Gene-flow in a mosaic hybrid zone: is local introgression adaptive?

Christelle Fraïsse*, Camille Roux*, John J. Welch†, Nicolas Bierne*

April 29, 2014

*Université Montpellier 2, Place E. Bataillon, 34095 Montpellier cedex 5, France.

§CNRS, Institut des Sciences de l’Evolution (ISEM UMR 5554), Station Méditerranéenne de l’Environnement Littoral, 2 rue des Chantiers, 34200 SETE, France.

†Department of Genetics, University of Cambridge, Downing Street, CB2 3EH Cambridge, UK.

‡Department of Ecology and Evolution, Lausanne University, Biopheore/Sorge, CH-1015 Lausanne, Switzerland.
Running title:

Local introgression in a mosaic hybrid zone

Key words:

outlier loci; soft sweep; approximate Bayesian computation; maximum likelihood; Mytilus

Author for correspondence: Christelle Fraïsse; christelle.fraisse@univ-montp2.fr; SMEL, 2 rue des Chantiers, 34200 Sète, France
Genomewide scans of genetic differentiation between hybridizing taxa can identify genome regions with unusual rates of introgression. Regions of high differentiation might represent barriers to gene flow, while regions of low differentiation might indicate adaptive introgression - the spread of selectively beneficial alleles between reproductively isolated genetic backgrounds. Here we conduct a scan for unusual patterns of differentiation in a mosaic hybrid zone between two mussel species, *Mytilus edulis* and *Mytilus galloprovincialis*. One outlying locus, *mac-1*, showed a characteristic footprint of local introgression, with abnormally high frequency of *edulis*-derived alleles in a patch of *M. galloprovincialis* enclosed within the mosaic zone, but low frequencies outside of the zone. Further analysis of DNA sequences showed that almost all of the *edulis* allelic diversity had introgressed into the *M. galloprovincialis* background in this patch. We then used a variety of approaches to test the hypothesis that there had been adaptive introgression at *mac-1*. Simulations and model fitting with Maximum Likelihood and Approximate Bayesian Computation approaches suggested that adaptive introgression could generate a “soft sweep”, which was qualitatively consistent with our data. Although the migration rate required was high, it was compatible with the functioning of an effective barrier to gene flow as revealed by demographic inferences. As such, adaptive introgression could explain both the reduced intraspecific differentiation around *mac-1* and the high diversity of introgressed alleles, although a localised change in barrier strength may also be invoked. Together, our results emphasize the need to account for the complex history of secondary contacts in interpreting outlier loci.
Genetic barriers between related lineages are often semi-permeable, varying in strength across the genome (Harrison 1986). As such, hybridization can lead to adaptive introgression, the transfer of selectively beneficial alleles between species (Parsons et al. 1993; Arnold 2004; The Heliconius Genome Consortium 2012; Hedrick 2013). However, the frequency and importance of such hybridization events remain hotly debated (Mallet 2007; Abbott et al., 2013; Barton 2013). It is now common to address such questions by scanning genome sequences for regions of enhanced or reduced differentiation. Universally advantageous alleles can usually cross barriers to gene flow without much delay (e.g. Pialek and Barton 1997), and neutral loci linked to such alleles are expected to display unusually low levels of differentiation. In contrast, barrier loci will delay the introgression of neutral alleles in proportion to their linkage (Barton 1979; Barton and Bengtsson 1986); that is the basis for the investigation of genomic islands of differentiation (Turner et al. 2005; Hohenlohe et al. 2010; Nadeau et al. 2012).

Interpreting regions of unusual differentiation can be difficult, not only because of variation in recombination across the genome (Nachman and Payseur 2012; Roesti 2013), but also because reduced differentiation might be caused by processes other than adaptive introgression. In particular, population history may promote multiple contacts between the same lineages in different locations, and so the resulting barriers to gene flow might vary from place to place, according to local ecological gradients and population connectivity. In accordance with this hypothesis, reversed or modified associations between genetic differentiation and habitat variation (Bierne et al. 2011; Jackson et al. 2012), and partial parallelism in genomic divergence (Gagnaire et al. 2013) have both been observed in replicated contact zones of the same species pair. Differentiating between these hypotheses is challenging, because adaptive introgression leaves a genomic signature that can be substantially different from the classic “hard sweep” (Maynard Smith and Haigh 1974) where adaptive substitution is associated with a single new mutation. Adaptive introgression may involve multiple
migrant copies of the same beneficial allele (Penningss and Hermisson 2006) leading to a “soft-sweep” signature which can be difficult to discriminate from a local increase in the neutral introgression rate (see Table 1).

Here we present a multilocus scan for local introgression between the marine mussels *Mytilus edulis* and *Mytilus galloprovincialis*. These species appear to have had a complex history of fragmentation and colonization during the Quaternary, as it is the case for almost all temperate organisms (Hewitt 2011). Recent analyses suggest that *M. edulis* and *M. galloprovincialis* diverged ~2.5 Ma, followed by a secondary contact beginning about 0.7 Ma (Roux et al. 2014). Today, they are isolated by multifarious pre- and post-zygotic mechanisms (Bierne et al. 2002; Bierne et al. 2003a; Bierne et al. 2006). Due to natural replications of contact zones, the *Mytilus* species complex is particularly suitable for studying the different possible outcomes of secondary contact. Here we study an original mosaic hybrid zone in Europe, extending from the Mediterranean Sea to the North Sea (Bierne et al. 2003b; Hilbish et al. 2012) and to British Isles (Skibinski et al. 1983). We focus on the French section of the mosaic zone that includes two patches per species, one peripheral and one enclosed within the zone (Figure 1).

To study introgression between these populations, we used previously published data at 440 loci (422 Amplified Fragment Length Polymorphism (AFLP) and 18 codominant nuclear markers, Gosset and Bierne 2012). Initial analyses identified a nuclear marker, *mac-1*, with high interspecific differentiation in the peripheral patch but an anomalously low level of differentiation in the enclosed patch in Brittany. This outlier locus was therefore a candidate for adaptive introgression in a geographically localised region.

To test the plausibility of this hypothesis, we first sequenced a 3.1 Kb region around *mac-1* and analysed how allele frequency varies along the chromosomal region. Unlike another *M. edulis* locus analyzed previously (Bierne 2010), we failed to identify variation in differentiation on such a small chromosomal scale. We then determined how well our data might be explained by a model of adaptive introgression (Penningss and Hermisson 2006). Infer-
ences indicated that the level of gene flow implied by the adaptive introgression hypothesis was high, but still compatible with the strength of the genetic barrier between the two species. We further propose that the barrier to gene flow in the vicinity of \textit{mac-1} could be simply more permeable in certain hybrid zones of the mosaic. Overall, we show that although the molecular signature of adaptive introgression is difficult to identify, dedicated methods that account for the history of speciation can help in interpreting outlying levels of introgression.

