Likelihood-Based Estimation of the Effective Population Size Using Temporal Changes in Allele Frequencies: A Genealogical Approach

Pierre Berthier,*,†,[1] Mark A. Beaumont,‡ Jean-Marie Cornuet§ and Gordon Luikart*

*Laboratoire de Biologie des Populations d'Altitude, UMR CNRS 5553, Université Joseph Fourier, F38041 BP53 Cedex 9 Grenoble, France, †Computational and Molecular Population Genetics Laboratory, Zoologisches Institut, Universität Bern, 3012 Bern, Switzerland, ‡School of Animal and Microbial Sciences, University of Reading, Reading RG6 6AJ, United Kingdom and §Laboratoire de Modélisation et de Biologie Évolutive, INRA-URLB, 34090 Montpellier, France

Manuscript received October 15, 2000
Accepted for publication October 19, 2001

ABSTRACT

A new genetic estimator of the effective population size (N_e) is introduced. This likelihood-based (LB) estimator uses two temporally spaced genetic samples of individuals from a population. We compared its performance to that of the classical *F*-statistic-based N_e estimator ($\widehat{N_{eFk}}$) by using data from simulated populations with known N_e and real populations. The new likelihood-based estimator ($\widehat{N_{eLB}}$) showed narrower credible intervals and greater accuracy than ($\widehat{N_{eFk}}$) when genetic drift was strong, but performed only slightly better when genetic drift was relatively weak. When drift was strong (*e.g.*, $N_e = 20$ for five generations), as few as ~10 loci (heterozygosity of 0.6; samples of 30 individuals) are sufficient to consistently achieve credible intervals with an upper limit <50 using the LB method. In contrast, ~20 loci are required for the same precision when using the classical *F*-statistic approach. The $\widehat{N_{eLB}}$ estimator is much improved over the classical method when there are many rare alleles. It will be especially useful in conservation biology because it less often overestimates N_e than does $\widehat{N_{eFk}}$ and thus is less likely to erroneously suggest that a population is large and has a low extinction risk.

THE effective size of a population can be defined as the size of an ideal population (Wright-Fisher model) in which genetic drift occurs at the same rate as in the studied population (Wright 1931). The effective population size (N_e) is an important parameter in evolution and conservation biology because it influences the amount of genetic drift in populations. Genetic drift influences the rate of loss of genetic diversity, the rate of fixation of deleterious alleles, and the efficiency of natural selection at maintaining beneficial alleles.

Unfortunately, N_e is notoriously difficult to estimate by demographic methods in the field (Frankham 1995), because it requires data such as variance in lifetime reproductive success, which is difficult to obtain for many wild populations. Genetic methods of estimating N_e are becoming more widely used because of the increasing availability of polymorphic molecular markers (for reviews, see Waples 1991; Schwartz *et al.* 1998). However, a serious problem with existing genetic estimators of N_e is their poor precision; *e.g.*, their confidence intervals often include infinity (Hill 1981; Waples 1991; Luikart and Cornuet 1999; Luikart *et al.* 1999).

The most widely used genetic method consists of measuring the variance of allele frequencies between generations to estimate the "variance effective size," N_{eV} (Robeards *et al.* 1991; Hedgecock *et al.* 1992; Husband and Barrett 1992; Taylor *et al.* 1993; Burczyk 1996; Jorde and Ryman 1996; Miller and Kapuscinski 1997; Saadreva 1997; Sitnikov *et al.* 1997; Laikre *et al.* 1998; Planes and Lecaillon 1998; Tarr *et al.* 1998; Fiumera *et al.* 1999; Funk *et al.* 1999; Kantanen *et al.* 1999). Given two genetic samples from one population, spaced by a known number of generations, estimation of N_{eV} can be conducted by moment-based methods (Waples 1989) and some recently published likelihood-based (LB) methods (Williamson and Slatkin 1999; Anderson *et al.* 2000).

LB estimators should provide better precision compared to moment-based estimators, because they use more of the information from the data (Edwards 1972). For example, the Williamson and Slatkin (1999) LB method showed better precision and accuracy compared to the *F*-statistic estimator of N_{eV} (Waples 1989). This initial study was restricted to diallelic loci because of the numerical complexity of using loci with a large number of alleles. However, Anderson *et al.* (2000) have developed an "importance sampling" approach that enables multiallelic loci to be analyzed. This approach is based on analysis of the Wright-Fisher model, in which the gene frequencies in the entire population in each generation are considered. We propose here a coalescent-based approach, which has different properties compared to that of Anderson *et al.* (2000). In many mating systems and life histories, in-

[1] *Corresponding author:* Computational and Molecular Population Genetics Laboratory, Zoologisches Institut, Universität Bern, Baltzerstrasse 6, 3012 Bern, Switzerland.
E-mail: pierre.berthier@zoo.unibe.ch

cluding that of the Wright-Fisher model, the gene genealogy tends to the coalescent as the population size becomes large (Donnelly and Tavaré 1995; Mohle 2000). Thus the coalescent-based approach will give very similar answers to that of Anderson *et al.* (2000) when the population sizes are large. Since the method deals only with the genealogy and not with the population gene frequencies it is potentially more efficient than the method of Anderson *et al.* for larger population sizes. In addition, in populations with low N_e that do not follow the Wright-Fisher model, the gene genealogy might be better approximated by the coalescent. In this article we demonstrate that, in fact, the method performs satisfactorily with data generated from a Wright-Fisher model with small N_e sampled over a single generation.

The objectives of this article are twofold: (i) to present a new likelihood-based estimator of N_e, on the basis of coalescent theory, which allows the use of multiallelic markers and can be extended to incorporate Bayesian prior information about the N_e value (*e.g.*, knowledge that $N_e < 500$) and (ii) to evaluate the accuracy and precision of this method in comparison to the classical estimator based on *F*-statistics (Krimbas and Tsakas 1971; Nei and Tajima 1981; Pollak 1983; Waples 1989), which uses the same genetic data from temporally spaced samples. Our evaluation is conducted empirically, using data from simulated populations with known N_e and from real populations. The model used to simulate populations is individual based, which provides realistic samples, complementary to the real data sets, for evaluating the usefulness of the estimators in natural populations.

The biggest problem with existing methods is their large confidence intervals (Waples 1991; Luikart *et al.* 1999). For example, the χ^2 approximation method (used for $\widehat{N_{eFk}}$) is known to be slightly too conservative (*i.e.*, confidence intervals are too wide; Waples 1989). Thus it is worth comparing the performance of the new method that we introduce to the widely used, "classical" method. We tested whether the precision of our LB method can be improved by incorporating Bayesian prior knowledge about N_e, *e.g.*, by setting N_{eMAX} (the maximum possible effective size) to 500 or 5000. Like numerous LB estimators, the method we present is computationally intensive. On a Pentium II, 400 MHz PC, it requires 2–10 hr to estimate N_e for one data set with 10–20 loci and samples of 30 individuals. The program is slower with more loci or individuals. The slow speed is mainly of concern for studies like ours involving thousands of population replicates (typical users can simply run the program overnight to get a single estimation). The speed compares favorably with that reported by Anderson *et al.* (2000). Because of the slow speed, we could not evaluate many scenarios [*i.e.*, combinations of N_e, sample size (S), time between the two samples (T), number of loci, and allele frequency distributions]. We tried to focus on realistic scenarios having the following characteristics: (i) multiallelic loci [*e.g.*, five alleles with heterozygosity (H) of 0.6] because they provide more precise and accurate estimations (Luikart *et al.* 1999) and are becoming increasingly available for natural populations (*e.g.*, microsatellites); (ii) sets of 5–20 unlinked loci, typical of most field studies; and (iii) samples of 30 or 60 individuals separated by one or five generations (*i.e.*, one or five episodes of genetic drift). However, we also conducted some evaluations with different allele frequencies (five alleles with equal frequencies, $H = 0.8$) and diallelic loci ($H = 0.2$) that more closely correspond to allozyme loci or single nucleotide polymorphisms (SNPs).

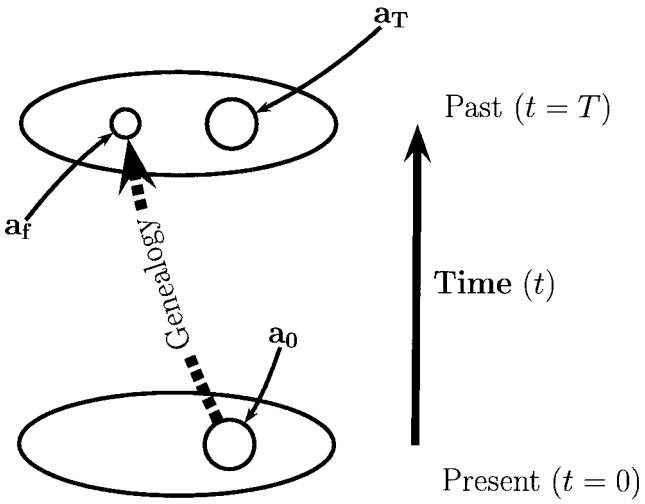

Figure 1.—Model used by the new likelihood-based estimator. We assume that samples are taken from a closed population at two different times. The population is represented by the dotted arrow. Following the conventions of genealogical modeling we take time to be increasing into the past, with the most recent sample ($\mathbf{a_0}$) taken at time $t = 0$ and the earlier sample ($\mathbf{a_T}$) taken at time $t = T$. $\mathbf{a_0}$ is assumed to have a genealogy, described by the standard coalescent model, and $\mathbf{a_f}$ represents its founders.

