Copyright © 2004 by the Genetics Society of America
DOI: 10.1534/genetics.103.026146

Comparative Evaluation of a New Effective Population Size Estimator Based on Approximate Bayesian Computation

David A. Tallmon,*,†,1 Gordon Luikart* and Mark A. Beaumont†

**Laboratoire d'Ecologie Alpine, UMR Centre National de la Recherche Scientifique 5553, Université Joseph Fourier, F38041 BP 53 Cedex 09, Grenoble, France and †School of Animal and Microbial Sciences, University of Reading, Whiteknights, Reading RG6 6AJ, United Kingdom*

Manuscript received December 26, 2003
Accepted for publication March 1, 2004

ABSTRACT

We describe and evaluate a new estimator of the effective population size (N_e), a critical parameter in evolutionary and conservation biology. This new "SummStat" N_e estimator is based upon the use of summary statistics in an approximate Bayesian computation framework to infer N_e. Simulations of a Wright-Fisher population with known N_e show that the SummStat estimator is useful across a realistic range of individuals and loci sampled, generations between samples, and N_e values. We also address the paucity of information about the relative performance of N_e estimators by comparing the SummStat estimator to two recently developed likelihood-based estimators and a traditional moment-based estimator. The SummStat estimator is the least biased of the four estimators compared. In 32 of 36 parameter combinations investigated using initial allele frequencies drawn from a Dirichlet distribution, it has the lowest bias. The relative mean square error (RMSE) of the SummStat estimator was generally intermediate to the others. All of the estimators had RMSE > 1 when small samples ($n = 20$, five loci) were collected a generation apart. In contrast, when samples were separated by three or more generations and $N_e \leq 50$, the SummStat and likelihood-based estimators all had greatly reduced RMSE. Under the conditions simulated, SummStat confidence intervals were more conservative than the likelihood-based estimators and more likely to include true N_e. The greatest strength of the SummStat estimator is its flexible structure. This flexibility allows it to incorporate any potentially informative summary statistic from population genetic data.

THE effective population size (N_e) plays a central role in how a population evolves because N_e affects the degree to which a population can respond to selection, as well as its sensitivity to inbreeding effects (Crow and Kimura 1970; Lande 1995; Lynch *et al.* 1995). As a result of the critical importance of N_e to evolution, a great deal of effort has focused upon estimating N_e accurately and precisely, and there is always a demand for efficient and useful N_e estimators. Currently, many different methods are available to infer N_e, including ones based on demographic or genetic data. These methods vary in the types of information they use, their accuracy, and the kinds of N_e estimates they provide (Crow and Denniston 1988; Harris and Allendorf 1989; Waples 1991; Schwartz *et al.* 1998; Storz *et al.* 2002).

N_e can be estimated from genetic data in one or more samples (Waples 1991). Most one-sample estimators use associations among alleles at different loci to infer N_e (Hill 1981; Vitalis and Couvet 2001). Multiple-sample methods infer N_e from temporal changes in allele frequencies or the rate of coalescence of alleles between sample periods (Nei and Tajima 1981; Wang 2001; Berthier *et al.* 2002). The most widely used estimator is a method-of-moments estimator (moment estimator), which infers N_e from the standardized variance in allele frequencies sampled one or more generations apart. The change in allele frequencies (F) between sample periods is an inverse function of N_e. Therefore, N_e can be derived from the amount of change in allele frequencies (Nei and Tajima 1981; Waples 1991). However, this estimator uses only the first two moments of the allele frequency distribution to obtain N_e and a number of approximations are made in its derivation. Several studies have noted that it is often biased high (Luikart *et al.* 1999; Wang 2001; Berthier *et al.* 2002).

Likelihood-based methods have been proposed to improve N_e estimation from temporally spaced samples and have become more feasible because fast computers required to calculate likelihoods are more generally available. In theory, these estimators should be more accurate and precise than the moment estimator because they use more of the information provided by the data. Williamson and Slatkin (1999) and Anderson *et al.* (2000) developed maximum-likelihood-based estimators that outperform the moment estimator. However, these methods are computationally extremely intensive, so they have been evaluated only with extremely

[1]*Corresponding author:* LECA-Génomique des Population et Biodiversité, UMR CNRS 5553, Université Joseph Fourier, F-38041 BP 53 Cedex 09, Grenoble, France. E-mail: dtallmon42@yahoo.com

small N_e and three or fewer alleles per locus. Recently, Wang (2001) used time-saving analytical approximations based on the method of Williamson and Slatkin (1999) to provide a more efficient likelihood N_e estimator. Unlike the latter study, which could be used only for biallelic markers, he assumed k alleles at the same locus could be treated as if from k independent, biallelic loci to generate the total likelihood, with an adjustment to take into account that there are only $k - 1$ independent allele frequencies. This pseudo-likelihood estimator appeared to perform well relative to the full-likelihood estimators, but was compared only with three alleles per locus and small N_e.

More recently, Berthier *et al.* (2002) developed a novel likelihood-based approach to obtain $\hat{N}_e$ from two samples using a genealogical representation from coalescent theory. Likelihoods were estimated by importance sampling. Markov chain Monte Carlo (MCMC) was then used to give a Bayesian posterior distribution for N_e. Using a prior distribution to set an upper limit for N_e, which ensured convergence of the MCMC estimates, Berthier *et al.* (2002) showed their estimator to be superior to the moment estimator when genetic drift was strong. Being based on the coalescent, it is only an approximation to the genealogy of the Wright-Fisher model, which assumes discrete generations. The approximation is most accurate when the sample size is small relative to N_e and N_e is large. Because their estimator relies upon Monte Carlo methods to provide the posterior distribution of N_e, it is computationally demanding and has not been extensively evaluated with simulations, nor has it been compared to the pseudo-likelihood estimator of Wang (2001).

We develop a new N_e estimator by combining simple summary statistics from multiple genetic samples with approximate Bayesian computation and compare this estimator to existing ones. Bayesian approaches are attractive because they allow for background information to be incorporated into the model, provide posterior probability distributions for parameters of interest, and integrate out cumbersome nuisance parameters that are common in population genetics data (Shoemaker *et al.* 1999). Potentially exact Bayesian computation (in the sense that the posterior distribution can be approximated to any desired level of accuracy) using MCMC is often very time consuming and requires a substantial amount of programming effort (Beaumont *et al.* 2002). These constraints make potentially exact Bayesian analyses impractical for many applications and difficult to evaluate in many cases, especially for large data sets. Thus, alternative, less time-consuming methods are desirable.

