Genome-wide insights into introgression and its consequences for genome-wide heterozygosity in the *Mytilus* species complex across Europe

David L. J. Vendrami | Michele De Noia | Luca Telesca | Eva-Maria Brodte | Joseph I. Hoffman

**Abstract**

The three mussel species comprising the *Mytilus* complex are widespread across Europe and readily hybridize when they occur in sympathy, resulting in a mosaic of populations with varying genomic backgrounds. Two of these species, *M. edulis* and *M. galloprovincialis*, are extensively cultivated across Europe, with annual production exceeding 230,000 tonnes. The third species, *M. trossulus*, is considered commercially damaging as hybridization with this species results in weaker shells and poor meat quality. We therefore used restriction site-associated DNA sequencing to generate high-resolution insights into the structure of the *Mytilus* complex across Europe and to resolve patterns of introgression. Inferred species distributions were concordant with the results of previous studies based on smaller numbers of genetic markers, with *M. edulis* and *M. galloprovincialis* predominating in northern and southern Europe respectively, while introgression between these species was most pronounced in northern France and the Shetland Islands. We also detected traces of *M. trossulus* ancestry in several northern European populations, especially around the Baltic and in northern Scotland. Finally, genome-wide heterozygosity, whether quantified at the population or individual level, was lowest in *M. edulis*, intermediate in *M. galloprovincialis*, and highest in *M. trossulus*, while introgression was positively associated with heterozygosity in *M. edulis* but negatively associated with heterozygosity in *M. galloprovincialis*. Our study will help to inform mussel aquaculture by providing baseline information on the genomic backgrounds of different *Mytilus* populations across Europe and by elucidating the effects of introgression on genome-wide heterozygosity, which is known to influence commercially important traits such as growth, viability, and fecundity in mussels.

**Keywords**

genetic variation, genome-wide heterozygosity, hybridization, introgression, *Mytilus*, Restriction site-associated DNA sequencing (RAD sequencing), stock structure
1 | INTRODUCTION

Mussel farming is one of the most important aquaculture sectors in Europe, with annual production exceeding 230,000 tons (FAO, 2015). The three mussel species present in Europe, collectively referred to as the Mytilus complex, readily hybridize when they occur in sympatry (Gosling, 1992). Two of them, *M. edulis* (hereafter referred to as ME) and *M. galloprovincialis* (hereafter referred to as MG), are extensively cultivated along the Atlantic coast of Europe as well as in the Mediterranean (Michalek, Ventura, & Sanders, 2016). By contrast, the third species, *M. trossulus* (hereafter referred to as MT) is undesirable for cultivation due to the possession of fragile shells and poor quality meat (Penney, Hart, & Templeman, 2007, 2008). Its hybridization with the other two species has been described as commercially damaging and has been linked to significant economic losses in regions of Scotland (Scott et al., 2010; Scottish Government, 2014).

Owing to the economic importance of mussels and because the *Mytilus* complex is ideally suited to exploring the processes that generate and maintain hybrid zones, several population genetic studies have sought to characterize the geographic distributions of these species across Europe and to identify areas in which introgression takes place. Studies using diverse genetic markers, from mitochondrial sequences through allozymes and microsatellites to single nucleotide polymorphisms (SNPs), have shown that MG is the dominant species in southern Europe (Daguin, Bonhomme, & Borsa, 2001; Sanjuan, Zapata, & Alvarez, 1994; Zbawicka, Drywa, Śmietanka, & Wenne, 2012), where it is found across the Mediterranean and along the coast of the Iberian Peninsula, whereas ME dominates the cooler northern European coastlines (Daguin et al., 2001; Zbawicka et al., 2012). By contrast, MT is better adapted to brackish conditions and occurs in the Baltic as well as in parts of Norway and Greenland (Kijewski, Zbawicka, Väinölä, & Wenne, 2006; Mathiesen et al., 2017; Stuckas et al., 2017; Väinölä & Strelkov, 2011; Wenne, Bach, Zbawicka, Strand, & McDonald, 2016; Zbawicka et al., 2012; Zbawicka, Sa’iko, Strand, & Wenne, 2014), while its hybrids have also been found in the Netherlands and Scotland (Beaumont, Hawkins, Doig, Davies, & Snow, 2008; Śmietanka, Zbawicka, Wołowicz, & Wenne, 2004; Zbawicka, Burzyński, Skibinski, & Wenne, 2010). Studies of hybridization between ME and MG have also revealed unforeseen complexities in France, the Netherlands, and the UK, where pure populations of the parental species coexist with mixed populations, resulting in complex species mosaics (Beaumont, Turner, Wood, & Skibinski, 2004; Bierne et al., 2003; Faure, David, Bonhomme, & Bierne, 2008; Hilbish, Carson, Plante, Weaver, & Gilg, 2002).

Although previous population genetic studies have uncovered clear evidence for hybridization in European mussels, estimates of the magnitude of introgression have tended to be somewhat crude due to the use of a single diagnostic marker (Me15/16, Inoue, Waite, Matsuoka, Odo, & Harayama, 1995) or small panels of diagnostic, partially diagnostic or otherwise informative loci. A related problem is that genetic markers designed to discriminate between species may suffer from ascertainment bias when used to quantify patterns of genetic variability across species (Heslot, Rutkoski, Poland, Jannink, & Sorrels, 2013; Lachance & Tishkoff, 2013). In principle, both of these issues can be circumvented by subjecting the pure species together with any potential hybrids to restriction site-associated DNA (RAD, Baird et al., 2008) sequencing, an approach for genotyping thousands of essentially random genome-wide distributed SNPs.

An important aspect of genetic variability that has been linked to fitness variation in many species is heterozygosity (Chapman, Nakagawa, Coltman, Slate, & Sheldon, 2009; David, 1998; Hansson & Westerberg, 2002; Szulkin, Bierne, & David, 2010). Literally hundreds of studies of wild organisms ranging from shellfish to birds and mammals have uncovered heterozygosity–fitness correlations (HFCs) for a wealth of traits ranging from early survival through growth to reproductive success (Coltman, Bowen, & Wright, 1998; Pujolar, Maes, Vancoillie, & Volckaert, 2005; Slate, Kruuk, Marshall, Pemberton, & Clutton-Brock, 2000). In *Mytilus*, over forty such studies have been conducted (reviewed in Koehn, 1991). These consistently point toward positive effects of heterozygosity on energy metabolism and protein synthesis, which in turn influence a multitude of commercially important traits including early growth and feeding rates, viability and reproductive output. However, HFCs remain poorly understood because most studies use too few genetic markers to accurately quantify variation in genome-wide heterozygosity, or inbreeding (Balloux, Amos, & Coulson, 2004; Kardos, Allendorf, & Luikart, 2014). Fortunately, the large genome-wide distributed SNP datasets generated by RAD sequencing have proven capable of quantifying inbreeding with far greater precision than small panels of classical genetic markers (Hoffman et al., 2014).