In materials and methods, we describe (i) the new likelihood-based estimator for N_e (noted $\widehat{N_{eLB}}$), (ii) the classical *F*-statistic-based estimator (noted $\widehat{N_{eFk}}$), (iii) the model used to simulate data sets, (iv) the real data sets, and (v) the methods we used to assess the relative performance of both estimators.

MATERIALS AND METHODS

Likelihood-based estimator: The method used to estimate likelihoods for N_e given the data is based on the genealogical approach described in O'Ryan *et al.* (1998) and Beaumont and Bruford (1999), which is very similar to that described in Nielsen *et al.* (1998) and Saccheri *et al.* (1999; see also Ciofi *et al.* 1999; Chikhi *et al.* 2001). The model is illustrated in Figure 1. We assume that samples are taken from a closed population at two different times. In principle the method can be easily extended to deal with samples taken at many times, but only a pair of samples is considered here. Following the conventions of genealogical modeling we take time to be increasing into the past, with the most recent sample taken

at time $t = 0$ and the earlier sample taken at time $t = T$. For a particular locus, the two samples at times $t = T$ and $t = 0$ consist of two vectors of counts, $\mathbf{a_T}$ and $\mathbf{a_0}$, of n_T and n_0 chromosomes distributed among k distinct alleles. The number of distinct alleles is taken to be that observed in both samples combined; each individual sample may have $<k$ alleles, and this number is denoted k_0 and k_T for the samples taken at times 0 and T. The samples are assumed to be sampled independently with probabilities $p(\mathbf{a_T}|n_T, k, \mathbf{x})$ and $p(\mathbf{a_0}|T, N_e, n_0, k, \mathbf{x})$, where N_e is the effective population size and $\mathbf{x}$ is the parametric gene frequency at time T. The samples at each of m loci are also assumed to be sampled independently and therefore an overall likelihood $\mathcal{L}(T, N_e, \mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_m})$ can be obtained by multiplying the probabilities across samples and across loci. Metropolis-Hastings simulation is then used to integrate out the $\mathbf{x_i}$ as described below.

Probability of the first sample: It is assumed that the n_T chromosomes are sampled with replacement from a population having unknown gene frequency $\mathbf{x}$. Hence $\mathbf{a_T}$ follows the multinomial distribution with probability $p(\mathbf{a_T}|n_T, k, \mathbf{x})$.

Probability of the second sample: The sample at $t = 0$ is assumed to have a genealogy, described by the standard coalescent model, with n_0 lineages at $t = 0$, n_c coalescent events between $t = 0$ and $t = T$, and $n_f = n_0 - n_c$ lineages at $t = T$. Since mutations are assumed not to occur, $n_f \geq k_0$. Let $\mathbf{a_f}$ be the vector of counts of the n_f distinct lineages that remain at $t = T$, distributed among the k allelic classes. Those lineages are sampled with replacement from the population with probability $p(\mathbf{a_f}|n_f, \mathbf{x})$. It is possible to calculate the probability of obtaining n_c coalescent events, $p(n_c|T, N_e, n_0)$ (Tavaré 1984), which depends on the effective population size, N_e. In the case of fluctuating populations, N_e is the harmonic mean effective population size over the interval T as discussed in O'Ryan *et al.* (1998). The distribution of coalescent events in an interval, and hence the joint likelihood of sample configurations, is identical between all models of population change, including stable populations, as long as N_e is measured as the harmonic mean N_e over the interval (Marjoram and Donnelly 1997).

Given the configuration in the founder lineages, $\mathbf{a_f}$, it is possible to calculate the probability of obtaining the configuration in the sample, $p(\mathbf{a_0}|\mathbf{a_f}, n_0, n_c)$, by n_c successive iterations of choosing a lineage uniformly at random and duplicating it (Slatkin 1996). Thus by enumerating all possible $\mathbf{a_f}$ for each value of $n_f = n_0 - n_c$, $n_c = 0 \ldots n_0 - k_0$ it is possible to calculate the probability of obtaining the sample configuration at time 0 as the sum

$$p(\mathbf{a_0}|T, N_e, n_0, k, \mathbf{x}) = \sum_{\mathbf{a_f}, n_c} p(\mathbf{a_0}|\mathbf{a_f}, n_0, n_c) p(\mathbf{a_f}|\mathbf{x}, n_f = n_0 - n_c) \times p(n_c|N_e, T, n_0). \quad (1)$$

Although it is feasible to evaluate this sum for small samples, it is impractical for typical cases. We use here an alternative, efficient method of evaluation described by Griffiths and Tavaré (1994) and used by O'Ryan *et al.* (1998). For a more detailed discussion of the method see Griffiths and Tavaré (1994), Felsenstein *et al.* (1999), and Stephens and Donnelly (2000).

We are interested in estimating $p(\mathbf{a_0}|T, N_e, n_0, k, \mathbf{x})$. Rewriting (1) in a more general way,

$$p(\mathbf{a_0}|T, N_e, n_0, k, \mathbf{x}) = \sum_{G} p(\mathbf{a_0}|G) p(G|\mathbf{a_f}, n_c) p(\mathbf{a_f}|\mathbf{x}, n_f = n_0 - n_c) \times p(n_c|T, N_e, n_0). \quad (2)$$

The summation is over all genealogical histories G, each of which can be represented as a sequence of configurations $G = (\mathbf{g_0}, \mathbf{g_1}, \ldots, \mathbf{g_{nc}})$, where, looking back in time from $t = 0$, $\mathbf{g_0}$ is the configuration prior to the first coalescent event and $\mathbf{g_{nc}}$ is the configuration prior to the sampling event at T (*i.e.*, $\mathbf{g_{nc}} = \mathbf{a_f}$). The states $(\mathbf{g_{nc}}, \ldots, \mathbf{g_1}, \mathbf{g_0})$ form a Markov chain where the number of lineages increases by 1 at each step. Let $\mathbf{b}$ denote a vector of length k with 1 at position b and 0 elsewhere, b being a particular allelic class. In the following, $g_r^{(b)}$ denotes the number of genes in the allelic class b within the rth state. The rth state in the sequence can be generated from

$$\Pr(\mathbf{g_{(r)}} = \mathbf{g_{(r+1)}} + \mathbf{b}|\mathbf{g_{(r+1)}}) = \frac{g_{(r+1)}^{(b)}}{n_{(r+1)}},$$

where $n_{(r+1)} = \Sigma_j g_{(r+1)}^{(j)}$. Thus $p(G|\mathbf{a_f}, n_c)$ is the product of these probabilities over n_c coalescent events, and $p(\mathbf{a_0}|G) = 1$ if $\mathbf{g_0} = \mathbf{a_0}$, 0 otherwise.

Equation 2 has the general form $p(x) = \Sigma_y p(x|y) p(y)$, which can be estimated by the classical Monte Carlo method as $1/s\Sigma_{i=1}^{s} p(x|y_i)$, where the y_i are drawn from $p(y)$ (see, *e.g.*, Tanner 1993). For example, we could simulate the number of coalescent events to obtain n_f, simulate a sample from $\mathbf{x}$, and simulate G. Unfortunately $p(\mathbf{a_0}|G) = 0$ for most G. An alternative approach is to use importance sampling (see, *e.g.*, Tanner 1993), where $\Sigma_y p(x|y) p(y)$ is estimated as $1/s\Sigma_{i=1}^{s} p(x|y_i) p(y_i)/p^*(y_i)$, where $p^*(y)$ is chosen to be as close as possible to $p(y|x)$; the variance due to importance sampling would be zero if it could be made equal to $p(y|x)$—this follows because $p(x|y) p(y)/p(y|x) = p(y|x) p(x)/p(y|x) = p(x)$ for any draw from $p(y|x)$. In general it is not possible to draw exactly from $p(y|x)$ but this can be approximated quite closely. Following the method of Griffiths and Tavaré (1994), we can simulate a sample sequence with known probability as follows. For a configuration at event r, $\mathbf{g_r}$, there are m prior configurations with one lineage less in each of the allelic classes, where m is the number of allelic classes in which there is at least one lineage. The configuration $\mathbf{g_{r-1}}$ with one lineage less in allelic class b is chosen with probability

