** Plot of the raw genotypes. Orange: homozygous *M. galloprovincialis*; blue: homozygous *M. edulis*; grey: heterozygous. The orientation of alleles and mean hybrid indexes for each population (h) 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). Loci identified by an asterisk are plotted in panels B2-6. **(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.

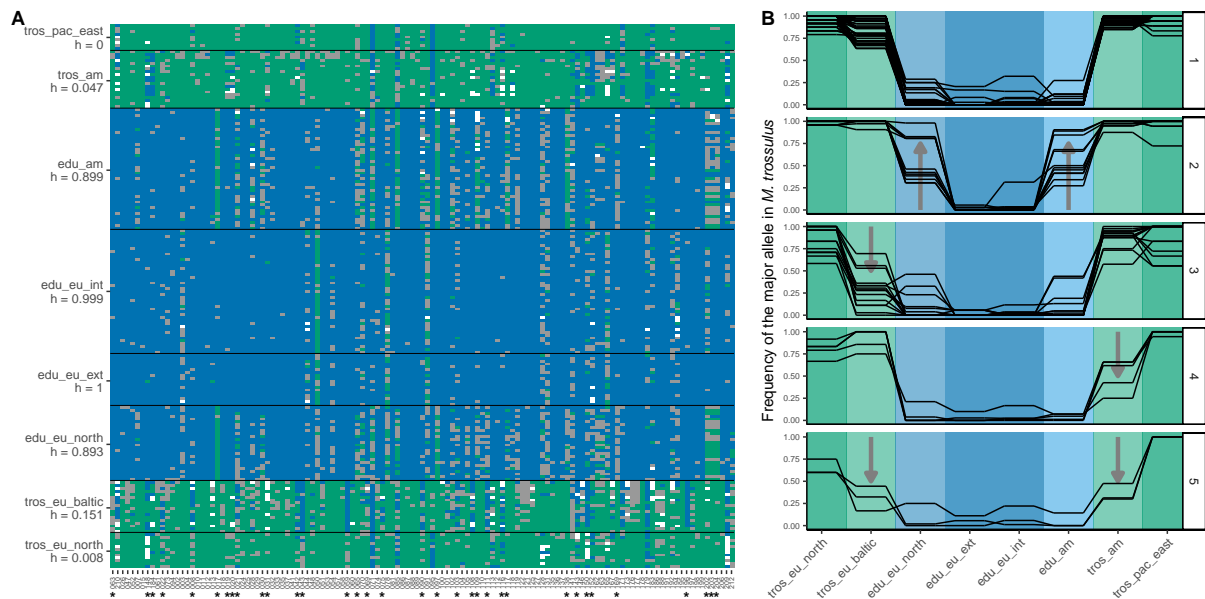


Figure 4: Visualisation of introgression between *M. trossulus* and *M. edulis* using the reference dataset. **(A)** Plot of the raw genotypes. Green: homozygous *M. trossulus*; blue: homozygous *M. edulis*; grey: heterozygous. The orientation of alleles and mean hybrid indexes for each population (h) were computed with the R package *introgress*. The *trospac_east* and *troseu_north* populations were taken as parent 1 (P1), and the *edueu_ext* and *edueu_int* ones were taken as parent 2 (P2). Loci identified by an asterisk are plotted in panels B2-5. **(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*.

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., 2020) with a limited number of excess ancestry markers both for α and β . On the other hand, analyses with distant parental populations taken as reference (Figure 5, right panels) exhibited an increase in outlier loci. We note that certain new outliers may be false positives due to a reduced convergence between chains in the latter analysis. For instance, the Aquitaine genomic cline analysis using distant parental populations appears to be a failure of the algorithm, where not enough intermediate genotypes are present to correctly estimate the parameters.

Three hybrid zones, in Brittany, Aquitaine and Scotland, involve contact between South-European *M. edulis* and Atlantic *M. galloprovincialis* (Figure 5, S6-S8). The Brittany and Aquitaine zones are on 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, 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 strong introgression compared with the distant population (P2_distant) (Figure S7). In Scotland, the central population exhibits variable hybrid indexes, and so resembles Brittany more than Aquitaine (Figures 5 and S8). The genetic composition of local *M. edulis* and *M. galloprovincialis* populations in Scotland are largely unknown. However the only *M. edulis* individual sampled in population 24 is indiscernible from the edu_eu_ext references (Figures 2 and 5).

