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

Alexis Simon[1], Christelle Fraïsse[1,2], Tahani El Ayari[1], Cathy Liautard-Haag[1], Petr Strelkov[3,4], John J. Welch[5] and Nicolas Bierne[1]

[1]ISEM, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France.
[2]Institute of Science and Technology Austria, Am Campus 1, Klosterneuburg 3400, Austria.
[3]St. Petersburg State University, Universitetskaya Emb. 7/9, St. Petersburg 199034, Russia.
[4]Laboratory of Monitoring and Conservation of Natural Arctic Ecosystems, Murmansk Arctic State University, Kapitana Egorova Str. 16, Murmansk 183038, Russia.
[5]Department 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.

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

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, Katolikova, & Väinolä, 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 Väinolä, 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
M. trossulus (tros)	America (am)	-	Saint Lawrence
	Europe (eu)	Baltic (baltic)	Baltic sea
	Pacific (pac)	East (east)	from Californian coast
M. edulis (edu)	America (am)	-	Long Island region
	Europe (eu)	external (ext)	-
		local (int)	-
		North (north)	Russia and Scandinavia
M. galloprovincialis (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 *M. edulis*	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_abyss_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	$AFD > 0.5$	$AFD > 0.7$	$AFD = 1$
M. edulis/*M. galloprovincialis*	79	62	1
M. edulis/*M. trossulus*	108	78	1
M. trossulus/*M. galloprovincialis*	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 (Fraïsse, Belkhir, et al., 2016).

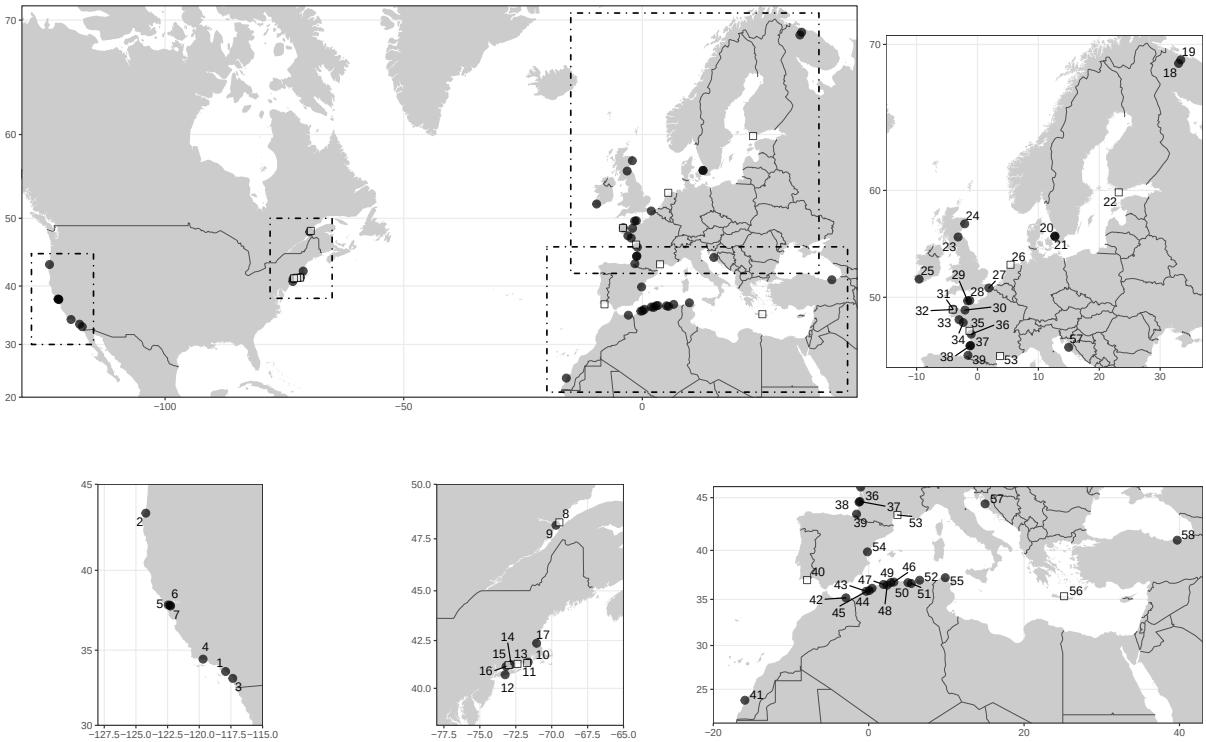

Figure 1: Maps of populations sampled. White squares: GBS. Black circles: BXP. 01-California-NEW, 02-Oregon-Coos, 03-California-Carlsbad, 04-SantaBarbara, 05-California-TBR, 06-California-BRK, 07-California-LAM, 08-StLawrence-CBD, 09-StLawrence-TAD, 10-Lgls-SHM, 11-Lgls-QNT, 12-Lgls-BYY, 13-Lgls-OST, 14-Lgls-BFM, 15-Lgls-MYC, 16-Lgls-SPM, 17-Boston, 18-NovyMost, 19-Retinskoye, 20-Oresund-Helsinborg, 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-ElMarsa, 47-Ghouraya, 48-Tipaza, 49-SidiFredj, 50-Bejaia, 51-Ziama, 52-Collo, 53-Thau, 54-Nules, 55-Bizerte, 56-Heraklion, 57-Croatia, 58-Trabzon

