Fine-grained habitat-associated genetic connectivity 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, Université 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 genetic-environment association 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 structuring infra-specific genetic variation, and possibly 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. It includes three taxa that hybridize 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). The Kerguelen mussels are of particular interest to investigate the potential role of admixture in enhancing micro-geographic structure, 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 has been suspected. Here, we extended a previous analysis by using targeted-sequencing data (51,878 SNPs) across the three Northern species and the Kerguelen population, coupled with a panel of 33 SNPs genotyped on 695 mussels across 35 sites in the Kerguelen Islands. 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 structure. We first showed that the Kerguelen population belongs to a divergent Southern lineage, most related to *M. edulis* mussels, that experienced secondary admixture with non-indigenous Northern species. We then demonstrated that the Kerguelen mussels were significantly differentiated over small spatial distance, and that this local genetic structure was associated with environmental variation and mostly revealed by ancestry-informative markers. Although local adaptation can explain the association with the environment we believe it more likely that environment variables better describe population connectivity than geographic distance. Our study

highlights genetic connectivity of populations is more easily revealed by non-equilibrium secondary introgression clines at a subset of loci, while association with the environment should not be hastily advocated to support adaptation from admixture variation.

Introduction

Genetic divergence at a fine-grained spatial scale is an interesting situation in which neutral processes can more easily be disentangled from adaptive ones. Yet, micro-geographic adaptation is expected to be rare because theory shows that 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 some evidence that micro-geographic genetic-environment associations occur 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. These are expected in geographic regions where environmental gradients promote local adaptation (Schmidt *et al.* 2008), or where semi-permeable genetic backgrounds form a contact zone that couples with environmental variation (Bierne *et al.* 2011).

Understanding how these local genetic patterns can be produced by a complex history of reticulated evolution is crucial to uncover the origins of genetic variation (i.e., new mutations, standing variation, or gene-flow), and especially adaptive variation (Welch & Jiggins 2014; Lee & Coop 2017). 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 semi-isolated 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). In addition, introgression creates a departure from equilibrium situations in which the influx of heterospecific alleles from one genetic background generates a transient gradient in allele frequencies within the other genetic background, revealing cryptic connectivity patterns (Gagnaire *et al.* 2015). Importantly, the reduction in gene flow between the backgrounds is expected to be visible only on a subset of markers localised at an intermediate linkage map distance to reproductive isolation genes. This potential effect of introgression on our capacity to detect connectivity breaks in apparently well-mixed populations is of central concern to conservation and species management (Gagnaire *et al.* 2015). In addition, because secondary contact zones often coincide with environmental transitions (Bierne *et al.* 2011), gradient of introgression may easily be confounded with local adaptation signatures, especially when environmental variables better describe the pattern of connectivity than geographic distance.

Mytilus mussels are an excellent system to address these issues, because they are subdivided into partially reproductively isolated species. Moreover, a recent study based on F_{ST} genome scans and small-scale gene genealogies demonstrated that local introgression is widespread, and it is the primary cause of outlying levels of genetic differentiation between conspecific populations (Fraïsse *et al.* 2016). *Mytilus* mussels have 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 display hybrid zones where their ranges overlap (Skibinski *et al.* 1983; Väinölä & Hvilsom 1991; Bierne *et al.* 2003). In the South, a reevaluation of allozyme data and a review of the results obtained with mtDNA and two nuclear DNA markers (Borsa *et al.* 2012) encouraged to group Southern mussels in two different taxa, namely *M. platensis* for those related to *M. edulis* (the South American and Kerguelen mussels), and *M. planulatus* for those related to *M. galloprovincialis* (the Australasian mussels). The presence of a mitochondrial clade endemic to the Southern Ocean further suggests Southern mussels are native rather than introduced by human-mediated activities (Hilbish *et al.* 2000; Gaitán-Espitia *et al.* 2016). 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 an ancestral lineage that gave rise to *M. edulis* in the North and *M. platensis* in the South, and one from an ancestral lineage that produced *M. galloprovincialis* in the North and *M. planulatus* after migrating to the South, followed by mitochondrial swamping in Northern populations*;* or (ii) a single 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 Islands harbor *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). Further analyses with nuclear markers strengthened the view of a 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*). 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). This micro-geographical variation in allele frequency suggests that admixture with Northern mussels contributed to the pattern observed at Glu-5', although shared ancestry certainly affects a large part of the genome in these closely-related species. 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 (if any) 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. Accordingly, reproductive isolation genes have not been reported so far between mussels in the Kerguelen Islands.

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). Only 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, which serves as substrata and refuge for many molluscs species. As such, local adaptation was invoked to explain the fine-scale maintenance of polymorphism at Glu-5'. Because Glu-5' and candidate allozymes are strongly differentiated between Northern species (Skibinski *et al.* 1983, Rawson *et al.* 1996), we might suspect that adaptation in Kerguelen populations may have been facilitated by gene exchange with Northern Hemisphere lineages. 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

6

146 complex interpretations (Bierne *et al.* 2011, Gagnaire *et al.* 2015).

147 Here, we investigated whether reticulate evolution actually contributed to micro-geographic
148 structure in the Kerguelen islands, and if so, whether (i) admixture facilitated local adaptation
149 in the island, or (ii) eased our investigation of the connectivity patterns thanks to the detection
150 of two genetic backgrounds that coexist in the island and introgress. We used published
151 genotyping-by-sequencing (GBS) data of the three Northern species (Fraïsse *et al.* 2016) and
152 new GBS data of a sample from a single Kerguelen population to reconstruct their genetic
153 relationships, and investigate whether reticulated patterns found with a handful of markers
154 hold genome-wide. Past introgression events between Northern and Southern mussels were
155 robustly inferred (on top of high rates of incomplete lineage sorting) by testing for admixture
156 with genome-wide allele frequency data and reconstructing gene genealogies at a small
157 chromosomal scale. In addition, a new SNP dataset from thirty-five Kerguelen populations
158 was produced by genotyping mussels with a KASpar (kompetitive allele specific PCR) SNP
159 assay, which was enriched for ancestry-informative loci (i.e., loci that are more differentiated
160 than the genomic average between reference samples in the Northern Hemisphere). These
161 ancestry-informative loci enabled us to infer genetic structure at a micro-geographical scale in
162 the islands and connect the evolution of the Kerguelen mussels with the history of the *Mytilus*
163 species in the Northern Hemisphere. We found that the Kerguelen Islands harbor a divergent
164 Southern lineage of mussels that we propose to consider as the native lineage, which is more
165 related to *M. edulis*, and was subsequently admixed with non-indigenous Northern species.
166 We then confirmed a significant fine-scale genetic differentiation between sites associated
167 with environmental variables. Notably, we found that loci with a more pronounced genetic-
168 environment association (GEA) also tended to be among the most ancestry-informative
169 markers in the *Mytilus spp*. We discuss the importance of introgression from past admixture
170 events with Northern lineages and its role on the current and local genetic structure 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. Info. M&M 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 bacterial artificial chromosomes (BAC) and cDNA sequences (see Supp. Info. M&M for details). The Northern samples 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 *et al.* (2016) (see Supp. Info. M&M for details). The final dataset across the twelve localities consisted of 1269 reference sequences (378 BAC contigs that come from a pool of 224 unlinked clones, and 891 cDNA contigs that correspond to unlinked coding sequences of known-functions or randomly selected) 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-typed Kerguelen individuals to analyse the fine-scale genetic structure in the