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 of coexistence of the two lineages extends to around 600 km (Figures 5 and S9). We did not sample Atlantic *M. galloprovincialis* populations close to this zone (P1_local). If the local population and hybrid zone were harbouring local introgressions, we would have expected discordant loci when using the P1_distant population as parental. As only one locus was shifted and not significantly, we can hypothesise that the local population of Atlantic *M. galloprovincialis* is not heterogeneously introgressed by the Mediterranean *M. galloprovincialis*.

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 and S10). On the other side of the hybrid zone, the Baltic *M. trossulus* population is introgressed as shown above (Fraisie et al., 2016; Väinölä and Strelkov, 2011; Figures 4 and 5).

To highlight the intermediate character of the North-European *M. edulis* lineage, we carried

out an analysis treating the North-European *M. edulis* individuals as admixed in the Russian populations (18 and 19, Fig. S5). As shown by the hybrid index and the Structure analyses, these mussels have a homogeneous admixture of around 60% South-European *M. edulis* and 40% American *M. edulis*.

When finally 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 (Permutation tests of the Spearman correlation coefficients $\rho = 0.38$ and $\rho = 0.36$, respectively, with p -values < 0.005 using 10^4 permutations), and β between the Brittany and Aquitaine hybrid zones ($\rho = 0.38$, p -value = 0.002).

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 highly differentiated, and potentially diagnostic markers for the *Mytilus* complex in the Northern Hemisphere. To this end, we started with 51,878 high-quality SNPs from 1269 sequences (Fraïsse et al., 2016), and selected SNPs with the greatest discriminatory power. We combined differentiated variants between species with intra-specific F_{ST} outliers to provide a more detailed identification. 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*; Figure 2).

Nevertheless, of 92 highly differentiated loci between *M. edulis* and *M. galloprovincialis* (Figure 3), only one was a fixed difference in our sample. Similarly, there was only one fixed difference among 125 highly differentiated 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. Our approach highlights that loci originally claimed to be diagnostic are likely to lose their diagnostic status in a large-scale assay.

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, Durmišić, 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 diagnosis with multilocus inference; and our results suggest