$$\frac{(g_r^{(b)} - 1)/(n_r - 1)}{\sum_{j=1}^{m}((g_r^{(j)} - 1)/(n_r - 1))} = \frac{g_r^{(b)} - 1}{n_r - m}, \quad (3)$$

where j subscripts are the allelic classes represented in $\mathbf{g_r}$. The ratio $p(G|\mathbf{a_f}, n_c)/p^*(G)$ is then given by the product of

$$\frac{g_r^{(b)} - 1}{n_r - 1} \Big/ \frac{g_r^{(b)} - 1}{n_r - m} = \frac{n_r - m}{n_r - 1} \quad (4)$$

over all coalescent events in G. Thus to estimate (2) we can use the following algorithm, which calculates $1/s \times S$, with $S = \Sigma_{i=1}^{s} p(\mathbf{a_0}|G_i) \times P_i \times p(\mathbf{a_f}|\mathbf{x}, n_f) p(n_c|T, N_e, n_0)$, and $P_i = p(G_i)/p^*(G_i)$. This algorithm involves two nested iterations: a first iteration, which builds the summation S over sampled genealogies, and another iteration, nested in the previous one, which calculates each P_i by sampling the number of coalescent events in the genealogy G_i. Let n_r be the number of lineages, m the number of allelic categories in the current configuration, and τ the total time elapsed since the start of the building of the current genealogy. The genealogies are built starting at $\tau = 0$, with the data configuration $\mathbf{a_0}$, so that for all i, $p(\mathbf{a_0}|G_i) = 1$. Then coalescent events are simulated using a waiting time, t, until the total time elapsed becomes $> T/N_e$. In the following algorithm, t is a random variable, sampled from an exponential distribution with scale $\binom{n_r}{2}$, which changes each time we refer to it:

1. set S to 0;
2. do s times:
   (a) set P_i to 1; set n_r to n_0; set τ to t;
   (b) while ($\tau \leq T/N_e$ and $n_r \neq m$) do:
      i. with probability given by Equation 3, choose an

allelic category in the current configuration and decrease the count by 1. Decrease n_r by 1;
ii. multiply P_i by the importance weight (Equation 4);
iii. set τ to $\tau + t$;
(c) if ($n_r = m$ and $\tau \leq T/N_e$) then
set $P_i = 0$ (*probability of obtaining the data is zero with this number of coalescences*) else
i. let $\mathbf{a}_{fi}$ be the current configuration (*i.e.*, allele counts among founder lineages), containing n_{fi} genes;
ii. add to S the product of P_i by $p(\mathbf{a}_{fi}|\mathbf{x}, n_{fi})$, the multinomial probability of choosing the current configuration from $\mathbf{x}$;
3. evaluate S/s.

Thus the method estimates $p(\mathbf{a_0}|T, N_e, n_0, k, \mathbf{x})$ with some error depending on the number of iterations, and the number we actually used is discussed below.

Posterior distribution: Assuming independence the probabilities can be multiplied across samples and loci as described above to give a likelihood $\mathcal{L}(N_e, T, \mathbf{x_1}, \ldots, \mathbf{x_m})$. We integrate out the $\mathbf{x_i}$ using Metropolis-Hastings sampling, following the approach described in O'Ryan *et al.* (1998).

In Metropolis-Hastings sampling we propose candidate values x' from some distribution conditional on the current value x, $p(x'|x)$ and calculate the ratio

$$X = \frac{\mathcal{L}(x')/p(x'|x)}{\mathcal{L}(x)/p(x|x')}, \quad (5)$$

which gives the likelihood of a candidate value, weighted by the probability of choosing it from the current value, relative to the current value, weighted by the probability of choosing it from the candidate value. If $X > 1$ we accept x'; otherwise we accept it with probability X, retaining x otherwise. The likelihoods include a weighting by priors. Provided certain conditions hold (see Tanner 1993), the resulting sequence of accepted and retained values has a stationary distribution that is $\mathcal{L}(x)/\int\mathcal{L}(x)\,dx$. In the case of jointly distributed variables marginal distributions can be estimated by looking at only one variable (*e.g.*, T/N_e here). Since T is known, using this approach we can estimate the posterior distribution $p(N_e|\mathbf{a_0}, \mathbf{a_T}, T)$, which will be proportional to the likelihood for N_e if a uniform prior for N_e is assumed. We need to specify the upper limit, N_{eMAX}, because convergence of the MCMC simulation is not otherwise guaranteed. Therefore our approach is more accurately described as Bayesian, with an informative rectangular prior on N_e. Here we assume rectangular priors on N_e between zero and some upper limit (N_{eMAX}—usually 500) and uniform Dirichlet, $\mathcal{D}(1, \ldots 1)$, priors on $\mathbf{x}$. However, since the tests of power and precision given in this article are inherently non-Bayesian (since the answer is known), we use the term "likelihood-based" to describe the general approach and restrict the use of the term Bayesian when comparing the effect on inference of different values of N_{eMAX}. When real data are analyzed, it is clearly sensible to take a fully Bayesian approach, with a prior that reflects background information (which is then unlikely to be rectangular).

Candidate values for N_e are proposed, using a lognormal distribution centered around the current value. Candidate values for $\mathbf{x}$ for each locus are proposed by randomly partioning the alleles into two groups and using a beta-distribution, as described in Ciofi *et al.* (1999) (although the Dirichlet method described in O'Ryan *et al.* 1998 also works well), and the likelihood $\mathcal{L}(N_e, T, \mathbf{x_1}, \ldots, \mathbf{x_m})$ is estimated as described above. The performance of Metropolis-Hastings sampling when the likelihood is estimated with error has not been intensively studied. As in O'Ryan *et al.* (1998) and Ciofi *et al.* (1999), we estimated $p(\mathbf{a_0}|N_e, T, n_0, k, \mathbf{x})$ for each locus, using 500 importance samples, and reestimated this for current and candidate values at each iteration of the Metropolis-Hastings algorithm. The number of 500, although giving appreciable error in estimation of $\mathcal{L}(N_e, T, \mathbf{x_1}, \ldots, \mathbf{x_m})$, had been suggested by pilot simulations in O'Ryan *et al.* (1998) in which it gave indistinguishable results from simulations with 10,000 iterations for subsets of the samples used in that study, and these latter simulations gave indistinguishable results (using smaller data sets) from Metropolis-Hastings sampling using exact likelihoods calculated according to Equation 1. We have not yet determined a lower limit of the number of iterations that will give reasonable performance. It should be noted that X (Equation 5) in the Metropolis-Hastings simulations is the ratio of importance weights and it is straightforward to show that the ratio of estimates using exactly one iteration of the Griffiths and Tavaré procedure is a valid acceptance criterion (in which case we are integrating over G as well). We have investigated this case and find that simulations using one iteration, however, have a very low acceptance rate.

TABLE 1

Data sets used to test for the convergence of the Metropolis-Hastings procedure embedded in N_{eLB}

True N_e	T	S	L	Gelman-Rubin criterion
10	1	30	10	(1.01, 1.02)
50	1	30	20	(1.01, 1.03)
20	5	30	10	(1.01, 1.04)
50	5	30	20	(1.04, 1.09)

All data sets use loci having five alleles each ($H = 0.6$). Results for the Gelman-Rubin criterion are presented in parentheses (value of statistic, upper 97.5% credible limit). T, number of generations between samples; S, sample size in individuals; L, number of loci used.

Convergence of the Metropolis-Hastings procedure: For the data sets described here, the Metropolis-Hastings procedure was run for single chains of 20,000 iterations, sampling every 5 iterations. The validity of this number was assessed for four representative data sets, by running five independent chains from different starting points for each set, and using the Gelman-Rubin criterion to assess convergence, implemented in CODA (Best *et al.* 1995). The Gelman-Rubin statistic takes the values from the last one-half of the sampled points from the chains and estimates the square root of the ratio of the variance in N_e when the chains are combined to the average of the variance within each chain. The idea is that initially, with independent starting points, the variance across chains will be substantially greater than the variance within, reflecting the different starting points. However, when the chains are run long enough the variances should converge and the ratio should tend to 1. Gelman (1996) suggests as a "rule of thumb" that the value of the ratio of variances should be $<$1.1–1.2 (*i.e.*, value of the statistic $<$1.05–1.1). The method also calculates the upper 97.5% credible limit for the statistic. The four data sets used and the results obtained for the Gelman-Rubin criterion are presented in Table 1. These results show that we can be reasonably confident that error due to Markov chain Monte Carlo (MCMC) estimation is a small fraction of the variability in estimation of N_e across data sets.