8

194 island, and its relation to the local environment. Moreover, as we wanted to determine if
195 adaptation in the Kerguelen was primarily driven by standing variation in the Northern
196 complex of species (i.e. SNPs fixed between Northern species), the selected SNPs were not a
197 random sample of the SNPs detected by GBS, otherwise they would have been mainly private
198 polymorphisms to the Kerguelen (60% of the Kerguelen SNPs were found private). As such,
199 we further enriched our SNP array with ancestry-informative markers, the most differentiated
200 SNPs between pairs of Northern Hemisphere species (representing 33% of the non-private
201 Kerguelen SNPs, 10% of the whole SNP dataset), namely the West-Mediterranean *M.*
202 *galloprovincialis* population, the North-Sea *M. edulis* population and the Baltic-Sea *M.*
203 *trossulus* population. F_{ST} values (Weir & Cockerham 1984) were calculated using the R
204 package hierfstat (Goudet 2005) for each SNP between pairs of populations (Text S1). SNPs
205 in the upper 15% of the empirical F_{ST} distribution were categorized as highly-differentiated.
206 Any SNPs with more than 25% of missing data were discarded. Retained SNPs were further
207 filtered-out based on Illumina Assay Design Tool scores (available on Illumina web page,
208 http://support.illumina.com) which predicts probes success based on the number of
209 degenerated sites in the flanking sequences (250 bp on each side of the focal SNP). The final
210 array comprised 58 SNPs out of which 30 were highly differentiated between Northern
211 species (11 *M. trossulus*-specific, 8 *M. edulis*-specific and 10 *M. galloprovincialis*-specific,
212 Table S2).

213 **KASPar genotyping in the Kerguelen Islands**

214 We used samples collected from 35 sites in the Kerguelen Islands by Gérard *et al.* (2015),
215 totalling 695 individuals (Supp. Info. M&M and Table S3). Pieces of mantle tissue were
216 preserved in 95% ethanol, and DNA was extracted with the Macherey-Nagel NucleoSPin 96
217 Tissue kit. A KASPar (Kompetitive Allele Specific PCR) 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 F_{ST} values between the eight GBS-typed 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 F_{IS} 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 Text 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-typed 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 Akaike information criterion (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. We also performed a supervised clustering analysis on the Kerguelen individuals with the KASPar SNPs (K=2 clusters and 50 replicates). We defined *M. edulis* and the Chilean mussels as reference populations from which the Kerguelen individuals derive their ancestry. Individual ancestries are provided in Text S3 for the GBS analysis and Text S4 for the KASPar analysis.

In complement to the *TreeMix* analysis, we used a model-based approach implemented in ∂a∂i v1.6.3 (Gutenkunst *et al.* 2009) to explicitly test for the presence of gene flow between Kerguelen mussels and Northern species during their divergence history. It was assessed in a pairwise manner based on their folded joint site frequency spectrum at the cDNA contigs

11

(provided in Text S5 in ∂a∂i format, and plotted in Figure S1): Kerguelen vs. *M. edulis* (represented by the european sample "EU – external"); Kerguelen vs. *M. galloprovincialis* (mediterranean sample "MED – west"); Kerguelen vs. *M. trossulus* (european sample "EU"). We defined eight demographic models following previous studies (Tine *et al.* 2014; Christe *et al.* 2017) to test: (i) the timing of gene flow during divergence (absence of gene flow "SI", continuous migration "IM", secondary contact "SC", ancient migration "AM"); (ii) the genomic heterogeneity in gene flow (presence "2M" or absence of interspecific genomic barriers); (iii) the genomic heterogeneity in effective population size (presence "2N" or absence of Hill-Robertson effects). All models began with the split of the ancestral population in two daughter populations, and then were followed by divergence in the absence or presence of gene flow. Each model was fitted to the observed joint site frequency spectrum (singletons were masked) using three successive optimization steps: "hot" simulated annealing, "cold" simulated annealing and BFGS (Tine *et al.* 2014). Model comparisons were made using AIC. A summary of the models is given in Table S5 and the script that defines the models in ∂a∂i is given in Text S6.

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 Text S7 and their length are indicated in Table S6 (4.5 kb in average, a

12

290 minimum length of 1 kb and a maximum length of 25 kb).

291 For each haplotype locus, the relationships between the Northern species and the Kerguelen
292 population were then quantified using *Twisst* (Van Belleghem *et al.* 2017), a tree weighting
293 approach. We tested the three possible unrooted topologies: (A) *M. edulis* grouped with the
294 Kerguelen population; (B) *M. galloprovincialis* grouped with the Kerguelen population and
295 (C) *M. trossulus* grouped with the Kerguelen population. Their exact weightings to the full
296 tree were estimated by considering all subtrees ("complete method"). Only contigs with a
297 resolved topology were analysed: 67 contigs for which one topology had a weight greater or
298 equal to 0.75. These topologies were further classified in two categories depending on
299 whether they most plausibly reflect: (i) ancient divergence of the Kerguelen population (i.e.
300 the Kerguelen and Northern individuals clustered into distinct monophyletic groups) or, (ii)
301 introgression with one of the Northern species (i.e., the Kerguelen individuals were
302 distributed within one or more Northern clades); "na" stands for topologies that we were
303 unable to classify in these two categories due to a lack of informative sites. Tree topology
304 weightings and classification are available in Table S6.

305 **Analyses of genetic variation in the Kerguelen Islands**

306 For each KASPar SNP, estimation of F_{ST} values (Weir & Clark Cockerham 1984) was
307 calculated over all sampling sites (Table S2), and in a pairwise manner across all SNPs (Table
308 S76) using Genetix 4.05 (Belkhir *et al.* 2002). Their significance was tested by a permutation
309 procedure (1000 permutations) and adjusted with the Bonferroni's correction for multiple
310 comparisons (Benjamini & Hochberg 2000).

311 **Analysis of habitat variables in the Kerguelen Islands**

312 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). It performs a multiple linear regression between a matrix of response variables (individual genotypic data) and a matrix of explanatory variables (environmental factors). Notably, the effect of partially confounded explanatory variables can be estimated separately. RDA is commonly used to estimate the relative contribution of spatial and environmental components on species communities, and it has been recently applied to analysis of population genetic structure (e.g., Legendre & Fortin 2010). 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 ~

338 significant environmental variables + condition(significant geographic variables).

339 **Simulations of secondary contact**

340 Following the methodology of Gagnaire *et al.* 2015, we modelled a secondary contact
341 between two semi-isolated genetic background (*M. edulis* vs. *M. platensis* Chilean
342 population) that meet twice on a circular stepping stone model (between demes n°18 / n°19
343 and demes n°19 / n°20), and start to exchange genes. At generation zero, the *M. edulis*
344 background settles in deme n°19 while the *M. platensis* background is located everywhere
345 else. Their initial allele frequency at the barrier loci was set-up to 0.10 for the *M. edulis*
346 background and 0.70 for the *M. platensis* background, which correspond to the average
347 foreign allele frequency at the four more differentiated loci observed in the *M. edulis* and
348 Chilean *M. platensis* mussels, respectively (see Figure 4B). The auto-recruitment rate was set
349 to 1-*m*, and migration to adjacent demes was *m*/2 (with *m*=0.5). A barrier to dispersal was set
350 between demes n°18 and n°19 (*m*=0.05), which corresponds to the genetic break observed
351 between sites PAF and RdA in the Kerguelen Islands. Strong and asymmetric selection (s=0.5
352 against the *M. platensis* allele in the *M. edulis* background vs. s=0.2 against the *M. edulis*
353 allele in the *M. platensis* background) acts on bi-locus haploid genotypes at a barrier locus,
354 which is linked to a neutral marker located 1cM away and unlinked to a second neutral
355 marker. Deme size was constant and set to 500 individuals.