Here, we RAD sequenced mussel populations of unknown ancestry along a European latitudinal cline together with putatively pure reference populations of the three *Mytilus* species in order to characterize genotype frequencies across Europe and to investigate introgression and its effects on genome-wide heterozygosity. We reconstructed local ancestries with high precision to test a number of hypotheses. First, while previous studies of introgression in mussels focused mainly on the frequency of hybrids in each population (e.g., Bierne et al., 2003; Daguin et al., 2001), our high-resolution data allowed us to quantify the magnitude of introgression at both the individual and population level. Given that all three *Mytilus* species readily hybridize, we hypothesized that introgression would be widespread, even if the fraction of introgressed alleles might be low in some populations. Second, we evaluated the geographical distribution of commercially damaging MT genotypes, which we hypothesized would be more abundant in areas of low salinity. Third, we expected to find a universally positive effect of hybridization on genome-wide heterozygosity, the overall aims of our study were (a) to inform the mussel industry about the genomic backgrounds of potential sources of mussel seed; and (b) to understand the consequences of introgression for genome-wide heterozygosity,
which has previously been linked to variation in commercially relevant traits.

2 | MATERIALS AND METHODS

2.1 | Sample collection

A total of 262 mussel samples were collected from 13 different sites along the Atlantic coastline of mainland Europe as well as from one site each from the Mediterranean and the Atlantic coast of Canada (Table 1 and Figure 1). These included 12 populations that were previously classified as “potentially introgressed” (Bierne et al., 2003; Daguin et al., 2001) plus three “putatively pure” reference populations of ME (GE1), MG (ITA), and MT (CAN) (Daguin et al., 2001; Stuckas, Stoof, Quesada, & Tiedemann, 2009; Wilson, Matejusova, McIntosh, Carboni, & Bekaert, 2018). All of the samples were of wild origin, with the exception of those from UK5, which originated from a mussel farm.

2.2 | DNA extraction, RAD sequencing, and bioinformatic analysis

Whole genomic DNA was extracted from the adductor muscle of each sample using an adapted phenol-chloroform protocol (Sambrook, Fritsch, & Maniatis, 1989) and shipped to the Beijing genomics institute for RAD sequencing. Libraries were constructed using the restriction enzyme PstI and sequenced on an Illumina HiSeq 4000 to generate a total of 292,239,549 50 bp single-end reads. After assessing the quality of the demultiplexed sequence reads using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), the reads were de novo assembled using the Stacks 2.2 pipeline (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). Values of the three main parameters \( -m \), \( -M \), and \( -n \) were chosen following the optimization procedure described by Rochette and Catchen (2017). Briefly, \( -n \) was set to three and a range of values for \( -M \) and \( -n \) were evaluated. The combination of these parameters for which the number of polymorphic loci present in at least 80% of individuals reached a plateau was defined as optimal. Two different strategies were employed: \( -n \) was either set as equal to \( -M \) or one unit greater to account for the possible presence of polymorphisms fixed in one of the three Mytilus species (Paris, Stevens, & Catchen, 2017). The optimal combination \( (m = 3, M = 5, n = 6) \) was then selected for analyzing the entire dataset. However, only the 51 samples belonging to the pure populations were used to generate the catalog in order to minimize the potential for noise (Rochette & Catchen, 2017). The raw genotypes were then filtered to retain only biallelic SNPs with genotype quality and depth of coverage greater than five using VCFtools (Danecek et al., 2011), as well as to retain only SNPs genotyped in at least 60% of the individuals. Subsequently, we discarded all SNPs with a depth of coverage greater than twice the mean depth of the raw SNP dataset (>34.4) in order to filter out potentially paralogous loci. Next,

| Population ID | Sampling location     | Population ancestry | Samples passing QC | \( H_o \) |
|---------------|-----------------------|---------------------|--------------------|--------|
| ITA           | Comacchio, Italy      | Pure MG             | 16                 | 0.11   |
| PO1           | Faro, Portugal        | Potentially introgressed | 18                 | 0.1    |
| PO2           | Lisbon, Portugal      | Potentially introgressed | 18                 | 0.1    |
| PO3           | Porto, Portugal       | Potentially introgressed | 18                 | 0.1    |
| FRA           | Brest, France         | Potentially introgressed | 15                 | 0.09   |
| UK1           | Exmouth, UK           | Potentially introgressed | 16                 | 0.09   |
| NET           | Texel, Netherlands    | Potentially introgressed | 17                 | 0.09   |
| GE1           | Helgoland, Germany    | Pure ME             | 17                 | 0.09   |
| GE2           | Kiel, Germany         | Potentially introgressed | 17                 | 0.09   |
| UK2           | Belfast, UK           | Potentially introgressed | 15                 | 0.09   |
| UK3           | St. Andrews, UK       | Potentially introgressed | 14                 | 0.09   |
| UK4           | Oban, UK              | Potentially introgressed | 18                 | 0.09   |
| SWE           | Kristeneberg, Sweden  | Potentially introgressed | 17                 | 0.09   |
| UK5           | Shetland Isl., UK     | Potentially introgressed | 18                 | 0.09   |
| CAN           | Bras d’Or Lake, Canada| Pure MT             | 18                 | 0.13   |

Note: Population ancestry was assigned a priori as putatively pure M. edulis (ME), putatively pure M. galloprovincialis (MG), putatively pure M. trossulus (MT), or potentially introgressed. The number of samples passing quality checks (QC) and mean observed heterozygosity (\( H_o \)) is reported for each population.
all individuals with more than 50% missing data were removed and only variants with MAF greater than 0.01 were retained. Finally, the software PLINK (version 1.9, Purcell et al., 2007) was used to prune out putatively linked loci using an r² threshold of .5.

2.3 | Genetic analysis of the putatively pure populations

Prior to analyzing the full dataset, we conducted a separate analysis of the putatively pure Mytilus samples to validate their use in subsequent introgression analyses. First, we used the R package “ape” (version 5.3, Paradis & Schliep, 2018) to construct a phylogenetic tree based on Euclidean distances followed by hierarchical clustering using Ward’s method. Second, we implemented principal component analysis (PCA) using the R package “adegenet” (version 2.1.1, Jombart, 2008; Jombart & Ahmed, 2011). Third, we explored patterns of polymorphism within and among species by quantifying the number of SNPs that were polymorphic within each species and across species. The results of this analysis were visualized in the form of a Venn diagram using the R package “venneuler” (version 1.1, Wilkinson, 2011).