Point estimates and confidence intervals: From the output of the MCMC simulations the following summary statistics are estimated: the mode and 0.05 and 0.95 quantiles (giving a 90% credible interval). The 0.05 and 0.95 quantiles are obtained from the ranked output in the standard way. The mode

is obtained by kernel density estimation. We used a logistic function as a kernel with a bandwidth (standard deviation) given by

$$\sigma = k\frac{0.5(q_2 - q_1)}{n^{0.2}},$$

where n is the sample size (the number of updates), q_2 is the value of the 0.625 quantile, and q_1 is the value of the 0.375 quantile. The constant, k, was set to be 1 for all the simulations. This formula was obtained by trial and error in pilot simulations and was set to tend toward under- rather than oversmoothing. A logistic kernel was used because it was easy to define truncated kernels at 0 and $N_{e\mathrm{MAX}}$, which reduced the degree of underestimation of the density at the boundaries. For appropriate bandwidths (slightly larger values of k) the method gives results very similar to those obtained using the program Locfit (Loader 1996) implemented in R (Ihaka and Gentleman 1996, http://www.r-project.org/), which is based on local-likelihood density estimation methods.

Many data sets may have little information on the upper limit of N_e (*i.e.*, the likelihood function tends to a nonzero constant for large N_e), in which case the 90% credible interval depends strongly on the prior used. We use a rectangular prior in the study here, but for real data a more smoothly varying prior may reflect background information more closely. This estimator was implemented in a computer program (TM3) available from http://www.rubic.rdg.ac.uk/~mab/software.html. A program for converting GENEPOP files to input format for the TM3 program is also available. The mode estimated from the MCMC output is referred to as the likelihood-based estimator $\widehat{N_{e\mathrm{LB}}}$.

***F*-statistic-based estimator:** We estimated N_{eF_k} for each population using the equation

$$\widehat{N_{eF_k}} = \frac{T}{2 \times (F - 1/2S_0 - 1/2S_t)} \quad (6)$$

(Nei and Tajima 1981), where T is the number of generations between the two samples and S_0 and S_T are sample sizes (of individuals) in the first and second sample, respectively. This equation is appropriate because the individuals sampled to estimate F are independent of the individuals reproducing to make the next generation (sampling scheme II of Waples 1989).

We estimated F as did Waples (1989) as the mean of F_k over loci. F_k was calculated for each locus from the equation

$$F_k = \frac{1}{A - 1} \cdot \sum_{i=1}^{A} \frac{(x_i - y_i)^2}{(x_i + y_i)/2} \quad (7)$$

(Krimbas and Tsakas 1971; Pollak 1983), where A is the number of alleles at the locus and x_i and y_i are the frequencies of the ith allele in the first and second samples, respectively.

Other estimators of F give generally similar results to those obtained with F_k (Waples 1989; Richards and Leberg 1996; Luikart *et al.* 1999). Confidence intervals (90%) for N_e were computed using a χ^2 approximation, known to be unbiased, but too conservative (Waples 1989; Luikart *et al.* 1999).

Model for simulating data sets: The simulated populations were generated using an individual-based model with Mendelian inheritance. Populations were initiated by randomly sampling alleles from independent loci (defined by an array of allelic frequencies; see Table 2). Simulations were based on a Wright-Fisher model, with two modifications: (i) separate sexes (with an equal sex ratio) and (ii) strict allogamy (no selfing). These modifications should lead to an N_e slightly larger than the actual number of breeders (N_b, individuals), as shown in Caballero (1994, Equation 16). We verified this fact by estimating the variance of reproductive success, S_k^2 by $V(k_i)$, which we found to be ~1.91 (instead of 2) for $N_b = 10$. In this case, $N_e \sim 10.2$. This difference between N_e and N_b is negligible and is even smaller when $N_b > 10$. Individuals were sampled under sampling scheme II of Nei and Tajima (1981; Pollak 1983; Waples 1989), in which, at given generation, the sample and the $2N_e$ gametes representing the effective size are both independent binomial samples from the pool of gametes of the preceding generation.

We studied only cases with relatively small N_e because it is known that estimators of N_e give reasonably small confidence intervals when N_e is large and when using realistic sample sizes (*e.g.*, 10–20 loci and 30–60 individuals; Luikart *et al.* 1999; Waples 2000). Because the N_e is usually $<N$ (the census size) in natural populations (see Frankham 1995; Schwartz *et al.* 1998), it is realistic to model a population where N_e is only 10 or 20 and to sample 30 or more individuals. This model was implemented in a computer program available from pierre.berthier@zoo.unibe.ch.

Real data: To test further the N_e estimators, we applied them to real data, which potentially follow a model very different from the ones used both by the estimators and by our simulation model. For example, real data can show migration (killifishes), selection and linkage among loci (Drosophila; see discussion), overlapping generations (otter), and nonstable population size (mosquito fish). Our purpose here is not to use these data sets to quantify the problems introduced by those violations of the model underlying the N_e estimators, but rather to illustrate that the violation of assumptions can lead to changes in the estimators' performance. We used microsatellite data from four animal populations:

TABLE 2

Allele frequency arrays used for the initiation of each locus for simulated data sets

		Allele						
Name	H	1	2	3	4	5	6	7
A	0.6	0.2	0.59	0.1	0.07	0.04	—	—
B	0.8	0.2	0.2	0.2	0.2	0.2	—	—
C	0.2	0.885	0.115	—	—	—	—	—
D	0.8	0.35	0.16	0.15	0.11	0.10	0.07	0.06
E	0.4	0.76	0.14	0.1	—	—	—	—
F[a]	0.6							

Arrays are given a name for further reference.

[a] Data sets with variation of H among loci as follows: A, one locus; D, two loci; and E, two loci ($\overline{H} = 0.6$).

TABLE 3

Effective population size estimates and confidence intervals for empirical data sets

					F-statistic based		Likelihood based	
Populations	True N_e	S	T	L	N_{eF_k}	χ^2	$N_{e_{LB}}$	Percentiles
Killifish	?	52	3	12	17	5–51	14.2	9.7–26.7
Drosophila								
100a	100.8	36	67	7	191.7	100.5–320.0	83.3	66.39–128.7
1spm24	18.8	18	11	7	44.0	21.6–80.7	11.8	7.9–17.6
Otter	?	17; 28	1	12	15.6	4.1–∞	15.6	9.6–452.8
Mosquito fish								
A5	2	40	2	8	12.4	9.4–16.1	2.6	2.4–3.5
B4	16	40	2	8	35.4	24.6–52.1	21.6	18–32.1

Empirical data sets used are *Aphanius iberus* (killifish), *Drosophila melanogaster*, *Lutra lutra* (otter), and *Gambusia affinis* (mosquito fish).True N_e was estimated with pedigrees for Drosophila populations (England 1998). For mosquito fish populations, true N_e was estimated to be the number of adult founders (see methods; Spencer *et al.* 2000). S, sample size (or S_0; S_1 when sizes of the two temporal samples are different); T, number of generations between samples; L, number of microsatellite loci used.

i. Spanish killifish *Aphanius iberus*: A captive population was founded in 1994 using 83 individuals from a wild population and a subsequent incorporation of 68 new individuals occurred 1 year later from the wild into the captive population (S. Schönhuth, unpublished data). The first sample was taken among the 83 founders (originating from the wild population) and the second sample was taken from the captive population after the incorporation of the 68 new individuals.

ii. Two *Drosophila melanogaster* populations (1spm24 and 100a), which originated from the wild (at the Tyrrells Winery near Sydney, Australia), were genotyped (first sample); submitted to a bottleneck for a duration of 1 and 57 generations for, respectively, 1spm24 and 100a (see Table 3); and then genotyped a second time 10 generations later, during a rapid expansion period to a population size of 500 and 750 individuals, respectively. The N_e value given in Table 3 is the harmonic mean of the N_e for the generations between samples. The N_e for each generation is estimated from the pedigrees (England 1998).

iii. Eurasian otter *Lutra lutra* (Dallas *et al.* 1999), for which samples from 1983 to 1988 were pooled to represent one sample, as were samples from 1991 to 1997.

iv. Two western mosquito fish (*Gambusia affinis*) populations (A5 and B4), established experimentally by a bottleneck of one pair and eight pairs, respectively, from a wild population "source" (Spencer *et al.* 2000). For each of these two populations, the first temporal sample was taken in the source population. Experimental populations were then allowed to expand after the founder event and the second temporal sample was taken two to three generations later. Female founders were obtained from a general laboratory stock and may have not been virgin at the release time, so that the N_e at establishment could be underestimated by the actual number of founders.