356 ## Results

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

359 An individual genetic network (Figure 1) was built from a subset of 51,878 high-quality SNPs
360 genotyped in eleven Northern populations and eight individuals from the Kerguelen Islands.

15

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 S2). The principal component analysis 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).

16

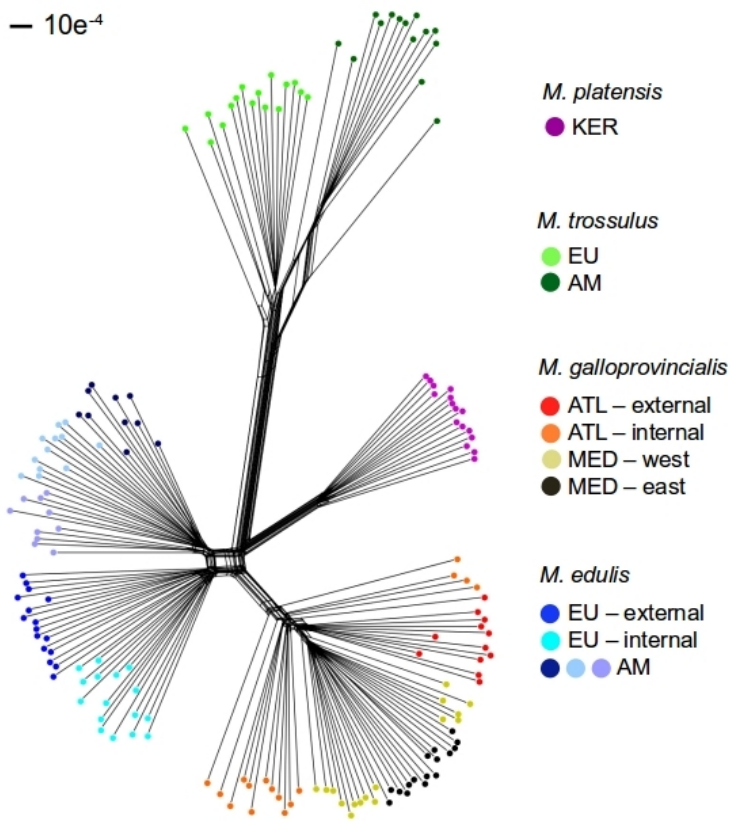


Figure 1. Genetic network of the Northern- and Southern-Hemisphere *Mytilus spp.* (twelve GBS-typed populations) produced with the neighbour-net method based on 51,878 high-quality GBS SNPs. The *M. platensis* sample is 'Baie de la Mouche' (KER, purple) in the Kerguelen Islands (Southern Ocean). *M. 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, three localities: dark blue, sky blue, slate 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 S3) 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 S3). The best fit to the data was obtained with seven migration events, which significantly improved the log-

17

likelihood of the model (Figure S3). 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 S8). The most robustly inferred migration event was between the Mediterranean *M. galloprovincialis* and the Kerguelen population (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 population and the American *M. trossulus* was additionally inferred for 38% of bootstrap replicates. Migration between the Kerguelen population and the European *M. edulis* was detected, but only in 4% of bootstrap replicates.

Reconstruction of the divergence history with δaδi consolidated evidence for ancient migration events between the Kerguelen mussels and each Northern species. In all three pairwise comparisons, the model of ancient migration with varying rates of introgression among loci ("*AM_2M*") received the strongest statistical support (Table S5). Migration occurred right after the split during a relatively short period (5% to 10% of the total divergence time) and it was asymmetric with a substantial fraction (90% to 95%) of the mussel genome in the Kerguelen permeable to *M. edulis* and *M. galloprovincialis* gene flow, while being mostly resistant to *M. trossulus*. Overall, these results suggest that the Kerguelen mussel is a Southern lineage related to *M. edulis* and that it secondarily admixed with all three Northern species (*M. edulis*, *M. galloprovincialis* and to a lesser extent with *M. trossulus*) in the past.

18

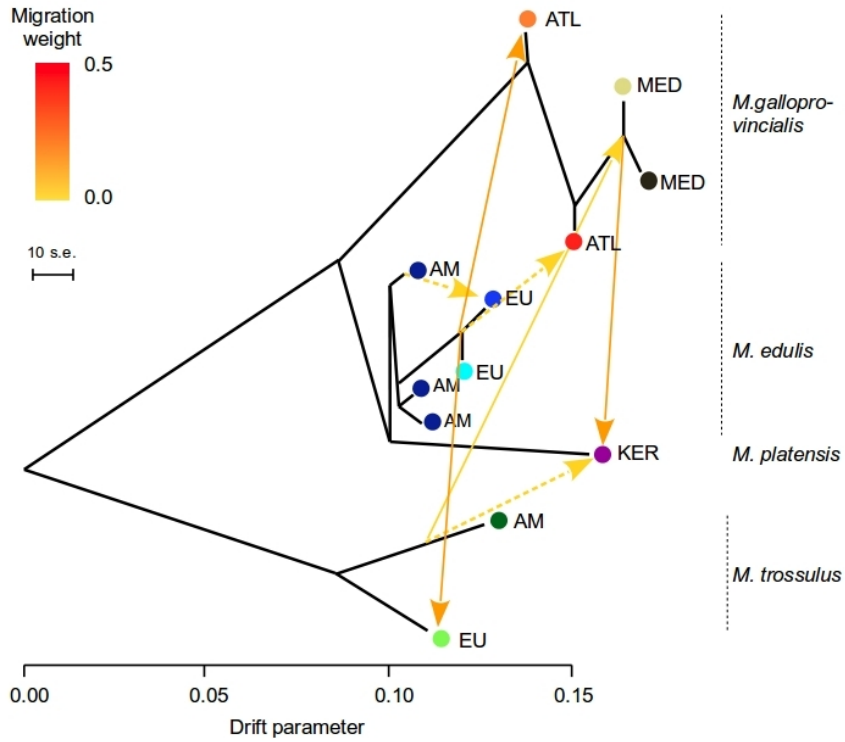


Figure 2. Maximum-likelihood population tree of the twelve GBS-typed *Mytilus* spp. populations inferred by *TreeMix* showing 7 migration events. ML estimation was made by using 32,162 GBS SNPs grouped in windows of 100 SNPs. Terminal nodes are labelled by locality abbreviation (Table S1) and colors 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 S8).