2.4 | Population genetic structure and introgression

A number of complementary approaches were applied to the full dataset in order to evaluate species composition and introgression. Initially, we investigated the overall pattern of population genetic structure by constructing a phylogenetic tree and by subjecting the full dataset to PCA as described above. We then explored patterns of introgression by conducting a genetic admixture analysis. Specifically, the R package LEA (version 2.6, Frichot & François, 2015) was used to perform sparse non-negative matrix factorization (sNMF) analysis, which determines the most likely number of genetic clusters (K) present in a dataset and outputs individual admixture coefficients (Q). sNMF efficiently handles large numbers of markers while in general performing as well as other methods that estimate admixture proportions (Wollstein & Lao, 2015). Five independent runs with an alpha regularization parameter of 100 were performed for each value of K, which was set to between one and five, and the best K value was determined by calculating cross-entropy values.

For comparison, we also conducted a formal analysis of introgression using the software ELAI (version 0.99, Guan, 2014). ELAI relies on a hidden Markov model to infer local ancestry for admixed individuals, which is quantified as “ancestry dosage” and has two important advantages over alternative approaches: (a) it does not require phased data and/or mapping distances; and (b) it is capable of dealing with three-way hybridization scenarios. Five independent replicate runs were implemented for all twelve of the potentially introgressed populations, while the putatively pure Mytilus populations were used as “baselines.” This analysis was performed by setting the number of upper clusters (-c) to three, corresponding to the number of Mytilus species present in Europe. The number of lower clusters (-c) was set to five, as recommended by Guan (2014), the number of admixture generations (-mg) was set to 40 to account for potential admixture over the last century, and the number of expectation-maximization (EM) steps (-s) was set to 40 following the recommendation of Guan (2014). Ancestry dosage values from each of the five runs were then averaged across all individuals within each of the populations to quantify population-specific fractions of ME, MG and MT ancestry.

Finally, data on salinity and sea surface temperature (SST) were extracted from the E.U. Copernicus Marine Service Information (http://marine.copernicus.eu) and used to test for associations between environmental conditions and the proportions of ME, MG, and MT ancestry across Europe. Specifically, we tested for correlations between each of the three Mytilus ancestry values and the two environmental variables while correcting the resulting p-values for the false discovery rate (FDR) as described in Benjamini and Hochberg (1995). We decided to focus on SST because ME and MG have often been referred to as cold-temperate and warm-temperate species respectively, and on salinity because adaptation to low salinity environments has been proposed for MT (Riginos & Cunningham, 2005).

2.5 | Effects of introgression on heterozygosity

To test for differences in genome-wide heterozygosity among populations with different genetic backgrounds, we used PLINK to calculate observed heterozygosity (H) averaged over all loci for each population. We then attributed “main ancestry” to each population depending on whether ME, MG, or MT had the highest mean ancestry dosage value and tested for significant differences in heterozygosity using a t test. Subsequently, we investigated the effect of introgression on heterozygosity among populations with different
main ancestries. To do so, we constructed a general linear model (GLM) where main ancestry, introgression (measured as the total proportion of ancestry dosage not attributable to the focal species), and their interaction were modeled as predictors of $H_o$.

Next, we replicated the above analyses at the individual level, this time expressing heterozygosity as standardized multilocus heterozygosity (sMLH), which was calculated using the R package InbreedR (Stoffel et al., 2016). Main ancestry was assigned to each individual according to whether the maximal ancestry dosage was attributable to ME, MG, or MT, and an ANOVA was implemented to test for differences in sMLH among individuals with different genetic backgrounds. We then investigated the effect of introgression on sMLH by constructing a general linear mixed model (GLMM). Here, the response variable was sMLH and individual main ancestry, introgression, and their interaction
were fitted as predictors, while sampling location was also included as random effect. The significance of the predictor variables was determined using a parametric bootstrap approach. Specifically, we constructed two alternative models, one including and one excluding the term of interest, and calculated an observed likelihood ratio statistic. We then simulated 1,000 bootstrap replicates based on the null model and used them to generate the null distribution of our test statistic and to calculate a p-value.

3 | RESULTS

To provide detailed insights into the structure of the *Mytilus* species complex across Europe and to investigate the impact of introgression on genome-wide heterozygosity, we RAD sequenced a total of 262 mussel samples from fifteen different populations (Table 1). These included twelve populations spanning a European latitudinal cline plus putatively pure reference populations of ME, MG, and MT from Helgoland in Germany, Comacchio in Italy and the Bras d’Or Lake in Canada, respectively. A total of 292,239,549 single-end 50 bp Illumina sequence reads were generated and de novo assembled into 926,383 RAD loci, which were used to call 1,773,643 raw SNPs. Application sequence reads were generated and de novo assembled into 926,383 RAD loci, which were used to call 1,773,643 raw SNPs. Application of the stringent filtering criteria described in the Materials and methods resulted in a final dataset consisting of 252 samples genotyped at 6,777 SNPs with an average depth of coverage of 12.8.

3.1 | Species-level relationships and genetic diversity

In order to verify the species identities of our putatively pure ME, MG, and MT samples, we constructed a phylogenetic tree and subjected the data to principal component analysis (PCA). The RAD dataset clearly resolved all three *Mytilus* species and showed that ME is phylogenetically more closely related to MG than MT (Figure 2a and b). All of the samples grouped as expected based on their putative species identities, suggesting that our species assignments are correct, and none of the samples occupied intermediate positions in the tree, implying an absence of hybrids. Observed heterozygosity was highest for MT, intermediate for MG, and lowest for ME (Table 1). Over half of the loci were polymorphic in a single *Mytilus* species (n = 3,544, 52.3%) while around a quarter were polymorphic in two species (n = 1,645, 24.3%) and fewer than 10% (n = 613) were polymorphic in all three species (Figure S1). 66 loci exhibited fixed differences between MT and the other two species and were therefore classified as putatively MT-diagnostic. Further information on these loci including their flanking sequences is provided in Table S1.

3.2 | Species composition across Europe

Phylogenetic analysis of the full dataset of fifteen populations uncovered three well-supported clades broadly corresponding to the three *Mytilus* species (Figure 2c). Consistent with the phylogenetic tree of the three pure populations, the MT clade (which included the Canadian samples plus a single mussel from the UK) was resolved as an outgroup. The remaining samples formed two clades, corresponding to ME and MG, respectively. The former did not show any evidence of sub-structure, while the latter was further divided into two groups comprising the pure MG samples and the remaining predominantly southern European samples.

