MOLECULAR ECOLOGY

Accounting for semi-permeability to gene flow using Approximate Bayesian Computation improves inference into the history of speciation: application to a mussel hybrid zone.

Journal:	*Molecular Ecology*
Manuscript ID:	MEC-13-0777
Manuscript Type:	Original Article
Date Submitted by the Author:	15-Jul-2013
Complete List of Authors:	Roux, Camille; UNIL, DEE Fraisse, Christelle; CNRS-Université de Montpellier 2, Castric, Vincent; CNRS-Université de Lille 1, ; Vekemans, Xavier; CNRS-Université de Lille 1, Pogson, Grant; University of California, Santa Cruz, Bierne, Nicolas; CNRS-Université de Montpellier 2,
Keywords:	Speciation, Population Genetics - Empirical, Molluscs, Molecular Evolution, Hybridization

Accounting for semi-permeability to gene flow using Approximate Bayesian Computation improves inference into the history of speciation: application to a mussel hybrid zone.

Camille Roux[1,2,3,4], *Christelle Fraïsse*[1,2], *Vincent Castric*[4], *Xavier Vekemans*[4], *Grant H. Pogson*[5] & *Nicolas Bierne*[1,2]

1) Université Montpellier 2, Place Eugène Bataillon, F34095 Montpellier, France.

2) CNRS-UMR5554 Institut des Sciences de l'Evolution, Station Méditerranéenne de l'Environnement Littoral, 2 rue des Chantiers, F34200 Sète, France.

3) Department of Ecology and Evolution, University of Lausanne, Biophore, Lausanne, Switzerland

4) Université Lille Nord de France, USTL, GEPV, CNRS, FRE 3268, F-59650 Villeneuve d'Ascq, France.

5) Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA 95064, USA.

Abstract

The use of molecular data to reconstruct the history of divergence and gene flow between populations of closely related taxa is a challenging problem that has recently received considerable attention. Barriers to gene flow observed in secondary contact hybrid zones, or between parapatric populations undergoing speciation with gene flow, are often semi-permeable -*i.e.,* genomic regions experience variable levels of introgression depending on their linkage to isolation genes. However, most demographic inference methods have neglected this source of variation and assumed that the gene flow parameter (*Nm*) is similar among loci. Here, we evaluate the improved performance of the Approximate Bayesian Computation (ABC) approach by analysing DNA sequences sampled from populations of the marine mussels *Mytilus edulis* and *M. galloprovincialis* across a well-studied mosaic hybrid zone in Europe where the patterns of introgression are highly variable among loci. A comparison of nested models revealed that a model allowing for heterogeneous gene flow across loci outperformed a model assuming equal migration rates. By incorporating this heterogeneity, our simulations suggest that the two mussel species had experienced a long period of allopatric isolation followed by recent secondary contact. By contrast, constraining migration to be homogeneous failed to discriminate among the different models of gene flow tested. Our results demonstrate that genomic variation in introgression rates can have profound impacts on the biological conclusions drawn from inference methods and that accounting for the semi-permeability of genetic barriers is an important step towards more realistic reconstructions of speciation scenarios.

Introduction

A number of recent approaches have been developed by evolutionary biologists to reconstruct the history of divergence and gene flow between populations or closely related taxa using molecular data (Hey & Nielsen 2004, 2007; Becquet & Przeworski 2007). However, this task is challenging because the true history of population divergence is often much more complex than the models fitted to the data. One difficulty that has received limited attention is that genetic barriers to gene flow observed between parapatric populations undergoing speciation with gene flow, or diverged populations experiencing secondary contact, are often semi-permeable. This leads to genome-wide heterogeneity (GWH) in the effective levels of gene flow due to the direct effect of selection on isolation genes as well as indirect effects on neutral loci depending on their linkage to selected genes (Barton 1979; Barton & Bengtsson 1986; Harrison 1993; Charlesworth *et al.* 1997; Nosil & Feder 2012). The indirect effects of selection produce patterns of gene flow ranging from low introgression in the neighbourhood of barrier loci (so-called genomic islands of differentiation) to basal introgression rates in regions devoid of selected loci, and maximal introgression around loci that experienced the fixation of an unconditionally favourable allele (i.e., adaptive introgression; Pialek & Barton 1997). Furthermore, heterogeneity in genomic patterns of differentiation could also result from temporal effects when successive fixations at new barrier loci sequentially lock up different genome regions at different times (Wu 2001).

Methods to infer the history of divergence and gene flow between closely related organisms from DNA sequence data have flourished during the last decade (Hey 2006; Becquet & Przeworski 2009; Pinho & Hey 2010) with a progressive increase in the complexity of the underlying scenarios. The first and most frequently used method, the so-called Isolation with Migration (IM), considers the divergence of two populations from T generations in the past that continue to exchange genes at a fixed rate, Nm (Nielsen & Wakeley 2001; Hey & Nielsen 2004, 2007). Although it has proved to be very useful, IM may be sensitive to violations of certain

assumptions, such as the absence of intragenic recombination or uninterrupted gene flow (Strasburg & Rieseberg 2010), and more complex scenarios have been proposed (Becquet & Przeworski 2009). Surprisingly, the fact that the divergence time and gene flow parameters are assumed to be shared among loci has rarely been questioned. However, two recent studies have proposed inference methods that explicitly account for the heterogeneity of gene flow among loci (Sousa *et al.* 2013; Roux *et al.* 2013). The approach of Sousa *et al.* (2013) extends the IM method by considering two or more groups of loci with different demographic parameters (migration rates and effective population sizes) and assigns each locus to a given group using a Bayesian method. The approach of Roux *et al.* (2013) takes advantage of the flexibility offered by Approximate Bayesian Computation (ABC) to investigate alternative demographic scenarios and to consider the migration rate parameter of each locus as a random variable drawn itself from a distribution that is estimated from the data (according to a hierarchical Bayesian approach with hyper-parameters). Both methods provide the ability to investigate GWH and have also suggested that neglecting such heterogeneity may lead to erroneous conclusions about speciation. For example, the best supported scenario involving GWH in the secondary contact between the types A and B species of *Ciona intestinalis* was not supported by a model assuming a single shared migration parameter because most of the genome was blocked from introgressing. This suggests that GWH may severely bias inferences when genetic barriers are porous and analysed with a handful of loci, which is typical in reconstructing the history of divergence and gene flow in non-model species.

A good system to test scenarios of speciation allowing GWH is the hybrid zone between the marine mussels *Mytilus edulis* and *M. galloprovincialis* where a semi-permeable barrier to gene flow has been previously demonstrated and extensively studied. The geographic structure of the zone is a mosaic of parental and hybrid populations along the Atlantic coasts of France (Bierne *et al.* 2003) and the British Isles (Skibinski *et al.* 1983). The interspecific barriers to gene

flow are due to a number of pre- and post-zygotic, intrinsic and extrinsic, isolating mechanisms including spawning asynchrony (Secor *et al.* 2001), habitat choice (Bierne *et al.* 2003), assortative fertilization (Bierne *et al.* 2002), directional selection (Gardner & Skibinski 1988; Hilbish *et al.* 2002), and hybrid fitness depression attributable to a large number of recessive genetic incompatibilities dispersed across the entire genome (Bierne *et al.* 2006). Finally, introgression rates have been shown to vary strongly among loci (Skibinski *et al.* 1983; Boon *et al.* 2009) and the *Mytilus* hybrid zone represents one clear example of genetic barriers where semi-permeability is strongly pronounced.

In this paper, new and previously published DNA sequence polymorphism data from eight nuclear loci were used to reconstruct the history of divergence and gene flow between the two mussel species. To allow for heterogeneity in migration rates among loci we used the hierarchical ABC approach with hyper-parameters developed by Roux *et al.* (2013). We were unable to apply the alternative method proposed by Sousa *et al.* (2013) because intragenic recombination was widespread in our data (e.g. Boon *et al.* 2009) and because the observed hybrid zone most likely results from secondary contact (Quesada *et al.* 1998; Boon *et al.* 2009), a scenario for which the IM method may provide misleading results (Becquet & Przeworski 2009). Indeed, one of our main objectives was to assess whether allowing GWH could improve our ability to discriminate among alternative scenarios (i.e., secondary contact *vs.* parapatric primary differentiation). We explicitly test the effect of allowing effective migration rates to vary among loci by comparing nested models with either homogeneous or heterogeneous migration and evaluate alternative scenarios of speciation. We then compare estimates of divergence times and the onset of secondary contact between scenarios and discuss the usefulness of these new approaches in studies on the evolutionary processes occurring in hybrid zones.

Materials and methods

DNA polymorphism

We used *Mytilus* spp. samples collected at two localities known to represent pure patches of *Mytilus edulis*, WS (Wadden Sea, Holland) and *M. galloprovincialis*, FA (Faro, Algarve, Portugal). The genetic composition of these samples have been analysed previously with DNA fragment length-polymorphism and AFLP markers (Bierne *et al.* 2003; Faure *et al.* 2008; Boon *et al.* 2009; Gosset & Bierne 2013). In addition to the previously published nucleotide sequence data at three loci (Faure *et al.* 2008; Boon *et al.* 2009), we obtained data from five new loci (average fragment length was ~900 bp). PCR primers are described in Supplementary Table S1. With the exception of locus mc125, which consisted exclusively of coding sequence (Addison *et al.* 2008), all other loci targeted a fragment of non-coding DNA (intron or intergenic). A standard protocol was used for the PCR reactions using the Promega GoTaq® DNA polymerase (Promega, Madison, WI, USA). Sequences were cloned following the mark-recapture (MR)-cloning protocol (Bierne *et al.* 2007; Faure *et al.* 2007, 2008; Boon *et al.* 2009). Individual PCR reactions were labelled with unique molecular tags using 5'-tailed primers. Tagged PCR products of similar quantities were mixed together and cloned into a pGEM-T vector by using Promega pGEM-T cloning kits and sequenced with the universal primers SP6 and T7 flanking the insert at the Genoscope platform (http://www.genoscope.cns.fr/). To avoid sampling bias and to minimise the number of artifactual mutations produced during PCR, cloning and sequencing we used a single allele per individual, chosen as the most frequently captured variant.