The use of summary statistics has been proposed as a means to avoid the problems presented by complex population genetics analyses (Tavaré *et al.* 1997). This requires the comparison of summary statistics from a real sample with an unknown parameter value to the same summary statistics generated under simulated conditions with known parameter values. Most applications have used a rejection sampling method (Pritchard *et al.* 1999), in which all summary statistic values that fall outside a given tolerance range are rejected, and only those summary statistics that fall within the tolerance range are used to estimate the target parameters (*e.g.*, Tishkoff *et al.* 2001). The approach we use here differs by using local linear regression and smooth weighting of summary statistics and associated N_e values falling within the tolerance range. Beaumont *et al.* (2002) showed that local linear regression and smooth weighting can be used to improve the accuracy and precision of parameter estimation from summary statistics over that provided by rejection sample methods.

We develop a novel N_e estimator using four simple summary statistics and local weighted regression in a Bayesian framework. The four summary statistics are the divergence between samples using Weir and Cockerham's θ (Weir 1990), the change in the number of alleles from the first to second sample (Δa), the change in within-sample gene diversity from the first to second sample (ΔH_s; Nei 1987), and the total expected heterozygosity between samples (H_t; Weir 1990). We evaluate the performance of this summary statistics (SummStat) estimator and compare its performance relative to three existing estimators. These estimators include the standard moment estimator, which has well-known properties over a wide range of parameter values, and promising likelihood estimators developed recently by Berthier *et al.* (2002) and Wang (2001) that are not as well known and have not been compared to each other. For this comparison we use simulated populations of known N_e and levels of genetic variation at marker loci typical of microsatellites, the preferred marker for N_e estimation (Luikart *et al.* 1999). We compare the accuracy and precision of these four N_e estimators, using a range of N_e's (20–100), numbers of loci (5 or 15), and sample sizes (20 or 60) separated by a range of generations (1–10) typical in studies of natural populations. Finally, we illustrate the use of the methods with a real microsatellite data set from an experimental population of mosquito fish (Spencer *et al.* 2000).

METHODS

SummStat estimator: We developed an N_e estimator using the summary statistics approach developed by several authors (Fu and Li 1997; Tavaré *et al.* 1997; Weiss and von Haeseler 1998) and modified by Beaumont *et al.* (2002) to incorporate weighted local regression in a Bayesian context. The general method is described in detail by Beaumont *et al.* (2002) and briefly here. This approach is especially useful when inferences about some parameter of interest, Φ, are difficult to make using full likelihoods. In this method, J values of Φ_i are simulated from a prior distribution, $\Phi_i \sim P(\Phi)$. For

each Φ_i a data set, D_i, is simulated using a Wright-Fisher model described below. Summary statistics, S_i, are then calculated from the data and scaled to have unit variance. Thus, the S_i and Φ_i are drawn from the joint distribution $P(S, \Phi)$. The posterior distribution $P(\Phi | S = S^*)$ is the conditional distribution of Φ given the target summary statistics S^*, calculated from the sample data. To approximate this, the simulated candidate value Φ_i and associated S_i are accepted when the Euclidean distance $||S_i - S^*|| < g$, where g defines a distance such that a proportion d_g of points closest to S^* are accepted (Fu and Li 1997; Tavaré *et al.* 1997; Pritchard *et al.* 1999). To improve the accuracy of the "rejection" scheme, in the method of Beaumont *et al.* (2002) each accepted Φ_i is given a weight that declines quadratically as a function of $||S_i - S^*||$ from 1 at distance 0 to 0 at distance g, and then weighted linear regression is used to adjust the values of Φ_i. The method fits a regression line such that each $\Phi_i = a + bS_i + e_i$, and then, assuming constant variance within the interval given by $||S_i - S^*|| < g$, makes the adjustment $\Phi_i' = a + b(S^* - S_i)$. These Φ_i' are then assumed to be random samples from the posterior distribution $P(S, \Phi)$, which, depending on how close to sufficient are the summary statistics, is itself assumed to be close to $P(D|\Phi)$.

To examine our SummStat N_e estimator, we created an individual-based, Wright-Fisher simulation model of diploid organisms using the programming language C. This model differs slightly from a Wright-Fisher model in that there are two allogamous sexes and equal numbers of each sex. In the present case, the model was initialized using genotypes drawn from a uniform Dirichlet distribution with eight alleles per locus, but it can also be initialized using a coalescent-based microsatellite distribution with any specified θ value or range of values. Following the collection of the first sample of n diploid individuals at time $t_1 = 0$, a breeding population of size N_e was created and randomly mated for t generations, when the second sample of n diploid individuals was collected from progeny of adults in generation t_2. Following the collection of the second sample, summary statistics were estimated over the L loci sampled. This sampling schedule follows plan 2 of Waples (1989). The summary statistics consisted of a common measure of divergence between samples, the coancestry coefficient,

$$\hat{\theta} = \frac{1}{L}\sum_{i=1}^{L}(Q_i - q_i)/(1 - q_i),$$

where Q_i is the estimated probability of identity of alleles in a sample at locus i and q_i is the estimated probability of identity of alleles in the two samples at locus i (Weir 1990); the mean change in the number of alleles from the first to second sample,

$$\Delta\hat{a} = \frac{1}{L}\sum_{i=1}^{L}(a_{1,i} - a_{2,i}),$$

where $a_{1,i}$ and $a_{2,i}$ are the estimated number of alleles at locus i in samples 1 and 2, respectively; the change in mean within-sample gene diversity from the first to second sample,

$$\Delta\hat{H}_s = \left(1 - \frac{1}{L}\sum_{i=1}^{L}p_{1,i}^2\right) - \left(1 - \frac{1}{L}\sum_{i=1}^{L}p_{2,i}^2\right),$$

where $p_{1,i}^2$ and $p_{2,i}^2$ are the squares of the estimated frequency of each allele present at locus i in the first and second samples, respectively (Nei 1987); and total expected heterozygosity between samples,

$$\hat{H}_t = 1 - \frac{1}{L}\sum_{i=1}^{L}p_i^2,$$

where p_i^2 is the estimated frequency of each allele in the combined samples (Weir 1990). All estimated values for these summary statistics also included appropriate sample size corrections.

It is important to note that we could have included any summary statistics that can be calculated from standard population genetic data, but limited ourselves to four that are straightforward to calculate, commonly used in population genetics studies, and thought to be related to our parameter of interest, N_e, on the basis of previous research and some preliminary simulations of our own.

The simulation model sampled from a uniform flat prior distribution of N_e between 4 and 400 to generate $J = 50{,}000$ values of the summary statistics. This prior is reasonable because N_e can fall in this range, even for some populations with thousands of individuals, and because most applications of N_e estimators are cases where N_e is small (Waples 1991). To test the performance of the SummStat estimator, we simulated independent populations of known N_e and calculated summary statistics for each target data set sampled from each population (see details of sampling conditions below). We used natural log of N_e in all regressions to adjust the values of Φ_i to ensure that the results were robust to changes in g. Values of N_e accepted within $d_g = 0.02$, as described above, were then regarded as samples from the posterior distribution of N_e. The mode and credible intervals of the posterior distribution of back-transformed N_e values were calculated using the density estimation method of Loader (1996), implemented in the statistical package R (R Development Core Team 2003). The log transformation did not prevent some regression-adjusted points from projecting beyond the upper bound of N_e.

