2007

Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates

Pierre Faubet
Laboratoire d'Ecologie Alpine, pierre.faubet@e.ujf-grenoble.fr

Robin Waples
NOAA, robin.waples@noaa.gov

Oscar Gaggiotti
Université Joseph Fourier, oscar.gaggiotti@ujf-grenoble.fr

Molecular Ecology (2007) **16**, 1149–1166 doi: 10.1111/j.1365-294X.2006.03218.x

Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates

PIERRE FAUBET,* ROBIN S. WAPLES† and OSCAR E. GAGGIOTTI*
**Laboratoire d'Ecologie Alpine (LECA), UMR CNRS 5553, B.P. 53, 38041 Grenoble cedex 09, France, †Northwest Fisheries Science Center, 2725 Montlake Blvd East, Seattle, WA 98112, USA*

Abstract

Bayesian methods have become extremely popular in molecular ecology studies because they allow us to estimate demographic parameters of complex demographic scenarios using genetic data. Articles presenting new methods generally include sensitivity studies that evaluate their performance, but they tend to be limited and need to be followed by a more thorough evaluation. Here we evaluate the performance of a recent method, BAYESASS, which allows the estimation of recent migration rates among populations, as well as the inbreeding coefficient of each local population. We expand the simulation study of the original publication by considering multi-allelic markers and scenarios with varying number of populations. We also investigate the effect of varying migration rates and F_{ST} more thoroughly in order to identify the region of parameter space where the method is and is not able to provide accurate estimates of migration rate. Results indicate that if the demographic history of the species being studied fits the assumptions of the inference model, and if genetic differentiation is not too low ($F_{ST} \geq 0.05$), then the method can give fairly accurate estimates of migration rates even when they are fairly high (about 0.1). However, when the assumptions of the inference model are violated, accurate estimates are obtained only if migration rates are very low ($m = 0.01$) and genetic differentiation is high ($F_{ST} \geq 0.10$). Our results also show that using posterior assignment probabilities as an indication of how much confidence we can place on the assignments is problematical since the posterior probability of assignment can be very high even when the individual assignments are very inaccurate.

Keywords: Bayesian methods, gene flow, MCMC, migration, multilocus genotypes, parameter estimation

Received 26 July 2006; revision received 17 October 2006; accepted 30 October 2006

Introduction

The study of dispersal processes (colonization and migration) is central to the fields of population genetics, molecular genetics and conservation and management of wildlife. Direct estimates of migration parameters can be obtained using purely ecological approaches such as mark-release–recapture methods (MRR), but they have the inconvenience of being time consuming and impractical for the study of large and/or spatially extended metapopulations. Indirect methods based on population genetics models are an attractive alternative because they are easy to implement in these situations and only require a carefully planned sampling programme aimed at collecting tissue samples for DNA extraction and analysis. For many decades these estimates were obtained from *F*-statistics, but more recently this practice has come under criticism due to the simplistic assumptions (constancy in demographic parameters and genetic equilibrium conditions) made by this approach (e.g. Whitlock & Mccauley 1999). Recent progress in population genetics theory and statistics has led to the development of sophisticated methods that avoid many (and sometimes most) of these unrealistic assumptions, and there is a growing interest in applying them to address practical questions in conservation and evolution.

Methods aimed at estimating migration parameters can be grouped into two types of approaches: (i) coalescent or genealogical approaches that use the genealogical information

Correspondence: Oscar E. Gaggiotti, Fax: 33 476 514 279; E-mail: oscar.gaggiotti@ujf-grenoble.fr

contained in DNA sequences, and (ii) multilocus genotype approaches that use gametic disequilibrium information. It is important to realize that these two types of methods differ not only in the type of information they use but also in the nature of the parameters they estimate. Coalescent methods (and those based on summary statistics) estimate long-term evolutionary parameters, while multilocus genotype methods estimate short-term ecological parameters.

It is a standard practice to publish the statistical genetic method with a limited validation study that is usually followed by a much more detailed one. This has indeed been the case for MIGRATE (first published by Beerli & Felsenstein 2001 and later evaluated by Abdo *et al*. 2004), the most well-known coalescent method for estimating migration rates. Here we evaluate the performance of a more recent method, BAYESASS (Wilson & Rannala 2003), which is the multilocus genotype counterpart of MIGRATE. It is based on a Bayesian approach and can estimate rates of recent immigration among populations. It also estimates the posterior probability distribution of individual immigrant ancestries, population allele frequencies and population inbreeding coefficients.

One of the most enticing features of Wilson & Rannala's (2003) method is that it has the potential for estimating contemporary migration rates among populations. It can thus be extremely useful for guiding conservation plans requiring the identification of demographically independent subpopulations. There is a paucity of studies that address the question of how small migration rates (m) should be to insure that subpopulations have independent dynamics (Waples & Gaggiotti 2006) but a study by Hastings (1993) suggests that two populations become demographically independent when m falls below about 0.10. The preliminary simulation study of Wilson & Rannala (2003) suggests that their method might be capable of accurately estimating migration rates of this order of magnitude, but a more thorough evaluation is required to confirm this possibility.

In their sensitivity study, Wilson & Rannala (2003) considered biallelic markers and a scenario with two populations and investigated the effect of varying migration rates (0.01, 0.05, 0.10, or 0.20) and F_{ST} (0.01, 0.10, or 0.25). They also studied the effect of varying sample sizes (20 or 100 individuals) and number of loci (5, 10 or 20). Here we expand this simulation study by considering multiallelic markers and scenarios with varying number of populations. We also investigate the effect of varying migration rates and F_{ST} more thoroughly in order to identify more precisely the region of parameter space where the method is and is not able to provide accurate estimates of migration rate. We studied the effect of deviations to the assumptions of BAYESASS by generating data using both the same approach as Wilson & Rannala (2003) and another method, EASYPOP, which simulates a different biological scenario.

Methods

BAYESASS implements a Bayesian approach using Markov chain Monte Carlo (MCMC) techniques. In the next two sections, we describe the probabilistic model implemented by BAYESASS and the simulation techniques we used to generate the synthetic data. We also provide details of the parameters used in the MCMC runs and the statistics used to evaluate the performance of the method.

BAYESASS

The inference model implemented by BAYESASS assumes linkage equilibrium but allows for deviations from Hardy–Weinberg equilibrium by estimating population-specific inbreeding coefficients. Migration rates among populations can be asymmetric but are constant over short periods of time (two generations). Additionally, it is assumed that migration rates are small (see Appendix A in Wilson & Rannala 2003). These two latter assumptions impose a constraint on the range of migration rates that can be considered by the method. More precisely, the total proportion of migrant individuals into a population per generation cannot exceed 1/3. Thus, nonmigrant proportions must be in the interval 2/3 to 1. The method also assumes that genetic drift and migration during the last few generations do not change subpopulation allele frequencies.

The Bayesian formulation implemented by Wilson & Rannala's (2003) method is,

$$f(\mathbf{m},\mathbf{M},\mathbf{t},\mathbf{F},\mathbf{p}/\mathbf{X};\mathbf{S}) \propto \Pr(\mathbf{X}/\mathbf{S};\mathbf{M},\mathbf{t},\mathbf{F},\mathbf{p})\Pr(\mathbf{M},\mathbf{t}/\mathbf{m})f_m(\mathbf{m})f_F(\mathbf{F})f_p(\mathbf{p}) \quad (1)$$

where the parameters to be estimated are $\mathbf{m} = (m_{ql})$, a matrix of migration rates between populations, $\mathbf{F} = \{F_l\}$ a vector of inbreeding coefficients, $\mathbf{M} = \{M_h\}$, a vector that contains the source of migrant ancestry of individuals in the sample, $\mathbf{t} = \{t_h\}$, a vector that gives the generation at which migrant ancestors of the sampled individuals arrived, and $\mathbf{p} = \{p_{lji}\}$, a matrix with the subpopulation allele frequencies. The estimation is based on the multilocus genotypes $\mathbf{X} = \{X_{hij}\}$ and population source $\mathbf{S} = \{S_h\}$ of individuals in the sample.

The prior densities $f_m(\mathbf{m})$, $f_F(\mathbf{F})$, $f_p(\mathbf{p})$ and $\Pr(\mathbf{M},\mathbf{t}/\mathbf{m})$, and likelihood function $\Pr(\mathbf{X}/\mathbf{S};\mathbf{M},\mathbf{t},\mathbf{F},\mathbf{p})$, are given in Wilson & Rannala (2003). The inference model is represented by the directed acyclic graph (DAG) in Fig. 1.

By default, BAYESASS provides means and variances of the parameters being estimated. In our case, we modified the program code in order to obtain the raw MCMC output and used *R* to estimate the probability density function, mean and mode of each parameter. The mode was estimated as the value that corresponds to the maximum of the

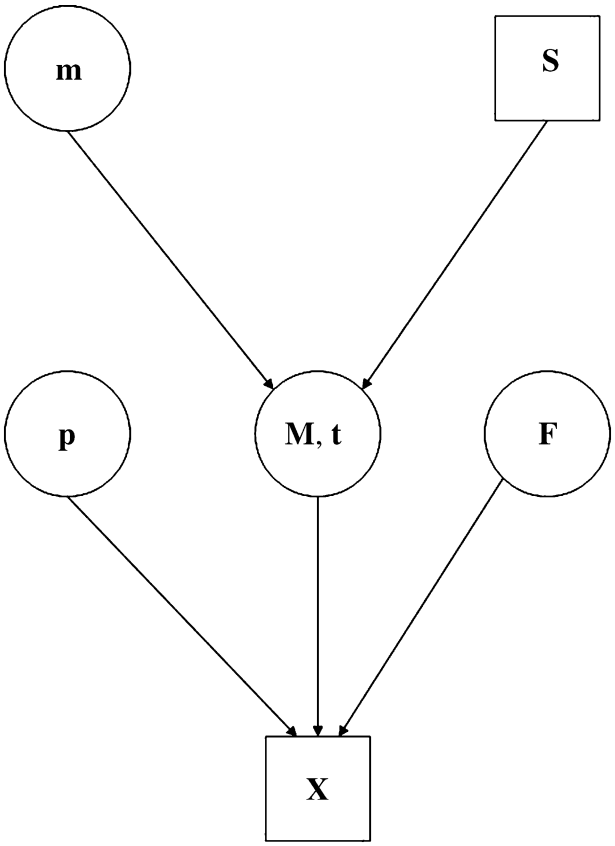

Fig. 1 The Directed Acyclic Graph (DAG) for the model given in equation 1. Square nodes denote known quantities (data) and circles represent parameters to be estimated. Lines between nodes represent direct stochastic relationships within the model. The variables within each node correspond to the different model parameters discussed in the text.

probability density function (pdf), i.e. the value with the highest posterior probability.

