



## Temporal-method estimates of $N_e$ from highly polymorphic loci

Thomas F. Turner<sup>1\*</sup>, Laura A. Salter<sup>2</sup> & John R. Gold<sup>3</sup>

<sup>1</sup>Department of Biology and Museum of Southwestern Biology, University of New Mexico, Albuquerque, New Mexico 87131; <sup>2</sup>Department of Mathematics and Statistics, University of New Mexico, Albuquerque, New Mexico 87131; <sup>3</sup>Center for Biosystematics and Biodiversity, Texas A&M University, College Station, Texas 77843, USA (\*Author for correspondence: turnert@unm.edu)

Received 1 December 2000; accepted 5 May 2001

**Key words:** coalescent, confidence intervals, Monte Carlo methods, red drum (*Sciaenops ocellatus*), statistical power, variance effective population size

### Abstract

The temporal method is used widely to estimate genetic effective population size ( $N_e$ ), a parameter of fundamental interest to studies of evolutionary and conservation biology. The statistical properties of temporal-method estimates have not been explored for highly polymorphic DNA markers that often contain many alleles occurring in very low frequencies. We used a Monte Carlo simulation approach to assess accuracy and precision of the temporal method when implemented with haplotypic/allelic data at mitochondrial (mt)DNA and nuclear-encoded microsatellite DNA loci. Estimates of  $N_e$  were between 2%–106% greater than their true values in 48 simulations parameterized using different demographic scenarios, models of mutation, and sample sizes. Overestimation of  $N_e$  results from a bias in the approximation used by Waples (1989) to derive the relationship between the expected temporal variance ( $F$ ) and  $N_e$  when allele frequencies are very close to 0 or 1. Our results show that one commonly applied solution to this problem, binning of low-frequency alleles, results in a trade-off of accuracy and precision in some cases. We show that both chi-square and normal approximations are appropriate for estimating 95% confidence intervals of  $N_e$  and we develop a power analysis based on the chi-square distribution to estimate sample sizes and allelic diversity required to evaluate specific hypotheses. For highly polymorphic loci like mtDNA and microsatellites, the increased precision afforded by the presence of rare alleles outweighs the upward bias in temporal-method estimates of  $N_e$ .

### Introduction

The advent of highly polymorphic DNA markers, such as variable number of tandem repeat (VNTR) loci or microsatellites, has sparked renewed interest in estimating population genetic parameters from molecular data. One such parameter, the genetic effective population size ( $N_e$ ), has received considerable theoretical and empirical attention because it is a fundamental determinant of genetic diversity in natural populations (Crow 1986).  $N_e$  is also recognized as a critical part of assessing long-term extinction risks of species or populations (Frankham 1995).

There are several methods commonly used to estimate  $N_e$  from molecular data. Most recently, atten-

tion has focused on methods that employ coalescent theory (Donnelly and Tavaré 1995). Methods based on the coalescent evaluate the long-term effective size (Avisé 2000), and can greatly overestimate  $N_e$  in large populations that have experienced a very recent bottleneck. The temporal method is an alternative approach that can provide very accurate estimates of current  $N_e$  after a recent bottleneck event (Luikart et al. 1998). The temporal method is widely used in studies of managed or conserved species and populations (Husband and Barrett 1992; Hedgecock et al. 1992; Jorde and Ryman 1996; Miller and Kapuscinski 1997; Laikre et al. 1998; Turner et al. 1999).

The temporal method relates the standardized variance of allele frequencies across generations to the

(variance) effective population size (Nei and Tajima 1981; Pollak 1983) The statistical properties of the method are known for loci with two to twelve alleles (Nei and Tajima 1981; Pollak 1983; Waples 1989; Luikart et al. 1999; Williamson and Slatkin 1999). What is not known is how the temporal method performs when loci with many alleles (>12) are used to estimate  $N_e$ . This is of more than passing interest, as many commonly used genetic markers such as mitochondrial (mt)DNA and microsatellite DNA loci possess greater than 12 haplotypes or alleles, most of which occur at frequencies less than 5%. Because temporal-method estimates of  $N_e$  based on di- or tri-allelic loci can be biased when one or more of the alleles are rare (Waples 1989), there is good reason to suspect that highly polymorphic loci also produce significantly biased estimates of  $N_e$ .