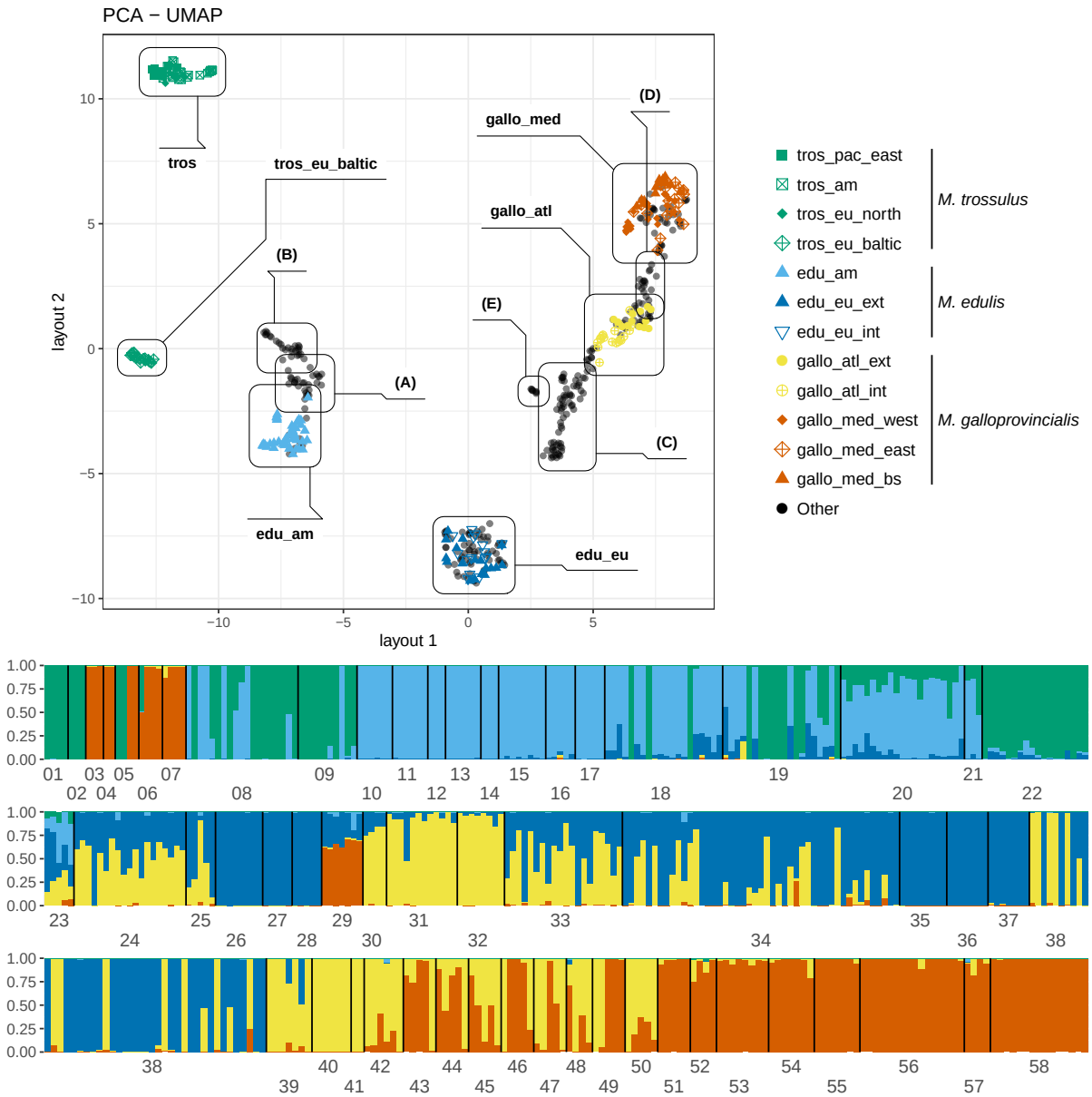

Figure 2: **top**: PCA-UMAP using the first 11 principal components. The reference level L3 (Table 1) is colour and shape coded. Note that this representation does not conserve distances and is designed to maximise groupings between similar entities, see Figure S2 for PCAs. Annotations show five groups 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. Nevetheless, 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

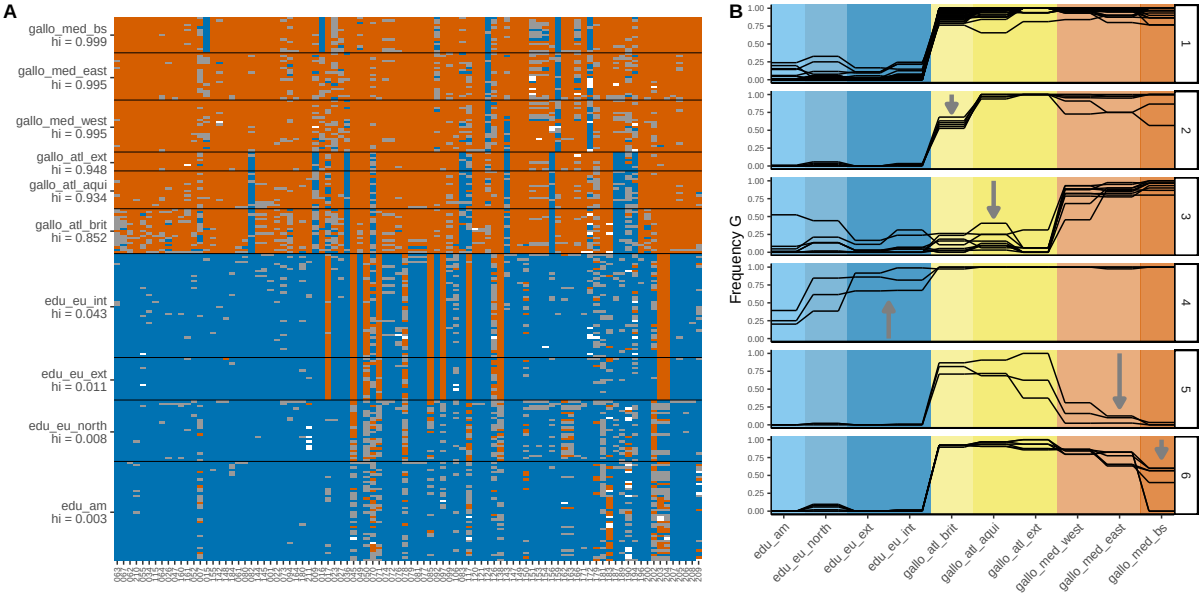

Figure 3: Visualisation of introgression between *M. galloprovincialis* and *M. edulis* the reference dataset. **(A)** Plot of the raw genotypes. Orange: homozygous *M. galloprovincialis*; blue: homozygous *M. edulis*; grey: heterozygous. The orientation of alleles and mean hybrid indexes for each population (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.

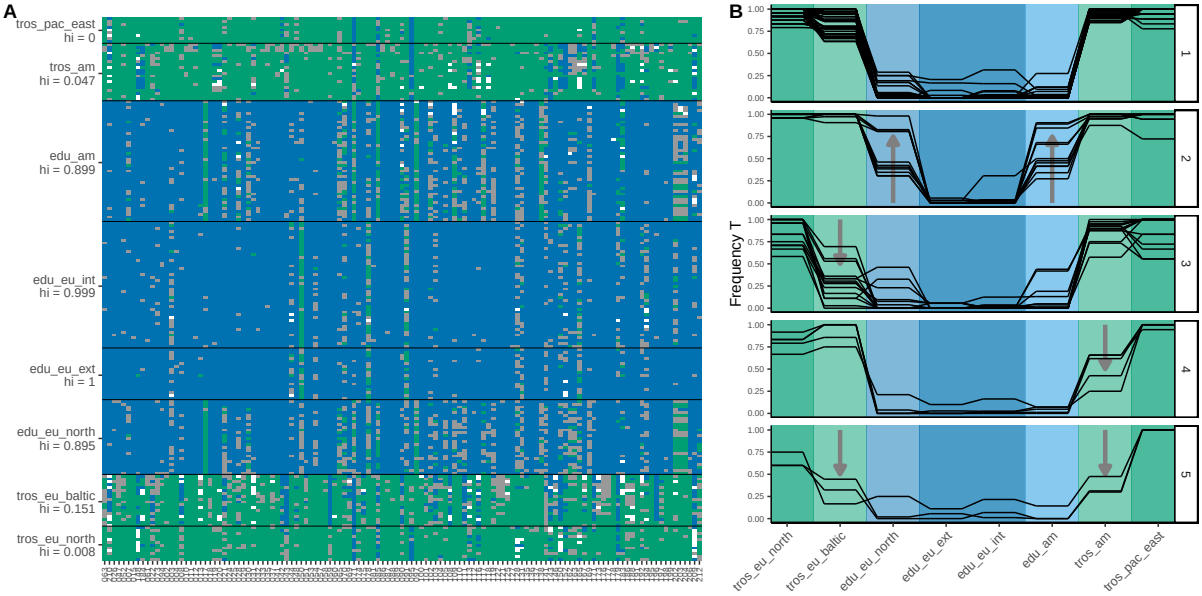

Figure 4: Visualisation of introgression between *M. trossulus* and *M. edulis* using the reference dataset. **(A)** Plot of the raw genotypes. Green: homozygous *M. trossulus*; blue: homozygous *M. edulis*; grey: heterozygous. The orientation of alleles and mean hybrid indexes for each population (hi) 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 $= 0.003$).