Generation of synthetic data for simulations

We use two different approaches to generate the synthetic data. In a first instance, we used the same approach as Wilson & Rannala (2003), in which the simulation model follows very closely the inference model. This allowed us to carry out a detailed sensitivity analysis of the method. In order to investigate how the method performs when the scenario considered deviates from the inference model, we also generated data using the software EASYPOP (Balloux 2001).

Simulations of the inference model. We simulated samples from subpopulations exchanging migrants according to the Wright island model at stationarity. We considered the general situation of a species with discrete generations inhabiting I islands of constant size and studied J marker loci with K_j alleles at any given locus j (i.e. the number of alleles can vary among loci). Each generation a fraction m of the individuals on each island is replaced by immigrants from a large mainland population with constant allele frequencies $\mathbf{q} = \{q_{ji}\}$, where q_{ji} is the frequency of allele i at locus j. Under these assumptions, the stationary distribution of allele frequencies in the islands, $\mathbf{p} = \{p_{lji}\}$, follows a Dirichlet distribution with parameters $4Nm\mathbf{q}$, i.e.

$$f(p_{lj1}, \ldots, p_{ljK_j}) = \Gamma(4Nm)\prod_{i=1}^{K_j} \frac{p_{lji}^{4Nmq_{ji}-1}}{\Gamma(4Nmq_{ji})} \quad (2)$$

where p_{lji} is the frequency of allele i at locus j in population l, N is the (constant and equal) size of the subpopulations, and m is the proportion of migrants exchanged among populations.

In equation 2, $4Nm$ can be replaced by its expected value at stationarity, $4Nm \approx 1/F_{ST} - 1$, to obtain a pdf for generating allele frequency distributions at each locus and each local population with a fixed F_{ST} value. This approach does not allow the simulation of local populations that differ in size. Thus, to simulate this latter scenario we used the sampling formula for F_{ST} as described in Balding & Nichols (1997). The global allele frequencies, $\mathbf{q}$, used to generate the simulated data were those of the grey seal metapopulation studied by Gaggiotti *et al*. (2004), and only data sets with F_{ST}s that were within 10% of the targeted value were kept. We generated samples of n_q individuals from each simulated local population using the multinomial distribution of equation 2 in Wilson & Rannala (2003), which gives the probability of observing $\mathbf{M}$ and $\mathbf{t}$ given $\mathbf{m}$. To reduce the number of parameters to be considered in the simulations, we used symmetric and equal migration rates, i.e.

$$\forall l \neq q,\ m_{lq} = m_{ql} = m^* \quad (3)$$

From equation 3 and the constraints on migration rates imposed by the method (see above) we have

$$0 \leq m^* \leq \frac{1}{3(I-1)} \quad (4)$$

The inference model assumes low migration rates and considers only possibilities involving at most a single migrant ancestor at some generation in the past. Thus, there are three types of individuals: nonmigrants, first generation migrants and second generation migrants (Wilson & Rannala 2003). The genotype of nonmigrants are generated by assigning alleles according to the Hardy–Weinberg proportions, conditional on the simulated allele frequency distributions of the population where the individual was sampled. Since the inference model assumes linkage equilibrium within each population, alleles are assigned independently at each locus. Genotypes of first generation migrants are generated according to Hardy–Weinberg

proportions conditional on the allele frequencies in their population of origin. Second generation migrant genotypes are assigned by drawing an allele from each population.

Simulations using EASYPOP. We considered a finite island model with I subpopulations, each of constant size N and equal sex ratio. Each generation, random mating was simulated to produce a diploid genotype for J independent gene loci for each individual, which then had a probability m of migrating to another subpopulation. All loci had the same mutation dynamics, which occurred according to the k-allele model (KAM; each mutation equally likely to lead to any of k possible allelic states). We considered 10 allelic states and a mutation rate $\mu = 5 \times 10^{-4}$, values that are representative of highly polymorphic markers like microsatellites. Simulations were initiated with maximal genetic diversity (genotypes in initial generation randomly drawn from all possible allelic states). We ran each replicate for 5000 generations before collecting data to attain an approximate mutation-migration-drift equilibrium. In the final generation of each replicate, samples of n_q individuals were taken from each subpopulation for genetic analysis.

Accuracy and bias. We are particularly interested in the ability of BAYESASS to accurately estimate migration rates, but we also investigated the accuracy of the estimated inbreeding coefficients and the individual assignments. We used the posterior means and modes of the posterior distributions of m_{ql} and F_l as estimators of these parameters and evaluated accuracy using the relative mean square error (RMSE) for estimates of m_{ql} and the mean square error (MSE) for estimates of F_l. This was carried out in order to be able to compare the accuracy when varying migration rate. In the case of inbreeding coefficient, we limited ourselves to scenarios that assumed Hardy–Weinberg equilibrium ($F_l = 0$) so we use MSE instead of RMSE. We also calculated the relative bias for the estimators of migration rates and bias for estimates of inbreeding coefficient. In order to calculate these statistics, we simulated $N = 10$ independent data sets for each scenario and used the following equations:

$$RBias(\tilde{\mathbf{m}}) = \frac{1}{N}\frac{1}{I(I-1)}\sum_k\sum_{q\neq l}\frac{\tilde{m}_{lq}^k - m^*}{m^*} \quad (5)$$

$$RMSE(\tilde{\mathbf{m}}) = \frac{1}{N}\frac{1}{I(I-1)}\sum_k\sum_{q\neq l}\left(\frac{\tilde{m}_{lq}^k - m^*}{m^*}\right)^2 \quad (6)$$

$$Bias(\tilde{\mathbf{F}}) = \frac{1}{N}\frac{1}{I}\sum_k\sum_l \tilde{F}_l^k \quad (7)$$

$$Bias(\tilde{\mathbf{F}}) = \frac{1}{N}\frac{1}{I}\sum_k\sum_l (\tilde{F}_l^k)^2 \quad (8)$$

where $\tilde{m}_{ql}^k$ is the estimated migration rate from population l into population q obtained for the replicate data set k, and $\tilde{F}_{ql}^k$ is the estimated inbreeding coefficients for population l obtained from data set k. Note that equations 5–8 give overall measures of bias and accuracy for the matrix of migration rates $\tilde{\mathbf{m}}$ and the vector of inbreeding coefficients $\mathbf{F}$, which are obtained by averaging across all the matrix/vector elements.

We obtained the 95% credible intervals (CI) for each element of the migration matrix and calculated its width. We also recorded the number of times that the true value fell within the CI. The results represent the average across all the elements of the migration matrix and replicates.

We evaluated the accuracy of migrant ancestry assignments using the proportion of individuals that were assigned to their correct (simulated) ancestral class and report the mean across all 10 replicates. We also use the maximum posterior probability with which these assignments were carried out. For each individual we recorded the population with the highest posterior assignment probability (irrespective of it being correct or false). These values were then averaged across all individuals in a data set and across all data sets. This was carried out only for data sets simulated under the inference model because in this case we knew how genotypes were drawn, which was not the case for data sets generated using EASYPOP.

Simulated scenarios. We chose a set of default values for the parameters of the simulation models and then studied the effect of varying only one of them at a time. For each simulation method and combination of parameters settings, we simulated 10 replicate data sets.

Table 1 presents the range of parameter values that we investigated with the simulations of the inference model. We looked at the influence of the level of genetic differentiation F_{ST}, number of individuals sampled per population n, number of loci J, number of alleles per locus K, number of populations I, and proportion of migrants $m = m^*$.

Table 2 presents the parameter sets considered using EASYPOP. In this case, we investigated the effect of varying population sizes N, migration rates m, and numbers of populations I. The characteristics of the samples were kept constant: sample size of 50 individuals per population, 20 loci each with 10 allelic classes.

