Gene-Flow in a Mosaic Hybrid Zone: Is Local Introgression Adaptive?

Christelle Fraïsse,^{*,†,‡,1} Camille Roux,^{*,†,§} John J. Welch,^{*} and Nicolas Bierne^{*,†}

^{*}Université Montpellier 2, 34095 Montpellier Cedex 5, France, [†]CNRS, Institut des Sciences de l'Évolution, ISEM Unité Mixte de Recherche 5554, 34200 SETE, France, [‡]Department of Genetics, University of Cambridge, CB2 3EH Cambridge, United Kingdom, and [§]Department of Ecology and Evolution, Lausanne University, Biophore/Sorge, CH-1015 Lausanne, Switzerland

ABSTRACT Genome-wide 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 *M. 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 localized 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 semipermeable, 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; 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., Piálek 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 *et al.* 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

Copyright © 2014 by the Genetics Society of America

doi: 10.1534/genetics.114.161380

Manuscript received January 10, 2014; accepted for publication April 17, 2014; published Early Online April 29, 2014.

Supporting information is available online at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.161380/-DC1>.

¹Corresponding author: SMEL, 2 rue des Chantiers, 34200 Sète, France.

E-mail: christelle.fraisse@univ-montp2.fr

is associated with a single new mutation. Adaptive introgression may involve multiple migrant copies of the same beneficial allele (Pennings 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 *M. galloprovincialis*. These species appear to have had a complex history of fragmentation and colonization during the Quaternary, as is the case for almost all temperate organisms (Hewitt 2011). Recent analyses suggest that *M. edulis* and *M. galloprovincialis* diverged ~2.5 million years (MY), followed by a secondary contact beginning about 0.7 MY (Roux *et al.* 2014). Today, they are isolated by multifarious pre- and postzygotic mechanisms (Bierne *et al.* 2002, 2003a, 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 the 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 localized region.

To test the plausibility of this hypothesis, we first sequenced a 3.1-kb region around *mac-1* and analyzed 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 (Pennings and Hermisson 2006). Inferences 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 *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

Table 1 Reduction in diversity after introgression in studies of adaptive introgression

Genus	Species pair		Locus	He _r /He _d
	Donor	Recipient		
<i>Canis</i> ^a	<i>latrans</i>	<i>lupus</i>	Mitochondria	0.752
<i>Eucalyptus</i> ^b	<i>cordata</i>	<i>globulus</i>	Chloroplast	0.751
<i>Eurema</i> ^c	<i>hecabe b</i>	<i>hecabe y</i>	<i>Wolbachia</i>	0.388
<i>Littorina</i> ^d	<i>obtusata</i>	<i>fabalis SS</i>	ARK	0.143
<i>Mytilus</i> ^e	<i>edulis</i>	<i>trossulus</i>	Mitochondria	0.608
<i>Mytilus</i> ^f	<i>edulis</i>	<i>galloprovincialis</i>	<i>mac-1</i>	0.957

d denotes the donor species; r denotes the recipient species. Only donor-derived alleles were considered. He_r/He_d, proportional diversity after introgression.

^a Lehman *et al.* (1991).

^b McKinnon *et al.* (2004).

^c Narita *et al.* (2006).

^d Kemppainen *et al.* (2011).

^e Quesada *et al.* (1999); Rawson and Hilbish (1998).

^f Present study.

of each species (Figure 1; Bierne *et al.* 2003c; Hilbish *et al.* 2012). We used two geographical samples of *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 *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 *et al.* (2008) and have been established to be representative of monospecific panmictic patches. In addition, we used a sample of *M. trossulus* (Tadoussac, Canada) to serve as an outgroup. Forty-eight 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 manufacturers protocol.

Multilocus scan

For the multilocus scan, we combined existing data, comprising 422 AFLP markers, 13 codominant nuclear markers (Gosset and Bierne 2012), and 5 allozyme markers (Faure *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 50 bp, 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 v. 1.0 (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

Figure 1 Localities of *Mytilus* spp. samples along the European coast. Panmictic patches of *M. edulis* are indicated by thick lines; their corresponding samples are (1) Wadden Sea (solid square) in the peripheral patch of the North Sea and (2) Lupin (solid circle) in the enclosed patch of the Bay of Biscay. Panmictic patches of *M. galloprovincialis* are indicated by double lines; their corresponding samples are (1) Faro (open square) in the peripheral patch of the Iberian Coast and (2) Roscoff (open circle) in the enclosed patch of Brittany. The three hybrid zones described by Bierne *et al.* (2003b) are indicated by dotted lines. HZ1, Iberian Coast/Bay of Biscay; HZ2, Bay of Biscay/Brittany; HZ3, Brittany/North Sea.

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 [Supporting Information, 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, personal communication) 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 diagram). 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, Brittany, Bay of Biscay, and Iberian Coast) for these four polymorphisms. See [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.* 1996–2004): the number of allelic classes (k), the heterozygosity (H_e , 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 neighbor-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 noncoding (S_{nc}) mutations, levels of nucleotide diversity estimated from the number of polymorphic sites (θ_s , Watterson 1975) and from pairwise differences (θ_π , Tajima 1983) and the minimum number of recombination events (R_m , 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 that 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 characterized 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 open circles in the *edulis* clade, Figure 3).

Maximum-likelihood approach: We first fit the data using a maximum-likelihood (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),

$$P(k_d | n_d, \theta) = \frac{\theta^{k_d} S_{n_d}^{k_d}}{\theta \times (\theta + 1) \dots (\theta + n_d - 1)} = \frac{\Gamma(\theta) \theta^{k_d} S_{n_d}^{k_d}}{\Gamma(\theta + n_d)}, \quad (1)$$

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). Pennings and Hermisson (2006) showed that adaptive introgression has some formal similarities to Equation 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 also approximate Ewens' distribution, but with the scaled migration rate, M , playing the role of the scaled mutation rate, θ (Pennings 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 θ . As such, the likelihood of observing our data are well approximated by

$$\begin{aligned} \mathcal{L}(\theta, M | k_d, k_r, n_d, n_r) &\approx P(k_d | n_d, \theta) \sum_{i=k_r}^{i=n_r} P(i | n_r, M) P(k_r | i, \theta) \\ &= \frac{\Gamma(M) \Gamma(\theta) \theta^{k_d+k_r} S_{n_d}^{k_d}}{\Gamma(M + n_r) \Gamma(\theta + n_d)} \sum_{i=k_r}^{n_r} \frac{\Gamma(\theta) M^i S_{n_r}^i S_i^{k_r}}{\Gamma(\theta + i)} \end{aligned} \quad (2)$$

(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 Equation 2 does not depend explicitly on the strength of selection, α , but this new result does allow us to estimate the parameters θ and M from our *mac-1* data. The parameters that maximize Equation 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 two units—obtained by maximizing the likelihood, conditional on the

parameter of interest taking a suboptimal value (Edwards 1992). An R script that implements Equation 2 is included as File S2.

Approximate Bayesian computation approach: Equation 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 (ABC) 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 (He_d and He_r ; Nei 1978), (3) the standard deviation in frequencies over all allelic classes (Sd_d and Sd_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 He_r/He_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 was estimated by means of weighted nonlinear multivariate regressions of the parameters on the summary statistics (Blum and Francois 2010). 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 were 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 data sets 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, $b = 1/n \sum_i [e_i/v_i]$, 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 th data set (Excoffier *et al.* 2005). We generated 100 pseudo-observed data sets

Figure 2 Genome scan of inter- and intraspecies differentiation. (A) 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 open circles while those outliers in four of six methods are shaded. The locus *mac-1* is the only outlier with all methods in the interspecific and intra-specific comparisons (top 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 File S1. (B) Frequency of the *galloprovincialis* allele at the three outlier loci in the four panmictic patches of the studied area. *M. edulis* patches are solid (North Sea and Bay of Biscay), while *M. galloprovincialis* patches are open (Iberian Coast and Brittany). We defined the *galloprovincialis* allele as the less frequent in the two patches of *M. edulis*.