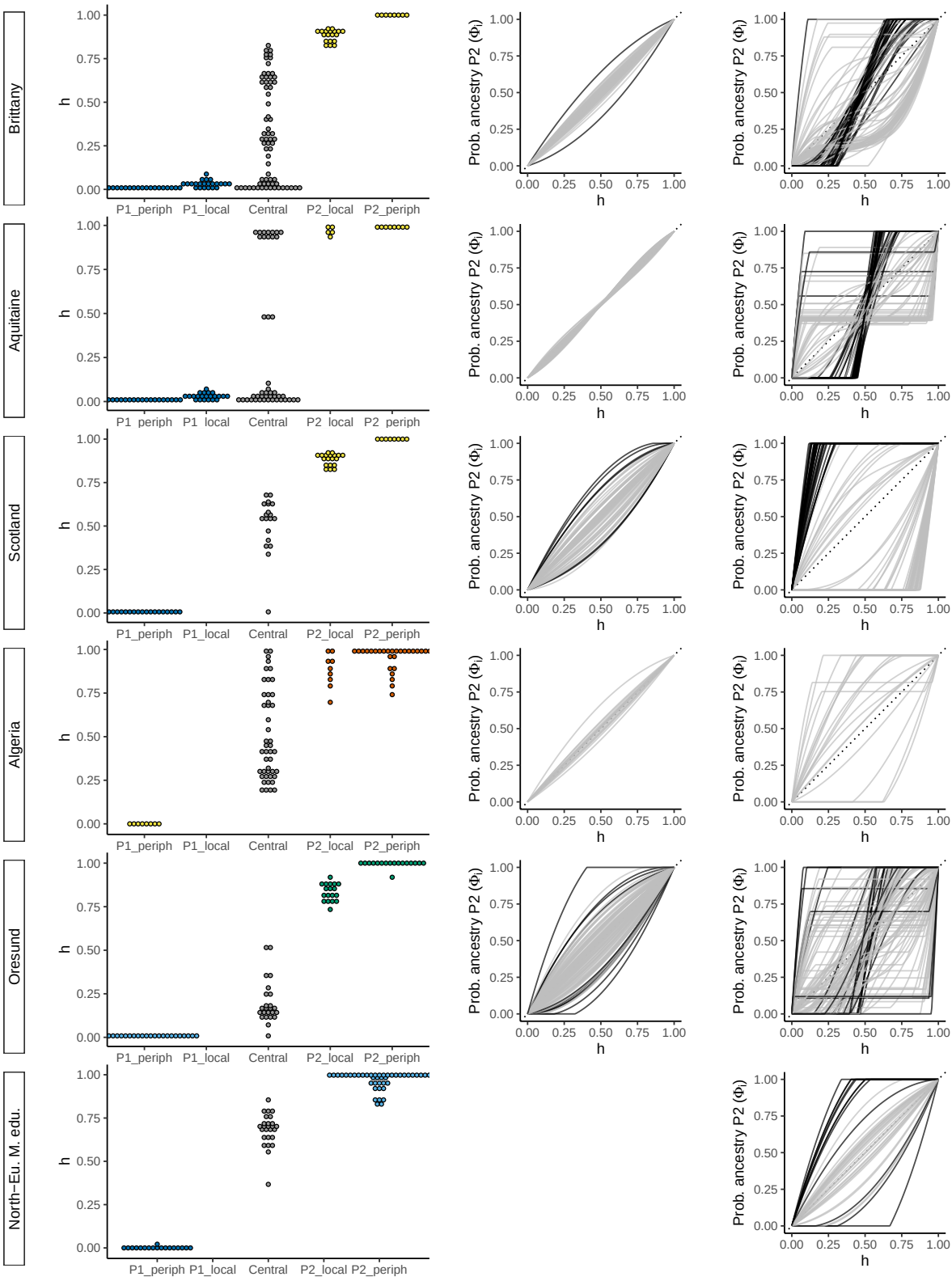

Figure 5: **Left panels:** Hybrid index distributions from `introgress`, for each group of the hybrid zone. **middle panels:** Genomic clines computed with the local parental groups (P1_local, P2_local or peripheral when not available) with clines presenting an excess in either α or β parameters highlighted in black. **right panels:** Genomic clines computed with external populations as parental groups and other groups as admixed. For genomic clines, 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; Luttikhuizen, 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 carryM. 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

Luttikhuizen, 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., & Väinolä, 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., & Väinolä, 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-LgIs-SHM	6	41.386917	-71.519167	Snug Harbor Marina, Charlestown, Long Island, USA
11-LgIs-QNT	6	41.332694	-71.713333	Quonochontaug, Long Island, USA
12-LgIs-BYY	3	41.346694	-71.967583	Brewer Yacht Yard, Mystic River, Long Island, USA
13-LgIs-OST	6	41.263278	-72.384194	Old Saybrook Town, Hartlands Drive, Long Island, USA
14-LgIs-BFM	3	41.260556	-72.820722	Branford Harbor, Long Island, USA
15-LgIs-MYC	8	41.210917	-73.051528	Milford Yacht Club, Long Island, USA
16-LgIs-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.

