Inferring recent migration rates from individual genotypes

THOMAS BROQUET,* JONATHAN YEARSLEY,* ALEXANDRE H. HIRZEL, JÉRÔME GOUDET and NICOLAS PERRIN

Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland

Abstract

We present a novel and straightforward method for estimating recent migration rates between discrete populations using multilocus genotype data. The approach builds upon a two-step sampling design, where individual genotypes are sampled before and after dispersal. We develop a model that estimates all pairwise backwards migration rates (m_{ij}, the probability that an individual sampled in population i is a migrant from population j) between a set of populations. The method is validated with simulated data and compared with the methods of BayesAss and Structure. First, we use data for an island model and then we consider more realistic data simulations for a metapopulation of the greater white-toothed shrew (*Crocidura russula*). We show that the precision and bias of estimates primarily depend upon the proportion of individuals sampled in each population. Weak sampling designs may particularly affect the quality of the coverage provided by 95% highest posterior density intervals. We further show that it is relatively insensitive to the number of loci sampled and the overall strength of genetic structure. The method can easily be extended and makes fewer assumptions about the underlying demographic and genetic processes than currently available methods. It allows backwards migration rates to be estimated across a wide range of realistic conditions.

Keywords: backwards migration, genetic inference, immigration, mixture model, philopatry

Received 11 June 2008; revision revised 8 November 2008; accepted 23 November 2008

Introduction

Migration plays a key role in determining the demographic and genetic future of fragmented populations because it governs the isolation of local demes. This is especially true when population sizes are small because demographic stochasticity and the loss of diversity induced by genetic drift will be important factors determining the extinction risk of isolated populations (Lande 1993; Frankham *et al.* 2002). Estimating migration rates in these systems on a short-term, ecological timescale will therefore not only document interesting aspects of a species' ecology but will also provide key information for guiding conservation.

Quantifying migration in the wild has proven extremely difficult. Approaches based on mark–recapture or telemetry allow individual dispersal behaviour to be explored in great detail, but the logistics often become intractable when estimating average migration rates for an entire population

*Thomas Broquet and Jonathan Yearsley contributed equally to this paper.
Correspondence: Thomas Broquet, Fax: +41 21 692 41 65;
E-mail: thomas.broquet@unil.ch

(Koenig *et al.* 1996; Berry *et al.* 2004). Population genetics provides several alternative approaches that are less reliant on intensive data collection, but suffer from the addition of confounding factors such as genetic drift and assumptions of equilibrium. These approaches fall into two broad classes: indirect estimation of migration from its effect on genetic variation between populations and direct estimation of migration from genetic data.

Indirect approaches for estimating migration rates focus on the consequences of migration upon the apportionment of genetic variation among populations. A variety of methods based on *F*-statistics, rare alleles, cladistics, coalescence or allelic covariance have been developed (reviewed in Slatkin & Barton 1989; Neigel 1997; Sork *et al.* 1999; Rousset 2001). All these indirect methods have properties in common. First, they all rely on population genetic models in order to relate genetic variation to migration rates and other demographic parameters. These population genetic models often make important simplifications based on strong assumptions (e.g. island model, Wright 1931; Whitlock & McCauley 1999), or weight the migration rates by additional parameters, such as the effective population size or the rate

of mutation (Hey & Nielsen 2004). If not known a priori, the value of these parameters can in some cases be jointly estimated (e.g. Beerli & Felsenstein 2001; Vitalis & Couvet 2001; Wang & Whitlock 2003). Second, all indirect methods measure effective gene flow. Effective gene flow only measures migrants that participate in local reproduction, because only these individuals affect the dynamics of the local gene pool. Finally, almost all these methods yield long-term, equilibrium estimates rather than recent, intrageneration measures of migration (but see Erwing *et al.* 2004).

Direct methods, on the other hand, focus either on the detection of migrants or on the consequences of migration rates upon changes in genetic structure over a short time interval (e.g. within one generation). These methods yield short-term estimates of (non-effective) dispersal (i.e. the proportion of immigrants in a population, irrespective of whether they will be reproductive). One approach towards estimating short-term, non-effective migration rates between pairs of populations is assignment, which aims to detect immigrants and assign them to potential source populations (reviewed in Manel *et al.* 2005). This approach is mainly limited by the level of genetic structure in the system, because the structure has to be high enough for assignment statistics to be powerful while migration itself tends to reduce genetic structure (Paetkau *et al.* 2004). Vitalis (2002) and Fontanillas *et al.* (2004) presented another approach which requires populations to be sampled at two time points which span the migration events that are of interest. With these data, the change in genetic structure between the two sampling events can be directly linked to average immigration rate, assuming that migration probabilities are homogeneous (i.e. island model for migration). Finally, mixture analysis can directly estimate pairwise rates of migration (Pella & Masuda 2001; Wilson & Rannala 2003; Reynolds & Templin 2004). For example, Wilson & Rannala (2003) used observed multilocus genotypes from each population to estimate the proportion of immigrants, as well as the inbreeding coefficient, allele frequencies in each population, and the ancestry of each individual (i.e. the origin of its immigrating ancestors and the number of generations since immigration occurred). One limitation of this method is that it assumes low migration rates and negligible drift over the last two generations.

Here we present a novel, direct method for estimating non-effective migration rates, which is related both to Vitalis' two-step sampling and to mixture analyses. Building upon the approaches of Vitalis (2002) and Fontanillas *et al.* (2004), we propose a model for estimating recent pairwise backwards migration rates between sampling locations. The method is the first to allow recent pairwise migration rates to be estimated over a clearly defined time period (the period between the two samples) with a minimum of assumptions (no requirements for an island model, migration–drift equilibrium, Hardy–Weinberg equilibrium or stationary

parameters). We investigate the method's performance using simulated data from both Wright's island model and a realistic metapopulation model of the greater white-toothed shrew (*Crocidura russula*), and compare it to results from BayesAss (Wilson & Rannala 2003) and Structure (Pritchard *et al.* 2000). The generation of simulated, yet realistic, data allows the statistical properties of our method to be tested across a range of scenarios commonly encountered with field data.

The model

Our goal is to estimate **M**, the matrix of backward rates of migration between times t and T (where t occurs before T) for all pairs of populations in a system of P populations. Specifically, we will estimate m_{xy} which is the proportion of individuals sampled in population y at time T that were present in population x at the earlier time t. We refer to time t as pre-dispersal and time T as post-dispersal. When relevant, we will use lower- and uppercase for variables sampled at time t and T, respectively. Following Vitalis (2002) and Fontanillas *et al.* (2004), we start with the genotype data, g and G, from n and N individuals sampled at times t and T, respectively. The main assumptions of our method is that drift is negligible between times t and T, and that the P populations form a closed system (i.e. no immigrants from outside the system). The former is best fulfilled when the two data sets (g and G) are collected within one generation (Vitalis 2002; Fontanillas *et al.* 2004): in this case, we would have a sample of pre-dispersal genotypes (i.e. genotypes of individuals sampled in the population where they were born) and a sample of post-dispersal genotypes (i.e. genotypes of older individuals sampled after natal dispersal). The method is therefore best suited to inferring recent migration events.

Each individual is genotyped at L loci that are assumed to be independent. Locus l is observed to have A_l alleles, and therefore, has a total of $S_l = A_l(A_l + 1)/2$ possible states [A_l homozygous states and $A_l(A_l − 1)/2$ heterozygous states], which need not be at Hardy–Weinberg equilibrium. The observed locations (or population identities) of the n pre-dispersal individuals (sampled at time t) are recorded in the vector **z**. Similarly, the locations of the N post-dispersal individuals (sampled at time T) are recorded in the vector **Z**. In contrast to these observed data, the pre-dispersal locations are unknown for the N post-dispersal individuals that were only sampled at time T. Let these unknown locations be contained in the vector, **z̃**, with N elements (one for each individual sampled at time T). Finally, let $\mathbf{\Phi}^{(l)}$ be an $S_l \times P$ matrix containing the frequency of genotype states for locus l in each population at time t (data on all loci are represented by $\mathbf{\Phi}$). The location data **z** and **z̃**, along with genotypes g and G, are observations, while **z̃** and $\mathbf{\Phi}^{(l)}$ are to be inferred from the observed data. The discrepancy

between inferred pre-dispersal locations $\tilde{\mathbf{z}}$ vs. the observed post-dispersal locations $\mathbf{Z}$ will give us an estimate of the backwards migration rates (see equation 6).