Comparison to other estimators: We compared the performance of the SummStat estimator to three existing two-sample N_e estimators, including a coalescent (TMVP; Beaumont 2003, which is based on the program TM3 in Berthier *et al.* 2002), a moment (Nei and Tajima 1981), and a pseudo-likelihood (MLNE; Wang 2001) estimator. The comparisons were conducted over a range of parameter values, including the effective population

sizes (N_e = 20, 50, 100), generations between samples (t = 1, 3, 5, 10), sample sizes (n = 20, 60 diploid individuals), and numbers of loci (L = 5, 15). Note that in some situations the sample size exceeded the effective size, which is common when sampling natural populations in which the number of juvenile or adult individuals can greatly exceed the number of breeders (Frankham 1995) and corresponds to sampling plan II of Waples (1989). Only one set of simulations with large N_e = 100 is included (t = 1, 3, 5, 10; n = 60; L = 15), because of the long time required to run all of the models on large N_e populations and because most applications of N_e estimators are to natural populations with small N_e. We ran 600 independent iterations of each combination of parameter values for the model comparisons. The models were compared in terms of bias and precision of $\hat{N}_e$, using a number of metrics: relative mean square error (RMSE) of the mode, median bias of the mode, 95th percentile of 95% confidence/credible intervals, and proportion of confidence/credible intervals that excluded true N_e. We also show the bootstrapped estimates of the standard errors of the RMSE estimates.

All sampled genotypes generated in our simulations were written to files to provide input for TMVP and MLNE. TMVP is an updated version of the TM3 program used in Berthier *et al.* (2002) and provides a posterior distribution of N_e using a MCMC approach with importance sampling (Beaumont 2003). For the simulations used here, the size of the importance sample was 20, 20,000 MCMC updates were used with 10 updates between estimate outputs, and an N_e ceiling was set at 400. The TMVP N_e evaluated in this article is the mode from the posterior MCMC distribution of values (except for an initial 10% of values discarded as burn-in).

MLNE provides a pseudo-likelihood N_e as described in Wang (2001). We used an updated version of the program, described in Wang and Whitlock (2003), which provides a ceiling for N_e. We used an N_e ceiling of 400 in these simulations. MLNE also provides N_e from the moment estimator of Nei and Tajima (1981), but no confidence intervals. Therefore, the 95% confidence intervals of the moment estimator are not reported and described here, but they have been explored previously and their precision and bias are well known (Luikart *et al.* 1999; Wang 2001; Berthier *et al.* 2002).

Robustness to prior probability distribution of N_e: We investigated the robustness of the SummStat estimator to changes in the prior probability distribution (using the same metrics described above) when the lower limit of the prior remained at 4 and upper limit of the prior probability distribution for N_e was 200, 400, or 1000. True N_e was set to 50. Diploid individuals were sampled 1 or 3 generations apart and genotyped at 15 loci. We did not directly compare the SummStat estimator to the others in these simulations. Results are based upon 600 iterations of each set of conditions.

Robustness to initial allele frequencies: Allele frequencies are drawn from a uniform Dirichlet distribution in most simulation-based studies of N_e estimators, including ours (Anderson *et al.* 2000; Wang 2001). This corresponds to specific assumptions about the mutation process—namely, that it is a k-allele model where each mutation occurs at rate Θ, and the probability of mutation to any specific allele is $1/k$. We investigated the robustness of our results to assumptions about the underlying allele frequency distribution by initializing a set of simulations and priors with allele frequency data drawn from the coalescent. For these simulations the Wright-Fisher population described above was used. Initial levels of polymorphism were determined by a randomly selected value of Θ = 5–15, and the population was first sampled t_1 = 1–10 generations following a change in effective size to N_e = 20. We used uniform flat priors for Θ and the t_1. The second sample was collected t_2 = 1, 3, 5, or 10 generations after t_1. In each sample, n = 60 individuals were genotyped at L = 15 loci. The estimators were evaluated as described above on the basis of 600 iterations of each set of conditions. With this slight modification we are now able to obtain the posterior distribution of N_e marginal to both Θ and t_1. This highlights the advantage of approximate Bayesian computation based on summary statistics, and it is straightforward to make changes in the model with minimal programming effort.

Real data set: To further illustrate the SummStat estimator, we estimated N_e for an experimentally bottlenecked population reported in Spencer *et al.* (2000) and evaluated by Berthier *et al.* (2002), using their coalescent-based N_e estimator and the moment estimator. The target data set was collected from a large source population of mosquito fish (*Gambusia affinis*) that was sampled and then experimentally reduced to eight pairs of founders, allowed to grow for two generations, and then resampled (Spencer *et al.* 2000). Forty individuals were genotyped at eight microsatellite loci in each sample. Although true N_e is unknown for this population, N_e = 16 is the hypothesized value. For this target data set we generated values for the same summary statistics used in our simulations: $\hat{\theta}$ = 0.283, $\Delta\hat{a}$ = −5.50, $\Delta\hat{H}_s$ = −0.096, and $\hat{H}_t$ = 0.708. We used the same Wright-Fisher model described above to simulate 50,000 populations, with an N_e randomly drawn from a uniform flat prior between 4 and 500 and initial levels of polymorphism determined by the coalescent with a uniform flat prior for θ between 5 and 15. In each simulation, 40 diploid individuals were sampled and genotyped at eight loci in generations 0 and 2. These genotypes were used to calculate the same summary statistics calculated for the target data set, with the tolerance set at 0.02. We report the mode and 95% credible intervals of the posterior distribution from SummStat and compare them to the other estimators. The estimates from TMVP and the moment method are taken directly from Berthier *et*

al. (2002; Table 3). MLNE estimates were obtained using a ceiling of 500.

RESULTS AND DISCUSSION

Estimator performance: The SummStat estimator provides accurate and precise $\hat{N}_e$ across a range of generations between samples and numbers of individuals and loci sampled. In general, this estimator shows a small, positive bias when only a single generation passes between samples (Figure 1). Only when $N_e = 50$ and sampling conditions are at their worst (only 20 individuals, 5 loci, and 1 generation) did SummStat show a substantial positive bias. The bias decreases rapidly with increases in the number of individuals sampled, loci sampled, and generations between samples. The relative mean square error (RMSE) of the SummStat estimator is small when true $N_e = 20$, but is noticeably larger when $N_e = 50$ and only 1 generation separates the samples (Figure 2). There is a consistent, striking decrease in the RMSE when the number of generations between samples increases from 1 to 3, regardless of the other sampling conditions. However, there is relatively little increase in accuracy and precision as the number of generations increases from 3 to 10.