In this study, we assessed accuracy and precision of the temporal method to estimate  $N_e$  from highly polymorphic genetic markers using a Monte Carlo simulation approach. Two demographic scenarios were examined. First, we evaluated performance of the temporal method under conditions of mutation-drift equilibrium, by simulating genetic drift of a rapidly evolving genetic marker in an idealized population of constant effective size. Second, we evaluated temporal shifts in allele frequencies estimated from mtDNA and microsatellite loci sampled in red drum (*Sciaenops ocellatus*) in the Gulf of Mexico. Genetic effective size of this long-lived and abundant marine fish species is on the order of  $N_e = 10^4$  (Gold et al. 1993; Turner et al. 1999), but simulated effective sizes were not greater than  $N_e = 5 \times 10^2$ . Such conditions would be expected in a large population that has undergone a moderate demographic bottleneck (severe bottlenecks have been examined by Richards and Leberg (1996) and Luikart et al. (1999)). In both demographic scenarios, molecular markers exhibited high levels of polymorphism and a number of rare alleles per locus, providing an excellent case study for evaluating performance of the temporal method to estimate  $N_e$  when employing highly polymorphic loci.

## Materials and methods

Monte Carlo simulations were parameterized to reflect the demographic scenarios presented above. For the first set of simulation studies, our aim was to examine accuracy and precision of temporal-method estimates of  $N_e$  using allelic frequency data gathered from a

rapidly evolving diploid locus (e.g. a microsatellite locus) in an idealized population at mutation-drift equilibrium. We began setting the effective population size to  $N_e = 500$  and the mutation rate to  $\mu = 0.001$ , which yields  $\theta = 4N_e\mu = 2$ . The mutation rate was chosen to reflect values reported for microsatellite DNA loci in other organisms (Jarne and Lagoda 1996). An expected allele frequency distribution was generated assuming an infinite alleles model (IAM) using the generating function

$$\Phi(x) = \theta(1-x)^{\theta-1}x^{-1}, \quad (1)$$

(Nei 1987: eq. 13.27), where  $x$  is the frequency of a particular allele in the population. A stepwise mutation model may also be appropriate for microsatellite loci, so we generated an allele frequency distribution assuming a stepwise mutation model (SMM) using

$$\Phi(x) = \frac{\Gamma(\theta + A + 1)}{\Gamma(\theta)\Gamma(A + 1)}(1-x)^{\theta-1}x^{A-1}, \quad (2)$$

where  $A = (1 - \theta - \sqrt{1 + 2\theta})/(\sqrt{1 + 2\theta} - 1)$  (Chakraborty et al. 1980: eq. A1). We then employed Ewens' (1972) sampling theorem to find the expected number of alleles  $E(k)$  for  $\theta = 2$  for  $n = 1000$  genes ( $N_e = 500$  for a diploid locus), which is approximately equal to 13 alleles. For each locus, number of alleles (rounded to whole numbers) was assigned by drawing from a normal distribution with  $E(k) = 13.0$  and  $\text{Var}(k) = 3.6$ , and frequencies were assigned to each allele by drawing values at random with replacement from the IAM and SMM allele frequency distributions. We then standardized allele frequencies within each locus so that they summed to one. This procedure was repeated eight times to create eight loci with a total of approximately 104 independent alleles. The number of loci was chosen to facilitate comparison with empirical red drum microsatellite data.

For the moderate bottleneck case, empirically derived haplotype/allele frequencies were used to parameterize simulations. DNA data from two different classes of highly polymorphic genetic markers were examined for red drum. MtDNA is maternally inherited as a single haploid locus, and often is highly variable because mutation rates are generally high relative to nuclear-encoded structural genes such as allozyme loci (Brown 1983). We assayed a total of 1369 red drum collected from 14 localities in the northern Gulf of Mexico for variation at 104 mtDNA restriction sites (data published in Gold et al. 1999).

The second class of markers examined was microsatellites, short simple-sequence repeats contained in nuclear genomes of most eukaryotes (Brooker et al. 1994). Microsatellites are inherited in a codominant, Mendelian fashion, and are thought to evolve much more rapidly than nuclear-encoded structural genes (Jarne and Lagoda 1996). As a result, microsatellite loci generally are highly polymorphic, have higher heterozygosities, and possess more alleles per locus than do most other nuclear-encoded markers. We screened eight consistently resolvable and polymorphic loci for allelic variation among 409 individuals that were a representative (geographic) subset of the individuals screened for variation in mtDNA restriction sites. PCR, screening, and allele-scoring methods followed Turner et al. (1998).