**MATERIALS AND METHODS**

**STUDY SITES AND SAMPLING**

The zone of study, along the European coast, is characterized by three successive transitions between panmictic patches of each species (Figure 1, BIERNE \textit{et al.} 2003c; HILBISH \textit{et al.} 2012). We used two geographical samples of \textit{M. edulis}: the peripheral patch of the North Sea (Wadden Sea, Holland); and the enclosed patch of the Bay of Biscay (Lupin, France). We also used two geographical samples of \textit{M. galloprovincialis}: the peripheral patch of the Iberian Coast (Faro, Portugal); and the enclosed patch of Brittany (Roscoff, France). These four samples have been described in FAURE \textit{et al.} (2008), and have been established to be representative of monospecific panmictic patches. In addition, we used a sample of \textit{Mytilus trossulus} (Tadoussac, Canada) to serve as an outgroup. 48 individuals per sample were examined, except for the Brittany sample which comprised 87 individuals. Genomic DNA was extracted from adults using the DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's protocol.

**MULTILOCUS SCAN**

For the multilocus scan, we combined existing data, comprising 422 AFLP markers and 13 codominant nuclear markers (GOSSET AND BIERNE 2012) and 5 allozyme markers (FAURE \textit{et al.} 2008). To combine loci with various level of diversity, nuclear codominant
markers were transformed into biallelic loci by pooling alleles according to their frequencies in the *M. galloprovincialis* and *M. edulis* reference samples (see McDonald 1994).

AFLP markers are subject to several well known caveats, notably fragment size homoplasy (Caballero et al. 2008; Whitlock et al. 2008). To reduce this problem, we excluded AFLP bands smaller than 50bp, which are most prone to homoplasy (Vekemans et al. 2002). Allele frequencies of AFLP markers were estimated with the Bayesian method of Zhivotovsky (1999) in AFLPsurv v1.0 (Vekemans et al. 2002). All data were combined for the outlier tests.

To ensure robustness, and following the recommendation of Pérez-Figueroa et al. (2010), we used six distinct methods to identify loci with unusual levels of genetic differentiation. These are the methods of Lewontin and Krakauer (1973), Beaumont and Nichols (1996), Vitalis et al. (2003), Beaumont and Balding (2004), Bonhomme et al. (2010) and a custom simulation method in which the neutral envelope of $F_{ST}$ is obtained from simulations with the parameter estimates of the best-supported demographic model. Full details are given in the Supp. Info. (File S1).

**DATA ANALYSES AT THE OUTLIER LOCUS MAC-1**

**Chromosomal walking along mac-1**

We sequenced a region of 3.1 kb around mac-1 taking advantage of published sequences (Daguin et al. 2001), an additional upstream sequence (M. Ohresser, pers. comm.) and looking for homology between coding regions and Expressed Sequence Tags of *Mytilus* (Tanguy et al. 2008). To describe the local variation of the *edulis* allele frequency around mac-1, we walked along the sequence and looked for evenly spaced genetic polymorphisms (see Figure S2 for a schematic representation). Four polymorphisms, two PCR product length polymorphisms and two SNPs, were chosen, based on the allelic genealogies, to be discriminant between the two species. We extensively genotyped our four samples (North Sea, Brit-
tany, Bay of Biscay and Iberian Coast) for these four polymorphisms. See Supp. Info. (File S1) for technical details.

**Genetic analysis**

For the PCR product length polymorphisms described above, we computed classical genetic statistics on the subset of *edulis* alleles (i.e., alleles that group within the diversity of *M. edulis* within the *mac-1* genealogy, see Figure 3) using Genetix 4.05.2 (Belkhir et al. 1999): the number of allelic classes (*k*), the heterozygosity (*He*, Nei 1978), departure from Hardy-Weinberg equilibrium and genetic differentiation were estimated by *f* and *Θ* (Weir and Cockerham 1984) respectively, and significance was tested with 1000 permutations. Finally, we evaluated the departure from the neutral expectation at mutation-drift equilibrium with the Ewens test (*ε*, Slatkin 1994; Slatkin 1996).

Sequence alignment was performed with Multalin (Corpet 1988), and verified by eye in BioEdit 7.5.3 (Hall 1999). Alignment gaps were excluded from further analyses. We inferred allelic genealogies for the three regions sequenced using the neighbour-joining algorithm implemented in Mega 5.0 (Kumar et al. 2004). We rooted the allele genealogy with the outgroup *M. trossulus*. We computed classical genetic statistics on DNA sequences using DNAsp (Rozas et al. 2003): the number of polymorphic sites (*S*), the number of synonymous (*S.s*), non-synonymous (*S.ns*) and non-coding (*S.nc*) mutations, levels of nucleotide diversity estimated from the number of polymorphic sites (*θ*<sub>S</sub>, Watterson 1975) and from pairwise differences (*θ*<sub>π</sub>, Tajima 1983) and the minimum number of recombination events (*Rm*, Hudson and Kaplan 1985). We also computed two indicators of the distortion of the allele frequency spectrum from the neutral mutation-drift equilibrium expectation: (1) Tajima’s *D* (Tajima 1989a; 1989b), and (2) Fay and Wu’s *H* (Fay and Wu 2000); significant departure from standard neutral model were assessed using standard-coalescent simulations.

**Model of adaptive introgression**
To assess the hypothesis of adaptive introgression at *mac-1*, we considered a model of adaptive introgression via recurrent migration, introduced by Pennings and Hermisson (2006).

This model considers a universally beneficial allele which introgresses from the donor population(s) to the recipient population via recurrent migration, and then sweeps to global fixation. Formally, the model considers two Wright-Fisher populations, each of *N* individuals. Each individual is characterised by two completely linked haploid loci. The first locus is biallelic, with alleles *B* and *b*. The *B* allele has a selective advantage *s* over the *b* allele. The second, linked locus, is neutral with an infinite number of possible alleles, and new mutations arising at rate *µ* (the per gene mutation rate). The donor population is initially fixed for the beneficial *B* allele, while the recipient population is initially fixed for the alternative *b* allele, and both populations are assumed to have reached mutation-drift equilibrium at the neutral locus. After equilibration, follows a period of secondary contact in which the populations exchange migrants at rate *m*. During this secondary contact, the beneficial *B* allele is introduced into the recipient population, where it will eventually reach fixation. The three most important parameters of this model are the scaled quantities $\theta = 2N\mu$, $M = 2Nm$ and $\alpha = 2Ns$ (Pennings and Hermisson 2006).

To fit this model to data, we used the *mac-1* alleles sampled from the donor populations (the two *M. edulis* patches) and the recipient population (the Brittany patch of *M. galloprovincialis*). Analytical results assume complete linkage, and recent completion of the sweep (see below), and so, to make our inferences robust to the violation of these assumptions, we take as data only those *M. galloprovincialis* alleles that group within the diversity of *M. edulis* within the genealogy (i.e., only the alleles denoted with white circles in the *edulis* clade, Figure 3).