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 (see 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 noncoding 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 data sets 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 (π , Tajima 1983); (2) Watterson's θ_w (Watterson 1975); (3) net interspecific divergence ($\text{netdiv}_{\text{edu-gal}}$); and (4) between-species differentiation computed as $1 - \pi_S/\pi_T$,

Table 2 Genetic statistics of the sequenced regions

Region	n_{edu}	n_{gal}	S_{edu}	S_{gal}	$S_{s_{edu}}$	$S_{s_{gal}}$	$S_{ns_{edu}}$	$S_{ns_{gal}}$	$S_{nc_{edu}}$	$S_{nc_{gal}}$
1	17	19	14	34	1	3	2	4	11	27
2	19	11	53	51	0	0	0	0	53	51
3	27	27	25	19	16	11	9	8	0	0

Region	$\theta_{s_{edu}}$	$\theta_{s_{gal}}$	$\theta_{\pi_{edu}}$	$\theta_{\pi_{gal}}$	$R_{m_{edu}}$	$R_{m_{gal}}$	D_{edu}	D_{gal}	H_{edu}	H_{gal}
1	0.004	0.006	0.005	0.011	2	2	-0.44	-1.78*	0.82	2.91
2	0.103	0.013	0.138	0.016	7	4	-1.03	-0.77	-0.25	6.4
3	0.005	0.004	0.009	0.007	2	2	-1.4	-1.37	2.58	-3.26

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 *D* and *H* were tested with standard-coalescent simulations: **P*-value < 0.05. *n*: number of sequences sampled; *S*: total number of polymorphic sites; *S_s*: number of synonymous mutations; *S_{ns}*: number of nonsynonymous mutations; *S_{nc}*: number of noncoding mutations; $\theta_{s_{}}$: diversity estimated from the number of polymorphic sites (Watterson 1975); $\theta_{\pi_{}}$: diversity estimated from pairwise differences (Tajima 1983); *R_m*: 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).

where π_s is the average pairwise nucleotide diversity within species and π_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, b). We also classified the variable genomic positions according to 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_p), 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.siteweb.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 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 (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 that 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_{ST} 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 (open circles within the “*edulis* clade”; Figure 3), with the related alleles found in the two *M. edulis* patches (solid squares and circles, Figure 3), suggests that very similar levels of genetic diversity are found in all three patches (see Table 2 and Table 3). For example, the proportion of diversity introgressed at the locus *indel-in0* is

Table 3 Genetic statistics of the PCR product length polymorphisms

Locus	n_{edu}	n_{gal}	k_{edu}	k_{gal}	He_{edu}	He_{gal}	f_{edu}	f_{gal}	Θ
<i>indel-in0</i>	92	28	9	5	0.364	0.244	0.049	-0.071	-0.006
<i>mac-1</i>	177	60	15	6	0.761	0.728	0.0962	-0.053	0.0097

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 Θ were tested with 1000 permutations: * P -value < 0.05. n , number of alleles sampled; k , number of allelic classes; He , heterozygosity (Nei 1978); f , departure from HW (Weir and Cockerham 1984); Θ , genetic differentiation (Weir and Cockerham 1984).

$He_{gal}/He_{edu} = 0.67$ and at *mac-1* is $He_{gal}/He_{edu} = 0.96$ (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 ~ 0.4 over several kilobase pairs.

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

Figure 3 Allele genealogy reconstructed by the neighbor-joining algorithm apply to the number of nucleotide differences (region 2). Sequences sampled in *M. edulis* patches (solid squares, North Sea; solid circles, Bay of Biscay) cluster within the "*edulis*" clade, while those sampled in the Iberian Coast patch (open squares) of *M. galloprovincialis* cluster within the "*galloprovincialis*" clade. In Brittany (open circles), sequences are found in both clades.

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 Equation 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 and $n_r = 60$ 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 solid 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 solid 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 shaded circles. In the following, we give their median and 95% credible intervals.

Figure 4 Frequency of the *edulis* allele at the four polymorphisms along the studied genomic region in the four patches. Position (in kilobase) is indicated by asterisks for SNPs and by inverted triangles for PCR product-length polymorphisms. *M. edulis* patches are solid (squares, North Sea; circles, Bay of Biscay), while *M. galloprovincialis* patches are open (squares, Iberian Coast; circles, Brittany). Sample size was $n = 48$, except in Brittany ($n = 87$).

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 that appear in the ML equations (namely, n_d , k_d , n_p and k_r). In this case, estimates of θ and M were very similar to the MLEs (Figure S4), and there was no information on α as expected. The results of the ABC approach therefore come from the additional use of summary statistics that contain information that is not used in the ML approach. We therefore conducted a goodness-of-fit test 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 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 understandable approach with a simpler model that retrieved the best fit to the data and consequently our interpretations rely on the ML estimates.

Finally, we used our simulations to assess the effects of violating some assumptions of our model: (1) the assumption

Figure 5 Expected sweep effect in the adaptive introgression model. The proportional diversity after introgression from d (donor species) to r (recipient species), He_r/He_d , is shown as a function of the scaled migration rate, M . Errors bars represent standard deviation calculated over 5000 replicates. Median values are depicted by circles. Parameter values were $N = 10000$, $\theta = 2$, $\alpha = 1000$, $r = 0$ (complete linkage), and $n_d = n_r = 50$ (sample size). Simulations were run until the beneficial allele reached fixation in the recipient background.

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.

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 analyzed 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 data sets 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

Figure 6 Parameter estimation in the adaptive introgression model. Maximum-likelihood estimation of (A) θ and (B) M are depicted by black curves. Maximum-likelihood values are denoted by solid vertical lines and confidence intervals by dotted lines. Posterior distributions of (A) θ , (B) M , and (C) α are depicted by shaded circles. 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).

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 open bars, Figure 7) and $M_{\text{edulis to gallo}} = 0.45$ [0.013 – 2.84] (see shaded bars, Figure 7). However, compared to the multilocus inference, introgression at *mac-1* was strongly unidirectional from *M. edulis* to *M. galloprovincialis* as suggested by the comparison of different scenarios of introgression (SC with introgression into *M. edulis*: $P_{\text{post}} = 0.0013$; SC with introgression into *M. galloprovincialis*: $P_{\text{post}} = 0.9997$; see Table S5). Furthermore introgression rate into *M. galloprovincialis* at *mac-1* was very high compared to the extent of gene flow across loci ($M_{\text{median}} = 2.50$, solid line, Figure 7) that is consistent with the *mac-1* genealogy (Figure 3). In comparison, our ML estimation of M at *mac-1* in the adaptive introgression model ($\hat{M} = 4.63$, Figure 6B) was outside the 95% quantile of the multilocus distribution, but its lower bound ($M_{\text{ML lower bound}} = 0.95$, 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. galloprovincialis* in the Iberian Coast and the patch enclosed in Brittany. In our analysis, AFLP markers were used mainly to obtain an estimation of the genomic distribution of F_{ST} values. We also have fitted a speciation model to an independent DNA sequence data set. We therefore checked whether the parameters inferred could reproduce the F_{ST} distribution observed at AFLP markers. The two distributions—AFLPs vs. 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 (open circles, 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

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 as shaded, while introgression in the opposite direction is shown as open. Distributions were obtained after randomly sampling 1000 values from each of the 2000 beta distributions retained by the ABC analysis. The 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 dotted line ($M_{ML, lower bound} = 0.95$).

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 analyzed.

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 hypothesized 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, Heliconius Genome Consortium (2012), Pardo-Diaz *et al.* (2012), and Smith and Kronforst (2013) provided

strong evidence for the adaptive introgression of wing color pattern genes involved in Müllerian mimicry between hybridizing species of *Heliconius*. In humans, Mendez *et al.* (2012a,b, 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 (Table 1, Table 2, and Table 3).

However, it remains possible that the *mac-1* alleles have introgressed via a soft sweep, with little loss of diversity (Pennings and Hermisson 2006). We tried two approaches to test for a soft sweep at this locus. First, we analyzed 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 ($M_{ML, lower bound} = 0.95$, dashed line, Figure 7), and so we cannot reject the hypothesis of adaptive introgression.