MCMC runs. We analysed the simulated data sets using MCMC runs of 21×10^6 iterations, to insure convergence. We discarded the first 10^6 iterations as burn-in and used a thinning interval of 2000 iterations. Instead of using the default values, we used delta values of 0.10 for all parameters because they resulted in acceptance rates that varied between 20 and 60%. We identified MCMC runs with convergence problems using two different approaches, depending on the method used to generate the data. In the case of data sets generated under the inference model, we considered as suspect any MCMC run that resulted in a

Table 1 Parameters for data generated with the inference model. The first column gives the parameter that was allowed to vary, and the range of values considered, the six that follow give the values assigned to the parameter that were fixed. The last column indicates the figure that show the results obtained for each scenario. F_{ST} = genetic differentiation, m = migration rate, I = number of populations, n = number of individuals sampled per population, J = number of loci, K = number of alleles per loci

Parameter ( ) true values considered	Fixed parameter						Figure
	F_{ST}	m	I	n	J	K	
F_{ST} (0.01, 0.02, 0.05, 0.075, 0.1, 0.25)		0.05	3	100	10	11	2, 4
m (0, 0.01, 0.02, 0.05, 0.1, 0.15)	0.1		3	100	10	11	2, 4
I (2, 3, 5, 7)	0.1	0.05		100	10	11	2, 4
n (20, 40, 60, 80, 100)	0.1	0.05	3		10	11	3, 5
J (5, 10, 15, 20)	0.1	0.05	3	100		11	3, 5
K (2, 5, 8, 11)	0.1	0.05	3	100	10		3, 5

Table 2 Parameters for data generated with EASYPOP. We generated data for $J = 20$ loci with $K = 10$ possible allelic classes. A number of $n = 50$ individuals was sampled per population. I = number of populations, N = common population size, m = migration rate. The last column indicates figures where corresponding results are shown

Parameter set	Input parameters				Figure
	I	N	m	Nm	
m5	4	500	0.01	5	6(a), 7(a)
m1n2	4	200	0.01	2	6(a), 6(b)
m1n5	4	50	0.01	0.5	6(a)
m2n2	4	200	0.05	10	6(b)
m3n2	4	200	0.10	20	6(b)
m3n5	4	50	0.10	5	7(a)
n5	4	100	0.05	5	7(a)
2–25	2	500	0.05	25	7(b)
m25	4	500	0.05	25	7(b)
8–25	8	500	0.05	25	7(b)

very low proportion of individuals correctly assigned (less than 40%). In the case of data sets generated using EASYPOP, we focused on the quadratic error defined as $\sum_{q \neq l}((\tilde{m}^k_{lq} - m^*)/m^*)^2$ and considered as suspect any MCMC run that resulted in a quadratic error one order of magnitude larger than that of the best run (i.e. the one with the lowest error). We discarded MCMC runs with converge problems and repeated the analysis using different starting conditions until the proportion of individuals correctly assigned (for simulations under the inference model) or the migration rate quadratic error (for EASYPOP simulations) was the same order of magnitude as that of the best run. We also calculated the Bayesian deviance (see Appendix and Discussion) for all MCMC runs in order to establish if it could be used as a criterion to identify suspect runs when BAYESASS is applied to real data sets (see below). Low deviance values indicate a good fit of the data to the model (see Spiegelhalter *et al*. 2002) and therefore it may be possible to identify runs with convergence problems as those that lead to a high deviance.

Results

Here we present separately the results for the two types of data sets generated. We start by discussing convergence problems and then discuss the quality of the estimates using bias and RMSE. For each simulation method and parameter set (Tables 1 and 2), we plot relative bias (or bias) and RMSE (or MSE) of posterior means and modes. For data simulated under the inference model, we also plot proportion of individuals correctly assigned and assignment probability.

Simulations of the inference model

We detected convergence problems in 31 MCMC runs out of a total of 290. In these 31 cases, the MCMC chain got trapped in a region of high posterior probability and did not sample the whole parameter space, leading to estimates that deviated strongly from the true parameter values. We observed convergence problems more frequently for scenarios with low genetic differentiation ($F_{ST} = 0.01$, 0.02) or high migration rates ($m = 0.15$).

As explained in the Appendix, in the present case the Bayesian deviance can be decomposed into a term based on the likelihood of a genotype given a particular migration ancestry, D_{gen}, and a term based on the probability of a particular assignment given a migration rate, D_{assign}. For each replicate, we estimated both components and also the overall deviance. Interestingly, lack of convergence was better identified using D_{assign} instead of the overall deviance. In all cases, D_{assign} of MCMC runs with convergence problems was much higher than that of 'good' runs (see Table S1, Supplementary material for an example), indicating that this statistic can be used for identifying suspect runs when the method is applied to real data sets.

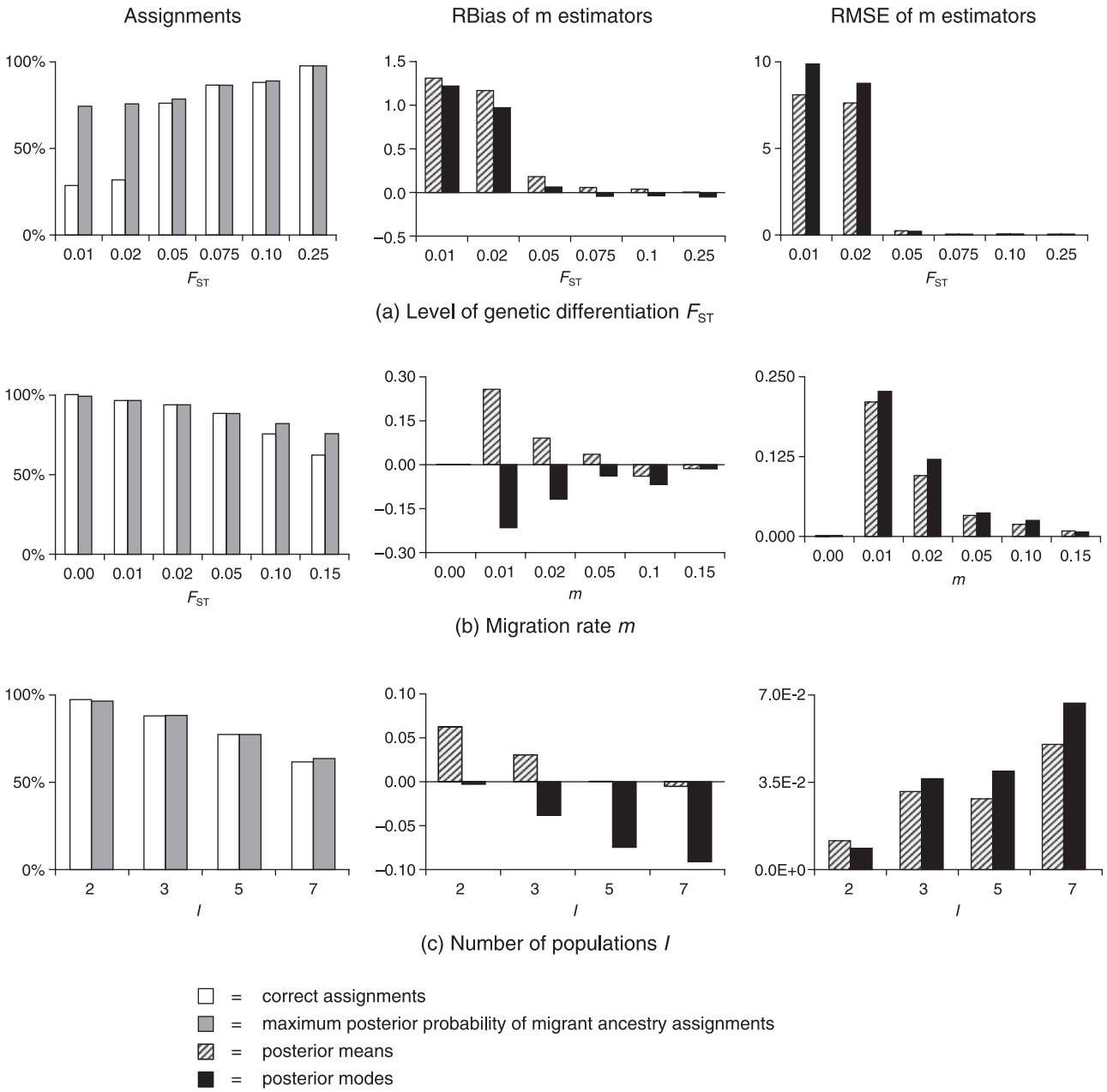

Fig. 2 Results for the data sets simulated under the inference model. Assignments, relative bias and RMSE of migration rate estimates when varying (a) level of genetic differentiation keeping $m^* = 0.05$, (b) migration rate with F_{ST} fixed at 0.10 and (c) number of populations I with $m^* = 0.05$ and $F_{ST} = 0.10$. Values of all other parameters are listed in Table 1.

The effect of genetic differentiation is very important; the accuracy of individual assignments and estimated migration rates increases with increasing F_{ST} values (Fig. 2a). Note that when genetic differentiation is low ($F_{ST} = 0.01$, 0.02), the individual assignments are very inaccurate but the maximum posterior probability with which individuals are wrongly assigned is very high. Thus, a high posterior assignment probability is not necessarily a good indication of how much confidence we can place on the assignments. As proportion of correct assignments increases, the bias of estimated migration rates decreases and their accuracy increases. In general, estimates of migration rates based on the mode are less biased than those based on the mean but their RMSE is larger, indicating that their variance is higher (RMSE = RBias^2 + variance).

The effect of varying migration rates (Fig. 2b) is less pronounced than that of varying F_{ST}, probably due to the fact that m^* and F_{ST} are decoupled in these simulations. As

migration rate increases, the proportion of correct assignments decreases but it is still above 60% for migration rates as high as 0.15 (when F_{ST} is fixed at 0.1). It is not possible to calculate the relative bias and RMSE when there is no migration ($m^* = 0$), so for this particular case we calculated the bias and MSE (results not shown), which show that the mean produces overestimates while the mode has no bias at all. For low and intermediate migration rates, the mean gives overestimates while the mode gives underestimates; for large values both underestimate the true value. The bias and RMSE of both estimators decrease as migration rate increases. The observed change of sign in the bias of estimates based on the mean is due to the fact that the method sets an upper limit of 1/3 for the total proportion of migrants in a population. Thus, when the true migration rate is close to this upper limit, the parameter space becomes very asymmetric around the true value and the MCMC will visit more often smaller than larger values. This also has the effect of decreasing the RMSE because the MCMC will not be able to visit values that are much larger than the true value.

Increasing the number of populations decreases the accuracy of individual assignments and estimates of migration rates (Fig. 2c). With only two populations, bias is much larger for the mean than for the mode but as more populations are added, the bias of the latter increases rapidly while that of the mean decreases. The accuracy of both estimators of migration rates decreases rapidly as the number of population increases but more so for the mode than for the mean.