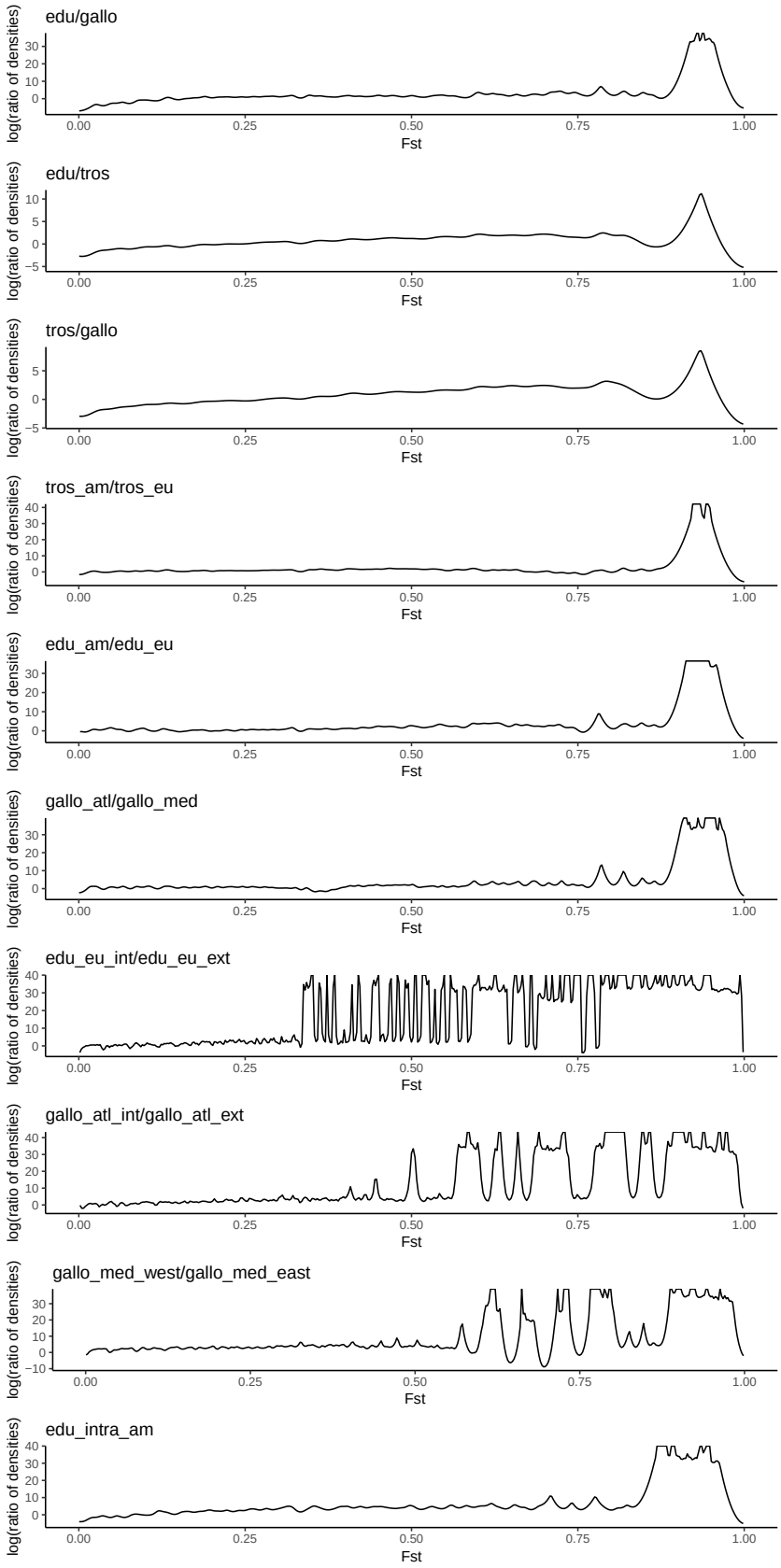

Figure S1: Enrichment in high F_{ST} markers for all comparisons. y-axis: natural logarithm of the ratio of F_{ST} densities between the ancestry informative dataset (present study) and the complete dataset (Fraïsse, Belkhir, Welch, & Bierne, 2016). Densities for F_{ST} between 0 and 1 are presented,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

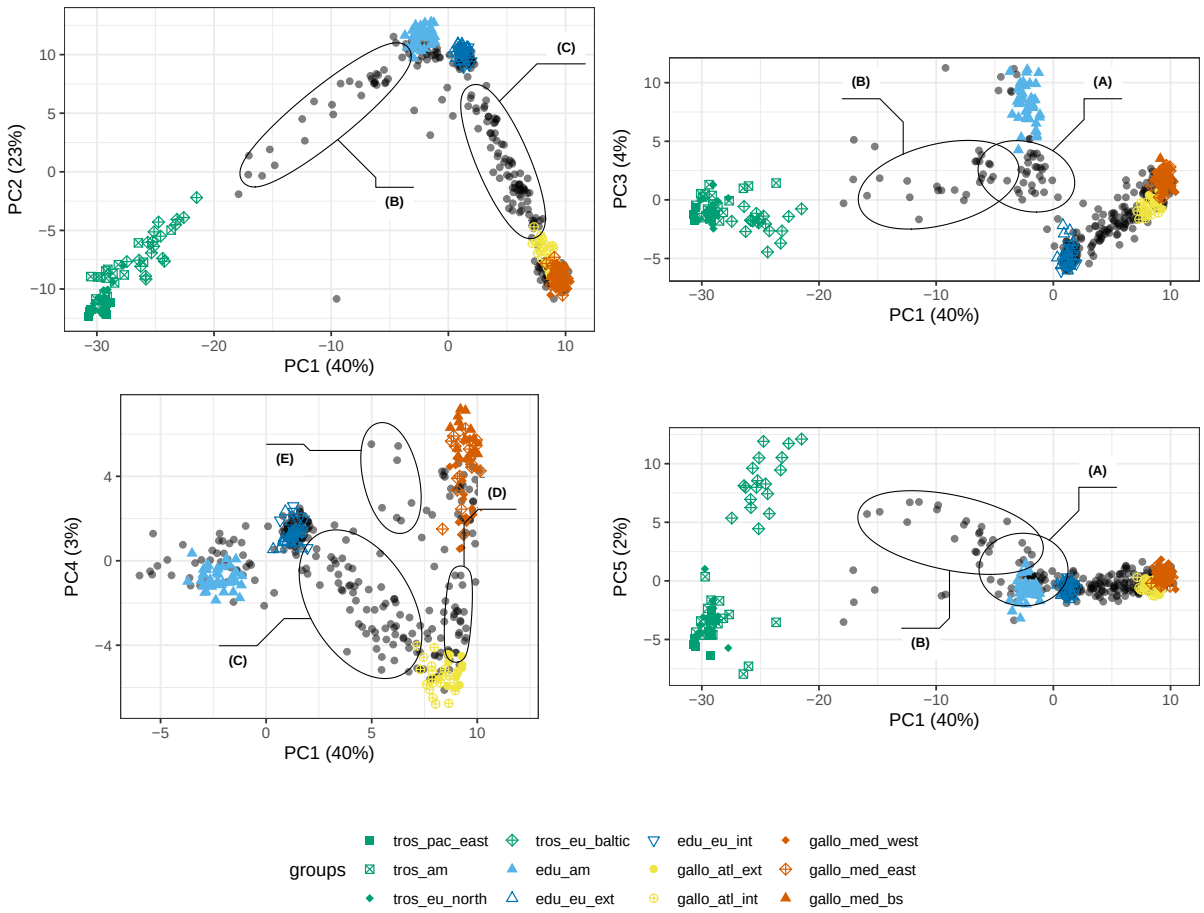

Figure S2: Principal Components Analysis presented for the first 5 components. In the bottom left PCA, *M. trossulus* individuals were cut out from the plot to zoom on the *M. edulis* and *M. galloprovincialis* interesting parts. Annotations show five groups of interest discussed in the main text: (A) 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.