While we cannot reject adaptive introgression, we must also recognize 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 localized 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 (Bierne *et al.* 2011, 2013a; Domingues *et al.* 2012; Strasburg *et al.* 2012; Abbott *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 (Observatoire de Recherche Méditerranéen de l'Environnement)] and the platform Génomique Environnementale of the LabEx CeMEB (Laboratoire d'Excellence Centre Méditerranéen de l'Environnement et de la Biodiversité). 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 (SU DOE, INTERREG IV B). This is article 2014-044 of Institut des Sciences de l'Evolution de Montpellier (ISEM).

Literature Cited

- Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird *et al.*, 2013 Hybridization and speciation. *J. Evol. Biol.* 26(2): 229–246.
- Arnold, M. L., 2004 Transfer and origin of adaptations through natural hybridization: Were Anderson and Stebbins right? *Plant Cell* 16(3): 562–570.
- Barton, N. H., 1979 The dynamics of hybrid zones. *Heredity* 43(3): 341–359.
- Barton, N. H., 1992 On the spread of new gene combinations in the third phase of Wright's shifting-balance. *Evolution* 46(2): 551–557.
- Barton, N. H., 2013 Does hybridization influence speciation? *J. Evol. Biol.* 26(2): 267–269.
- Barton, N. H., and B. O. Bengtsson, 1986 The barrier to genetic exchange between hybridising populations. *Heredity* 57: 357–376.
- Beaumont, M. A., and D. J. Balding, 2004 Identifying adaptive genetic divergence among populations from genome scans. *Mol. Ecol.* 13(4): 969–980.
- Beaumont, M. A., and R. A. Nichols, 1996 Evaluating loci for use in the genetic analysis of population structure. *Proc. Biol. Sci.* 263(1377): 1619–1626.
- Beaumont, M. A., W. Zhang, and D. J. Balding, 2002 Approximate Bayesian computation in population genetics. *Genetics* 162: 2025–2035.
- Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste, and F. Bonhomme, 1996–2004 GENETIX, version 4.05. <http://kimura.univ-montp2.fr/genetix/>
- Bierne, N., 2010 The distinctive footprints of local hitchhiking in a varied environment and global hitchhiking in a subdivided population. *Evolution* 64(11): 3254–3272.
- Bierne, N., P. David, P. Boudry, and F. Bonhomme, 2002 Assortative fertilization and selection at larval stage in the mussels *Mytilus edulis* and *M. galloprovincialis*. *Evolution* 56(2): 292–298.
- Bierne, N., F. Bonhomme, and P. David, 2003a Habitat preference and the marine-speciation paradox. *Proc. Biol. Sci.* 270(1522): 1399–1406.
- Bierne, N., P. Borsa, C. Daguin, D. Jollivet, F. Viard *et al.*, 2003b Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. *Mol. Ecol.* 12(2): 447–461.
- Bierne, N., C. Daguin, F. Bonhomme, P. David, and P. Borsa, 2003c Direct selection on allozymes is not required to explain heterogeneity among marker loci across a *Mytilus* hybrid zone. *Mol. Ecol.* 12(9): 2505–2510.
- Bierne, N., F. Bonhomme, P. Boudry, M. Szulkin, and P. David, 2006 Fitness landscapes support the dominance theory of post-zygotic isolation in the mussels *Mytilus edulis* and *M. galloprovincialis*. *Proc. Biol. Sci.* 273(1591): 1253–1260.
- Bierne, N., J. Welch, E. Loire, F. Bonhomme, and P. David, 2011 The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Mol. Ecol.* 20(10): 2044–2072.
- Bierne, N., P.-A. Gagnaire, and P. David, 2013a The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. *Curr. Zool.* 59(1): 72–86.
- Bierne, N., D. Roze, and J. J. Welch, 2013b Pervasive selection or is it. . .? Why are F_{ST} outliers sometimes so frequent? *Mol. Ecol.* 22(8): 2061–2064.
- Blum, M. G. B., and O. Francois, 2010 Non-linear regression models for approximate Bayesian computation. *Stat. Comput.* 20(1): 63–73.
- Bonhomme, M., C. Chevalet, B. Servin, S. Boitard, J. Abdallah *et al.*, 2010 Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics* 186: 241–262.
- Caballero, A., H. Quesada, and E. Roldan-Alvarez, 2008 Impact of amplified fragment length polymorphism size homoplasy on the estimation of population genetic diversity and the detection of selective loci. *Genetics* 179: 539–554.
- Charalambides, C. A., and J. Singh, 1988 Review of the stirring numbers, their generalizations and statistical applications. *Commun. Stat. Theory* 17(8): 2507–2532.
- Charlesworth, B., M. Nordborg, and D. Charlesworth, 1997 The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res.* 70(2): 155–174.
- Cornuet, J.-M., V. Ravigné, and A. Estoup, 2010 Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics* 11(1): 401.
- Corpet, F., 1988 Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* 16(22): 10881–10890.
- Csilléry, K., O. François, and M. G. B. Blum, 2012 abc: an R package for approximate Bayesian computation (ABC). *Methods Ecol. Evol.* 3(3): 475–479.
- Daguin, C., F. Bonhomme, and P. Borsa, 2001 The zone of sympatry and hybridization of *Mytilus edulis* and *M. galloprovincialis*, as described by intron length polymorphism at locus *mac-1*. *Heredity* 86: 342–354.
- Domingues, V. S., Y.-P. Poh, B. K. Peterson, P. S. Pennings, J. D. Jensen *et al.*, 2012 Evidence of adaptation from ancestral variation in young populations of beach mice. *Evolution* 66(10): 3209–3223.
- Edwards, A. W. F., 1992 *Likelihood*. Johns Hopkins University Press, Baltimore, MD.

