Fine-grained genetic-environment association in an admixed population of mussels in the small isolated Kerguelen island

Christelle Fraïsse[1,2,*], Anne Haguenauer[3], Karin Gérard[4], Alexandra Anh-Thu Weber[3,5], Nicolas Bierne[1] and Anne Chenuil[3]

[1] Institut des Sciences de l'Evolution UMR5554, University Montpellier, CNRS, IRD, EPHE, 34095 Montpellier cedex 05, France.

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

[3] Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale - CNRS UMR 7263, Aix-Marseille Université, Station marine d'Endoume, 13007 Marseille, France.

[4] Laboratorio de Ecosistemas Marinos Antárticos y Subantárticos, Instituto de la Patagonia, Universidad de Magallanes, Avenida Bulnes #01890, Punta Arenas, Chile.

[5] Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland.

[*]*Corresponding author: E-mail: christelle.fraisse@ist.ac.at.*

Running title: Admixture and adaptation in a small island

Keywords: *Mytilus* mussels, Adaptation, Native / non indigenous species, Admixture, Introgression, Genetic - environment association.

Abstract

Reticulated evolution -i.e. secondary introgression/admixture between sister taxa- is increasingly recognized as a key evolutionary process that may play a role in promoting adaptation. *Mytilus spp.* is an ideal system to assess its importance, because these marine mussels form semi-isolated species that remain reproductively compatible over large time-scales. They have an antitropical distribution, which includes three hybridizing taxa in the Northern Hemisphere (*M. edulis, M. galloprovincialis* and *M. trossulus*) and two taxa of uncertain ancestry in the Southern Hemisphere (*M. platensis*: South America and the Kerguelen Islands; and *M. planulatus*: Australasia) that originated following transequatorial migrations during the Pleistocene. The Kerguelen mussels are of particular interest to investigate the potential role of admixture in micro-geographic adaptation, as they inhabit a small and isolated island in the Southern Ocean characterized by a highly heterogeneous environment, and genomic reticulation between Northern and Southern lineages have been suspected. Here, we extended a previous analysis by using targeted-sequencing data (1269 contigs; 51,878 SNPs) across the three Northern species and the Kerguelen, coupled with a panel of 33 SNPs genotyped on 695 mussels across 35 sites in the Kerguelen. The panel was enriched with ancestry-informative SNPs, i.e. SNPs that were more differentiated than the genomic average between Northern lineages, to evaluate whether reticulated evolution contributed to micro-geographic adaptation. We first showed that the Kerguelen is a divergent Southern lineage that most probably derived from a proto-edulis population and subsequently experienced admixture with non-indigenous *M. galloprovincialis* and *M. trossulus* mussels. We then demonstrated that the Kerguelen mussels were significantly differentiated over small spatial distance and that the genetic structure was associated with environmental variation.

Introduction

Adaptive divergence at a fine-grained spatial scale is an interesting situation in which neutral processes can easily be disentangled from adaptive ones. Yet, micro-geographic adaptation is expected to be rare because theory shows local adaptation is limited by gene flow when the scale of dispersal is large relative to habitat patch size (Lenormand 2002). Marine species with planktonic larvae are high-dispersal organisms, with large effective population size and living in a highly connected environment (Cowen & Sponaugle 2009), thereby they generally show low level of genetic differentiation across the species range (Palumbi 1992). Nevertheless, there is increasing evidence that local recruitment occurs in marine environments (Swearer *et al.* 2002), highlighting the role of high population size instead of high levels of dispersal in explaining the weak genetic structure of marine species (Gagnaire *et al.* 2015; Bierne *et al.* 2016). Accordingly, micro-geographic genetic-environment associations (GEAs) have been reported at specific loci in marine species, such as barnacles (Schmidt & Rand 1999), mussels (Koehn *et al.* 1980) or Atlantic killifishes (Reid *et al.* 2017), despite genome-wide genetic homogeneity.

Mytilus mussels are subdivided into semi-differentiated taxa distributed worldwide with an antitropical distribution, i.e. they occur in high latitudes of the Northern and Southern Hemispheres, as a result of transequatorial migration during the Pleistocene (Hilbish *et al.* 2000; Gérard *et al.* 2008). In the North, *M. edulis* and *M. galloprovincialis* are closely-related species which started to diverge about 2.5 mya (Roux *et al.* 2014), while *M. trossulus* is clearly an outgroup to them with a divergence dated at 3.5 mya (Rawson & Hilbish 1995). The three species have experienced a complex history of divergence punctuated by periods of gene flow (Roux *et al.* 2014); and nowadays they produce hybrid zones where their ranges overlap (Skibinski *et al.* 1983; Väinölä & Hvilsom 1991; Bierne *et al.* 2003). In the South, allozyme studies initially suggested that mussels from South America and the Kerguelen were related to *M. edulis*, while those from Australasia were related to Mediterranean *M. galloprovincialis* (McDonald *et al.* 1991). However, a reevaluation of the allozyme data and a review of the results obtained with mtDNA and two nuclear DNA markers encouraged to consider them as different taxonomic entities (Borsa *et al.* 2012). The presence of a mitochondrial clade endemic to the Southern Ocean suggests Southern mussels are native rather than introduced by human-mediated activities (Hilbish *et al.* 2000; Gaitán-Espitia *et al.*

2016). This is corroborated by mitochondrial inference of the divergence time between the two hemispheres which is estimated to date from the Pleistocene between 0.5 and 1.3 mya (Hilbish *et al.* 2000; Gérard *et al.* 2008). So far, two alternative scenarios of transequatorial migration have been proposed to explain their origin (Gérard *et al.* 2008): (i) two independent migration events, one from proto-*M. edulis* (producing the South American and Kerguelen mussels, *M. platensis*) and one from proto-*M. galloprovincialis* (producing the Australasian mussels, *M. planulatus*), followed by mitochondrial swamping in Northern populations*;* or (ii) a unique migration event older than the divergence between *M. edulis* and *M. galloprovincialis* followed by geographical differentiation between *M. platensis* and *M. planulatus* with incomplete lineage sorting at nuclear genes.

In the Southern Indian ocean, the isolated Kerguelen Island harbors *Mytilus* mussels which are polymorphic for allozyme alleles characteristic of all three Northern species (Blot *et al.* 1988), although they are most similar to *M. edulis* at a few allozyme loci (McDonald *et al.* 1991). In addition, some mitochondrial haplotypes typical of both *M. edulis* and *M. galloprovincialis* were found in the Kerguelen suggesting the possibility of a recent migration in the island (Hilbish *et al* 2000), although only the southern lineage was subsequently observed (Gérard *et al.* 2008). Further analyses with nuclear markers (mac-1 and Glu-5') revealed the mixed genome ancestry of the Kerguelen mussels (Borsa *et al.* 2007): at Glu-5', a Northern diagnostic marker, mussels carry a heterospecific polymorphism (*M. edulis* / *M. galloprovincialis*) in Hardy-Weinberg equilibrium. Surprisingly, and as opposed to admixed mussels in the Northern hybrid zones (Bierne *et al.* 2003), this polymorphism is not in linkage disequilibrium with the Northern genetic backgrounds, although genetic differentiation is maintained between micro-habitats (Gérard *et al.* 2015). These preliminary results suggest either that reproductive isolation genes responsible of the interspecific barrier in the North were not yet evolved at the time of admixture in the Kerguelen, or that isolation is not as strong in the demographic, ecological and genetic context of the Kerguelen Islands as it is in the Northern Hemisphere hybrid zones.

The geomorphology of the Kerguelen Islands has been shaped by volcanic activity and glacial erosion which resulted in a carved coast with sheltered bays and fjords (Gérard *et al.* 2015). Micro-geographic adaptation in the islands has first been evoked by Blot *et al.* (1989) who reported genetic differences between populations at three allozymes (Lap, Pgm, Pgd) whose

frequencies correlated with salinity and wave exposure. Recently, Gérard *et al.* (2015) have investigated the genetic-environment associations in the island with four nuclear markers (mac-1, Glu-5', EFbis and EFprem's) and a mitochondrial gene (COI). Glu-5' revealed significant genetic differentiation among and within geographic regions, and between habitats. In particular, allele frequencies at Glu-5' were significantly correlated with the presence/absence of the kelp *Macrocystis* in the island. As such, local adaptation was invoked to explain the fine-scale maintenance of polymorphism at Glu-5'. However, we do not usually expect adaptive polymorphisms to be found easily with few markers (Hoban *et al.* 2016) and the ease with which this micro-geographic signal of differentiation has been identified calls for more complex interpretations (Bierne *et al.* 2011).

Because Glu-5' and candidate allozymes are ancestry-informative in the Northern Hemisphere (i.e. they are strongly differentiated between Northern taxa, Skibinski *et al.* 1983, Rawson *et al.* 1996), we might suspect that adaptation in Kerguelen populations may have been enhanced by gene exchange with Northern Hemisphere lineages. A recent study argues that local introgression is widespread in *Mytilus* mussels, and it is the primary cause of outlying levels of genetic differentiation between conspecific populations (Fraïsse *et al.* 2016). Actually, introgression is increasingly acknowledged as an important source of adaptation with many examples collected in plants (Arnold 2004) and animals (Hedrick 2013). Adaptation from hybridizing sister species (or conspecific populations) has been argued to be potentially faster than from new mutations because: (i) incoming beneficial alleles usually start at higher frequencies, (ii) multiple changes within a gene or across multiple loci can be introgressed at once and (iii) adaptive variants coming from a sister-species are generally older than new mutations, so they may have already been tested by selection in the past. Indeed, in the context of an invading lineage experiencing new environmental conditions already faced by the native lineage, introgression of adaptive alleles seems a likely outcome (Wang *et al.* 2014). The influx of heterospecific alleles also creates a departure from equilibrium situations that can better reveal the genetic connectivity within species (Gagnaire *et al.* 2015).