Comparison of estimators: The first goal of this article is to describe a new estimator for N_e. The second goal is to provide an evaluation of its performance in comparison with a widely used estimator, so that researchers can use it with some confidence, on the basis of realistic examples. We take a frequentist approach and use simulations to compare the two methods. Although this is inconsistent with the Bayesian paradigm behind the new estimator, it is the most practicable way to compare the two approaches. There are two issues to consider here. Having an upper limit of $N_e = 500$ in the likelihood-based method will in itself increase the accuracy of the method. However, this effect is generally small, and cases where it has an effect are clear in the results and discussion. In addition, the credible intervals from the likelihood estimator are *a priori* not expected to have the same coverage properties as the confidence intervals from the *F*-statistic-based estimator—the former is a fixed interval, and the "truth" is a random variable, while the latter is a random interval, and the truth is fixed. However, since the assessment is frequency based (where the truth is fixed) it seems reasonable to assess the coverage properties of the two estimators, while accepting that they mean different things. It is also worth noting that a general result in classical theory is that the credible intervals and confidence intervals will converge to have the same coverage properties, providing that the posterior distribution is asymptotically normal (see, *e.g.*, Gelman *et al.* 1995, Chap. 4), and this is likely to be the case here with increasing numbers of loci (but not sample size).

In assessing the performance of each estimator we attempted to answer the following questions: (i) Are the point estimates biased?, (ii) how accurate are the two estimators?, (iii) how large can we expect the confidence intervals to be?, and (iv) how often do the confidence intervals contain the true value? We answer these questions by using simulated data sets and compare the performance of the estimators using summary statistics computed from the set of replicates obtained with each tested scenario. The parameters investigated include the true N_e value, the sample size (S), the number of loci (L), the allele frequencies in the population from which is taken the first sample (AF; Table 2), and the number of generations between samples (T). Each scenario (*i.e.*, parameter set) was evaluated using 200 populations simulated independently (*i.e.*, replicates).

The bias of the point estimates was investigated by reporting the median of the 200 point estimates from the simulations. The median was used rather than the mean because the distribution of point estimates can be strongly skewed (for example, in the case of the *F*-statistic-based estimator, there are occasionally point estimates of infinity). We also report the standard error of the mean. The accuracy of the point estimators was investigated by reporting the square root of the mean of the squared differences between the estimation values and the true value (the square root of the mean square error, $\sqrt{\text{MSE}}$).

The properties of the estimated confidence intervals (or credible intervals in the case of the likelihood-based estimator)

TABLE 4

N_e estimates and confidence intervals for data simulated with true N_e = 10, 20, or 50 for one or five generations (T)

		F-statistic based			Likelihood based		
L	AF	N_{eFk} (SE)	$\sqrt{\text{MSE}}$	Summary of CIs	$N_{e_{LB}}$ (SE)	$\sqrt{\text{MSE}}$	Summary of CIs
				N_e = 10, T = 1			
5	A	11.1 (500)	7071.6	2.8–∞	8.4 (2.5)	34.3	2.9–463
5[a]	A	10.0 (0.3)	4.4	3.0–49.4	7.3 (0.3)	4.2	2.8–57.7
10	A	9.7 (0.6)	8.4	3.9–49.1	7.8 (0.3)	4.3	3.6–130.6
10	C	11.2 (1209)	17323.5	2.9–∞	12.7 (6.6)	93.5	2.1–475.3
20	A	10.4 (0.2)	2.5	5.0–26.4	7.8 (0.1)	2.6	4.4–21.2
				N_e = 20, T = 5			
5	C	23.8 (1683)	24491.4	1.9–∞	19.3 (8.1)	119.9	3.6–478.2
5[a]	A	25.1 (0.6)	10.4	8.3–86.3	17.3 (0.5)	6.3	7.8–57.0
5	B	20.8 (0.6)	8.7	7.1–79.4	19.4 (0.6)	9.3	8.9–124.9
5[a]	B	19.6 (0.5)	6.7	6.7–66.8	18.3 (0.5)	6.8	8.4–62
5	A	26.3 (1.1)	17	7.6–124.4	18.1 (0.5)	7.7	6.3–105.6
10	A	26.9 (0.5)	10.6	11.3–74.7	18.4 (0.4)	5.4	9.7–53.9
20	A	25.9 (0.3)	7.8	14.0–52.0	17.7 (0.3)	3.9	10.6–35.3
5	F	24.5 (0.7)	12.3	7.4–103.2	18.8 (0.6)	8.1	6.9–117.3
				N_e = 50, T = 5			
10[a]	A	55.7 (1.3)	20	22.2–181.2	47.4 (1.0)	17.3	23.3–136.7
10	A	56.9 (2.6)	39.4	19.6–475.1	49.9 (2.1)	30.0	20.4–435.5
20	A	57.6 (1.6)	26.0	26.6–193.7	49.5 (1.3)	18.4	26.4–183.0

The point estimates distributions obtained with 200 replicates are described for each parameter set in the columns $N_{e_{F_k}}$ and $N_{e_{LB}}$ (*i.e.*, mode of the likelihood curve for $N_{e_{LB}}$) by the median of the 200 point estimates and the standard error (SE) in parentheses. The square root of the mean squared error is also given in the following column ($\sqrt{\text{MSE}}$) as a measure of accuracy. $N_{e_{LB}}$ estimates were computed using $N_{e_{MAX}}$ = 500. The spreads of the confidence and credible intervals (CIs) are summarized by the 5th percentile of the lower limits and the 95th percentile of the upper limits obtained from 200 simulation replicates. For example, *x*–*y* means 95% of the 200 lower limits independently obtained were above *x*, and 95% of the 200 upper limits independently obtained were below *y*. *L*, number of loci used; *T*, number of generations between samples; AF, allele frequencies array used to simulate populations (see Table 2).

[a] 60 individuals sampled (all other simulations have 30).

were assessed by reporting two types of summary statistic. First, to summarize the overall spread of the estimated intervals, we report the 5th percentile of the lower limits of the intervals (that is, 5% of the 200 confidence or credible intervals had lower bounds that were lower than this) and the corresponding 95th percentile of the upper limits. The interval defined by these two percentiles is a summary of the overall spread of confidence or credible intervals in the 200 simulations. Second, to summarize the coverage properties of the estimated intervals, we report the proportion of times the true N_e was below the lower limit and the proportion of times the true N_e was above the upper limit.

RESULTS AND DISCUSSION

Accuracy of estimates: The likelihood-based (LB) method appeared generally more accurate than the *F*-statistic-based method (Table 4). For example, in 13 out of 16 sets of simulations $\sqrt{\text{MSE}}$ was lower in the former than the latter ($P \sim 0.02$). The performance of the estimators seems rather variable and may partially depend on the base frequencies used. For example, two of the cases where $\sqrt{\text{MSE}}$ is higher in the LB method come from the only two sets of simulations that use frequency set B. There appear to be no other clear-cut effects on $\sqrt{\text{MSE}}$ in Table 4.

On the basis of the medians of the estimates from each set of 200 simulations, patterns in the bias of the estimates are discernible in Table 4. In general, $\widehat{N_{eFk}}$ appears to systematically overestimate N_e, whereas $\widehat{N_{eLB}}$ appears to slightly underestimate it. The overestimation by $\widehat{N_{eFk}}$ is due mainly to the loss of alleles in early generations of the time period between the two samples (see Richards and Leberg 1996; Luikart *et al.* 1999). These earlier observations suggest that the bias is greater with increasing drift and is also greater when there are many rare alleles, which are likely to be lost through drift (Waples 1989; Luikart *et al.* 1999). Table 4 supports the first observation: With N_e = 20 and T = 5, $\widehat{N_{eFk}}$ overestimates N_e by up to 25%, whereas it generally

overestimates by <10% in the other two sets of simulations. Interestingly, the degree of overestimation appears also to be affected by N_e or T, because the amount of drift is otherwise the same in these two sets of simulations. The Drosophila data in Table 3 support the second observation on the effect of rare alleles on $\widehat{N_{eFk}}$. Here $\widehat{N_{eFk}}$ substantially overestimates N_e, whereas $\widehat{N_{eLB}}$ does not ($\widehat{N_{eFk}}$ was 191.7 and $\widehat{N_{eFk}}$ was 83.3 when true N_e was 100.8).

The underestimation by $\widehat{N_{eLB}}$ is probably due to the violation of the assumptions of the coalescent in the simulated data sets and appears to be strongest when N_e is small. For example, looking at the simulations where $\sqrt{\text{MSE}}$ is at reasonable levels (*i.e.*, ignoring the three extreme cases), the underestimate is ~25% for $N_e = 10$, 10% for $N_e = 20$, and <5% for $N_e = 50$. Although the $\sqrt{\text{MSE}}$ is too variable for any clear patterns to be discerned, it is reasonable to assume that the reduction in bias will tend to increase the relative accuracy of $\widehat{N_{eLB}}$ over $\widehat{N_{eFk}}$.

Precision of estimates: Because the accuracy is generally good (see Table 4 and examples above), it is usually less problematic than the poor precision (see, for example, the N_e estimates for the otter data in Table 3). Thus, it is very informative to assess the relative performance of the methods used to compute confidence or credible intervals (CIs): the χ^2 approximation (Waples 1989) for the F_k-based estimator of N_e(N_{eFk}) and the 5–95th percentiles for the likelihood-based estimator of N_e(N_{eLB}).