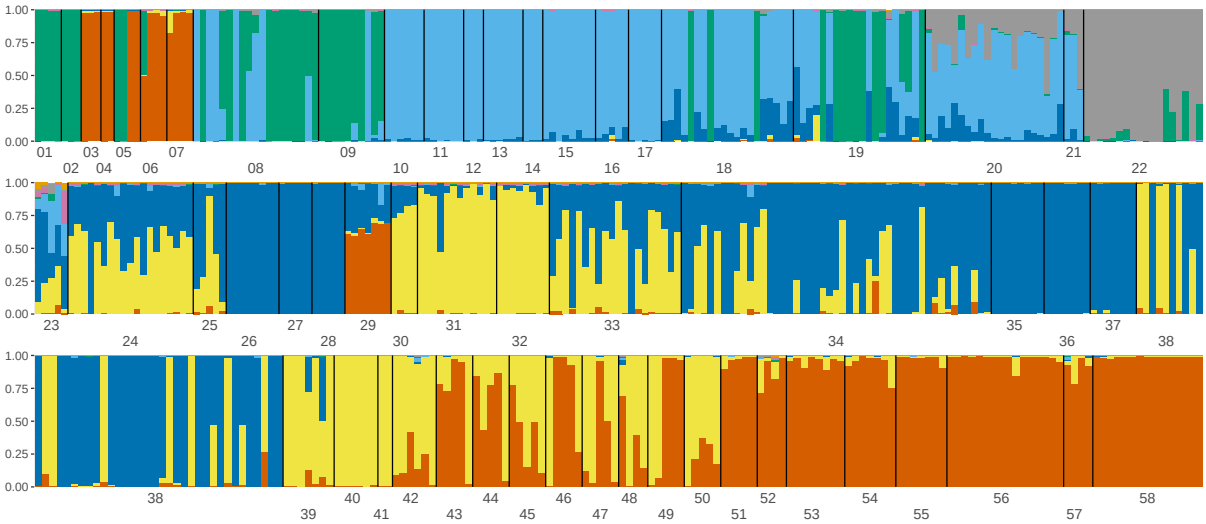

Figure S3: `Structure` results with $K = 8$, for all populations. The major mode obtained with `Clumpak` is presented.

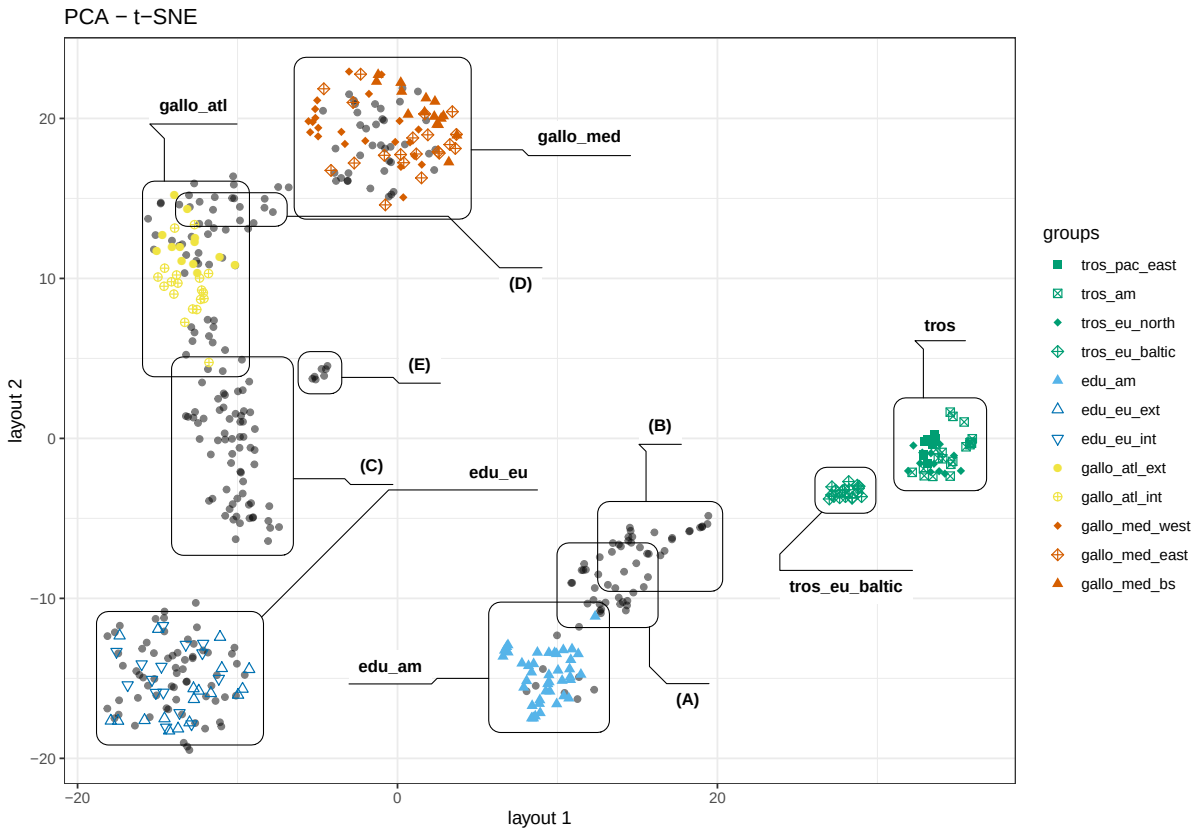

Figure S4: tSNE method applied to the first informative PCA components, as was done for Figure 2.