To investigate whether reticulate evolution contributes to micro-geographic adaptation in the Kerguelen islands, we used published GBS data of the three Northern species (Fraïsse *et al.* 2016), new GBS data of a sample from a single Kerguelen population, and a new SNP dataset

from thirty-five Kerguelen populations. Mussels from the Kerguelen were genotyped with a KASpar SNP assay, which was enriched for ancestry-informative loci (i.e., loci that are more differentiated than the genomic average between reference samples in the Northern Hemisphere). We found that the Kerguelen Islands harbor a divergent Southern lineage of mussels that we propose to consider as the native lineage and that most probably derived from a proto-edulis ancestral population and was subsequently admixed with non-indigenous *M. galloprovincialis* as well as with *M. trossulus*. We then confirmed a significant fine-scale genetic differentiation between sites associated with environmental variables. Notably, we found that loci with a more pronounced GEA also tended to be among the most differentiated loci between *M. edulis* and *M. galloprovincialis* in the Northern Hemisphere. We discuss the importance of introgression from past admixture events with Northern lineages and its evolutionary role on the adaptive history of the Kerguelen mussels.

Materials and Methods

Genotyping-by-sequencing of the *Mytilus* spp.

We used samples collected from eleven localities in the Northern Hemisphere (Supp. Materials & Methods and Table S1) to investigate the patterns of admixture between Northern and Southern genetic backgrounds in the Kerguelen Islands. The genetic composition of these samples has been analysed in Fraïsse *et al.* (2016) with target enrichment sequencing of BAC and cDNA sequences. They have been shown to be representative of populations of the *Mytilus edulis* species complex, which comprises three species that hybridize at several places in the Northern Hemisphere: *M. galloprovincialis*, *M. edulis* and *M. trossulus*. In addition to these previously published samples, eight individuals from the Kerguelen Islands (Baie de la Mouche, Table S1) were included in the target enrichment experiment. These individuals were treated together with the Northern samples following the genotyping-by-sequencing (GBS) method described in Fraïsse (2016) (see Supp. Materials & Methods for details). The final dataset across the twelve localities consisted of 1269 reference sequences (378 BAC contigs, 891 cDNA contigs) and 129,346 SNPs. DNA sequences and VCF files including GBS genotypes are available on Dryad doi: 10.5061/dryad.6k740 (Fraïsse *et al.* 2016)

KASPar SNP panel

Based on the SNP database generated by GBS, we specifically selected SNPs segregating in the eight GBS Kerguelen individuals to analyse the fine-scale genetic structure in the island, and its relation to the local environment. Moreover, as we wanted to determine if adaptation in the Kerguelen was primarily driven by standing variation in the Northern complex of species (i.e. SNPs fixed between Northern species), the selected SNPs were not a random sample of the SNPs detected by GBS, otherwise they would have been mainly private polymorphisms to the Kerguelen (60% of the Kerguelen SNPs are private). As such, we further enriched our SNP array with ancestry-informative markers, the most differentiated SNPs between pairs of Northern Hemisphere species (representing 33% of the non-private Kerguelen SNPs, 10% of the whole SNP dataset), namely the West-Mediterranean *M. galloprovincialis* population, the North-Sea *M. edulis* population and the Baltic-Sea *M. trossulus* population. FST values (Weir & Cockerham 1984) were calculated using the R package hierfstat (Goudet 2005) for each SNP between pairs of populations (File S1). SNPs in the upper 15% of the empirical FST distribution were categorized as highly-differentiated. Any SNPs with more than 25% of missing data were discarded. Retained SNPs were further filtered-out based on Illumina Assay Design Tool scores (available on Illumina web page, http://support.illumina.com) which predicts probes success based on the number of degenerated sites in the flanking sequences (250 bp on each side of the focal SNP). The final array comprised 58 SNPs out of which 30 were highly differentiated between Northern species (11 *trossulus*-specific, 8 *edulis*-specific and 10 *galloprovincialis*-specific, Table S2).

KASPar genotyping in the Kerguelen Islands

We used samples collected from 35 sites in the Kerguelen Islands by Gérard *et al.* (2015), totalling 695 individuals (Supp. Materials & Methods and Table S3). Pieces of mantle tissue were preserved in 95% ethanol, and DNA was extracted with the Macherey-Nagel NucleoSPin 96 Tissue kit. A KASPar genotyping assay (Smith & Maughan 2015) was used to genotype the 58 SNPs, of them, 44 SNPs were successfully amplified. We removed seven loci which showed significant FST values between the eight GBS Kerguelen individuals and the KASPar individuals. These may be due to error in the genotyping-by-sequencing, typically the assembly of paralogous loci in two alleles of the same locus, or alternatively to problem of

amplification in the KASPar assay as a consequence of primer design. We further eliminated two loci with null alleles (significant Fis values in most of the sampling sites) and two loci physically linked to one another. The final dataset was composed of 33 KASPar SNPs (Table S2). Additionally, we included allele frequency data of a length-polymorphism locus in the adhesive plaque protein gene, Glu-5', previously scored in the same sampling sites (Gérard *et al.* 2015). Genotypes for all individuals at each KASPar SNP is available in File S2, and population allele frequencies are given in Table S4.

Genetic network of the *Mytilus* spp.

Genotypes of the GBS dataset were statistically phased with beagle v3.3.2 (Browning & Browning 2007) using genotype likelihoods provided by bcftools. All individuals were included in the analysis to maximize linkage disequilibrium, and 20 haplotype pairs were sampled for each individual during each iteration of the phasing algorithm to increase accuracy. Phased sequences (haplotypes) were then generated using a custom perl script. An individual genetic network analysis was conducted with splitstree4 v4.12.6 (Hureson & Bryant 2006) to get insight into the population relationships across the three Northern Hemisphere species and the eight individuals sampled in the Kerguelen Islands. All haplotype loci were compiled to create an artificial chromosome of 51,878 high-quality SNPs and analysed using the neighbour-net method.

Analyses of admixture in the *Mytilus* spp.

An estimation of the historical relationships among the eleven Northern populations and the GBS Kerguelen population was performed with *TreeMix* v.1.1 (Pickrell & Pritchard 2012). A maximum-likelihood population tree was estimated based on the matrix of GBS allele frequency covariance between population pairs, and admixture events were sequentially added. To account for linkage disequilibrium, variants were grouped together in windows of size k=100 SNPs. Trees were rooted with the two *M. trossulus* populations and no sample size correction (option "-noss") was applied. We tested for a range of migration events from m=0 to m=12, and looked for an asymptotic value of the log-likelihood. The number of significant migration events was assessed by stepwise comparison of AIC values. Finally, we made 100

bootstrap replicates (option "–bootstrap") of the maximum-likelihood tree to assess statistical support of migration edges.

Additionally, we performed model-based clustering analysis of these populations based on the GBS genotypes. Ancestry of each individual was estimated using the Maximum-likelihood approach implemented in ADMIXTURE v1.23 (Alexander *et al.* 2009). We ran 50 replicates for a number of clusters from K=2 to K=8 and chose the maximum log-likelihood run for each K.

Topology weighting of the *Mytilus* spp.

The distinct haplotype loci of the GBS dataset were also individually analysed with the neighbour-net method. Allele genealogies were inferred with the R package APE (Paradis 2010) using a neighbour-joining algorithm with F84 distances (Felsenstein & Churchill 1996). Haplotype loci were filtered based on the following excluding criteria: scale < 0.00005 ; 0.00005 =< scale < 0.0005 & length < 10000 bp ; 0.0005 =< scale < 0.001 & length < 5000 bp ; and scale >= 0.001 & length >= 1000 bp, where "scale" is the scale of the gene tree and "length" is the length of the sequence. Neighbour-joining trees of the 395 retained sequences are available in File S3.

For each haplotype locus, the relationships between the Northern species and the Kerguelen were then quantified using *Twisst* (Van Belleghem *et al.* 2017), a tree weighting approach. We tested the three possible unrooted topologies: *M. trossulus* grouped with the Kerguelen, *M. edulis* grouped with the Kerguelen and *M. galloprovincialis* grouped with the Kerguelen. Their exact weightings to the full tree were estimated by considering all subtrees ("complete method"). Only contigs with a resolved topology were analysed: 69 contigs for which one topology had a weight greater or equal to 0.75. These topologies were further classified in two categories depending on whether they most plausibly reflect: (i) ancient divergence of the Kerguelen clade (i.e. the Kerguelen and Northern individuals clustered into distinct monophyletic groups) or, (ii) introgression with one of the Northern species (i.e., the Kerguelen individuals were distributed within one or more Northern clades); "na" stands for topologies that we were unable to classify in these two categories. Tree topology weightings and classification are available in Table S5.

Analyses of genetic variation in the Kerguelen Islands

For each KASPar SNP, estimation of FST values (Weir & Clark Cockerham 1984) was calculated over all sampling sites (Table S2), and in a pairwise manner across all SNPs (Table S6) using Genetix 4.05 (Belkhir *et al.* 2002). Their significance was tested by a permutation procedure (1000 permutations) and adjusted with the Bonferroni's correction for multiple comparisons (Benjamini & Hochberg 2000).

Analysis of habitat variables in the Kerguelen Islands

To evaluate how much of the genetic variation among sites was explained by local environmental factors, we used redundancy analysis (RDA), a constrained ordination method implemented in the R package vegan (Oksanen *et al.* 2017). Geographic coordinates and five qualitative factors were measured in each site to describe the local habitat (Table S3): (i) Substrate (rock: R, blocks: B, gravels: G, or sand: S); (ii) Wave Exposure (sheltered: Sh, or exposed: E); (iii) Slope (flat: F, steep: St, or hangover: H); (iv) Salinity (oceanic water: OW, or low-salinity water: LSW); (v) *Macrocystis* (presence: P, or absence: A).