The credible intervals for the likelihood-based estimator of N_e($\widehat{N_{eLB}}$) gave better precision than the *F*-statistic-based estimator ($\widehat{N_{eFk}}$) in the cases with high drift, whereas the difference in performance appeared variable (depending on N_e and on the number of loci) in the cases with relatively low drift. In the high drift cases, for example, when true $N_e = 20$ (and $T = 5$, $L = 10$, AF = A), the summaries of CIs were 9.7–53.9 for N_{eLB} and 11.3–74.7 for $\widehat{N_{eFk}}$ (Table 4). With real microsatellite data from a Drosophila population of known N_e ($N_e^{\text{pedigree}} = 18.8$), the confidence interval for $\widehat{N_{eLB}}$ (7.9–17.6) was also much smaller than that for $\widehat{N_{eFk}}$ (21.57–80.75; Table 3). We note that the N_e^{pedigree} was computed assuming no selection and no association among loci. However, selection is possible (*e.g.*, inbreeding depression, see Saccheri *et al.* 1998) and associations among loci are possible as Drosophila have only four chromosomes. Violations of these assumptions in the Drosophila population could decrease the actual N_e below 18.8 (computed from pedigrees).

When drift is weak, it is known that the lack of drift signal in allele frequency data leads to bad performance of the N_e estimators studied here (*e.g.*, Luikart *et al.* 1999). In those weak drift cases, $\widehat{N_{eFk}}$ gave more reliable confidence intervals than $\widehat{N_{eLB}}$, but both estimators are similar when at least ~20 loci are used (see Table 4). When considering the scenario with $N_e = 10$, $T = 1$, and 5 loci (first line in Table 4), it is noteworthy that N_{eLB} performed better when doubling the sample size (second line) than when doubling the number of loci used (third line). On the other hand, $\widehat{N_{eFk}}$ showed similar benefits from these two possible strategies for increasing the sampling effort. It is also important that the $\widehat{N_{eLB}}$ CIs performed slightly worse as N_e became smaller (from 50 to 20 to 10). This probably results from violation of the assumption in coalescent methods that N_e is large.

Because our simulations used a fixed heterozygosity (H) level for all loci, we investigated if a variance of H among loci changes precision of N_e estimates. We used a set of five loci having a mean $\overline{H} = 0.6$, described in Table 2, under the name F. The CIs obtained from populations simulated using variable interlocus H were only slightly different than those from populations having all loci with the same H (0.6; Table 4). Thus the results in Table 4 can be used as very rough guidelines for the number of loci and sample size needed to achieve reasonable precision.

Table 5 shows that the N_{eLB} CIs are more often too low than too high (they are biased low), whereas N_{eFk} CIs are more often too high than too low (they are biased high). The same pattern appears with real data (see Table 3); for example, with Drosophila popula-

TABLE 5

Percentage of confidence intervals lower (*L*) and higher (*H*) than the true N_e for 200 simulation estimates

		N_{eFk} CIs		N_{eLB} CIs	
L	AF	*L*	*H*	*L*	*H*
		$N_e = 10$, $T = 1$			
5	A	0.5	4.0	4.1	19.5
5[a]	A	3.0	0.5	16.6	3.5
10	A	0.5	4.0	9.6	4.2
10	C	1.5	3.0	3.2	41.1
20	A	1.0	2.5	16.6	0.0
		$N_e = 20$, $T = 5$			
5	C	1.5	6.5	1.0	25.8
5[a]	A	0.0	7.0	5.9	3.8
5	B	0.0	1.0	1.5	12.0
5[a]	B	3.0	1.0	5.5	6.5
5	A	1.2	5.9	8.0	7.0
10	A	0.0	12.0	3.0	7.0
20	A	0.0	15.0	12.0	2.0
5	F	0.6	3.8	6.5	7.0
		$N_e = 50$, $T = 5$			
10[a]	A	3.5	6.5	5.0	4.4
10	A	1.0	8.5	4.5	17.0
20	A	0.5	7.0	4.0	8.0

The data used here are the same as those in Table 4. *L*, number of loci used. *T*, number of generations between samples. AF, allele frequencies array used to simulate populations (see Table 2). CIs, confidence or credible intervals.

[a] 60 individuals sampled (all other simulations have 30).

tions, when the actual N_e was 18.8, the LB method credible interval was 7.9–17.6, whereas the *F*-statistic-based method CI was 21.6–80.7. This is interesting for conservation biology applications because overestimation is likely to delay the detection of a small N_e.

Guidelines: To improve precision, one has four possibilities: (i) Use more loci or loci with more alleles, as this provides more independent observations of drift; (ii) sample more individuals, as it gives better allele frequency estimates; (iii) increase the time between samples, as this increases the number of drift events observed (T); and (iv) use prior knowledge on N_e in the case of the likelihood-based method. The first three should increase the signal-to-noise ratio in the data (Waples 1991).

Typical users wishing to increase the precision of N_e estimators would like to know whether increasing the number of individuals sampled or the number of loci is more likely to help. Indeed, new promising molecular markers (*e.g.*, SNPs) may allow use of dozens of loci in the near future and could double or triple the number of loci typically used today (5–20). However, the cost and time needed for developing them might counterbalance the cost of sampling more individuals while keeping classical genetics markers. This is why performance analyses like ours are needed to help make such decisions about sampling investment. It is difficult to find dozens of unlinked loci in most real populations, and a linkage disequilibrium between loci is possible in small populations. Thus, more study is needed to quantify the potential problems caused by statistical association between loci on the N_e estimators we studied.

For improving precision, the total number of independent alleles ($\Sigma_i A_i - 1$, where A_i is the number of alleles at the locus i) is more important than the number of loci. For example, using five diallelic loci (thus 5 independent alleles) leads to *infinite CIs* with both N_{eF_k} and N_{eLB}, whereas using five loci with 5 alleles each (20 independent alleles) provided far smaller CIs (Table 4, lines 6 and 10, for example). In this study, adding more loci or more individuals in the sample dramatically increased the precision on N_{eF_k} and N_{eLB} estimates (Figure 2), but doubling the number of loci did not lead to obviously better results than doubling the number of individuals (for loci having 5 alleles), except for N_{eLB} in the cases of low drift. Our results provide rough ideas about this issue when using realistic data sets.

It is interesting that increasing the number of alleles in the samples by having more loci with rare alleles is now less problematic with $\widehat{N_{eLB}}$ than with $\widehat{N_{eF_k}}$. For example, $\widehat{N_{eLB}}$ is less biased and more precise than $\widehat{N_{eF_k}}$ when using loci with some rare alleles (*e.g.*, scheme A in Table 2). This is important because most alleles in natural populations are at low frequency except in severely bottlenecked populations (Luikart *et al.* 1998).

Bayesian methods give more direct information than classical statistics about the estimated parameter N_e, as one gets a posterior distribution for N_e (Figure 3). One advantage of a posterior distribution is that it allows the visualization of the information brought by the data. Thus, a flat posterior means that the data contain no information in connection with the model used: See, for example, Figure 3c. The symmetry of the curve allows us to roughly estimate the signal-to-noise ratio (large if symmetrical, for example, Figure 3a; small if skewed or flat, for example, Figures 3, b and c).