Another important aspect to investigate is the effect of size differences among local populations, since large differences are likely to increase the strength of genetic drift, and therefore have an effect on the accuracy of migration rate estimates. Table 3 compares the results for a scenario with equal local sizes (200 individuals) and another with two populations of size 50, one of size 200 and two others of size 500. When all local populations are equal in size, 77% of individuals are correctly assigned with a posterior probability of 0.77. However, when they differ in size, the proportion of individuals correctly assigned drops to 51% but the posterior probability remains high (0.75). The estimates of migration rate are also strongly affected and more so for the mean than for the mode. The relative bias of the mean for the scenario with unequal population sizes is one order of magnitude higher than that with equal sizes. Note that in the case of the mode, there is an underestimation of migration rates when all populations have equal sizes but an overestimation when they differ in size. The RMSE of both mean and mode is one order of magnitude larger when populations differ in size.

It is also important to investigate the effect of the amount of the data used for the estimation, which can be characterized in terms of sample sizes, number of loci scored and their degree of polymorphism (number of allelic classes). Sample size does not seem to have much of an effect on the accuracy of individual assignments, but this is not the case for estimates of migration rates (Fig. 3a). As sample size increases, the bias and RMSE of both the mode and the mean decrease. The mode always underestimates migration rates while the mean overestimates them, but the absolute value of the bias is more or less the same for both. The RMSE is much larger for the mode for sample sizes of 20 individuals, but for larger sample sizes it is the same as that of the mean. The quality of the estimates does not seem to improve a lot for sample sizes of 60 or more.

Increasing the number of loci increases the accuracy of the individual assignments and sharply decreases the bias of the mean but does not have much of an effect on the mode; on the other hand, the RMSE of both estimators decreases sharply initially but does not change much after 15 loci (Fig. 3b). Again, the mode underestimates the migration rates while the mean overestimates them. The accuracy of individual assignments is higher for multiallelic markers than for biallelic ones but not much is gained by using loci with more than five alleles (Fig. 3c). This is also true for the bias and RMSE of both the mode and the mean. As was the case before, the mode underestimates migration rates wile the mean overestimates them. It should be noted that these results correspond to a scenario with strong

Table 3 Posterior estimates obtained when varying local population sizes in both simulation schemes with overlapping parameter spaces. We compare two scenarios: the first one with equal local sizes (200 individuals) and another with two populations of size 50, one of size 200 and two others of size 500. We report both relative bias and RMSE of mean and mode estimates and credible interval statistics

Simulation scheme	Island sizes	RBias(m)		RMSE(m)		95% CI width	Proportion of times true value falls within CI
		mean	mode	mean	mode	Migration rate	Migration rate
Inference model EASYPOP	Equal	4.1E−01	−4.0E−01	1.2E+00	1.2E+00	0.07	99%
	Unequal	1.2E+00	6.3E−01	1.2E+01	1.4E+01	0.07	75%
	Equal	8.5E−01	−1.2E−01	1.2E+01	1.1E+01	0.07	66%
	Unequal	1.2E+00	7.4E−01	1.8E+01	2.0E+01	0.06	63%

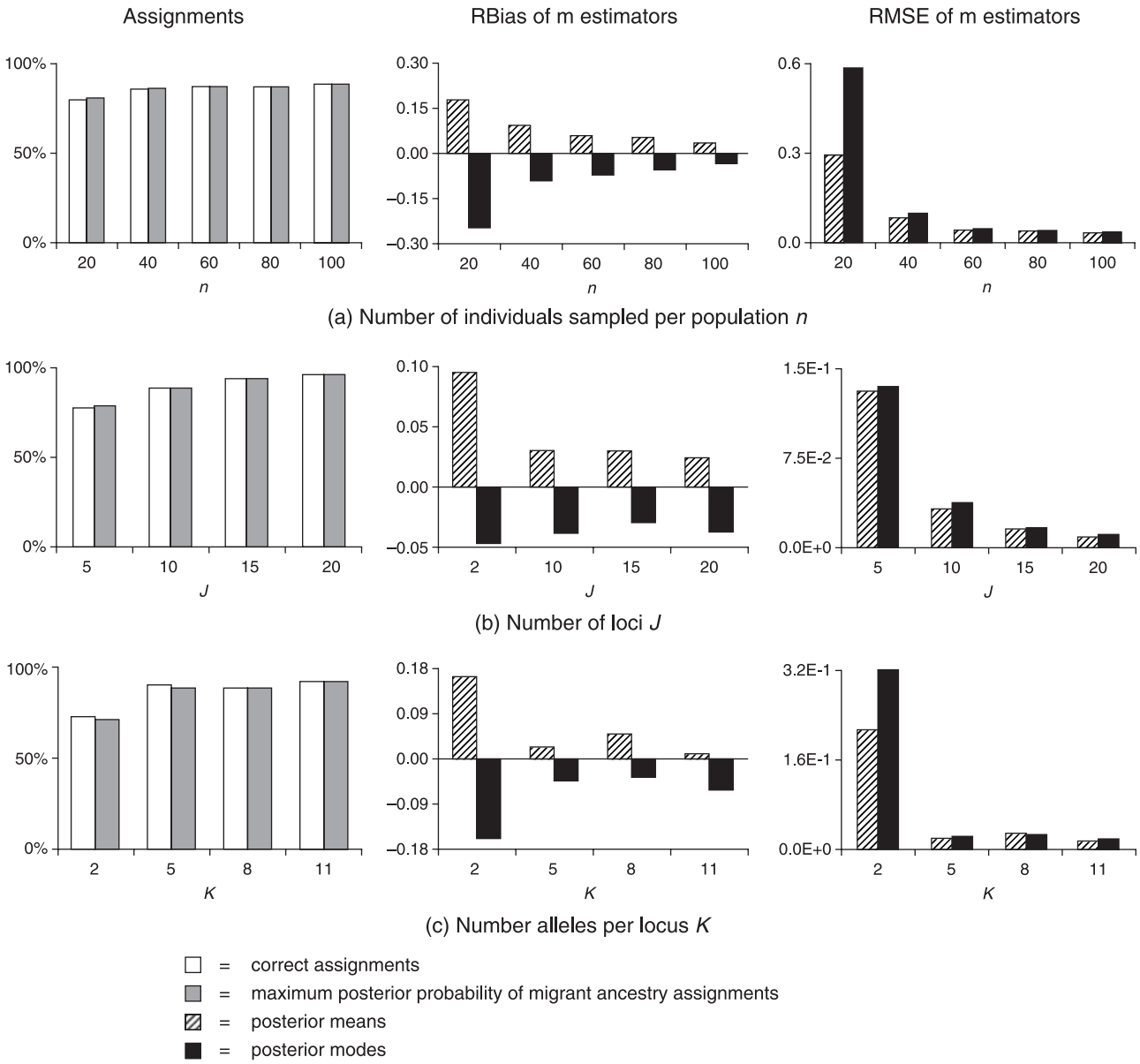

Fig. 3 Results for the data sets simulated under the inference model. Assignments, relative bias and RMSE of migration rate estimates when varying (a) number of individuals sampled per population, (b) number of loci and (c) number alleles per locus. We fixed $F_{ST} = 0.10$ and $m^* = 0.05$. Values of all other parameters are listed in Table 1.

genetic differentiation ($F_{ST} = 0.1$); with lower F_{ST} values, accuracy is likely to continue to increase as the number of loci and their variability increases.

We also investigated the effect of varying the different model parameters on the width of credible intervals, CIs, of immigration rate estimates and on the proportion of times the true value fells within the CIs. As expected, increasing the information content of the data set (i.e. increasing F_{ST}, sample size, number of loci and/or number of alleles per locus) decreases the width of the CIs (Table 4). The proportion of times the true value is within the CIs is almost always 100%; only very low F_{ST}s (less than 0.05) can lead to much lower values. Increasing migration rates increases the width of the CIs but does not have an effect on the proportion of times they contain the true value (Table 4). The number of populations does not seem to have an effect on either measure (Table 4). Finally, size differences among local populations do not influence the width very much but it can greatly decrease the proportion of times the true value falls within the CI (Table 3).

Overall, these results indicate that if the assumptions of the inference model are not violated, the method can estimate migration rates fairly accurately when genetic differentiation is at least moderate ($F_{ST} \geq 0.05$) and samples

Table 4 Credible intervals (CI) of migration rates for data simulated with the inference model. We report the width of the 95% CIs and the proportion of times the real value of the parameter falls within them when varying parameters

Parameter	Values	CI width	Proportion of times true value falls within CI	Figure
F_{ST}	0.010	0.10	38%	2(a)
	0.020	0.07	27%	
	0.050	0.08	92%	
	0.075	0.06	100%	
	0.100	0.06	100%	
	0.250	0.05	100%	
m*	0.01	0.03	100%	2(b)
	0.02	0.04	100%	
	0.05	0.06	100%	
	0.10	0.07	100%	
	0.15	0.08	100%	
I	2	0.05	100%	2(c)
	3	0.06	100%	
	5	0.06	100%	
	7	0.06	100%	
n	20	0.13	100%	3(a)
	40	0.09	100%	
	60	0.07	100%	
	80	0.06	100%	
	100	0.06	100%	
J	5	0.08	100%	3(b)
	10	0.06	100%	
	15	0.05	100%	
	20	0.05	100%	
K	2	0.11	98%	3(c)
	5	0.06	100%	
	8	0.06	100%	
	11	0.05	100%	

are of good quality (40 individuals or more and 15 multiallelic markers). In general, it is preferable to use as estimator the posterior mean migration rate, which is more accurate than the mode of the migration rate (but see Discussion).

We investigated the quality of estimators of the inbreeding coefficient only for the case of scenarios with random mating within populations ($F_I = 0$) and therefore we report the results using the MSE instead of the RMSE (Figs 4 and 5). Contrary to what was observed for migration rates, the mode is a much better estimator than the mean because posterior distributions of F are very asymmetric. Also, both the mean and the mode overestimate F. As F_{ST} increases, the bias and MSE decrease, being fairly low for an F_{ST} of 0.05 or more (Fig. 4a). Increasing migration rates increases the bias and decreases the accuracy of the estimates of F (Fig. 4b). On the other hand, increasing the number of populations does not have much of an effect on the mode but does increase the bias and MSE of the mean (Fig. 4c). The effect of the quality of the samples on estimates of F is less important than for the estimates of migration rates (see Fig. 5). Increasing the sample size does improve the estimates based on the mean but does not have much effect on those based on the mode (Fig. 5a). A similar pattern is observed when increasing the number of loci (Fig. 5b). However, the effect of increasing the number of allelic classes is rather different, since the bias does not seem to depend on how polymorphic the markers are, while the MSE is much lower for multiallelic markers than for biallelic ones (Fig. 5c).