The SummStat 95% credible intervals contain true N_e consistently and do not depart greatly from the expectation of 2.5% of the true N_e values falling above or below the credible intervals (Table 1). There is a slight tendency to exclude true N_e from the lower credible interval with sparse data and a single generation between samples, which probably is due to the slight positive bias of the SummStat estimator under these conditions noted above. In addition, the 95th percentiles of the 95% credible intervals tend to be conservative when data are sparse or drift is weak (Table 2). The SummStat 95% credible intervals narrow rapidly with increasing numbers of generations between samples. For the parameter combinations evaluated, a threefold increase in the number of loci sampled provides a markedly greater increase in precision than a threefold increase in the number of individuals sampled.

The SummStat estimator appears robust to changes in the upper limit for the prior probability distribution of N_e. The bias, RMSE, and lower credible interval of the estimator do not change much with a fivefold change in the upper limit of the N_e prior from 200 to 1000 and true $N_e = 50$ (Table 3). However, while the lower credible interval is relatively stable, the upper credible interval does appear to be sensitive to the choice of a prior when only a single generation separates samples. For example, the upper credible interval more than doubled with a fivefold increase in the upper limit of the N_e prior. With three generations between samples the upper credible limit changed by <25% with a fivefold increase in the upper limit of the N_e prior. This limit changes less with more generations between samples (results not shown).

Comparison to other estimators: The SummStat estimator has the lowest bias and generally performs well relative to the other estimators. When only 1 generation passes between samples ($t = 1$), TMVP and MLNE tend to underestimate N_e, whereas the SummStat estimator slightly overestimates it (Figure 1). In general, TMVP shows the largest bias of the four estimators when 1–5 generations pass between sampling events. In contrast, with 10 generations passing between sampling events, the bias of TMVP is greatly reduced, whereas the moment estimator has the greatest bias. MLNE generally shows the second lowest bias after the SummStat estimator, and this bias decreases with increasing generations between samples. Wang (2001) reported a slight positive bias for MLNE that is based on the mean of the N_e estimates, rather than on the median as reported here. In fact, the distribution of N_e estimates is skewed (as illustrated in Wang and Whitlock 2003, Figure 3), and the mean is generally larger than the median, as reported here.

All of the estimators generally show reduced bias and RMSE with increasing sample sizes, loci, and generations between samples. The exception is the moment estimator, which shows increasing bias with increasing generations between samples. In most cases simulated, SummStat RMSE is intermediate to that of the other estimators, and the relative performance of the estimators changes with sampling conditions. However, the RMSE of the SummStat estimator is generally larger than that of MLNE, which often provides the smallest RMSE, and is consistently smaller than the RMSE of the moment estimator. TMVP shows very small RMSE and little bias when 10 generations pass between samples. In contrast, the moment estimator consistently has the largest RMSE when there are 10 generations between samples.

TMVP and MLNE tend to exclude true N_e from the upper 95% credible/confidence interval much more than expected by chance (Table 1). This tendency is most notable with TMVP, which consistently provides the lowest upper and lower confidence intervals. The observed bias of the confidence intervals generally worsens with the number of loci or individuals sampled, if the number of generations between samples is held constant. In contrast, the observed downward biases of the 95% confidence/credible intervals of TMVP and MLNE improve rapidly with increasing generations between samples. Although the 95% credible intervals of the SummStat estimator are less likely to exclude true N_e across the entire range of sampling conditions, the 95% confidence/credible intervals of the other estimators are often narrower when there are few generations between samples (Table 2). Consequently, there is a fairly consistent trade-off in the bias and precision of these estimators, especially when data are sparse, N_e large, and few generations pass between samples. TMVP and MLNE are more precise, but there is a risk of underestimating