Under the assumptions of our model, the genotypes of our sampled individuals at times t and T are independent draws from a distribution which is completely specified by $\Pr(g|\Phi, \tilde{\mathbf{z}})$ and $\Pr(G|\Phi, \tilde{\mathbf{z}})$, respectively. The posterior distribution of $\tilde{\mathbf{z}}$ and Φ in light of the observed genotypes g and G and the observed sampling locations $\mathbf{z}$ at time t is given by Bayes rule

$$\Pr(\tilde{\mathbf{z}}, \Phi | g, G, \mathbf{z}) = \Pr(G|\Phi, \tilde{\mathbf{z}})\Pr(g|\Phi, \mathbf{z})\Pr(\Phi)\Pr(\tilde{\mathbf{z}}), \quad \text{(eqn 1)}$$

where $\Pr(\Phi)$ and $\Pr(\tilde{\mathbf{z}})$ are the prior distributions for the frequency of genotype states and the individual locations at time t. The likelihoods $\Pr(g|\Phi, \mathbf{z})$ and $\Pr(G|\Phi, \tilde{\mathbf{z}})$ in equation 1 are given by

$$\Pr(g|\Phi, \mathbf{z}) = \prod_{l=1, i=1}^{L,n} \Pr(g_i^{(l)}|\Phi^{(l)}, \mathbf{z}) \quad \text{(eqn 2)}$$

$$\Pr(G|\Phi, \tilde{\mathbf{z}}) = \prod_{l=1, i=1}^{L,N} \Pr(G_i^{(l)}|\Phi^{(l)}, \tilde{\mathbf{z}}), \quad \text{(eqn 3)}$$

where

$$\Pr(g_i^{(l)} = s|\Phi^{(l)}, \mathbf{z}) = f_{z_i l s} \quad \text{(eqn 4)}$$

$$\Pr(G_i^{(l)} = s|\Phi^{(l)}, \tilde{\mathbf{z}}) = f_{\tilde{z}_i l s}, \quad \text{(eqn 5)}$$

where z_i and $\tilde{z}_i$ are the ith elements of vectors $\mathbf{z}$ and $\tilde{\mathbf{z}}$ respectively, and f_{xls} is the frequency at time t of state s on locus l in population x. Since it is assumed that the dispersal of all individuals can be described by the same backwards migration rate matrix, the probability model for $\tilde{z}_i$ can be written in terms of the backwards migration rate matrix

$$\Pr(\tilde{\mathbf{z}}_i = x|\Phi, \mathbf{M}, Z_i = y) = m_{xy} \quad \text{(eqn 6)}$$

where Z_i is the ith element of vector $\mathbf{Z}$. Equation 6 considers the observed location, Z_i, of the ith individual sampled at time T and the inferred location, $\tilde{z}_i$, of that individual at the earlier time t. It defines the backwards migration rate, m_{xy}, as the probability of an individual being in population x at time t given that it was sampled in population y at time T. Equation 6 therefore provides the link between the data, Z_i, and inference, $\tilde{z}_i$, upon individuals and the backwards migration rates which we wish to estimate.

To complete the model it remains necessary to specify the distributions for Φ and $\mathbf{M}$. We note that Φ is the distribution of marker frequencies, which are the frequencies of genotype states as we have described here. However, the same model can be used for many different types of markers (e.g.

allele frequencies, allozymes, morphological traits) and each type of marker will have its own particularities when selecting a distribution for Φ. A common choice for a distribution of frequencies is the Dirichlet distribution. For allele frequency data, this choice would have theoretical support (Balding & Nichols 1995). Another possible choice, which gives more flexibility than the Dirichlet, is the logistic-normal distribution (Aitchison 1986). For simplicity, we take Φ to be the observed frequency of genotype states in our pre-dispersal data, g, for each locus and population. This assumes a good sampling of possible states, and any biases in the sampling of genotype states will lead to biases in the estimates of migration rates. For poor sampling, the number of states can be reduced by considering allele rather than genotype frequencies since there are fewer allelic states compared to genotype states. In general, however, the use of allele frequencies does not capture all the information from the data, because there are fewer allelic states compared to genotype states. For data where the number of genotype states is low (e.g. single nucleotide polymorphism data), it would be feasible to extend the method so as to estimate the parameters of Φ's distribution as part of the model, similar to the parameters for $\mathbf{M}$'s distribution.

The distribution of backwards migration rates for population y (the vector $\mathbf{m}_{\cdot y}$ with P elements) could also take a Dirichlet distribution (e.g. Pritchard *et al*. 2000; Pella & Masuda 2001). However, we use a logistic-normal distribution (Aitchison 1986) to describe $\mathbf{m}_{\cdot y}$ because it performs well and allows our method to be generalized to describe covariances between migration rates when data are collected from more than two time points. The Dirichlet distribution is unable to describe such covariance (Aitchison 1986). Our prior for $\mathbf{m}_{\cdot y}$ is therefore

$$\Pr(\mathbf{m}_{\cdot y}|\mathbf{\mu}_y, \mathbf{\Gamma}_y) = \left(\sqrt{(2\pi)^{P-1}\det(\mathbf{\Gamma}_y)}\prod_{x=1}^{P}m_{xy}\right)^{-1}$$
$$\exp\left[-\frac{1}{2}(\phi(\mathbf{m}_{\cdot y}) - \mathbf{\mu}_y)^{T}\mathbf{\Gamma}_y^{-1}(\phi(\mathbf{m}_{\cdot y}) - \mathbf{\mu}_y)\right], \quad \text{(eqn 7)}$$

where $\phi(\mathbf{m}_{\cdot y})$ is the log-ratio transform of $\mathbf{m}_{\cdot y}$,

$$\phi(\mathbf{m}_{\cdot y}) = \left[\log\left(\frac{m_{1y}}{m_{Py}}\right), \ldots, \log\left(\frac{m_{P-1y}}{m_{Py}}\right)\right]^{T} \quad \text{(eqn 8)}$$

while $\mathbf{\mu}_y$ and $\mathbf{\Gamma}_y$ are hyperparameters giving the mean and precision matrix of $\phi(\mathbf{m}_{\cdot y})$. The hyperparameter $\mathbf{\mu}_y$ can be related to parameters of a Dirichlet distribution in a certain limit (the Kullback–Leibler logistic normal approximation Aitchison 1986), while $\mathbf{\Gamma}_y$ describes the covariance among the pairwise migration rates, which are parameters not available for a Dirichlet distribution. The hyperprior for $\mathbf{\mu}_y$ is taken to be a non-informative multivariate normal distribution, meaning that we intend parameter estimates to be only affected by the data, with negligible effect of the

prior. An objective method for constructing a non-informative prior is still a problematic area (Kass & Wasserman 1996; Gelman *et al*. 2004; Ghosh *et al*. 2006). If non-informative priors are to be used, each analysis should verify that the chosen prior has little effect upon the estimates by varying the prior distribution, although this approach also has its problems (Kass & Wasserman 1996). However, the quality of the estimates as evaluated through simulations can ultimately confirm that the prior truly had little influence. For Γ_y, an inverse Wishart distribution is used,

$$\Gamma_y^{-1} \sim \text{Wishart}(a\Omega, P-1), \tag{eqn 9}$$

where Ω is a matrix set to specify a correlation structure where the compositions are independent, apart from the summation constraint (i.e. the Kullback-Leibler logistic normal approximation to a Dirichlet, Aitchison 1986; Billheimer 2001). Following Billheimer (2001), we set a diffuse prior by using $a = 0.1$.