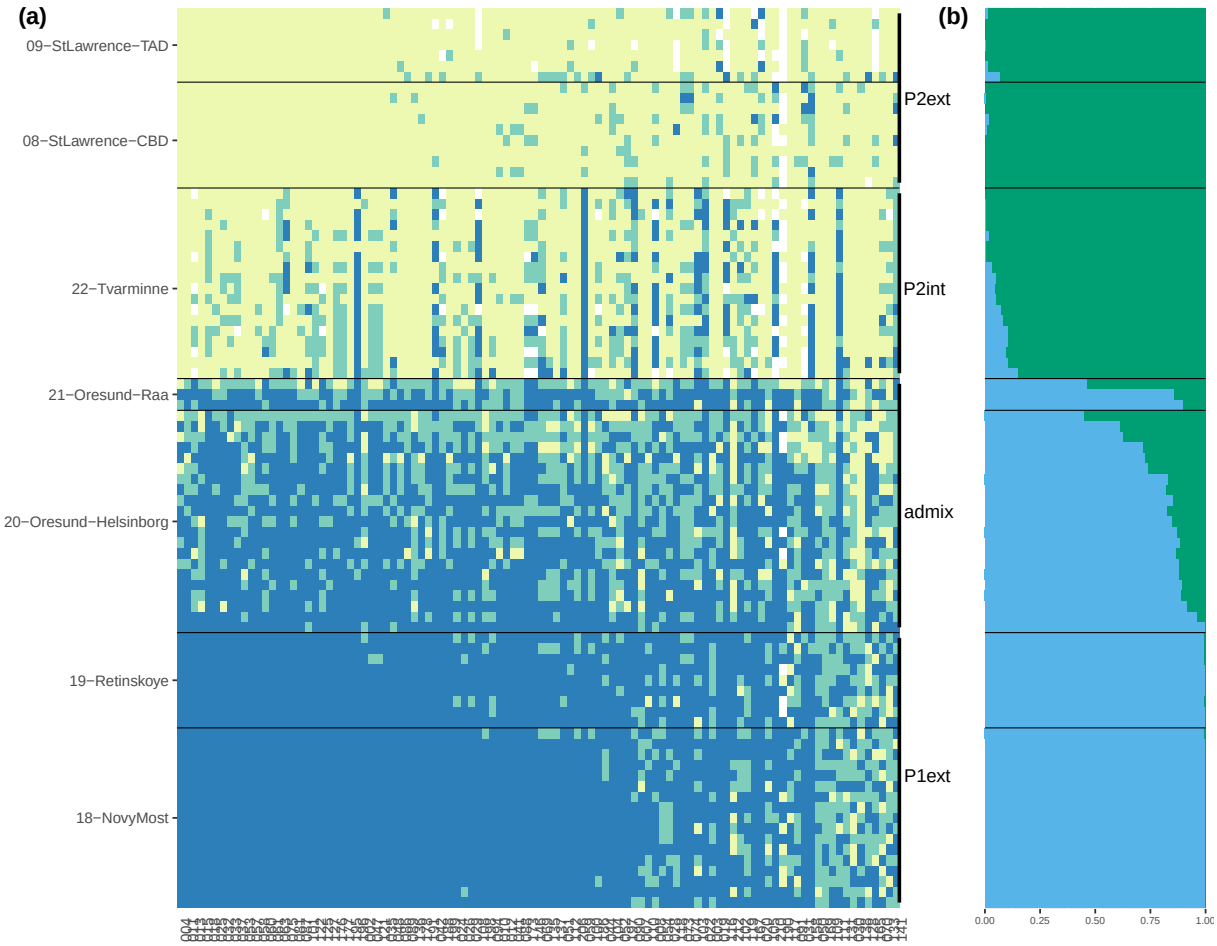

Figure S5: Oresund hybrid zone. (a) `introgress` plot of genotypes for each individual in line and each marker in column (AFD > 0.5). Individuals ar ordered by hybrid index in each population while markers are ordered by degree of differentiation. (b) Corresponding `Structure` ancestry compositions for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

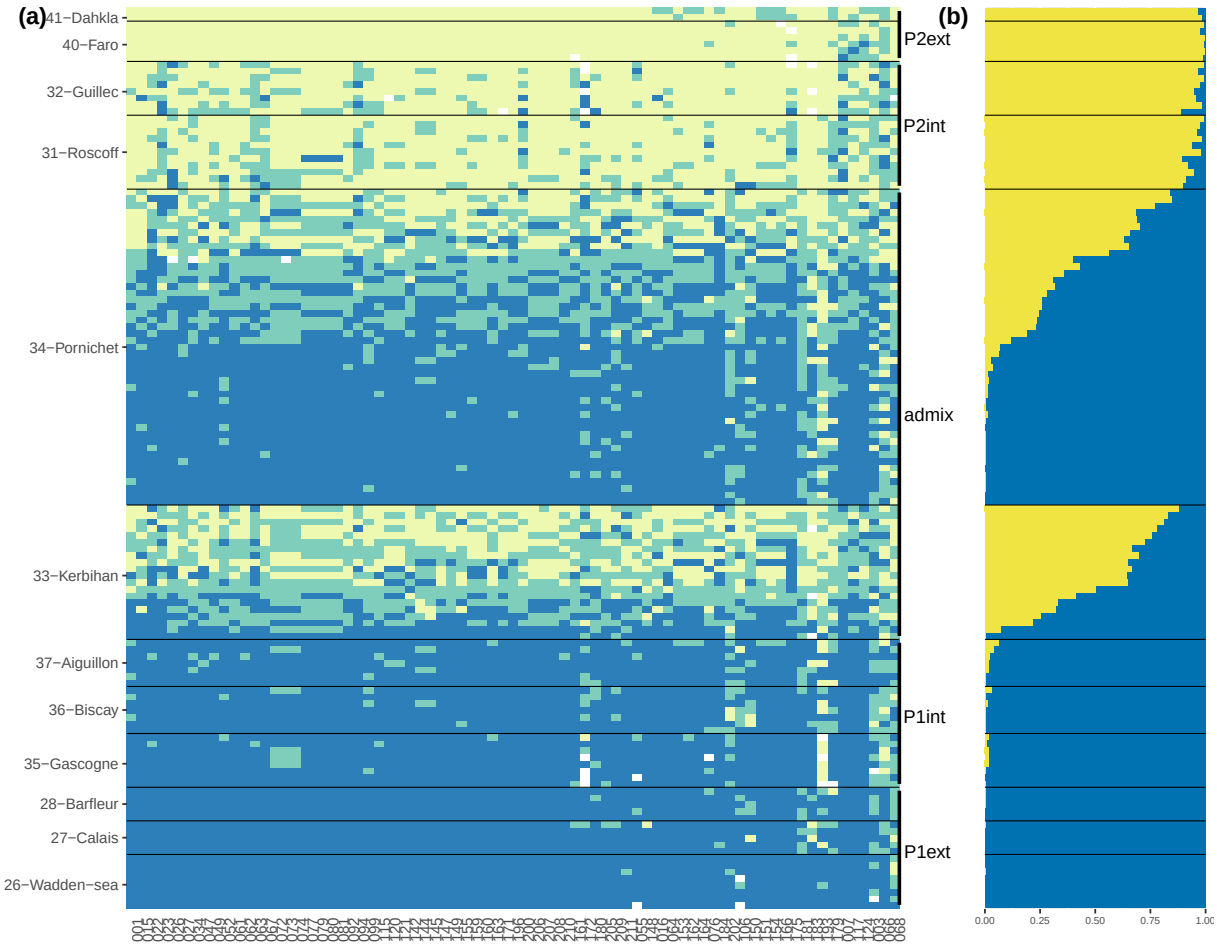

Figure S6: Brittany hybrid zone. (a) `introgress` plot of genotypes for each individual in line and each marker in column (AFD > 0.5). Individuals ar ordered by hybrid index in each population while markers are ordered by degree of differentiation. (b) Corresponding `Structure` ancestry compositions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