Simulations using EASYPOP

In the case of EASYPOP data sets, we observed convergence problems even for runs with the lowest quadratic error. We observed that the MCMC chain got trapped in regions that corresponded to the bounds of the prior distribution used for the migration rates. More precisely, the proportion of nonmigrants was either close to 2/3 or to 1; conversely, the proportion of immigrants from deme q into deme l was either very close to 0 or very close to 1/3 (see examples in Figure S1, Supplementary materials). The results we present in what follows correspond to MCMC runs that had a quadratic error of the same order of magnitude as the run with the lowest error of the corresponding scenario, but it should be noted that this does not guarantee convergence. Moreover, we found only one scenario (*m1n2*, see Table 2) for which the RMSE of data sets generated with EASYPOP is of the same order of magnitude as those observed for data sets generated under the inference model. This scenario corresponds to $N = 200$ and $m = 0.01$ in which case, the F_{ST} is high (0.11). The RMSE observed for all other scenarios are at least one order of magnitude larger than those obtained for data sets simulated under the inference model. It should be noted that even if there were convergence problems, the relationship between the quadratic error and the Bayesian deviance for the assignments, D_{assign}, was as expected, that is, runs with the lowest quadratic error had the lowest deviance (see Table S2, Supplementary material).

In the simulations of the inference model, we could fix F_{ST} and the migration rates separately because it is assumed that we start with subpopulations with a certain level of genetic differentiation, which then exchange migrants for two generations. In the case of EASYPOP, this is not possible since migrants are exchanged from the very beginning of the simulations and the degree of differentiation (at equilibrium) is determined by Nm, the effective number of migrants. Thus, increasing subpopulation sizes, N, while keeping migration rates fixed at $m = 0.01$, decreases genetic differentiation and this leads to an increase in bias and RMSE (Fig. 6a). Nm can also be increased by increasing m while keeping $N = 200$ constant.

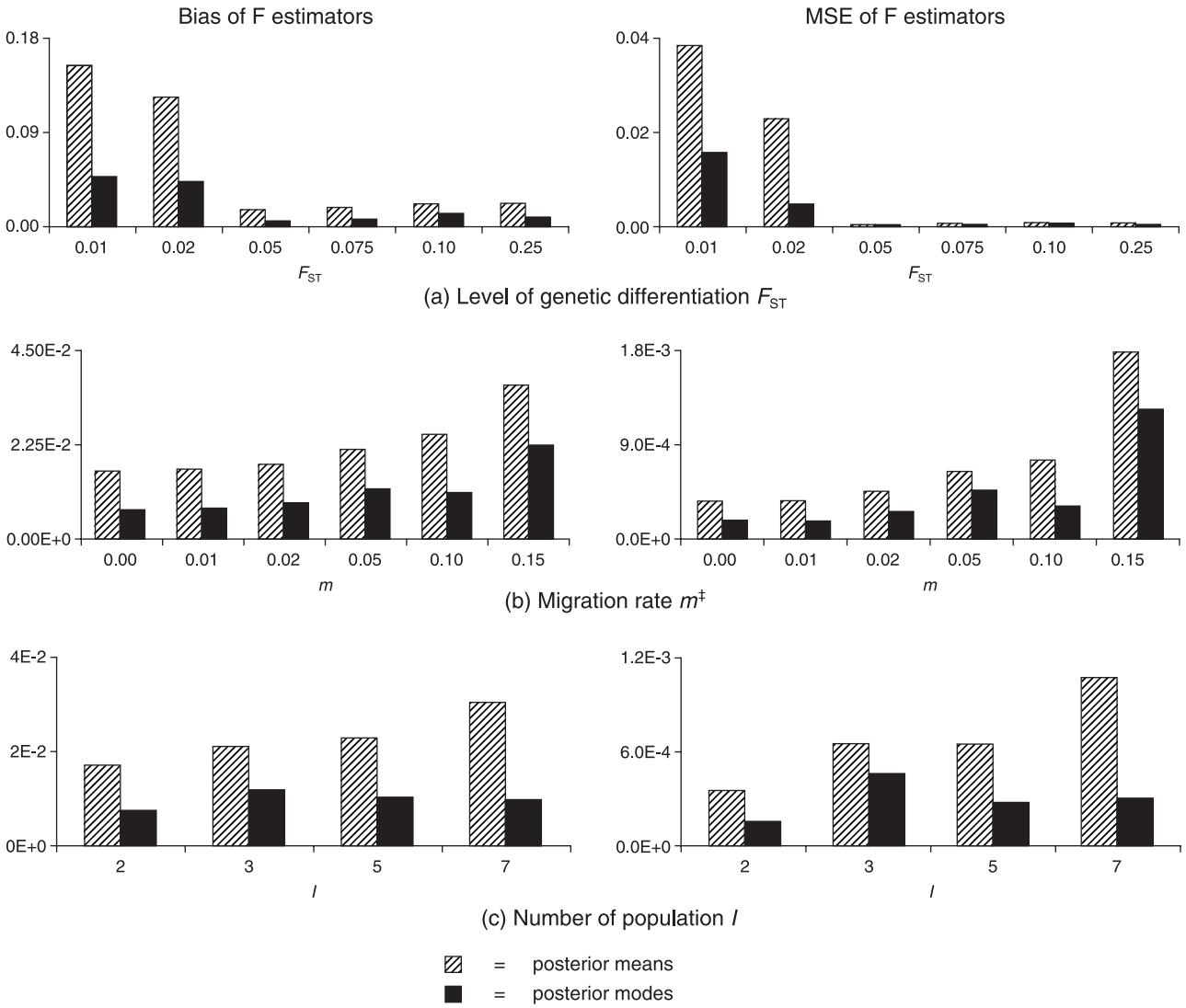

Fig. 4 Results for the data sets simulated under the inference model. Bias and MSE of inbreeding coefficient estimates when varying (a) level of genetic differentiation keeping $m^* = 0.05$, (b) migration rate with F_{ST} fixed at 0.10 and (c) number of populations with $m^* = 0.05$ and $F_{ST} = 0.10$. Values of all other parameters are listed in Table 1.

In this case, however, the results differ from those obtained when N increases and m is kept constant. As m increases, the relative bias and RMSE first increase and then decrease (Fig. 6b). If we keep Nm constant by increasing N while decreasing m, then relative bias increases while the RMSE first increase and then decrease (Fig. 7a). Thus, the quality of the estimates does not necessarily depend on F_{ST}. In fact, the explanation for these results (Figs 6b and 7a) is that, as mentioned before, convergence problems result in estimates of m_{gl} that tend to be either very close to 0 or very close to 1/3. Thus, the distance between the estimate and the true value is larger for $m^* = 0.05$ than for $m^* = 0.01, 0.10$. We also explored the effect of increasing the number of populations when the effective number of migrants per generation Nm equals 25 (Fig. 7b). As I increases, the bias and the RMSE of estimates based on both the mean and the mode decrease.

Finally, we explored the effect of unequal population sizes on migration rate estimates (Table 3). The relative bias of the mean increases with respect to that of the scenario with equal sizes but remains within the same order of magnitude. The bias of the mode goes from negative with equally sized populations to positive with unequal sizes. The RMSE of both mean and mode increases with unequal population sizes but remains within the same order of magnitude.

Varying m or N does not have much of an effect on the width of the CIs; on the other hand the proportion of times the true value falls within them is more sensitive to the migration rate than to the population size (Table 5). In