- Ewens, W. J., 1972 The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* 3(1): 87–112.
- Excoffier, L., A. Estoup, and J.-M. Cornuet, 2005 Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics* 169: 1727–1738.
- Fagundes, N. J. R., N. Ray, M. Beaumont, S. Neuenschwander, F. M. Salzano *et al.*, 2007 Statistical evaluation of alternative models of human evolution. *Proc. Natl. Acad. Sci. USA* 104(45): 17614–17619.
- Faure, M. F., P. David, F. Bonhomme, and N. Bierne, 2008 Genetic hitchhiking in a subdivided population of *Mytilus edulis*. *BMC Evol. Biol.* 8: 164.
- Fay, J. C., and C. I. Wu, 2000 Hitchhiking under positive Darwinian selection. *Genetics* 155: 1405–1413.
- Gagnaire, P.-A., S. A. Pavey, E. Normandeau, and L. Bernatchez, 2013 The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution* 67(9): 2483–2497.
- Gosset, C. C., and N. Bierne, 2012 Differential introgression from a sister species explains high F_{ST} outlier loci within a mussel species. *J. Evol. Biol.* 26(1): 14–26.
- Hall, T., 1999 BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41: 95–98.
- Hamilton, G., M. Stoneking, and L. Excoffier, 2005 Molecular analysis reveals tighter social regulation of immigration in patrilocal populations than in matrilineal populations. *Proc. Natl. Acad. Sci. USA* 102(21): 7476–7480.
- Harrison, R. G., 1986 Pattern and process in a narrow hybrid zone. *Heredity* 56(3): 337–349.
- Harrison, R. G., and D. M. Rand, 1989 Mosaic hybrid zone and the nature of species boundaries, pp. 111–133 in *Speciation and Its Consequences*, edited by D. Otte, and J. A. Endler. Sinauer Associates, Sunderland, MA.
- Hedrick, P. W., 2013 Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol. Ecol.* 22(18): 4606–4618.
- Heliconius Genome Consortium, 2012 Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487(7405): 94–98.
- Hewitt, G. M., 2011 Quaternary phylogeography: the roots of hybrid zones. *Genetica* 139(5): 617–638.
- Hilbish, T. J., F. P. Lima, P. M. Brannock, E. K. Fly, R. L. Rognstad *et al.*, 2012 Change and stasis in marine hybrid zones in response to climate warming. *J. Biogeogr.* 39(4): 676–687.
- Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson *et al.*, 2010 Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.* 6(2): e1000862.
- Hudson, R. R., 2002 Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* 18(2): 337–338.
- Hudson, R. R., and N. L. Kaplan, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111: 147–164.
- Jackson, B., T. Kawakami, S. Cooper, J. Galindo, and R. Butlin, 2012 A genome scan and linkage disequilibrium analysis among chromosomal races of the Australian grasshopper *Vandiemennella viatica*. *PLoS ONE* 7(10): e47549.
- Kempainen, P., T. Lindskog, R. Butlin, and K. Johannesson, 2011 Intron sequences of arginine kinase in an intertidal snail suggest an ecotype-specific selective sweep and a gene duplication. *Heredity* 106(5): 808–816.
- Kumar, S., K. Tamura, and M. Nei, 2004 MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* 5(2): 150–163.
- Lehman, N., A. Eisenhawer, K. Hansen, L. D. Mech, R. O. Peterson *et al.*, 1991 Introgression of coyote mitochondrial DNA into sympatric North American gray wolf populations. *Evolution* 45(1): 104–119.
- Lewontin, R. C., and J. Krakauer, 1973 Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74: 175–195.
- Mallet, J., 2007 Hybrid speciation. *Nature* 446(7133): 279–283.
- Maynard Smith, J., and J. Haigh, 1974 The hitchhiking effect of a favourable gene. *Genet. Res.* 23(01): 23–35.
- McDonald, J. H., 1994 Detecting natural selection by comparing geographic variation in protein and DNA polymorphisms, pp. 88–100 in *Non-neutral Evolution*, edited by B. Golding. Springer, New York.
- McKinnon, G. E., R. E. Vaillancourt, D. A. Steane, and B. M. Potts, 2004 The rare silver gum, *Eucalyptus cordata*, is leaving its trace in the organellar gene pool of *Eucalyptus globulus*. *Mol. Ecol.* 13(12): 3751–3762.
- Mendez, F. L., J. C. Watkins, and M. F. Hammer, 2012a Global genetic variation at OAS1 provides evidence of archaic admixture in melanesian populations. *Mol. Biol. Evol.* 29(6): 1513–1520.
- Mendez, F. L., J. C. Watkins, and M. F. Hammer, 2012b A haplotype at STAT2 introgressed from neanderthals and serves as a candidate of positive selection in Papua New Guinea. *Am. J. Hum. Genet.* 91(2): 265–274.
- Mendez, F. L., J. C. Watkins, and M. F. Hammer, 2013 Neandertal origin of genetic variation at the cluster of OAS immunity genes. *Mol. Biol. Evol.* 30(4): 798–801.
- Nachman, M. W., and B. A. Payseur, 2012 Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philos. Trans. R. Soc. B* 367 (1587): 409–421.
- Nadeau, N. J., A. Whibley, R. T. Jones, J. W. Davey, K. K. Dasmahapatra *et al.*, 2012 Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philos. Trans. R. Soc. B* 367(1587): 343–353.
- Narita, S., M. Nomura, Y. Kato, and T. Fukatsu, 2006 Genetic structure of sibling butterfly species affected by *Wolbachia* infection sweep: evolutionary and biogeographical implications. *Mol. Ecol.* 15(4): 1095–1108.
- Nei, M., 1973 Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* 70(12): 3321–3323.
- Nei, M., 1978 Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583–590.
- Pardo-Díaz, C., C. Salazar, S. W. Baxter, C. Merot, W. Figueiredo-Ready *et al.*, 2012 Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genet.* 8(6): e1002752.
- Parsons, T. J., S. L. Olson, and M. J. Braun, 1993 Unidirectional spread of secondary sexual plumage traits across an avian hybrid zone. *Science* 260(5114): 1643–1646.
- Payseur, B. A., 2010 Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Mol. Ecol. Resour.* 10(5): 806–820.
- 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): 1076–1084.
- Pérez-Figueroa, A., M. J. García-Pereira, M. Saura, E. Rolán-Alvarez, and A. Caballero, 2010 Comparing three different methods to detect selective loci using dominant markers. *J. Evol. Biol.* 23(10): 2267–2276.
- Piálek, J., and N. H. Barton, 1997 The spread of an advantageous allele across a barrier: the effects of random drift and selection against heterozygotes. *Genetics* 145: 493–504.
- Quesada, H., R. Wenne, and D. O. Skibinski, 1999 Interspecies transfer of female mitochondrial DNA is coupled with role-reversals

- and departure from neutrality in the mussel *Mytilus trossulus*. *Mol. Biol. Evol.* 16(5): 655–665.
- Rawson, P. D., and T. J. Hilbish, 1998 Asymmetric introgression of mitochondrial DNA among european populations of blue mussels (*Mytilus spp.*). *Evolution* 52(1): 100–108.
- Robertson, A., 1975 Gene frequency distributions as a test of selective neutrality. *Genetics* 81: 775–785.
- Roesti, M., D. Moser, and D. Berner, 2013 Recombination in the threespine stickleback genome-patterns and consequences. *Mol. Ecol.* 22(11): 3014–3027.
- Ross-Ibarra, J., S. I. Wright, J. P. Foxe, A. Kawabe, L. DeRose-Wilson *et al.*, 2008 Patterns of polymorphism and demographic history in natural populations of *Arabidopsis lyrata*. *PLoS ONE* 3(6): e2411.
- Roux, C., V. Castric, M. Pauwels, S. I. Wright, P. Saumitou-Laprade *et al.*, 2011 Does speciation between *Arabidopsis halleri* and *Arabidopsis lyrata* coincide with major changes in a molecular target of adaptation? *PLoS ONE* 6(11): e26872.
- Roux, C., G. Tsagkogeorga, N. Bierne, and N. Galtier, 2013 Crossing the species barrier: genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. *Mol. Biol. Evol.* 30(7): 1574–1587.
- 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. *J. Evol. Biol.* (in press).
- Rozas, J., J. C. Sánchez-DelBarrio, X. Messeguer, and R. Rozas, 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 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. Soc. Lond.* 19(2): 137–183.
- Slatkin, M., 1994 An exact test for neutrality based on the Ewens sampling distribution. *Genet. Res.* 64(1): 71–74.
- Slatkin, M., 1996 A correction to the exact test based on the Ewens sampling distribution. *Genet. Res.* 68(3): 259–260.
- Smith, J., and M. R. Kronforst, 2013 Do *Heliconius* butterfly species exchange mimicry alleles? *Biol. Lett.* 9, 20130503.
- Song, Y., S. Endepols, N. Klemann, D. Richter, F.-R. Matuschka *et al.*, 2011 Adaptive introgression of anticoagulant rodent 33 poison resistance by hybridization between old world mice. *Curr. Biol.* 21(15): 1296–1301.
- 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 adaptation and speciation? *Philos. Trans. R. Soc. B* 367(1587): 364–373.
- Tajima, F., 1983 Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437–460.
- Tajima, F., 1989a The effect of change in population size on DNA polymorphism. *Genetics* 123: 597–601.
- Tajima, F., 1989b Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
- Tanguy, A., N. Bierne, C. Saavedra, B. Pina, E. Bachère *et al.*, 2008 Increasing genomic information in bivalves through new EST collections in four species: development of new genetic markers for environmental studies and genome evolution. *Gene* 408(1–2): 27–36.
- Turner, T. L., M. W. Hahn, and S. V. Nuzhdin, 2005 Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol.* 3: e285.
- Vekemans, X., 2010 What's good for you may be good for me: evidence for adaptive introgression of multiple traits in wild sunflower. *New Phytol.* 187(1): 6–9.
- Vekemans, X., T. Beauwens, M. Lemaire, and I. Roldán-Ruiz, 2002 Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Mol. Ecol.* 11(1): 139–151.
- Vitalis, R., K. Dawson, P. Boursot, and K. Belkhir, 2003 DetSel 1.0: a computer program to detect markers responding to selection. *J. Hered.* 94(5): 429–431.
- Wakeley, J., and J. Hey, 1997 Estimating ancestral population parameters. *Genetics* 145: 847–855.
- Watterson, G., 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 7(2): 256–276.
- Weir, B. S., and C. C. Cockerham, 1984 Estimating F-statistics for the analysis of population structure. *Evolution* 38(6): 1358–1370.
- Whitlock, R., H. Hipperson, M. Mannarelli, R. K. Butlin, and T. Burke, 2008 An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Mol. Ecol. Res.* 8(4): 725–735.
- Whitney, K. D., R. A. Randell, and L. H. Rieseberg, 2010 Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. *New Phytol.* 187(1): 230–239.
- Wright, S., 1949 Population structure in evolution. *Proc. Am. Philos. Soc.* 93(6): 471–478.
- Zhivotovsky, L. A., 1999 Estimating population structure in diploids with multilocus dominant DNA markers. *Mol. Ecol.* 8(6): 907–913.