Haplotype and allele frequencies at each locus were calculated by pooling individuals across sampling localities. We took this approach to increase sample sizes, which provided more robust frequency estimates. Such an approach is only justified when levels of spatial genetic variation are very small and do not appreciably alter estimates of global  $N_e$  based on an assumption of panmixis. Previous studies of mtDNA variation have indicated weak population structure in red drum that is consistent with an isolation-by-distance model of migration (Gold et al. 1999). To evaluate potential effects of spatial genetic structure in microsatellites, we conducted regression of pairwise genetic divergence ( $R_{ST}$ ) and geographic distance (km) between samples. This analysis revealed a pattern of isolation by distance, but the slope of regression was very small ( $1.65 \times 10^{-6}$ ,  $P < 0.001$ ). Theoretical studies have shown that isolation by distance does not significantly alter global effective size when the slope of regression is substantially less than one (Chambers 1995).

We assessed accuracy and precision of the temporal method separately for each demographic scenario and each class of genetic markers by extending the Monte Carlo simulation approach developed by Waples (1989). Briefly, simulations entailed constructing initial gamete pools containing allele frequencies generated under an equilibrium model (described above), or observed in red drum. From initial pools, gametes were drawn at random with replacement to comprise generation  $t(0)$  of size  $N_e$ . A second generation,  $t(1)$ , of size  $N_e$ , was drawn at random with replacement from generation  $t(0)$ . A recently developed model for estimating  $N_e$  from species with overlapping generations focuses on

evaluating temporal shifts in allele frequency over temporally adjacent year classes and then applying a correction factor based on demographic information (Jorde and Ryman 1996). The time interval for estimation of temporal variance in allele frequencies across adjacent cohorts is short ( $\leq 1$  generation), and so we were particularly interested in the performance of the temporal method over short time spans (i.e. one generation).

Individuals were sampled with replacement from generations  $t(0)$  and  $t(1)$  (following sampling plan II of Waples (1989)), yielding sample sizes  $S_{t(0)} = S_{t(1)}$  ( $S = 50$  and  $S = 200$  were used in simulations). Haplotype/allele frequencies were tabulated for  $S_{t(0)}$  and  $S_{t(1)}$ , respectively. Within each marker class, separate simulations were conducted for three different effective sizes:  $N_e = 50, 100, \text{ and } 500$ .

Simulations were conducted to assess accuracy of estimates of the quantity  $F$ , the standardized genetic variance attributable to gene frequency changes across generations  $t(0)$  and  $t(1)$ . For each marker class, we examined two computational methods for estimating  $F$ : the first,  $\hat{F}_C$  (the carat denotes an estimate throughout the manuscript), was developed by Nei and Tajima (1981); the second,  $\hat{F}_K$ , was developed by Pollak (1983). Waples (1989) provides a detailed account of each computational method. Means and distributions of  $\hat{F}_C$  and  $\hat{F}_K$  were estimated from 10000 randomized replicates.

The value  $F$  is inversely related to  $N_e$  in an idealized, diploid population by

$$N_e \approx \frac{t}{2 \left( F - \frac{1}{2S_0} - \frac{1}{2S_t} \right)}, \quad (3)$$

where  $t$  is the number of generations separating the two samples, and  $S_{t(0)}$  and  $S_{t(1)}$  are sample sizes at generations  $t(0)$  and  $t(1)$ , respectively (Waples 1989). Equation 3 was adjusted for mtDNA data to account for haploid inheritance to give

$$N_{ef} \approx \frac{t}{\left( F - \frac{1}{S_0} - \frac{1}{S_t} \right)}, \quad (4)$$

which yields the effective female population size,  $N_{ef}$ . We multiplied  $N_{ef}$  by 2 for comparison to  $N_e$  values based on microsatellites. By substituting  $\hat{F}_C$ ,  $\hat{F}_K$ , and  $S$  values into Equations 3 and 4, we solved for  $\hat{N}_e$ . We then evaluated accuracy of these estimates by computing the percentage difference between  $\hat{N}_e$  and  $N_e$ .