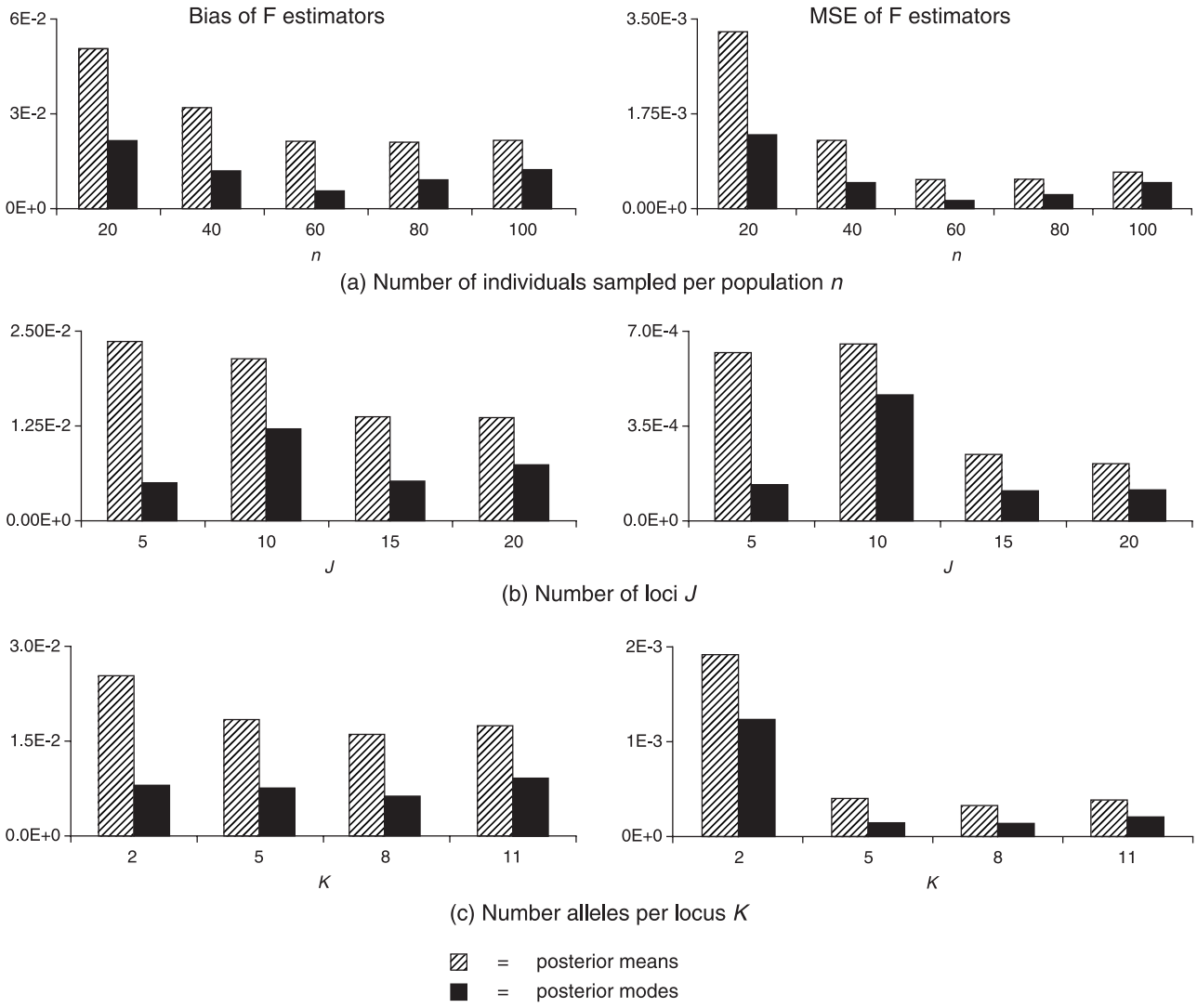

Fig. 5 Results for the data sets simulated under the inference model. Bias and MSE of inbreeding coefficient estimates when varying (a) number of individuals sampled per population, (b) number of loci and (c) number alleles per locus. We fixed $F_{ST} = 0.10$ and $m^* = 0.05$. Values of all other parameters are listed in Table 1.

Table 5 Credible interval (CI) of migration rates for data sets generated with EASYPOP. We report the width of the 95% CIs and the proportion of times the real value of the parameter falls within them

Parameter		Values	CI width	Proportion of times true value falls within CI	Figure
N		50	0.02	78%	6(a)
		200	0.02	97%	
		500	0.05	83%	
m		0.01	0.02	97%	6(b)
		0.05	0.02	0%	
		0.10	0.02	1%	
$Nm = 5$	$N = 50$	$m = 0.10$	0.08	22%	7(a)
	$N = 100$	$m = 0.05$	0.07	23%	
	$N = 500$	$m = 0.01$	0.05	83%	
I		2	0.03	15%	7(b)
		4	0.02	0%	
		8	0.02	0%	

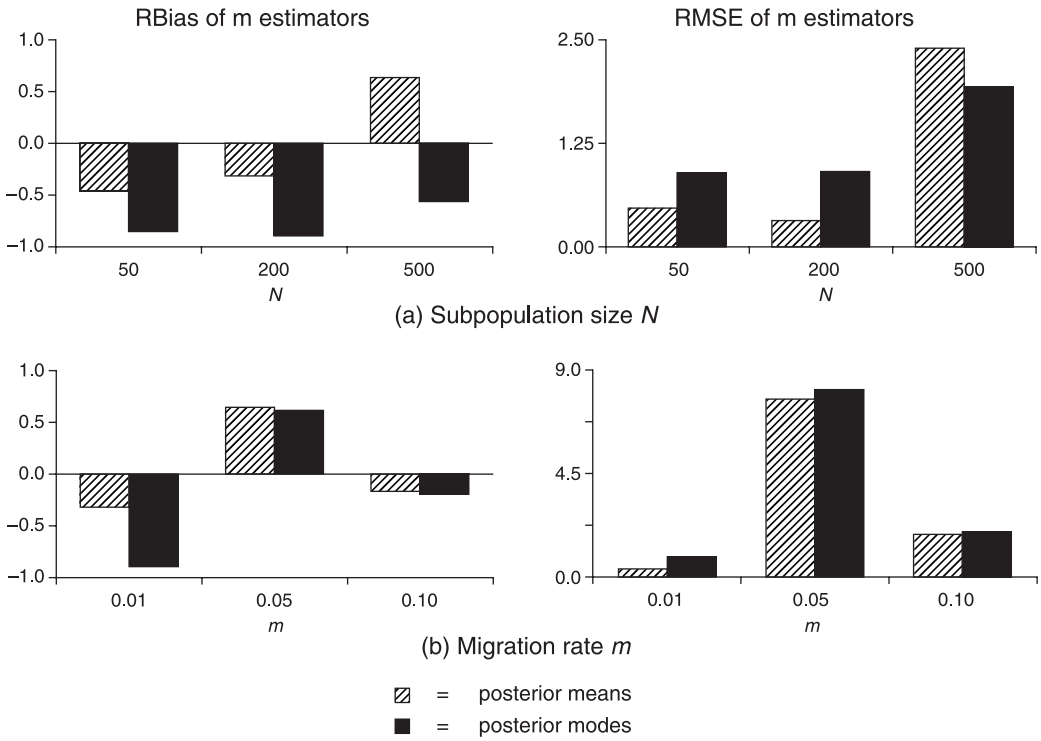

Fig. 6 Results for the data sets generated using EASYPOP. Relative bias and RMSE of migration rate estimates when varying (a) subpopulation sizes while keeping migration rate constant ($m = 0.01$) and varying (b) migration rates while keeping subpopulation sizes ($N = 200$). Values of all other parameters are listed in Table 2.

particular, this latter measure is close to 0% for migration rates higher than 0.05. Keeping $Nm = 5$ while decreasing m and increasing N decreases the width of the CIs and increases sharply the proportion of times they contain the true value. Increasing the number of populations does not change much the width of the CIs while the proportion of time the true value falls within them is very low and drops to 0% when more than two populations are considered (Table 5). It should be noted here that Nm was kept constant at 25, which explains the low values observed for this measure. Finally, size differences among local populations do not change much either measures (Table 3).

EASYPOP does not allow the user to choose a fixed value for the inbreeding coefficient. Instead, it provides three choices for the mating system: random, polygyny, and monogyny. We chose random mating but it is clear that small populations will exhibit inbreeding even under random mating. Similarly, exchanging migrants can lead to a Wahlund effect increasing F. Thus, it is difficult to establish whether a positive bias in the estimates of this parameter is not in fact due to real inbreeding and Wahlund effects. For all these reasons, we do not present results for F for data sets generated with EASYPOP.

There is little agreement between the results obtained for the data sets simulated under the inference model and those generated using EASYPOP. For example, when varying the number of populations, the RMSE increased in the first case but decreased in the second. Moreover, as previously mentioned, there is only one scenario where the quality of the estimates obtained for EASYPOP data sets was similar to those observed for data sets simulated under the inference model. Even when parameters values were the same for both sets of simulations (Table 3), the quality of estimates were better for data simulated under the inference model than for data generated with EASYPOP. The deterioration due to unequal population sizes is more pronounced for data sets generated under the inference model than for those from EASYPOP. However, RMSE is still lower for the former than for the latter. These differences suggest that if the assumptions of the inference model are violated, the estimations of migration rate obtained should be interpreted with caution.

Discussion

The results indicate that if the demographic history of the species being studied fits the assumptions of the inference

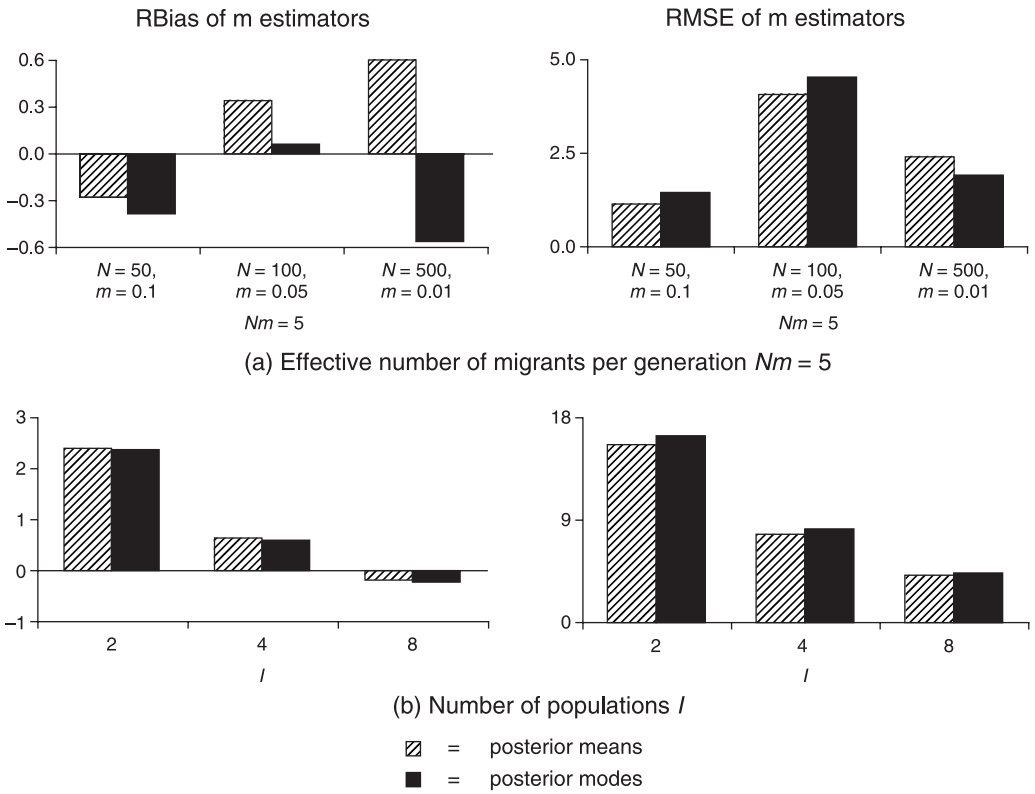

Fig. 7 Results for the data sets generated using EASYPOP. Relative bias and RMSE of migration rate estimates when varying (a) migration rate and subpopulation size while effective number of migrants per generation is fixed ($Nm = 5$) and varying (b) number of populations with $m = 0.05$ and $N = 500$. Values of all other parameters are listed in Table 2.