Communicating editor: M. A. Beaumont

GENETICS

Supporting Information

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.161380/-/DC1>

Gene-Flow in a Mosaic Hybrid Zone: Is Local Introgression Adaptive?

Christelle Fraïsse, Camille Roux, John J. Welch, and Nicolas Bierne

Supplementary methods

Outlier detection in the multilocus scan:

Below are the details of the six methods used to detect F_{ST} outlier loci.

(1) We used a development of the Lewontin and Krakauer (1973) test for between-locus heterogeneity in Wright's fixation index, F_{ST} . To minimize false positives resulting from complex population structure, we followed the conservative suggestion of Fourcade et al. (2013) and fitted the data to a Generalized Extreme Value distribution rather than a Chi-squared distribution. This analysis used the "evir" package (version 1.7-3; Bernhard Pfaff) in R (R Core Development Team 2011), with an outlier threshold of $p < 0.01$ for this conservative test.

(2) We used the method of Beaumont and Nichols (1996) that estimates the null distribution of the test statistic by simulation. Simulations used the FDIST2 software for the codominant markers (Beaumont and Nichols 1996) and DFDIST software for the dominant AFLP markers (Beaumont 2008). The mean neutral F_{ST} value was calculated by an iterative procedure, removing loci outlying the 95% quantile. We obtained a null distribution using a two-deme model, generating 100,000 loci sampled for 48 individuals per population. The global critical frequency of the most common allele was set to 0.99, and the outlier threshold set at $p < 0.01$.

(3) We used the method implemented in DETSEL (Vitalis et al. 2003) which includes another form of population structure in which a common ancestral population of size N splits into two daughter populations that diverge by drift for T generations, and derives its null distribution using coalescent simulations. Four sets of parameters without a bottleneck were computed: (i) $T=100$, $N=100,000$; (ii) $T=1000$, $N=100,000$; (iii) $T=100$, $N=1,000,000$; and (iv) $T=1000$, $N=1,000,000$. The maximum allele frequency was set to 0.99, and the outlier threshold set at $p < 0.01$.

(4) We used the approach of Beaumont and Balding (2004), which includes a more complex model of population structure, as implemented in BAYESCAN (Foll and Gaggiotti 2008). Departure from neutrality at a given locus is inferred with a reversible-jump MCMC between models with and without selection. The MCMC used a sample size of 5000, and 20 pilot runs of the same length, a thinning interval of 20, and a burn-in of 100,000 were used. For the AFLP data, we used a Beta prior for F_{IS} with a mean of 0.005 and standard deviation of 0.001. We used prior odds for the neutral model of 5 (i.e., we assumed that neutrality was 5 times more likely *a priori*). The outlier threshold was defined as a Bayes Factor > 10 .

(5) We used a method that accounts for the possible inflation of the variance in neutral F_{ST} due to co-ancestry between individuals (Robertson 1975). In particular, we used the method of Bonhomme et al. (2010), and estimated the co-ancestry matrix from subset of the AFLP markers (removing any outlier loci identified with the method of Beaumont and Nichols

1996), and the *M. trossulus* sample as outgroup. We performed 1,000,000 simulations conditioned on the estimated matrix to derive the empirical null distribution of the test statistic, and the outlier threshold was set at $p < 0.001$.

(6) *Custom simulation method*: We developed a custom method to estimate the neutral envelope of F_{ST} both between and within species using the best-supported demographic models (see below “Models of speciation”). These are Secondary Contact with heterogeneous rates of introgression for the between-species comparisons and a two-deme model with homogeneous gene flow for the within-species comparison. We simulated 100 datasets of 422 SNPs in a 3-patch model (*M. edulis*, *M. galloprovincialis* of Iberian Coast and *M. galloprovincialis* of Brittany) in which *M. edulis* and the ancestral lineage of *M. galloprovincialis* patches split T_{split} My ago; and came back into contact T_{SC} My ago. We assumed that the two *M. galloprovincialis* patches split T_{SC} My ago from which the three patches started to exchange migrants at rates $M_{X \text{ to } Y}$ (in other words we made the assumption that the mosaic structure as currently observed established at the same time as the secondary contact). We set up the values of demographic parameters to their posterior estimate (see below “Models of speciation”): effective population sizes fixed to $N_{anc} \sim 1,000,000$; $N_{edulis} \sim 200,000$; $N_{gallo_Iberian_Coast} \sim 200,000$; $N_{gallo_Brittany} \sim 200,000$; time of speciation $T_{split} \sim 2.5$ My and time of secondary contact $T_{SC} \sim 0.7$ My; heterogeneous gene flow between *M. edulis* and *M. galloprovincialis* of the Iberian Coast (see Roux *et al.* 2014) with median values of $M_{edulis \text{ to } Iberian_Coast} = 1.22$ and $M_{Iberian_Coast \text{ to } edulis} = 0.85$; heterogeneous gene flow between *M. edulis* and *M. galloprovincialis* of Brittany with median values of $M_{edulis \text{ to } Brittany} = 0.3$ and $M_{Brittany \text{ to } edulis} = 0.9$; symmetric and homogeneous gene flow within *M. galloprovincialis* with median values of $M_{Brittany \text{ to } Iberian_Coast} = M_{Iberian_Coast \text{ to } Brittany} = 23$.

Chromosomal walking along *mac-1*:

(1) *Sequencing along mac-1*: To avoid the amplification of paralogues, at least one primer of a pair was defined inside an intron. To directly sequence the PCR products, allele-specific primers were defined on intra-species polymorphism (PCR product length polymorphisms or SNP). Then, we genotyped the presence/absence of the two versions of the primers, and sequenced the heterozygous individuals (see Bierne 2010). We corrected the sampling bias by re-sampling the sequences according to the frequency of the allele in each sample. PCRs were performed as described in Daguin *et al.* (2001). PCR products were purified with an ExoSAP kit (USB) and subjected to a sequencing reaction using the Big Dye Terminator kit (Applied Biosystems). Sequencing used an ABI 3130XL instrument at the LabEx CeMEB sequencing platform, Montpellier. The 880-bp region upstream of *mac-1* was amplified with the forward primer, *region1-F* and a set of reverse primers that prime each of the two alleles. For the *M. galloprovincialis* samples, sets were: (i) *region1-galA-R1* and *region1-galA-R2*, (ii) *region1-galB-R1* and *region1-galB-R2*, (iii) *region1-galC-R1* and *region1-galC-R2*. For the *M. edulis* samples, the set was: *region1-edu-R1* and *region1-edu-R2*. A second 1.5-kb region is located between *mac-1* and the second exon of the sequenced segment. It was amplified using a set of biallelic forward primers and a reverse primer, *region2-R*. For the *M. galloprovincialis* samples, the set was: *region2-gal-F1* and *region2-gal-F2*. For the *M. edulis* samples, the set was: *region2-*

edu-F1 and *region2-edu-F2*. To obtain high quality sequences, we defined new primers in the middle of the second region (*region2-middle-F* and *region2-middle-R*) and re-sequenced from the beginning to the middle (600 bp) and from the middle to the end (900 bp) of the PCR product. The third region (750 bp) corresponds to the second exon of the sequenced segment. The low level of polymorphism of the third region allowed us sequencing without separating the two chromosomes. To avoid sequencing paralogues, we first used the reverse primer *SNP-R* with a set of locus-specific forward primers: (i) *region3-edu-F* for the *M. edulis* samples, (ii) *region3-gal-F* for the *M. galloprovincialis* samples. Then, we re-sequenced from the beginning to the end of the second exon with the forward primer *SNP-ex2-F*. Primers are listed in Table S6.