Data analyses

Only silent positions (*i.e.,* synonymous polymorphisms in coding regions and non-coding polymorphisms in introns or intergenic regions) were used to study the demographic history of *Mytilus* populations. The data were summarized by a widely used array of statistics for demographic inference (Wakeley & Hey 1997; Becquet & Przeworski 2007; Ross-Ibarra *et al.* 2008; Roux *et al.* 2011). We computed classical diversity estimators (nucleotide diversity, π, and Watterson's θ_W) (Watterson 1975; Tajima 1983), between-species differentiation measured by F_{ST} (computed as 1- π_s/π_T where π_s is the average pairwise nucleotide diversity within population and π_T is the total pairwise nucleotide diversity of the pooled sample across populations), and the departure of site frequency spectrum from mutation/drift equilibrium by Tajima's *D* (Tajima 1989) using a routine written in C (MScalc, available from http://www.abcgwh.sitew.ch/; Roux *et al.* 2011). In addition, we classified the observed polymorphic sites into four distinct categories: (1) polymorphisms exclusive to *M. edulis* noted Sx_{edu}, (*i.e,*. polymorphic sites for which only one allele was found in *M. galloprovincialis*, but two alleles segregate in *M. edulis*); (2) polymorphisms exclusive to *M. galloprovincialis* noted Sx_{gal}; (3) fixed differences between species (noted *Sf*); and (4) shared polymorphic sites (noted *Ss*) (*i.e.,* sites for which the same two alleles were segregating in both species). To estimate intragenic recombination rate ρ (=$4Nr$, with *N* the effective population size and *r* the recombination rate per nucleotide site), we used a composite-likelihood approach (McVean *et al.* 2002) implemented in the PAIRWISE program of the LDhat 2.1 package.

Inferring ancestral demography

Coalescent simulations

We used an ABC framework (Tavaré *et al.* 1997; Beaumont *et al.* 2002) to investigate three scenarios of speciation with gene flow (Isolation with Migration, IM; Ancient Migration, AM; and Secondary Contact, SC; Fig. 1) and one without gene flow (Strict Isolation, SI). All scenarios assume an instantaneous split of an ancestral population into two daughter populations of constant sizes. The IM scenario assumes continuous gene flow between populations. In the AM scenario the migration events are restricted to the initial phase of speciation, whereas in the SC scenario the two daughter populations begin to evolve in strict isolation and then experience secondary contact. For scenarios with gene flow, we used the scaled migration rates M=4Nm (with *M1* the migration rate from *M. galloprovincialis* to *M. edulis* and *M2* the migration rate to *M. galloprovincialis*, time being defined forward), where m is the fraction of the population that is composed of migrants from the other population each generation. For the IM, AM, and SC scenarios, two alternative models were compared representing the hypotheses of identical *vs* variable effective migration rates among loci ("homogeneous" vs. "heterogeneous" models, respectively). We thus compared a total of seven models. Following Roux *et al.* (2013) the "heterogeneous" models consisted of hierarchical Bayesian models with migration rate parameters for each locus drawn from a scaled-Beta distribution characterized by three hyper-parameters (the alpha and beta shape parameters of the Beta distribution and a scalar “c” to which the Beta distribution is multiplied). This distribution accommodates a large variety of distinct shapes, while avoiding the pitfalls of over-parameterization. Five million multilocus simulations were performed for each model. We used large uniform prior distributions for all parameters, with identical prior distributions for parameters common to all models. Prior distributions for $\theta_{edu}/\theta_{ref}$, $\theta_{gal}/\theta_{ref}$ and θ_A/θ_{ref} were uniform on the interval 0-20 with θ_{ref}=4.N_{ref}.μ. 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. This rate was estimated from analysis of divergence between *M. californianus* and species from the *Mytilus edulis* complex, assuming a divergence time of 7.6 MY (Ort & Pogson 2007) and a generation time of 2 years. θ_{edu}, θ_{gal} and θ_A are values for θ of the *M. edulis*, *M. galloprovincialis* and ancestral populations respectively. We sampled $T_{split}/4.N_{ref}$ from the interval 0-25 generations, $0\text{-}10^7$ generations in demographic units. The parameters T_{iso} and T_{SC} were drawn from a uniform distribution on the interval $0\text{-}T_{split}$. Prior distributions for scaled migration rates in both directions were uniform on the interval 0-20.

In the "homogeneous" models, values of the two migration rate parameters *M1* and *M2* were randomly sampled from the uniform prior interval 0-30 for all loci. For the alternative "heterogeneous" models, a single combination of the shape parameters from the Beta distribution was first randomly and independently sampled for each multilocus simulation from the uniform intervals 0-5 for alpha and 0-200 for beta. Then, for each locus the two migration rate parameters *M1* and *M2* were randomly sampled from the Beta distribution. Prior distributions were computed using a modified version of the Priorgen software (Ross-Ibarra *et al.* 2008), and coalescent simulations were run using Msnsam (Ross-Ibarra *et al.* 2008), a modified version of the ms program (Hudson 2002) that allows for different sample sizes at each locus.

Model testing

In order to statistically evaluate alternative models of speciation, we followed a two-step hierarchical procedure (Fagundes *et al.* 2007). First, for each scenario allowing migration (IM, AM and SC), we evaluated posterior probabilities for the two alternative models (homogeneous and heterogeneous). Next, we compared the best models from these scenarios in addition to the SI scenario. 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 under the ABC framework using the R package "abc" (Csillery *et al.* 2012). 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 $S_{obs}=S_{sim}$. Computations were performed using 50 trained neural networks and 15 hidden networks in the regression.

To perform model checking, we randomly sampled 1,000 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 1,000 replicates were then used to compute the probability that the best-supported model (*i.e.,* the one with the highest value of the posterior probability obtained from the simulated dataset) is indeed the true simulated model (Fagundes *et al.* 2007; Cornuet *et al.* 2008). Hence, one minus this probability gives the probability of type I error (*i.e.,* the probability of rejecting a true hypothesis = *P*-value).

Parameter estimation

We first estimated parameters shared by all loci and then inferred migration rates for each locus under the heterogeneous models.

Parameters were log-tangent transformed (Hamilton *et al.* 2005) and only the 2,000 replicate simulations with the smallest associated Euclidean distance $\delta=||S_{obs}-S_{sim}||$ were considered. The joint posterior distribution of parameters describing the best model was then obtained by weighted non-linear multivariate regressions of the parameters on the summary-statistics (Blum & François 2009). For each regression, 50 feed-forward neural networks and 15

hidden networks were trained using the R package "abc" (Csillery *et al.* 2012). When the best model involved heterogeneous gene flow, we then estimated the locus-specific migration rate parameters *M1* and *M2*. Hence, for each locus we ran $1.5x10^6$ random coalescent simulations using parameter values sampled in the joint-posterior distribution for the five parameters common to all loci (N_A, N_B, N_{anc}, T_{split}, and T_{SC}) 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.

Results

Levels of DNA polymorphism and distribution of variable sites

Twelve to 24 multiply captured individual sequences in *M. edulis*, and 14 to 20 in *M. galloprovincialis* were obtained from the cloning experiment resulting in ~5.3 Kb of alignable silent sites including 737 biallelic positions (Table 1). Both species exhibited similar levels of silent nucleotide diversity when measured with either π (π_{edu}=0.0213 and π_{gal}=0.0256, Wilcoxon signed-rank test V=17, p=0.9453) or Watterson's θ (θ_{edu}=0.0256 and θ_{gal}=0.0317, V=8; p=0.3525). Silent segregating sites were mostly specific to each species (246 and 295 sites were exclusively polymorphic in *M. edulis* and *M. galloprovincialis,* respectively), but a large proportion of the polymorphic positions were shared by the two species (196 sites). Remarkably, the two species exhibited no fixed silent differences but the distributions of pairwise nucleotide divergence between species were clearly not unimodal (Fig. 2; unimodality was rejected for each locus by the Hartigan's DIP-test; Hartigan & Hartigan 1985). For several loci the distributions of pairwise nucleotide divergence between species appeared to be bimodal (Fig. 2) as expected when two diverged species actively exchange a small category of genes following secondary contact.

Variation in migration rates among loci and the timing of gene flow between *M. edulis* and *M. galloprovincialis*.

Using an ABC approach with explicit modelling of intralocus recombination (as measured using the LDhat 2.1 package (McVean *et al.* 2002), Table S2), we first applied a model choice procedure for each demographic scenario implementing migration (Fig. 1) to evaluate alternative scenarios of gene flow and test if the heterogeneous model explained the data better than the homogeneous model. For all three scenarios incorporating gene flow we observed unambiguous support in favour of the heterogeneous migration over the commonly used homogeneous alternative (Table 2). The posterior probabilities of the heterogeneous models were always higher than the homogeneous models and analyses using pseudo-observed datasets obtained by simulations indicated that these differences were highly significant. Figure 3 shows that the ABC approach had considerable power to detect a semi-permeable barrier to gene flow for each demographic scenario: 100%, 99.9% and 100% of pseudo-observed datasets simulated under the IM, AM, and SC scenarios with heterogeneous gene flow were correctly supported by our model choice procedure (*i.e.*, associated with posterior probabilities above 0.5). Indeed, supporting either of the two alternative models did not require a very high posterior probability (Fig. S1).

We then studied the temporal pattern of migration by applying the model choice procedure to comparisons between the SI scenario and the IM, AM and SC scenarios with heterogeneous migration (Table 2). Scenarios allowing for ongoing migration (IM and SC) had an elevated cumulated posterior probability (0.92), which strongly rejected the hypothesis that *M. edulis* and *M. galloprovincialis* represent fully isolated species. More importantly, the SC scenario was the best supported scenario suggesting that migration between the two species is recent evolutionary event through secondary contact following a period of allopatric isolation.