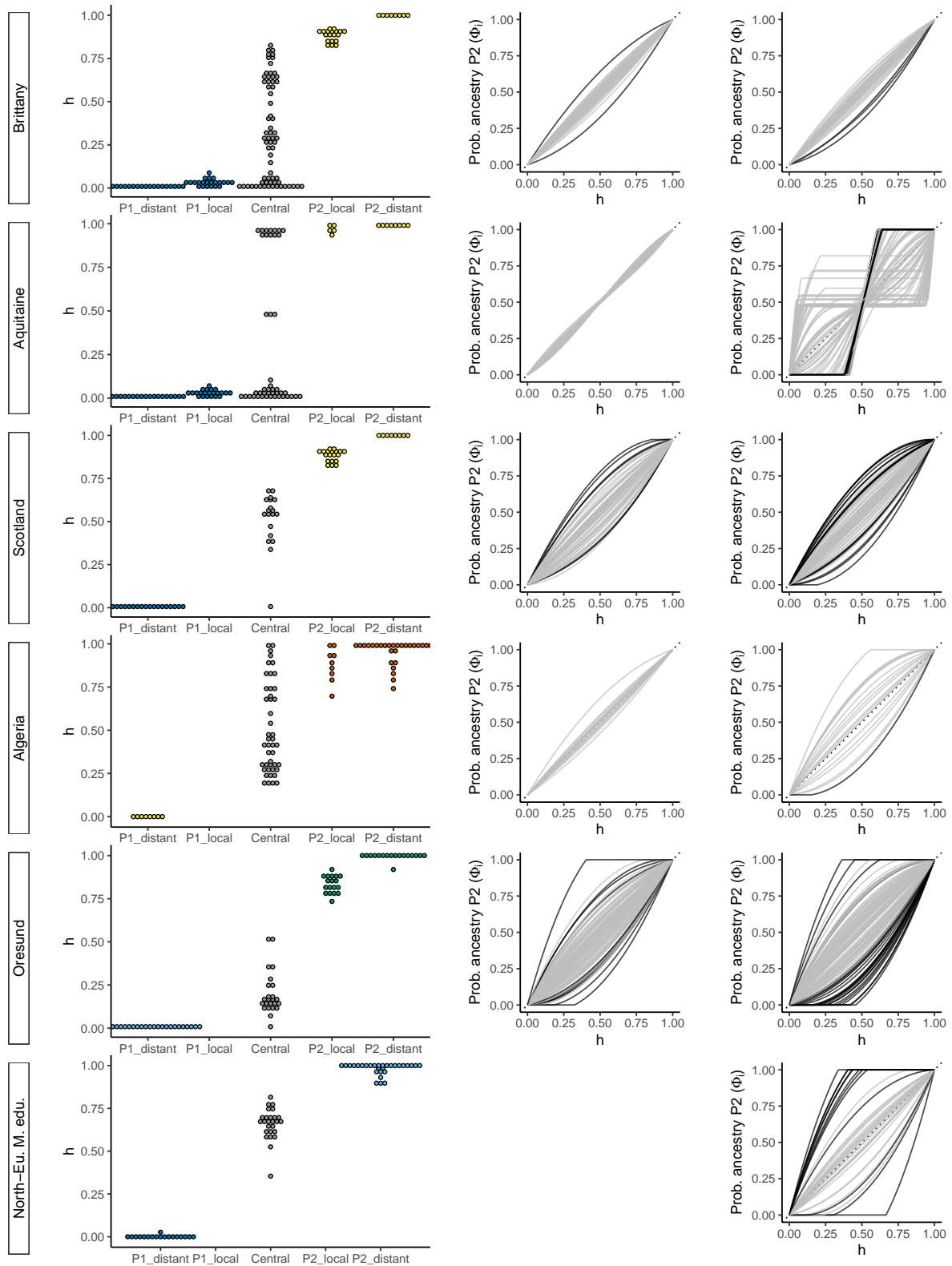


Figure 5: **Left panels:** Hybrid index (h) distributions from introgress, for each group of the hybrid zone. **middle panels:** Genomic clines computed with the local parental groups (P1_local, P2_local or distant when not available) with clines presenting a significant excess in either α or β parameters highlighted in black. **right panels:** Genomic clines computed with distant populations as parental groups. For genomic clines, only markers with an allele frequency difference > 0.3 between P1_distant and P2_distant are drawn. The probability of the locus-specific ancestry for P2 against the genome wide hybrid index (h) is plotted. When distant parental populations are used in place of local ones, the number of discordant loci increases. See Table 3 for a list of the populations used in the analyses.

that 5-10 well chosen loci should be sufficient for this purpose, but not less.

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, Bierne, & Viard, 2016). As for the *Mytilus* system, *Ciona* species have very high nucleotide diversities and have been shown to go through a history of secondary contact (Roux, Tsagkogeorga, Bierne, & Galtier, 2013) resulting in a higher proportion of high frequency derived alleles either shared between populations or private to one population (i.e. top left and bottom right of the unfolded joint allele frequency spectrum) than expected for well isolated and low N_e species.

4.2 Enhanced local introgression is widespread and has several possible causes

A notable feature of our data was the appearance of heterospecific alleles that were present at high frequency in one or a small number of populations. These alleles were very common (at least one such allele was found in each population of each lineage of all three species), but were found heterogeneously across the genome (Figures 3 and 4).

Fraïsse et al. (2016) also found this pattern, in their much smaller sample. Furthermore, their use of genome-wide data, combined with analyses of gene-genealogies, suggested that the pattern was due to local introgression, rather than, say, incomplete lineage sorting (e.g. the local sweep of a shared ancestral variant). While our data do not allow us to distinguish between these possibilities, and while both have no doubt taken place, our use of the data of Fraïsse et al. (2016) to choose our SNPs, combined with the finding that these alleles are preferentially found in contiguous populations (Figures 3 and 4), strongly suggest high levels of local introgression in the *Mytilus* complex (Fraïsse, Haguénauer, et al., 2018; Fraïsse, Roux, Gagnaire, et al., 2018).

The biogeography of the complex also allowed us to study introgression of a very particular kind: 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 enhanced local 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 the heterogeneity of local introgression. First, introgression could be adaptive (Hedrick, 2013; Pardo-Díaz et al., 2012; Racimo, Marnetto, & Huerta-Sánchez, 2017; Staubach et al., 2012). In this case, the fact that introgression was 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). However, Fraïsse et al. (2016) found no clear links between the biological function of outlier genes and the environmental conditions of introgressed populations. Alternatively, if the markers are not the direct target of selection – as seems highly plausible (Fraïsse et al., 2016) – a marker might hitchhike through one barrier, but be halted at a second (Barton, 2000; Faure et al., 2008).