PCA of the full dataset revealed a similar pattern, with the samples clustering into three distinct groups (Figure 2d). Specifically, the first principal component separated the pure MT samples from the remaining samples, with the exception of a single British sample, while the second principal component separated ME from MG. Furthermore, when PC3 was also taken into account, the pure MG samples from the Adriatic separated apart from the other populations belonging to the MG cluster (Figure S2). Additionally, an appreciable number of individuals occupied intermediate positions between the pure ME and MG samples, providing a first indication of the presence of introgressed mussels in our dataset.

3.3 | Patterns of introgression

To investigate geographical patterns of introgression, we used sNMF to assign individuals to genetic clusters and to derive admixture coefficients. The most likely number of genetic clusters (K) in our dataset was three (Figure S3), corresponding to the three *Mytilus* species. Membership coefficients (Q) for the inferred clusters are summarized in Figure 3a, where each vertical bar represents a different individual and the relative proportions of the different colors indicate the probabilities of being assigned to each cluster. Samples from the three Portuguese populations (PO1, PO2 and PO3) and from northern France (FRA) were predominantly assigned to the MG cluster, with the remaining ancestry being largely attributable to ME. The remaining northern European populations were predominantly assigned to the ME cluster, although two populations from the east coast of Scotland and the Shetland islands carried somewhat larger MG contributions and a single individual from Kiel in Germany was assigned as pure MG.

Small amounts of MT ancestry were also detected in several northern European populations (Figure 3a). The fraction of ancestry attributable to MT was generally below 0.1, although two populations closest to the Baltic Sea (GE2 and SWE) had higher Q values in the order of 0.1–0.2 and a single individual from Oban (UK4) on the west coast of Scotland had a Q value of 0.72. To investigate further, we examined the subset of 66 loci that were found to be diagnostic of MT in our previous analysis of the pure populations. We found that mussel populations from Oban, Kiel, and Kristeneberg carried MT-diagnostic alleles at between ten and 25 of these loci (Figure S4), suggesting that the signal of introgression is not an artifact of allele frequency differences at loci that are not actually diagnostic of MT. Slightly fewer diagnostic alleles were detected in the Shetland Islands, around the UK, and in the Netherlands, while only
ancestry dosage values. We then tested for differences in observed
between SST and the mean ancestry dosage values of
tion, as significant associations were found at the population level
-Sea. These patterns are partly explained by environmental varia-
the coast of the UK and in proximity to the entrance to the Baltic
ries. Figure 4a shows that populations dominated by
had the highest overall
coasts of the UK, the Netherlands, Germany and Sweden. Small frac-
northern France, while
for a large proportion of the ancestry of mussels from Portugal and
attributable to the other two Mytilus species. Finally, we constructed
the population with main ancestry, introgression, and their interaction
p
428
introgression showed a significant correlation with salinity
2–3 diagnostic alleles were found in northern France and Portugal,
either suggesting that MT introgression occurs in a limited way as
far south as Faro, or that a small fraction of these loci may not be
consistent with the results of the clustering analysis, MG accounted
for a large proportion of the ancestry of mussels from Portugal and
northern France, while ME was the dominant species around the
coasts of the UK, the Netherlands, Germany and Sweden. Small frac-
tions of MT ancestry were also detected in populations from around
the coast of the UK and in proximity to the entrance to the Baltic
Sea. These patterns are partly explained by environmental varia-
tion, as significant associations were found at the population level
between SST and the mean ancestry dosage values of ME ($r = -0.832$,
$p < .01$), MG ($r = .84$, $p < .01$), and MT ($r = -.778$, $p < .01$), while only
MT introgression showed a significant correlation with salinity
($r = -.676$, $p < .05$).

3.4 | Effects of introgression on heterozygosity

In order to investigate the influence of introgression on genome-
wide heterozygosity, we first assigned “main ancestry” to each
population based on whether ME, MG, or MT had the highest mean
ancestry dosage values. We then tested for differences in observed
heterozygosity ($H_o$) among populations with different main ances-
tries. Figure 4a shows that populations dominated by ME ancestry
had significantly lower $H_o$ than populations whose main ancestry
was MG (unpaired t test, $t = -5.45$, $p < .01$). The pure MT population
had the highest overall $H_o$ although we did not test for statistical
significance given that none of the other populations had majority
MT contributions. For each population, we then quantified the mag-
nitude of introgression as the total proportion of ancestry dosage
attributable to the other two Mytilus species. Finally, we constructed
a GLM of $H_o$ with main ancestry, introgression and their interaction
fit as predictor variables. All three were highly significant (main
ancestry: $F = 139.28$, $p < .01$; introgression: $F = 7.36$, $p < .05$; interac-
tion: $F = 21.7$, $p < .01$), confirming species-level differences and sug-
gesting that the influence of introgression on $H_o$ depended on the
primary genomic background. Specifically, the introgression of ME
alleles into populations whose main ancestry was MG had a highly
significant negative influence on $H_o$ ($b = -0.02$, $t = -6.36$, $p < .01$)
whereas the introgression of MG alleles into populations whose main
ancestry was ME had a weakly positive but nonsignificant effect
($b = 0.003$, $t = -0.977$, $p = .36$, Figure 4b).

Next, we repeated the analysis at the individual rather than the
population level. Each individual was assigned main ancestry
according to the maximal ancestry dosage attributable to ME, MG,
or MT, and individual genome-wide heterozygosity was quanti-
fied as standardized multilocus heterozygosity (sMLH). Again, we
found clear differences in heterozygosity among individuals with
different main ancestries, with mean sMLH being lowest for indi-
viduals whose main ancestry was ME, intermediate for individuals
whose main ancestry was MG, and highest for individuals whose
main ancestry was MT (ANOVA: $F = 544.33$, $p < .01$, Figure 4c). To
investigate the interplay between introgression and heterozygosity
at the individual level, we constructed a GLMM where sMLH was the
response variable, individual main ancestry, introgression, and
their interaction were fitted as predictor variables, and sampling lo-
cation was included as a random effect. Again, all three predictors
were statistically significant (main ancestry, $p < .01$; introgression,
p = .018; interaction, p < .01) with ME introgression into individuals whose main ancestry was MG having a negative influence on sMLH (b = −0.05, p < .01), whereas MG introgression into individuals whose main ancestry was ME had a positive influence on sMLH (b = 0.04, p < .01, Figure 4d).