(2) *Genotyping at four polymorphisms*: For genotyping the PCR product length polymorphisms (*indel_in0*, an indel polymorphism and *mac-1*, an “exon-primed intron-crossing PCR marker”), the fluorescent dye 5' end-labelled-primer technique was used with dye HEX (Sigma Genosys). We first amplified *indel_in0* with primers *indel-in0-F/indel-in0-R* and *mac-1* with primers *mac-1-F/mac-1-R*. Then, the PCR products were separated and detected using an Applied ABI Prism 3130XL (Applied Biosystems) and sized by GeneScan™ 500 ROXTM Size Standard (Applied Biosystems). GeneMapper® v4.5 software (Applied Biosystems) was used to read the resulting chromatograms. For *SNP_in1*, we used a presence/absence genotyping on agarose gel by defining a reverse primer (*SNP-R*) and species-specific forward primers: *SNP-in1-F1* for the *edulis* allele and *SNP-in1-F2* for the *galloprovincialis* allele. For *SNP_ex2*, we genotyped by sequencing the second exon of the sequenced segment as described above. Primers are listed in Table S7.

Models of speciation:

(1) *Coalescent simulations*: We carried out simulations under various models of speciation with varying patterns of gene flow: Strict Isolation, Isolation with Migration, Ancient Migration and Secondary Contact (Figure S8). Migration rates were allowed to be asymmetrical between the populations, and were allowed to be either homogeneous or heterogeneous between loci. In the “homogeneous” scenarios, scaled migration rates in each direction were randomly sampled from the uniform prior interval [0 - 10]. For the “heterogeneous” scenarios, the migration rates at each locus were drawn from a scaled Beta distribution, $pdf(x) = (x/c)^{\alpha-1}(1 - (x/c))^{\beta-1} / B(\alpha, \beta)$, and the parameters of this distribution were randomly sampled from the uniform intervals [0 - 5] (α) and [0 - 200] (β), and [0 - 10] (c). We also placed priors on the population mutation rates in the ancestral and two extant populations. Each of these values was scaled by the reference rate, $\theta_{ref} = 4N_{ref}\mu$ where N_{ref} is the effective number of individuals of a reference population used in coalescent simulations, arbitrarily fixed at 100,000 and μ the mutation rate of 2.763×10^{-8} /bp/generation. Each of these scaled quantities was assigned a uniform prior on the interval [0 -20]. We sampled the scaled speciation time $T_{split}/4N_{ref}$ from the interval [0 - 25] corresponding to [0 - 10^7] generations in demographic units. The time of Secondary Contact T_{sc} and the isolation time after Ancient Migration T_{iso} were drawn from a uniform distribution on the interval [0 - T_{split}].

(2) *Model choice*: To statistically compare the different models of speciation, we first evaluated the relative posterior probabilities for the two alternative scenarios (“homogeneous” vs “heterogeneous”) within models implementing gene flow. Next, we compared the best-supported version from these models, i.e., the ones with the highest value of the posterior probability, to the Strict Isolation model. Posterior probabilities for each candidate model were estimated using a feed-forward neural network implementing a non-linear multivariate regression by considering the model itself as an additional parameter to be inferred. The 2,000 x n replicate simulations nearest to the observed values for the summary statistics were selected (where n is the number of compared models), and these were weighted by an Epanechnikov kernel that reaches a maximum when $\text{Statistic}_{\text{obs}} = \text{Statistic}_{\text{sim}}$. Computations were performed using 50 trained neural networks and 15 hidden networks in the regression (R package “abc”, Csilléry *et al.* 2012). In a similar way, we carried out pairwise comparisons of different scenarios of introgression at the *mac-1* locus, in the secondary contact model; these were (1) no introgression; (2) introgression into *M. edulis* only; (3) introgression into *M. galloprovincialis* only, and (4) introgression in both directions at different rates. The posterior probability of each scenario was estimated as described above. For parameter estimation, we log-tangent transformed the parameters (Hamilton *et al.* 2005) and only the 2,000 replicate simulations producing the smallest associated Euclidean distance were considered.

(3) *Model checking*: Model checking was performed according to the procedure described in Fagundes *et al.* (2007). We randomly sampled 500 replicates from the five million simulations performed for each model, and used them as “pseudo-observed” datasets. For each dataset we applied the same model choice procedure to compute the posterior probabilities of each of the compared models. The relative distributions of these probabilities over the 500 replicates were then used to compute the probability that the best-supported model, obtained during model-choice procedure, is indeed the true simulated model (Fagundes *et al.*, 2007; Cornuet *et al.*, 2010; Roux *et al.*, 2011).

(4) *Parameter estimation*: The joint posterior distribution of parameters describing the best model was obtained by weighted non-linear multivariate regressions of the parameters on the summary-statistics (Blum and Francois 2010). When the best model involved heterogeneous gene flow, we then estimated the locus-specific migration rate parameters. Hence, for each locus we ran 1.5×10^6 random coalescent simulations using parameter values sampled in the joint-posterior distribution for the five parameters common to all loci (population mutation rates and times) obtained using the procedure described above. Finally, we applied the described rejection/regression analysis to the simulations performed for each locus to jointly estimate both effective migration rate parameters. Computations were performed using 50 trained neural networks and 15 hidden networks in the regression (R package “abc”, Csilléry *et al.* 2012).

File S2

Code implementing inferential methods

File S2 is available for download as a zip file.

It contains source code implementing log-likelihood inference (see folder Code_InL) as well as a simple version of the ABC inference (see folder Code_ABC).

Literature Cited

Beaumont, M. A., 2008 DFDIST. <http://www.maths.bris.ac.uk/~mamab/stuff/>

Beaumont, M. A., and R. A. Nichols, 1996 Evaluating loci for use in the genetic analysis of population structure. *Proc. Biol. Sci.* 263(1377): 1619–1626.

Beaumont, M. A., and D. J. Balding, 2004 Identifying adaptive genetic divergence among populations from genome scans. *Mol. Ecol.* 13(4): 969–980.

Bierne, N., 2010 The distinctive footprints of local hitchhiking in a varied environment and global hitchhiking in a subdivided population. *Evolution* 64(11): 3254–3272.

Blum, M. G. B., and O. Francois, 2010 Non-linear regression models for approximate Bayesian computation. *Stat. Comput.* 20(1): 63–73.

Bonhomme, M., C. Chevalet, B. Servin, S. Boitard, J. Abdallah et al., 2010 Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics* 186: 241–262.

Cornuet, J.-M., V. Ravigné, and A. Estoup, 2010 Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics* 11(1): 401.

Csilléry, K., O. François, and M. G. B. Blum, 2012 abc: an R package for approximate Bayesian computation (ABC). *Methods Ecol. Evol.* 3(3): 475–479.

Daguin, C., F. Bonhomme, and P. Borsa, 2001 The zone of sympatry and hybridization of *Mytilus edulis* and *M. galloprovincialis*, as described by intron length polymorphism at locus *mac-1*. *Heredity* 86: 342–354.

Fagundes, N. J. R., N. Ray, M. Beaumont, S. Neuenschwander, F. M. Salzano et al., 2007 Statistical evaluation of alternative models of human evolution. *Proc. Natl. Acad. Sci. USA* 104(45): 17614–17619.

Foll, M., and O. Gaggiotti, 2008 A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180: 977–993.

Fourcade, Y., A. Chaput-Bardy, J. Secondi, C. Fleurant, and C. Lemaire, 2013 Is local selection so widespread in river organisms?: fractal geometry of river networks leads to high bias in outlier detection. *Mol. Ecol.* 22(8): 2065–2073.