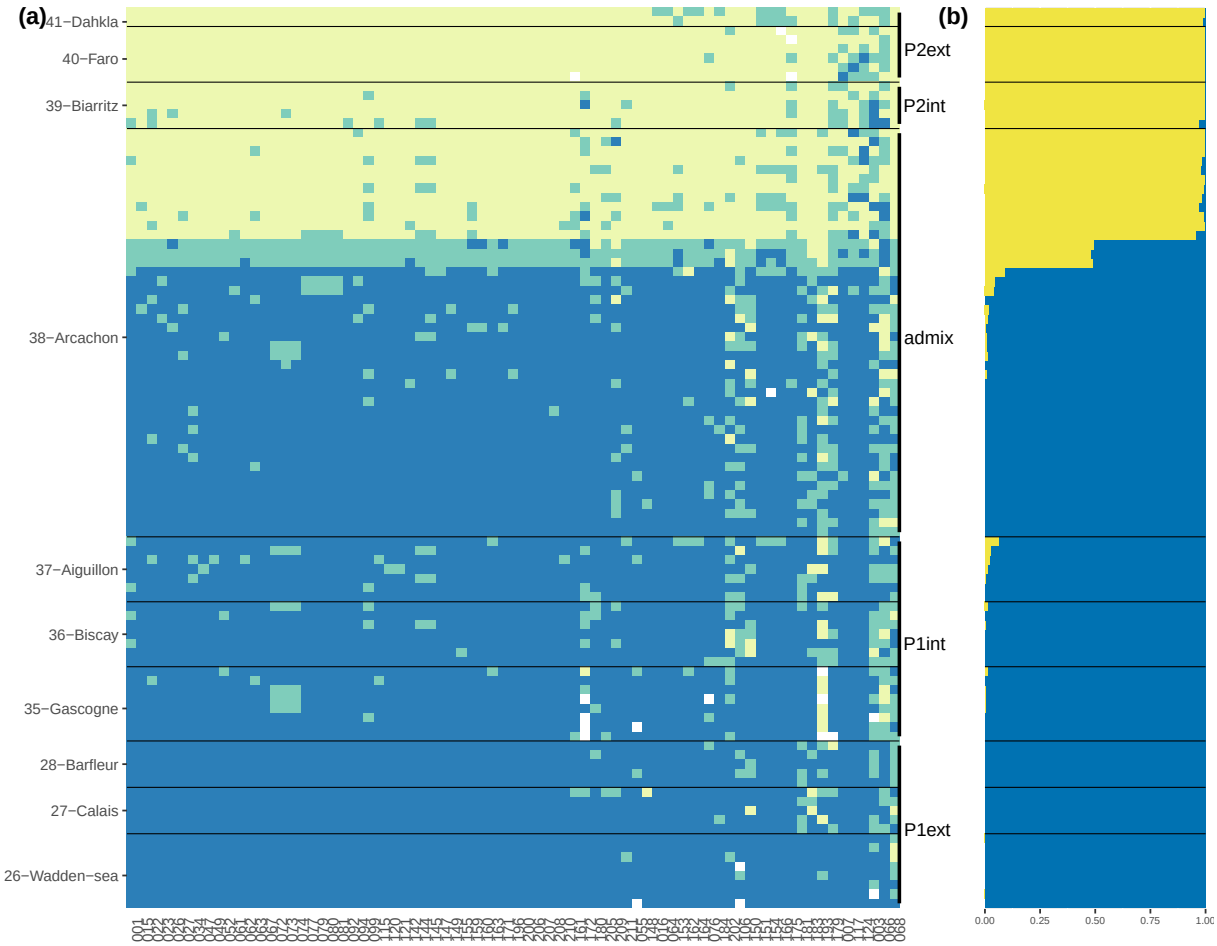

Figure S7: Aquitaine hybrid zone. (a) `introgress` plot of genotypes for each individual in line and each marker in column (AFD > 0.5). Individuals ar ordered by hybrid index in each population while markers are ordered by degree of differentiation. (b) Corresponding `Structure` ancestry compositions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

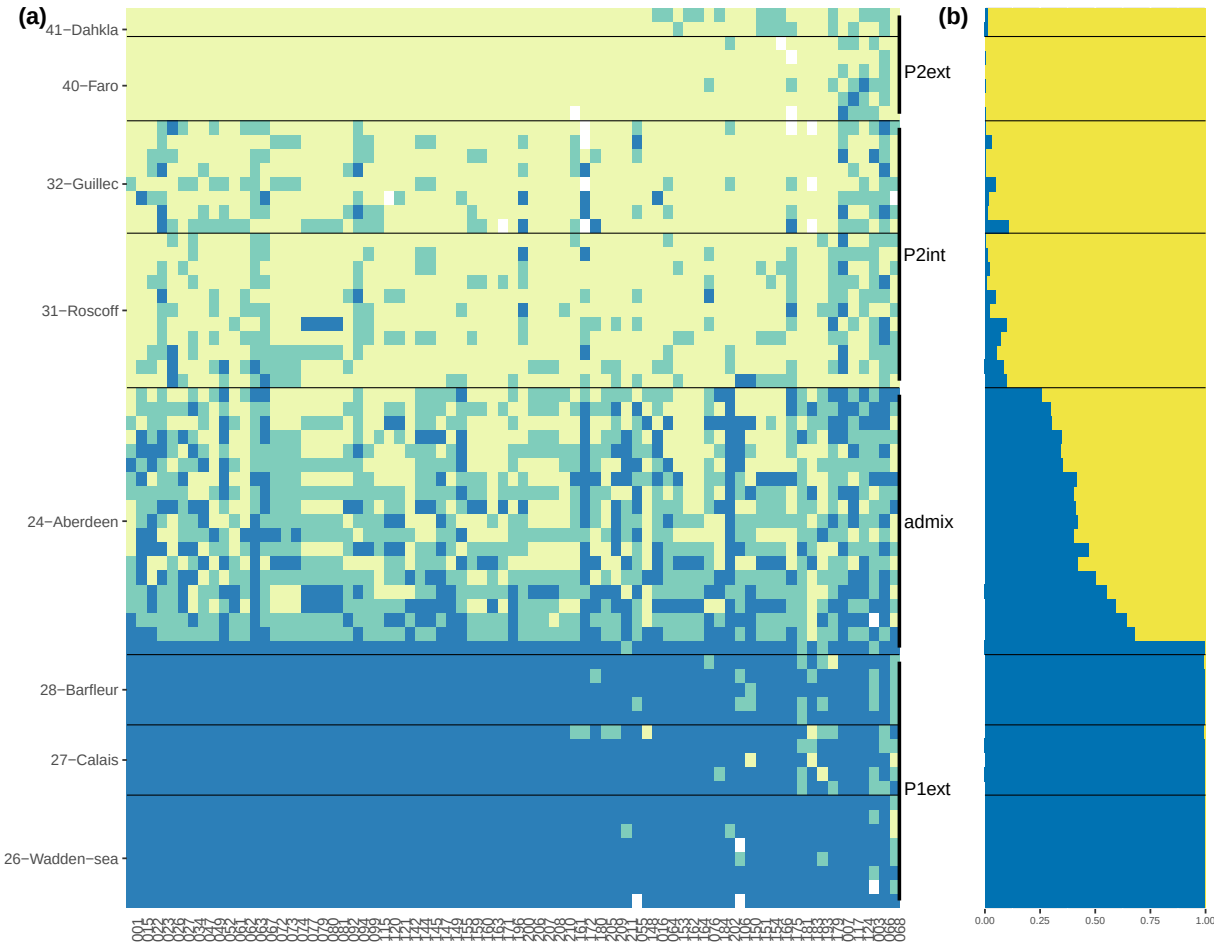

Figure S8: Scotland hybrid zone. (a) `introgress` plot of genotypes for each individual in line and each marker in column (AFD > 0.5). Individuals ar ordered by hybrid index in each population while markers are ordered by degree of differentiation. (b) Corresponding `Structure` ancestry compositions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