**Maximum Likelihood approach:** We first fit the data using an ML approach. Our data in this case comprised the number of alleles sampled in the donor and recipient populations ($n_d$ and $n_r$) and the number of distinct allelic classes in these samples ($k_d$ and $k_r$). Under the
assumptions described above, the probability of observing the data from the donor population
is given by the Ewens’ sampling formula (Ewens 1972):

\[ P(k_d | n_d, \theta) = \frac{\theta^{k_d} S_{n_d}^{k_d}}{\theta \times (\theta + 1) \times \ldots \times (\theta + n_d - 1)} = \frac{\Gamma(\theta) \theta^{k_d} S_{n_d}^{k_d}}{\Gamma(\theta + n_d)} \]

where \( S_{n}^{k} \) is an unsigned Stirling number of the first kind, i.e., the number of permutations
of \( n \) elements in \( k \) disjoint cycles (Charalambides and Singh 1988). Penning and
Hermisson (2006) showed that adaptive introgression has some formal similarities to eq. (1). In particular, if \( \alpha = 2Ns \gg 1 \), then after the beneficial allele has reached fixation in the
donor population, the number of distinct immigrant alleles that contribute to the sweep will
also approximate a Ewens’ distribution, but with the scaled migration rate, \( M \), playing the
role of the scaled mutation rate, \( \theta \) (Penning and Hermisson 2006). We cannot observe
the number of immigrant alleles directly, but we do know that these migrants were drawn at
random from the donor population, and that the diversity in the donor population is described
by Ewens’ distribution with parameter \( \theta \). As such, the likelihood of observing our data is
well approximated by:

\[ L(\theta, M | k_d, k_r, n_d, n_r) \approx P(k_d | n_d, \theta) \sum_{i=k_r}^{n_r} P(i | n_r, M) P(k_r | i, \theta) \]

\[ = \frac{\Gamma(M) \Gamma(\theta) \theta^{k_d+k_r} S_{n_d}^{k_d} S_{n_r}^{k_r}}{\Gamma(M+n_r) \Gamma(\theta+n_d)} \sum_{i=k_r}^{n_r} \frac{\Gamma(\theta) M^{i} S_{n_r}^{i} S_{i}^{k_r}}{\Gamma(\theta+i)} \]

(where the summation is over the number of possible migrants from the donor population
which appear in the sample from the recipient population). Note that the approximation of
eq (2) does not depend explicitly on the strength of selection, \( \alpha \), but this new result does
allow us to estimate the parameters \( \theta \) and \( M \) from our mac-l data. The parameters which
maximise eq. (2), \( \hat{\theta} \) and \( \hat{M} \), are the Maximum Likelihood Estimates. Confidence intervals
on a given parameter were the values that reduced the log likelihood by 2 units - obtained
by maximizing the likelihood, conditional on the parameter of interest taking a suboptimal
value (EDWARDS 1992). An R script that implements eq. (2) is included as supplementary material (see Supp. Info., File S2).

**Approximate Bayesian Computation approach:** Eq. (2) is an approximation (PENNINGS AND HERMISSON 2006), furthermore, it considers only the number of allelic classes \((k_d\) and \(k_r\)), and therefore ignores some of the information in the data, such as the allele frequency spectrum. Accordingly, we also estimated model parameters using an Approximate Bayesian Computation approach (BEAUMONT et al. 2002). This approach uses forward simulation of the adaptive introgression model, and then compares summary statistics of the simulated and observed data.

To simulate the model, mutation-drift equilibrium was first achieved by sampling directly from the full EWENS (1972) distribution, as implemented in the program MONTECARLO (SLATKIN 1994). Then, we ran forward-in-time simulations of the secondary contact, ending each simulation when the beneficial allele had reached fixation in the recipient population.

To compare the fit of the simulations to our data, we used several summary statistics. First, for each of the two species where \(d\) denotes the donor species and \(r\) denotes the recipient species, we calculated (1) the number of allelic classes \((k_d\) and \(k_r\)), (2) the expected heterozygosity at the locus \((H_{e_d}\) and \(H_{e_r};\) NEI 1978), (3) the standard deviation in frequencies over all allelic classes \((S_{d_d}\) and \(S_{d_r}\)) and (4) the proportion of private alleles (not found in the other species, \(\%P_{d}\) and \(\%P_{r}\)). We also used (5) the between-species diversity \(F_{ST}\) (NEI 1973) and (6) the proportional diversity after introgression described by \(H_{e_r}/H_{e_d}\). All of these summary statistics for the true data are given in Table S1.

To estimate the parameters, we performed 950,000 forward simulations in total. For each simulation, parameters were log-tangent transformed (HAMILTON et al. 2005), and we then calculated the Euclidean distance between the observed and simulated transformed parameters. The 1000 simulations with the smallest associated Euclidean distance were then retained, and the posterior distribution of the scaled parameters estimated by means of weighted non-linear multivariate regressions of the parameters on the summary statistics (BLUM AND...
For each regression, 50 feed-forward neural networks and 25 hidden networks were trained using the R package *abc* (Csilléry et al. 2012) and results averaged over the replicate networks. Priors on the model parameters were all uniform (flat), and used the following ranges: \( \theta \sim [0.2, 40] \), \( M \sim [0.2, 200] \), \( \alpha \sim [2, 2000] \).

We evaluated the performance of the ABC method in two ways. First, we performed a goodness-of-fit test (Cornuet et al. 2010). This involved simulating 50,000 datasets with parameters drawn from their estimated posterior distributions, and then comparing the summary statistics of these simulated data sets to the statistics of the observed data. Second, we computed the mean bias statistic, \( b = 1/n \sum [e_i/v_i] \) where \( n \) is the number of summary statistics, \( e_i \) is the median estimation and \( v_i \) is the true median value of the \( i^\text{th} \) dataset (Excoffier et al. 2005). We generated 100 pseudo-observed datasets with known parameter values drawn from the prior distributions and computed \( b \) for each scaled-parameter.

Forward simulations of adaptive introgression written in C++ are available as supplementary material (see Supp. Info., File S2) together with R scripts implementing a simple version of the ABC approach.

**Models of speciation**

To compare our estimate of \( M \) at *mac-1* to the extent of gene flow between the Brittany patch of *M. galloprovincialis* and *M. edulis*, we took advantage of the ABC approach described in Roux et al. (2013) that accounts for a putative genome-wide heterogeneity in introgression rates. We obtained DNA sequence data at the eight nuclear loci described in Roux et al. (2014) for the North Sea patch of *M. edulis* and the Brittany patch of *M. galloprovincialis*. Only silent positions (i.e., synonymous polymorphisms in coding regions and non-coding polymorphisms in introns or intergenic regions) were retained in the analysis.

We investigated four models of speciation differentiated by their temporal patterns of introgression (see Figure S8): (1) Strict Isolation (SI) between the two daughter species, (2) Isolation with Migration (IM) assuming continuous gene flow since the two species started
to diverge; (3) Ancient Migration (AM) where migration is restricted to the early period of speciation; and (4) Secondary Contact (SC) where the two daughter populations first evolve in strict isolation and then experience gene flow in a secondary contact. For models with gene flow (IM, AM and SC), we compared two alternative scenarios in which the effective migration rate was either homogeneous or heterogeneous among loci.