Our model was written and analysed with openBUGS (Thomas *et al*. 2006) and R (R Development Core Team 2005) using the BRugs interface (Thomas *et al*. 2006). We performed three Markov chain Monte Carlo (MCMC) runs with over-dispersed initial conditions, each with a burn-in period of 20 000 iterations. Convergence of the chains was analysed using the R package CODA (Plummer *et al*. 2006). After the burn-in period, backwards migration rates, m_{xy}, were recorded for a further 2000 iterations (with a thin of 100, resulting in a total chain length of 200 000 after burn-in). The Gelman and Rubin convergence diagnostic (Gelman & Rubin 1992), trace plots of the variables, and comparison of the migration estimates from each chain independently were used to detect obvious chain convergence problems. Each backwards migration rate estimate is taken to be the median of the m_{xy}, reported along with 95% highest posterior density interval (HPDI). While real migration rates can take the values of 0 or 1, the high level of numerical precision used throughout MCMC sampling makes it unlikely for HPDI region to include these extreme exact values. HPDI limits were thus rounded to three decimal places (e.g. a lower HPDI of 0.0004 was rounded to zero). The lowest possible non-zero HPDI is therefore 0.001, which, given our population sizes, corresponds to less than one migrant. We define the 'coverage quality' of our method as the proportion of HPDI that include true migration values for each simulation scenario. R scripts are available from the authors on request.

Data simulation

Island model data

We first tested our method with Wright's island model at a migration–drift equilibrium (Wright 1931). This allows us to test the method across a range of expected migration rates and levels of population structure. It also helps us to compare our method with previously published methods that were only tested on island simulations. Assuming that each locus carries two alleles (e.g. single nucleotide polymorphic markers) and the overall frequency of allele 1 at locus i is q_i, then under Wright's island model, the stationary distribution of frequencies for allele 1 is approximated by a beta distribution,

$$f(p_i) = \frac{\Gamma(4Nm)}{\Gamma(4Nmq_i)\Gamma(4Nm[1-q_i])} p_i^{4Nmq_i-1}(1-p_i)^{4Nm(1-q_i)-1},$$

$$\tag{eqn 10}$$

where N is the effective population size and m is the migration rate (migration is isotropic and homogeneous). We generated pre-dispersal allele frequencies for three populations by sampling from equation 10 for 20 independent loci, assuming $q_i = 0.5$ (other values of q_i gave very similar results). Post-dispersal genotypes were then generated for a sample of 100 individuals assuming all migration rates are equal to m. Migration rate took values of 0.01, 0.05, 0.1 and 0.2 while the expected $F_{ST} = 1/(4Nm+1)$ took values of 0.01, 0.05, 0.1 and 0.25. A migration rate of $m = 0.01$ implies that on average, one individual among the sample of 100 will be a migrant. Simulations were performed for all combinations of m and F_{ST}, and simulations for each parameter set were replicated 20 times. The migration rate, m, was estimated from the estimated philopatry across the three populations.

Realistic metapopulation data

The data from an island model is quick to generate and has few parameters, but few populations in the field are expected to behave like an island model. Moreover, the beta distribution of underlying allelic frequencies used above for the island model is a very special case, which need not be assumed with our method. To further test our method, we generated realistic genetic data sets from individual-based simulations of a greater white-toothed shrew (*Crocidura russula*) metapopulation (parameters shown in Table 1). These data include realistic descriptions of the life cycle for a diploid species with zygotic dispersal, microsatellite markers, population sizes, and the stochastic variation due to the individual-based nature of simulations. The metapopulation was based upon 10 of the villages studied by Ehinger *et al*. (2002) and Fontanillas *et al*. (2004) (the 10 low-altitude villages in figure 1 of Ehinger *et al*. 2002). We took the shrew life cycle to be: birth, dispersal, population regulation, reproduction and death. Generations were non-overlapping, mating was random within each village with a Poisson distribution of fecundity, and population regulation was by truncation according to a

Table 1 Details of life-cycle parameters used in simulations of a shrew metapopulation

Parameters	Default value	Range
Dispersal*		
Probability of leaving a village†,c	0.3 (♂); 0.43 (♀)	
Preference for dispersal to nearby villages, β	$1/2/\text{km}^{-1}$	
Survival rate during dispersal, α	$1/3/\text{km}^{-1}$	
Population regulation		
Carrying capacity	50	
Reproduction‡		
Mean fecundity	3.5	3.5–28
Sex ratio of offspring	0.5	
Sampling		
Number of demes sampled (simulated)	10 (10)	4–10 (10–16)
Proportion of juveniles (adults) sampled	100% (100%)	25–100% (25–100%)
Number of microsatellite loci§	7	2–8

*parameters reproduce average dispersal rates found by Fontanillas *et al*. (2004).
†female (♀) and male (♂) values according to Favre *et al*. (1997) and Fontanillas *et al*. (2004).
‡values based upon data from Churchfield (1990), Genoud & Perrin (1994), Favre *et al*. (1997), Bouteiller & Perrin (2000), Duarte *et al*. (2003).
§values based upon the study by Fontanillas *et al*. (2004).

village's carrying capacity. Each juvenile individual has a probability of leaving village i and dispersing towards village j given by

$$P_{ij} = c\exp(-\beta d_{ij})\Big/\sum_{k \neq i}\exp(-\beta d_{ik}), \quad \text{(eqn 11)}$$

where d_{ij} is the distance from village i to village j (which has a median of 3.7 km for our landscape, and ranges from 0.8 to 9.8 km), c is the probability of leaving a village, and β is the preference for dispersing to nearby villages. Increasing β therefore increases the preference for migration to nearby villages. The probability of surviving dispersal between villages i and j was given by

$$S_{ij} = \exp(-\alpha d_{ij}), \quad \text{(eqn 12)}$$

where α is the survival rate during dispersal. The expected number of migrants dispersing from village i to village j is therefore given by the number of juveniles produced in i multiplied by the product of P_{ij} and S_{ij}. Dispersal parameters, α, β and c, were chosen to reproduce the average dispersal rates found by Fontanillas *et al*. (2004, average immigration probability = 0.15 for males and 0.24 for females). This gave a maximum of 13 migrants between two populations (median = 0, mean = 1.1).

Imitating the genetic data collected by Fontanillas *et al*. (2004), each individual had a diploid genotype with seven biparentally inherited microsatellite loci (meiosis and syngamy are simultaneously simulated at the reproduction step, during which each offspring receives randomly chosen alleles from each of its parents). Each locus had a maximum of 18 alleles and a mutation rate of 10^{-4} (stepwise mutation model). We started the simulations with each population at its carrying capacity and a random assignment of alleles for each individual. The default parameters gave an average population size of 80 individuals before dispersal and 50 individuals after dispersal and population regulation. Unless otherwise stated, we ran all simulations for 25 generations, and each parameter set was replicated 20 times (parameter sets shown in Table 1). We recorded the genotypes of juveniles (i.e. before dispersal) and adults (i.e. after dispersal and population regulation) from the final generation. The pairwise rates of backwards migration were estimated and compared to the simulation's true migration rates. All simulations were performed with Ecogenetics (individual-based model, version 1.5.3.60, Hirzel & Jaquiéry 2007).

Sampling issues

It is unlikely that a whole metapopulation can be perfectly sampled. Using the shrew metapopulation simulations, we therefore recreated four sampling issues that will be commonly encountered in the field. First, we independently reduced the sampling proportion of pre-dispersal individuals or post-dispersal individuals (i.e. reducing the proportion of one category of individuals while sampling all individuals from the other), and the sampling proportion of both pre- and post-dispersal individuals together. Second, we investigated the effect of not sampling all of the populations by either sampling a subset of the 10 villages, or introducing additional 'ghost' populations but only

sampling the original 10 villages (with perfect sampling within each village in both cases). Third, since time and money commonly limit the total sampling effort, we simulated a trade-off between the proportion of populations sampled and the proportion of individuals sampled per population. Finally, we varied the intensity of the genetic sampling by changing the number of microsatellite loci used in the analysis. The availability of molecular markers may differ among specific study cases, and it may greatly affect the performance of the method since it controls the amount of genetic information for distinguishing different populations.