A second possibility is that introgression acts to reduce genetic load in the recipient population (Harris & Nielsen, 2016; Juric, Aeschbacher, & Coop, 2016; 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, introgression could involve bi-stable variants in a tension zone (Barton & Hewitt, 1985). Such variants can individually move via an asymmetry in parental fitness, i.e. when one parental genotype is fitter than the other and both are fitter than hybrids (Barton, 1979a; Barton & Turelli, 2011). However, they can easily be trapped for long periods by a density trough, a barrier to dispersion, and coupling (Barton and Turelli, 2011; Bierne, Welch, Loire, Bonhomme, and David, 2011, see El Ayari et al., 2019 for a possible example in *Mytilus*). 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). In other words, rare stochastic events can free the “pushed” wave of advance (Piálek & Barton, 1997). This hypothesis was already proposed in *Mytilus* by Fraïsse, Roux, Welch, and Bierne (2014), and has also been proposed for other systems such as *Bathymodiolus* mussels (Plouviez et al., 2013), sea bass (Duranton et al., 2020), killer whales (Foote et al., 2019) and house mouse (Macholán et al., 2008).

4.3 High overall concordance of barrier loci at hybrid zones

With high levels of local introgression, it can be difficult to study genomic clines. With our data this was evident in the occasional failures of our cline analyses, for instance in the Aquitaine hybrid zone. Nevertheless, in most cases these analyses yielded clear results: a high level of concordance with respect to local parental populations (Figure 5 middle panels), and increased discordance with respect to distant parental populations (again, consistent with local introgression).

Even with local parental populations, we observed a few SNPs that deviated from the

genomic average. The anomalous “excess ancestry” loci observed in Scotland and Øresund in the local analysis could still be a consequence of enhanced local introgression, as P1_local populations were missing from our sampling. Nonetheless deviating loci are still observed, e.g. in Brittany (Figure 5). These loci tend to introgress more than the others in the centre of the zone, but – unlike established introgression – have not escaped the genetic barrier. We suspect that these loci might contribute to bi-stable variation that is asymmetric, as described above, but with a lower fitness difference between parental genotypes.

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 centre of the hybrid zones, but not in flanking parental populations. If a force is opposing the spread of the adaptive variant from a barrier, it could be selection against heterozygous or recombinant hybrid genotypes (Barton, 1979b).

Finally, the observation of a small but sometimes significant correlation of the genomic cline parameters can point us to parallel processes that may be at play in the hybridisation of *Mytilus* species. Here, correlations were only found in hybrid zones implicating the same species and lineages. However, Simon et al. (2020) already showed that parallel deviations in admixed populations of mussels exist, and that it can also sometimes happen between different lineages. Such cases call for further investigations of potential selective effects in the regions of the studied markers, either due to positive selection or barrier loci.

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, local introgression of *M. edulis* alleles into *M. galloprovincialis* from the Black Sea and Mediterranean is consistent with ancient contacts between these populations (Fraïsse et al., 2016; Gosset & Bierne, 2013). The two private introgressions into the Black sea were nevertheless unexpected. Fraïsse et al. (2016) did not sample these populations, and so our data set was not enriched for discriminatory SNPs. Therefore 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 3A, B2) are relatively recent. This is probably because, given its position enclosed in the mosaic, this population only became established after 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., 2020), 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 the *Mytilus* complex. 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 S5) and hybrid index analyses (Figure 5). This observation previously made in Simon et al. (2020), has recently been confirmed by Wenne et al. (2020). 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 the North Sea and populations North of the Kattegat region; Bierne, Daguin, Bonhomme, David, and Borsa, 2003; Stuckas, Stoof, Quesada, and Tiedemann, 2009). Given the shared introgressions, therefore, the contemporary biogeography could reflect a recolonisation of America after the last glacial maximum, 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 refugium (Riginos & Henzler, 2008), might have colonised the North-Atlantic and Scandinavia, and then been introgressed by the South-European *M. edulis* in Europe. Finally, while North-European *M. edulis* and *M. trossulus* live in sympatry in several locations around Scandinavia and northern Russia (this study; Katoikova, Khaitov, Väinölä, Gantsevich, and Strelkov, 2016; Simon et al., 2020), our analyses show that introgression is not higher than in American populations.