Five million multilocus simulations were performed for each model using the coalescent simulator Msnsam (Hudson 2002; Ross-Ibarra et al. 2008). We then compared the simulated and observed datasets by using an array of summary statistics widely used in the literature (Wakeley and Hey 1997; Fagundes et al. 2007; Ross-Ibarra et al. 2008) (1) nucleotide diversity ($\pi$, Tajima 1983); (2) Watterson’s $\theta_w$ (Watterson 1975); (3) net interspecific divergence (netdiv$_{edu-gal}$) and (4) between-species differentiation computed as $1 - \pi_S / \pi_T$, where $\pi_S$ is the average pairwise nucleotide diversity within species and $\pi_T$ is the total pairwise nucleotide diversity of the pooled sample across species. We assessed departure from mutation/drift equilibrium using Tajima’s $D$ (Tajima 1989a; 1989b). We also classified the variable genomic positions according whether they are exclusively polymorphic in $M. edulis$ ($S_{x.edu}$); exclusively polymorphic in $M. galloprovincialis$ ($S_{x.gal}$); with different alleles fixed in both species ($S_f$) or sharing the same polymorphism ($S_s$). We computed the average and standard deviation of each of these statistics across the surveyed loci by using MScalc (available from http://www.abcgwh.sitew.ch/, see Roux et al. 2011). We provide the observed values computed from the sequenced data in Table S2.

Full detail of priors, model choice and model checking procedures, as well as parameter estimation are described in the Supp. Info. (File S1).

RESULTS

DETECTION OF INTERSPECIFIC AND INTRASPECIFIC OUTLIERS

Figure 2A compares interspecific and intraspecific differentiation at screened loci. The
corresponding figure in *M. edulis* is provided in Supp. Info. (see Figure S1). Differentiation between patches of the same species was generally low (in *M. galloprovincialis* $F_{ST} = 0.021$; in *M. edulis* $F_{ST} = 0.024$), but highly variable between loci, and a few loci were identified as outliers using several methods of detection. In particular, all methods agreed that three loci were outliers within *M. galloprovincialis* (ACG.CGA.206 $F_{ST} = 0.189$; mac-1 $F_{ST} = 0.192$ and CAG.CTC.317 $F_{ST} = 0.224$, Figure 2A), and two loci within *M. edulis* (ACG.CGA.152 $F_{ST} = 0.294$ and CAG.CGA.363 $F_{ST} = 0.374$, Figure S1). In the comparison between species, differentiation was also low ($F_{ST} = 0.068$) but the distribution of $F_{ST}$ values was overdispersed, as one would expect when a semipermeable barrier to gene flow isolates two species. The locus mac-1, which was highly differentiated within *M. galloprovincialis*, was also highly differentiated between species ($F_{ST} = 0.507$) and was the only locus to be detected as an outlier in both kinds of comparison.

To see this in more detail, Figure 2B shows allele frequencies in all four patches at the three loci identified as outliers within *M. galloprovincialis*. While all three loci show allele frequency differences within *M. galloprovincialis*, only mac-1 shows allele frequency differences between the Brittany patch of *M. galloprovincialis* and the neighboring patches of *M. edulis*.

**DETAILED ANALYSIS OF MAC-1**

Because the pattern of differentiation at mac-1 was unique, this locus was investigated in more detail. Specifically, we sequenced three regions, comprising a total of 3.1 Kb in the vicinity of mac-1 (Figure S2). All three regions contained several SNPs, and the first region also contained indels that were used to define PCR primers which generate length polymorphism of the PCR product, *indel-in0*. Summary statistics of the genetic diversity within each sequenced region are found in Table 2; while Table 3 describes the genetic diversity of the two PCR product length polymorphisms.

We next constructed genealogies of the three regions. Figure 3 shows the genealogy of region 2 where alleles from the Brittany patch of *M. galloprovincialis* are found in two distinct
clades, grouping with (1) alleles from both *M. edulis* patches (“edulis clade”, Figure 3) and (2) with the other *galloprovincialis* alleles (“galloprovincialis clade”, Figure 3). Similar patterns are also found in regions 1 and 3 (Figure S3), and were robust to the exclusion of possible recombinants (Table 2).

Together, these patterns are consistent with the results of *F*<sub>ST</sub> outlier tests and suggest that there has been an introgression of mac-1 alleles from *M. edulis* into the Brittany patch of *M. galloprovincialis*, which lies between them. Furthermore, comparison of the Brittany alleles that group in the edulis clade (white circles within the “edulis clade”, Figure 3), with the related alleles found in the two *M. edulis* patches (squares and circles in black, Figure 3), suggests that very similar levels of genetic diversity are found in all three patches (see Tables 2 and 3). For example, the proportion of diversity introgressed at the locus indel-in0 is *H*<sub>gal</sub>/ *H*<sub>edu</sub> = 0.67 and at mac-1 is *H*<sub>gal</sub>/ *H*<sub>edu</sub> = 0.96 (see Table 3). Additionally, there is no evidence of departure from neutrality in any of the three sets of alleles (see the results of Tajima’s *D* and Fay and Wu’s *H* in Table 2).

One possible interpretation of these results is that mac-1 has been subject to adaptive introgression from *M. edulis* into Brittany patch of *M. galloprovincialis*. The following sections use several methods to test the validity of this hypothesis.

**TESTING FOR THE CHROMOSOMAL SIGNATURE OF ADAPTIVE INTROGRESSION**

One genomic signature of local adaptation is a single sharp peak in the frequencies of linked neutral alleles, centered on the adaptive polymorphism (Charlesworth *et al.* 1997). To test for this signature we identified four polymorphisms, two indels and two SNPs, that mapped to an internal branch of their corresponding genealogy (Figure 3 for region 2, and Figure S3 for regions 1 and 3). These polymorphisms were evenly spaced along the 3.1 Kb sequence, and their locations are shown in Figure S2 (triangles for indels and stars for SNPs). Figure 4 shows the variation of the edulis allele along the sequence in the four patches.
Whereas the frequency of the *edulis* allele remains consistently low in the Iberian Coast patch \((F_{q_{edu}} < 0.05)\), and consistently high in the North Sea and Bay of Biscay patches \((F_{q_{edu}} > 0.95)\), we noted a slight increase of the *edulis* allele toward the 5' side of *mac-1* in the Brittany patch. However, there was a surprising lack of variation along the whole sequence: the *edulis* allele frequency remains around 0.4 over several Kb.

**MODEL FITTING OF ADAPTIVE INTROGRESSION**

The limited scale of our chromosomal walking does not allow us to reject any hypotheses with high confidence. Accordingly, we next tested whether adaptive introgression might plausibly explain the pattern of diversity observed at *mac-1*.

First, to illustrate the effects of adaptive introgression on diversity, Figure 5 plots the expected reduction in diversity against the rate of gene flow (see also PENNINGS AND HERMISSON 2006). When \(M < 0.01\) we see a clear “hard sweep” with a single haplotype reaching high frequency in the recipient deme, while when \(M > 1\), most of the heterospecific diversity introgresses into the recipient background (a classic “soft sweep” with no detectable reduction in diversity). For intermediate values of \(M\), the effect is also intermediate, with both a soft sweep and a reduction in diversity.