model, and if genetic differentiation is not too low ($F_{ST} \geq 0.05$), then the method can give fairly accurate estimates of migration rates even when they are close to the threshold (about 0.1) that leads to correlated dynamics between populations. However, when the assumptions of the inference model are violated, accurate estimates are obtained only if migration rates are very low ($m = 0.01$) and genetic differentiation is high ($F_{ST} \geq 0.10$). Our results also show that using posterior assignment probabilities as an indication of how much confidence we can place on the assignments is problematical since the individual assignments can be very inaccurate but the maximum posterior probability with which individuals are (wrongly) assigned can still be very high, as illustrated by the results for scenarios with low genetic differentiation ($F_{ST} = 0.01, 0.02$; cf. Fig. 2a). This is a rather unexpected result since in principle, when genetic differentiation is low ($F_{ST} = 0.01, 0.02$), the actual conditional marginal likelihood function for different assignments is relatively flat. The true conditional posterior assignment probabilities should reflect this, by being pushed towards the uniform distribution among the local populations. However, BAYESASS results indicate high certainty in the assignments. A closer look at the MCMC output indicates that there is always a population with a very low immigration rate to which individuals from the other populations are assigned mainly as second-generation migrants. The other populations tend to have a proportion of nonmigrant individuals close to 2/3 (which corresponds to the lower bound of the prior distribution for m). Thus, the population to which an individual is assigned does not change much during an MCMC run, leading to a high posterior probability. It is important to note that as opposed to other methods such as STRUCTURE (Pritchard *et al*. 2000), BAYESASS is not only carrying out assignments but is also estimating migration rates. Thus, the prior used for the migration rate can have an effect on the assignment of individuals. More precisely, although the prior for the vector of migration rates for any given population is uninformative, the marginal prior distribution for any given migration rate, m_{lq}, is not flat at all but L-shaped with a mode at zero (see Figure S2, Supplementary material). This type of prior might limit the mixing of the MCMC chain, forcing it to remain for very long periods of time at the same value of **M** (origin of migrant ancestor). Such a problem could be avoided by running extremely long MCMCs, or (probably more realistically) by improving the procedure with Metropolis-coupled MCMC (Geyer 1991).

We extended the simulation study of Wilson & Rannala (2003) by considering a larger number of populations and multiallelic markers. We also considered more values of migration rate and F_{ST} values. Our results confirm their suggestion that the use of multiallelic markers should increase accuracy of the estimates. However, as the number of populations increases, accuracy decreases. Within limits, increasing the quantity of information contained in the sample by increasing the number of loci and/or sample sizes also increases accuracy of estimates. Note, however, that with up to three populations, not much is gained by using more than 15 loci and/or more than about 40 individuals.

The results of our simulations with EASYPOP suggest that the performance of the method is rather sensitive to deviations from the assumption of negligible change in allele frequencies due to migration and/or genetic drift over a few generations. This assumption is likely to be violated when migration rates are close to the 0.10 value considered as threshold for demographic independence. Thus, BAYESASS is unlikely to be useful for the identification of demographically independent units for borderline cases, which are the most interesting since it is very easy to identify demographic independence when migration rates are much smaller than 0.1.

The posterior mean of the migration rates seems to be a better estimator than the posterior mode because in general its RMSE is lower and its bias only a little bit higher than that of the mode. However, sometimes there is a need to be conservative. For example, we might prefer to err on the side of keeping two populations as separate management units rather than combining them; if this is the case, then it is better to use the mode. Additionally, when only two populations are involved (as is often the case in applications to management), the mode is always a better estimator of migration rates than the mean.

In general, although users of methods for the estimation of demographic parameters focus on point estimates rather than CIs, the latter can be a better way of evaluating the performance of Bayesian methods. In general, we expect that when data sets are highly informative, the width of the CIs will be narrow, while poor data sets will produce very wide CIs. In both cases, however, we expect that the proportion of times the true value falls within the CIs be very high. This is in general what we observe for the data sets simulated under the inference model; however, EASYPOP data sets give rather narrow CIs that in general do not contain the true value. This is another indication that we should be extremely cautious in the interpretation of results provided by BAYESASS when we suspect that the species being studied does not fit very well the assumptions of the method. One way of identifying unreliable results is to verify if the CIs are narrow and very close to one of the boundaries of the prior used for the migration rates. For example, one may obtain immigration rate estimates that are very close to 1/3 or 0, and correspondingly estimates of the proportion of nonmigrants that are close to either 2/3 or 1. If this is indeed the case, it is necessary to carry out many replicate analyses using very long MCMC runs (see below).

A practical problem associated with the use of MCMC is that of establishing whether or not the chain has converged. The basic principle implemented by the MCMC method is to construct an aperiodic and irreducible Markov chain whose stationary distribution (the 'target' distribution) is that given by the Bayesian formulation (in our case eq. 1). The estimation procedure consists of running the chain for 'sufficiently' long and treating the simulated values as a dependent sample from the target distribution (Brooks 1998). The underlying logic here is that the chain will visit more often regions of parameter space with a high posterior probability. In principle, the initial state of the chain (i.e. the initial values of the parameters we need to estimate) is arbitrary because we only start collecting data after the chain has reached equilibrium (i.e. converged). In practice, however, it is difficult to be sure that the chain has indeed converged. This is particularly the case with complex data sets and models, in which case the posterior probability is likely to be multimodal. The chain can then converge to one of the modes and remain in its vicinity for extremely long periods of time, giving the impression that it has converged. Running a second MCMC on the same data but with a different initial state can give very different results. Running longer chains is unlikely to solve this problem; for example, in our case we used runs of 21×10^6 iterations and still observed that many of them produced estimates that were very different from those obtained from runs that gave estimates very close to the true parameter values. In a simulation study such as ours it is easy to identify MCMC chains that did not converge because we know the true parameter values. However, in real applications this is not possible. One potential solution is to carry out multiple MCMC runs of the same data set and then compute a measure of model fit for each one of the runs, discarding those that provide a poor fit. One such measure of model fit is the Bayesian deviance (see References in Spiegelhalter *et al*. 2002); we explain how it is calculated in the Appendix.

In the present study, we analysed several replicates for each scenario and found that some posterior estimates departed strongly from the real values. Repeating the MCMC run on the same data set but with different initial conditions led to estimates that were much closer to the true values used as input for the simulations and to much lower statistical deviances. Thus, we suggest that in order to minimize convergence problems, it is advisable to carry out many MCMC runs, say 10, and select the one with the lowest deviance for obtaining the parameter estimates. Given that using extremely long MCMC runs does not seem to solve the convergence problem, we suggest using runs of 21×10^6, discarding the first 2×10^6 as burn-in. We

have applied this strategy to one of the scenarios generated using EASYPOP (data set *m1n2*; see Table 2). Of the 10 replicate runs, three (6, 7 and 10) have a very high deviance for the assignment component, D_{assign} (Table S3, Supplementary material) and relative bias and RMSE at least one order of magnitude larger than the other runs (all of which have very similar low values). Table 6 presents the results taken from the best and from one of the worst runs (runs 1 and 6, respectively). In the chosen example, the true migration rate was 0.01 and the best run provides estimates that are almost identical to this value. On the other hand, the estimates obtained from run 6 contain many migration rate estimates that are very different from the true value (m_{11}, m_{14}, m_{41} and m_{44}). In both cases, the CIs are very narrow regardless of whether the estimates are accurate or not. In this case, 10 replicates allowed discrimination between good and bad runs. However, if there are reasons to think that the species under study departs strongly from the assumptions of the inference model, then it would be appropriate to increase the number of replicate runs.

The most likely cause of the convergence problem we observed is the prior used for the migration rates, which sets bounds of 0 and 1/3 for the immigration rates and, equivalently, 2/3 and 1 for the proportion of nonimmigrants. Our simulation study shows that the chain gets trapped in regions of parameter space that correspond to these values. In this regard, it is interesting to note that in the example of the grey wolf provided by Wilson & Rannala (2003; Table 2), the estimated proportion of nonimmigrants into each population is close to either 1/3 or to 1. It is possible to avoid this convergence problem if the assumptions of the model (e.g. migration does not change the allele frequencies over the two generations considered) are not violated, in which case, following the advice provided above will suffice to insure convergence. However, if the assumptions are violated it is very difficult to avoid the biases introduced by the convergence problem.