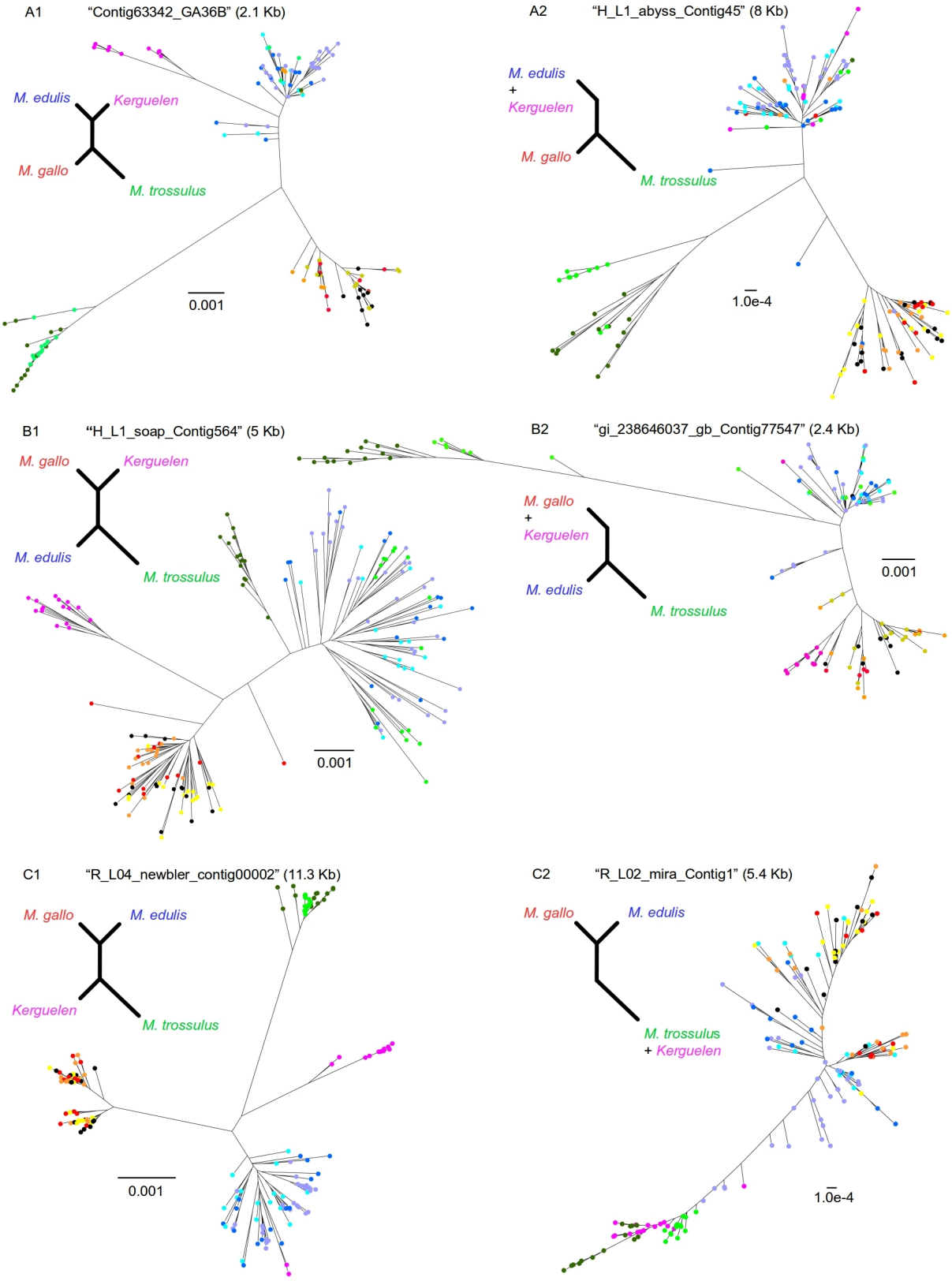
Variation of admixture histories across the genome

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% (67) of them showed resolved relationships, i.e. one of the unrooted topology weighted 75% or more, among which 40% (27) 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 (39 contigs) put the Kerguelen individuals together with *M. edulis*, while they were grouped with *M. trossulus* in 19 contigs (i.e. ancestral to the *M. edulis* / *M. galloprovincialis* subgroup) and with *M. galloprovincialis* in 9 contigs (Table 1).

19

When classifying the topologies in categories, 18 contigs supported the « ancient Kerguelen divergence » scenario while 19 supported an « introgression » scenario among which 4 were from *M. trossulus*, 12 from *M. edulis* and 3 from *M. galloprovincialis*; 30 contigs could not be classified. Figure 3 illustrates representative cases of the different *Twisst* categories, including candidate loci for introgression. Panel C2 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, a gene that has been previously suggested to be involved in adaptation in *M. edulis* (Bierne 2010). A similar pattern is shown on panel A2 where *M. edulis* haplotypes have totally replaced their Southern counterparts in the Kerguelen. Panel B2 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. 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 population 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% (Table 1).

20



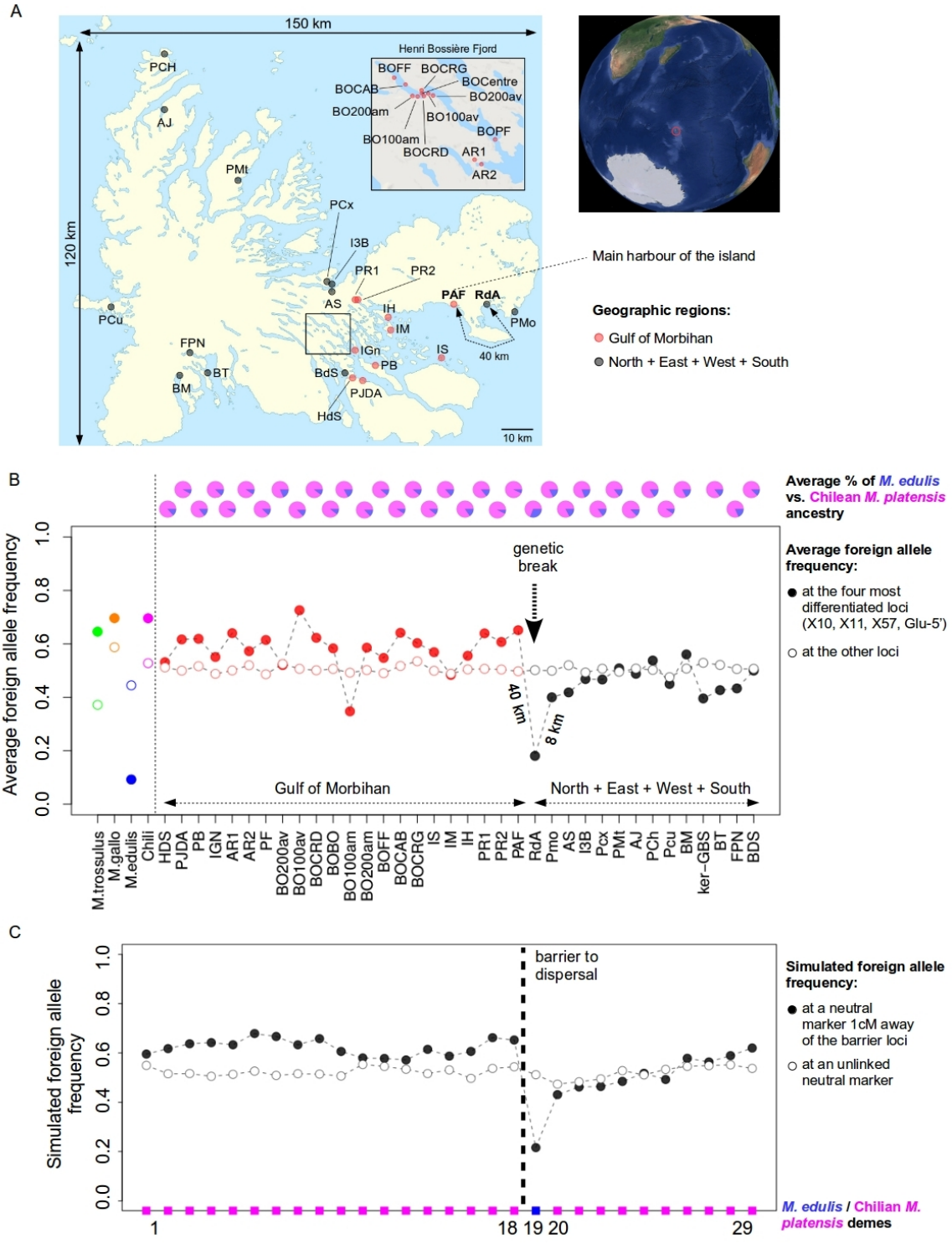
21

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 mussels 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". For each category, a typical neighbour-joining tree computed on the longest non-recombining block of the contig is shown (defined with the Difference of Sums of Squares method of McGuire & Wright 2000). Contig IDs correspond to Table S6, and sizes of non-recombining blocks are given in brackets. Colors match Figure 1.