Building on the results of PENNINGS AND HERMISSON (2006), we fitted a model of adaptive introgression to our *mac-1* data. We first developed an ML approach to estimate the population mutation rate, \(\theta = 2N\mu\), in the *M. edulis* populations (North Sea and Bay of Biscay), and the effective number of migrants, \(M = 2Nm\), from these populations to the Brittany patch of *M. galloprovincialis* (see eq. (2)). Considering only *edulis*-derived alleles, the total number of alleles sampled, and the number of distinct allelic classes observed were \(n_d = 177\) and \(k_d = 15\) for the North Sea and Bay of Biscay patches, \(n_r = 60\) and \(k_r = 6\) for the Brittany patch (Table 3). From these data, we obtained a relatively precise estimate of the population mutation rate: \(\hat{\theta} = 3.79 [1.95 – 6.66]\) (see black lines, Figure 6A). However, the effective number of migrants was imprecisely estimated, with low migration rates rejected
but no upper bound: $\hat{M} = 4.63 [0.95- ]$ (see black lines, Figure 6B). The lack of an upper bound on the estimate reflects the fact that nearly all of the allelic diversity from the edulis patches had been introgressed into the Brittany patch of *M. galloprovincialis*.

In an attempt to improve our inference, we next used an ABC approach, which allowed us to relax the assumption of strong selection, and to consider more of the information in the data. Parameter posterior distributions, corrected for mean bias, are shown in Figure 6 as grey dots. In the following, we give their median, and 95% credible intervals. We obtained an even more precise estimate of the scaled mutation rate (Figure 6A): $\theta_{\text{median}} = 2.82 [1.42 - 4.53]$ whose bounds are similar to the ML estimates. Estimation of the scaled migration rate (Figure 6B): $M_{\text{median}} = 99 [24 - 154]$ showed a clear discrepancy between the two methods with respect to the lower bound. Compared to ML, the lower bound estimation was $M \gg 1$ while its upper bound was poorly estimated (value approximates the prior upper bound). Finally, the estimation of selective strength, $\alpha_{\text{median}} = 402 [203 - 594]$ (Figure 6C) was imprecise but compatible with the assumption of strong selection used in the ML inference.

To better understand the discrepancy between the two methods, we reran the ABC estimation, using only the summary statistics which appear in the ML equations (namely, $n_d$, $k_d$, $n_r$ and $k_r$). In this case, estimates of $\theta$ and $M$ were very similar to the MLEs (Figure S4), and there was no information on $\alpha$ as expected. The results of the ABC approach therefore come from the additional use of summary statistics that contain information which is not used in the ML approach. We therefore conducted a goodness-of-fit test in order to evaluate how well the model fit the data at *mac-1* when estimating the parameters by each method (ML and ABC using all statistics). Figure S5 shows the goodness-of-fit of ML estimation (blue dots) and ABC estimation (red dots) for each statistic. In the main, both methods fit the data well (see Supp. Info., Table S1 for quantiles and $p$-values) but looking at the median values, ML performs better. In particular, ABC tends to overestimate $k_r$ and $He_r$ which basically explains the discrepancy in the estimation of $M$. The ML method proved to be the most understand-able approach with a simpler model that retrieved the best fit to the data and consequently
our interpretations will rely on the ML estimates.

Finally, we used our simulations to assess the effects of violating some assumptions of our model: (1) the assumption of complete linkage between the selected and neutral loci and (2) the assumption that migration rates remained constant. Figure S6 shows that the sweep pattern was indeed robust to including recombination. Figure S7 suggests that our results will also be fairly robust to temporal variation in migration rates, particularly in mussels where dispersal rates are very high (see figure legend for full details).

INSIGHTS INTO SECONDARY CONTACT HISTORY

Both estimation methods agree that $M > 1$ best explains our data if adaptive introgression has taken place. To assess the plausibility of our estimation, we analysed several models of speciation (Figure S8) between *M. edulis* and the Brittany patch of *M. galloprovincialis*. We first applied a model choice procedure for each of the three speciation models with gene flow to test whether heterogeneity in introgression rates was supported. We observed clear support for heterogeneous migration (Table S3, “within model”). We then applied the model choice procedure to the speciation models, assuming heterogeneous migration (Table S3, “between-models”). Secondary Contact was the best-supported model with posterior probability, $P_{\text{post}} = 0.61$; confirming a previous study (Roux et al. 2014). We tested the robustness of this result by simulating 500 pseudo-observed datasets for each model. We empirically estimated that given a posterior probability of 0.61, the probability that Secondary Contact with heterogeneous introgression rates is the correct model was 0.987 (Table S4). It is worth mentioning that when migration rate was assumed to be homogeneous among loci, the method failed to infer Secondary Contact; instead Ancient Migration was the best-supported model (SC: $P_{\text{post}} = 0.26$ and AM: $P_{\text{post}} = 0.65$).

We then estimated the multilocus distributions of introgression rates in the best-supported SC model with heterogeneous migration (Figure 7). We found an asymmetry of introgression in favor of *M. edulis*: $M_{\text{gallo to edulis}} = 0.15 \ [0.002 - 1.06]$ (see white bars, Figure 7)
and $M_{\text{edulis to gallo}} = 0.45 \ [0.013 - 2.84]$ (see grey bars, Figure 7). However, compared to the multilocus inference, introgression at $\text{mac-1}$ was strongly unidirectional from $M. \text{edulis}$ to $M. \text{galloprovincialis}$ as suggested by the comparison of different scenarios of introgression (SC with introgression into $M. \text{edulis}$: $P_{\text{post}} = 0.0013$; SC with introgression into $M. \text{galloprovincialis}$: $P_{\text{post}} = 0.9997$, see Table S5). Furthermore introgression rate into $M. \text{galloprovincialis}$ at $\text{mac-1}$ was very high compared to the extent of gene flow across loci ($M_{\text{median}} = 2.50$, grey solid line, Figure 7); that is consistent with the $\text{mac-1}$ genealogy (Figure 3). In comparison, our ML estimation of $M$ at $\text{mac-1}$ in the adaptive introgression model ($\hat{M} = 4.63$, Figure 6B) was outside of the 95%-quantile of the multilocus distribution, but its lower bound ($M_{\text{ML lower bound}} = 0.95$, grey dashed line, Figure 7) remained consistent with the Secondary Contact inference.

DISCUSSION

Adaptive introgression between hybridizing species has long been proposed as a potentially important source of new adaptations in plants, and more recently, in animals too (Arnold 2004; Hedrick 2013). Most of all, compared to adaptation from new mutations or standing variation, adaptive introgression might allow big evolutionary jumps by the acquisition of complex variants (Wright 1949; Mallet 2007; Abbott et al. 2013).

Genome scan surveys of differential introgression among loci can help to localize genomic regions that introgress adaptively (Payseur 2010). Here, we took advantage of the mosaic structure of the hybrid zone of mussels along the western European coastline, to perform a multilocus scan of local introgression. As pointed out by Harrison and Rand (1989), an exciting feature of mosaic hybrid zones is that they involve independent contacts between species, each with a unique evolutionary trajectory. In addition to random fluctuations, the outcomes will be contingent to the relative abundance of the two species, local environmental conditions and the genetic architecture of their reproductive isolation.