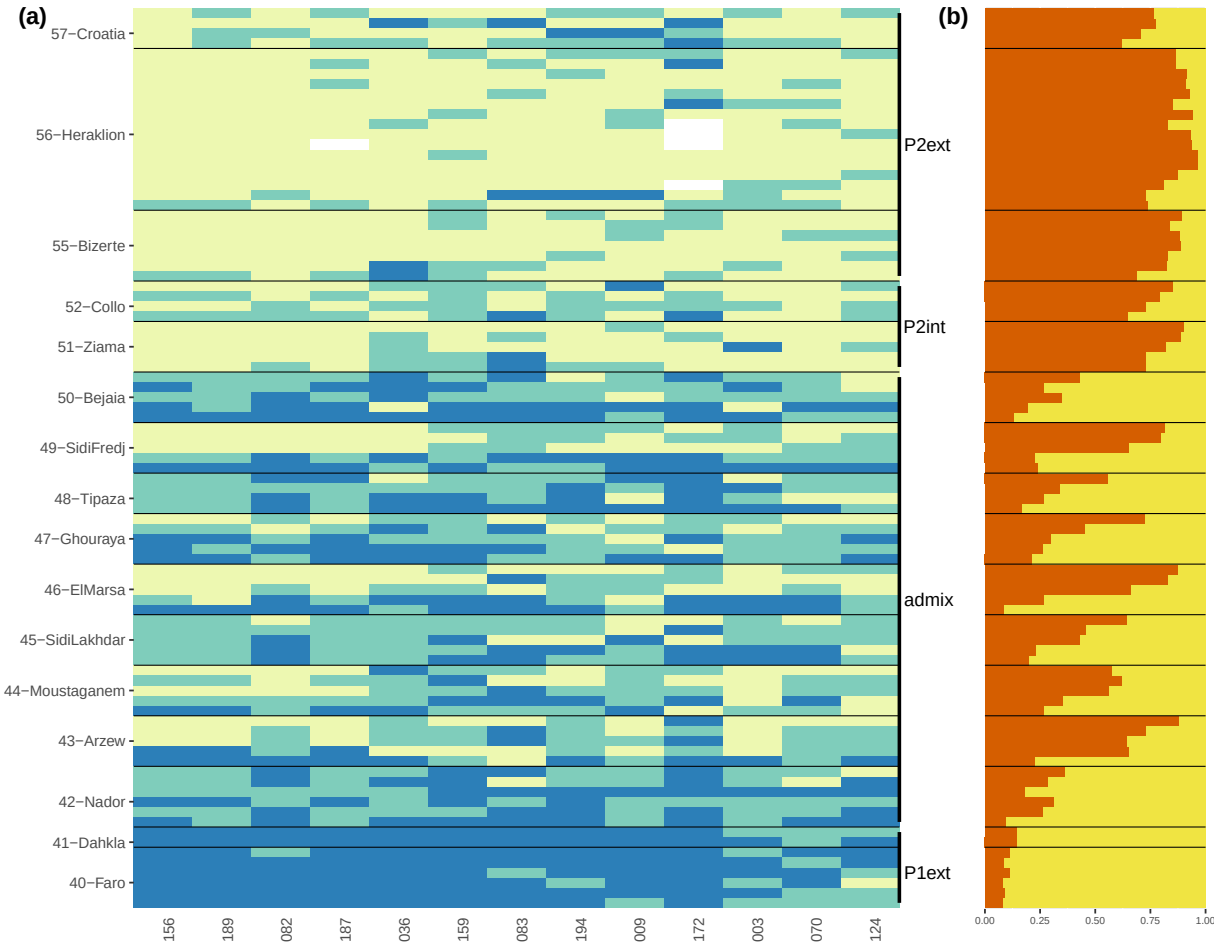

Figure S9: Algerian hybrid zone. (a) `introgress` plot of genotypes for each individual in line and each marker in column (AFD > 0.5). Individuals ar ordered by hybrid index in each population while markers are ordered by degree of differentiation. (b) Corresponding `Structure` ancestry compositions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

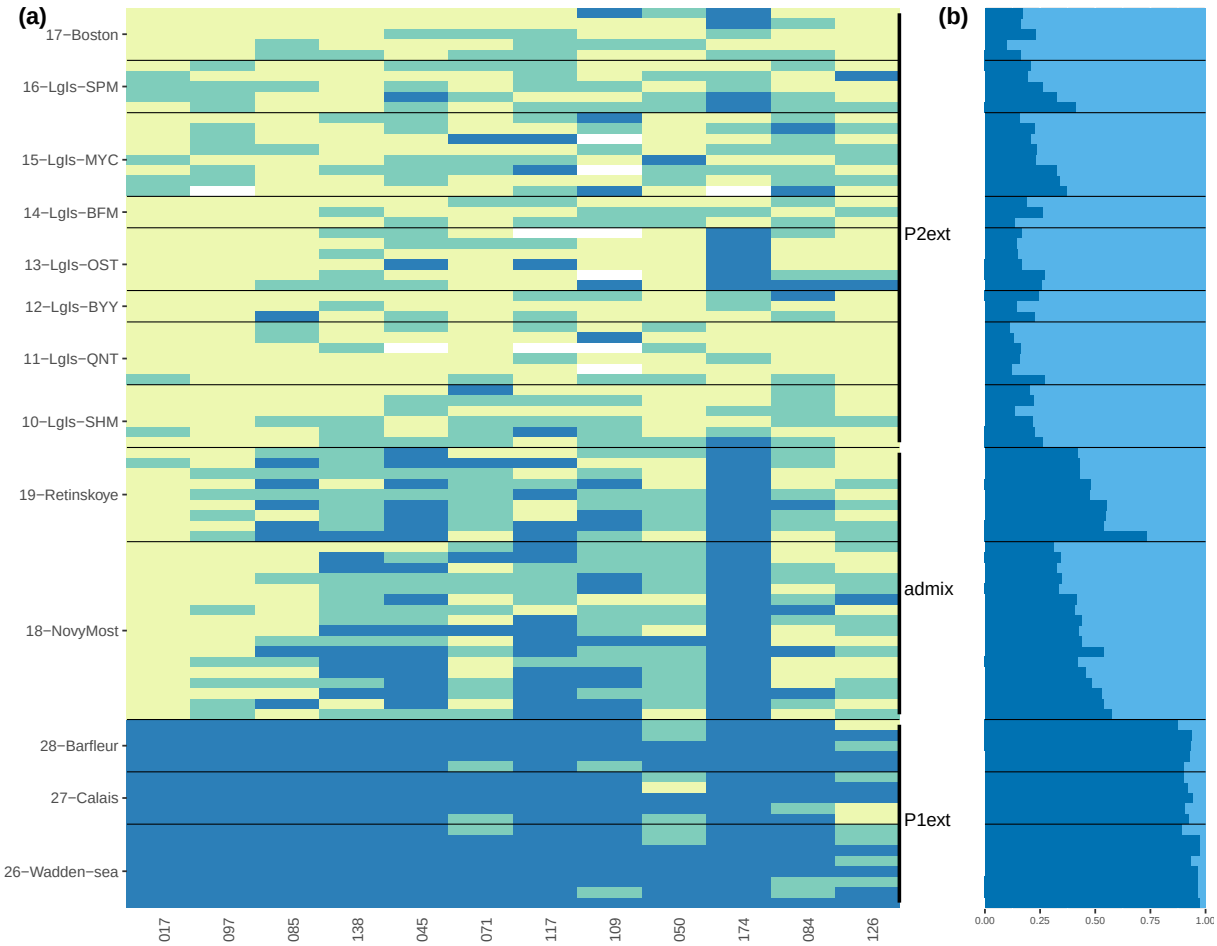

Figure S10: Northern European *M. edulis* hybrid zone. (a) `introgress` plot of genotypes for each individual in line and each marker in column (AFD > 0.5). Individuals ar ordered by hybrid index in each population while markers are ordered by degree of differentiation. (b) Corresponding `Structure`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