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 F_{ST} values across all SNPs (Table S7) 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. Further, they indicate the existence of spatial heterogeneity in dispersal-driven connectivity at the scale of the island.

22



504 **Figure 4**. Geographic variation of the average foreign allele frequency across the four most
505 differentiated loci in the Kerguelen Islands (filled symbols, X10, X11, X57 and Glu-5' from
506 Gérard et al. (2015)) and the other loci (open symbols). Alleles were labelled based on their
507 frequencies in the *M. galloprovincialis* Atlantic population of Iberian Coast (Table S4). Red

23

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 (the main harbour of the island) and RdA are separated by 40 kms of coasts, RdA and Pmo by 8 kms, and they show a genetic break on Panel B. Sampling details are provided in Table S3. **B.** Frequency of the average foreign allele across sampling sites in the Kerguelen Islands (ordered by geography), in the three reference Northern species and in the Chilean mussels. See Figure S4 for the detailed pattern at each locus. On top, the average *M. edulis* vs. Chilean ancestries from an ADMIXTURE analysis (K=2) are shown for each site. **C.** Simulation results after 300 generations of a secondary contact between *M. edulis* and Chilean *M. platensis* mussels that meet twice on a circular stepping-stone model (between demes n°18 / n°19 and demes n°19 / n°20). A physical barrier to dispersal was modelled between the *M. edulis* deme (n°19) and the Chilean deme n°18 (dashed vertical bar, migration rate=0.05 instead of 0.5 as everywhere else). Deme size was fixed to 500 individuals, and the species barrier was asymmetric (at barrier loci, the selection coefficient against *M. edulis* in the *M. platensis* background was set to 0.2 while it was set to 0.5 against the *M. platensis* allele in the *M. edulis* background).

Global F_{ST} 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. In contrast, it was at higher frequency in the Gulf of Morbihan and at intermediate to low frequency in the South and West. These trends were similar to those at Glu-5' (Table S4), 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). 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),

24

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). These clearly show a genetic break between two geographically close sites, PAF and RdA (40 kms apart), and to a lesser extent between RdA and Pmo, which are separated by only 8 kms of coasts. 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 increased gradually 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.

An admixture analysis using all KASPar SNPs and defining *M. edulis* and the Chilean mussels as reference populations (Figure 4B, pie charts), suggests that the Kerguelen Island is occupied by mussels related to Chilean mussels (*M. platensis*), and that RdA has by far the highest level of *M. edulis* ancestry (69% compared to >81% elsewhere). We therefore hypothesize that two genetic backgrounds may be present in the island, one related to *M. edulis* and trapped at site RdA close to a potential density trough in the the Prince of Wales' Peninsula (as theoretically expected, Barton 1979, Barton & Hewitt 1985), and the other related to Chilean mussels and present everywhere else.

We illustrated this scenario by simulating a secondary contact between these two backgrounds (including a physical barrier to dispersal between demes n°18 and n°19), and tracking the frequency of the foreign allele at two neutral markers (linked and unlinked to the barrier locus) a few hundreds generations after the contact (Figure 4C). Simulations often fitted well with the observed variation in allele frequency across sites, as predicted by Gagnaire *et al.*'s model (2015), providing that the species barrier was asymmetric in order to protect the small *M. edulis* patch to be quickly swamped by *M. platensis* introgression. This

25

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 associations in the Kerguelen, i.e. for a correlation between genetic differentiation and environmental factors, independently of geographic structure. As such, we performed a redundancy analysis, RDA, (i.e., a multivariate constrained ordination) on the 695 individual genotypes sampled from the 35 sites characterized by different habitats, and estimated the relative contribution of each environmental factor on population genetic structure. Among the seven constrained factors (five qualitative variables, plus geographic coordinates), three were not significant in the initial model (Salinity, Exposure and Latitude, Table S9) and were removed from 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 S10). 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

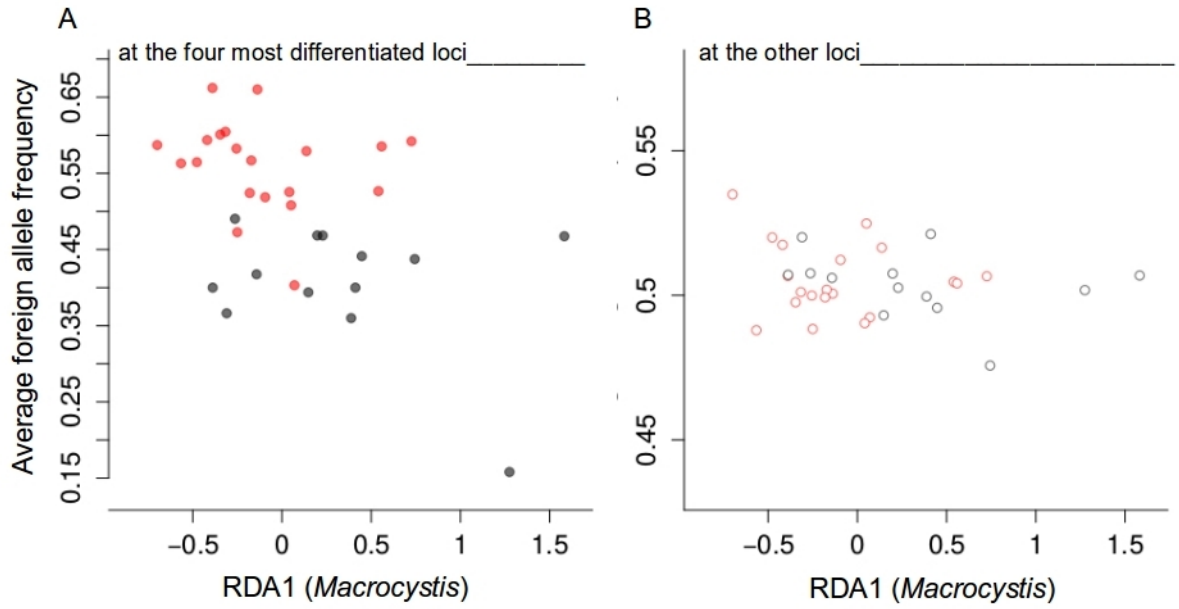
26

585 Glu-5' was also significantly correlated with *Macrocystis* (Gérard *et al.* 2015). In agreement,
586 we found a significant correlation between the average foreign allele frequency at the four
587 most differentiated loci in the Kerguelen Islands and the presence/absence of *Macrocystis*
588 (Figure 5A), whereas there was no correlation with the other loci (Figure 5B). This suggests
589 either that the environment constraints a moderate connectivity, or that adaptation may be
590 polygenic and connectivity extensive at the scale of the island, such that outlier-based
591 methods are not suitable in the Kerguelen (Le Corre & Kremer 2012).