To understand why introgression does not increase genome-wide heterozygosity in both species, we conducted a more detailed analysis focusing on ME and MG. For this, we exploited information from the pure populations to select three mutually exclusive subsets of SNPs. The first comprised SNPs that were only polymorphic in pure ME individuals ("ME-SNPs," shown in green), for loci that were only polymorphic in pure MG individuals ("MG-SNPs," shown in blue) and for loci that were polymorphic in both pure ME and pure MG individuals ("ME/MG-SNPs," shown in gold). The regression lines show the fit of generalized linear models constructed separately for each subset of loci.
ME individuals (hereafter termed “ME-SNPs,” n = 1,592). The second comprised loci that were polymorphic in both pure ME and pure MG individuals (hereafter termed “ME/MG-SNPs,” n = 1,415). The third comprised loci that were only polymorphic in pure MG individuals (hereafter termed “MG-SNPs,” n = 1,501). We then calculated sMLH separately for each class of SNP and investigated how the resulting values were influenced by introgression separately for each species. For individuals whose main ancestry was ME, the decrease of sMLH at ME-SNPs with increasing introgression was less pronounced than the increase of sMLH at MG-SNPs. ME-SNPs: \( b = -0.51, \ p < .01; \) ME/MG-SNPs: \( b = -0.03, \ p = .047; \) MG-SNPs: \( b = 0.61, \ p < .01; \) Figure 4e). By contrast, for individuals whose main ancestry was MG, the decrease of sMLH at MG-SNPs with increasing introgression was more pronounced than the increase of sMLH at ME-SNPs (ME-SNPs: \( b = 0.66, \ p < .01; \) ME/MG-SNPs: \( b = -0.12, \ p < .01; \) MG-SNPs: \( b = -1.51, \ p < .01; \) Figure 4f). This suggests that the balance of the contributions of ME-SNP and MG-SNP heterozygosity toward genome-wide heterozygosity may shift depending on the main genetic background and the magnitude of introgression.

4 | DISCUSSION

We used RAD sequencing to obtain detailed insights into the genetic composition of the Mytilus species complex in Europe. We found evidence for widespread introgression, particularly between ME and MG, although small contributions of MT ancestry were also detected across much of northern Europe. Moreover, introgression had opposing effects on genome-wide heterozygosity depending on the primary genetic background. As MT is considered a commercially damaging species (Scott et al., 2010; Scottish Government, 2014) and heterozygosity is known to impact commercially important traits in Mytilus (Koehn, 1991), our findings may have implications for mussel aquaculture.

4.1 | Reference populations

Many of our analyses relied upon inferences derived from pure reference populations of ME, MG, and MT. We therefore carefully selected reference populations that had been described as pure in the literature (Daguin et al., 2001; Stuckas et al., 2009; Wilson et al., 2018). We specifically chose a Canadian MT reference population as opposed to a Baltic one because MT and ME have extensively mixed in the Baltic, leading to complete replacement of mitochondrial genomes and a hybrid swarm structure (Kijewski et al., 2006; Väinölä & Strelkov, 2011). To confirm the validity of our reference samples, we conducted a phylogenetic analysis, which resolved ME and MG as sister groups and MT as an outgroup. This pattern is consistent with previous molecular and morphological studies (Barsotti & Meluzzi, 1968; Heath, Rawson, & Hilbish, 1995; Vermeij, 1991). Furthermore, the pattern of grouping of individual samples suggested they had all been correctly assigned to species. This is important because a similar study clustered one out of five pure MT reference individuals from Penn Cove in the USA together with pure ME individuals from Scotland (Wilson et al., 2018), implying that it may be relatively easy to incorrectly assign individual mussels to species based solely on their provenance.

4.2 | Species distributions and introgression

In line with previous studies based on smaller numbers of mitochondrial or nuclear genetic markers, clear species partitioning was found between southern and northern Europe. Specifically, MG ancestry predominated in the Mediterranean, along the coast of the Iberian Peninsula and in Brittany, consistent with Faure et al. (2008), Bierne et al. (2003) and Sanjuan et al. (1994), whereas ME was the dominant species across much of northern Europe, as previously shown by Zbawicka et al. (2012) and Daguin et al. (2001). Although we did not find any pure MT individuals in our dataset, small fractions of MT ancestry were apparent across much of northern Europe and in particular around the entrance to the Baltic as well as in northern Scotland. Prevailing environmental conditions are likely to play a role in explaining these distributions, as ancestry dosage was associated with SST in all three species as well as with salinity in MT, consistent with previous work by Riginos and Cunningham (2005). However, the transport of spat in ocean currents or via shipping may also contribute toward the local composition of mussel populations (Stuckas et al., 2017), as appears to be the case in Svalbard where mussels carry large amounts of MG ancestry despite ME dominating the surrounding areas (Mathiesen et al., 2017).

Among the southern, predominantly MG populations, we found a general tendency for ME introgression to increase with increasing latitude. However, this pattern may be an artifact of our sampling design, as our dataset does not have the spatial resolution to capture complexities that are known to be present in this system. For example, instead of a single transition occurring from MG to ME along the western Atlantic seaboard, a mosaic hybrid zone is present that comprises three separate transitions (Bierne et al., 2003; Faure et al., 2008). Our dataset was unable to capture this fine-scale heterogeneity, although we did find that MG ancestry was relatively high in Brittany, consistent with this part of northwestern France constituting an “MG island” surrounded by predominantly ME populations (Bierne et al., 2003; Faure et al., 2008).

Among the northern, predominantly ME populations, considerable geographical variation was found in the magnitude of MG introgression. This is again consistent with previous studies documenting a mosaic structure across the UK (Gardner & Skibinski, 1988; Gosling & McGrath, 1990; Hilbish et al., 2002; Skibinski, Beardmore, & Cross, 1983; Wilhelm & Hilbish, 1998). Given that MG is considered a warm-temperate species (Michalek et al., 2016), we were initially surprised to find relatively large amounts of MG ancestry in two of the most northerly UK populations, St. Andrews and the Shetland Islands. However, high frequencies of MG alleles have been documented at even higher latitudes, possibly due to oceanographic
features or human-mediated transport (Mathiesen et al., 2017). A role of humans also cannot be discounted in northern Scotland as our sample from the Shetland Islands originated from a farmed population that has been augmented with spat from other localities (Michael Tait, personal communication).