We specifically tested the effect of each of these constrained factors (explanatory variables) on the distribution of genotypes at the 33 KASPar SNPs (response variables). The following initial model was used: Genotypes ~ *Macrocystis* + Salinity + Slope + Exposure + Substrate + Longitude + Latitude. The significance of the global model was first established by permutation test, in which the genotypic data were permuted randomly and the model was refitted (1000 permutations). Marginal effect permutation tests were then performed to assess the significance of each factor by removing each term one by one from the model containing all other terms (1000 permutations). Nonsignificant factors were removed from the final model. Based on that model, we performed a conditioned RDA analysis for each factor to assess its contribution to the genotypic variance independently from the other explanatory variables. These co-variables were removed from the ordination by using a condition function: Genotypes ~ tested variable + condition(all other variables). Finally, we performed a conditioned RDA on geography to specifically control its confounding effect: Genotypes ~ significant environmental variables + condition(significant geographic variables).

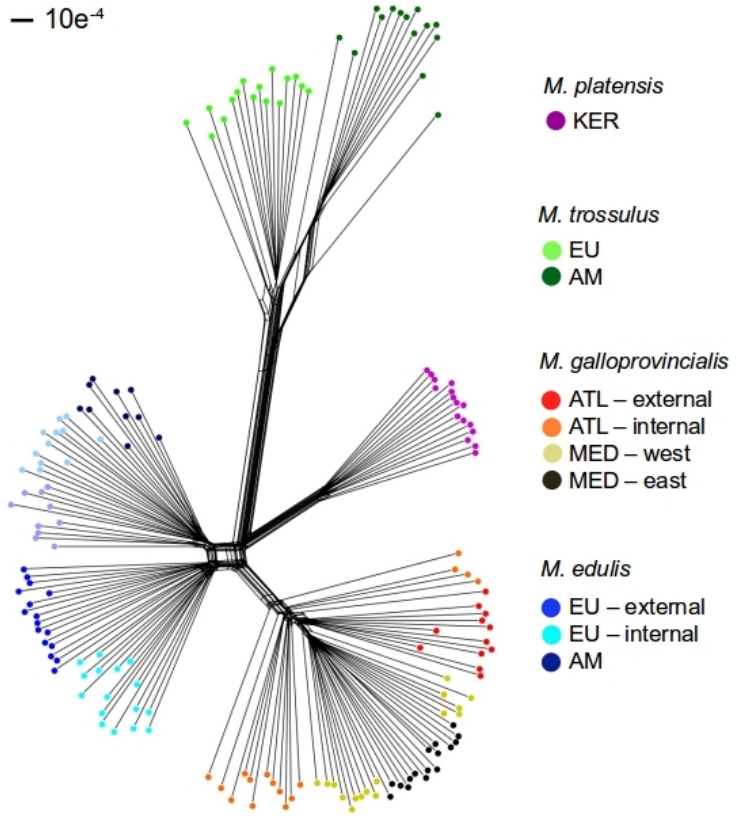
Results

The Kerguelen mussels: signal of divergence of a Southern lineage after transoceanic migration and secondary admixture with Northern lineages

An individual genetic network (Figure 1) was built from a subset of 51,878 high-quality SNPs genotyped in eleven Northern populations and eight individuals from the Kerguelen Islands. We observed that the Northern populations formed three distinct clusters, corresponding to the three Northern species: *M. edulis*, *M. galloprovincialis* and *M. trossulus*. Accordingly, the majority of SNPs fixed between populations (295 in total) were species-specific: *M. edulis*=6, *M. galloprovincialis*=62 and *M. trossulus*=224. The Kerguelen individuals clustered together into a single divergent clade. Indeed, the proportion of SNPs which were private to the Kerguelen Islands amounted to 60% (3805 private for a total of 6297 SNPs in Kerguelen, after removing singletons). In comparison, the number of private SNPs in *M. trossulus* was 3070, and it was only 492 in *M. galloprovincialis* and 48 in *M. edulis* (indicative of introgression between the two latter species). Among the 2492 SNPs shared by the Kerguelen mussels with Northern species, 33% (830) were highly differentiated between at least two Northern species. When considering Northern species-specific SNPs, 83% of those fixed in *M. edulis* were segregating in the Kerguelen (5 for a total of 6 fixed). These numbers were 16% for *M. galloprovincialis* (10 for a total of 62 fixed) and 12% in *M. trossulus* (27 for a total of 224 fixed). A multivariate analysis on KASpar-typed SNPs, including the Northern samples, the Kerguelen Islands and other samples from the Southern Hemisphere that were also genotyped in our SNP assay, is provided as a supplementary figure (Figure S1). The PCA clearly shows that the Chilean mussels (MAU) group with the Kerguelen mussels in accordance with them being both named *M. platensis*; while the Australasian samples (Australia, Tasmania and New-Zealand) usually named *M. planulatus* cluster with the Northern *M. galloprovincialis*. These findings corroborate previous results based on mitochondrial DNA (Gérard *et al.* 2008) and nuclear markers (Borsa *et al.* 2012).

Figure 1. Genetic network of the Northern- and Southern- Hemisphere *Mytilus spp.* (12 GBS populations) produced with the neighbour-net method based on 51,878 high-quality GBS SNPs. The *M. platensis* sample is 'Baie de la Mouche' (KER, purple) in the Kerguelen Islands (Southern Ocean). *M. trossulus* samples are 'Tvarminne' (EU, light green) in the European population of the Baltic Sea and 'Tadoussac' (AM, dark green) in the American population of the Saint Lawrence River. *M. galloprovincialis* samples are 'Faro' (ATL – external, red) in the

Atlantic population of Iberian Coast, 'Guillec' (ATL – internal, orange) in the Atlantic population of Brittany, 'Sete' (MED – west, yellow) in the Occidental Mediterranean basin and 'Crete' (MED – east, black) in the Oriental Mediterranean basin. *M. edulis* samples are 'Wadden Sea' (EU – external, light blue) in the European population of the North Sea, 'Lupin/Fouras' (EU – internal, cyan) in the European population of the Bay of Biscay and 'Quonochontaug/Old Saybrook Town' (AM, dark blue) in the American population of Rhode Island.



The species relationships found in the genetic network (Figure 1) were generally supported by the maximum-likelihood population tree inferred by *TreeMix* (Figure 2), except that the Kerguelen population was inferred as the sister-group of *M. edulis*. The pairwise population residuals in a model without admixture (Figure S2) suggested substantial migration between species. So, we sequentially allowed from 0 to 12 migration events in the analysis, and assessed their significance by stepwise comparison of AIC values (Figure S2). The best fit to the data was obtained with seven migration events, which significantly improved the log-likelihood of the model (Figure S2). This population tree was bootstrapped 100 times to assess statistical support of migration edges. Three migration edges had more than 50% bootstrap support (Figure 2 and Table S7). The more robustly inferred migration event was between the Mediterranean *M. galloprovincialis* and the Kerguelen (81 % of bootstrap replicates). The two others included migration between Northern species as expected: the European populations of *M. edulis* and *M. galloprovincialis*, and the European populations of *M. edulis* and *M. trossulus*. A migration event was also inferred between the Mediterranean *M. galloprovincialis* and the European *M. trossulus*. An edge between the Kerguelen and the

American *M. trossulus* was additionally inferred for 38% of bootstrap replicates. Eventually, migration between the Kerguelen and the European *M. edulis* was detected, but only in 4% of bootstrap replicates. Overall, these results suggest that the Kerguelen mussel is a *M. edulis*-derived Southern lineage secondarily admixed with *M. galloprovincialis* and to a lesser extent with *M. trossulus*.

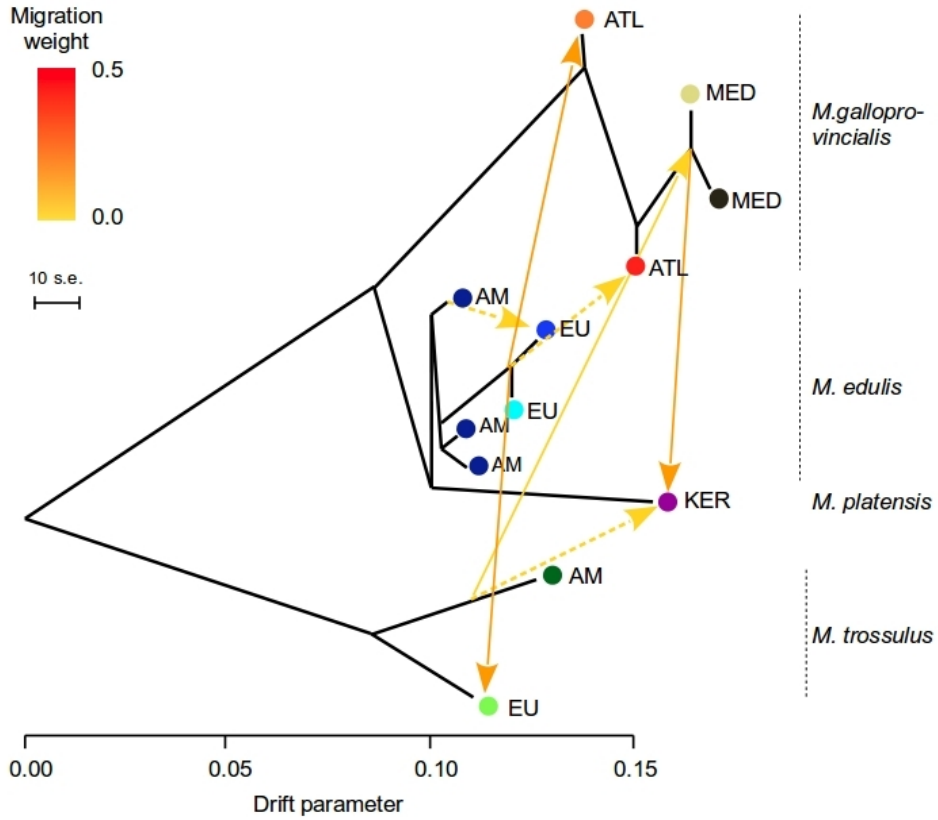


Figure 2. Maximum-likelihood population tree of the 12 GBS *Mytilus* spp. populations inferred by *TreeMix* showing 7 migration events. ML estimation was made by using 32,162 GBS SNPs grouped in windows of 100 SNPs. Terminal nodes are labelled by locality abbreviation (Table S1) and colors match Figure 1. The drift parameter is shown on the x-axis. Horizontal branch lengths are proportional to the amount of genetic drift that has occurred in each branch. Admixture arrows are coloured according to the migration weight, which represents the fraction of ancestry derived from the migration edge. Migration edges with bootstrap support less than 50% are shown with dotted lines, and those with bootstrap support less than 20% are not shown (Table S7).