Hamilton, G., M. Stoneking, and L. Excoffier, 2005 Molecular analysis reveals tighter social regulation of immigration in patrilocal populations than in matrilocal populations. *Proc. Natl. Acad. Sci. USA* 102(21): 7476–7480.

Lewontin, R. C., and J. Krakauer, 1973 Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74: 175–195.

Robertson, A., 1975 Gene frequency distributions as a test of selective neutrality. *Genetics* 81: 775–785.

Roux, C., V. Castaic, M. Pauwels, S. I. Wright, P. Saumitou-Laprade et al., 2011 Does speciation between *Arabidopsis halleri* and *Arabidopsis lyrata* coincide with major changes in a molecular target of adaptation? *PLoS ONE* 6(11): e26872.

Roux, C., C. Fraise, V. Castaic, 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. *J. Evol. Biol.* *in press*.

Vitalis, R., K. Dawson, P. Boursot, and K. Belkhir, 2003 DetSel 1.0: a computer program to detect markers responding to selection. *J. Hered.* 94(5): 429–431.

FIGURE S1. Genome scan of inter and intra-species differentiation. (A) shows F_{ST} values between species against values within *M. edulis* (North Sea vs Bay of Biscay). All other details match Figure 2A. (B) shows frequency of the *edulis* allele at the two outlier loci in the four panmictic patches of the studied area. All other details match Figure 2B.

FIGURE S2. Studied sequence around *mac-1*. Position of the four polymorphisms chosen for extensive genotyping is indicated by stars for SNPs and by triangles for PCR product length polymorphisms. The second polymorphism, *mac-1*, corresponds to the outlier locus identified by genome scan. The sequence has been segmented in three regions (region 1 of 880 bp; region 2 of 1500 bp ; region 3 of 750 bp) for gene tree reconstruction.

FIGURE S3. Allele genealogies reconstructed by the neighbour-joining algorithm apply to the number of nucleotide differences. (A) region 1 and (B) region 3 of the studied sequence. Sequences sampled in *M. galloprovincialis* patches are in red (Iberian Coast) and orange (Brittany); while those sampled in *M. edulis* patches are in dark blue (North Sea) and cyan (Bay of Biscay).

FIGURE S4. Parameter estimation in the adaptive introgression model. Maximum likelihood estimation of θ (A) and M (B) is depicted in black. Posterior distribution of θ (A), M (B) and α (C) is depicted by grey dots. ML and ABC inferences were based on 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). All other details match Figure 6.

FIGURE S5. Goodness-of-fit to the adaptive introgression model. For each statistic, (1) the distribution in red was obtained from 50,000 simulations using the posterior parameter estimate of N , M and θ ; (2) the distribution in blue was obtained from 1000 simulations using the maximum likelihood value of θ and M , other parameters were set at $N = 50000$, $\alpha = 700$, $n_d = 177$ and $n_r = 60$. A dashed vertical line shows the observed value for each statistic. Statistics used were: k , number of allelic classes; He , expected heterozygosity; Sd , standard deviation in frequencies over all allelic classes; $\%P$, proportion of private alleles; F_{ST} , level of differentiation of *edulis* alleles; He_r / He_d , proportional diversity after introgression where d denotes the donor species and r denotes the recipient species.

FIGURE S6. Effect of recombination in the adaptive introgression model. The proportional diversity after introgression from d (donor species) to r (recipient species), He_r / He_d , is shown as a function of the scaled migration rate, M . Different colours correspond to different values of the recombination rate, r (see legend). Error bars represent standard deviations calculated over 5,000 replicates. Median values are depicted by dots. Other parameters were set at: $N = 10000$, $\theta = 2$, $\alpha = 1000$ and $n_d = n_r = 50$ (sample size). Simulations were run until the beneficial allele reached fixation in the recipient background.

FIGURE S7. Effect of temporal variability in migration rate in the adaptive introgression model. The proportional diversity after introgression from d (donor species) to r (recipient species), He_r / He_d , is shown as a function of the scaled migration rate, M . Green dots indicate a constant migration rate every generation ($M=2Nm$). Blue dots refer to a pulsed migration regime in which $M=10 \cdot 2Nm$ every 10 generations, otherwise $M=0$; so that on average $M=2Nm$. Error bars represent standard deviation calculated over 1,000 replicates. Median values are depicted by dots. Other parameters were set at: $N = 10000$, $\theta = 2$, $\alpha = 1000$ and $n_d = n_r = 50$ (sample size). Simulations were run until the beneficial allele reached fixation in the recipient background.

FIGURE S8. Models of speciation: Strict Isolation (SI), Isolation with Migration (IM), Ancient Migration (AM), and Secondary Contact (SC). Four parameters are shared by all models: T_{split} is the number of generations since the speciation time; N_{anc} , N_{edulis} and N_{gallo} are the number of effective individuals in the ancestral population, *M. edulis* and *M. galloprovincialis* respectively. T_{iso} is the number of generations since the two nascent species stopped exchanging migrants in the AM model. T_{sc} is the number of generations since the two daughter species experienced secondary contact after a period of isolation in the SC model. The asymmetrical migration rates $M_{gallo \text{ to } edulis}$ and $M_{edulis \text{ to } gallo}$ are expressed in $2Nm$ units, where m is the proportion of a population made up of migrants from the other population per generation.

FIGURE S9. Goodness-of-fit of the best supported demographic model to the pairwise F_{ST} comparisons. (A) and (B) show the distribution of F_{ST} between species given respectively for the *M. galloprovincialis* patch of the Iberian Coast and Brittany; while (C) shows the distribution of F_{ST} within *M. galloprovincialis*. Observed F_{ST} values at the 422 AFLPs markers are shown in black. Simulated F_{ST} values (in grey) were obtained based on 100 simulations of 422 SNPs in a 3-patch model using the posterior estimates of the secondary contact model with heterogeneity of introgression rates (see File S1). Dashed vertical lines show mean F_{ST} values; red vertical lines denote the F_{ST} values observed at *mac-1*. In each pairwise comparison, *mac-1* is out of the 95%-quantile of the simulated distribution: (A) p -value = 0.0037, (B) p -value = 0.1010 (C), p -value = 0.0049.

TABLE S1 Goodness-of-fit in the adaptive introgression model

statistic	data	ML				ABC			
		quantiles			<i>p-value</i>	quantiles			<i>p-value</i>
		5%	50%	95%		5%	50%	95%	
k_d	15	10	15	21	0.42	8	13	21	0.31
k_r	6	4	8	12	0.13	5	10	15	0.05
He_d	0.76	0.6	0.81	0.89	0.29	0.51	0.78	0.89	0.42
He_r	0.73	0.39	0.72	0.84	0.46	0.49	0.78	0.89	0.34
Sd_d	0.11	0.06	0.09	0.17	0.33	0.06	0.11	0.22	0.49
Sd_r	0.15	0.09	0.15	0.31	0.48	0.06	0.12	0.26	0.31
$\%P_d$	0.6	0.37	0.56	0.72	0.31	0.14	0.35	0.55	0.02*
$\%P_r$	0	0	0.14	0.29	0***	0	0.09	0.25	0***
F_{ST}	0.016	0.01	0.06	0.19	0.12	-0.02	0.01	0.05	0.24
He_r/He_d	0.96	0.54	0.9	1.11	0.28	0.85	1	1.14	0.26

The donor population is represented by the two patches of *M. edulis* (North Sea and Bay of Biscay). The recipient population is represented by the Brittany patch of *M. galloprovincialis*. Only edulis-derived alleles were considered.

k: number of allelic classes; ***He***: expected heterozygosity (Nei 1978); ***Sd***: standard deviation in frequencies over all allelic classes; ***%P***: proportion of private alleles; ***F_{ST}***: level of differentiation of edulis alleles (Nei 1973); ***He_r / He_d***: proportional diversity after introgression.