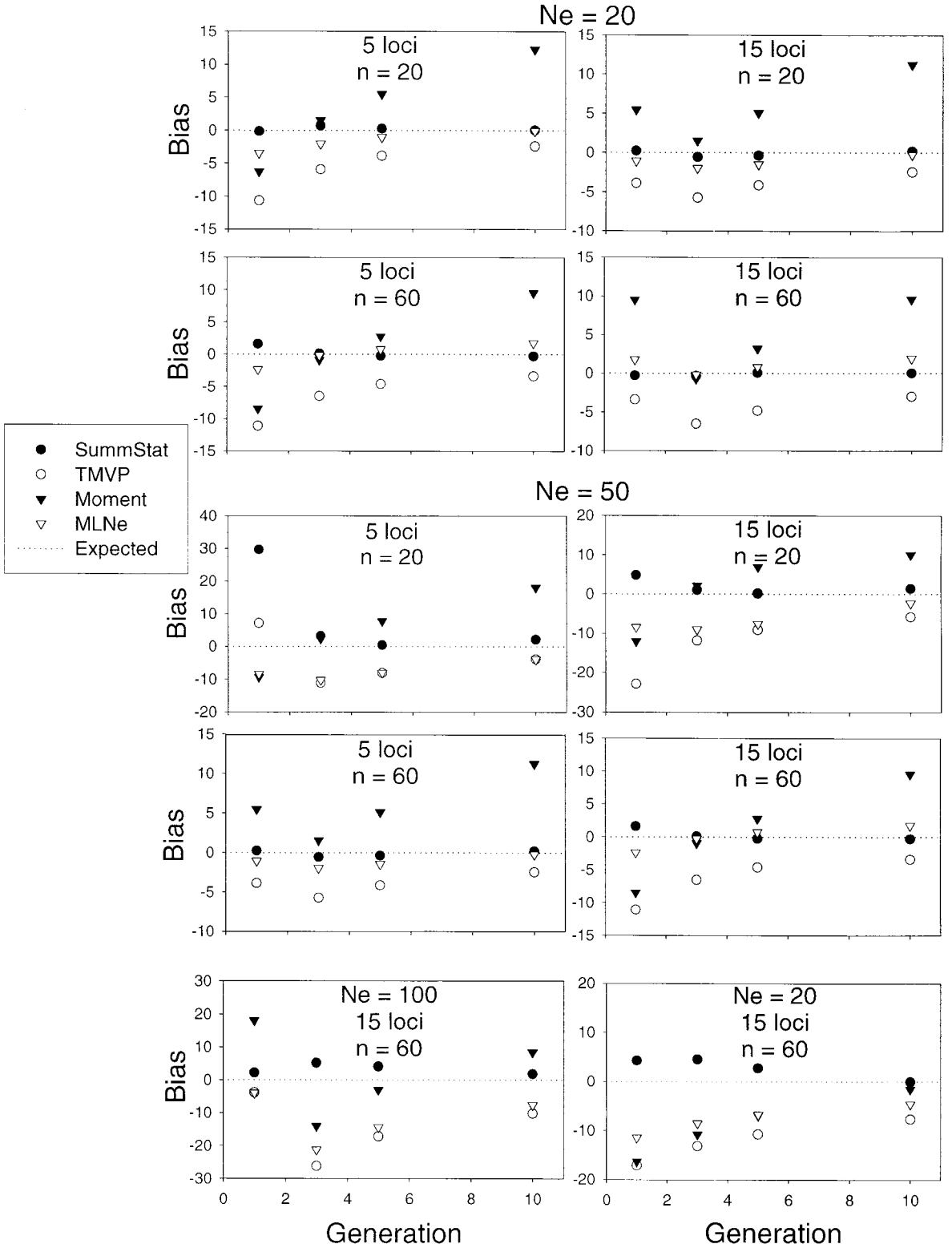

FIGURE 1.—Bias of four effective population size (N_e) estimators, including the new SummStat estimator, for known N_e values (20, 50, or 100) and various sampling conditions. Bias is reported as the median difference between the mode and true N_e from 600 simulations. The bottom right shows bias when initial allele frequencies are drawn from the coalescent (all others are from Dirichlet). The *y*-axis changes throughout the figure to emphasize differences among methods.

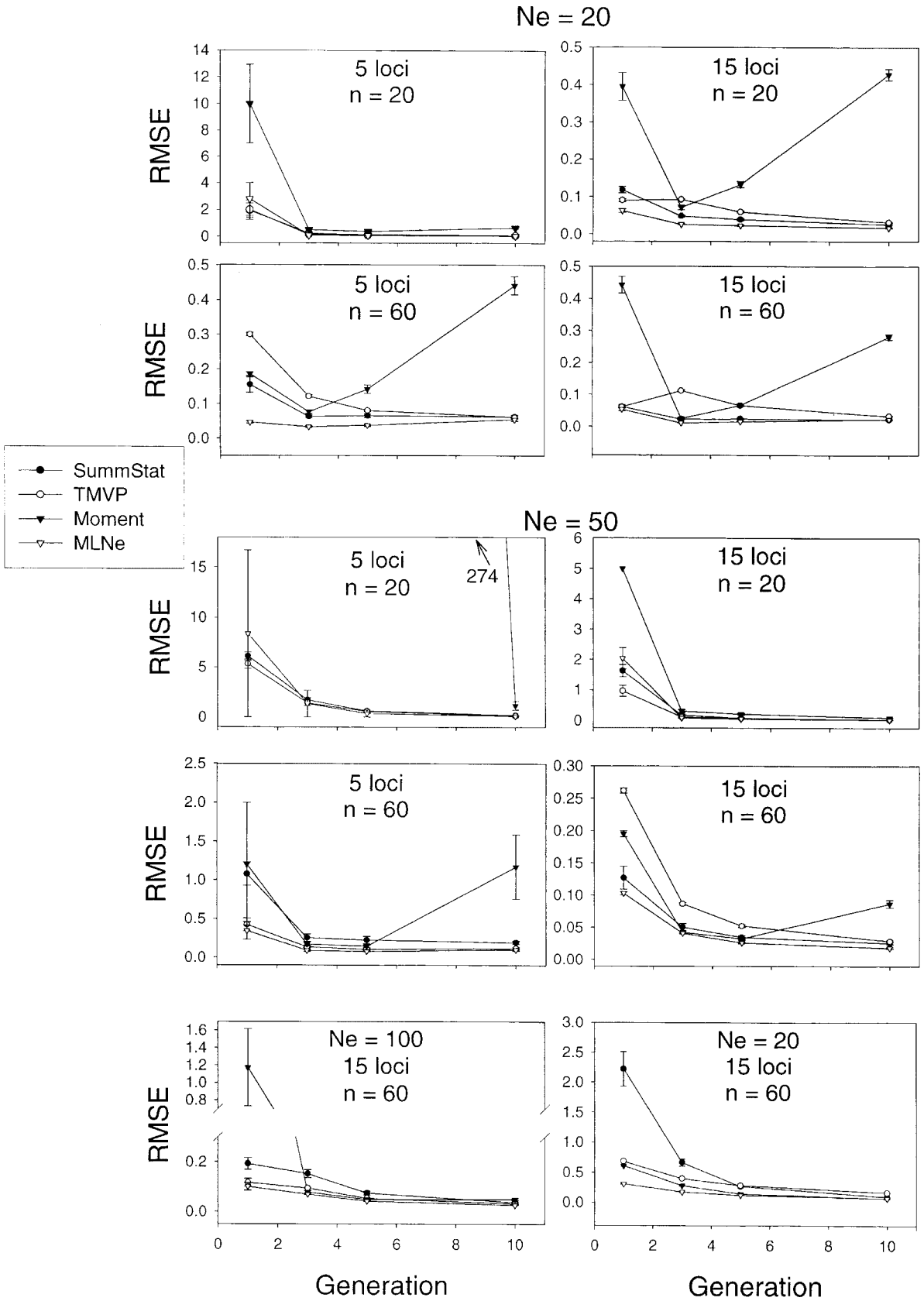

FIGURE 2.—Relative mean square error (RMSE) of four effective population size (N_e) estimators, including the new SummStat estimator, for true N_e (20, 50, or 100), and various sampling conditions. RMSE values reported are means from 600 simulations with SE bars from 1000 bootstraps. The bottom right shows estimator RMSE when initial allele frequencies are drawn from the coalescent (all others are from Dirichlet). The *y*-axis changes throughout the figure to emphasize differences among methods.