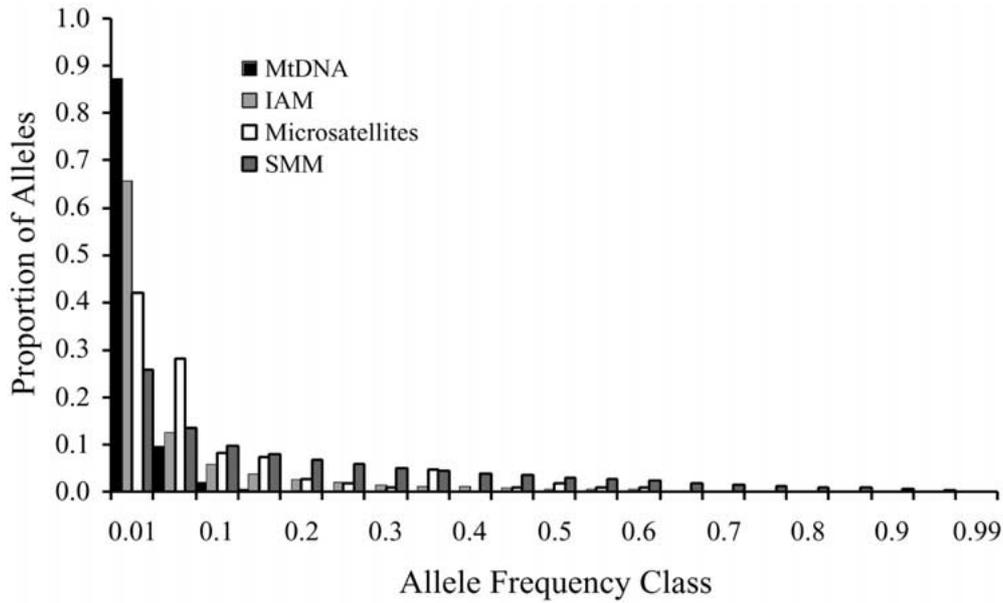


Figure 1. Equilibrium allele frequency distributions expected under infinite alleles (IAM) and stepwise mutation models (SMM) for a diploid locus segregating in a population of effective size  $N_e = 500$  and a mutation rate  $\mu = 0.001$  (yielding  $\theta = 4N_e\mu = 2.0$ ). Observed frequencies of haplotypes/alleles for mtDNA-RFLPs and eight microsatellite loci surveyed from red drum are included for comparison (Gold et al. 1999). MtDNA has the highest proportion of rare alleles ( $< 0.01$ ). SMM is expected to have the fewest rare alleles.

A second goal of the study was to compare precision of  $\hat{F}$  (and thus  $\hat{N}_e$ ) between marker classes. The theoretical distribution of  $F$  (the variance of a variance) is not known but has been approximated both by chi-square (Pollak 1983; Waples 1989) and normal distributions (Nei and Tajima 1981, Jorde and Ryman 1996). We asked whether the distribution of 10000 randomized replicates of  $\hat{F}$  (obtained from simulation) was best described by a chi-square or a normal distribution. Expected values at a given probability under a chi square distribution were determined by

$$F'_\alpha = \left( \frac{k\hat{F}}{\chi^2_{[\alpha,k]}} \right) - \frac{1}{2S_0} - \frac{1}{2S_t}, \quad (5)$$

where  $F'_\alpha$  is the expected value of  $\hat{F}$  (minus sampling error) at a given probability ranging from  $\alpha = 0.001$  to  $\alpha = 0.999$ ,  $k$  is the number of independent haplotypes/alleles,  $\hat{F}$  is the mean of 10000 randomized replicates, and  $\chi^2_{[\alpha,k]}$  is the value of chi-square at probability  $\alpha$  with degrees of freedom equal to  $k$ . Equation 5 was adjusted for mtDNA using the formula

$$F'_\alpha = \left( \frac{k\hat{F}}{\chi^2_{[\alpha,k]}} \right) - \frac{1}{S_0} - \frac{1}{S_t}. \quad (6)$$

Expected values under the normal distribution were determined by evaluating  $F'_\alpha$  using the inverse normal density function with mean and standard deviation of  $\hat{F}$  (minus sampling error) determined from 10000 randomized replicates. Goodness of fit was determined qualitatively by plotting the simulated  $\hat{F}$  distribution to the expected  $F'$  distribution and examining departures from equality.