Variation of admixture histories across the genome

To further investigate how genetic relationships varied across the genome, we quantified the contribution of three unrooted topologies (Figure 3) to the full tree at 395 GBS contigs with *Twisst*. Only 17% (69) of them showed resolved relationships, i.e. one of the unrooted topology weighted 75% or more, among which 42% (28) were highly resolved (weight >= 90%). A first result of the analysis is therefore a high rate of incomplete lineage sorting. The most represented resolved topology (38 contigs) put the Kerguelen individuals together with *M. edulis*, while they were grouped with *M. trossulus* in 19 contigs (i.e. ancestral to the *edulis*/*galloprovincialis* subgroup) and with *M. galloprovincialis* in 10 contigs (Table 1). When classifying the topologies in subcategories (Figure 3), 14 contigs supported the « ancient Kerguelen divergence » scenario while 24 supported an « introgression » scenario among which 4 were from *M. trossulus*, 16 from *M. edulis* and 4 from *M. galloprovincialis*; 29 contigs could not be classified. Figure S3 illustrates representative cases of the different *Twisst* subcategories, including candidate loci for introgression. Panel A2 represents a complete introgression of *M. trossulus* haplotypes into the Kerguelen Islands. The clearest case is observed for a contig containing the Elongation Factor 1 alpha gene (Figure S3 panel A2), a gene already involved in adaptation in *M. edulis* (Bierne 2010). A similar pattern is shown on panel B2 where *M. edulis* haplotypes have totally replaced their Southern counterparts in the Kerguelen. Panel C2 suggests a more ancient introgression of *M. galloprovincialis* haplotypes given that all haplotypes sampled in the Kerguelen form a distinct cluster within the *M. galloprovincialis* clade. Interestingly, some contigs showed multiple ancestry with both *M. edulis* and *M. galloprovincialis* or *M. trossulus* alleles segregating in the Kerguelen. These results suggest that the Kerguelen mussels have a genome of mixed ancestry, mainly dominated by *M. edulis*-related alleles from which they probably derive, but with which they also have probably secondarily admixed again. This is in contrast with the negligible *M. edulis* introgression found in the *TreeMix* analysis where the Kerguelen was inferred to be the sister-clade of *M. edulis*. In fact, it may have been hard to fully distinguish migration from shared ancestral polymorphism only based on allele frequencies in the ML population tree. Moreover, it should be noted that all these patterns hold when using a minimal weight of 90%, though the proportion of "introgression" cases drop down from 63% to 43% (Table 1).

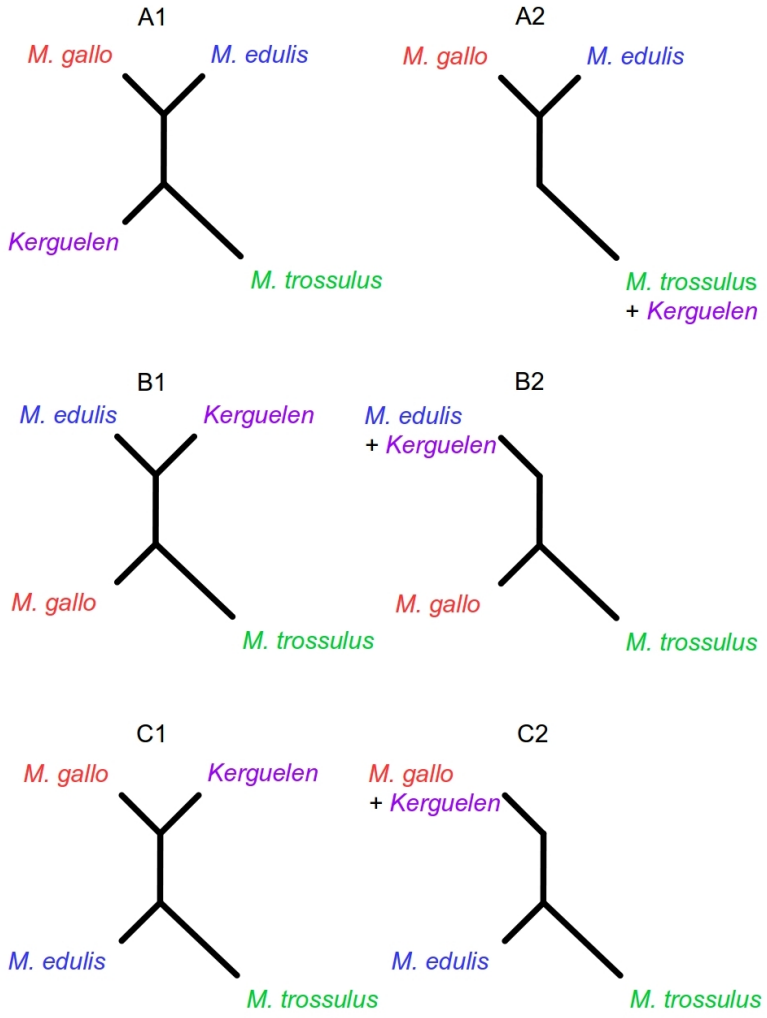


Figure 3. Summary of the different topologies. Three topologies have been weighed with *Twisst* for each of the 395 contigs, and classified in different categories depending whether the Kerguelen individuals branched as a sister-clade to a Northern species ("ancient divergence"), or were distributed within a Northern species ("introgression"). A. Kerguelen clustered with *M. trossulus*: A1 "ancient Kerguelen divergence" and A2: "introgression"; B. Kerguelen clustered with *M. edulis*: B1 "ancient Kerguelen divergence" and B2: "introgression"; C. Kerguelen clustered with *M. galloprovincialis*: C1 "ancient Kerguelen divergence" and C2: "introgression".

Substantial genetic structure in the Kerguelen Islands

Mussels were collected from 35 sampling sites all around the Kerguelen Islands (Figure 4A, Table S3) and successfully genotyped at 33 KASpar SNPs. Pairwise FST values across all SNPs (Table S6) revealed significant fine-scale genetic differentiation between sites from different geographic regions. Remarkably, RdA (North-East) and PCu (West) were significantly differentiated with nearly all other sites. Sites from the South, especially BdS, and from the North, especially AS, were differentiated from the Gulf of Morbihan. At a smaller scale within the Gulf of Morbihan, several sites showed genetic structure among them, but their significance level did not pass the correction for multiple tests. These results extend the study by Gérard *et al.* (2015) to many SNPs and substantiate their finding of significant genetic differentiation at different scales in the island.