We also captured a pervasive but low-level signal of MT ancestry across most of northern Europe. Previous studies have shown that MT alleles occur at high frequency in the Baltic (Kijewski et al., 2006, 2019; Väinölä & Strelkov, 2011; Zbawicka et al., 2012; Zbawicka et al., 2014) as well as in some parts of Norway and Greenland (Mathiesen et al., 2017; Wenne et al., 2016). Additionally, MT introgression was implicated in the recent collapse of the mussel farming industry at Loch Etive in western Scotland (Beaumont et al., 2008) and has since been documented at several Scottish locations including Highland and Argyll (Beaumont et al., 2008; Michalek et al., 2016). However, our data are suggestive of small contributions of MT ancestry not only around the Baltic and the Scottish coasts, but also in the Netherlands, southwest England, Northern Ireland, and the Shetlands. In addition, a single mussel from Oban (out of a total of 18 individuals from this location) had almost 75% MT ancestry. Overall, our results imply that MT alleles may be more widespread than was previously appreciated.

4.3 | Introgression and heterozygosity

Previous studies of the Mytilus complex in Europe have often neglected genetic diversity, partly due to the unsuitability of diagnostic markers for quantifying genome-wide patterns, but also due to the risk of ascertainment bias when markers developed in one species are applied to another (Heslot et al., 2013; Lachance & Tishkoff, 2013). We avoided these issues by simultaneously de novo assembling RAD sequencing data from all three species and their putative hybrids. This approach should produce relatively unbiased estimates of genome-wide variation and thus allow comparative analysis of populations with varying ancestries. As a wealth of previous studies have linked heterozygosity to variation in commercially relevant traits in mussels (reviewed by Koehn, 1991), we focused specifically on the influence of introgression on genome-wide heterozygosity, which under most circumstances can be reliably inferred from several thousand unlinked SNPs (Hoffman et al., 2014; Kardos, Taylor, Ellegren, Luikart, & Allendorf, 2016).

Highly concordant results were obtained regardless of whether the data were analyzed at the level of populations or individuals, with ME having the lowest heterozygosity, MG having intermediate heterozygosity, and MT having the highest heterozygosity. This is in line with Gardner (1994) who found that allozyme heterozygosity was higher in MG than ME, as well as with Zbawicka et al. (2014) who reported higher levels of SNP heterozygosity in MT relative to ME. However, our results are at odds with two other studies documenting comparatively low levels of heterozygosity in MT (Mathiesen et al., 2017; Zbawicka et al., 2012). One possible explanation for this discrepancy could be ascertainment bias, as Mathiesen et al. (2017) used SNPs that were mainly discovered in ME. Similarly, the majority of SNPs analyzed by Zbawicka et al. (2012) were fixed for a single allele in a pure MT population and it is therefore unclear to what extent these loci are representative of the genetic variability of MT. Our study should be relatively unaffected by issues relating to preascertained markers, both because our pools of pure individuals were equally large and because RAD sequencing allows thousands of SNPs to be genotyped regardless of the main genetic background or degree of introgression.

One potential drawback of our approach, however, was that the flanking sequences of our RAD loci were too short (approx. 45bp) to allow reliable mapping to a reference genome. Consequently, our SNPs are not accompanied by positional information and functional annotations are lacking for any SNPs that may reside in genes. We do not see this as a major drawback of our study as we were primarily interested in genome-wide patterns as opposed to the role of specific genomic regions. Nevertheless, more detailed studies of the genomic landscape of introgression are essential for improving our understanding of adaptive phenotypic variation and selection in Mytilus.

Somewhat counterintuitively, introgression appears to have contrasting effects on genome-wide heterozygosity depending on the Mytilus species in question. In populations or individuals whose main ancestry was ME, we found that the introgression of MG alleles was associated with an increase in heterozygosity. By contrast, the introgression of ME alleles into populations or individuals whose main ancestry was MG was associated with a decrease in heterozygosity. This pattern appears to be a reflection of species-level differences in heterozygosity and of the balance between the increase of heterozygosity caused by the introgression of new alleles versus the loss of heterozygosity as the primary genetic background is progressively diluted.

4.4 | Implications for mussel aquaculture

In Europe, seed supply for mussel production relies either on natural local recruitment or on the transfer of spat from shellfish farms (Michalek et al., 2016; Smetanka et al., 2004). Molecular genomic tools such as RAD sequencing could therefore assist the aquaculture industry by providing information in support of decisions such as where to locate mussel farms and where to source mussel spat. Our study suggests that RAD sequencing is capable of providing detailed information on stock structure and introgression in Mytilus. Although we focused primarily on wild mussel populations, it is not difficult to envisage how reduced representation sequencing or related approaches might be applied in an industrial setting, for example to characterize the genetic composition of commercial mussel stocks, to assist in the selection of genetic material for cultivation, or to improve our understanding of how commercially important traits vary among mussels with different genomic backgrounds.

Particularly undesirable for cultivation are mussels carrying appreciable fractions of MT ancestry (Scott et al., 2010; Scottish
Government, 2014) due to their poor quality meat and fragile shells (Beaumont et al., 2008). Consequently, information on the geographic distribution and magnitude of introgression of MT alleles will be of interest to the mussel industry. RAD sequencing allowed us to detect small amounts of MT ancestry in several northern European populations, including localities where the presence of MT had not previously been reported. Data such as these may contribute toward efforts to minimize the spread of MT by helping mussel producers to make more informed decisions about where to source their spat. Our findings also highlight the need for further screening for the presence of MT genotypes, particularly around northern European coastlines.

Finally, we uncovered evidence for widespread introgression between ME and MG and could show that introgression between these species can have rather complex effects on overall levels of genome-wide heterozygosity. While the commercial implications of these findings may not be immediately obvious, it is important to recognize that the primary genetic background (Bierne, Bonhomme, Boudry, Szulkin, & David, 2006; Coustau, Renaud, Maillard, Pasteur, & Delay, 1991; Gardner, 1994; Hilbish, Bayne, & Day, 1994; Skibinski, 1983), hybridization (Bierne et al., 2006; Gardner, 1994; Gardner, Skibinski, & Bajdik, 1993), and heterozygosity (Koehn, 1991) all have substantial effects on fitness traits such as growth rate, viability, and productivity in European mussels. Furthermore, although heterozygosity tends to be positively associated with fitness within species, the increase in heterozygosity that occurs when two species interbreed can result in a variety of outcomes ranging from hybrid vigor to outbreeding depression (Chapman et al., 2009). Even in Mytilus where hybridization has been extensively investigated, contrasting fitness outcomes have been described, with one study finding that introgression between ME and MG increased fitness (Gardner, 1994), another documenting intermediate fitness in F1 hybrids relative to the two parental species (Gardner et al., 1993), and a third study reporting high levels of larval mortality in F2 hybrids (Bierne et al., 2006). Given the degree of admixture observed in many of the sampled populations in the current study, we believe that a strong case could be made for further studies of the phenotypic effects of introgression in the Mytilus complex. RAD sequencing would offer an alternative to using artificial crosses by allowing mussels with different proportions of ME, MG, and MT ancestry (selected on the basis of their ancestry dosage values) to be raised in a common-garden setup.