## Results and discussion

Rare haplotypes/alleles (frequencies lower than 0.01) were present in initial gamete pools of equilibrium models (IAM and SMM) and in empirical red drum data (Figure 1). Eighty-eight percent of haplotypes observed in the mtDNA data were rare by this criterion, as were approximately 66% of expected alleles under IAM at equilibrium. Microsatellites had proportionally fewer rare alleles (40%) and possessed more moderately frequent and common alleles (Figure 1), as did the SMM case with 26% of alleles expected to be rare. Our mtDNA data set was highly polymorphic and contained numerous rare alleles because of the large number of restriction enzymes (13) used to characterize molecular variation.

Table 1. Accuracy of estimates of  $N_e$  (denoted by the carat) as determined from Monte Carlo simulations for a population at migration/drift equilibrium (IAM, SMM), and a population that has experienced a moderate bottleneck (mtDNA, microsatellites). Three initial effective population sizes ( $N_e$ ), two sample sizes ( $S$ ), and two computational methods ( $\hat{F}_K$  (Pollak 1983) and  $\hat{F}_C$  (Nei and Tajima 1981)) were examined. Percent deviation is calculated as the difference of  $\hat{N}_e$  and  $N_e$  divided by  $N_e$  multiplied by 100

Simulation Case	$N_e$	$S$	$\hat{N}_e$ (Pollak)	% deviation	$\hat{N}_e$ (Nei-Tajima)	% deviation
IAM	50	50	71	42	69	38
	100	50	146	46	142	42
	500	50	1031	106	993	99
	50	200	61	22	60	20
	100	200	121	21	118	18
	500	200	630	26	596	19
SMM	50	50	59	18	62	24
	100	50	119	19	129	29
	500	50	567	13	596	19
	50	200	56	12	56	12
	100	200	108	8	109	9
	500	200	503	1	516	3
mtDNA	50	50	67	34	73	46
	100	50	139	39	160	60
	500	50	689	38	983	97
	50	200	54	8	55	10
	100	200	118	18	121	21
	500	200	643	29	725	45
Microsatellites	50	50	64	27	63	26
	100	50	131	31	129	29
	500	50	778	56	767	53
	50	200	59	19	58	17
	100	200	118	18	115	15
	500	200	603	21	568	14

Microsatellite loci exhibited patterns similar to those observed in other marine teleost fishes (Broughton and Gold 1997; DeWoody and Avise 2000). Given the same total number of alleles in the data set, microsatellites are expected to have proportionally fewer rare alleles than mtDNA because many independent loci (with relatively few alleles per locus) can be employed. Increasing polymorphism in mtDNA surveys requires identifying alleles from a single locus, so any new alleles identified (by increasing the number of restriction enzymes assayed) are likely to be rare.

Regardless of initial conditions, Monte Carlo simulations indicated that the temporal method

consistently underestimated  $F$ , and thus overestimated  $N_e$ , in all cases.  $\hat{N}_e$  exceeded true values by as little as 2% or as much as 106%, depending on the kind of genetic marker evaluated, whether an equilibrium or moderate bottleneck case was examined, and the initial size of the population under simulation (Table 1). The magnitude of bias appears to be slight in nearly all cases (except mtDNA and IAM,  $N_e = 500$ ,  $S = 50$ ) and similar to that observed by Waples (1989) for triallelic loci with one rare allele (frequency = 0.01). The magnitude of bias did not appear to differ in a systematic fashion between Pollak's (1983) and Nei and Tajima's (1981) methods for estimation of  $\hat{F}$ , and thus  $\hat{N}_e$  (Table 1).