Our multilocus scan, based on AFLP and codominant markers, showed some genetic differences between the peripheral patch of $M. \text{galloprovincialis}$ in the Iberian Coast and the
patch enclosed in Brittany. In our analysis, AFLP markers were mainly used to obtain an estimation of the genomic distribution of $F_{ST}$ values. We also have fitted a speciation model to an independent DNA sequence dataset. We therefore checked whether the parameters inferred could reproduce the $F_{ST}$ distribution observed at AFLP markers. The two distributions - AFLPs versus simulations under the best demographic model inferred - were very similar (Figure S9) suggesting that we have a good estimate of the genomic distribution of $F_{ST}$. To mitigate the problem of false positives, due to an underestimation of the neutral variance of $F_{ST}$ (ROBERTSON 1975), we cross-checked the results of six methods, all of which assume different population structures. In this way, we identified three loci with a robust deviation from the neutral expectations (see white dots, Figure 2A). When compared to other methods, simulations of the inferred secondary contact model with heterogeneous gene flow produced an elevated variance of $F_{ST}$ in both the between-species comparison, and - more surprisingly - in the within-species comparison. Nevertheless, even under this scenario, the combination of a high differentiation within and between species observed at mac-1 was not observed under neutrality (see dashed lines, Figure 2A). The genealogy of mac-1 confirmed that this was due to local introgression of $M. edulis$ alleles into the Brittany patch of $M. galloprovincialis$, but not into the patch on the Iberian Coast (Figure 3). This pattern contrasts strongly with results at the other loci analysed.

The locus-specific nature of the introgression at mac-1 suggested that selective processes were acting rather than purely demographic ones. The highly asymmetrical introgression rates from $M. edulis$ (Table S5) contrasts with the multilocus pattern (Figure 7), and so we hypothesised that mac-1 might have been subject to adaptive introgression, with edulis alleles sweeping to fixation in the Brittany patch of $M. galloprovincialis$.

Demonstrating adaptive introgression requires a combination of evidence that is hard to obtain in most species (VEKEMANS 2010). Few studies have undertaken experiments to confirm the fitness advantage of the introgressed traits/alleles in the recipient backgrounds. Exceptions include the adaptive introgression of drought tolerance in Helianthus annuus.
(Whitney et al. 2010) and the adaptive introgression of rodent poison resistance in the house mouse, *Mus musculus domesticus* (Song et al. 2011). In contrast, most studies rely on indirect evidence involving geographical variation in allele frequencies, analysis of DNA polymorphism and the reconstruction of gene genealogies. For example, The Heliconius Genome Consortium (2012), Paro-Diaz et al. (2012) and Smith and Kron-forst (2013) provided strong evidence for the adaptive introgression of wing colour pattern genes involved in Müllerian mimicry between hybridizing species of *Heliconius*. In humans, Mendez et al. (2012a; 2012b; 2013) found a high frequency of divergent haplotypes at two immune genes, STAT2 and OAS1, in modern Melanesians. These haplotypes share recent common ancestry with archaic hominin sequences (STAT2: Neanderthal; OAS1: Denisovan). This result is suggestive of ancient introgressions, but their adaptive nature remains hypothetical. Similarly Roux et al. (2013) identified genomic hotspots of introgression between two highly divergent sea squirt species but could draw no conclusions about their adaptive nature, illustrating that specific tests need to be developed.

The patterns that we observe at mac-1 differ in one respect from previously studied cases of outlying introgression. In particular, in most previous cases, there has been a reduction in the genetic diversity in the recipient population, when compared to the donor population (see Table 1). In contrast, at the mac-1 locus, 96% of the *M. edulis* diversity was introgressed to *M. galloprovincialis*, and there were no other signs of a classical selective sweep (Tables 1-3).

However, it remains possible that the mac-1 alleles have introgressed via a “soft sweep”, with little loss of diversity (Penning and Hermisson 2006). We tried two approaches to test for a soft sweep at this locus. First, we analysed four polymorphisms along the sequence (Figure 4) with the hope of observing a gradient of introgression. Unfortunately, the chromosomal scale proved to be too small. Second, we used two new inference methods to test whether the rate of migration required to generate the soft sweep effect was consistent with the existence of an efficient genetic barrier at mac-1 elsewhere in the range. Our inference
suggested that $M = 2Nm > 1$ was required to explain our data under the adaptive introgression model, and the ML estimate ($\hat{M} = 4.63$; Figure 6B) was outside of the 95%-quantile of the multilocus distribution of introgression rates between $M. edulis$ and the Brittany patch of $M. galloprovincialis$ (Figure 7). However, its lower bound remained consistent with the Secondary Contact inference ($M_{ML lower bound} = 0.95$, grey dashed line, Figure 7), and so we cannot reject the hypothesis of adaptive introgression.

While we cannot reject adaptive introgression, we must also recognise alternative hypotheses. Migration pressure can easily spread gene combinations, even if these are universally selected against (e.g., Barton 1992). It is therefore possible that our results could be explained by a geographically and genomically localised change in the barrier strength between $M. edulis$ and $M. galloprovincialis$. The genomic region around mac-1 may contain barrier loci that were maintained in some patches, but not in the Brittany patch - perhaps due to differences in local ecological selection, or because an asymmetric intrinsic incompatibility managed to uncouple from the tension zone.

Together, our results emphasize the difficulties in interpreting outlier loci in populations that have undergone secondary contact. This is unsurprising given that patterns of introgression after such a contact can vary dramatically, depending on the way divergence occurred in allopatry, the resulting genetic architecture of the barrier, and the local landscape where the contact takes place - including its population density, connectivity and environmental variation (Biern et al. 2011; Domingues et al. 2012; Strasburg et al. 2012; Abbott et al. 2013; Biern et al. 2013; Gagnaire et al. 2013). Nevertheless, our best hope of understanding these issues is to apply dedicated methods, such as those developed here, to much larger genome-wide data sets.

ACKNOWLEDGMENTS

The authors are grateful to Marie-Thérèse Augé and Célia Gosset for their technical help with this work, and Marc Ohresser who kindly provided the sequence of the mac-1
region. Molecular data were produced through the ISEM platform “Génomique marine” at the Station Méditerranéenne de l’Environnement Littoral (OSU OREME) and the platform “Génomique Environnementale” of the LabEx CeMEB. Numerical results presented in this article were carried out using the ISEM computing cluster at the platform “Montpellier Bioinformatique et Biodiversité” of the LabEx CeMEB, and we are grateful to its staff. This work was funded by the Agence National de la Recherche (HYSEA project, ANR-12-BSV7-0011) and the project Aquagenet (SUDOE, INTERREG IV B). This is article 2014-044 of Institut des Sciences de l’Evoluation de Montpellier.

LITERATURE CITED


ARNOLD, M. L., 2004 Transfer and origin of adaptations through natural hybridization: were Anderson and Stebbins right? Plant Cell 16(3): 562–570.