Effect of the level of genetic structure

The amount of genetic structure present in a metapopulation is an important factor affecting the performance of genetic-based migration estimates (e.g. Paetkau *et al.* 2004). A very high level of structure may indicate isolated populations with very low migration rates that will be difficult to detect. On the other hand, a low level of structure makes it difficult to identify individual populations from the genetic data, which then makes migrants difficult to identify even if they are common. We tested for the effect of genetic structure on our migration estimates for the shrew metapopulation simulations while keeping all other parameters constant (with 50% of pre- and post-dispersal individuals sampled in each population). Genetic structures (measured by global F_{ST}) of 0.015, 0.03, 0.04 and 0.05 were simulated. These values correspond to simulations of 2, 4, 8 and 25 generations (because initial allelic frequencies followed the same uniform distribution within each population at the onset of a simulation, reducing simulation time allows us to simulate data with identical life-cycle parameters but reduced genetic differentiation). Since our method compares pre- and post-dispersal individuals from the same generation, the number of simulated generations prior to data collection is not a confounding factor. There is no alternative way to vary the level of structure in the system while keeping everything else constant (except in an island case, where F_{ST} depends entirely on N_e and m; see above).

Comparison with other methods

First, we compared the results obtained with our approach to that obtained using the method of Vitalis (2002). While this method is based on the island model and therefore aims at estimating the average migration in a system of populations, the comparison is interesting since Vitalis pioneered the idea of using a two-sample design. We then compared the efficiency of our method to that of two published methods that can estimate recent pairwise migration events: Structure (Pritchard *et al.* 2000) and BayesAss (Wilson & Rannala 2003). Our goal here was not to test all available methods in a wide range of conditions,

but to provide a comparison of our method with some of the 'state of the art' methods in average conditions. We chose to test the different approaches by re-analysing five replicates from a simulation with 50% sampling of pre- and post-dispersal data (from the realistic metapopulation case described above). Both methods were intended for use with samples taken at a single time point, but Structure can actually make use of both pre- and post-dispersal data sets using the following approach. Pre- and post-dispersal data must appear in the same input file, but with distinct population labels. Then, options must be defined so that only the population information from pre-dispersal data is used to update allelic frequencies (set USEPOPINFO selection flag to 1 for the pre-dispersal genotypes only; select 'use population information', 'no admixture model' and set MIGRPRIOR to 0 in the ancestry model options; and select 'update allele frequencies with POPFLAG = 1 data' in the advanced options). Following this procedure, estimates of pairwise migration rates are given by the 'proportion of membership of each pre-defined population (i.e. pre-dispersal) in each of the clusters' (post-dispersal). The length of the burn-in period was set to 50 000 repeats, and the number of MCMC iterations used in the estimation was 10^6. The software BayesAss was run using the simulated post-dispersal data only with all parameters set to their default values.

Simulation results

Island model data

We quantified the bias and precision of our estimates with the median and the 95% interquantile range of the absolute error, respectively. When the genetic structure is strong (F_{ST} = 0.1 and 0.25), the method gives excellent estimates of migration rate, with a slight underestimation for all but the strongest migration rate (Fig. 1C, D). The precision of the estimates increased as migration rate decreased, due to the fact that migration rates are bounded between zero and one. For example, when m = 0.1 and F_{ST} = 0.01 (Fig. 1A), the distribution of errors is concentrated at –0.1 because our method finds no evidence of migration for the vast majority of the data. Decreasing F_{ST} makes the population allele frequencies more similar, and thus harder for migration to be detected. Consistent with this, we found that the precision of our estimates decreased with decreasing F_{ST} and that this decrease was most severe for the smallest migration rates (Fig. 1). The bias also became more negative with decreasing F_{ST} due to an overestimation of philopatry. For instance, for an F_{ST} = 0.05 (Fig. 1B), only the estimates of m = 0.2 were reliable. Most estimates returned by our method for lower migration rates (m = 0.01, 0.05 and 0.1, respectively) were equal to zero, which makes the absolute error equal to the migration rate (for instance for m = 0.01,
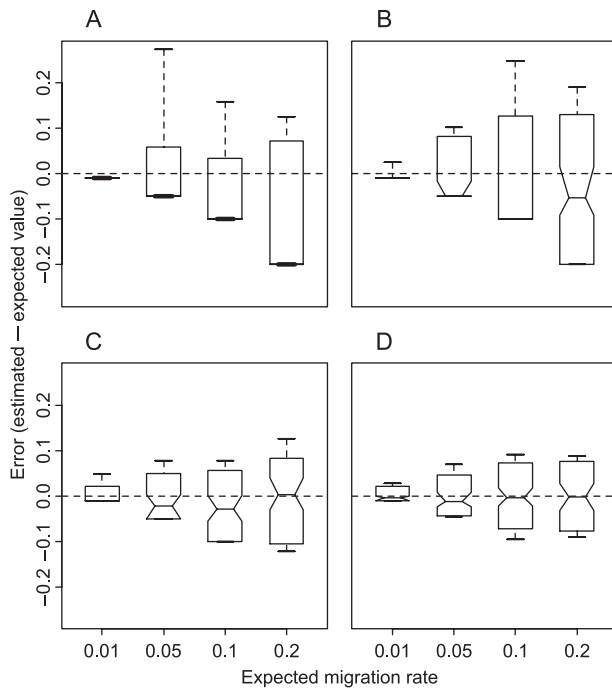
Fig. 1 The error in the estimation of the expected migration rate, *m*, from three populations of an island model. (A) $F_{ST} = 0.01$, (B) $F_{ST} = 0.05$, (C) $F_{ST} = 0.1$, (D) $F_{ST} = 0.25$ ($4Nm = 1/F_{ST} - 1$ where N is the population effective size). In each of the three populations 100 post-dispersal individuals are sampled. Results are for 20 simulations, giving 60 estimates of *m*. The solid line in each box shows the median of the error distribution, while whiskers show the full error range, the box shows the 5% and 95% quantiles, and the notch shows 25% and 75% quantiles.

the method returned a majority of estimates equal to 0, and thus an error of –0.01, meaning that *m* is not properly estimated). This is shown by the corresponding boxplots in Fig. 1B where the lower half of the distribution of errors is condensed (e.g. when *m* = 0.01 the lower half of the boxplot is condensed to –0.01; see also Fig. 1A for a similar pattern with $F_{ST} = 0.01$).