The Bayesian aspect of using $N_{e_{MAX}} = 500$ (*i.e.*, setting

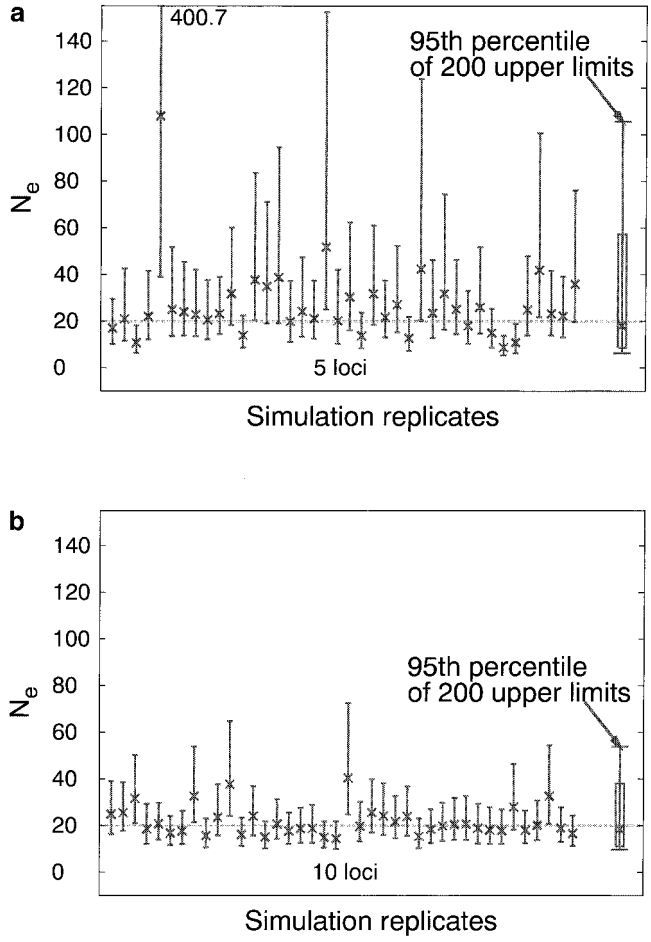

Figure 2.—Confidence intervals for N_{eLB} estimates from independently simulated populations with the same effective population size (N_e = 20). All estimates used loci with five alleles, an initial heterozygosity of 0.6, and samples of 30 individuals separated by five generations. a was generated using only 5 loci, whereas b uses 10 loci. Dashed lines show the true N_e. (a and b) The single vertical bars represent the confidence intervals (the 5–95th percentiles of the posterior distribution) for 40 independent simulation estimates. The summary box chart at the right is made from 200 independent simulations. Ninety-five percent of the 200 upper support limits obtained are below the upper horizontal bar (see "95th percentile" arrow), and 95% of the 200 lower support limits obtained are above the lower horizontal bar. Similarly, 80% of the upper support limits are below the upper edge of the box, and 80% of the lower support limits are above the lower box edge.

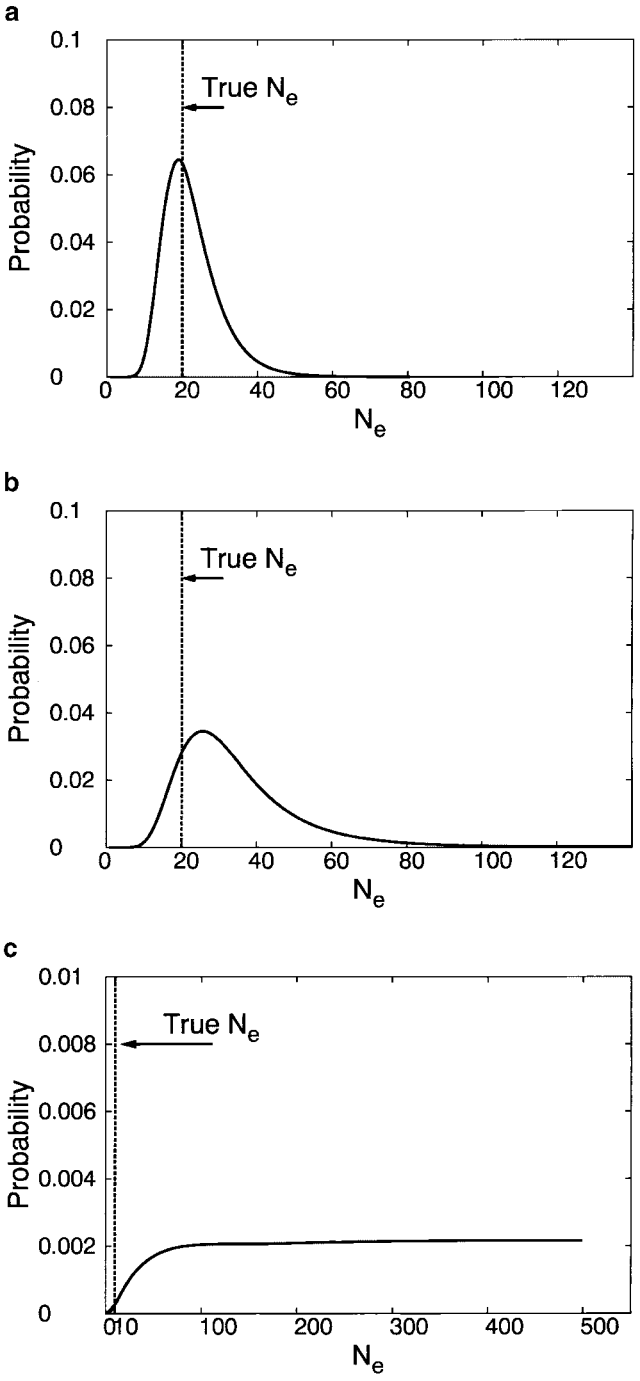

FIGURE 3.—Posterior distributions (likelihood curves) for samples from three independent simulated populations with (a and b) $N_e = 20$ and 5 loci with five alleles ($H = 0.6$) and with (c) $N_e = 10$ and 10 loci with two alleles ($H = 0.2$). All estimates use samples of 30 individuals separated by (a and b) five generations or by (c) one generation. Confidence intervals are often narrow (a and b), but the confidence intervals are occasionally very large (c). Please note that c uses different scales on the axis than a and b.

a narrow prior probability distribution for N_e: $N_{e_{MAX}} = 500$ instead of 5000) provided smaller CIs when drift was relatively weak, but no obvious difference when drift was strong. For example, when N_e was 10 (and $T = 1$, with 10 loci), setting $N_{e_{MAX}}$ to 5000 gave a summary of CIs of 3.6–3732.2 for $\widehat{N_{eLB}}$ percentiles, whereas it gave 3.6–130.6 with $N_{e_{MAX}} = 500$. This is a far better improvement than in the case of high drift ($N_e = 20$, $T = 5$, with 10 loci), for which case we obtained 6.3–107.8, when $N_{e_{MAX}}$ was 5000 and 6.3–105.6 when $N_{e_{MAX}}$ was 500. The better improvement observed under low drift situations was expected because likelihood curves obtained in these cases are skewed toward high N_e, and using the $N_{e_{MAX}}$ prior information is likely to remove a greater amount of the area under the curve (*i.e.*, in the right tail of the curve) than in the high drift cases. Further research is needed to study the influence of using different prior distributions for N_e, *e.g.*, smaller $N_{e_{MAX}}$ values.

In the context of management the Bayesian methodology presented here may be further refined to incorporate the methods of decision theory. For example, a loss function can be defined, which enables the posterior risk of management decisions based on N_e to be quantified. Nevertheless, for conservation biology purposes, the $\widehat{N_{eLB}}$ method should already be useful as it gives narrower credible limits, with less chance to be biased high than low. In conservation biology, it is critical to not overestimate N_e because overestimation could lead managers to not detect a small N_e (and an associated excessive loss of genetic variation) and to consequently underestimate the population extinction risk.

We sincerely thank P. England for the Drosophila data, S. Schönhuth for the killifish data, J. Dallas for the otter data, and C. Spencer for the mosquito fish data. Many thanks to E. Anderson, M. Schwartz, L. Excoffier, N. Valière, S. Manel, and M. Gaudeul for their helpful comments on the manuscript. P.B. was funded by a DEA grant from French Research Ministry. M.A.B. is supported by a University of Reading RETF fellowship. The Bureau des Ressources Génétiques (INRA) provided funding for G.L. and J.-M.C. G.L. was also supported by a grant from the European Union (BIO4 CT96 1189).

LITERATURE CITED

ANDERSON, E. C., E. G. WILLIAMSON and E. A. THOMPSON, 2000 Monte Carlo evaluation of the likelihood for N_e from temporally spaced samples. Genetics **156:** 2109–2118.

BEAUMONT, M. A., and M. W. BRUFORD, 1999 Microsatellites in conservation genetics, pp. 165–182 in *Microsatellites: Evolution and Applications*, edited by D. B. GOLDSTEIN and C. SCHLÖTTERER. Oxford University Press, Oxford.

BEST, N. G., M. K. COWLES and S. K. VINES, 1995 CODA: convergence diagnostics and output analysis software for Gibbs sampler output, version 0.3. Technical report, Cambridge, UK, 1995. (http://www.stat.ufl.edu/system/man/BUGS/cdaman03/).

BURCZYK, J., 1996 Variance effective population size based on multilocus gamete frequencies in coniferous populations: an example of a scots pine clonal seed orchard. Heredity **77:** 74–82.

CABALLERO, A., 1994 Developments in the prediction of effective population size. Heredity **73:** 657–679.

CHIKHI, L., M. W. BRUFORD and M. A. BEAUMONT, 2001 Estimation of admixture proportions: a likelihood-based approach using Markov chain Monte Carlo. Genetics **158:** 1347–1362.

CIOFI, C., M. A. BEAUMONT, I. R. SWINGLAND and M. W. BRUFORD, 1999 Genetic divergence and units for conservation in the Komodo Dragon *Varanus komodoensis*. Proc. R. Soc. Lond. Ser. B **266:** 2269–2274.

DALLAS, J. F., P. J. BACON, D. N. CARSS *et al.* 1999 Genetic diversity in the

Eurasian otter, *Lutra lutra*, in Scotland: evidence from microsatellite polymorphism. Biol. J. Linn. Soc. **68:** 73–86.