592 The sharp genetic break between RdA and PAF further indicates that two genetic
593 backgrounds may have been locally trapped by an ecological boundary or a region of reduced
594 dispersal (Bierne *et al.* 2011). Accordingly, there is an oceanic threshold at the entrance of the
595 Gulf of Morbihan that impedes exchanges with water masses from outside; and at a larger
596 scale, the Antarctic circumpolar current moves the water masses from West to East causing
597 gyres and turbulences on the North-Eastern coast and pushing water masses far to the East
598 (Karin Gerard, pers. comm.). Thus, the water masses between the Gulf of Morbihan and the
599 North Coast do not mix well, suggesting that exchanges between the two sites are limited.
600 Moreover, these two sites differ at all five ecological variables (Table S3), but not in the
601 direction predicted: RdA shows an habitat characteristic of the Gulf of Morbihan while being
602 located on the East coast and having the lowest foreign allele frequency (and the reverse is
603 true for PAF). This imperfect correlation between genotypes and habitats suggests that
604 enhanced genetic drift and intense gene flow in the island grambled the signal at our markers.

605

27



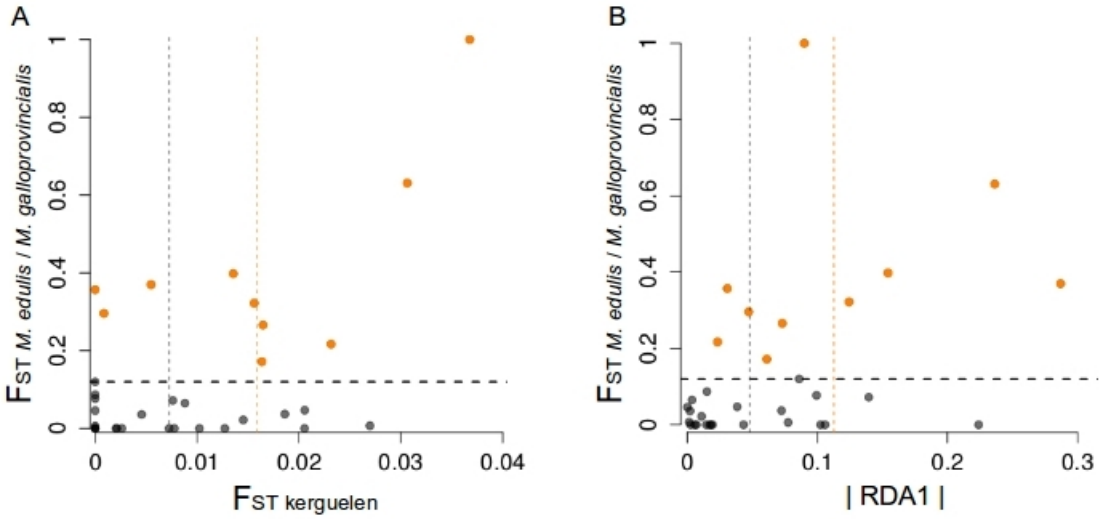
606 **Figure 5.** Correlation between the average foreign allele frequency of the four most
607 differentiated loci (A) and other loci (B) at each sampling site and the presence/absence of
608 *Macrocystis* (i.e., the average site coordinates on the first axis of the *Macrocystis* RDA, see
609 Table S121). Pearson correlation coefficient: on the left panel, 0.494 (p=0.003); on the right
610 panel, 0.107 (p=0.541). Red points indicate sites located in the Gulf of Morbihan (see Figure
611 4A).

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

614 In the total sample, the average allele frequency of the foreign allele was 0.417 at Glu-5',
615 0.503 at X10, 0.619 at X11 and 0.480 at X57. These polymorphisms were surprisingly well-
616 balanced in the island, despite being species-specific in the Northern species (Table S2, also
617 see Gérard *et al.* 2015 for Glu-5'). To investigate whether local adaptation in the island was
618 primarily depending on ancestry-informative loci in the Northern complex of species, we
619 compared the degree of differentiation between sites in the Kerguelen Islands and that of the
620 Northern species, *M. edulis* and *M. galloprovincialis*, at the 33 KASPar SNPs (Figure 6).
621 Panel A shows that the level of genetic differentiation among sites in the Kerguelen (global

28

622 F_{ST}, Table S2) was significantly higher (p-value=0.021) for the ancestry-informative loci
623 (mean=0.015, in orange) compared to the control loci (mean=0.007, in grey). Importantly, the
624 difference between the two categories was also significant when considering the genetic-by-
625 environment association across all variables (Panel B: mean_orange=0.112,
626 mean_grey=0.048, p-value=0.006), which was measured by the locus coordinates on the first
627 axis of the conditioned RDA (Table S11); or only including *Macrocystis* (Table S12:
628 mean_orange=0.089, mean_grey=0.041, p-value=0.016). Moreover, these patterns hold when
629 adding the locus Glu-5' (Figure S6) in the case of genetic differentiation (Panel A, p-
630 value=0.01), and genetic-environment associations (Panel B, p-value=0.004) measured by the
631 F_{CT} from an AMOVA analysis performed by grouping sites according to the presence/absence
632 of *Macrocystis* (see Gérard *et al.* 2015).



633 **Figure 6**. Correlation between the level of differentiation among the Kerguelen Islands (x-
634 axis) and the Northern species (F_{ST}, y-axis) at each KASPar SNP. Panel (**A**) shows the genetic
635 differentiation between Kerguelen populations (global F_{ST}); and panel (**B**) shows the genetic-
636 by-environment association (locus coordinates on the first axis of the conditioned RDA
637 (absolute values), see Table S11). Northern species are *M. edulis* (European population of the
638 North Sea) and *M. galloprovincialis* (Mediterranean population of the West basin). Ancestry-
639 informative loci, i.e., F_{ST} *M.edulis_M.galloprovincialis* > 0.120 (horizontal dashed line, see Table

29

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 that the reticulated evolution of the Southern Hemisphere Kerguelen mussels, which was suggested by a handful of nuclear markers (Borsa *et al.* 2007) and mitochondrial DNA (Hilbish *et al.* 2000; Gérard *et al.* 2008), holds at a genome-wide scale. We first analyzed their genetic relationship with the three Northern species (*M. edulis*, *M. galloprovincialis* and *M. trossulus*) at 1269 contigs (51,878 SNPs), and we demonstrated that mussels in the Kerguelen Islands belong to a Southern lineage. We further showed based on a maximum-likelihood population tree that the Kerguelen population is the sister clade of *M. edulis*, which suggests that Kerguelen mussels originated from an ancestor of *M. edulis* and *M. platensis* that migrated to the South. This Atlantic-Pleistocene scenario predicts that the divergence between mussel populations in the two hemispheres is relatively recent (~0.5 to

30

~1.3 mya, Gérard *et al.* 2008), and thus explains their large shared ancestry. Furthermore, 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).