Shrew metapopulation data

Figure 2 shows the migration estimates using all individuals from each simulation of the shrew metapopulation (average $F_{ST} = 0.05$, average immigration rate = 0.2, average backwards migration rate between two given distinct populations $m_{xy} = 0.022$). Since dispersal events are a stochastic process, the true migration rates show considerable variation from one generation to the next, which is well estimated by the method. Figure 2(B) shows that the bias in both philopatry and immigration is negligible, although the estimates are slightly skewed towards lower immigration (and hence higher philopatry).
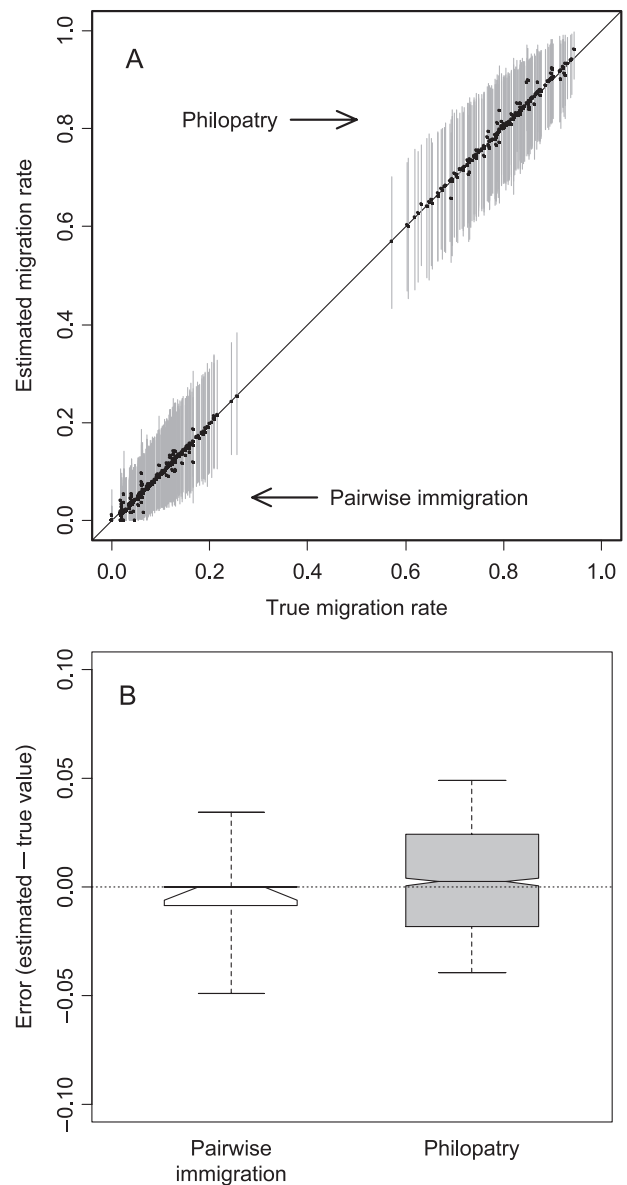

Fig. 2 Backwards migration rate estimates for a shrew metapopulation. Data from 20 simulations are shown. (A) Estimated backward migration rates against the true simulated values with 95% highest posterior density intervals. Each black dot represents the backwards migration m_{xy} between two given populations of a given replicate (philopatry data refer to m_{xx} values, i.e. the proportion of individuals in a given population that originate from the same population). (B) The error in the estimation of philopatry (grey box) and immigration (clear box). The solid line in each box shows the median of the error distribution, while whiskers show the full range of the errors, the box shows the 5% and 95% quantiles, and the notch shows 25% and 75% quantiles.

Sampling issues

Using the data simulated from the shrew metapopulation, we found that the sampling intensity in each population
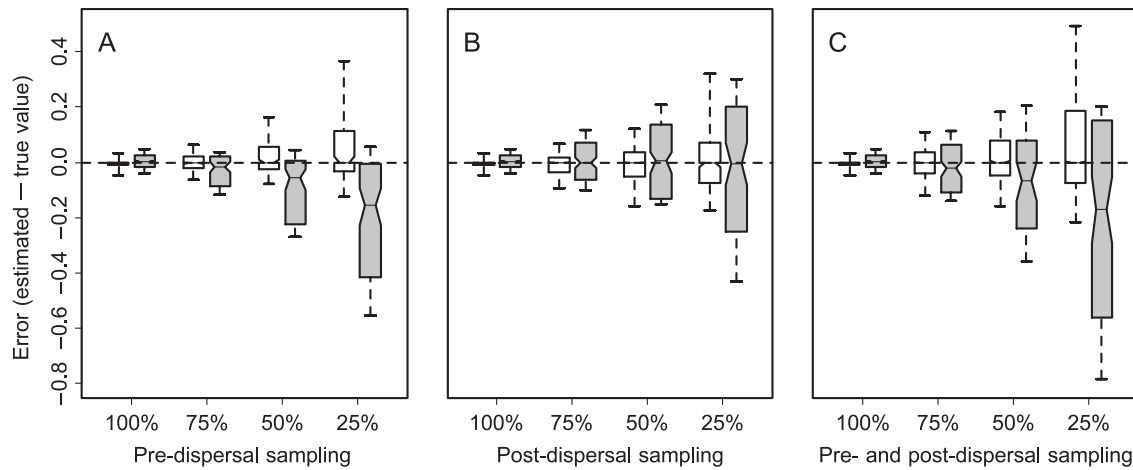
Fig. 3 The error in philopatry (grey bars) and immigration (clear bars) estimations for various population samples. (A) Pre-dispersal sampling, (B) post-dispersal sampling, or (C) both. Box plots are as in Fig. 2B.
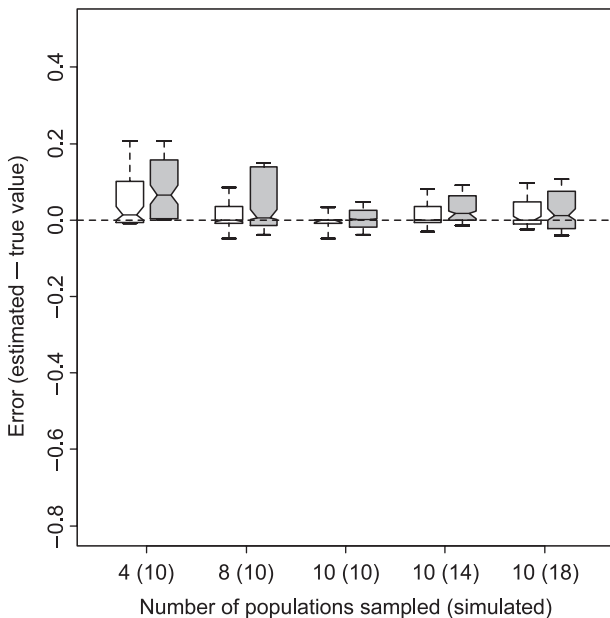

Fig. 4 The error in philopatry (grey bars) and immigration (clear bars) estimations for various numbers of sampled populations. On the left hand side, the 10 populations are sub-sampled (on the far left, four out of 10 populations are sampled). On the right hand side, additional ghost populations are added but not sampled (on the far right, six ghost populations are added). Box plots as in Fig. 2B.

strongly affected both the bias and precision of the backwards migration rate estimates (Fig. 3). Imperfect sampling of pre-dispersal individuals caused an underestimation of philopatry and decreased the precision of both immigration and philopatry estimates (Fig. 3A). Imperfect sampling of post-dispersal individuals decreased the precision of the estimates but introduced no bias (Fig. 3B).

The effects of pre- and post-dispersal sample intensity combined additively, implying no interaction between the two sampling periods (Fig. 3C). Also, the proportion of HPDI containing the true value (i.e. the coverage quality) decreased as the proportion of individuals sampled within each population got smaller (e.g. coverage quality was 100% in case of perfect sampling, and decreased to 87% when half of the individuals were sampled in each population).

Unsampled populations did not have a strong effect upon the precision of the migration estimates but they did result in a slight overestimation of philopatry (Fig. 4). This positive bias depended more upon the number of unsampled populations, rather than their proportion (i.e. sampling four populations out of 10 gave a higher bias in both philopatry and migration than sampling 10 populations out of 18, despite the latter case having many more parameters to estimate). The same statement applies to HPDI coverage, which remained larger than 96% in all but the case where four populations out of 10 were sampled (HPDI coverage = 80%).

Comparing Figs 3 and 4 shows that our method is more sensitive to the sampling of individuals rather than the sampling of populations. Given a fixed sampling effort (i.e. a constant number of sampled individuals across the metapopulation), our method had the greatest precision when sampling intensity was focused upon sampling many individuals within a few populations (Fig. 5). However, the bias was least (and HPDI coverage highest: 92%) when approximately 60% of individuals and populations were sampled.