Donnelly, P., and S. Tavaré, 1995 Coalescents and genealogical structure under neutrality. Annu. Rev. Genet. **29:** 401–421.

Edwards, A. W. F., 1972 *Likelihood*. Cambridge University Press, Cambridge, UK/London/New York.

England, P. R., 1998 The conservation genetics of population bottlenecks. Ph.D. Thesis, Macquarie University, Sydney, Australia.

Felsenstein, J., M. K. Kuhner, J. Yamato and P. Beerli, 1999 Likelihoods on coalescents: a Monte Carlo sampling approach to inferring parameters from population samples of molecular data. IMS Lect. Notes Monogr. Ser. **33:** 163–185.

Fiumera, A. C., L. Wu, P. G. Parker and P. A. Fuerst, 1999 Effective population size in the captive breeding program of the Lake Victoria cichlid *Paralabidichromis chilotes*. Zoo Biol. **18:** 215–222.

Frankham, R., 1995 Effective population size/adult population size ratios in wildlife: a review. Genet. Res. **66:** 95–106.

Funk, W. C., D. A. Tallmon and F. W. Allendorf, 1999 Small effective population size in the long-toed salamander. Mol. Ecol. **8:** 1633–1640.

Gelman, A., 1996 Inference and monitoring convergence, pp. 131–144 in *Markov Chain Monte Carlo in Practice*, edited by W. R. Gilks, S. Richardson and D. J. Spiegelhalter. Chapman & Hall, London/New York.

Gelman, A., J. B. Carlin, H. S. Stern and D. B. Rubin, 1995 *Bayesian Data Analysis*. Chapman & Hall, London/New York.

Griffiths, R. C., and S. Tavaré, 1994 Simulating probability distributions in the coalescent. Theor. Popul. Biol. **46:** 131–159.

Hedgecock, D., V. Chow and R. S. Waples, 1992 Effective population numbers of shellfish broodstocks estimated from temporal variance in allelic frequencies. Aquaculture **108:** 215–232.

Hill, W. G., 1981 Estimation of effective population size from data on linkage disequilibrium. Genet. Res. **38:** 209–216.

Husband, B. C., and S. C. H. Barrett, 1992 Effective population size and genetic drift in tristylous *Eichhornia paniculata* (Pontederiacee). Evolution **46:** 1875–1890.

Ihaka, R., and R. Gentleman, 1996 R: A language for data analysis and graphics. J. Comput. Graph. Stat. **5:** 299–314.

Jorde, P. E., and N. Ryman, 1996 Demographic genetics of brown trout (*Salmo trutta*) and estimation of effective population size from temporal change in allele frequencies. Genetics **143:** 1369–1381.

Kantanen, J., I. Olsaker, S. Adalsteinsson and K. Sandberg, 1999 Temporal changes in genetic variation of North European cattle breeds. Anim. Genet. **30:** 16–27.

Krimbas, C. B., and S. Tsakas, 1971 The genetics of *Dacus oleae* V: changes of esterase polymorphism in a natural population following insecticide control—selection or drift? Evolution **25:** 454–460.

Laikre, L., P. E. Jorde and N. Ryman, 1998 Temporal change in mitochondrial DNA haplotype frequencies and female effective size in a brown trout (*Salmo trutta*) population. Evolution **52:** 910–915.

Loader, C. R., 1996 Local likelihood density estimation. Ann. Stat. **24:** 1602–1618.

Luikart, G., and J.-M. Cornuet, 1999 Estimating the effective number of breeders from heterozygote excess in progeny. Genetics **151:** 1211–1216.

Luikart, G., F. W. Allendorf, J.-M. Cornuet and W. B. Sherwin, 1998 Distortion of allele frequency distributions provides a test for recent population bottlenecks. J. Hered. **89:** 238–247.

Luikart, G., J.-M. Cornuet and F. W. Allendorf, 1999 Temporal changes in allele frequencies provide estimates of population bottleneck size. Conserv. Biol. **13:** 523–530.

Marjoram, P., and P. Donnelly, 1997 Human demography and the time since mitochondrial Eve, pp. 107–131 in *Progress in Population Genetics and Human Evolution*, edited by P. Donnelly and S. Tavaré. Springer-Verlag, New York.

Miller, L. M., and A. R. Kapuscinski, 1997 Historical analysis of genetic variation reveals low effective population size in a northern pike (*Esox lucius*) population. Genetics **147:** 1249–1258.

Mohle, M., 2000 Ancestral processes in population genetics—the coalescent. J. Theor. Biol. **204:** 629–638.

Nei, M., and F. Tajima, 1981 Genetic drift and estimation of effective population size. Genetics **98:** 625–640.

Nielsen, R., J. L. Mountain, J. P. Huelsenbeck and M. Slatkin, 1998 Maximum-likelihood estimation of population divergence times and population phylogeny in models without mutation. Evolution **52:** 669–676.

O'Ryan, C., E. H. Harley, M. W. Bruford, M. Beaumont, R. K. Wayne *et al.*, 1998 Microsatellite analysis of genetic diversity in fragmented South African buffalo populations. Anim. Conserv. **1:** 85–94.

Planes, S., and G. Lecaillon, 1998 Consequences of the founder effect in the genetic structure of introduced island coral reef fish populations. Biol. J. Linn. Soc. **63:** 537–552.

Pollak, E., 1983 A new method for estimating the effective population size from allele frequency changes. Genetics **104:** 531–548.

Richards, C., and P. L. Leberg, 1996 Temporal changes in allele frequencies and a population's history of severe bottlenecks. Conserv. Biol. **10:** 832–839.

Roberds, J. H., S. T. Friedman and Y. A. Elkassaby, 1991 Effective number of pollen parents in clonal seed orchards. Theor. Appl. Genet. **82:** 313–320.

Saadreva, C., 1997 Low effective sizes in hatchery populations of the European oyster (*Oskea edulis*): implications for the management of genetic resources. J. Shellfish Res. **16:** 441–446.

Saccheri, I., M. Kuussaari, M. Kankare, P. Vikman, W. Fortelius *et al.*, 1998 Inbreeding and extinction in a butterfly metapopulation. Nature **392:** 492–494.

Saccheri, I. J., I. J. Wilson, R. A. Nichols, M. W. Bruford and P. M. Brakefield, 1999 Inbreeding of bottlenecked butterfly populations: estimation using the likelihood of changes in marker allele frequencies. Genetics **151:** 1053–1063.

Schwartz, M. K., D. A. Tallmon and G. Luikart, 1998 Review of DNA census and effective population size estimators. Anim. conserv. **1:** 293–299.

Sitnikov, A. V., Y. F. Kartavtsev and S. M. Nikiforov, 1997 Estimation of the effective size of a grass shrimp (*Pandalus kessleri*) population by a temporal method and direct census of breeders during breeding season. Genetika **33:** 1354–1361.

Slatkin, M., 1996 Gene genealogies within mutant allelic classes. Genetics **143:** 579–587.

Spencer, C. C., J. E. Neigel and P. L. Leberg, 2000 Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. Mol. Ecol. **9:** 1517–1528.

Stephens, M., and P. Donnelly, 2000 Inference in molecular population genetics. J. R. Stat. Soc. B **62:** 605–655.

Tanner, M. A., 1993 *Tools for Statistical Inference*, Ed. 2. Springer-Verlag, New York.

Tarr, C. L., S. Conant and R. C. Fleischer, 1998 Founder events and variation at microsatellite loci in an insular passerine bird, the Laysan finch (*Telespiza cantans*). Mol. Ecol. **7:** 719–731.

Tavaré, S., 1984 Lines-of-descent and genealogical processes, and their application in population genetics models. Theor. Popul. Biol. **26:** 119–164.

Taylor, C. E., Y. T. Toure, M. Coluzzi and V. Petrarca, 1993 Effective population size and persistence of *Anopheles arabiensis* during the dry season in West Africa. Med. Vet. Entomol. **7:** 351–357.

Waples, R. S., 1989 A generalized approach for estimating effective population size from temporal changes in allele frequency. Genetics **121:** 379–391.

Waples, R. S., 1991 Genetic methods for estimating the effective size of cetacean populations, pp. 279–300 in *Report of the International Whaling Commission*, special issue 13, edited by A. R. Hoezel. International Whaling Commission, UK.

Waples, R. S., 2000 Some issues associated with the definition and estimation of effective population size in the conservation of endangered species, in *Population Viability Analysis: Assessing Models for Recovering Endangered Species*, edited by S. R. Beissinger and D. R. McCullough (in press).

Williamson, E. G., and M. Slatkin, 1999 Using maximum likelihood to estimate population size from temporal changes in allele frequencies. Genetics **152:** 755–761.

Wright, S., 1931 Evolution in Mendelian populations. Genetics **16:** 96–159.

Communicating editor: G. A. Churchill