BEAUMONT, M. A., 2008 DFDIST. Available at http://www.maths.bris.ac.uk/~mamab/stuff/


24


Kumar, S., K. Tamura, and M. Nei, 2004 MEGA3: integrated software for


PENNINGS, P. S. AND J. HERMISSON, 2006 Soft sweeps II: molecular population genetics of adaptation from recurrent mutation or migration. Mol. Biol. Evol. 23(5):


Roux, C., V. Castric, M. Pauwels, S. I. Wright, P. Saumitou-Laprade et al., 2011 Does speciation between Arabidopsis halleri and Arabidopsis lyrata coincide

ROUX, C., C. FRAISSE, V. CASTRIC, X. VEKEMANS, G. H. POGSON et al., 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. *in press."

ROUX, C., G. TSAGKOGEORGA, N. BI ErNE, AND N. GALTIER, 2013 Crossing the
species barrier: genomic hotspots of introgression between two highly divergent *Ciona

DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinfor-
matics 19(18): 2496–2497.

SKIBINSKI, D. O. F., J. A. BEARDMORE, AND T. F. CROSS, 1983 Aspects of the
population genetics of *Mytilus* (Mytilidae; Mollusca) in the British Isles. Biol. J. Linn.

SLATKIN, M., 1994 An exact test for neutrality based on the Ewens sampling distribu-

SLATKIN, M., 1996 A correction to the exact test based on the Ewens sampling distribu-

SMITH, J. AND M. R. KRONFORST, 2013 Do *Heliconius* butterfly species exchange

SONG, Y., S. ENDEPOLs, N. KLEMMANN, D. RICHTER, F.-R. MATUSCHKA et al.,
2011 Adaptive introgression of anticoagulant rodent 33 poison resistance by hybridiza-

STRASBURG, J. L., N. A. SHERMAN, K. M. WRIGHT, L. C. MOYLE, J. H. WILLIS
*et al.*, 2012 What can patterns of differentiation across plant genomes tell us about


Figure 1.—Localities of *Mytilus* spp. samples along the European coast. Panmictic patches of *M. edulis* are indicated by black solid lines; their corresponding samples are (1) Wadden Sea (black square) in the peripheral patch of the North Sea and (2) Lupin (black circle) in the enclosed patch of the Bay of Biscay. Panmictic patches of *M. galloprovincialis* are indicated by white solid lines; their corresponding samples are (1) Faro (white square) in the peripheral patch of the Iberian Coast and (2) Roscoff (white circle) in the enclosed patch of Brittany. The three hybrid zones described by Bierne et al. (2003) are indicated by dotted lines. HZ1: Iberian Coast / Bay of Biscay; HZ2: Bay of Biscay / Brittany; HZ3: Brittany / North Sea.
Figure 2.—Genome scan of inter and intra-species differentiation. (A) shows $F_{ST}$ values between species against values within *M. galloprovincialis* (Iberian Coast vs Brittany). Values for the $F_{ST}$ between species were based on the mean allele frequency of the two patches for each species. Dotted lines represent the threshold at 99% of the Lewontin and Krakauer test for the two comparisons. Dashed lines represent the 95% and 99% quantiles of the neutral envelope of $F_{ST}$ obtained from our custom simulation method. Loci detected by all methods as outliers within *M. galloprovincialis* are depicted by white dots while those outliers in 4 out of 6 methods are in light grey. The locus mac-1 is the only outlier with all methods in the interspecific and intraspecific comparisons (upper right square). On the margins are represented the corresponding $F_{ST}$ distributions. Genome scan was based on 440 loci (422 AFLPs and 18 codominant nuclear markers). All methods are described in Supp. Info. (File S1). (B) shows frequency of the *galloprovincialis* allele at the three outlier loci in the four panmictic patches of the studied area. *M. edulis* patches are in black (North Sea and Bay of Biscay); while *M. galloprovincialis* patches are in white (Iberian Coast and Brittany). We defined the *galloprovincialis* allele as the less frequent in the two patches of *M. edulis*.
Figure 3.—Allele genealogy reconstructed by the neighbour-joining algorithm apply to the number of nucleotide differences (region 2). Sequences sampled in *M. edulis* patches (black squares: North Sea; black circles: Bay of Biscay) cluster within the “*edulis*” clade; while those sampled in the Iberian Coast patch (white squares) of *M. galloprovincialis* cluster within the “*galloprovincialis*” clade. In Brittany (white circles), sequences are found in both clades.
Figure 4.—Frequency of the *edulis* allele at the four polymorphisms along the studied genomic region in the four patches. Position (in Kb) is indicated by stars for SNPs and by triangles for PCR product length polymorphisms. *M. edulis* patches are in black (squares: North Sea; circles: Bay of Biscay); while *M. galloprovincialis* patches are in white (squares: Iberian Coast; circles: Brittany). Sample size was $n = 48$, except in Brittany ($n = 87$).
FIGURE 5.—Expected sweep effect in the adaptive introgression model. The proportional diversity after introgression from \(d\) (donor species) to \(r\) (recipient species), \(H_{e_r}/H_{e_d}\), is shown as a function of the scaled migration rate, \(M\). Errors bars represent standard deviation calculated over 5,000 replicates. Median values are depicted by dots. Parameter values were \(N = 10000\), \(\theta = 2\), \(a = 1000\), \(r = 0\) (complete linkage) and \(n_d = n_r = 50\) (sample size). Simulations were run until the beneficial allele reached fixation in the recipient background.
Figure 6.—Parameter estimation in the adaptive introgression model. Maximum likelihood estimation of $\theta$ (A) and $M$ (B) are depicted in black. Maximum likelihood values are denoted by solid vertical lines and confidence intervals by dotted lines. Posterior distributions of $\theta$ (A), $M$ (B) and $\alpha$ (C) are depicted by grey dots. Medians are denoted by solid vertical lines and 95% highest posterior density values by dotted lines. Values are corrected for parameter bias in the ABC inference (see text for more details).
Figure 7.—Multilocus distributions of introgression rates estimated in the Secondary Contact model between *M. edulis* (North Sea) and *M. galloprovincialis* (Brittany). Introgression from *M. edulis* into *M. galloprovincialis* is shown in grey, while introgression in the opposite direction is shown in white. Distributions were obtained after randomly sampling 1,000 values from each of the 2,000 Beta distributions retained by the ABC analysis. The grey solid line shows the median introgression rate into *M. galloprovincialis* estimated at *mac-1* in the Secondary Contact model (*M_{SC} median* = 2.50). For comparison, the lower bound of the Maximum Likelihood estimation in the adaptive introgression model is reported as a grey dotted line (*M_{ML lower bound} = 0.95*).
**TABLE 1**

Reduction in diversity after introgression in studies of adaptive introgression

<table>
<thead>
<tr>
<th>Genus</th>
<th>Species pair</th>
<th>Locus</th>
<th>H_e / H_d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canis⁵</td>
<td>latrans</td>
<td>lupus</td>
<td>mitochondria</td>
</tr>
<tr>
<td>Eucalyptus⁶</td>
<td>cordata</td>
<td>globulus</td>
<td>chloroplast</td>
</tr>
<tr>
<td>Eurema⁷</td>
<td>hecabe b</td>
<td>hecabe y</td>
<td>Wolbachia</td>
</tr>
<tr>
<td>Littorina⁸</td>
<td>obtusata</td>
<td>fabalis SS</td>
<td>ARK</td>
</tr>
<tr>
<td>Mytilus⁹</td>
<td>edulis</td>
<td>trossulus</td>
<td>mitochondria</td>
</tr>
<tr>
<td>Mytilus¹⁰</td>
<td>edulis</td>
<td>galloprovincialis</td>
<td>mac-1</td>
</tr>
</tbody>
</table>

* d denotes the donor species; r denotes the recipient species. Only donor-derived alleles were considered. H_r/H_d: proportional diversity after introgression.

