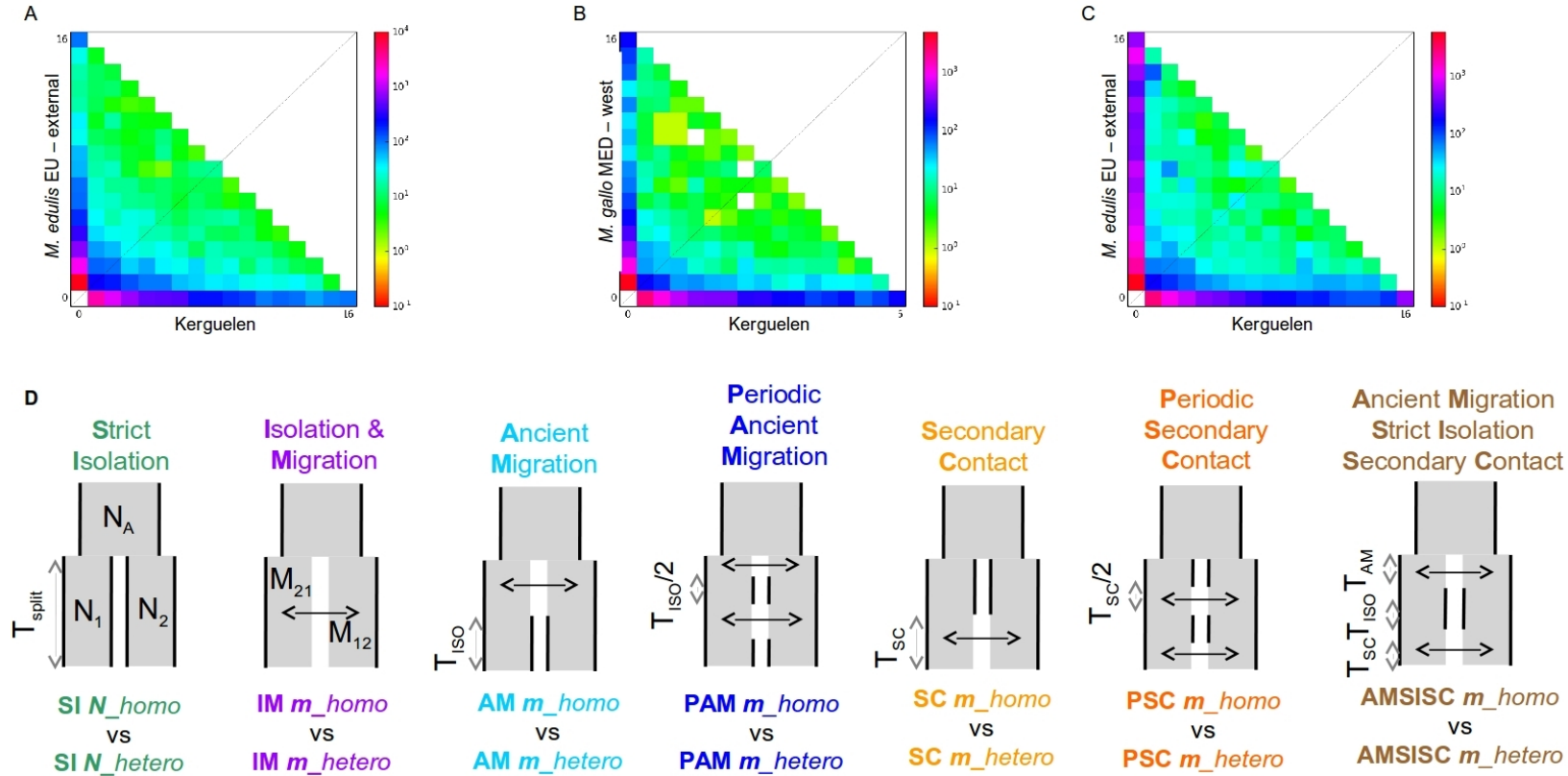
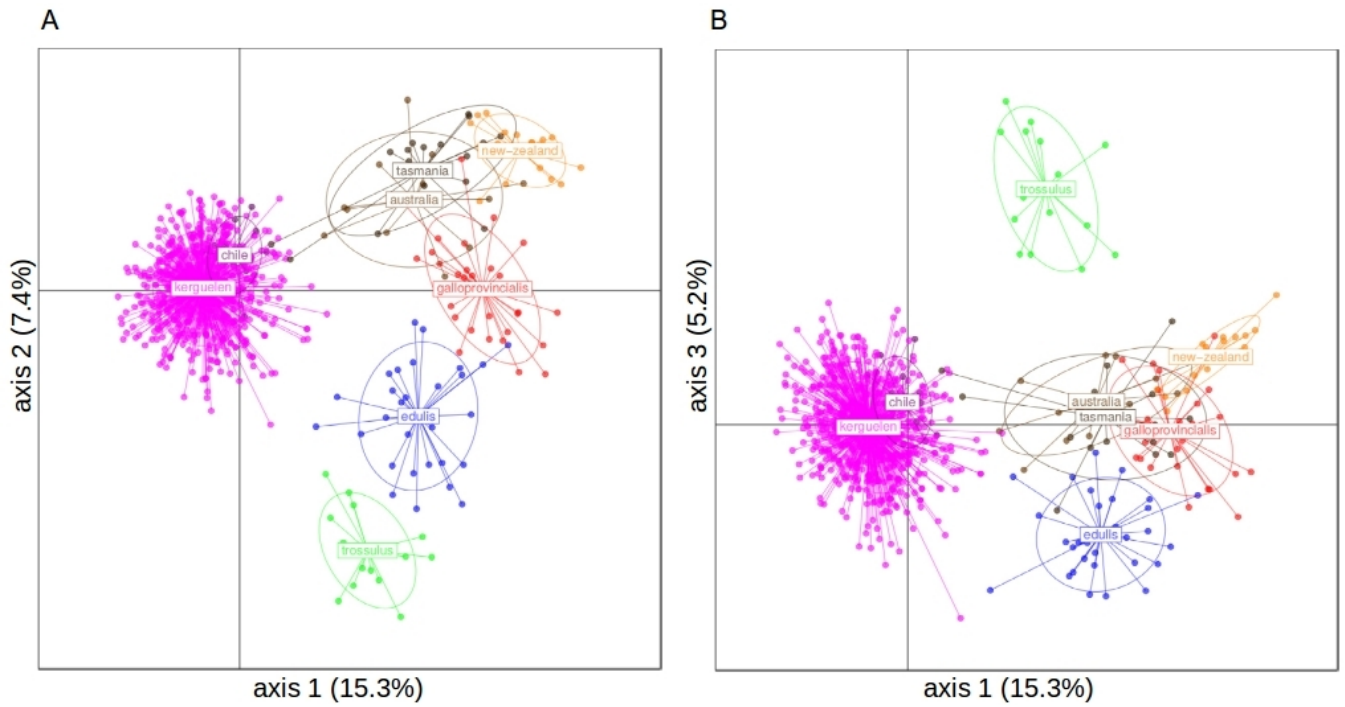


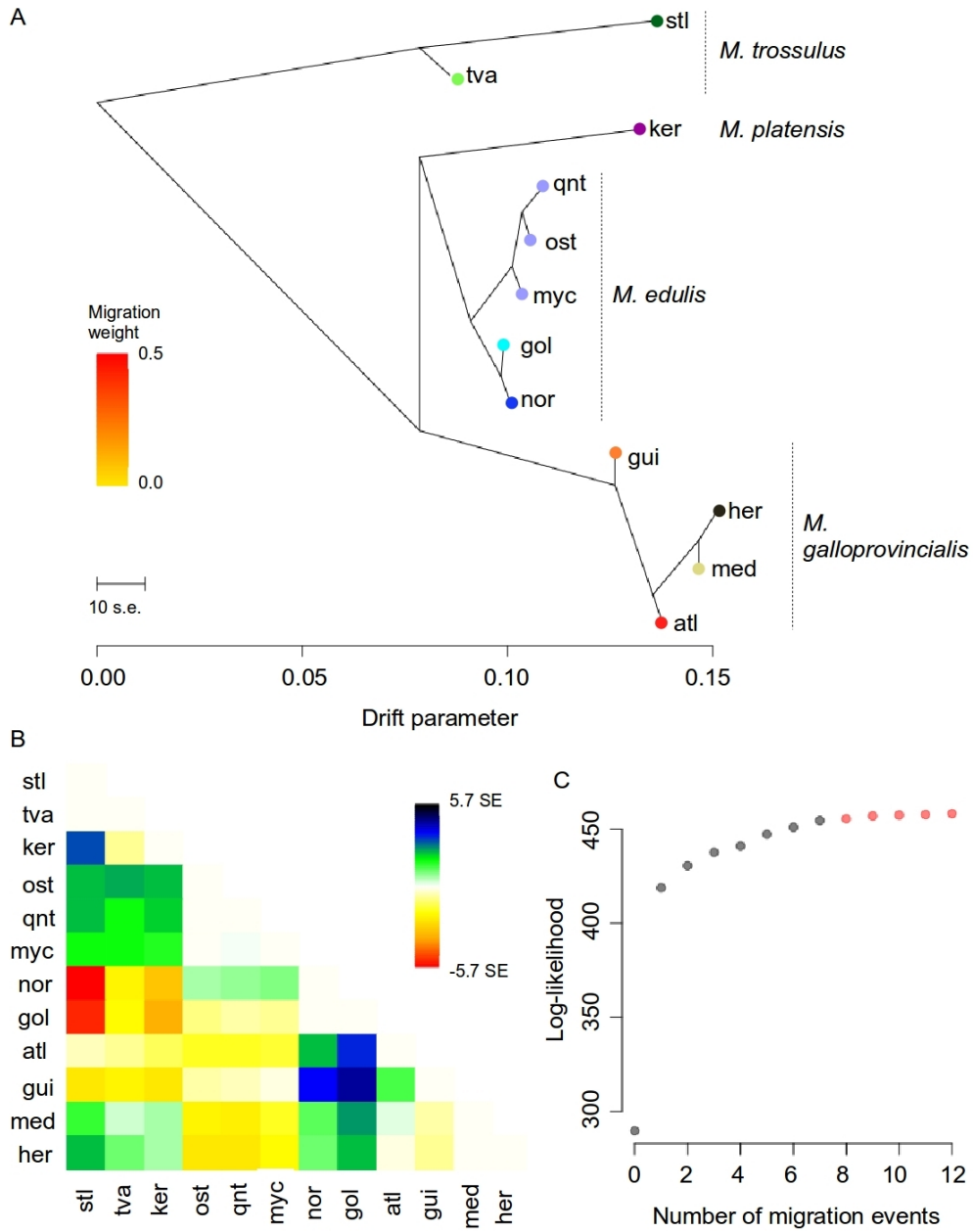
## Supplementary Figures



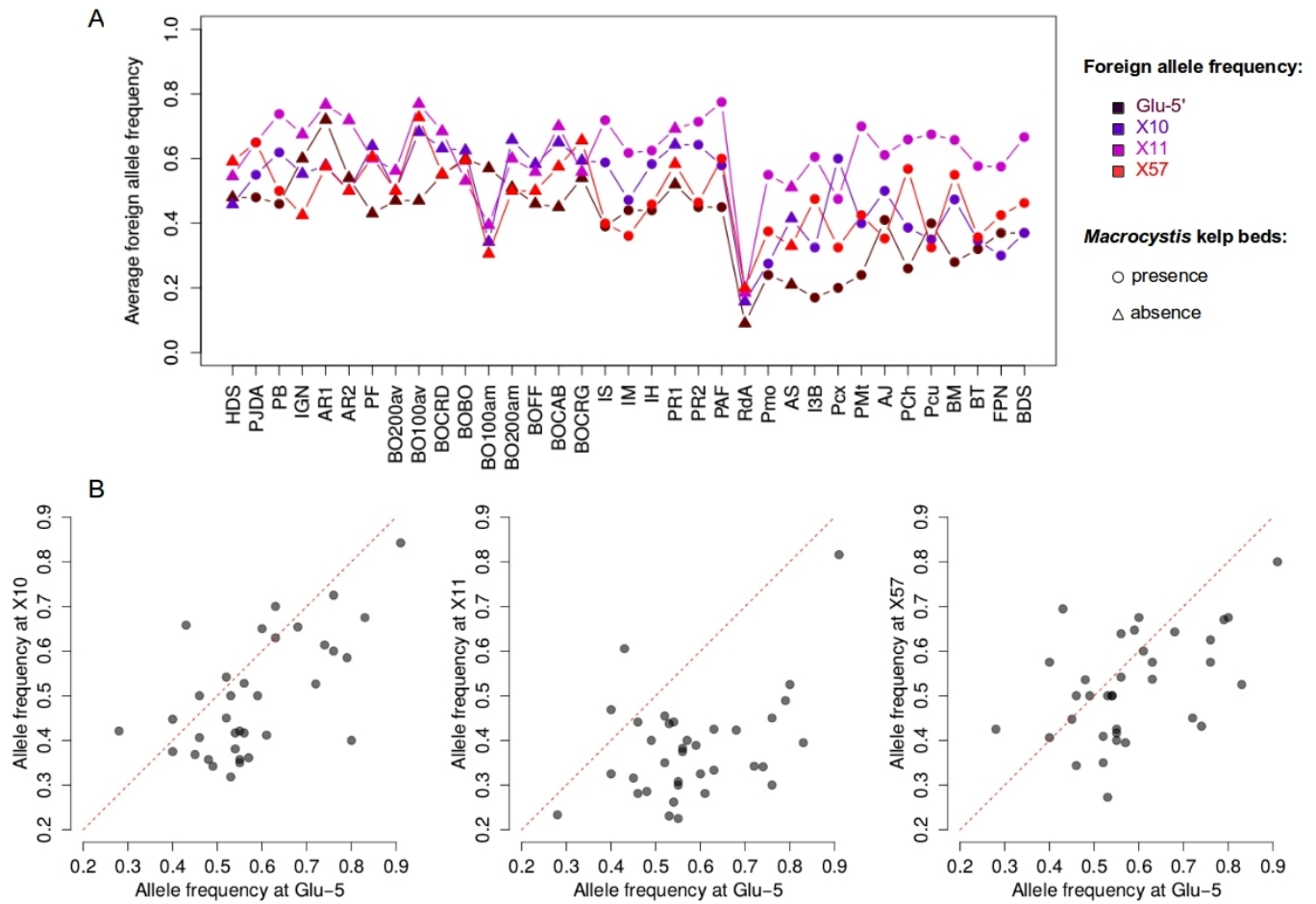
**Figure S1.** Observed folded joint site frequency spectrum between the Kerguelen mussels and Northern mussels based on the GBS dataset. The color scale indicates the number of SNPs falling in each bin. (A) *Mytilus edulis* (european sample “EU – external”, 24,524 SNPs). (B) *Mytilus galloprovincialis* (mediterranean sample “MED – west”, 15,920 SNPs). (C) *Mytilus trossulus* (european sample “EU”, 28,090 SNPs). (D) The fourteen demographic models implemented in  $\partial a \partial i$ . Each model with gene flow (IM, AM, PAM, SC, PSC, AMSISC) is shown in two versions: genomic homogeneity ( $m_{homo}$ ) or heterogeneity ( $m_{hetero}$ ) in the migration rate. The Strict Isolation model (SI) is also shown in two versions: genomic homogeneity ( $N_{homo}$ ) or heterogeneity ( $N_{hetero}$ ) in the effective population size.