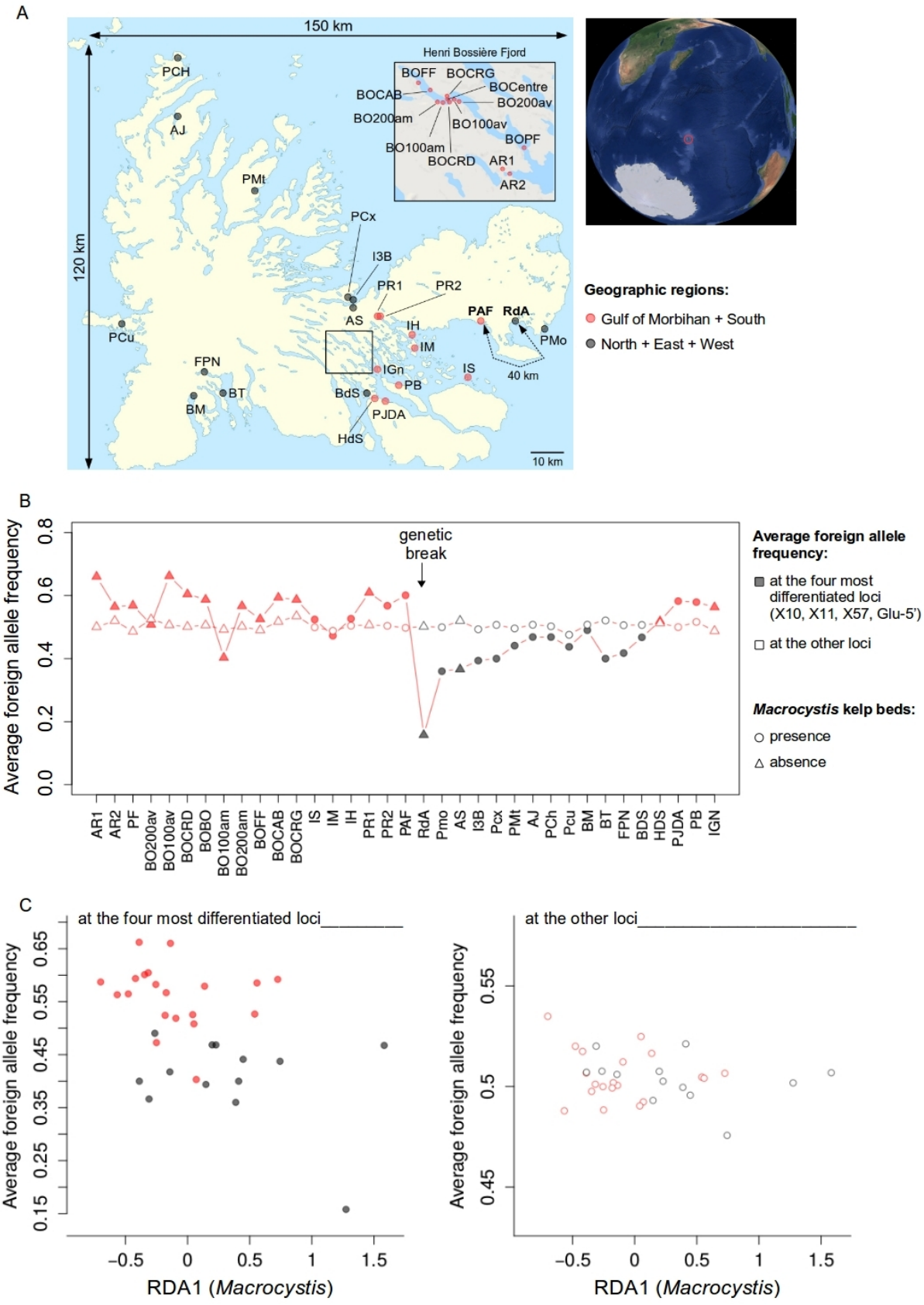


Figure 4. Geographic variation of the average foreign allele frequency across the four most differentiated loci in the Kerguelen Islands (filled symbols, X10, X11, X57 and Glu-5' from Gérard et al. (2015)) and the other loci (open symbols). Alleles were labelled based on their frequencies in the *M. galloprovincialis* Atlantic population of Iberian Coast (Table S4). Red points indicate sites located in the Gulf of Morbihan. **A.** Map of the Kerguelen Islands (150 km East to West; 120 km North to South) together with a world map indicating its location in the Southern Ocean (surrounded in red), and an enlarged map of the Henri Bossière Fjord. Sites PAF and RdA are separated by 40 km of coasts and show a genetic break on Panel B. Sampling details are provided in Table S3. **B.** Frequency of the average foreign allele across sampling sites (ordered by geography). Points (resp. triangles) represent sites characterized by the presence (resp. absence) of *Macrocystis* kelp beds. See Figure S4 for the detailed pattern at each locus. **C.** Correlation between the average foreign allele frequency of most differentiated loci (left panel) and other loci (right panel) at each sampling site and the presence/absence of *Macrocystis* (i.e., the average site coordinates on the first axis of the *Macrocystis* RDA, see Table S11). Pearson correlation coefficient: on the left panel, 0.494 (p=0.003); on the right panel, 0.107 (p=0.541).

Global FST across all sites was calculated for each SNP and tested with 1000 permutations (Table S2). Values were non-significant after Bonferroni's correction, except at the three most differentiated loci: X10, X11 and X57. Their foreign allele, oriented based on its frequency in the Northern species (*M. galloprovincialis* Atlantic population of Iberian Coast, Table S4), was at low frequency in the North of the island, especially in the North-East sites, RdA and PMo (average frequencies: X11=0.217, X10=0.368, X57=0.287). In contrast, it was at higher frequency in the Gulf of Morbihan (X11=0.576, X10=0.646, X57=0.529) and at intermediate to low frequency in the South (X11=0.373, X10=0.620, X57=0.449) and West (X11=0.35, X10=0.675, X57=0.325). These trends were similar to those at Glu-5', a nuclear marker suspected to be affected by selection in the island (Gérard *et al.* 2015) and at candidate allozymes although with fewer sampling sites (Blot *et al.* 1989). Three genetically differentiated regions were identified with average foreign allele frequency of 0.165 in the North-East, 0.501 in the Gulf of Morbihan, and 0.335 in the South/West (Table S4). Similarly, the RdA site had the lowest frequency of the foreign allele at Glu-5' in the whole data set. Across all sites, the frequencies of the foreign allele at Glu-5' were significantly correlated with those at X10 ($r=0.61$, p-value < 0.001), X11 ($r=0.419$, p-value=0.012), and X57 ($r=0.49$, p-value=0.003), but they were globally higher at Glu-5' (Figure S4). The foreign allele frequency at those four loci is represented in Figure S4 and the average over the four loci in Figure 4B (filled symbols), and it clearly shows a genetic break between two geographically

close sites (PAF and RdA). The average frequency was the highest in the Gulf of Morbihan (from HdS to PAF), then it abruptly dropped down (in 40 km) between PAF and RdA (respectively on the West and East coast of the Prince of Wales' Peninsula), and finally slowly increased along the coast from North-East to South-West. This is in sharp contrast with the pattern observed at the other loci (open symbols) of which the average frequency remained similar across all sites. This suggests that the genetic break at the boundary of the Gulf of Morbihan and the North-East region is better revealed by the frequency of foreign alleles at ancestry-informative loci implying a role of admixture either in the maintenance or in the detection of the genetic structure.

Environment-associated genetic structure in the Kerguelen Islands

We then tested for genetic-environment association (GEA) in the Kerguelen, i.e. the correlation between genetic differentiation and environmental factors, independently of geographic structure. As such, we performed redundancy analysis on the 695 individual genotypes sampled from the 35 sites characterized by different habitats. Among the seven constrained factors (five qualitative variables, plus geographic coordinates), three were not significant in the initial model (Salinity, Exposure and Latitude, Table S8) and were removed of further analyses. The proportion of total genotypic variance explained by all constrained factors was highly significant in the global model (p-value=0.001, Table 2A, left panel), but quite low (2.32%). The first RDA axis, which explained 61% of the constrained variance, was mainly contributed by *Macrocystis* (Figure S5 and Table S9). Accordingly, it was the only factor whose marginal effect remained significant (p-value=0.032, Table 2B, left panel).

We statistically controlled for the effect of geography by performing a conditioned RDA analysis on Longitude (Table 2A, right panel). The combined effect of the three environmental variables remained significant (p-value=0.041), explaining 1.1% of the total genotypic variance. Individually, *Macrocystis* and Substrate still showed significant effects, after removing all other confounding factors (p-value=0.011 and 0.001, respectively, Table 2B, right panel). Interestingly, it has been previously shown that the fine-scale genetic variation at Glu-5' was also significantly correlated with *Macrocystis* (Gérard *et al.* 2015). In agreement, we found a significant correlation between the average foreign allele frequency at the four most differentiated loci in the Kerguelen Islands and the presence/absence of *Macrocystis*

(Figure 4C, left panel), whereas there was no correlation with the other loci (right panel). This suggests either that the environment constraints a moderate connectivity, or that adaptation may be polygenic and connectivity extensive at the scale of the island, such that outlier-based methods are not suitable in the Kerguelen (Le Corre & Kremer 2012). The sharp genetic break between RdA and PAF further indicates that two genetic backgrounds may have been locally trapped by an ecological boundary or a region of reduced dispersal (Bierne *et al.* 2011). Accordingly, these two sites differ at all five ecological variables (Table S3), and the water masses between the Gulf of Morbihan and the North Coast do not mix well (Karin Gerard, pers. comm.), suggesting that exchanges between the two sites are limited.

Most differentiated SNPs in the Kerguelen Islands are primarily ancestry-informative in the Northern Hemisphere

In the total sample, the average allele frequency of the foreign allele was 0.417 at Glu-5', 0.503 at X10, 0.619 at X11 and 0.480 at X57. These polymorphisms were surprisingly well-balanced in the island, despite being species-specific in the Northern species (Table S2, also see Gérard *et al.* 2015 for Glu-5'). To investigate whether local adaptation in the island was primarily depending on ancestry-informative loci in the Northern complex of species, we compared the degree of differentiation between sites in the Kerguelen Islands and that of the Northern species, *M. edulis* and *M. galloprovincialis*, at the 33 KASPar SNPs (Figure 5). Panel A shows that the level of genetic differentiation among sites in the Kerguelen (global FST, Table S2) was significantly higher (p-value=0.021) for the ancestry-informative loci (mean=0.015, in orange) compared to the control loci (mean=0.007, in grey). Importantly, the difference between the two categories was also significant when considering the genetic-by-environment association across all variables (Panel B: mean_orange=0.112, mean_grey=0.048, p-value=0.006), which was measured by the locus coordinates on the first axis of the conditioned RDA (Table S10); or only including *Macrocystis* (Table S11: mean_orange=0.089, mean_grey=0.041, p-value=0.016). Moreover, these patterns hold when adding the locus Glu-5' (Figure S6) in the case of genetic differentiation (Panel A, p-value=0.01), and GEAs (Panel B, p-value=0.004) measured by the FCT from an AMOVA analysis performed by grouping sites according to the presence/absence of *Macrocystis* (see Gérard *et al.* 2015).