References: ¹Lehman et al. 1991; ²McKinnon et al. 2004; ³Narita et al. 2006; ⁴Kemppainen et al. 2011; ⁵Quesada et al. 1999; Rawson and Hilbish 1998; ⁶Present study.
### TABLE 2

Genetic statistics of the sequenced regions

<table>
<thead>
<tr>
<th>Region</th>
<th>n&lt;sub&gt;edu&lt;/sub&gt;</th>
<th>n&lt;sub&gt;gal&lt;/sub&gt;</th>
<th>S&lt;sub&gt;edu&lt;/sub&gt;</th>
<th>S&lt;sub&gt;gal&lt;/sub&gt;</th>
<th>S.&lt;i&gt;s&lt;/i&gt;&lt;sub&gt;edu&lt;/sub&gt;</th>
<th>S.&lt;i&gt;s&lt;/i&gt;&lt;sub&gt;gal&lt;/sub&gt;</th>
<th>S.&lt;i&gt;ns&lt;/i&gt;&lt;sub&gt;edu&lt;/sub&gt;</th>
<th>S.&lt;i&gt;ns&lt;/i&gt;&lt;sub&gt;gal&lt;/sub&gt;</th>
<th>S.&lt;i&gt;nc&lt;/i&gt;&lt;sub&gt;edu&lt;/sub&gt;</th>
<th>S.&lt;i&gt;nc&lt;/i&gt;&lt;sub&gt;gal&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17</td>
<td>19</td>
<td>14</td>
<td>34</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>11</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>11</td>
<td>53</td>
<td>51</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>53</td>
<td>51</td>
</tr>
<tr>
<td>3</td>
<td>27</td>
<td>27</td>
<td>25</td>
<td>27</td>
<td>16</td>
<td>11</td>
<td>9</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Region</th>
<th>θ&lt;sub&gt;S&lt;sub&gt;edu&lt;/sub&gt;&lt;/sub&gt;</th>
<th>θ&lt;sub&gt;S&lt;sub&gt;gal&lt;/sub&gt;&lt;/sub&gt;</th>
<th>θ&lt;sub&gt;R&lt;sub&gt;edu&lt;/sub&gt;&lt;/sub&gt;</th>
<th>θ&lt;sub&gt;R&lt;sub&gt;gal&lt;/sub&gt;&lt;/sub&gt;</th>
<th>R&lt;sub&gt;m&lt;sub&gt;edu&lt;/sub&gt;&lt;/sub&gt;</th>
<th>R&lt;sub&gt;m&lt;sub&gt;gal&lt;/sub&gt;&lt;/sub&gt;</th>
<th>D&lt;sub&gt;edu&lt;/sub&gt;</th>
<th>D&lt;sub&gt;gal&lt;/sub&gt;</th>
<th>H&lt;sub&gt;edu&lt;/sub&gt;</th>
<th>H&lt;sub&gt;gal&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.004</td>
<td>0.006</td>
<td>0.005</td>
<td>0.011</td>
<td>2</td>
<td>2</td>
<td>-0.44</td>
<td>-1.78*</td>
<td>0.82</td>
<td>2.91</td>
</tr>
<tr>
<td>2</td>
<td>0.103</td>
<td>0.013</td>
<td>0.138</td>
<td>0.016</td>
<td>7</td>
<td>4</td>
<td>-1.03</td>
<td>-0.77</td>
<td>-0.25</td>
<td>6.4</td>
</tr>
<tr>
<td>3</td>
<td>0.005</td>
<td>0.004</td>
<td>0.009</td>
<td>0.007</td>
<td>2</td>
<td>2</td>
<td>-1.4</td>
<td>-1.37</td>
<td>2.58</td>
<td>-3.26</td>
</tr>
</tbody>
</table>

<sub>edu</sub> denotes the patches of <i>M. edulis</i> (North Sea and Bay of Biscay); <sub>gal</sub> denotes the Brittany patch of <i>M. galloprovincialis</i>. Only <i>edulis</i>-derived alleles were considered. Significance of <i>D</i> and <i>H</i> were tested with standard-coalescent simulations (* p-value < 0.05). <i>n</i>: number of sequences sampled; <i>S</i>: total number of polymorphic sites; <i>S.<i>s</i></i>: number of synonymous mutations; <i>S.<i>ns</i></i>: number of non-synonymous mutations; <i>S.<i>i</i></i>: number of non-coding mutations; θ<sub>S</sub>: diversity estimated from the number of polymorphic sites (Watterson 1975); θ<sub>W</sub>: diversity estimated from pair-wise differences (Tajima 1983); R<i>m</i>: minimum number of recombination events (Hudson and Kaplan 1985); D: Tajima's D (Tajima 1989a; 1989b); H: Fay and Wu's H (Fay and Wu 2000).
**TABLE 3**

<table>
<thead>
<tr>
<th>Locus</th>
<th>(n_{edu})</th>
<th>(n_{gal})</th>
<th>(k_{edu})</th>
<th>(k_{gal})</th>
<th>(H_{edu})</th>
<th>(H_{gal})</th>
<th>(f_{edu})</th>
<th>(f_{gal})</th>
<th>(\Theta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>indel-in0</td>
<td>92</td>
<td>28</td>
<td>9</td>
<td>5</td>
<td>0.364</td>
<td>0.244</td>
<td>0.049</td>
<td>-0.071</td>
<td>-0.006</td>
</tr>
<tr>
<td>mac-1</td>
<td>177</td>
<td>60</td>
<td>15</td>
<td>6</td>
<td>0.761</td>
<td>0.728</td>
<td>0.0962</td>
<td>-0.053</td>
<td>0.0097</td>
</tr>
</tbody>
</table>

*edu* denotes the patches of *M. edulis* (North Sea and Bay of Biscay); *gal* denotes the Brittany patch of *M. galloprovincialis*. Only *edulis*-derived alleles were considered. Significance of \(f\) and \(\Theta\) were tested with 1000 permutations (* p-value < 0.05). \(n\): number of alleles sampled; \(k\): number of allelic classes; \(H_{e}\): heterozygosity (Nei 1978); \(f\): departure from HWE (Weir and Cockerham 1984); \(\Theta\): genetic differentiation (Weir and Cockerham 1984).