Finally, decreasing the number of microsatellite loci had surprisingly little effect on the quality of migration estimates (Fig. 6). As expected, the precision of the estimation increased as the number of loci increased, but this effect was negligible for more than six loci. The bias (coming
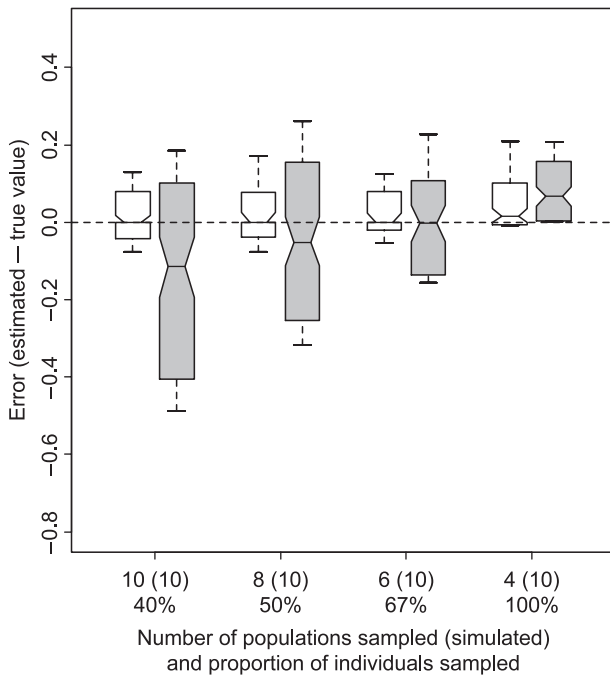
Fig. 5 The error in philopatry (grey bars) and immigration (clear bars) estimations for various trade-offs between number of populations sampled and proportion of individuals sampled per population (i.e. the total number of individuals sampled across the entire metapopulation is constant). Box plots are as in Fig. 2B.
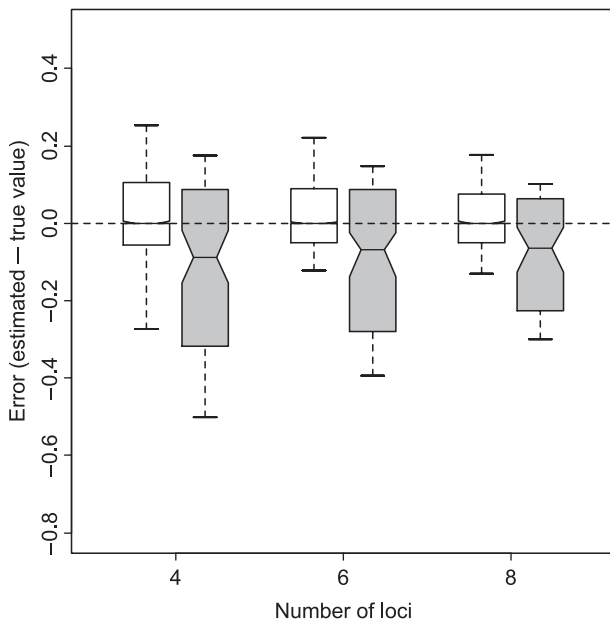

Fig. 6 The error in philopatry (grey bars) and immigration (clear bars) estimations for various numbers of sampled microsatellite loci. These analyses use 50% sampling of pre- and post-dispersal individuals. Box plots are as in Fig. 2B.
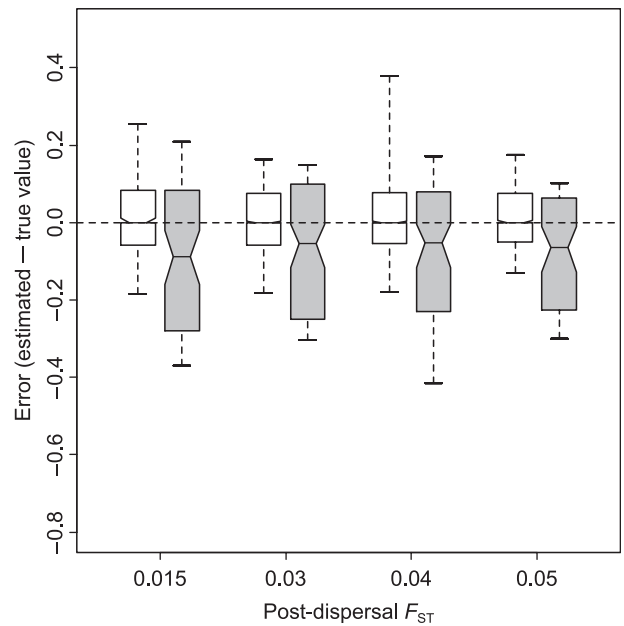

Fig. 7 The error in philopatry (grey bars) and immigration (clear bars) estimations for different genetic structures. An F_{ST} of 0.015 is a weak genetic structure for a real metapopulation, while an F_{ST} of 0.05 is an average value commonly observed in field studies. These analyses use 50% sampling of pre- and post-dispersal individuals. Box plots are as in Fig. 2B.

from 50% subsampling of post-dispersal data) only started to increase for estimates of philopatry with four loci, while HPDI coverage quality remained unaffected.

Population-genetic issues

For our default parameters, the equilibrium genetic structure in our realistic metapopulation was moderate (pre-dispersal $F_{ST} \sim 0.07$; post-dispersal $F_{ST} \sim 0.05$). Reducing the level of structure had very little effect on the precision and bias of estimated migration rates (Fig. 7). Even the lowest level of genetic structure (pre-dispersal $F_{ST} \sim 0.024$; post-dispersal $F_{ST} \sim 0.015$) gave no noticeable decrease in the accuracy and precision of the estimates, and only a slight reduction in the quality of HPDI coverage (it dropped from 87 to 85% for post-dispersal F_{ST} values of 0.05 and 0.015, respectively).

Comparison with other approaches

The approach of Vitalis (2002) appears to introduce a positive bias on philopatry estimates compared to our method of approximately 25% (data not shown). The bias and precision of the results from Structure and BayesAss is nearly indistinguishable from our method (Fig. 8). Our method gave slightly more accurate results than Structure,
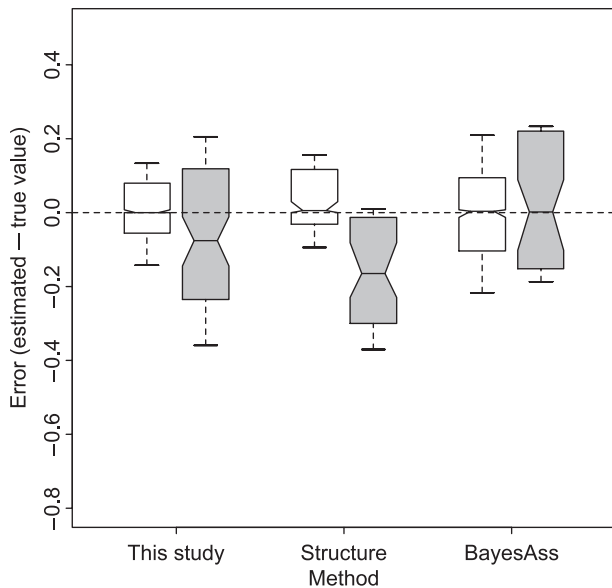
Fig. 8 The error in philopatry (grey bars) and immigration (clear bars) estimates using our method and two alternative approaches implemented in Structure and BayesAss. These results were obtained from five replicates, using 50% sampling of pre- and post-dispersal individuals. Box plots are as in Fig. 2B.

while BayesAss returned the least-biased estimates (especially for philopatry). However, for BayesAss, the confidence intervals for the estimates with the largest error were often too small, which meant that they did not include the true value (this occurred in 17% of immigration estimates and 58% of philopatry estimates, as compared to 13% and 14%, respectively, with our approach).