By analysing pseudo-observed datasets with simulations, we found that the probability that SC was the correct scenario given the posterior probability of 0.52 was 0.957 (*P*-value = 0.043, Fig. S2). It is worth emphasising that the statistical distinction between the SC and IM scenarios was only found when migration rates were allowed to vary among loci; support for the SC scenario disappeared when gene flow was assumed to be homogeneous among loci.

Inference of the historical parameters describing the best-supported SC scenario

The length of time spent in allopatry relative to the initial time of divergence is an important factor determining the global strength of barriers to gene flow under the SC scenario of speciation for the two mussel species. To explore the timing of these events we first inferred the five parameters common to all loci (N_{edu}, N_{gal}, N_{anc}, T_{split} and T_{SC}), then estimated the parameters of the genomic distribution of migration rates. The joint-posterior distribution from 2,000 accepted simulations was strongly differentiated from the prior of each parameter suggesting that our data provides sufficient information and that we explored the correct parameter space (Fig. 4). In Table 3 we report the 95% highest posterior density interval for the five parameters shared by all loci as well as the mode and median of each posterior distribution. Our estimates of the effective population size of *M. edulis* (195,538, HPD95: 76,968-359,460) is slightly, but not significantly, lower than *M. galloprovincialis* (318,462, HPD95: 78,032-1,225,095) but our analysis suggests that the ancestral population was substantially larger than the two daughter populations (964,827, HPD95: 527,464-1,429,032). Interestingly, the less supported "homogeneous" scenario is less informative about the demographic history with the exception of the two current population sizes which are the only two parameters well differentiated from their prior distributions (Fig. 5, Table 3). Parameter estimates of the best-supported scenario suggests an ancestral subdivision about 2.5 MY ago followed by secondary contact beginning around 0.7 MY ago (Table 3). Under this scenario, both species would thus

have remained isolated for approximately three-quarters of their history. We then investigated the predicted distributions of genomic introgression rates from *M. galloprovincialis* into *M. edulis* (M1) and from *M. edulis* into *M. galloprovincialis* (M2) by $2x10^6$ random samples from re-scaled Beta distributions using the estimated joint shape parameters (Fig. 6). According to our estimates, introgression into *M. edulis* occurred at a lower rate (average *M1*=0.85) than introgression into *M. galloprovincialis* (average *M2*=1.22). We then compared observed and simulated summary statistics under a goodness-of-fit procedure using the joint-posterior distributions and found that the SC scenario with heterogeneous migration fit the data well except for variation among loci in F_{ST}, which was slightly underestimated by the scenario. (Table S4).

Finally, we obtained locus-specific estimates of both migration rates (Table 4). We note that posterior distributions for these two parameters were informative only for a few loci (Fig. 7). Therefore, locus-specific inferences of introgression rates are qualitatively informative but do not allow precise quantification of the mean number of migrants per generation. Nevertheless, heterogeneity in migration rates across loci were clearly apparent, ranging from below one for *EF1α* to values substantially greater than one for *mytilin B*, *mc125* and *glucanase* (Fig. 7; Table 4). Consistent with the multilocus inference, most locus-specific introgression rates tend to be close to the lower bound of the prior distribution, with the introgression spectrum into *M. galloprovincialis* deviating slightly toward higher values than into *M. edulis*.

Discussion

Detecting heterogeneity in migration rates among loci using a hierarchical ABC approach

Models of divergence with gene flow have increased in sophistication since their initial development by Wakeley & Hey (1997) and have provided important insights into the process of speciation (Pinho & Hey 2010; Feder & Nosil 2010). Here, we document how the recently-developed hierarchical ABC approach of Roux *et al.* (2013) that incorporates heterogeneous gene flow had considerable power in detecting a semi-permeable barrier to introgression between two mussel species (*M. edulis* and *M. galloprovincialis*) across a well characterized hybrid zone, even though our dataset was limited to 8 loci. Models incorporating variable rates of migration across loci outperformed models assuming equal levels of gene flow among loci thus confirming the highly variable patterns of introgression documented in previous studies on the mussel hybrid zone (Skibinski *et al.* 1983; Bierne *et al.* 2003; Boon *et al.* 2009). The superiority of the heterogeneous models to account for the patterns of polymorphism and divergence observed between *Mytilus* species was apparent by comparing posterior probabilities of the alternative models using a model choice procedure and statistical support was provided by a model checking procedure involving pseudo-observed datasets obtained by simulations (Fagundes *et al.* 2007; Cornuet *et al.* 2008). These simulations highlighted the very small rate of false positives and false negatives in comparisons between homogeneous and heterogeneous models. They also demonstrated that the procedure can be applied efficiently for biological models corresponding to the isolation with migration scenario (IM, Hey & Nielsen 2004), the ancient migration scenario (AM, analogous to a sympatric speciation model with no secondary contact), and the secondary contact scenario (SC).

Two recent studies have attempted to test for heterogeneity in migration rates across loci using related approaches. Sousa *et al.* (2013) proposed a modified version of the IM method

(Hey & Nielsen 2004, 2007) that allows the clustering of loci into distinct groups defined by their effective migration rates. This approach allows for the groups of loci to experience different levels of genetic drift, which is an important advancement because different genomic regions do not share the same effective population size (Charlesworth 2009). Sousa *et al.* (2013) conducted a simulation study on pseudo-observed datasets of 10 loci and showed how their method is conservative (only a small proportion of datasets sharing one migration rate was supported by any GWH model). To illustrate their method, GWH was tested between two subspecies of the European rabbit (*Oryctolagus cuniculus spp.*) for which a bimodal distribution of F_{ST}-values had been previously described (Geraldes *et al.* 2008) and a strong association was found between the levels of differentiation and the assignment to two groups of loci with shared migration parameters. Although the method of Sousa *et al.* (2013) proved to be efficient in the context of the IM scenario, it remains unclear if it can be applied when the history of gene flow deviates from the assumption that migration is continuous through time (Becquet & Przeworski 2009; Strasburg & Rieseberg 2010). The risk in only considering the IM scenario is that it may lead to biases in parameter estimates if the species under study have experienced different demographic histories (Becquet & Przeworski 2009). It may also miss information provided by the model itself. For example, the mode of speciation between the two *Oryctolagus* subspecies was not explicitly tested and it remains unclear if any gene flow occurred after initial divergence between lineages or whether secondary introgression occurred after a period of strict isolation. In contrast to the *Oryctolagus* example, the ongoing hybridization between two highly divergent *Ciona intestinalis* species (≈14.4% of synonymous divergence) investigated by Roux *et al.* (2013) was unlikely to have occurred continuously in time since their original split. Using an ABC-based model choice procedure, Roux *et al.* (2013) showed that the introgression involves a minority of loci (≈20%) between species that have recently experienced secondary contact following complete isolation for more than three million years. Combined with the results presented here

for mussels, the ABC approach appears to be an effective approach for testing alternative speciation scenarios and further highlights the need for incorporating heterogeneous gene flow to improve model testing and parameter estimation.

Due to the hierarchical Bayesian design of the heterogeneous models, our procedure allowed us to estimate the shape of the genomic distribution of migration rates (determined by the values of the hyper-parameters *alpha* and *beta*). This distribution is expected to depend in a complex way on the demographic scenarios (*e.g.,* the time since the two species diverged, the time since the onset of migration, and the level of gene flow), the patterns of natural selection that determine the realized gene flow around isolation genes, and the genomic patterns of linkage disequilibrium that determine the effect of genetic linkage. For the *Mytilus* dataset, the migration rate distributions were mostly L-shaped suggesting a predominance of genomic regions loosely permeable to introgression, which is consistent with the estimates of a long divergence time between the two species. It is also consistent with results of a previous study on the genetic basis of post-zygotic isolation between the two mussel species that suggested the existence of a large number of recessive Bateson-Dobzhansky-Muller incompatibilities across the genome (Bierne *et al.* 2006). With broader genomic coverage, it might be possible to determine whether introgression acts over a large fraction of the genome or is restricted to small genomic regions similar to the genomic hotspots of introgression in the *Ciona* species (Roux *et al.* 2013).

It is informative to compare the results of the ABC approach between *Ciona* and *Mytilus spp.* For *C. intestinalis,* Roux *et al.* (2013) observed small median rates of introgression with the migration rate into *C. intestinalis* species A being marginally higher than that into *C. intestinalis* species B (M_A=0.079 *vs.* M_B=0.0501, respectively). For *Mytilus*, the initial divergence began ~1.2MY later than *C. intestinalis* and the period of time when the lineages experienced gene flow following secondary contact was ~45 times longer. Consistent with this shorter period of species differentiation, gene flow between *M. edulis* and *M. galloprovinvialis* occurred at higher

rates than between *Ciona* species (*Medu*=0.4546 and *Mgallo*=0.7932). The genomic distribution of introgression rates for both studies were both L-shaped but differed in magnitude suggesting that most of the genome quickly isolated during the initial phase of speciation followed by a slower accumulation of barriers in the remaining regions with time. The next step will now be confirm this relationship between the time of differentiation and the shape of the genomic distribution of introgression rates for a given geographical context by using high-throughput sequencing technologies.

Inferring the history of divergence and gene flow between M. edulis and M. galloprovincialis

By allowing heterogeneity of effective gene flow among loci our analyses confirmed that the best demographic scenario corresponded to the subdivision of an ancestral population in two isolated gene pools followed by secondary contact and subsequent gene exchange. This scenario confirms the predictions made from previous preliminary investigations (e.g., Boon *et al.* 2009). Our simulations suggest that the subdivision of the ancestral mussel population occurred ~2.5 MY ago and was followed by a ~1.8 MY long period during which both *Mytilus* lineages remained isolated. This long period of allopatry is favorable for the accumulation of loci contributing to genetic incompatibilities (Navarro & Barton 2003; Matute *et al.* 2010; Moyle & Nakazato 2010; Nachman & Payseur 2012) and it is likely that a majority of the multifarious barriers to gene flow became fixed during this time. Following secondary contact it is unclear whether gene flow has been continuous or intermittent due to distributional shifts caused by glacial oscillations. Although the latter seems likely, our dataset does not provide sufficient power to test for intermittent gene flow since secondary contact (data not shown).