4.5 Implications for speciation research

The *Mytilus* species complex is characterised by diverse levels of genetic differentiation between lineages and species, and by pervasive introgression due to a dense history of secondary contacts and mosaic geographical distribution. As such, it constitutes a powerful model to investigate the maintenance of species barriers in the marine environment. Based on our results and previous studies (Bierne, Borsa, et al., 2003; Fraïsse et al., 2016), we suggest that secondary contacts are not only periods of neutral introgression; but also periods during which genetic incompatibilities can be swamped resulting in a punctuated decay of the barrier to gene flow. This vision is rooted in the observation that gene flow after secondary contacts in the *Mytilus* complex is shown to be heterogeneous along the genome (Fraïsse, Roux, Gagnaire, et al., 2018; Roux et al., 2014; Roux et al., 2016). In regions of higher gene flow, some alleles have locally introgressed. Those can either be missed when distant parental populations are not taken

into account, or observed and interpreted as local adaptive introgression. Another explanation exists, considering that bi-stable variants, which could be associated with or responsible for genetic incompatibilities, can punctually escape a zone of coupling. 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.

Data availability: All data and scripts for analyses are available in the Dryad archive <https://doi.org/10.5061/dryad.r4xgxd29n>.

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 2020-240 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. *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
- 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. (2000). Genetic hitchhiking. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *355*(1403), 1553–1562. doi:10.1098/rstb.2000.0716
- 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., & 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
- Benjamini, Y., & Yekutieli, D. (2001). The Control of the False Discovery Rate in Multiple Testing Under Dependency. *The Annals of Statistics*, 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. *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*. *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., Daguin, C., Bonhomme, F., David, P., & Borsa, P. (2003). Direct selection on allozymes is not required to explain heterogeneity among marker loci across a *Mytilus* hybrid zone. *Molecular Ecology*, 12(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
- 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
- Borsa, P., Daguin, C., Ramos Caetano, S., & Bonhomme, F. (1999). Nuclear-DNA evidence that northeastern Atlantic *Mytilus trossulus* mussels carry *M. edulis* genes. *Journal of Molluscan Studies*, 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. *Evolutionary Ecology*, 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. *Molecular Ecology*, 25(21), 5527–5542. doi:10.1111/mec.13854

- Chaturvedi, S., Lucas, L. K., Buerkle, C. A., Fordyce, J. A., Forister, M. L., Nice, C. C., & Gompert, Z. (2020). Recent hybrids recapitulate ancient hybrid outcomes. *Nature Communications*, *11*(1), 2179. doi:10.1038/s41467-020-15641-x
- Diaz-Papkovich, A., Anderson-Trocme, L., & Gravel, S. (2019). UMAP reveals cryptic population structure and phenotype heterogeneity in large genomic cohorts. *PLoS Genetics*, *15*(11), e1008432. doi:10.1371/journal.pgen.1008432
- Durantou, 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
- Durantou, M., Allal, F., Valière, S., Bouchez, O., Bonhomme, F., & Gagnaire, P.-A. (2020). The contribution of ancient admixture to reproductive isolation between European sea bass lineages. *Evolution Letters*. doi:10.1002/evl3.169
- 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
- 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.
- 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).
- Foote, A. D., Martin, M. D., Louis, M., Pacheco, G., Robertson, K. M., Sinding, M.-H. S., ... Morin, P. A. (2019). Killer whale genomes reveal a complex history of recurrent admixture and vicariance. *Molecular Ecology*, *28*, 3427–3444. doi:10.1111/mec.15099
- 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
- Fraïsse, C., Roux, C., Welch, J. J., & Bierne, N. (2014). Gene-flow in a mosaic hybrid zone: Is local introgression adaptive? *Genetics*, *197*, 939–951. doi:10.1534/genetics.114.161380