TABLE 1

Percentage of lower and upper credible/confidence limits that exclude true N_e

		SummStat			TMVP			MLNE		
Conditions	t	L	U	>400	L	U	>400	L	U	>400
					$N_e = 20$					
5 loci										
$n = 20$	1	17.1	24.9	1.0	9.8	37.8	0.0	1.2	24.0	28.5
	3	1.0	1.0	0.0	0.0	16.3	0.0	0.6	4.5	0.3
	5	0.1	2.7	0.0	0.0	12.8	0.0	1.3	4.5	0.0
	10	1.3	1.8	0.0	0.2	6.8	0.0	1.2	1.8	0.0
$n = 60$	1	0.3	1.0	0.0	0.0	71.0	0.2	0.6	8.7	0.0
	3	0.7	1.2	0.0	0.0	38.2	0.0	1.0	2.7	0.0
	5	1.5	2.3	0.0	0.0	25.0	0.0	2.3	0.5	0.0
	10	2.0	2.2	0.0	0.7	11.0	0.0	4.8	0.5	0.0
15 loci										
$n = 20$	1	0.2	2.7	0.0	0.0	12.8	0.0	1.3	4.5	0.0
	3	0.2	1.3	0.0	0.0	43.5	0.0	0.3	9.2	0.0
	5	3.2	3.3	0.0	0.0	35.0	0.0	0.3	7.5	0.0
	10	1.3	2.8	0.0	0.3	16.5	0.0	0.8	3.8	0.0
$n = 60$	1	2.0	2.2	0.0	0.7	11.0	0.0	4.8	0.5	0.0
	3	1.3	1.7	0.0	0.0	78.8	0.0	1.8	4.3	0.0
	5	3.2	2.8	0.0	0.0	55.2	0.0	5.3	2.0	0.0
	10	2.0	2.5	0.0	0.0	25.7	0.0	12.2	0.3	0.0
					$N_e = 50$					
5 loci										
$n = 20$	1	14.0	0.0	21.2	4.7	2.0	0.0	0.0	3.2	83.2
	3	8.3	0.5	6.0	4.5	4.3	0.0	1.2	4.2	32.0
	5	4.3	0.6	1.6	3.0	3.5	0.0	1.3	4.2	9.8
	10	2.2	1.3	0.2	0.8	4.0	0.0	0.8	4.2	0.8
$n = 60$	1	6.3	0.6	3.2	1.0	23.2	0.0	0.3	12.7	12.0
	3	3.0	1.2	0.2	0.3	14.5	0.0	0.3	5.7	0.6
	5	1.8	1.8	0.0	0.5	13.2	0.0	1.0	7.0	0.2
	10	1.3	2.0	0.0	0.7	8.2	0.0	1.2	3.2	0.0
15 loci										
$n = 20$	1	10.7	0.2	6.5	2.3	10.2	0.0	0.0	3.2	53.0
	3	2.3	0.8	0.2	0.3	15.3	0.0	0.0	11.1	1.2
	5	0.6	1.5	0.0	0.2	14.2	0.0	0.3	12.8	0.0
	10	4.6	1.3	0.0	0.6	17.8	0.0	1.5	6.3	0.0
$n = 60$	1	1.8	0.3	0.0	0.0	72.0	0.0	0.0	35.7	0.0
	3	1.7	1.5	0.0	0.0	43.3	0.0	0.0	20.8	0.0
	5	1.7	1.8	0.0	0.0	31.8	0.0	0.2	12.3	0.0
	10	3.8	1.5	0.0	0.8	17.8	0.0	1.3	5.5	0.0
					$N_e = 100$					
15 loci										
$n = 60$	1	2.1	1.3	0.2	0.8	4.0	0.0	0.8	4.2	0.8
	3	4.0	1.0	0.2	0.3	22.0	0.0	0.3	17.6	0.5
	5	2.0	1.2	0.0	0.0	15.4	0.0	0.0	12.0	0.0
	10	1.5	1.9	0.0	0.5	13.3	0.0	0.2	9.0	0.0

Percentages of lower (L) and upper (U) limits of 95% confidence/credible intervals for the SummStat, TMVP, and MLNE estimators that do not include true $N_e = 20$, 50, or 100 and the percentage of upper limits that exceed the N_e ceiling (>400) of 400 are shown. Values were generated from 600 independent iterations of n diploid individuals sampled t generations apart.

TABLE 2

Ninety-fifth percentiles of the lower and upper confidence/credible limits for the N_e estimators

		SummStat		TMVP		MLNE	
Conditions	t	L	U	L	U	L	U
				$N_e = 20$			
5 loci							
$n = 20$	1	3.50	355.97	2.76	369.19	7.27	>400
	3	6.57	173.20	5.22	126.16	8.41	79.61
	5	7.32	101.95	6.17	58.45	8.95	56.47
	10	7.91	53.67	7.26	45.05	9.66	45.93
$n = 60$	1	7.46	135.50	3.50	36.33	10.1	50.49
	3	8.78	62.50	5.83	35.82	11.29	43.50
	5	9.26	51.88	6.85	37.15	11.91	43.40
	10	8.60	45.20	7.49	38.39	12.26	44.53
15 loci							
$n = 20$	1	7.32	101.95	6.17	58.45	8.95	56.47
	3	10.66	47.64	7.81	30.00	10.97	34.42
	5	11.02	40.85	8.91	31.08	11.49	33.30
	10	11.49	34.99	10.41	30.17	12.54	32.70
$n = 60$	1	8.60	45.20	7.49	38.39	12.26	44.53
	3	11.84	36.21	8.03	23.11	13.89	29.73
	5	12.03	35.00	9.35	35.47	14.54	31.50
	10	12.25	32.94	10.66	27.34	15.14	32.46
				$N_e = 50$			
5 loci							
$n = 20$	1	6.01	418.88	3.82	395.82	9.73	>400.00
	3	9.17	404.96	8.84	388.23	13.65	>400.00
	5	12.42	366.70	12.99	367.70	16.10	>400.00
	10	18.56	268.35	17.29	232.99	18.82	195.88
$n = 60$	1	10.12	390.67	6.80	365.54	14.27	>400.00
	3	16.70	279.41	13.83	191.48	19.53	163.09
	5	18.44	216.59	15.85	136.04	20.10	134.71
	10	22.16	139.62	19.80	111.39	22.97	115.33
15 loci							
$n = 20$	1	10.28	406.64	7.20	388.48	14.51	>400.00
	3	18.20	298.50	16.33	279.80	19.51	193.06
	5	23.71	188.41	20.02	136.26	22.03	119.83
	10	30.48	93.13	27.45	74.99	30.65	79.64
$n = 60$	1	21.14	227.58	11.51	91.64	19.80	105.38
	3	26.33	130.72	19.51	75.30	24.81	80.33
	5	28.73	102.66	23.12	72.81	27.71	77.82
	10	30.73	90.41	27.38	73.70	30.65	77.81
				$N_e = 100$			
15 loci							
$n = 60$	1	18.56	268.35	17.29	232.99	18.82	195.88
	3	40.95	338.54	35.06	271.64	40.53	250.12
	5	49.13	276.98	43.94	192.89	47.53	191.12
	10	55.62	210.41	51.22	168.61	54.37	172.06

The 95th percentile of the 95% confidence/credible lower (L) and upper (U) limits of three N_e estimators is shown. See Table 1 for details.

true N_e. The SummStat estimator provides relatively conservative credible intervals and is less biased.