Although secondary contact scenarios has been implicated for the *M. edulis* complex of species for some time (e.g., Hilbish *et al.* 2002), it is worth emphasizing that our ABC approach strongly supported the SC scenario only when we allowed heterogeneity in introgression rates.

Alternative models with homogeneous migration rates consistently led to ambiguous results. Neglecting genomic variation in introgression rates failed to distinguish between the IM and SC scenarios and parameter estimates for the SC-Homogeneous scenario exhibited large variances in the posterior distributions of biologically relevant parameters (the times of speciation and secondary contact). As previously shown in *Ciona* by Roux *et al.* (2013), neglecting GWH can also sometimes lead to the statistical support of an incorrect scenario. Therefore, it appears that GWH must be taken into account by future studies investigating divergence with gene flow, especially when estimating historical parameters and testing for alternative speciation scenarios. Since variable patterns of gene flow have been widely documented in numerous taxa including *Helianthus* sunflowers (Whitney *et al.* 2010), *Heliconius* butterflies (Pardo-Diaz *et al.* 2012), *Mus* mice (Song *et al.* 2011), and *Ficedula* flycatchers (Ellegren *et al.* 2012), it might prove useful to test these case studies with methods that explicitly account for GWH. Furthermore, the statistical evaluation of alternative models proposed in the hierarchical ABC framework should help strengthen the case for specific modes of speciation that may have been overlooked by evaluating the IM scenario with homogeneous migration.

Acknowledgements

Numerical results presented in this article were carried out using the ISEM computing cluster, and the authors highly appreciate the helpful support of its staff. We also highly appreciate and thank the technical staff of the CRI-Lille 1 center. We acknowledge Matthieu Faure for technical assistance and Nicolas Galtier for useful discussions. This study was supported by the Agence National de la Recherche (HySea project ANR-12-BSV7-0011) and the project Aquagenet (SUDOE, INTERREG IV B). This is article 2013-XXX of Institut des Sciences de l'Evolution de Montpellier.

References

Addison JA, Ort BS, Mesa KA, Pogson GH (2008) Range-wide genetic homogeneity in the California sea mussel (*Mytilus californianus*): a comparison of allozymes, nuclear DNA markers, and mitochondrial DNA sequences. *Molecular Ecology*, **17**, 4222–4232.

Barton NH (1979) The dynamics of hybrid zones. *Heredity*, **43**, 341–359.

Barton N, Bengtsson BO (1986) The barrier to genetic exchange between hybridising populations. *Heredity*, **57**, 357–376.

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.

Becquet C, Patterson N, Stone AC, Przeworski M, Reich D (2007) Genetic structure of chimpanzee populations. *PLoS Genet*, **3**, e66.

Becquet C, Przeworski M (2007) A new approach to estimate parameters of speciation models with application to apes. *Genome Res*, **17**, 1505–1519.

Becquet C, Przeworski M (2009) Learning about modes of speciation by computational approaches. *Evolution*, **63**, 2547–2562.

Bierne N, Bonhomme F, Boudry P, Szulkin M, David P (2006) Fitness landscapes support the dominance theory of post-zygotic isolation in the mussels Mytilus edulis and M. galloprovincialis. *Proceedings. Biological sciences / The Royal Society*, **273**, 1253–60.

Bierne N, Bonhomme F, David P (2003) Habitat preference and the marine-speciation paradox. *Proceedings. Biological sciences / The Royal Society*, **270**, 1399–406.

Bierne N, Borsa P, Daguin C *et al.* (2003) Introgression patterns in the mosaic hybrid zone between Mytilus edulis and M. galloprovincialis. *Mol Ecol*, **12**, 447–461.

Bierne N, David P, Boudry P, Bonhomme F (2002) Assortative fertilization and selection at larval stage in the mussels Mytilus edulis and M. galloprovincialis. *Evolution*, **56**, 292–298.

Bierne N, Tanguy A, Faure M *et al.* (2007) Mark-recapture cloning: a straightforward and cost-effective cloning method for population genetics of single-copy nuclear DNA sequences in diploids. *Molecular Ecology Notes*, **7**, 562–566.

Blum MGB, François O (2009) Non-linear regression models for Approximate Bayesian Computation. *Statistics and Computing*, **20**, 63–73.

Boon E, Faure MF, Bierne N (2009) The flow of antimicrobial peptide genes through a genetic barrier between Mytilus edulis and M. galloprovincialis. *J Mol Evol*, **68**, 461–474.

Charlesworth B (2009) Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nature reviews. Genetics*, **10**, 195–205.

Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet Res*, **70**, 155–174.

Cornuet J-M, Santos F, Beaumont MA *et al.* (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics*, **24**, 2713–2719.

Csillery K, Francois O, Blum MGB (2012) abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*.

Ellegren H, Smeds L, Burri R *et al.* (2012) The genomic landscape of species divergence in Ficedula flycatchers. *Nature*, **491**, 756–60.

Fagundes NJR, Ray N, Beaumont M *et al.* (2007) Statistical evaluation of alternative models of human evolution. *Proc Natl Acad Sci U S A*, **104**, 17614–17619.

Faure B, Bierne N, Tanguy A, Bonhomme F, Jollivet D (2007) Evidence for a slightly deleterious effect of intron polymorphisms at the EF1alpha gene in the deep-sea hydrothermal vent bivalve Bathymodiolus. *Gene*, **406**, 99–107.

Faure MF, David P, Bonhomme F, Bierne N (2008) Genetic hitchhiking in a subdivided population of *Mytilus edulis*. *BMC evolutionary biology*, **8**, 164.

Feder JL, Nosil P (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*, **64**: 1729-1747.

Gardner JPA, Skibinski DOF (1988) Historical and size-dependent genetic variation in hybrid mussel populations. *Heredity*, **61**, 93–105.

Geraldes A, Carneiro M, Delibes-Mateos M *et al.* (2008) Reduced introgression of the Y chromosome between subspecies of the European rabbit (Oryctolagus cuniculus) in the Iberian Peninsula. *Molecular Ecology*, **17**, 4489–99.

Gosset CC, Bierne N (2013) Differential introgression from a sister species explains high F(ST) outlier loci within a mussel species. *Journal of evolutionary biology*, **26**, 14–26.

Hamilton G, Currat M, Ray N *et al.* (2005) Bayesian estimation of recent migration rates after a spatial expansion. *Genetics*, **170**, 409–417.

Harrison RG (1993) *Hybrid zones and the evolutionary process*. Oxford University Press.

Hartigan JA, Hartigan PM (1985) The Dip test of unimodality. *The Annals of Statistics*, **13**, 70–84.

Hey J (2006) Recent advances in assessing gene flow between diverging populations and species. *Current opinion in genetics & development*, **16**, 592–6.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–760.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc Natl Acad Sci U S A*, **104**, 2785–2790.

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

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, **18**, 337–338.

Matute DR, Butler IA, Turissini DA, Coyne JA (2010) A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science*, **329**, 1518–1521.

McVean G, Awadalla P, Fearnhead P (2002) A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics*, **160**, 1231–1241.

Moyle LC, Nakazato T (2010) Hybrid incompatibility "snowballs" between Solanum species. *Science*, **329**, 1521–1523.

Nachman MW, Payseur BA (2012) Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, **367**, 409–21.

Navarro A, Barton NH (2003) Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution*, **57**, 447–459.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.

Nosil P, Feder JL (2012) Widespread yet heterogeneous genomic divergence. *Molecular Ecology*, **21**, 2829–32.

Ort BS, Pogson GH (2007) Molecular population genetics of the male and female mitochondrial DNA molecules of the California sea mussel, *Mytilus californianus*. *Genetics*, **177**, 1087–99.

Pardo-Diaz C, Salazar C, Baxter SW *et al.* (2012) Adaptive introgression across species boundaries in Heliconius butterflies. *PLoS Genet*, **8**, e1002752.

Pialek J, Barton NH (1997) The Spread of an Advantageous Allele Across a Barrier: The Effects of Random Drift and Selection Against Heterozygotes. *Genetics*, **145**, 493–504.

Pinho C, Hey J (2010) Divergence with gene flow: Models and data. *Annu Rev. Ecol. Evol. Syst.,* **41**: 215-230.

Quesada H, Warren M, Skibinski DO (1998) Nonneutral evolution and differential mutation rate of gender-associated mitochondrial DNA lineages in the marine mussel Mytilus. *Genetics*, **149**, 1511–1526.

Ross-Ibarra J, Wright SI, Foxe JP *et al.* (2008) Patterns of polymorphism and demographic history in natural populations of Arabidopsis lyrata. *PLoS One*, **3**, e2411.

Roux C, Castric V, Pauwels M *et al.* (2011) Does speciation between *Arabidopsis halleri* and *Arabidopsis lyrata* coincide with major changes in a molecular target of adaptation? *PLoS One*, **6**, e26872.

Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species barrier: Genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. *Molecular biology and evolution*, mst066–.

Secor DH, Rooker JR, Zlokovitz E, Zdanowicz VS Identification of riverine, estuarine, and coastal contingents of Hudson River striped bass based upon otolith elemental fingerprints. *Marine ecology. Progress series*, **211**, 245–253.

Skibinski DOF, Beardmore JA, Cross TF (1983) Aspects of the population genetics of Mytilus (Mytilidae; Mollusca) in the British Isles. *Biological Journal of the Linnean Society*, **19**, 137–183.

Song Y, Endepols S, Klemann N *et al.* (2011) Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Curr Biol*, **21**, 1296–1301.

Sousa VMC, Carneiro M, Ferrand N, Hey J (2013) Identifying loci under selection against gene flow in Isolation-with-Migration models. *Genetics*, **194**: 211-233.