- 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. (2009). A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology*, *18*(6), 1207–1224. doi:10.1111/j.1365-294X.2009.04098.x
- 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
- Gompert, Z., & Buerkle, C. A. (2011). Bayesian estimation of genomic clines. *Molecular Ecology*, *20*(10), 2111–2127. doi:10.1111/j.1365-294X.2011.05074.x
- Gompert, Z., & Buerkle, C. A. (2012). Bgc : Software for Bayesian estimation of genomic clines. *Molecular Ecology Resources*, *12*(6), 1168–1176. doi:10.1111/1755-0998.12009.x
- 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
- 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. *Cahiers de Biologie Marine*, *53*(1), 35–44.
- Harris, K., & Nielsen, R. (2016). The Genetic Cost of Neanderthal Introgression. *Genetics*, *203*(2), 881–891. doi:10.1534/genetics.116.186890
- Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, *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. *Molecular Ecology*, *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. *Journal of Biogeography*, *39*(4), 676–687. doi:10.1111/j.1365-2699.2011.02633.x
- Jombart, T. (2008). Adegnet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405. doi:10.1093/bioinformatics/btn129
- Juric, I., Aeschbacher, S., & Coop, G. (2016). The Strength of Selection against Neanderthal Introgression. *PLoS Genetics*, *12*(11), e1006340. doi:10.1371/journal.pgen.1006340
- 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
- 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 Genetics*, *14*(10), e1007741. doi:10.1371/journal.pgen.1007741
- Konopka, T. (2019). *Umap: Uniform Manifold Approximation and Projection*. R package version 0.2.2.0.
- 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
- 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
- Macholán, M., Baird, S. J. E., Dufková, P., Munclinger, P., Bímová, B. V., & Piálek, J. (2011). Assessing Multilocus Introgression Patterns: A Case Study on the Mouse X Chromosome in Central Europe. *Evolution*, *65*(5), 1428–1446. doi:10.1111/j.1558-5646.2011.01228.x
- Macholán, M., Baird, S. J., Munclinger, P., Dufková, P., Bímová, B., & Piálek, J. (2008). Genetic conflict outweighs heterogametic incompatibility in the mouse hybrid zone? *BMC Evolutionary Biology*, *8*, 271. doi:10.1186/1471-2148-8-271
- 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
- McInnes, L., & Healy, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv*, 1802.03426v2.
- 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
- Paradis, E. (2010). Pegas: An R package for population genetics with an integrated–modular approach. *Bioinformatics*, *26*(3), 419–420. doi:10.1093/bioinformatics/btp696
- Pardo-Díaz, 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 Genetics*, *8*(6), e1002752. doi:10.1371/journal.pgen.1002752

- 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.
- Plouviez, S., Faure, B., Le Guen, D., Lallier, F. H., Bierne, N., & Jollivet, D. (2013). A New Barrier to Dispersal Trapped Old Genetic Clines That Escaped the Easter Microplate Tension Zone of the Pacific Vent Mussels. *PLoS ONE*, *8*(12), e81555. doi:10.1371/journal.pone.0081555
- 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.
- Racimo, F., Marnetto, D., & Huerta-Sánchez, E. (2017). Signatures of Archaic Adaptive Introgression in Present-Day Human Populations. *Molecular Biology and Evolution*, *34*(2), 296–317. doi:10.1093/molbev/msw216
- Riginos, C., & Cunningham, C. W. (2005). Local adaptation and species segregation in two mussel (*Mytilus edulis* × *Mytilus trossulus*) hybrid zones. *Molecular Ecology*, *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*. *Marine Biology*, *155*(4), 399–412. doi:10.1007/s00227-008-1038-4
- 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
- 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
- Roux, C., Tsagkogeorga, G., Bierne, N., & Galtier, N. (2013). Crossing the Species Barrier: Genomic Hotspots of Introgression between Two Highly Divergent *Ciona intestinalis* Species. *Molecular Biology and Evolution*, *30*(7), 1574–1587. doi:10.1093/molbev/mst066
- 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

- 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. (2020). Replicated anthropogenic hybridisations reveal parallel patterns of admixture in marine mussels. *Evolutionary Applications*, *13*(3), 575–599. 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. *Evolution Letters*, *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 Genetics*, *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. *Marine Biology*, *164*(11), 214. 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. *Marine Biology*, *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. *Marine Biology*, *98*(1), 51–60. doi:10.1007/BF00392658
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2020). Rank-normalization, folding, and localization: An improved \hat{R} for assessing convergence of MCMC. *Bayesian Analysis, Advance publication*. doi:10.1214/20-BA1221
- 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
- Wenne, R., Zbawicka, M., Bach, L., Strelkov, P., Gantsevich, M., Kukliński, P., ... Kotta, J. (2020). Trans-Atlantic Distribution and Introgression as Inferred from Single Nucleotide Polymorphism: Mussels *Mytilus* and Environmental Factors. *Genes*, *11*(5), 530. doi:10.3390/genes11050530

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. *Journal of Molluscan Studies*, 69(1), 61–66. doi:10.1093/mollus/69.1.61