Our population tree inference provides evidence of secondary genetic exchanges with Northern mussels that occurred after the first establishment in the Southern Hemisphere. This was confirmed by reconstructing their divergence history, which further suggested that introgression occurs at variable rates across the genome with some genomic regions resistant to gene flow (and carrying interspecific barriers) while others are essentially permeable. 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. The majority of the genome showed evidence of incomplete lineage sorting, with only 17% of the regions that have a resolved topology. However, most of these regions (51%) show 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,* whereas *M. trossulus* and *M. galloprovincialis* contributed to a lesser extent. 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.

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. Therefore, contrary to

31

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-Müller 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.

Finally, 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. 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. Although enhanced genetic drift is expected in this small (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), the fine-scale genetic structure observed in such a high-dispersal species as mussels is at first sight at odds with selective neutrality. So, we explicitly tested the role of habitat heterogeneity in explaining this differentiation. Our RDA analysis shows that genetic variation was associated with habitats, even after controlling for spatial effects; and the most important factors were the presence of

715 *Macrocystis* kelps, substrate type and slope. Despite being low, this significant habitat-driven
716 genetic differentiation could suggest a role of selection.

717 Firstly, it could be due to local adaptation of the mussels opposed by gene flow between
718 habitats. Accordingly, we observed a significant correlation between the presence/absence of
719 *Macrocystis* and the average foreign allele frequency at the four most differentiated loci. This
720 points toward a primary effect of *Macrocystis*, which is a keystone species in marine
721 ecosystems that forms kelp forest serving as substrata and refuge for many molluscs species,
722 including *Mytilus* (Adami & Gordillo, 1999), in areas exposed to wave action. Nevertheless,
723 the RdA site, which has the lowest foreign allele frequency, is not occupied by *Macrocystis*
724 kelp, weakening this local adaptation hypothesis.

725 Alternatively, the consistent genetic patterns observed across several physically unlinked
726 loci indicate the possible existence of two genetic backgrounds maintained at the scale of the
727 island. We propose, and illustrate by simulations, that a genetic background related to *M.*
728 *edulis* is trapped at RdA and surrounded by another genetic background related to Chilean *M.*
729 *platensis* mussels and present everywhere else. The enclosed location of the genetic
730 background at RdA explains that it is strongly introgressed at most markers and thus hard to
731 detect. However, the physical barrier to dispersal between sites PAF and RdA produces a clear
732 genetic break on the West side of the contact. Introgression between the two backgrounds
733 generates gradients in allele frequencies, which are better correlated with habitat variation
734 than geographical distance. The foreign allele (as defined by its frequency in the *M.*
735 *galloprovincialis* Atlantic population of Iberian Coast) tended to be at higher frequency in
736 shallow sites sheltered from the influence of open marine waters with a low salinity and flat-
737 sandy bottoms, mainly in the inner part of the Gulf of Morbihan. These sites are characterized
738 by an absence of *Macrocystis* kelp beds, as opposed to exposed rocky shores. Port aux
739 Français (PAF) is also the harbour where ships arrive and it is the best place for the arrival of

non-native genetic backgrounds. Interestingly, *M. 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 genetic-environment associations between hemispheres as predicted by the coupling hypothesis (Bierne *et al.* 2011). This hypothesis proposes that genetic-environment associations can easily be revealed by intrinsically maintained genetic backgrounds in linkage disequilibrium with local adaptation genes, and that the phase of the disequilibrium can reverse when contacts are replicated as could have happened in Southern Hemisphere mussels. Overall, these findings reinforce the idea that 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.

Although we had the hypothesis that secondary contact could be an explanation of the micro-structure observed in the Kerguelen (Gérard *et al.* 2015), we could not know which backgrounds were in interaction on the sole basis of the GBS data of a single sample. However, our procedure of identifying SNPs that were both polymorphic in the Kerguelen and highly differentiated between Northern Hemisphere species proved to be an interesting procedure to enrich for loci able to reveal the micro-geographic structure in the Kerguelen. Luckily the sample we used for the GBS analysis was localised in the introgression cline (sample "ker-GBS", Figure 4B), and this can also explain why the enrichment procedure was successful. This is exemplified with the genealogy around locus X10 (Figure S8), which shows that a SNP that differentiates *M. edulis* from other Northern species, and was polymorphic in the Kerguelen Islands, is able to reveal the cline of introgressed *M. edulis* allele we observed in the island.

In this work, we show that the most differentiated SNPs in the Kerguelen and those that

765 most strongly drive the genetic-environment associations are primarily ancestry - informative,
766 suggesting that maintenance of genetic differentiation at a small spatial scale, and possibly
767 adaptation to fine-scale environmental variations in the island, may have been facilitated by
768 secondary admixture and introgression of alleles from Northern species. These foreign alleles
769 may have adaptively introgressed the Southern background in the Kerguelen, as it has been
770 already reported at *mac-1* between *M. edulis* and *M. galloprovincialis* along the French coast
771 (Fraïsse *et al.* 2014) and at many other loci in the whole complex of species of the Northern
772 Hemisphere (Fraïsse *et al.* 2016). However, the signal is probably erasing because of
773 recombination between adaptive alleles and our neutral markers, and is also probably further
774 blurred by genetic drift. A number of examples of adaptive introgression of complex traits
775 have been documented in plants (e.g., resistance to drought in *Helianthus*, Vekemans 2010),
776 and terrestrial animals (e.g., mimicry-determining loci in *Heliconius*, Heliconius Genome
777 Consortium 2012; or insecticide resistance loci in *Anopheles*, Norris *et al.* 2015). Such
778 adaptive variation could even serve as a source of genetic variation that subsequently became
779 recombined into novel trait and favoured the emergence of new lineages, as proposed in
780 cichlid fishes of Africa's Lake Victoria (Meier *et al.* 2017).

781 A central question is whether admixture is a simple source of variation on which local
782 selection can effectively act or if the initial linkage disequilibria between foreign alleles in the
783 donor background are required for the successful emergence of micro-geographic adaptation
784 (or speciation in the case of cichlids) and are maintained rather than built-on. In the case of
785 Kerguelen mussels, the evidences we gained here for the maintenance of linkage
786 disequilibrium are limited and indeed rather support extensive recombination. However our
787 markers have likely lost too much signal to answer the question. Our results are very
788 promising that a genome-wide survey in which the direct targets of selection will be identified
789 should bring insightful information about the issue of adaptation from admixture-enhanced

790 standing variation. For now, we can simply say that admixture between native and non-
791 indigenous mussels has something to do with the enhancement of a micro-geographic
792 structure in the small isolated island of Kerguelen. Maybe local adaptation is operating at loci
793 linked to the candidate SNPs, but most probably these markers simply better reveal a genome-
794 wide signal of habitat constrained connectivity (Gagnaire *et al.* 2015).

795 ## Acknowledgements

796 Jean-Pierre Féral and Christian Marschal provided the eight Kerguelen specimens used for
797 GBS during the PROTEKER campaign (Programme IPEV n°1044). Other Kerguelen samples
798 (>600) were collected during scientific program IPEV-MACROBENTHOS n° 195 (1999-
799 2003) by the technical volunteers (VATs) from IPEV missions 49–53 in Kerguelen. This work
800 was funded by the Research Network GDR 3445 cnrs ifremer MarCo and the Agence
801 Nationale de la Recherche (HYSEA project, ANR-12-BSV7-0011). This is article 2015-130 of
802 Institut des Sciences de l'Evolution de Montpellier.