The general patterns seen in the data generated using a Dirichlet distribution to draw initial allele frequencies also appear to be robust to changes in the underlying allele frequency distribution and lend credibility to our

TABLE 3

Robustness of SummStat estimator to changes in the N_e prior

t	Prior	Bias	RMSE (SE)	Lower C.I.	Upper C.I.
1	4–200	4.04	0.13 (0.017)	22.11	166.07
	4–400	4.58	0.13 (0.017)	21.06	233.81
	4–1000	2.15	0.093 (0.013)	14.94	374.42
3	4–200	2.83	0.050 (0.0040)	26.79	122.19
	4–400	2.54	0.048 (0.0045)	26.31	131.48
	4–1000	2.99	0.048 (0.0046)	25.80	151.72

Metrics shown are the same as those described in Figures 1 and 2 and Table 1. Values are based upon 600 iterations with true N_e = 50 and 15 loci from 60 diploid individuals sampled t = 1 or 3 generations apart. All prior distributions are uniform flat between the values shown. C.I., confidence interval.

TABLE 4

Coverage of the N_e estimators with initial allele frequencies drawn from the coalescent

	SummStat		TMVP		MLNE	
t	L	U	L	U	L	U
			% N_e excluded			
1	8.0	0.2	0.0	98.6	0.0	93.6
3	3.2	0.6	0.0	94.3	0.0	81.0
5	3.0	0.8	0.0	89.7	0.1	70.3
10	2.3	2.0	0.0	76.8	0.3	46.8
			95th percentile of C.L.			
1	4.00	179.64	1.10	13.13	5.26	20.90
3	4.38	100.49	2.78	20.71	6.48	25.50
5	4.97	75.96	3.99	22.10	7.38	26.66
10	6.39	54.75	5.73	25.00	8.57	28.64

Coverage includes the percentage of true N_e values excluded (% N_e excluded) from the lower (L) and upper (U) confidence/credible limits and the 95th percentiles of the 95% confidence/credible limits (95th percentile of C.L.). Values are based upon 600 iterations with true N_e = 20 and 15 loci genotyped from 60 diploid individuals sampled t generations apart.

analyses. The patterns of bias and RMSE of the estimators are similar whether initial allele frequencies are drawn by sampling from a Dirichlet or from the coalescent. However, the tendencies of each estimator noted above appear to be accentuated. The SummStat estimator continued to show the lowest bias (Figure 1) as well as a relatively large RMSE when few generations passed between samples (Figure 2). TMVP and MLNE were precise, but showed considerably strong bias. A noticeable change in the relative performance of these estimators is the improved performance of the moment estimator in the coalescent simulations, especially with 5–10 generations between samples. The positive bias and low accuracy of the moment estimator when drift is strong have been noted elsewhere to result from the fixation of low-frequency alleles present in the first sample (Richards and Leberg 1996; Luikart *et al.* 1999; Williamson and Slatkin 1999; Wang 2001; Berthier *et al.* 2002). The improved performance here is probably due to there being fewer rare alleles because the simulated populations were first sampled as many as 10 generations following the initial reduction in N_e. This long period of strong drift before drawing the first sample could decrease the number of rare alleles, thus reducing this source of bias for the moment estimator.

The patterns seen in the 95% confidence/credible intervals when initial allele frequencies are drawn from the coalescent are also similar to the patterns in simulations that draw from the Dirichlet. Namely, the SummStat credible intervals are conservative and are slightly more likely to fall above than below true N_e (Table 4). In contrast, the confidence/credible intervals of TMVP and MLNE tend to fall below true N_e and provide underestimates of true N_e quite frequently.

TMVP provides point estimates and upper credible intervals that are biased low when there are few generations between samples, and this bias has been noted previously (Berthier *et al.* 2002). However, TMVP is very accurate and precise when 10 generations pass between samples, despite the fact that it suffers from a disadvantage in our comparisons in that it assumes a coalescent model to simulate drift and a Wright-Fisher model was assumed for our simulations. Indeed, TMVP was very accurate in a subset of simulations using the coalescent rather than a Wright-Fisher model to simulate drift (data not shown).

Empirical data: The application of SummStat to data collected from an experimental mosquito fish population studied by Spencer *et al.* (2000) demonstrates its utility in an empirical setting. The SummStat point estimate of N_e = 8.93 falls below the hypothesized true N_e = 16 for this population. In contrast, the coalescent, moment, and MLNE estimates are 21.6, 35.4, and 32.51, respectively. More importantly, the SummStat N_e 95% credible intervals contain the hypothesized true N_e value and are reasonably precise, whereas all of the others excluded N_e from their credible/confidence intervals (Figure 3). In this case, only eight loci and 40 individuals were sampled, which is a modest sampling effort to expect for most natural populations. Although this is only a single application of the SummStat estimator, the results are encouraging and suggest the estimator may work well when applied to microsatellite data sampled from real populations.

Other considerations: It has been suggested by Wang and Whitlock (2003) that measurement of bias and precision should be based on $1/(2N_e)$ rather than on N_e. The relative advantages or disadvantages of different point estimators and transformations can be assessed only in a decision-theoretic framework (O'Hagan

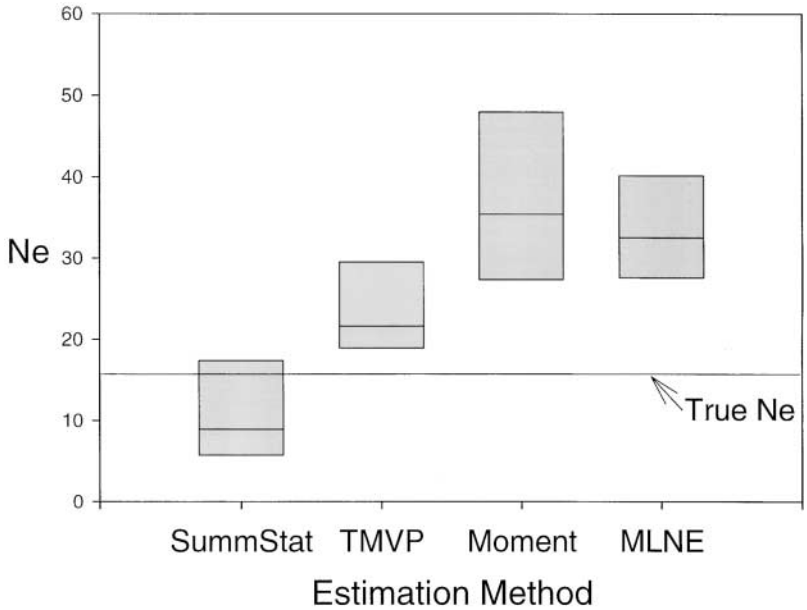

Figure 3.—Comparison of four effective population size (N_e) estimators using data from Richards and Leberg (1996) for an experimental population of mosquito fish. Lines indicate point estimates and shaded boxes indicate 95% confidence/credible intervals.

1994), where a utility function can be specified for N_e, and in the absence of this we prefer to measure bias and precision in terms of N_e. Examination of a number of data sets indicates that the general patterns in the data reported here are robust whether one examines estimates of N_e or $1/(2N_e)$.