Discussion

Method's performance

Our results show that backward migration rates can be satisfactorily estimated under a wide range of realistic conditions. These estimates are for a clearly defined time period, making them of interest for ecological and conservation studies. Furthermore, breeding populations need not be identified since the method makes no assumptions about the distribution of genotypes in a local population (e.g. HW equilibrium). Therefore, for the purposes of the method, the spatial locations of the genetic samples can be used to define what we call a 'local population'. With the realistic data from the simulated shrew metapopulation, the estimated migration rates correctly tracked the stochastic variation in migration rates between populations and between replicates, showing the power of a two-step sampling design for estimating recent migration rates. Estimates for the island

model data were of lower quality than those from the realistic metapopulation data. Further simulations showed this to be due to the smaller number of segregating alleles used in the island simulations (data not presented). Using the island model to study a range of migration rates and F_{ST} values showed a trade-off between the amount of genetic structure between populations and the minimum amount of immigration that can be detected. With an F_{ST} of ~ 0.05, the minimum detectable immigration rate was approximately m ~ 0.2. Further simulations (not presented) showed that this trade-off could be significantly improved by using more bi-allelic loci.

Sampling intensity was by far the most important determinant of estimation quality. Pre- and post-dispersal sampling intensity of individuals had distinctly different effects. Poor pre-dispersal sampling reduced the probability of correctly associating post-dispersal individuals with their source population. At worst, association of individuals will be random, which for 10 populations would give a philopatry of 0.1. This low value of philopatry for random association explains the negative bias in philopatry as pre-dispersal sampling intensity declined (Fig. 3A). On the other hand, poor post-dispersal sampling reduced the sample size of immigrants, which decreased precision but introduced no bias into the estimates (Fig. 3B, and see Fontanillas *et al.* 2004 for discussion). Poor individual sampling also strongly decreased HPDI coverage quality, a point that deserves special attention. When the sampling is poor the 95% HPDI should not be used.

Unsampled populations introduce a positive bias in philopatry because some migrants from ghost populations will be incorrectly identified as nonmigrants. A similar, but weaker effect is seen for pairwise immigration estimates. Empirical studies on *Crocidura russula* report local capture rates of almost 100% (e.g. Favre *et al.* 1997; Vogel 1999; Bouteiller & Perrin 2000; Fontanillas *et al.* 2004). However, such high capture rates are obtained in local sampling units (classically in gardens around human habitations) that are smaller than the populations considered in our simulations. Our reduced sampling intensities are realistic for population-scale studies (e.g. a sampling intensity of 25% gives 20 pre- and 13 post-dispersal individuals per population).

Estimating migration in the wild using our approach requires a proper sampling design in space (i.e. a few focal populations intensively sampled) and time (i.e. capture of pre- and post-dispersal individuals, although each particular individual need not be caught twice). It can be applied to any species that can be sampled at two well-identified life stages and for which variable markers are available (although a limited number of microsatellite loci will be enough). Our two-step sampling design has the advantage that it allows our method to associate the migration rate estimates with a precise time period, and to

work with any distribution of genotype frequencies (i.e. there is no need to make assumptions such as Hardy–Weinberg, migration–drift, or linkage equilibrium). The method is, therefore, very general for estimating short-term migration rates, it is not limited to genetic markers (i.e. any set of measurable traits that show variation between populations), and can use information from several types of marker in the same analysis. More importantly, migration can be estimated without the usual confounding factors (e.g. genetic drift), because with an appropriate sampling of pre- and post-dispersal individuals, migration will be the main force affecting the genetic state of the system. If this sampling scenario cannot be respected (e.g. in species with overlapping generations or asynchronous reproduction and dispersal), there is a risk that genetic drift may affect short-term changes in genotype frequency. In such cases, our method may still perform acceptably and method validation is possible by a simulation-based pilot study for the specific species of interest.

The two-step sampling also has limitations. While our method makes no assumptions about the pre-dispersal genotype frequencies, this also makes it hard for our method to correct any sampling errors in the pre-dispersal genotype frequencies. If a multinomial distribution were used for genotype counts, then the variance of the estimated genotype frequencies could be estimated from the distribution. For more general distributions, variances and covariances of the genotype frequencies can be estimated from repeated sampling. This information could be incorporated directly into the estimation model (e.g. by using a logistic normal distribution for the genotype frequencies and estimating the covariance matrix). The error introduced by imperfect sampling of pre-dispersal populations can be seen in our simulations because we can vary the pre-dispersal sampling effort (Fig. 3A). This shows that estimates of philopatry are more sensitive to the pre-dispersal sampling intensity compared to estimates of immigration.

Methodological issues

Since our approach makes few assumptions, it is possible to modify the model over a wide range of possibilities. For instance, the model can be modified to estimate only philopatry by averaging pre-dispersal genotype frequencies from nonfocal populations. Such a model may be more robust to missing populations and data imperfections, and since there is only one parameter, it is quick and simple to estimate. Increasing the number of loci does not seem to be a promising way to improve significantly the quality of the estimates. Although this result must be taken with caution since we did not simulate large numbers of loci (and the number of loci was varied only for one simulation scenario), the genetic information content seems to quickly reach an optimum, depending upon loci polymorphism. Alternatively,

allele frequencies rather than genotype frequencies can be used, with some loss of information at the level of individuals but greater sample sizes in the allelic counts. However, we found that these simplified models gave no improvement in philopatry estimates for all but the weakest sampling intensities (data not presented).

Performing MCMC analysis of a mixture model is notoriously difficult because sampling the Markov chain can be inefficient, especially for small migration rates (Gelman & Rubin 1992). This can lead to excessive computational time, poor convergence of the Markov chains, and ultimately a poor estimation of the posterior distribution. It is likely that the most severely underestimated migration rates from our model (those where estimates and HPDI values were erroneously close to zero) were due to inefficient sampling of the Markov chain. Specific strategies for overcoming these problems have been developed (Neal 2000; Jain & Neal 2004), but are not implemented in the version of openBUGS (release 2.1.1, Thomas *et al.* 2006) used in this study. Another refinement of the method would be to use reversible jump Markov chain methods (Gelman & Rubin 1992; Brooks 2001) to remove populations that have provided no immigrants (i.e. populations that do not contribute to the mixture of individuals). This should improve the convergence of the Markov chains and provide better migration rate estimates. All these issues require that great care be taken to ensure that estimations are based upon Markov chains that have properly converged.

Comparison with other approaches

While it was not our intention here to identify the advantages and weaknesses of each method in a wide variety of conditions, applying different approaches for one basic simulation set allowed us to evaluate the performance of our method relative to others. Vitalis (2002) and Fontanillas *et al.* (2004) showed that the average immigration in an island system can be estimated from the genetic differentiation between pre- and post-dispersal individuals. Our method builds upon this approach, allowing estimates to be found for all pairwise migration rates and philopatry to be estimated with less positive bias. However, the estimation quality of our method is expected to decrease when pre- and post-dispersal samples are not from the same generation. This is partly because genetic drift will be more important across generations and partly because migration rate will generally be more variable between generations compared to within a generation. In contrast, this is not a problem for the approach of Vitalis, provided the genetic structure is at equilibrium (see Fontanillas *et al.* 2004 for an example).

Recent pairwise migration can also be estimated with BayesAss (Wilson & Rannala 2003) and Structure (Pritchard *et al.* 2000). Like our approach, these methods ignore the effect of genetic drift for short-term estimates of migration

rates. All methods performed with comparable precision. The strengths of each method lie in the implicit assumptions and the required data. Compared to our method, BayesAss requires only one sample, although at the cost of assuming that migration is weak and constant over the last two generations. In agreement with the study by Faubet *et al.* (2007), the confidence intervals from BayesAss were unreliable (in the particular conditions simulated here). Our method makes few assumptions about the pre-dispersal genotype distribution and gives migration estimates over a well-defined time period. Although Structure was not written with migration rate estimation in mind, its logic is very similar to our method. It provides a user-friendly interface, albeit at the expense of an assumption about Hardy–Weinberg equilibrium and greater underestimation of philopatry. Finally, other individual assignment techniques can estimate pairwise rates of migration (Paetkau *et al.* 2004; Manel *et al.* 2005). These techniques only require one sample, but their power decreases when the genetic differentiation between populations becomes weak.