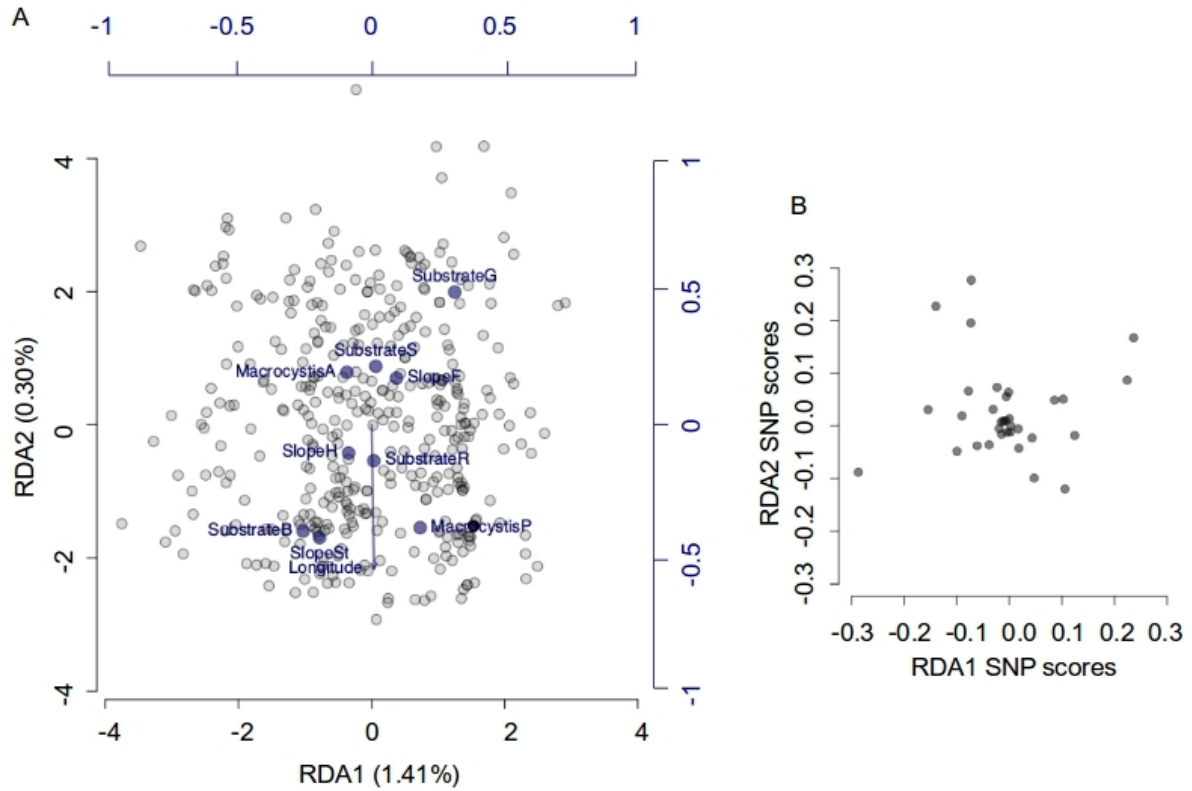
**Figure S2.** Principal component analysis based on the matrix of genotypes at 33 KASPar SNPs. Projections of samples on the plane defined by the first two axes (**A**) and by the first and the third axes (**B**) are shown. The proportion of genotypic variance explained by each axis is given in brackets. Samples from the Kerguelen Islands (in purple) are projected together with Northern Hemisphere samples (*Mytilus trossulus* in green; *Mytilus edulis* in blue; *Mytilus galloprovincialis* in red) and other Southern Hemisphere samples (Chile in dark purple; Tasmania in dark brown; Australia in brown and New-Zealand in orange).