ACKNOWLEDGMENTS

We are grateful to Kirti Ramesh, Joanna Wilson, Barry McDonald, Ellen Kenchington, Frederico Batista, Rob Dekker, and Mark Hamilton for sample provision. The research leading to these results has received funding from the European Union Marie Curie Seventh Framework Programme under grant agreement no. 605051. We acknowledge support for the Article Processing Charge by the Deutsche Forschungsgemeinschaft and the Open Access Publication Fund of Bielefeld University.

CONFLICT OF INTEREST

None declared.

DATA ACCESSIBILITY

The raw sequence reads used to generate the results of this study are available at the Short Read Archive (SRA accession: PRJNA615219).

REFERENCES

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., ... Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE, 3(10), e3376. https://doi.org/10.1371/journal.pone.0003376

Balloux, F., Amos, W., & Coulson, T. (2004). Does heterozygosity estimate inbreeding in real populations? Molecular Ecology, 13(10), 3021–3031. https://doi.org/10.1111/j.1365-294X.2004.02318.x

Barsotti, G., & Meluzzi, C. (1968). Osservazioni su Mytilus edulis L. e Mytilus galloprovincialis Lamarck. Conchiglie, 4, 50–58.

Beaumont, A. R., Hawkins, M. P., Doig, F. L., Davies, I. M., & Snow, M. (2008). Three species of Mytilus and their hybrids identified in a Scottish Loch: Natives, relicts and invaders? Journal of Experimental Marine Biology and Ecology, 367(2), 100–110. https://doi.org/10.1016/j.jembe.2008.08.021

Beaumont, A. R., Turner, G., Wood, A. R., & Skibinski, D. O. (2004). Hybridisations between Mytilus edulis and Mytilus galloprovincialis and performance of pure species and hybrid veliger larvae at different temperatures. Journal of Experimental Marine Biology and Ecology, 302(2), 177–188. https://doi.org/10.1016/j.jembe.2003.10.009

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), 289–300.

Bierne, N., Bonhomme, F., Boudry, P., Szulkin, M., & David, P. (2006). Fitness landscapes support the dominance theory of postzygotic isolation in the mussels Mytilus edulis and M. galloprovincialis. Proceedings of the Royal Society B: Biological Sciences, 273(1591), 1253–1260.

Bierne, N., Borsa, P., Daguin, C., Jollivet, D., Viard, F., Bonhomme, F., & David, P. (2003). Introgression patterns in the mosaic hybrid zone between Mytilus edulis and M. galloprovincialis. Molecular Ecology, 12(2), 447–461. https://doi.org/10.1046/j.1365-294X.2003.01730.x

Catchen, J., Hohenlohe, P., Bassham, S., Amores, A., & Cresko, W. (2013). Stacks: An analysis toolset for population genomics. Molecular Ecology, 22(11), 3124–3140.

Chapman, J. R., Nakagawa, S., Colman, D. W., Slate, J., & Sheldon, B. C. (2009). A quantitative review of heterozygosity–fitness correlations in animal populations. Molecular Ecology, 18(13), 2746–2765. https://doi.org/10.1111/j.1365-294X.2009.04247.x

Colman, D. W., Bowen, W. D., & Wright, J. M. (1998). Birth weight and neonatal survival of harbour seal pups are positively correlated with spring temperatures. Proceedings of the Royal Society of London. Series B: Biological Sciences, 265(1398), 803–809.

Coustau, C., Renaud, F., Maillard, C., Pasteur, N., & Delay, B. (1991). The zone of sympatry and genetic variation measured by microsatellites. Molecular Ecology, 22(11), 3124–3140.

Coutaud, V., Altenberg, A., Bapst, C., Dussin, T., & Delay, B. (1993). Introgression pattern and population structure of Mytilus edulis and M. galloprovincialis. Molecular Ecology, 2(3), 342.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... Durbin, R. (2011). The variant call format and VCFtools. Bioinformatics, 27(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

David, P. (1998). Heterozygosity–fitness correlations: New perspectives on old problems. Heredity, 80(5), 531. https://doi.org/10.1046/j.1365-2540.1998.00393.x
FAO (2015). FAO Yearbook. Fishery and Aquaculture Statistics.

Gosling, E. (1992). Genetics of Mytilus. In E. M. Gosling (Ed.), The mussel Mytilus: Ecology, physiology. Genetics and culture (pp. 309–382). Amsterdam: Elsevier Science Publisher.

Gosling, E. M., & McGrath, D. (1990). Genetic variability in exposed-shore mussels, Mytilus spp., along an environmental gradient. Marine Biology, 104(3), 413–418. https://doi.org/10.1007/BF01314344

Guan, Y. (2014). Detecting structure of haplotypes and local ancestry. Genetics, 196(3), 625–642. https://doi.org/10.1534/genetics.113.160697

Hansson, B., & Westerberg, L. (2002). On the correlation between heterozygosity and fitness in natural populations. Molecular Ecology, 11(12), 2467–2474. https://doi.org/10.1046/j.1365-294X.2002.01644.x

Heath, D. D., Rawson, P. D., & Hilbish, T. J. (1995). PCR-based nuclear markers identify alien blue mussel (Mytilus spp.) genotypes on the west coast of Canada. Canadian Journal of Fisheries and Aquatic Sciences, 52(12), 2621–2627.

Heslot, N., Rutkoski, J., Poland, J., Jannink, J. L., & Sorrells, M. E. (2013). Impact of marker ascertainment bias on genomic selection accuracy and estimates of genomic diversity. PLoS ONE, 8(9), e74612. https://doi.org/10.1371/journal.pone.0074612

Hilbish, T. J., Bayne, B. L., & Day, A. (1994). Genetics of physiological differentiation within the marine mussel genus Mytilus. Evolution, 48(2), 267–286.

Hilbish, T., Carson, E., Plante, J., Weaver, L., & Gilg, M. (2002). Distribution of Mytilus edulis, M. galloprovincialis, and their hybrids in open-coast populations of mussels in south-western England. Marine Biology, 140(1), 137–142.

Hoffman, J. I., Simpson, F., David, P., Rijks, J. M., Kuiken, T., Thorne, M. A. S., ... DasMahapatra, K. K. (2014). High-throughput sequencing reveals inbreeding depression in a natural population. Proceedings of the National Academy of Sciences, 111(10), 3775–3780. https://doi.org/10.1073/pnas.1318945111

Inoue, K., Waite, J. H., Matsuoka, M., Odo, S., & Harayama, S. (1995). Interspecific variations in adhesive protein sequences of Mytilus edulis, M. galloprovincialis, and M. trossulus. The Biological Bulletin, 189, 370–375.