Perspectives

We have presented a new method that gives a detailed description of the recent migration rates within a fragmented population. We have compared our approach with other published methods and find that all give acceptable results. Our approach is distinguished by few underlying assumptions, migration estimates for a clearly defined time period and its ease of adaptability to new scenarios. For instance, ghost populations could be taken into account by introducing unsampled populations (see Beerli 2004 and Slatkin 2005 for discussion), or estimation of sex-specific migration rates (as in Vitalis 2002) and migration at several levels of spatial hierarchy (Fontanillas *et al.* 2004) could be added, giving further ecologically relevant information on migration.

Genetic-based methods of migration inference offer a set of tools that are complementary to more field-intensive approaches, such as mark–recapture or telemetry. Application of genetic-based methods to complex landscape is commonly limited by the simplicity of their underlying population-genetic models. However, methods such as the one presented here can give detailed quantitative information about a species' recent migration in a complex landscape. This information can be used to guide the conservation of fragmented populations and opens up the interesting possibility of quantifying the relative effects of landscape features upon a species' dispersal behaviour.

References

Aitchison J (1986) *The Statistical Analysis of Compositional Data*. Blackburn Press, Caldwell, New Jersey.

Balding DJ, Nichols RA (1995) A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica*, **96**, 3–12.

Beerli P (2004) Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Molecular Ecology*, **13**, 827–836.

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.

Berry O, Tocher MD, Sarre SD (2004) Can assignment tests measure dispersal? *Molecular Ecology*, **13**, 551–561.

Billheimer D (2001) Compositional receptor modeling. *Environmetrics*, **12**, 451–467.

Bouteiller C, Perrin N (2000) Individual reproductive success and effective population size in the greater white-toothed shrew *Crocidura russula*. *Proceedings of the Royal Society B: Biological Sciences*, **267**, 701–705.

Brooks SP (2001) On Bayesian analyses and finite mixtures for proportions. *Statistics and Computing*, **11**, 179–190.

Churchfield S (1990) *The Natural History of Shrews*. Cornell University Press, New York.

Duarte LC, Bouteiller C, Fontanillas P, Petit E, Perrin N (2003) Inbreeding in the greater white-toothed shrew, *Crocidura russula*. *Evolution*, **57**, 638–645.

Ehinger M, Fontanillas P, Petit E, Perrin N (2002) Mitochondrial DNA variation along an altitudinal gradient in the greater white-toothed shrew, *Crocidura russula*. *Molecular Ecology*, **11**, 939–945.

Erwing G, Nicholls G, Rodrigo A (2004) Using temporally spaced sequences to simultaneously estimate migration rates, mutation rate and population sizes in measurably evolving populations. *Genetics*, **168**, 2407–2420.

Faubet P, Waples RS, Gaggiotti OE (2007) Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. *Molecular Ecology*, **16**, 1149–1166.

Favre L, Balloux F, Goudet J, Perrin N (1997) Female biased dispersal in the monogamous mammal *Crocidura russula*: evidence from field data and microsatellite patterns. *Proceedings of the Royal Society B: Biological Sciences*, **264**, 127–132.

Fontanillas P, Petit E, Perrin N (2004) Estimating sex-specific dispersal rates with autosomal markers in hierarchically structured populations. *Evolution*, **58**, 886–894.

Frankham R, Ballou JD, Briscoe DA (2002) *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge, UK.

Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–511.

Gelman A, Carlin JB, Stearn HS, Rubin DB (2004) *Bayesian Data Analysis*, 2nd edn. Chapman & Hall/CRC, New York.

Genoud M, Perrin N (1994) Fecundity versus offspring size in the greater white-toothed shrew, *Crocidura russula*. *Journal of Animal Ecology*, **63**, 328–336.

Ghosh JK, Delampady M, Samanta T (2006) *An Introduction to Bayesian Analysis*. Springer, New York.

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.

Hirzel AH, Jaquiéry J (2007) *Ecogenetics Computer Program*. University of Lausanne, Lausanne, Switzerland. URL: http://www.unil.ch/biomapper/ecogenetics/.

Jain S, Neal RM (2004) A split–merge Markov chain Monte Carlo procedure for the Dirichlet process mixture model. *Journal of Computational and Graphical Statistics*, **13**, 158–182.

Kass RE, Wasserman L (1996) The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, **91**, 1343–1370.

Koenig W, Van Vuren D, Hooge P (1996) Detectability, philopatry, and the distribution of dispersal distances in vertebrates. *Trends in Ecology & Evolution*, **11**, 514–517.

Lande R (1993) Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *The American Naturalist*, **142**, 911–927.

Manel S, Gaggiotti OE, Waples RS (2005) Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology & Evolution*, **20**, 136–142.

Neal RM (2000) Markov chain sampling methods for Dirichlet process mixture. *Journal of Computational and Graphical Statistics*, **9**, 249–265.

Neigel JE (1997) A comparison of alternative strategies for estimating gene flow from genetic markers. *Annual Review of Ecology and Systematics*, **28**, 105–128.

Paetkau D, Slade R, Burdens M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*, **13**, 55–65.

Pella J, Masuda M (2001) Bayesian methods for analysis of stock mixtures from genetic characters. *Fisheries Bulletin*, **99**, 151–167.

Plummer M, Best N, Cowles K, Vines K (2006) CODA: convergence diagnosis and output analysis for MCMC. *R News*, **6**, 7–11. http://CRAN.R-project.org/doc.

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

R Development Core Team (2005) *R: A Language and Environment for Statistical Computing*. 2.2.1 edn. R Foundation for Statistical Computing, Vienna, Austria.

Reynolds JH, Tremplin WD (2004) Detecting specific populations in mixtures. *Environmental Biology of Fishes*, **69**, 233–243.

Rousset F (2001) Genetic approaches to the estimation of dispersal rates. In: *Dispersal* (eds Clobert J, Danchin E, Dhondt AA, Nichols JD), pp. 18–28. Oxford University Press, New York.

Slatkin M (2005) Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. *Molecular Ecology*, **14**, 67–73.

Slatkin M, Barton NH (1989) A comparison of three indirect methods for estimating average levels of gene flow. *Evolution*, **43**, 1349–1368.

Sork VL, Nason J, Campbell DR, Fernandez JF (1999) Landscape approaches to historical and contemporary gene flow in plants. *Trends in Ecology & Evolution*, **14**, 219–224.

Thomas A, Hara BO, Ligges U, Sturtz S (2006) Making BUGS open. *R News*, **6**, 12–17.

Vitalis R (2002) Sex-specific genetic differentiation and coalescence times: estimating sex-biased dispersal rates. *Molecular Ecology*, **11**, 125–138.

Vitalis R, Couvet D (2001) Estimation of effective population size and migration rate from one- and two-locus identity measures. *Genetics*, **157**, 911–925.

Vogel P (1999) Colonisation capacity of the greater white-toothed shrew *Crocidura russula*: an experimental study. *Säugetierkundliche Mitteilungen*, **44**, 37–47.

Wang J, Whitlock MC (2003) Estimating effective population size and migration rates from genetic samples over space and time. *Genetics*, **163**, 429–446.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration $F_{ST} \neq 1/(4Nm + 1)$. *Heredity*, **82**, 117–125.

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

Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.

Thomas Broquet is a post-doc interested in the interactions between life-history traits and the dynamics of neutral genetic variance, with particular emphasis on the genetic aspects of dispersal. Jonathan Yearsley is interested in the invasion dynamics of genes and species in heterogeneous environments where population sizes or competition coefficients vary spatially. He is also interested in how habitat heterogeneity affects the maintenance of genetic variation. Alexandre Hirzel is interested in modelling various aspects of spatial ecology, from habitat suitability mapping to metapopulation processes and landscape connectivity. He has developed several computer programs along these lines (among which ecogenetics was used for the present study). Jérôme Goudet is a population geneticist interested in the consequences of population structure and dynamics for neutral and adaptive traits. Nicolas Perrin is interested in the ecology and evolution of dispersal, with a focus on sex-specific patterns.