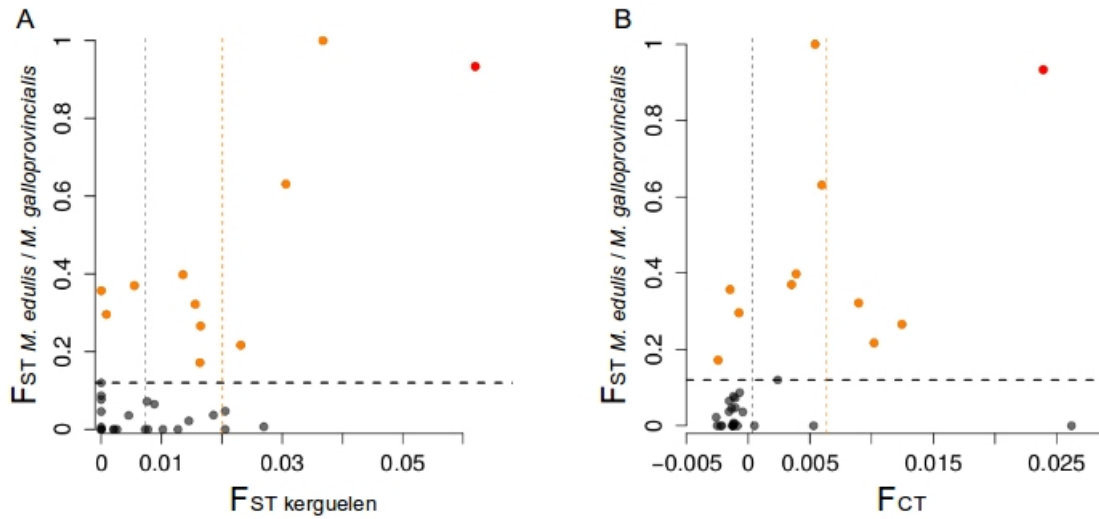
**Figure S3.** **A.** Maximum-likelihood tree without admixture events generated by *TreeMix* based on the GBS dataset. Terminal nodes are labelled by locality abbreviation (Table S1) and colors match Figure 1. The drift parameter is shown on the x-axis. Horizontal branch lengths are proportional to the amount of genetic drift that has occurred in each branch. **B.** Pairwise population residual fit. Residuals above zero depict populations that are more closely related to each other in the data than in the maximum-likelihood tree, and thus are indicative of admixture events. **C.** Increase in log-likelihood of the population tree model as a function of the number of migration events (from  $m=0$  to  $m=12$ ) in *TreeMix*. Red points indicate the migration events for which stepwise comparisons of log-likelihood become non-significant (i.e., the difference in Akaike information criterion is below 2).



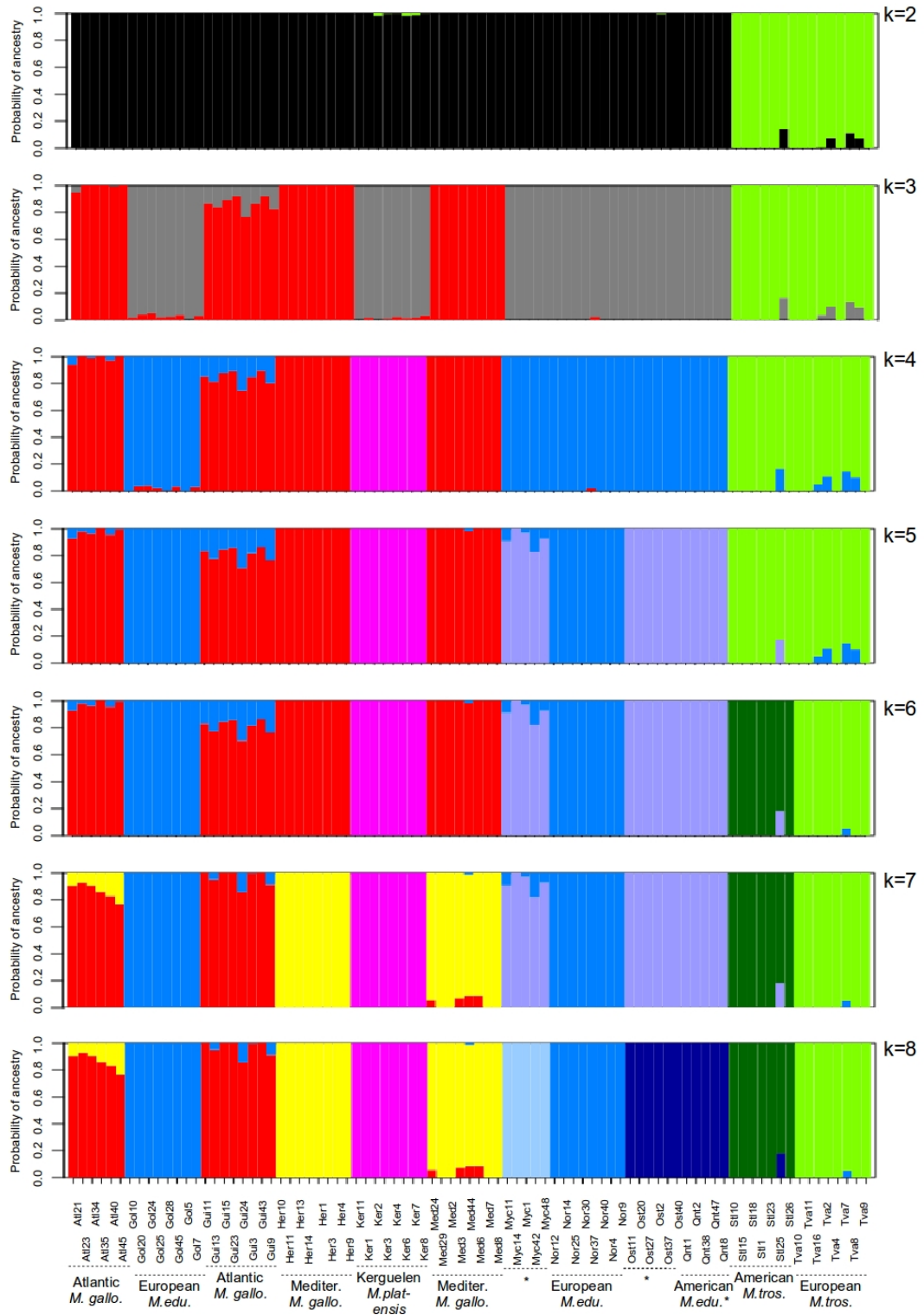
**Figure S4. A.** Geographic variation of foreign allele frequencies across sampling sites (ordered by geography) at the four loci most differentiated in the Kerguelen Islands: X10, X11, X57 and Glu-5' from Gérard *et al.* (2015). Alleles were labelled based on their frequencies in the *Mytilus galloprovincialis* Atlantic population of Iberian Coast (Table S4). Points (resp. triangles) represent sites characterized by the presence (resp. absence) of *Macrocystis* kelp beds. **B.** Correlation of the foreign allele frequency between each SNP and Glu-5'. Pearson correlation coefficient: Glu-5' vs X10:  $r=0.615$  (p-value < 0.001); Glu-5' vs X11:  $r=0.419$  (p-value=0.012); Glu-5' vs X57:  $r=0.49$  (p-value=0.003).



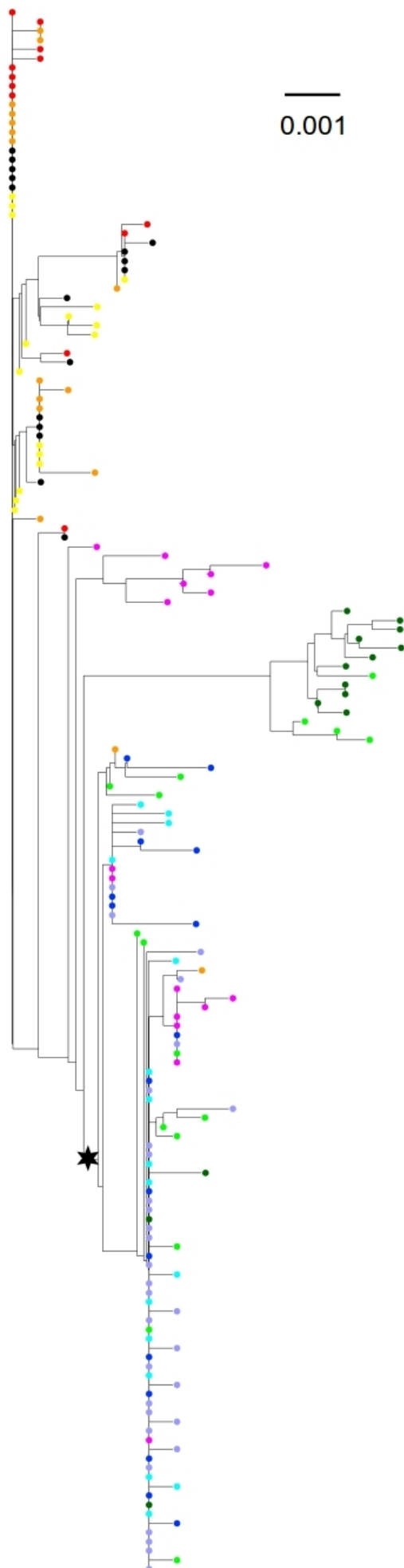