An important choice in Bayesian inference is the method used to generate the prior probability distribution. We have found our estimator to be robust whether initial allele frequencies are drawn from the Dirichlet or the coalescent. However, as with any Bayesian approach, it is important to recognize that large differences between the biological conditions that gave rise to the data and the method used to generate the prior can affect the validity of the results. Common effects of small population size, such as small departures from Hardy-Weinberg proportions or gametic phase equilibrium, should not create problems in our simulations because the assumed model incorporates these effects adequately. However, if the target data had been collected from a population with high rates of undetected immigration, for example, and the model used to generate the priors did not incorporate this, then biological inferences could be misleading. Consequently, researchers should temper their interpretations with careful consideration of the assumptions (*e.g.*, immigration, no substructuring, neutral markers) used in creating the prior probability distribution and how differences within the biological context that created the data set of interest might affect their inferences.

It is also important to note that the quality of the information provided by each summary statistic will vary with sampling conditions, amounts of genetic drift, and type of genetic markers. For example, if molecular markers or populations with few alleles are studied, then the mean loss of alleles per locus may be a less informative summary statistic than, for example, divergence between samples. Thus, careful attention should be paid to choosing statistics that are likely to be informative, as one seeks to maximize the amount of information that can be extracted from a data set while avoiding the curse of dimensionality created by using many different summary statistics (Beaumont *et al.* 2002). Ideally, both the literature and simulations should be used to help choose the best summary statistics for a given set of biological conditions and sampling constraints.

Conclusions: The SummStat estimator performs well, relative to the others, using only four summary statistics. It is the least biased method over the full range of parameter values investigated, has an RMSE intermediate to the others in most of the scenarios investigated, and performed well when applied to a real data set. However, MLNE generally has the smallest RMSE of the estimators compared, despite a negative bias and tendency to exclude true N_e from the upper confidence intervals when few generations pass between samples. The complementary properties of MLNE (precision) and SummStat (accuracy) suggest that it would be wise to use both to estimate N_e. TMVP may be more appropriate when reproduction occurs continuously rather than in discrete generations.

The SummStat estimator is best viewed as a flexible approach to N_e estimation, in terms of both modeling and choice of summary statistics, as shown by the ease with which it is modified to consider a very different prior from a Dirichlet. An advantage of the SummStat estimator is that one can combine any potentially informative summary statistics calculated from sample data into a single approach to N_e estimation. For example, it could conceivably be extended beyond the allele frequency-based information used in the present example to include genotypic information, thus extracting more

information provided by the data (Hill 1981; Waples 1991). The time-saving attributes of approximate Bayesian approaches relative to exact Bayesian methods make them an appealing alternative for population genetics applications. New molecular technologies provide ever-increasing numbers of molecular markers, which, in turn, make full-likelihood calculations very time consuming. The performance of the SummStat N_e estimator suggests further exploration and expansion using approximate Bayesian computation may be a fruitful means to improve efficient N_e estimation.

We thank L. Excoffier, J. Wang, and R. Waples for helpful reviews and suggestions that improved this manuscript. We also thank J. Wang for graciously recompiling slightly altered versions of his program for us. This work was supported by a National Science Foundation postdoctoral grant (INT 0202707) to D.A.T. and by the European Commission (Econogene contract QLK5-CT-2001-02461) to G.H.L. M.A.B. is supported by a Natural Environment Research Council Advanced Fellowship.

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., 2003 Estimation of population growth or decline in genetically monitored populations. Genetics **164:** 1139–1160.

Beaumont, M. A., W. Zhang and D. J. Balding, 2002 Approximate Bayesian computation in population genetics. Genetics **162:** 2025–2035.

Berthier, P., M. A. Beaumont, J.-M. Cornuet and G. Luikart, 2002 Likelihood-based estimation of the effective population size using temporal changes in allele frequencies: a genealogical approach. Genetics **160:** 741–751.

Crow, J. F., and C. Denniston, 1988 Inbreeding and variance effective population effective numbers. Evolution **42:** 482–495.

Crow, J. F., and M. Kimura, 1970 *An Introduction to Population Genetics Theory.* Harper & Row, New York.

Frankham, R., 1995 Effective population size/adult population size ratios in wildlife: a review. Genet. Res. **16:** 95–107.

Fu, Y.-X., and W.-H. Li, 1997 Estimating the age of the common ancestor of a sample of DNA sequences. Mol. Biol. Evol. **14:** 195–199.

Harris, R. B., and F. W. Allendorf, 1989 Genetically effective population size of large mammals: an assessment of estimators. Conserv. Biol. **3:** 181–191.

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

Lande, R., 1995 Mutation and conservation. Conserv. Biol. **9:** 782–791.

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

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

Lynch, M., J. Conery and R. Burger, 1995 Mutational meltdowns in sexual populations. Evolution **49:** 1067–1080.

Nei, M., 1987 *Molecular Evolutionary Genetics.* Columbia University Press, New York.

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

O'Hagan, A., 1994 *Bayesian Inference.* Arnold, London.

Pritchard, J. K., M. T. Seielstad, A. Perez-lezaun and M. W. Feldman, 1999 Population growth of human Y-chromosomes: a study of Y chromosome microsatellites. Mol. Biol. Evol. **16:** 1791–1798.

R Development Core Team, 2003 *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna (http://www.R-project.org).

Richards, C., and P. L. Leberg, 1996 Evaluation of the effects of bottlenecks on changes in allele frequencies. Conserv. Biol. **10:** 832–839.

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

Shoemaker, J. S., I. Painter and B. S. Weir, 1999 Bayesian statistics in genetics. Trends Genet. **15:** 354–358.

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.

Storz, J. F., U. Ramakrishnan and S. C. Alberts, 2002 Genetic effective size of a wild primate population: influence of current and historical demography. Evolution **56:** 817–829.

Tavaré, S. J., D. J. Balding, R. C. Griffiths and P. Donnelly, 1997 Inferring coalescence times from DNA sequence data. Genetics **145:** 505–518.

Tishkoff, S. A., R. Varkonyi, N. Cahinhinan, S. Abbes and G. Argyropoulos, 2001 Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. Science **293:** 455–462.

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

Wang, J., 2001 A pseudo-likelihood method for estimating effective population size from temporally spaced samples. Genet. Res. **78:** 243–257.

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

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, Cambridge, UK.

Weir, B. S., 1990 *Genetic Data Analysis: Methods for Discrete Population Genetic Data.* Sinauer, Sunderland, MA.

Weiss, G., and A. von Haeseler, 1998 Inference of population history using a likelihood approach. Genetics **149:** 1539–1546.

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

Communicating editor: L. Excoffier