Strasburg JL, Rieseberg LH (2010) How robust are "isolation with migration" analyses to violations of the IM model? A simulation study. *Molecular biology and evolution*, **27**, 297–310.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.

Tajima F (1989) The effect of change in population size on DNA polymorphism. *Genetics*, **123**, 597–601.

Tavaré S, Balding DJ, Griffiths RC, Donnelly P (1997) Inferring coalescence times from DNA sequence data. *Genetics*, **145**, 505–518.

Templeton A (2006) *Population Genetics and Microevolutionary Theory*. Wiley-Liss.

Wakeley J, Hey J (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**, 256–276.

Whitney KD, Baack EJ, Hamrick JL *et al.* (2010) A role for nonadaptive processes in plant genome size evolution? *Evolution*, **64**, 2097–2109.

Wu C-I (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

Data accessibility

DNA sequences: GenBank accessions AAXXXXXX-BBYYYYYY

DNA sequence alignments: DRYAD doi:XX.XXXX/dryad.xaxxx

Author Contributions

C.R and N.B designed the study with the contribution of V.C and X.V, analysed the data and wrote a first draft of the manuscript. N.B, C.F and G.H.P generated the genetic data. C.R performed ABC modelling and analysis with the input of V.C, X.V and N.B. V.C, X.V, C.F and G.H.P contributed to the writing of the manuscript.

Figure legends

Figure 1: Alternative scenarios of speciation for *M. edulis* and *M. galloprovincialis*

Four classes of models with different temporal pattern of migration are compared: strict isolation (SI), constant migration (IM), isolation with 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_A , N_{gal} and N_{edu} are the number of effective individuals in the ancestral population, *M. galloprovincialis* and *M. edulis*, 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 migration rates M_1 and M_2 are expressed in $4.N.m$ units, where m is the proportion of a population made up of migrants from the other population per generation.

Figure 2: Distribution of pairwise inter-specific molecular divergence among loci

Figure 3: Empirical distributions of estimated relative posterior probabilities in 'homo' versus 'hetero' model comparisons.

611 Each distribution was obtained from ABC analysis of 1,000 simulated pseudo-observed datasets.
612 The area under each curve above 0.5 represents the fraction of times that the true model is
613 correctly recovered by our estimation procedure.
614
615

616 **Figure 4: Parameter estimates of the best support model of speciation SC.**
617 Prior and posterior distributions are represented by open and shaded symbols, respectively.
618 Effective population sizes and times must be multiplied by 100,000 and 400,000 respectively to
619 be converted in demographic units (Table 3).
620
621

622 **Figure 5: Posterior distributions of parameters for the two homo and hetero alternative SC**
623 **models.**
624 Homo and hetero posterior distributions are represented by open and shaded symbols,
625 respectively.
626

627 **Figure 6: Estimated genomic distributions of introgression rates into *M. edulis* and *M.***
628 ***galloprovincialis*.**
629 Distributions are obtained after randomly sampling 1,000 values from each 2,000 Beta
630 distributions retained by ABC analysis.
631

632 **Figure 7: Locus-specific estimates of introgression rates for the sequenced loci.**
633 The eight colored lines represent posterior distributions for the sequenced loci. The dotted line
634 represents the prior distribution.
635

Figure S1: Empirical relationship between the relative posterior probability of hetero or homo alternative model for the three models with migration and the associated probability to support the correct model.

1,000 pseudo-observed datasets were analyzed for each of the six pairwise homo/hetero comparisons. The probability to correctly support model-A was computed as the ratio P(model-A | model-A)/[P(model-A | model-A)+P(model-A | model-B)], where P(model-A | model-A) on the x-axis is the relative posterior probability in favor of model-A when analyzing a pseudo-observed dataset simulated under model-A, and P(model-A | model-B) is the estimated relative posterior probability in favor of the model-A when analyzing a pseudo-observed dataset simulated under model-B. The red line indicates a probability of supporting the correct model of 0.95.

Figure S2: Empirical distributions of the estimated relative probabilities of the SC model when the SI (red line), the IM (blue line), the AM (green line) and the SC (black line) models are the true models. The density estimates of the four models at the SC posterior probability = 0.5194 (vertical line) were used to compute the probability that SC is the correct model given our observation that $P_{SC} = 0.5194$. This probability is equal to 0.957.

Tables

657

Loci	n *M. edulis*[a]	n *M. galloprovincialis*[b]	L[c]	π_{edu}[d]	π_{gal}[e]	θ_{edu}[f]	θ_{gal}[g]	D_{edu}[h]	D_{gal}[i]	Sf[j]	Sx_{edu}[k]	Sx_{gal}[l]	Ss[m]	F_{ST}[n]	netdivAB[o]
EF1	20	20	639	0.0195	0.0146	0.0278	0.0278	-1.9317	0.0575	0	49	49	14	0.5491	0.0405
EF2	20	20	1,133	0.0026	0.0090	0.0070	0.0109	-0.6937	0.0117	0	28	44	0	0.3426	0.0059
Glucanase	18	14	468	0.0211	0.0126	0.0230	0.0188	-1.4043	0.0184	0	20	11	17	0.0712	0.0015
mac1	12	20	825	0.0043	0.0164	0.0040	0.0174	-0.2290	0.0195	0	6	47	4	0.3633	0.0091
Mannanase2	20	18	564	0.0194	0.0223	0.0255	0.0304	-1.1121	0.0222	0	30	38	21	0.0296	0.0014
mc125	19	18	108	0.0260	0.0737	0.0371	0.0705	0.1777	0.0586	0	6	19	8	0.0931	0.0088
mgd2	24	16	1,100	0.0387	0.0339	0.0411	0.0433	-0.9419	0.0392	0	78	67	91	0.0495	0.0029
MytilinB	16	16	535	0.0385	0.0223	0.0394	0.0344	-1.4939	0.0308	0	29	20	41	0.0064	0.0004
Average	-	-	-	0.0213	0.0256	0.0256	0.0317	-0.9536	0.0322	-	-	-	-	0.1881	0.0088
Standard deviation	-	-	-	0.0135	0.0209	0.0141	0.0188	0.6917	0.0180	-	-	-	-	0.2018	0.0132
Sum	-	-	5,372	-	-	-	-	-	-	0	246	295	196	-	-

Table1. Single locus statistics

[a]Total number of sequences in *M. edulis*

[b]Total number of sequences in *M. galloprovincialis*

[c]silent length excluding all gaps from the total alignment

[d]Average number of pairwise differences in *M. edulis*

[e]Average number of pairwise differences in *M. galloprovincialis*

[f]Watterson's θ measured in *M. edulis*

[g]Watterson's θ measured in *M. galloprovincialis*

[h]Tajima's D in *M. edulis*

[i]Tajima's D in *M. galloprovincialis*

[j]Number of fixed differences between *M. edulis* and *M. galloprovincialis*

[k]Number of exclusive polymorphic sites in *M. edulis*

[l]Number of exclusive polymorphic sites in *M. galloprovincialis*

[m]Number of shared polymorphic sites between *M. edulis* and *M. galloprovincialis*

[n]Level of species differentiation

[o]Net molecular divergence measured at synonymous positions.

658

659

660

8 loci		within scenarios	between scenarios
SI	x	x	0.0084
IM	homo	0.0045	x
	hetero	**0.9955**	0.4016
AM	homo	0.3206	x
	hetero	**0.6794**	0.0705
SC	homo	0.0262	x
	hetero	**0.9738**	**0.5194**

Table 2. Relative posterior probabilities of investigated model

The 'homo' and 'hetero' alternative models were first compared within the three CM, IM and SC scenarios.

Each of the best alternative were then compared together with the SI scenario

661

Parameters	Alternative SC Scenarios	Median	Mode	95% HPD
Current *M. edulis* population *size*	**hetero**	**195,538**	**200,755**	**76,968-359,460**
	homo	87,651	58,691	26,217-320,743
Current *M. galloprovincialis* population size	**hetero**	**318,462**	**221,996**	**78,032-1,225,095**
	homo	208,560	159,566	86,690-547,228
Size of the ancestral population	**hetero**	**964,827**	**1,059,216**	**527,464-1,429,032**
	homo	903,004	414,775	71,016-1,911,930
T_{split}	**hetero**	**2,524,781**	**2,083,954**	**1,041,507-6,413,986**
	homo	3,469,549	2,594,116	773,465-8,625,005
T_{SC}	**hetero**	**676,108**	**589,892**	**390,612-1,153,517**
	homo	1,964,932	1,476,798	572,885-5,538,929

Table3. Demographic and historical parameters estimated under the favored SC model with variable migration rates among loci

The estimates where calibrated by assuming a generation time of two years and a mutation rate of 2.763×10^{-8} /pb/generation