**Figure S5. A.** RDA analysis of the KASPar dataset (35 sampling sites, 33 SNPs). The proportion of genotypic variance explained by each RDA axis is provided in brackets. Grey data points correspond to the projection of individual genotypes on RDA axes 1 and 2. Significant constraining variables were projected on RDA axes 1 and 2: *Macrocystis*, Substrate, Longitude and Slope. The Longitude vector (shown in blue) points to the strongest gradient of variation, and its length indicates the strength of its contribution to each axis. Blue data points correspond to the centroids of the categorical variables (*Macrocystis*, Substrate and Slope). **B.** SNP contributions to the first and second axes of the conditioned RDA on geography (Table S11).



**Figure S6.** Correlation between the level of differentiation among the Kerguelen Islands (x-axis) and the Northern species ( $F_{ST}$ , y-axis) at each KASPar SNP. Panel (A) shows the genetic differentiation between Kerguelen populations (global  $F_{ST}$ ); and panel (B) shows the genetic-by-environment association ( $F_{CT}$  from an AMOVA analysis performed by grouping sites according to the presence/absence of *Macrocystis*, see Gérard *et al.* 2015). Northern species are *Mytilus edulis* (EU – peripheral, European population of the North Sea) and *Mytilus galloprovincialis* (MED – west, Mediterranean population of the West basin). Ancestry-informative loci, i.e.,  $F_{ST}M.edulis\_M.galloprovincialis} > 0.120$  (horizontal dashed line, see Table S2), are depicted in orange, and they include the locus *Glu-5'* from Gérard *et al.