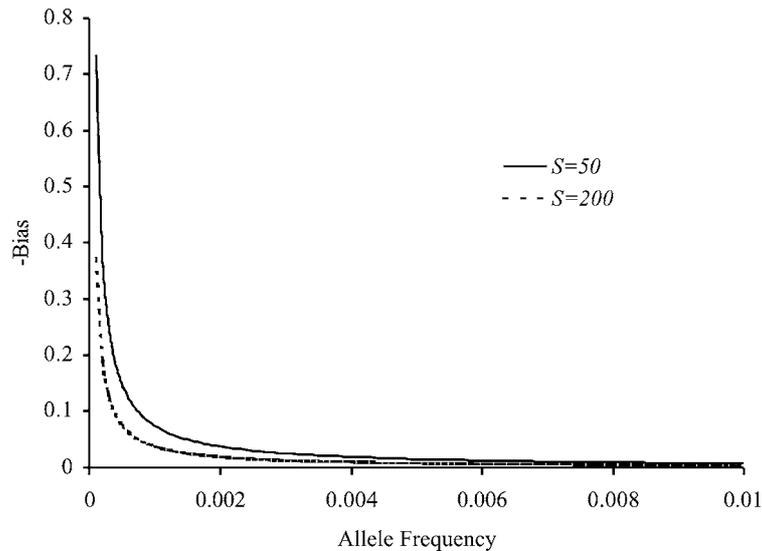


Figure 2. Graph of negative bias in the Waples (1989) approximation of the expectation of  $F$  as a function of the allele frequency when  $N_e = 50$ ,  $S = 50$  and  $S = 200$  based on Equation 7 in the text. Note that bias decreases as the sample size  $S$  increases.

At least two factors appear to cause the temporal-method to overestimate  $N_e$  when employing highly polymorphic loci: (i) the proportion of rare alleles in the data set and; (ii) the number of individuals sampled ( $S$ ) which determines the sampling variance of alleles. We further examined how these two factors might be affecting bias by considering the approximation used by Waples (1989) to derive the relationship between the expected value of  $F$  and  $N_e$ . Waples used a first-order Taylor Series Expansion to approximate the expected value of  $F$ , and found the relationship given in Equation 3, with  $F$  replaced by its expectation. To study the adequacy of this approximation, we used a second-order Taylor Series Expansion to approximate the expected value of  $F$ , and we examined the difference between our approximation and that given by Waples. We call this difference the bias, and denote it by  $Bias(S_0, S_t, N_e, P, t)$  since it is a function of the sample sizes at generations 0 and  $t$ , the effective population size, the allele frequency,  $P$ , and the number of generations,  $t$ . When Sampling Plan II of Waples is used,  $Bias(S_0, S_t, N_e, P, t)$  can be computed. In this case, Equation 3 becomes

$$N_e \approx \frac{t}{2\left(F - \frac{1}{2S_0} - \frac{1}{2S_t} - Bias(S_0, S_t, N_e, P, t)\right)}, \quad (7)$$

(see Appendix for details). Thus, if bias is positive,  $N_e$  will be underestimated by Equation 3, and if bias is negative,  $N_e$  will be overestimated by Equation 3.

We computed the bias term as a function of  $P$  for the combinations of  $N_e$  and  $S$  used in our simulations. Bias in these cases is in fact negative and quite small, except for values of  $P$  which are very close to 0 or to 1. Figure 2 shows the negative of the bias as a function of  $P$  when  $N_e = 50$ ,  $S = 50$ , and  $t = 1$ . Graphs in the other cases are similar, except that bias is smaller when  $S$  is larger, as expected (Figure 2). Thus, the temporal method is nearly unbiased, but bias increases as allele frequencies get close to 0 or to 1, and decreases as the sample size increases.

Simulations are generally consistent with this result. Accuracy of  $\hat{N}_e$  decreases as the proportion of rare alleles increase in a particular data set: on average, IAM and mtDNA estimates are most biased (and contain the highest proportion of rare alleles), followed by microsatellites, and then SMM (Table 1). Bias decreases with increasing  $S$  in nearly all cases. We note also that bias due to low allele frequency alleles increases with increasing  $t$  ( $t = 2$  and  $t = 3$  cases evaluated but not shown). Such an increase in bias might be expected since low frequency alleles are constrained in how far they can drift downward, i.e. low frequency alleles can go extinct in very short time periods via genetic drift.

To examine whether the bias term above accounted for overestimation of  $N_e$  in our simulation studies, we applied Equation 7 to the mtDNA simulation case ( $N_e = 500$ ,  $S = 200$ ,  $\hat{F}_C$ ). We started by finding the mean

bias term for the first 50 replicates from simulation and subtracting that value from the estimate of  $\hat{F}_C$  across 10,000 replicates. The resulting corrected value of  $\hat{N}_e = 480$  was much closer to the true value ( $N_e = 500$ ) than uncorrected  $\hat{N}_e = 725$  (from Table 1), suggesting that the bias term largely accounts for overestimation of  $\hat{N}_e$  in this case.