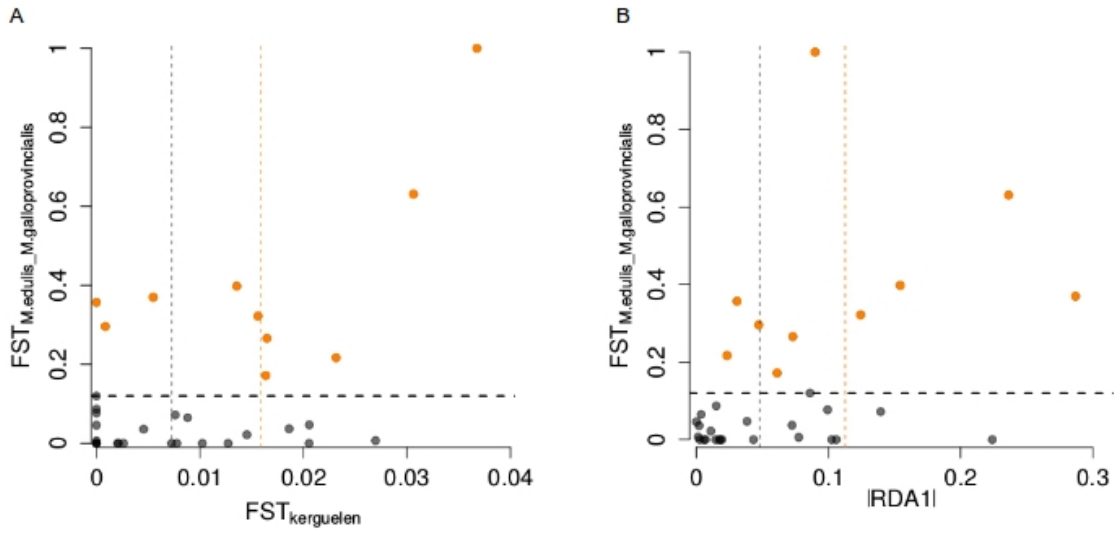


Figure 5. Correlation between the level of differentiation among the Kerguelen Islands (x-axis) and the Northern species (FST, y-axis) at each KASPar SNP. Panel (**A**) shows the genetic differentiation between Kerguelen populations (global FST); and panel (**B**) shows the genetic-by-environment association (locus coordinates on the first axis of the conditioned RDA (absolute values), see Table S10). Northern species are *M. edulis* (European population of the North Sea) and *M. galloprovincialis* (Mediterranean population of the West basin). Ancestry-informative loci, i.e., $FST_{M.edulis\_M.galloprovincialis} > 0.120$ (horizontal dashed line, see Table S2), are depicted in orange. Wilcoxon's test between the ancestry-informative loci (orange) and the controls (grey): (**A**) mean_orange=0.015, mean_grey=0.007 (p-value=0.021); (**B**) mean_orange=0.112, mean_grey=0.048 (p-value=0.006). Their respective means are depicted by vertical dashed lines. Pearson correlation coefficient: (**A**) r=0.540 (p-value=0.001); (**B**) r=0.439 (p-value=0.011).

Discussion

Gene trees are not species trees (Nichols 2001), and the primary cause in eukaryotes is thought to be incomplete lineage sorting between closely-related species (Mallet *et al.* 2016). Nevertheless, recent genomic studies, e.g., *Anopheles gambiae* mosquitoes (Fontaine *et al.* 2015), *Xiphophorus* fishes (Cui *et al.* 2013), African lake cichlids (Meier *et al.* 2017), Caribbean *Cyprinodon* pupfishes (Richards & Martin 2017) or *Heliconius* butterflies (Martin *et al.* 2013), recognized a prominent role of introgressive hybridization as a source of reticulate phylogenies. This is particularly true in species complexes, such as *Mytilus* mussels, in which incompletely isolated species with overlapping ranges commonly exchange genes via introgressive hybridization (Fraïsse *et al.* 2016).

Here we confirmed the reticulated history of evolution of the Southern Hemisphere Kerguelen mussels by analysing their genetic relationship with the three Northern species (*M. edulis*, *M. galloprovincialis* and *M. trossulus*) at 1269 contigs (51,878 SNPs). A genetic network of eleven samples of Northern populations and a Kerguelen sample (Figure 1) supports previous studies based on a handful of nuclear markers (Borsa *et al.* 2007) and mitochondrial DNA (Hilbish *et al.* 2000; Gérard *et al.* 2008) suggesting that the Kerguelen Islands belong to a Southern lineage which originated either from an ancestor of *M. edulis* and *M. galloprovincialis* (scenario 1) or from proto-*M. edulis* that migrated to the South (scenario 2). This Atlantic-Pleistocene scenario predicts that Southern mussels are closely related to species endemic to the Atlantic Ocean, i.e. *M. edulis* and *M. galloprovincialis*, and that the divergence between the two hemispheres is relatively recent (~0.5 to ~1.3 mya, Gérard *et al.* 2008). The maximum-likelihood population tree (Figure 2), which shows Kerguelen as a sister clade to *M. edulis*, reinforces scenario 2 in which the Kerguelen mussels differentiated from proto-*M. edulis* following a transequatorial migration to the South. The deep branching of the Southern mtDNA clade observed by Gérard *et al.* (2008) would therefore be explained either by ancestral polymorphism or more likely by introgression swamping at the time of contact between *M. edulis* and *M. galloprovincialis* (Smietanka *et al.* 2010). In addition, we further confirmed with our SNP assay previous findings (Gérard *et al.* 2008, Borsa *et al.* 2012) that the Kerguelen mussels and South American mussels from Chile clustered together (Figure S1), maybe due to a displacement of the polar front in the past which would have connected them (Holliday & Read 1998), or because of an anthropogenically-driven colonization of the Kerguelen from South America by maritime vessel transport. On the contrary, samples from Australasia were more related to *M. galloprovincialis* (Figure S1).

The population tree inference provides evidence of secondary genetic exchanges with the Northern Hemisphere that occurred after the first establishment in the Southern Hemisphere (Figure 2, arrows). The resulting genome-wide ancestry variation was estimated by applying a new topology weighting method to each GBS sequence (Martin & Van Belleghem 2016), which weighted the contribution of three topologies to the full tree (Figure 3). The majority of the genome (63 % of the classified loci, Table 1) shows clear evidence of admixture, i.e., the Kerguelen haplotypes were all (or part of) nested within a Northern clade. Most of the cases involved introgression from *M. edulis* (16 contigs), whereas *M. trossulus* and *M.*

galloprovincialis contributed to a lesser extent (4 contigs each), and their introgression was probably more ancient (as suggested by the clustering of the Kerguelen haplotypes within each of these two Northern clades). It is also possible that Australasian *M. planulatus* mussels that are related to *M. galloprovincialis* according to our SNP data, could have contributed to the reticulated history. Moreover, at some GBS loci, Kerguelen mussels possessed alleles characteristic of both *M. edulis* and *M. galloprovincialis* or *M. trossulus* indicating polymorphism for Northern species-specific alleles in the Kerguelen. Importantly, these loci did not depart from Hardy-Weinberg and linkage equilibrium as exemplified by an ADMIXTURE analysis (Figure S7) in which the Kerguelen mussels appeared as a well demarcated panmictic cluster. Their admixture with Northern genomes only became apparent with this type of method when considering two or three ancestral clusters (out of eight).

Genomic reticulation was already observed by Borsa *et al.* (2007) at the Glu-5' nuclear marker; and contrary to what is known in Northern hybrid zones (Bierne *et al.* 2003), there is no evidence of reproductive barriers impeding admixture in the Kerguelen Islands. Several hypotheses can be proposed: (i) a weaker reproductive barrier between Northern backgrounds at the time of contact in the south; or (ii) an insufficient barrier to gene flow under the demographic and environmental conditions, specifically strong genetic drift, high-potential for hybridization in this small isolated island, or a strong demographic asymmetry between the native and the introduced populations. The first hypothesis highlights the importance of Dobzhansky-Muller incompatibilities for reproductive isolation (Coyne & Orr 2004), and may explain how *M. edulis*, *M. galloprovincialis* and *M. trossulus* alleles at different loci can co-exist into a single Southern population that did not evolve their incompatible interactors, as opposed to the Northern populations. The second hypothesis highlights that the outcome of hybridization can be highly dependent on the demographic context.

Contrary to the large continental *Mytilus* populations we expect enhanced genetic drift in this small island (150 km East to West; 120 km North to South, that should be compared to a dispersal distance of 50 km per generation on average) and isolated island (4,100 km from South Africa; 4,000 km from Australia). By collecting mussels from contrasted habitats, we demonstrated significant differentiation across 35 sampling sites at the scale of the island, and at a smaller scale between geographically close sites, especially within the Gulf of Morbihan (Table S6). As found by Gérard *et al.* (2015) at Glu-5', and previously at allozyme loci (Blot

et al. 1989), the most significant structure was observed between the North-South coasts and the Gulf of Morbihan, with a genetic break between RdA and PAF at the three most differentiated loci (X10, X11, X57, Figure 4B and Figure S4). Altogether, these consistent genetic patterns across several physically unlinked loci indicate the possible existence of two genetic backgrounds maintained at the scale of the island. The foreign allele (as defined by its frequency in Northern reference populations) tended to be at higher frequency in shallow sites sheltered from the influence of open marine waters with a low salinity and flat-sandy bottoms, mainly in the inner part of the Gulf of Morbihan. These sites are characterized by an absence of *Macrocystis* kelp beds, as opposed to exposed rocky shores. Interestingly, *galloprovincialis* alleles are found more frequent in exposed, rather than sheltered sites in the hybrid zone between *M. edulis* and *M. galloprovincialis* in Europe, which would suggest inverted GEAs between hemispheres as predicted by the coupling hypothesis (Bierne *et al.* 2011). This hypothesis proposes that GEA can easily be revealed by intrinsically maintained genetic backgrounds in linkage disequilibrium with local adaptation genes, and that the phase of the disequilibrium can inverse when contacts are replicated as could have happened in the Southern Hemisphere mussels.