662

663

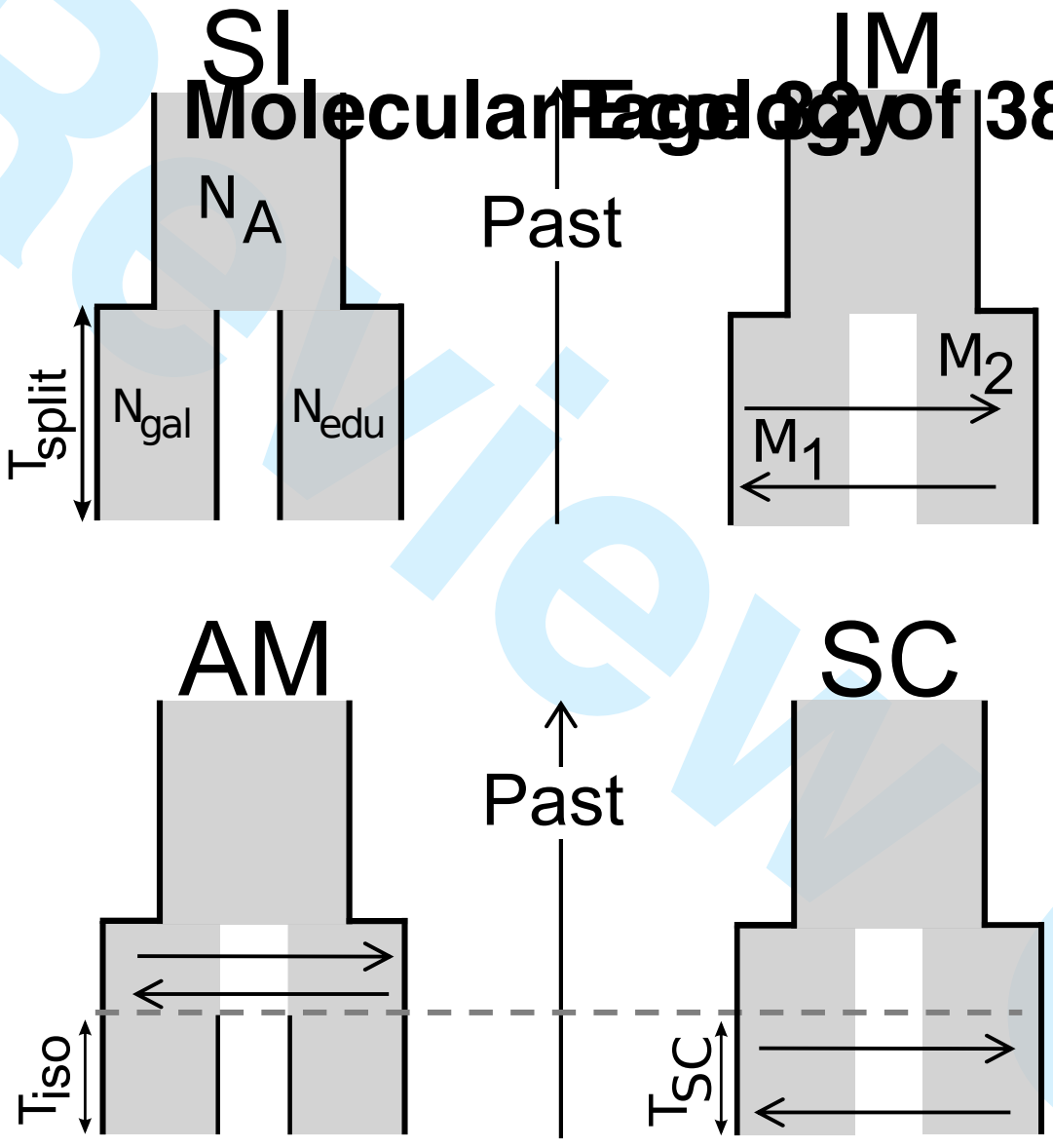
664

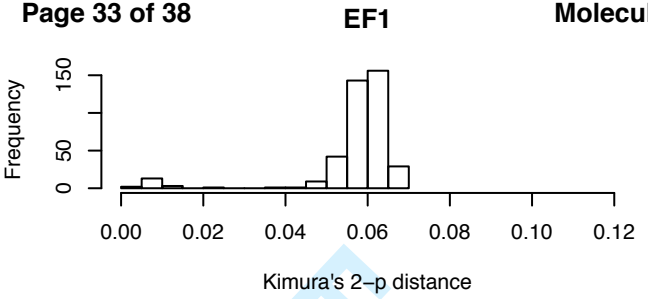
	$M_{from\ gal\ to\ edu}$		$M_{from\ edu\ to\ gal}$	
	Median	95% HPD	Median	95% HPD
EF1	0.5423	0.0975-2.0843	0.2012	0.0445-0.4997
EF2	0.3049	0.0791-1.2071	2.3026	0.6346-5.5531
glucanase	8.8162	0.6537-19.3147	5.9771	0.3159-15.7622
mac1	1.2847	0.0653-11.5563	2.9716	0.7434-12.0017
mannanase2	5.3827	0.1411-19.0486	4.7532	0.4208-16.248
mc125	8.6226	3.5079-16.1347	5.8908	3.6006-10.3749
mgd2	4.1736	0.8767-12.6816	4.1708	1.1034-8.5489
mytilinB	11.9486	4.1625-18.6808	9.6341	1.8621-17.6082

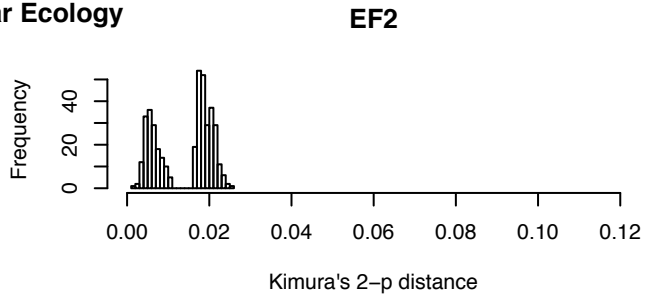
Table4. Locus specific estimates of migration rates