Jombart, T. (2008). adegenet: A package for the multivariate analysis of genetic markers. Bioinformatics, 24(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. Bioinformatics, 27, 3070–3071.

Kardos, M., Allendorf, F. W., & Luikart, G. (2014). Evaluating the role of inbreeding depression in heterozygosity-fitness correlations: How useful are tests for identity disequilibrium? Molecular Ecology Resources, 14(3), 519–530. https://doi.org/10.1111/1755-0998.12193

Kardos, M., Taylor, H. R., Ellegren, H., Luikart, G., & Allendorf, F. W. (2016). Genomics advances the study of inbreeding depression in the wild. Evolutionary Applications, 9(10), 1205–1218. https://doi.org/10.1111/evo.12414

Kijewski, T., Zbawicka, M., Strand, J., Kautsky, H., Kotta, J., Rätsep, M., & Wenné, R. (2019). Random forest assessment of correlation between environmental factors and genetic differentiation of populations: Case of marine mussels Mytilus. Oceanologia, 61(1), 131–142. https://doi.org/10.1016/j.oceano.2018.08.002

Kijewski, T. K., Zbawicka, M., Väinölä, R., & Wenné, R. (2006). Introgression and mitochondrial DNA heteroplasmy in the Baltic populations of mussels Mytilus trossulus and M. edulis. Marine Biology, 149(6), 1371–1385. https://doi.org/10.1007/s00227-006-0316-2

Koehn, R. K. (1991). The genetics and taxonomy of species in the genus Mytilus. Aquaculture, 94(2–3), 125–145. https://doi.org/10.1016/0044-8486(91)90114-M

Lachance, J., & Tishkoff, S. A. (2013). SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it. BioEssays, 35(9), 780–786. https://doi.org/10.1002/bies.201300014

Mathiesen, S. S., Thyrring, J., Hemmer-Hansen, J., Berge, J., Sukhotin, A., Leopold, P., ... Nielsen, E. E. (2017). Genetic diversity and connectivity within Mytilus spp. in the subarctic and Arctic. Evolutionary Applications, 10(1), 39–55.

Michalek, K., Ventura, A., & Sanders, T. (2016). Mytilus hybridisation and impact on aquaculture: A minireview. Marine Genomics, 27, 3–7. https://doi.org/10.1016/j.margen.2016.04.008

Paradis, E., & Schliep, K. (2018). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics, 35(3), 526–528.

Paris, J., Stevens, J., & Catchen, J. (2017). Lost in parameter space: A road map for STACKS. Methods in Ecology and Evolution, 8(10), 1360–1373.

Penney, R. W., Hart, M. J., & Templeman, N. D. (2007). Shell strength and appearance in cultured blue mussels Mytilus edulis, M. trossulus, and M. edulis × M. trossulus hybrids. North American Journal of Aquaculture, 69, 281–295.

Penney, R. W., Hart, M. J., & Templeman, N. D. (2008). Genotype-dependent variability in somatic tissue and shell weights and its effect on meat yield in mixed species [Mytilus edulis L., M. trossulus (Gould), and their hybrids] cultured mussel populations. Journal of Shellfish Research, 27(4), 827–835.

Pujolar, J. M., Maes, G. E., Vancollie, C., & Volckaert, F. A. M. (2005). Growth rate correlates to individual heterozygosity in the European eel. Anguilla Anguilla L. Evolution, 59(1), 189–199.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics, 81(3), 559–575. https://doi.org/10.1086/519795

Riginos, C., & Cunningham, C. W. (2005). Invited review: Local adaptation and species segregation in two mussel (Mytilus edulis × Mytilus trossulus) hybrid zones. Molecular Ecology, 14(2), 381–400.

Rochette, N., & Catchen, J. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. Nature Protocols, 12, 2640–2659.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). Molecular cloning: A laboratory manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Sanjuan, A., Zapata, C., & Alvarez, G. (1994). Mytilus galloprovincialis and M. edulis on the coasts of the Iberian Peninsula. Marine Ecology Progress Series, 131–146. https://doi.org/10.3354/meps113131

Scott, D., McLead, D., Young, J., Brown, J., Immink, A., & Bostock, J. (2010). A study of the prospects and opportunities for shellfish farming in Scotland.

Scottish Government (2014). The aquaculture and fisheries (Scotland) Act 2013 (Specification of commercially damaging species) Order, 0214.

Skibinski, D. O. F. (1983). Natural selection in hybrid mussel populations. Protein Polymorphism: Adaptive and Taxonomic Significance, 24, 283–298.
Skibinski, D. O. F., Beardmore, J. A., & Cross, T. F. (1983). Aspects of the population genetics of Mytilus (Mytilidae; Mollusca) in the British Isles. *Biological Journal of the Linnean Society*, 19(2), 137–183. https://doi.org/10.1111/j.1095-8312.1983.tb00782.x

Slate, J., Kruuk, L. E. B., Marshall, T. C., Pemberton, J. M., & Clutton-Brock, T. H. (2000). Inbreeding depression influences lifetime breeding success in a wild population of red deer (*Cervus elaphus*). *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1453), 1657–1662.

Śmietanka, B., Zbawicka, M., Wolowicz, M., & Wenne, R. (2004). Mitochondrial DNA lineages in the European populations of mussels (*Mytilus spp.*). *Marine Biology*, 146(1), 79–92. https://doi.org/10.1007/s00227-004-1418-3

Stoffel, M. A., Esser, M., Kardos, M., Humble, E., Nichols, H., David, P., & Hoffman, J. I. (2016). inbreedR: An R package for the analysis of inbreeding based on genetic markers. *Methods in Ecology and Evolution*, 7(11), 1331–1339.

Wilkinson, L.. (2011). *venneuler: Venn and Euler Diagrams*. R package version 1.1-0. https://CRAN.R-project.org/package=venneuler

Zbawicka, M., Sańko, T., Strand, J., & Wenne, R. (2014). New SNP markers reveal largely concordant clinal variation across the hybrid zone between *Mytilus* spp. in the Baltic Sea. *Aquatic Biology*, 21(1), 25–36. https://doi.org/10.3354/ab00566.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section.

*How to cite this article:* Vendrami DLJ, De Noia M, Telesca L, Brodte E-M, Hoffman JI. Genome-wide insights into introgression and its consequences for genome-wide heterozygosity in the *Mytilus* species complex across Europe. *Evol Appl*. 2020;13:2130–2142. https://doi.org/10.1111/eva.12974