This fine-scale genetic structure in such a high-dispersal species as mussels is at odds with selective neutrality, so we explicitly tested the role of habitat heterogeneity in explaining this differentiation. Our RDA analysis (Table 2) shows that genetic variation was associated with habitats, even after controlling for spatial effects; and the most important factors were the presence of *Macrocystis* kelps, substrate type and slope. Despite being low, this significant habitat-driven genetic differentiation suggests a role of selection. Additionally, we observed a significant correlation between the presence/absence of *Macrocystis* and the average foreign allele frequency at the four most differentiated loci (Figure 4C, left panel). These findings reinforce the idea that adaptive genetic variation can be maintained at fine geographical scales in high-dispersal organisms, as recently shown in Chilean mussels (Araneda *et al.* 2016) or in passerine birds (Szulkin *et al.* 2016, Perrier *et al.* 2017). In these examples however the link with a possible history of admixture has not been investigated. Together with understanding the genomic architecture of adaptation, the relative importance of its source (new mutations, standing variation, or gene-flow) is a fundamental issue in evolutionary biology (Welch & Jiggins 2014; Lee & Coop 2017). Here, we show that the most differentiated SNPs in the

Kerguelen and those that most strongly drive the GEAs are primarily ancestry-informative (Figure 5), suggesting that adaptation to fine-scale environmental variations in the island may have been facilitated by secondary admixture and introgression of alleles from Northern species. These favorable alleles may have adaptively introgressed the Southern background in the Kerguelen, as it has been already reported at mac-1 between *M. edulis* and *M. galloprovincialis* along the French coast (Fraïsse *et al.* 2014) and at many other loci in the whole complex of species of the Northern Hemisphere (Fraïsse *et al.* 2016). However, the signal is probably erasing because of recombination between adaptive alleles and our neutral markers, and is also probably further blurred by genetic drift. A number of examples of adaptive introgression of complex traits have been documented in plants (e.g., resistance to drought in *Helianthus*, Vekemans 2010), and terrestrial animals (e.g., mimicry-determining loci in *Heliconius*, Heliconius Genome Consortium 2012; or insecticide resistance loci in *Anopheles*, Norris *et al.* 2015). Such adaptive variation could even serve as a source of genetic variation that subsequently became recombined into novel trait and favoured the emergence of new lineages, as proposed in cichlid fishes of Africa's Lake Victoria (Meier *et al.* 2017). A central question is whether admixture is a simple source of variation on which local selection can effectively act or if the initial linkage disequilibria between foreign alleles in the donor background are required for the successful emergence of micro-geographic adaptation (or speciation in the case of cichlids) and are maintained rather than built-on. In the case of Kerguelen mussels, the evidences we gained here for the maintenance of linkage disequilibrium are limited and indeed rather support extensive recombination. However our markers have likely lost too much signal to answer the question. Our results are very promising that a genome-wide survey in which the direct targets of selection will be identified should bring insightful information about the issue of adaptation from admixture-enhanced standing variation. For now, we can simply say that admixture between native and non-indigenous mussels has something to do with micro-geographic adaptation in the small isolated island of Kerguelen. Possibly these markers simply better reveals a genome-wide signal of habitat constrained connectivity (Gagnaire *et al.* 2015). But we cannot know yet if micro-geographic adaptation would have been possible in the absence of admixture.

Acknowledgements

Jean-Pierre Féral and Christian Marschal provided the eight Kerguelen specimens used for GBS during the PROTEKER campaign (Programme IPEV n°1044). Other Kerguelen samples (>600) were collected during scientific program IPEV-MACROBENTHOS n° 195 (1999-2003) by the technical volunteers (VATs) from IPEV missions 49–53 in Kerguelen. This work was funded by the Research Network GDR 3445 cnrs ifremer MarCo and the Agence Nationale de la Recherche (HYSEA project, ANR-12-BSV7-0011). This is article 2015-130 of Institut des Sciences de l'Evolution de Montpellier.

References

- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.*, 19, 1655–1664.
- Araneda C, Larraín MA, Hecht B, Narum S (2016) Adaptive genetic variation distinguishes Chilean blue mussels (*Mytilus chilensis*) from different marine environments. *Ecol. Evol.*, 6, 3632–3644.
- Arnold ML (2004) Transfer and origin of adaptations through natural hybridization: were Anderson and Stebbins right? *The Plant cell*, 16, 562–570.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2002) GENETIX 4.03, Logiciel Sous Windows™ Pour la Génétique Des Populations. Laboratoire Génome, Populations, Interactions: CNRS UMR. 5000. Université Montpellier 2, Montpellier, France.
- Benjamini Y, Hochberg Y (2000) On the Adaptive Control of the False Discovery Rate in Multiple Testing with Independent Statistics. *Journal of educational and Behavioral Statistics*, 25, 60–83.
- Bierne N, Bonhomme F, Arnaud-Haond S (2016) Editorial Dedicated population genomics for the silent world: the specific questions of marine population genetics. *Current Zool.*, 62, 545–550.
- Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Mol. Ecol.*, 20, 2044–2072.
- Bierne N (2010) The distinctive footprints of local hitchhiking in a varied environment and global hitchhiking in a subdivided population. *Evolution*, 64, 3254–3272.
- 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, 447–461.

- Blot M, Thiriot-Quiévreux C, Soyer J (1988) Genetic relationships among populations of *Mytilus desolationis* from Kerguelen, *M. edulis* from the North Atlantic and *M. galloprovincialis* from the Mediterranean. *Mar. Ecol. Prog. Ser.*, 44, 239–247.
- Blot M, Thiriot-Quivreux C, Soyer J (1989) Genetic differences and environments of mussel populations in Kerguelen Islands. *Polar Biol.*, 10, 167–174.
- Borsa P, Rolland V, Daguin-Thiébaut C (2012) Genetics and taxonomy of Chilean smooth-shelled mussels, *Mytilus spp.* (Bivalvia: Mytilidae). *Comptes rendus biologies*, *335*(1), 51–61.
- Borsa P, Daguin C, Bierne N (2007) Genomic reticulation indicates mixed ancestry in Southern-Hemisphere *Mytilus* spp. mussels. *Biol. J. Linn. Soc.*, 92, 747–754.
- Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.*, 81, 1084–1097.
- Cowen RK, Sponaugle S (2009) Larval Dispersal and Marine Population Connectivity. *Annu. Rev. Mar. Sci.*, 1, 443–466.
- Coyne JA, Allen Orr H (2004) Speciation. Sinauer Associates Incorporated.
- Cui R, Schumer M, Kruesi K, Walter R, Andolfatto P, Rosenthal GG (2013) Phylogenomics reveals extensive reticulate evolution in *Xiphophorus* fishes. *Evolution*, 67, 2166–2179.
- Felsenstein J, Churchill GA (1996) A Hidden Markov Model approach to variation among sites in rate of evolution. *Mol. Biol. Evol.*, 13, 93–104.
- Fontaine MC, Pease JB, Steele A *et al.* (2015) Mosquito genomics. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, 347, 1258524.
- Fraïsse C, Belkhir K, Welch JJ, Bierne N (2016) Local interspecies introgression is the main cause of extreme levels of intraspecific differentiation in mussels. *Mol. Ecol.*, 25, 269–286.
- Fraïsse C, Roux C, Welch JJ, Bierne N (2014) Gene-flow in a mosaic hybrid zone: is local introgression adaptive? *Genetics*, 197, 939–951.
- Gagnaire P-A, Broquet T, Aurelle D, Viard F, Souissi A, Bonhomme F, Arnaud-Haond S, Bierne N (2015) Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evol. App.*, 8, 769–786.
- Gaitán-Espitia JD, Quintero-Galvis JF, Mesas A, D'Elía G (2016) Mitogenomics of southern hemisphere blue mussels (Bivalvia: Pteriomorphia): Insights into the evolutionary characteristics of the *Mytilus* edulis complex. *Sci. Rep.*, 6, 26853.

- Gérard K, Bierne N, Borsa P, Chenuil A, Féral J-P (2008) Pleistocene separation of mitochondrial lineages of *Mytilus* spp. mussels from Northern and Southern Hemispheres and strong genetic differentiation among southern populations. *Mol. Phylogenet. Evol.*, 49, 84–91.
- Gérard K, Roby C, Bierne N, Borsa P, Féral JP, Chenuil A (2015) Does natural selection explain the fine scale genetic structure at the nuclear exon Glu-5′ in blue mussels from Kerguelen? *Ecol. Evol.*, 5, 1456–1473.
- Goudet J (2005) hierfstat, a package for r to compute and test hierarchical F-statistics. *Mol. Ecol. notes*, 5, 184–186.
- Hedrick PW (2013) Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol. Ecol.*, 22, 4606–4618.
- Heliconius Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487: 94–8.
- Hilbish TJ, Mullinax A, Dolven SI, Meyer A, Koehn RK, Rawson PD (2000) Origin of the antitropical distribution pattern in marine mussels (*Mytilus spp*.): routes and timing of transequatorial migration. *Mar. Biol.*, 136, 69–77.
- Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, Poss ML, Reed LK, Storfer A, Whitlock MC (2016) Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *Am. Nat.*, 188, 379–397.
- Holliday NP, Read JF (1998) Surface oceanic fronts between Africa and Antarctica. *Deep Sea Research Part I: Oceanographic Research Papers*, 45, 217–238.
- Koehn RK, Bayne BL, Moore MN, Siebenaller JF (1980) Salinity related physiological and genetic differences between populations of *Mytilus edulis*. *Biol. J. Linn. Soc.*, 14, 319–334.
- Le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Mol. Ecol.*, 21, 1548–1566.
- Lee KM, Coop G (2017) Distinguishing among modes of convergent adaptation using population genomic data. *Genetics*, doi:10.1534/genetics.117.300417.
- Lenormand T (2002) Gene flow and the limits to natural selection. *Trends Ecol. Evol.*, 17, 183–189.
- Mallet J, Besansky N, Hahn MW (2016) How reticulated are species? *BioEssays*, 38, 140–149.
- Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, Blaxter M, Manica A, Mallet J, Jiggins CD (2013) Genome-wide evidence for speciation with

gene flow in *Heliconius* butterflies. *Genome Res.*, 23, 1817–1828.