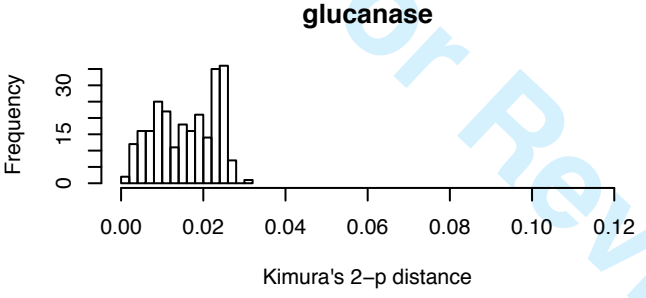
665

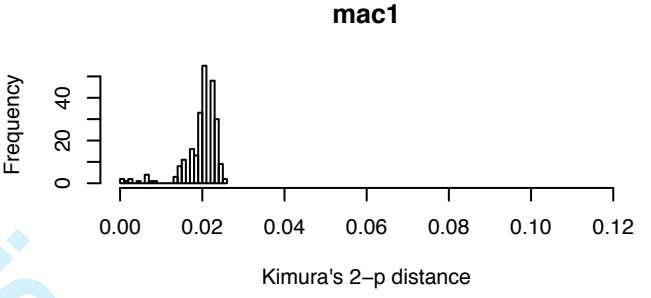
666 **Figures**

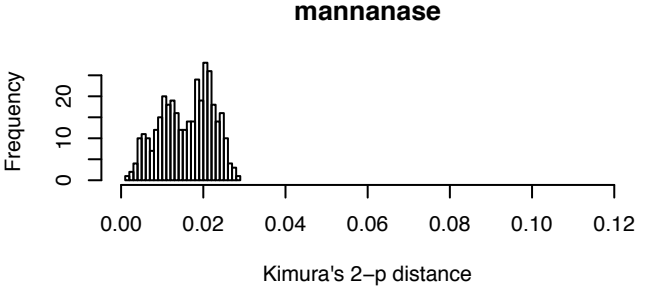

SI
IM
AM
SC
Past
Past
N_A
N_{gal}
N_{edu}
T_{split}
M_2
M_1
T_{iso}
T_{SC}

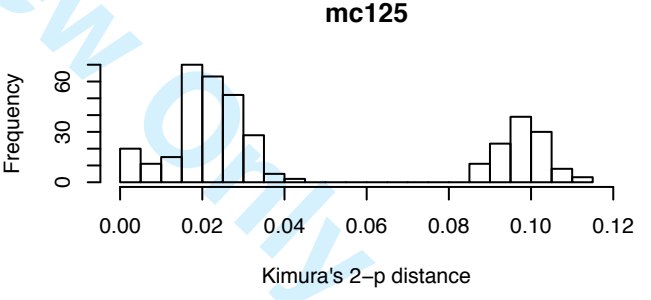


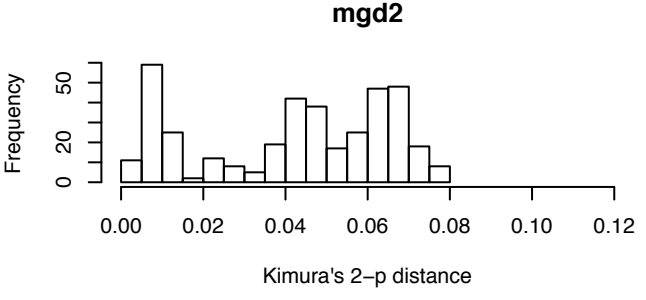


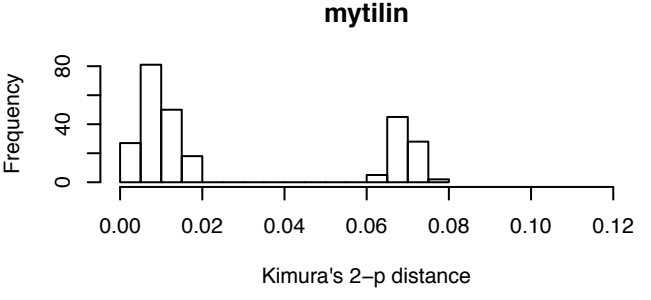


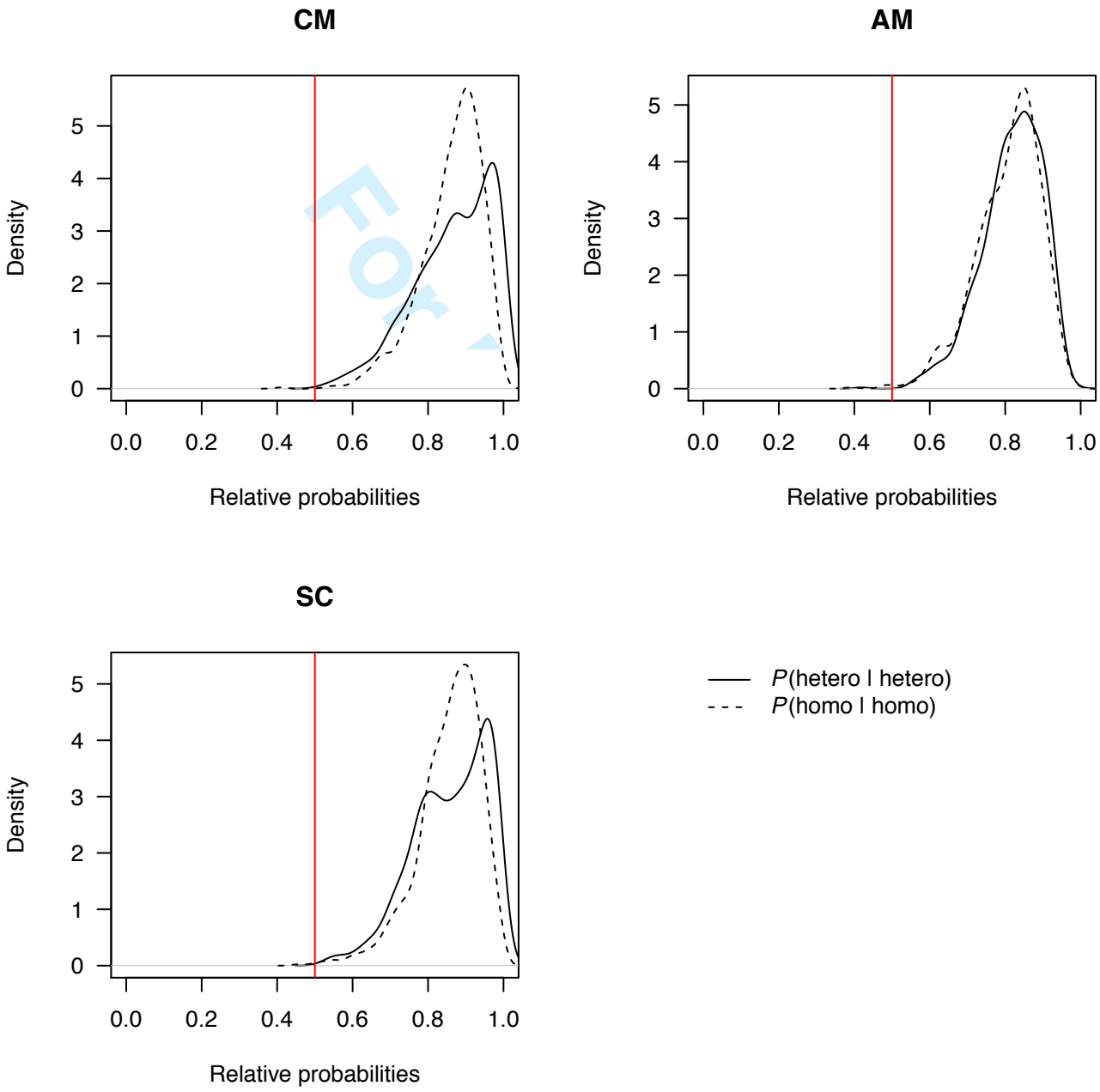


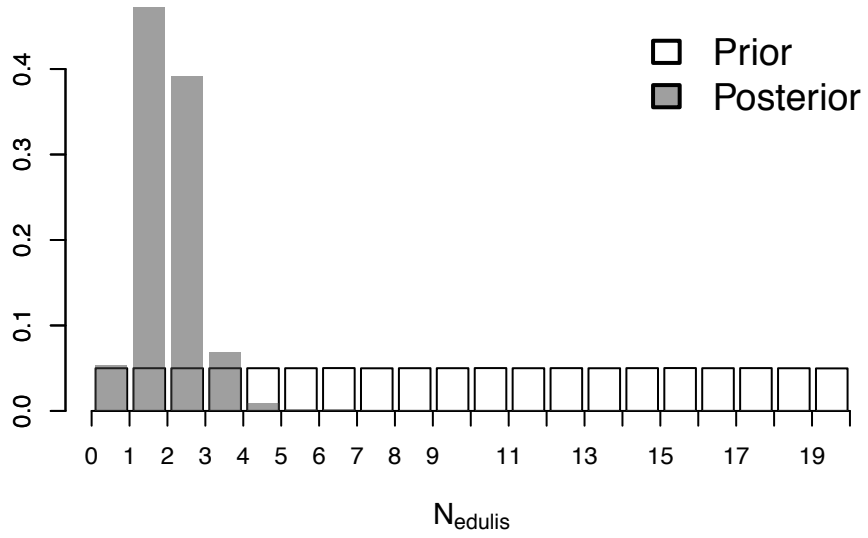


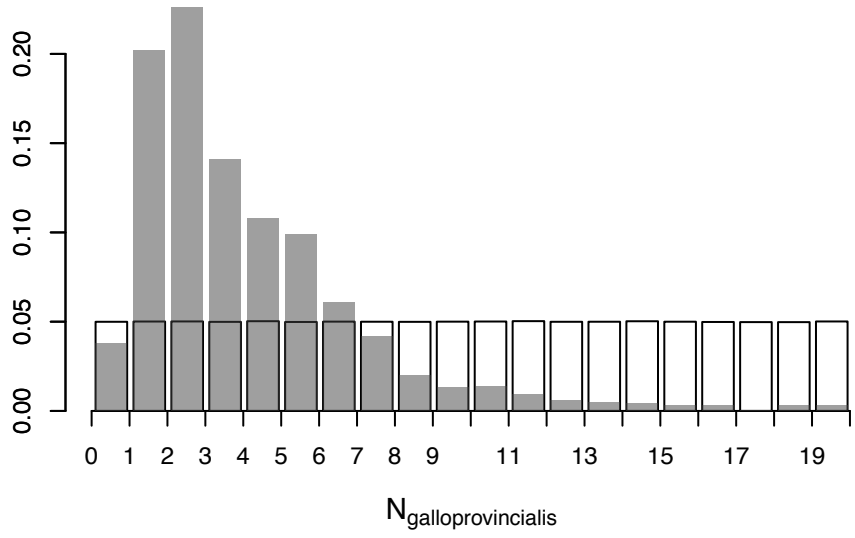


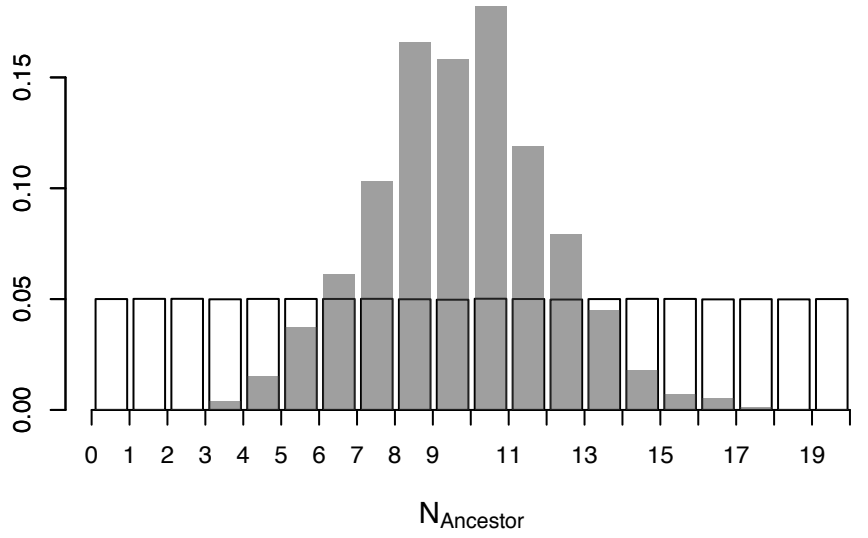


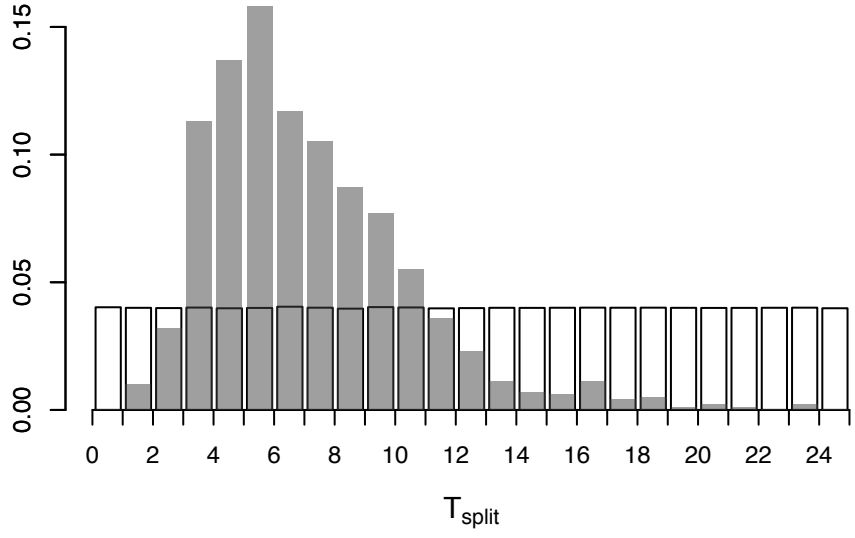


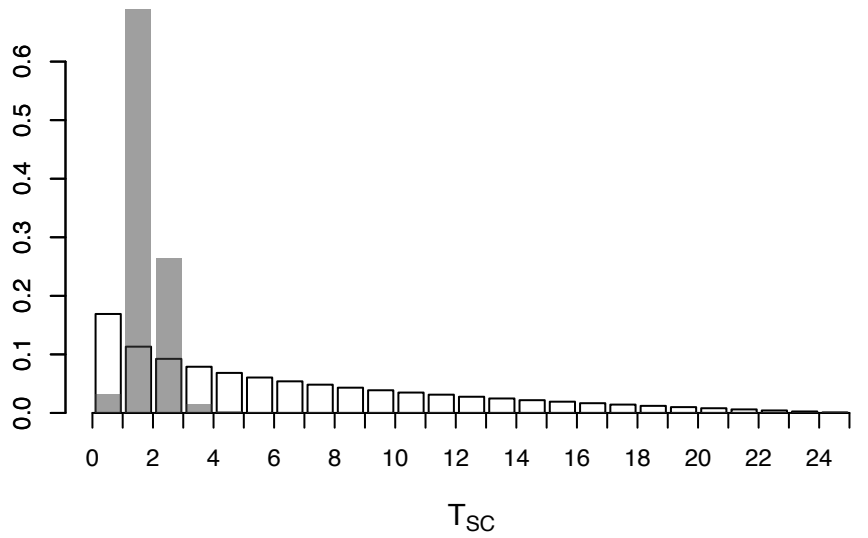

CM
AM
SC
Density
Relative probabilities
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
P(hetero | hetero)
P(homo | homo)