* 2015 (red point). Wilcoxon's test between the ancestry-informative loci (orange) and the background loci (grey): (A) mean\_orange=0.02, mean\_grey=0.007 (p-value=0.01); (B) mean\_orange=0.006, mean\_grey=0.0003 (p-value=0.004). Their respective means are depicted by vertical dashed lines. Pearson correlation coefficient: (A)  $r=0.695$  (p-value<0.0001) ; (B)  $r=0.486$  (p-value=0.004).



**Figure S7.** ADMIXTURE analysis based on 32,162 GBS SNPs in the 10 GBS *Mytilus* spp. populations. Each individual is assigned to one or more ancestral cluster (K). The maximum log-likelihood run over 50 replicates is shown for each K, from 2 to 8 (ML\_K2=-1,697,206; ML\_K3=-1,447,640; ML\_K4=-1,338,272; ML\_K5=-1,291,696; ML\_K6=-1,246,387; ML\_K7=-1,214,566; ML\_K8=-1,199,842). The ancestral clusters corresponding to the four species appear from K=2 to 4. Afterwards (from K=5 to 8), each Northern species is divided into different ancestral clusters revealing intraspecies differentiation. Importantly, the Kerguelen individuals never appear to be admixed with Northern backgrounds. Abbreviations are as follows: *M. gallo* = *Mytilus galloprovincialis*; *M. edu* = *Mytilus edulis*; *M. tro* = *Mytilus trossulus*. Colors match Figure 1.



**Figure S8.** Neighbour-joining tree (F84 distances) of the contig containing the candidate SNP "X10". Gene tree was computed on a 2 Kb block flanking the focal SNP (position 1,739 on contig "H\_L1\_abyss\_Contig216", see Table S2). The black star indicate the likely position of X10 on the tree, based on its allele frequency in the Northern species and the Kerguelen Islands (Table S2). Colors match Figure 1.