TABLE S2 Summary statistics observed at the eight nuclear loci

Locus	n_{edu}	n_{gal}	L_{tot}	L_{sil}	average across the sampled sequences											
					π_{edu}	π_{gal}	Θ_{edu}	Θ_{gal}	D_{edu}	D_{gal}	$(1-\pi_s/\pi_T)$	$\text{netdiv}_{\text{edu-gal}}$	$S_{x,\text{edu}}$	$S_{x,\text{gal}}$	S_f	S_s
EF1	20	20	928	608	0.0229	0.0133	0.032	0.0246	-1.1546	-1.8565	0.5887	0.0505	57	41	0	12
EF2	20	20	1108	781	0.0033	0.0126	0.009	0.0162	-2.4481	-0.8926	0.3888	0.0099	24	44	1	1
glucanase	18	14	889	460	0.0269	0.018	0.0278	0.0267	-0.1324	-1.4093	0.0679	0.0021	21	16	0	23
mac-1	12	20	824	824	0.0043	0.0157	0.004	0.0216	0.2908	-1.103	0.1826	0.0017	4	57	0	6
Mannanase	20	18	983	558	0.0194	0.0236	0.0253	0.0328	-0.9344	-1.1709	0.0284	0.0014	29	42	0	21
mc125	19	20	502	104	0.027	0.074	0.0385	0.084	-1.1099	-0.4689	0.0463	0.0036	7	24	0	7
mgd2	20	20	1083	950	0.0359	0.042	0.0362	0.0587	-0.0396	-1.1856	0.0145	0.0011	50	126	0	72
Mytilin	16	16	754	547	0.041	0.0419	0.0424	0.0331	-0.1427	1.1394	0.0902	0.008	27	10	0	50

M. edulis is represented by the North Sea patch. *M. galloprovincialis* is represented by the Brittany patch.

n : total number of sequences; L_{tot} : total length; L : silent length excluding all gaps from the total alignment; π : number of pairwise differences (Tajima, 1983); Θ : Watterson's θ (Watterson, 1975); D : Tajima's D (Tajima, 1989a, 1989b); $1-\pi_s/\pi_T$: level of species differentiation, where π_s is the average pairwise nucleotide diversity within species and π_T is the total pairwise nucleotide diversity of the pooled sample across species; $\text{netdiv}_{\text{edu-gal}}$: net molecular divergence measured at synonymous positions; S_x : number of exclusive polymorphic sites; S_f : number of fixed differences; S_s : number of shared polymorphic sites.

TABLE S3 Relative posterior probabilities of different models of speciation

Models		Comparison	
		within models (homo vs hetero)	between models
Stric Isolation (SI)	x	x	0.0017
Isolation with Migration (IM)	homo	0.0675	x
	hetero	0.9325	0.3065
Ancient Migration (AM)	homo	0.4226	x
	hetero	0.5774	0.0794
Secondary Contact (SC)	homo	0.1177	x
	hetero	0.8823	0.6125

We first compare homogeneous ('homo') vs heterogeneous ('hetero') scenarios of introgression rates within the IM, AM and SC models. We then compare the best version of each model with gene flow (IM, AM and SC) to the SI model.

TABLE S4 Test of robustness for the secondary contact model with heterogeneous introgression rates

	True model				P _{type I error}
	Stric Isolation (SI)	Isolation with Migration (IM)	Ancient Migration (AM)	Secondary Contact (SC)	
Prob (SC True model)	0.0002	0.1785	0.0076	0.8137	0.013

Values are the estimated relative probabilities of the SC model when the SI, the IM, the AM and the SC models are the true models. They were calculated at the SC posterior probability that we observed ($P_{\text{post}} = 0.61$). Estimations are based on 500 pseudo-observed datasets for each model.

We used them to compute the probability that SC is the correct model given that $P_{\text{post}} = 0.61$. Probability of type I error is one minus this probability so that $1 - 0.987 = 0.013$.

We assumed heterogeneity of introgression rates for models including migration (IM, AM, SC) as they are strongly supported compared to their homogeneous alternatives (Table S3).

TABLE S5 Relative posterior probabilities of different scenarios of introgression at *mac-1* in the secondary contact model

	SC ₁	SC ₂	SC ₃
SC ₀	SC ₀ : 0.3005	SC ₀ : 0.0003	SC ₀ : 0.0025
	SC₁: 0.6995	SC₂: 0.9997	SC₃: 0.9975
SC ₁		SC ₁ : 0.0013	SC ₁ : 0.0251
		SC₂: 0.9987	SC₃: 0.9749
SC ₂			SC₂: 0.9138
			SC ₃ : 0.0862

SC₀: no introgression;

SC₁: introgression into *M. edulis* only;

SC₂: introgression into *M. galloprovincialis* only;

SC₃: introgression in both directions at different rates.

TABLE S6 List of primers used for sequencing around *mac-1*

Region	Forward primers	Reverse primers
1	<i>region1-F</i> : 5'-AAGATTTATTTATATATGCGTGTG	<i>region1-galA-R1</i> : 5'-ATAAGCAGATTGATAAAATAAATCG <i>region1-galA-R2</i> : 5'-AATATAATTATACAATATAAGCATG <i>region1-galB-R1</i> : 5'-CTGGGTACCATAGCAGATTAGTTTAG <i>region1-galB-R2</i> : 5'-TACTGGGTACCATAGCAGATTGATAA <i>region1-galC-R1</i> : 5'-TTTGCTTATTGTCGTCAAGGAATTG <i>region1-galC-R2</i> : 5'-TATGATTTGCTTATTGTCGTGGAAG <i>region1-edu-R1</i> : 5'-TTAATCAATTTAAGTACTACGCTAA <i>region1-edu-R2</i> : 5'-TTAATCAATTTAAGTACTACGCTAG
2 start - middle	<i>region2-gal-F1</i> : 5'-AATTCGATTTATTTTATCAATCTGC <i>region2-gal-F2</i> : 5'-GACATAAATTATGAATAACATGCT <i>region2-edu-F1</i> : 5'-TTTTACTAAATAAGTCGAAATTACC <i>region2-edu-F2</i> : 5'-TTTTACTAATTAAGTCGAAATTACG	<i>region2-middle-R</i> : 5'-AGGCTATATTCGGTATGTACT
2 middle - end	<i>region2-middle-F</i> : 5'-AGTACATACCGAATATAGCC	<i>region2-R</i> : 5'-TGTGACAATTCGGTGTCAATTGGG
3	<i>region3-gal-F</i> : 5'-TATAAACCCCACTTCCGTTTT <i>region3-edu-F</i> : 5'-TATAAACCCCACTTCCTTTCT	<i>SNP-R</i> : 5'-CTGTTTGCTGATCCACATCTGTTGG
3 middle - end	<i>SNP-ex2-F</i> : 5'-GGTATGGGACAGAAAGACAGCTACG	<i>SNP-R</i> : 5'-CTGTTTGCTGATCCACATCTGTTGG

TABLE S7 List of primers used for genotyping each of the four polymorphisms along the sequence

Locus	Forward primers	Reverse primers
<i>indel-in0</i>	<i>indel-in0</i> -F: 5'-CATGTCTCTATACCATTGTACTTTGG	<i>indel-in0</i> -R: 5'-AGACCATACGTAAAGGAAATGTATG
<i>mac-1</i>	<i>mac-1</i> -F: 5'-CGTCTAGCGTAGTACTTAAATTG	<i>mac-1</i> -R: 5'-CGAAAATTGTAGTCTAGTTTTGTG
<i>SNP-in1</i>	<i>SNP-in1</i> -F1 5'-TATAAACCCCACTTCCTTTCT	<i>SNP</i> -R: 5'-CTGTTTGCTGATCCACATCTGTTGG
	<i>SNP-in1</i> -F2: 5'-TATAAACCCCACTTCCGTTTT	
<i>SNP-ex2</i>	<i>SNP-ex2</i> -F: 5'-GGTATGGGACAGAAAGACAGCTACG	<i>SNP</i> -R: 5'-CTGTTTGCTGATCCACATCTGTTGG