- Martin SH, Van Belleghem SM (2016) Exploring evolutionary relationships across the genome using topology weighting. *Genetics*, 206, 429-438.
- McDonald JH, Seed R, Koehn RK (1991) Allozymes and morphometric characters of three species of *Mytilus* in the Northern and Southern Hemispheres. *Marine Biol.*, 111, 323–333.
- Meier JI, Sousa VC, Marques DA, Marques DA, Selz OM, Wagner CE, Excoffier L, Seehausen O (2017) Demographic modelling with whole-genome data reveals parallel origin of similar *Pundamilia* cichlid species after hybridization. *Mol. Ecol.*, 26, 123–141.
- Nichols, R. (2001). Gene trees and species trees are not the same. *Trends Ecol. Evol.*, 16, 358–364.
- Norris LC, Main BJ, Lee Y, Collier TC, Fofana A, Cornel AJ, Lanzaro GC (2015) Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. *Proc. Natl. Acad. Sci.*, 112, 815–820.
- Oksanen J, *et al.* (2017) vegan: Community Ecology Package. R package version 2.4-2. https://CRAN.R-project.org/package=vegan
- Palumbi SR (1992) Marine speciation on a small planet. *Trends Ecol. Evol.*, 7, 114–118.
- Paradis E (2010) pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics,* 26: 419–420.
- Perrier C, Lozano del Campo A, Szulkin M, Demeyrier V, Gregoire A, Charmantier A (2017) Great tits and the city: distribution of genomic diversity and gene - environment associations along an urbanization gradient. *Evol. App.*, doi:10.1111/eva.12580
- Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.*, 8, e1002967.
- Rawson PD, Hilbish TJ (1995) Evolutionary relationships among the male and female mitochondrial DNA lineages in the *Mytilus edulis* species complex. *Mol. Biol. Evol.*, 12, 893–901.
- Rawson PD, Joyner KL, Meetze K, Hilbish TJ (1996) Evidence for intragenic recombination within a novel genetic marker that distinguishes mussels in the *Mytilus edulis* species complex. *Heredity*, 77 (Pt 6), 599–607.
- Reid NM, Jackson CE, Gilbert D, Minx P, Montague MJ, Hampton TH, Helfrich LW, King BL, Nacci DE, Aluru N, Karchner SI, Colbourne JK, Hahn ME, Shaw JR,

Oleksiak MF, Crawford DL, Warren WC, Whitehead A (2017) The landscape of extreme genomic variation in the highly adaptable Atlantic killifish. *Genome Biol. Evol.*, *9*, 00–00.

- Richards EJ, Martin CH (2017) Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes. *PLoS Genet.* 13(8): e1006919.
- Roux C, Fraïsse C, Castric V, Vekemans X, Pogson GH, 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. *J. Evol. Biol.*, 27, 1662–1675.
- Schmidt PS, Rand DM (1999) Intertidal microhabitat and selection at mpi: interlocus contrasts in the Northern acorn barnacle, *Semibalanus balanoides*. *Evolution*, 53, 135–146.
- Śmietanka B, Burzyński A, Wenne R (2010) Comparative genomics of marine mussels (*Mytilus spp.*) gender associated mtDNA: rapidly evolving atp8. *J. Mol. Evol.*, 71, 385–400.
- Skibinski DOF, Beardmore JA, Cross TF (1983) Aspects of the population genetics of *Mytilus* (Mytilidae; Mollusca) in the British Isles. *Biol. J. Linn. Soc.*, 19, 137–183.
- Smith SM, Maughan PJ (2015) SNP genotyping using KASPar assays. *Methods in Mol. Biol.*, 1245, 243–256.
- Swearer SE, Stephen E, Jeffrey S, Michael E, Simon R, Geoffrey P, Ross D, Steven G, Kimberly A, Gregory M, Robert R (2002) Evidence of self-recruitment in demersal marine populations. *Bull. Mar. Sci.*,70.1 251–271.
- Szulkin M, Gagnaire P-A, Bierne N, Charmantier A (2016) Population genomic footprints of fine-scale differentiation between habitats in Mediterranean blue tits. *Mol. Ecol.*, 25, 542–558.
- Väinölä R, Hvilsom MM (1991) Genetic divergence and a hybrid zone between Baltic and North Sea *Mytilus* populations (Mytilidae: Mollusca). *Biol. J. Linn. Soc.*, 43, 127–148.
- Van Belleghem SM, Rastas P, Papanicolaou A, Martin SH, Arias CF, Supple MA, Hanly JJ, Mallet J, Lewis JJ, Hines HM, Ruiz M, Salazar C, Linares M, Moreira GRP, Jiggins CD, Counterman BA, McMillan WO, Papa R (2017) Complex modular architecture around a simple toolkit of wing pattern genes. *Nat. Ecol. Evol.*, 1, 0052.
- Vekemans X (2010) What's good for you may be good for me: evidence for adaptive introgression of multiple traits in wild sunflower. *New Phytol.*, 187, 6–9.
- Wang G-D, Fan R-X, Zhai W, Liu F, Wang L, Zhong L, Wu H, Yang HC, Wu SF, Zhu

CL, Li Y, Gao Y, Ge RL, Wu CI, Zhang YP (2014) Genetic convergence in the adaptation of dogs and humans to the high-altitude environment of the tibetan plateau. *Genome Biol. Evol.*, 6, 2122–2128.

- Welch JJ, Jiggins CD (2014) Standing and flowing: the complex origins of adaptive variation. *Mol. Ecol.*, *23*, 3935–3937.
- Weir BS, Clark Cockerham C (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, 38, 1358.

Data Accessibility

File S1. Pairwise FST values between Northern species at the GBS SNPs. (A) FST between Med and Nor; (B) FST between Med and Tva; (C) FST between Nor and Tva; Nor: North Sea *M. edulis*; Med: West-Mediterranean *M. galloprovincialis*; Tva: Baltic Sea *M. trossulus*.

File S2. KASPar genotypes of each individual in the Kerguelen Islands (35 sampling sites), plus those of the additional individuals from other Southern Hemisphere populations (6 sampling sites).

File S3. Neighbour-joining trees of the 395 retained GBS sequences.

Dryad doi: 10.5061/dryad.6k740. DNA sequences and VCF files including GBS genotypes of each individual in the 12 GBS populations (11 Northern populations and eight Kerguelen mussels).

Author Contributions

Data acquisition: A. Haguenauer, A. Weber and K. Gérard.

Data analysis: C. Fraïsse, A. Chenuil and N. Bierne.

Writing: C. Fraïsse, A. Chenuil and N. Bierne.

Conceptualization: C. Fraïsse, A. Chenuil and N. Bierne.

Funding acquisition: A. Chenuil and N. Bierne.

Table 1. Counting the different topologies (395 contigs)

topology	categories	counts	
		weight >= 0.90	weight >= 0.75
A. *M. trossulus* with Kerguelen	A1 ancient Kerguelen divergence	5	7
	A2 introgression	3	4
	NA	1	8
	total	**9**	**19**
B. *M. edulis* with Kerguelen	B1 ancient Kerguelen divergence	3	5
	B2 introgression	6	16‡
	NA	4	17
	total	**13**	**38**
C. *M. galloprovincialis* with Kerguelen	C1 ancient Kerguelen divergence	2	2
	C2 introgression	4†	4†
	NA	0	4
	total	**6**	**10**
total		**28**	**67**

topology: three topologies have been weighted with *Twisst*.

categories: “ancient divergence”: the Kerguelen branched as a sister-clade to a Northern species; “introgression”: Kerguelen mussels were distributed within a Northern species; "na": topologies that we were unable to classify in these two categories. Categories including double introgression from *M. edulis* and *M. galloprovincialis*: † one contig, ‡ two contigs.

weight: exact weightings of each topology to the full tree.

Table 2. RDA analysis for the KASPar dataset (695 individuals, 33 SNPs)

	RDA (N=695)			
	variance	%variance	d.f.	p-value
A. Global effect				
Model	0.5288	2.324	7	0.001 ***
Residual	22.2183	97.676	687	-
B. Marginal effect				
Longitude	0.0128	0.0563	1	0.664
Substrate	0.0594	0.261	3	0.133
Macrocystis	0.0275	0.121	1	0.032 *
Slope	0.0379	0.167	2	0.217
Residual	10.9621	48.191	687	-

	conditioned RDA (N=695)			
	variance	%variance	d.f.	p-value
A. Environmental effect				
Substrate+*Macrocystis*+Slope	0.1213	1.0929	6	0.041 *
Residual	10.9621	98.7588	687	-
B. Individual effect				
Longitude	0.1424	0.626	1	0.001 ***
Substrate	0.2322	1.021	3	0.001 ***
Macrocystis	0.0691	0.304	1	0.011 *
Slope	0.0912	0.401	2	0.072 .
Residual	22.2183	97.676	687	-

Significance tests are shown on the left for the global model with nonsignificant terms removed (**A**). The marginal effect of each constraining variable (**B**) was tested through permutation tests by removing each term one by one from the model containing all other terms.

Conditioned RDA significance tests are shown on the right for the combined effect of environmental variables (**A**) after conditioning on Longitude; and for each term (**B**) after conditioning on other constraining variables to remove their confounding effects.

N: number of individual genotypes; **variance**: genotypic variance explained by each factor; **%variance**: percent of genotypic variance; **d.f.**: degrees of freedom; **p-value**: p-values obtained through 1000 permutations (* ≤ 0.05; ** ≤ 0.01; *** ≤ 0.001).