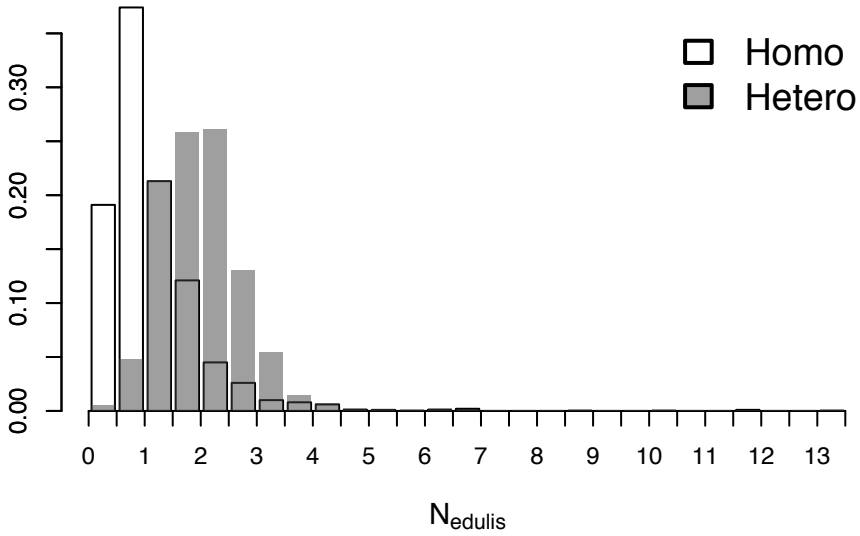

Prior
Posterior
0.0 0.1 0.2 0.3 0.4
0 1 2 3 4 5 6 7 8 9 11 13 15 17 19
N_{edulis}

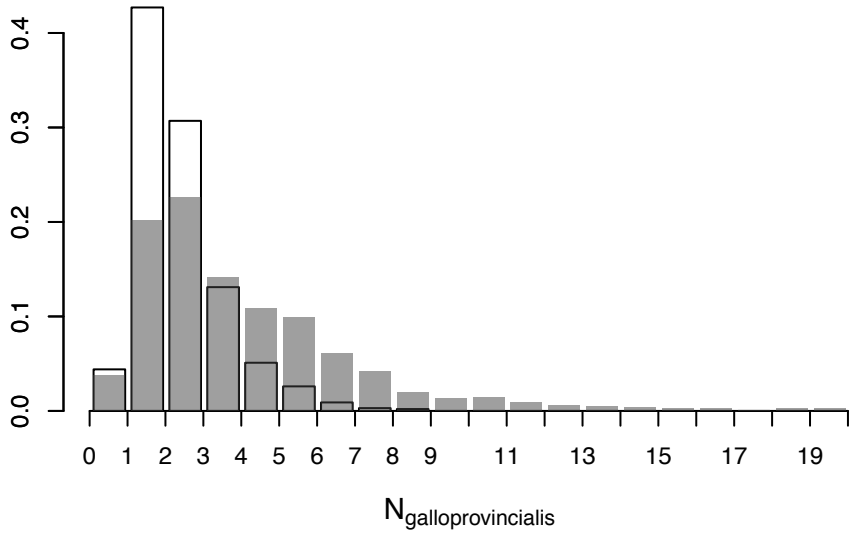

0.00 0.05 0.10 0.15 0.20
0 1 2 3 4 5 6 7 8 9 11 13 15 17 19
$N_{galloprovincialis}$

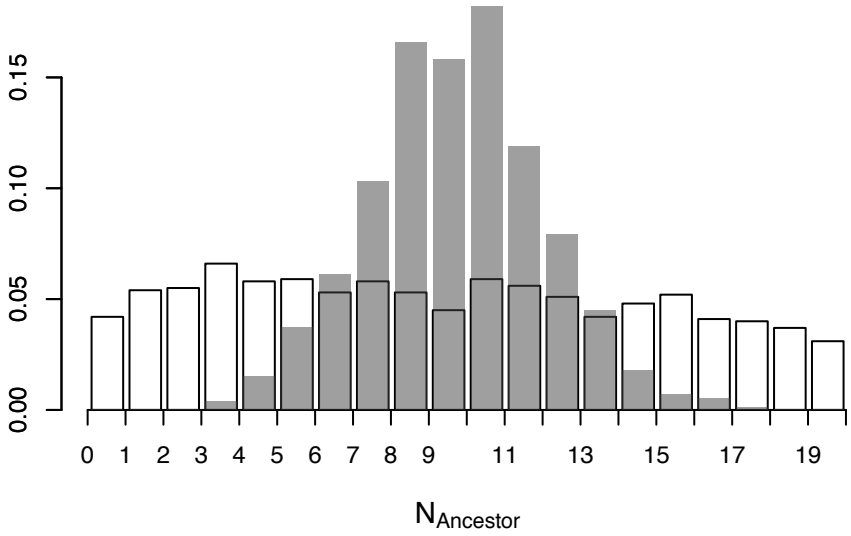

0.00 0.05 0.10 0.15
0 1 2 3 4 5 6 7 8 9 11 13 15 17 19
$N_{Ancestor}$

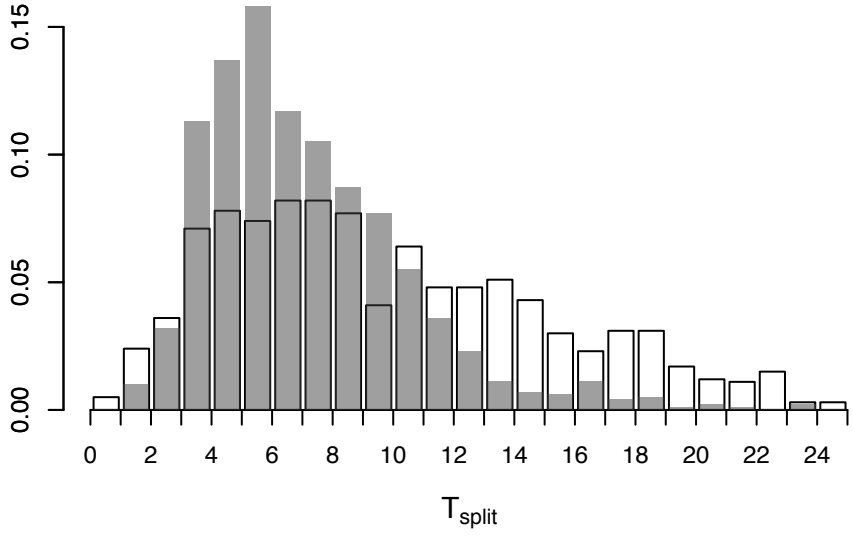

0.00 0.05 0.10 0.15
0 2 4 6 8 10 12 14 16 18 20 22 24
T_{split}

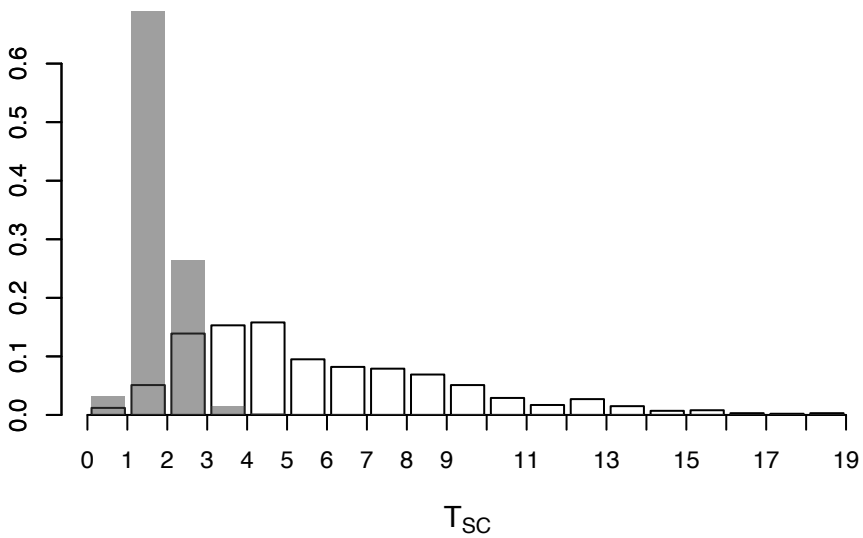

0.0 0.1 0.2 0.3 0.4 0.5 0.6
0 2 4 6 8 10 12 14 16 18 20 22 24
T_{SC}


Homo
Hetero
0.00
0.10
0.20
0.30
0 1 2 3 4 5 6 7 8 9 10 11 12 13
N_{edulis}


0.0
0.1
0.2
0.3
0.4
0 1 2 3 4 5 6 7 8 9 11 13 15 17 19
$N_{galloprovincialis}$


0.00
0.05
0.10
0.15
0 1 2 3 4 5 6 7 8 9 11 13 15 17 19
$N_{Ancestor}$


0.00
0.05
0.10
0.15
0 2 4 6 8 10 12 14 16 18 20 22 24
T_{split}


0.0
0.1
0.2
0.3
0.4
0.5
0.6
0 1 2 3 4 5 6 7 8 9 11 13 15 17 19
T_{SC}