Convergence problems have been reported for many recently developed Bayesian methods such as STRUCTURE, GENELAND (Guillot *et al*. 2005) and BAPS (Corander *et al*. 2004). In the case of STRUCTURE, Evanno *et al*. (2005) proposed a method based on running several MCMCs and calculating an ad-hoc statistic, Δk, based on the rate of change in the log probability of data between successive k values. The problem with this method is that there is always the potential of including in the calculation of Δk several chains that have not converged, leading to results that are unreliable. We observed this type of behaviour in a previous study (Waples & Gaggiotti 2006) and concluded that, for the simple finite island model that we considered, Evanno *et al*.'s (2005) method does not perform better than the original approach proposed by Pritchard *et al*. (2000). We think that it is better to use the same strategy used by Pritchard *et al*. (2000), namely run several chains for each value of k, say 20, and for each select the MCMC run that gives the smallest value of $-2 \log \Pr(X/k)$. Using these

Table 6 Comparison of two runs of the same replicate of data set *m1n2* (see Table 2). True value for migration rate is $m = 0.01$. We used MCMC runs 1 and 6 of Table S3. The former provide accurate estimates while the latter presents convergence problems. Migration rate estimates with such problems are highlighted

	Chain 1				Chain 6			
	Estimates		Credible interval bounds		Estimates		Credible interval bounds	
Migration matrix	Mean	Mode	Lower	Upper	Mean	Mode	Lower	Upper
m11	0.98	0.99	0.95	1.00	**0.67**	**0.67**	**0.67**	**0.69**
m12	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.03
m13	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.02
m14	0.00	0.00	0.00	0.02	**0.31**	**0.32**	**0.29**	**0.33**
m21	0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.02
m22	0.99	1.00	0.97	1.00	0.99	1.00	0.97	1.00
m23	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01
m24	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.01
m31	0.01	0.00	0.00	0.02	0.00	0.00	0.00	0.01
m32	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01
m33	0.99	1.00	0.97	1.00	0.99	1.00	0.97	1.00
m34	0.02	0.00	0.00	0.01	0.01	0.00	0.00	0.02
m41	0.01	0.00	0.00	0.01	**0.31**	**0.32**	**0.29**	**0.33**
m42	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01
m43	0.00	0.01	0.00	0.04	0.00	0.00	0.00	0.03
m44	0.98	0.99	0.95	1.00	**0.67**	**0.67**	**0.67**	**0.69**

chains one can then select the value of k that best fits the data set and base all estimations on the results of the best MCMC run. It should be noted that, as stated by Pritchard *et al*. (2000), $-2 \log \Pr(X/k)$ is simply the mean of the Bayesian deviance penalized by a quarter of its variance.

In the case of GENELAND, Guillot *et al*. (2005) proposed a similar approach to that used by Pritchard *et al*. (2000) for STRUCTURE, but in this case they used the mode of the posterior distribution for the number of populations as the criterion to choose the best MCMC runs. Finally, in the case of BAPS, Corander *et al*. (2004) proposed a similar strategy to that used by Pritchard *et al*. (2000) and Guillot *et al*. (2005) but using the posterior probability of the partition as the basis to select the best run.

Clearly, Bayesian methods such as the one we evaluate in this article are very powerful and offer an opportunity for answering difficult questions in ecology, population genetics, evolution and conservation biology, but we should be aware that their application is not as straightforward as that of the frequentist methods that have been used in past. Thus, users of these new methods should endeavour to follow very closely the recommendations provided by the software manuals and also seek the advice of colleagues competent in Bayesian methods. Furthermore, users should be aware that the models can have limited power to provide meaningful estimates under many realistic real-world scenarios, especially those that involve low levels of genetic differentiation.

Acknowledgements

We thank Bruce Rannala for helpfully discussing with us the results of this study. Three anonymous reviewers made very useful comments that greatly improved the manuscript. This work was supported by the Fond National de la Science (grant ACI-Impbio-2004–42-ADGP). P.F. holds a Ph.D. studentship from the Ministère de la Recherche.

Supplementary material

The supplementary material is available from
http://www.blackwellpublishing.com/products/journals/suppmat/MEC/MEC3218/MEC3218sm.htm

Fig. S1 Results for data generated with EASYPOP. We represent posterior distributions of proportion of (a) nonmigrants and (b), (c) and (d) various immigration rates from typical output files corresponding to the three parameter sets presented in Fig. 6(b). Subpopulation sizes are fixed at $N = 200$ while varying m (0.01, 0.05, and 0.10). The chains with migration rates equal to 0.05 or 0.10 are trapped in modes corresponding to the bounds of the prior for migration rates (0 and 1/3). This explains the pattern observed on Figs 6(b) and 7(a) where RMSE is larger for a moderate value (0.05) of true migration rate than for extreme values (0.01, 0.10).

Fig. S2 Plot of the Dirichlet prior distribution used by BAYESASS for the migration rates. In this case, we consider a scenario with $I = 4$ populations so that the migration matrix contains 16 elements. Although the prior is uniform in the multidimensional space and nonmigrant proportion uniformly distributed on the interval (2/3,1), the marginal prior for the immigration rates are L-shaped with a mode at 0.

Table S1 Results of the analyses of data sets simulated with the inference model. Simulated data sets consist of 10 repetitions with $I = 3$ populations with high level of genetic differentiation ($F_{ST} = 0.10$) and low migration rate ($m^* = 0.05$). We used $J = 10$ loci with $K = 11$ allele states and $n = 80$ individuals per population. Highlighted replicate presents convergence problems

Table S2 Results of the analyses of EASYPOP *m1n2* data sets. Simulated data sets consist of $I = 4$ populations of constant size $N = 200$. We used low migration rate $m = 0.01$, $J = 20$ loci with $K = 10$ allele states and $n = 50$ individuals per population. Highlighted replicates present convergence problems

Table S3 Results for 10 MCMC runs for the same replicate of m1n2 data set (see Table 2). We highlight runs that have a large value for the assignment component of the Bayesian deviance D_{assign}. The corresponding chains and their posterior estimates or credible intervals (CI) provide very different results than the others.

References

Abdo Z, Crandall KA, Joyce P (2004) Evaluating the performance of likelihood methods for detecting population structure and migration. *Molecular Ecology*, **13**, 837–851.

Balding DJ, Nichols RA (1997) Significant genetic correlations among Caucasians at forensic DNA loci. *Heredity*, **78**, 583–589.

Balloux F (2001) EASYPOP (version 1.7): a computer program for population genetics simulations. *Journal of Heredity*, **92**, 301–302.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences, USA*, **98**, 4563–4568.

Brooks SP (1998) Markov chain Monte Carlo method and its application. *Journal of the Royal Statistical Society Series D — the Statistician*, **47**, 69–100.

Corander J, Waldmann P, Marttinen P, Sillanpää MJ (2004) BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics*, **20**, 2363–2369.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Gaggiotti OE, Brooks SP, Amos W, Harwood J (2004) Combining demographic, environmental and genetic data to test hypotheses about colonization events in metapopulations. *Molecular Ecology*, **13**, 811–825.

Geyer CJ (1991) Markov chain Monte Carlo maximum likelihood. In: *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface* (ed. Keramidas EM), pp. 156–163. Interface Foundation, Fairfax Station,Virginia.

Guillot G, Estoup A, Mortier F, Cosson JF (2005) A spatial statistical model for landscape genetics. *Genetics*, **170**, 1261–1280.

Hastings A (1993) Complex interactions between dispersal and dynamics – lessons from coupled logistic equations. *Ecology*, **74**, 1362–1372.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measure of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**, 583–616.

Waples RS, Gaggiotti OE (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1439.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: F_{ST} not equal to $1/(4Nm + 1)$. *Heredity*, **82**, 117–125.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177–1191.

Pierre Faubet is a PhD student working on statistical methods for the study of the genetical structure of populations. Robin Waples is interested in developing and applying population genetic principles to real-world problems in ecology, conservation, and management. His research focuses on population genetics and conservation genetics of marine and anadromous fishes. Oscar Gaggiotti's research focuses on developing theory and statistical methods aimed at bridging the gap between population ecology, population genetics and evolution. Much of his research is applied to the study of metapopulations.

Appendix: Bayesian deviance

In this section we outline the calculation of the Bayesian deviance, which we use to discriminate between MCMC runs that converged from those that did not. We base our discussion on the work of Spiegelhalter *et al*. (2002), who used the deviance statistic to define the DIC, a measure for choosing the model that provides the best fit among a group of alternative models. In our case we are not comparing models and therefore we simply use the Bayesian deviance, which has been proposed as a measure of model fit by a number of authors (see References in Spiegelhalter *et al*. 2002).

In Bayesian statistical modelling of data y we specify a prior distribution $f(\theta)$, $\theta \in \Theta$, and a likelihood $\Pr(y/\theta)$, which give rise to a marginal distribution

$$p(y) = \int_{\Theta} \Pr(y/\theta) f(\theta)\, d\theta \tag{9}$$

The Bayesian deviance is then defined as:

$$D(\theta) = -2 \log \Pr(y/\theta) + 2 \log g(y), \tag{10}$$

where $g(y)$ is some fully specified standardizing term which is function of the data alone. We can assume without loss of generality that $g(y) = 1$, so

$$D(\theta) = -2 \log \Pr(y/\theta) \tag{11}$$

We can thus estimate the expected deviance, $E_{\theta/y}[D(\theta)]$ from a MCMC run by taking the sample mean, $\overline{D(\theta)}$ of the simulated values of $D(\theta)$.

In order to calculate the deviance for a hierarchical model such as that implemented in BAYESASS, we need to define the parameter on which we want to focus. Hierarchical Bayesian models further parameterize the prior(s) with unknown 'hyper-parameters' Ψ to obtain a full probability model

$$p(y, \theta, \psi) = p(y,\theta)\Pr(\theta/\psi) f(\psi) \tag{12}$$

Then, depending on the parameters in focus, we can specify the model in terms of the likelihood $\Pr(y/\theta)$ and prior $f(\theta) = \int \Pr(\theta/\psi) f(\psi)\, d\psi$, or in terms of the likelihood $\Pr(y/\psi) = \int_{\Theta} \Pr(y/\theta)\Pr(\theta/\psi) d\theta$ and prior $f(\Psi)$. In our case, we are interested in using bayesass to estimate migration rates so we will focus on $\mathbf{m}$ and thus consider the likelihood, which is $\Pr(\mathbf{X}/\mathbf{S}; \mathbf{M},\mathbf{t},\mathbf{F},\mathbf{p})\ \Pr(\mathbf{M},\mathbf{t}/\mathbf{m})$, and the prior $f_m(\mathbf{m})$. Thus, the deviance is composed of two terms, the first one, D_{gen}, concerns the likelihood of the genotypes and the second one, D_{assign}, the probabilities of assignments:

$$D(\mathbf{m}) = \underbrace{\overline{-2 \log \Pr(\mathbf{X}/\mathbf{S},\mathbf{M},\mathbf{t},\mathbf{F},\mathbf{p})}}_{D_{\text{gen}}} - \underbrace{\overline{2 \log \Pr(\mathbf{M},\mathbf{t}/\mathbf{m})}}_{D_{\text{gen}}} \tag{13}$$