803 ## References

804 • Adami ML, Gordillo S (1999) Structure and dynamics of the biota associated with
805 *Macrocystis pyrifera (*Phaeophyta) from the Beagle Channel, Tierra del Fuego. *Sci.*
806 *Mar.*, 63(S1), 183–191.

807 • Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry
808 in unrelated individuals. *Genome Res.*, 19, 1655–1664.

809 • Araneda C, Larraín MA, Hecht B, Narum S (2016) Adaptive genetic variation
810 distinguishes Chilean blue mussels (*Mytilus chilensis*) from different marine
811 environments. *Ecol. Evol.*, 6, 3632–3644.

812 • Arnold ML (2004) Transfer and origin of adaptations through natural hybridization:
813 were Anderson and Stebbins right? *The Plant cell*, 16, 562–570.

- Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annu. Rev. Ecol. Evol. Syst.*, 16(1), 113–148.
- Barton NH (1979) The dynamics of hybrid zones. *Heredity*, 43(3), 341.
- 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, 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.

- Christe C, Stölting KN, Paris M, Fraïsse C, Bierne N, Lexer C (2017) Adaptive evolution and segregating load contribute to the genomic landscape of divergence in two tree species connected by episodic gene flow. *Mol Ecol*, 26, 59–76.
- 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, 769–78
- 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.
- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. PLOS Genetics 5(10), e1000695.
- 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*, 207(4), 1591–1619.

- Legendre P, Fortin MJ (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Mol. Ecol. Resour.*, 10(5), 831–844.
- 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.
- McGuire G, Wright F (2000) TOPAL 2.0: Improved Detection of Mosaic Sequences within Multiple Alignments. *Bioinformatics*, 16(2), 130–134.
- 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(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(3), 659–676.
- 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.
- Schmidt PS, Serrao EA, Pearson GA, Riginos C, Rawson PD, Hilbish TJ, Brawley

960 SH, Trussell GC, Carrington E, Wethey DS, Grahame JW, Bonhomme F, Rand DM
961 (2008). Ecological genetics in the North Atlantic: environmental gradients and
962 adaptation at specific loci. *Ecology*, 89, 91–107.

963 • Śmietanka B, Burzyński A, Wenne R (2010) Comparative genomics of marine mussels
964 (*Mytilus spp.*) gender associated mtDNA: rapidly evolving atp8. *J. Mol. Evol.*, 71,
965 385–400.

966 • Skibinski DOF, Beardmore JA, Cross TF (1983) Aspects of the population genetics of
967 *Mytilus* (Mytilidae; Mollusca) in the British Isles. *Biol. J. Linn. Soc.*, 19, 137–183.

968 • Smith SM, Maughan PJ (2015) SNP genotyping using KASPar assays. *Methods in*
969 *Mol. Biol.*, 1245, 243–256.

970 • Szulkin M, Gagnaire P-A, Bierne N, Charmantier A (2016) Population genomic
971 footprints of fine-scale differentiation between habitats in Mediterranean blue tits.
972 *Mol. Ecol.*, 25, 542–558.

973 • Tine M, Kuhl H, Gagnaire P-A, Louro B, Desmarais E, Martins RST, Hecht J, Knaust
974 F, Belkhir K, Klages S, Dieterich R, Stueber K, Piferrer F, Guinand B, Bierne N,
975 Volckaert FAM, Bargelloni L, Power DM, Bonhomme F, Canario AVM, Reinhardt R
976 (2014) European sea bass genome and its variation provide insights into adaptation to
977 euryhalinity and speciation. *Nat. Comm.*, 5, 1–10.

978 • Väinölä R, Hvilsom MM (1991) Genetic divergence and a hybrid zone between Baltic
979 and North Sea *Mytilus* populations (Mytilidae: Mollusca). *Biol. J. Linn. Soc.*, 43, 127–
980 148.

981 • Van Belleghem SM, Rastas P, Papanicolaou A, Martin SH, Arias CF, Supple MA,
982 Hanly JJ, Mallet J, Lewis JJ, Hines HM, Ruiz M, Salazar C, Linares M, Moreira GRP,
983 Jiggins CD, Counterman BA, McMillan WO, Papa R (2017) Complex modular
984 architecture around a simple toolkit of wing pattern genes. *Nat. Ecol. Evol.*, 1, 0052.

985 • Vekemans X (2010) What's good for you may be good for me: evidence for adaptive
986 introgression of multiple traits in wild sunflower. *New Phytol.*, 187, 6–9.

987 • Wang G-D, Fan R-X, Zhai W, Liu F, Wang L, Zhong L, Wu H, Yang HC, Wu SF, Zhu
988 CL, Li Y, Gao Y, Ge RL, Wu CI, Zhang YP (2014) Genetic convergence in the
989 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

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

Text 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).

Text S3. Individual ancestries estimated with ADMIXTURE for the GBS samples. (A) K=2; (B) K=3; (C) K=4; (D) K=5; (E) K=6; (F) K=7; (G) K=8.

Text S4. Individual ancestries estimated with ADMIXTURE for the KASPar samples in the Kerguelen (K=2; defining reference populations as *M. edulis* and Chilean mussels).

Text S5. Joint site frequency spectrum (in ∂a∂i format) between the Kerguelen mussels and (A) M. edulis; (B) M. galloprovincialis; (C) M. trossulus.

Text S6. Definition of the eight models of divergence used in our inferences with ∂a∂i.

Text S7. 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 twelve GBS-typed populations (eleven Northern populations and eight Kerguelen mussels).

Supplementary Information

Supplementary Information M&M

Supplementary Information Tables (S1 – S12)

Supplementary Information Figures (S1 – S8)

Supplementary Information Texts (S1 – S7).

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.

44

Tables

Table 1. Counting the different topologies

topology	categories	counts	
		weight >= 0.90	weight >= 0.75
A. *M. edulis* with Kerguelen	A1 ancient Kerguelen divergence	3	5
	A2 introgression	5	12†
	NA	5	22
	total	**13**	**39**
B. *M. galloprovincialis* with Kerguelen	B1 ancient Kerguelen divergence	2	2
	B2 introgression	3	3
	NA	0	4
	TOTAL	**5**	**9**
C. *trossulus* with Kerguelen	C1 ancient Kerguelen divergence	5	11
	C2 introgression	3	4
	NA	1	4
	TOTAL	**9**	**19**
TOTAL		**27**	**67**

topology: three topologies have been weighted with *Twisst*; **categories**: “ancient divergence”: the Kerguelen population 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. **weight**: exact weightings of each topology to the full tree.

45

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

	RDA (N=695)					conditioned RDA (N=695)			
	variance	%variance	d.f.	p-value		variance	%variance	d.f.	p-value
A. Global effect					**A. Environmental effect**				
Model	0.5288	2.324	7	0.001 ***	Substrate+*Macrocystis*+Slope	0.1213	1.0929	6	0.041 *
Residual	22.2183	97.676	687	-	Residual	10.9621	98.7588	687	-
B. Marginal effect					**B. Individual effect**				
Longitude	0.0128	0.0563	1	0.664	Longitude	0.1424	0.626	1	0.001 ***
Substrate	0.0594	0.261	3	0.133	Substrate	0.2322	1.021	3	0.001 ***
Macrocystis	0.0275	0.121	1	0.032 *	*Macrocystis*	0.0691	0.304	1	0.011 *
Slope	0.0379	0.167	2	0.217	Slope	0.0912	0.401	2	0.072 .
Residual	10.9621	48.191	687	-	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).

46