Another potential solution to bias introduced by numerous rare alleles is to bin rare alleles, thereby reducing the number of rare alleles and increasing the number of moderately frequent alleles. We examined the efficacy of this approach by summing alleles that were less than or equal to a prescribed frequency in initial gamete pools to create a single new allele (Figure 3). Binning alleles in this fashion, however, results in a trade-off of accuracy and precision in the case where  $S = 200$  (Figure 3a). When  $S = 50$ , accuracy did not increase substantially when alleles were binned, but precision decreased at a rate similar to the case where  $S = 200$  (Figure 3b). Given that bias is relatively small in most cases, binning alleles at the expense of precision is probably not warranted. It is worth noting that increased sample sizes resulted in an increase of both accuracy and precision of temporal method estimates of  $N_e$  (Table 1, Figures 3a, b).

Bivariate plots of observed and expected values suggested that  $F'$  is approximated closely by both chi-square and normal distributions (Figure 4). There were differences between chi-square and normal approximations: values of  $F'$  under a chi-square distribution were larger (in all simulation cases) than values generated from a normal distribution and from observed values (Figure 4). Confidence intervals of  $\hat{N}_e$  calculated from the chi-square distribution tended to be slightly wider than those computed from the normal distribution, and are thus more conservative for hypothesis testing. However, the normal approximation produced confidence intervals that were more similar to those obtained from simulation analysis (Figure 4). In practice, a normal approximation would not be applicable to mtDNA because variance of  $F$  across independent loci cannot be computed.

Determining a close approximation for the distribution of  $F$  is important for two reasons. First, appropriate confidence intervals can be computed for  $\hat{F}$  and  $\hat{N}_e$  without resampling; and second, rough estimates of statistical power to test hypotheses can be produced rapidly and conducted with preliminary data. For example, a researcher could compare statistical power of different marker classes to test the hypothesis that lower bound  $\hat{N}_e$  values differ from benchmark  $N_e$

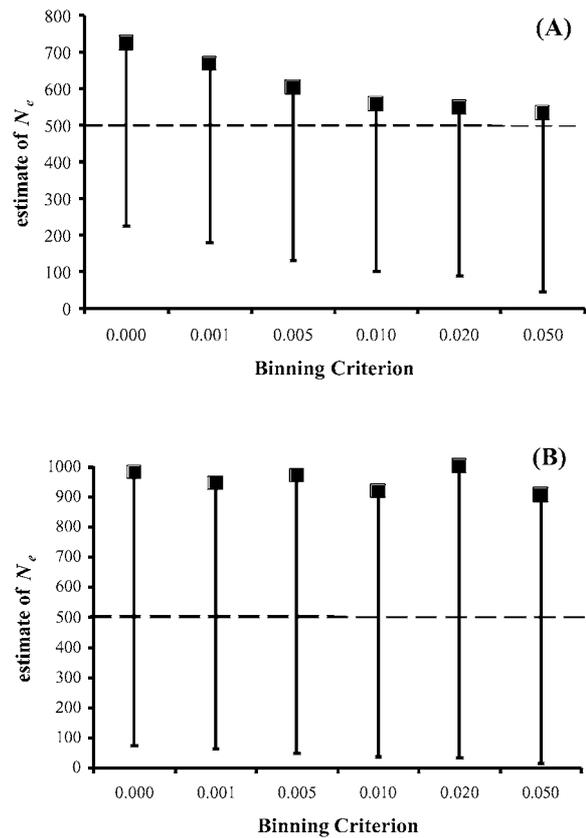


Figure 3. Effects of binning alleles on accuracy and precision of  $\hat{N}_e$  for the mtDNA data set with  $N_e = 500$ , (A)  $S = 200$ , and (B)  $S = 50$  using Nei and Tajima's (1981) computational method for  $\hat{F}$ . The number of independent alleles was reduced from the original mtDNA data set by summing (binning) haplotypes with frequencies less than or equal to the frequency depicted on the x-axis (Binning criterion) to make a single new allele in the initial gamete pool. Zero indicates that no alleles were summed. The dashed line indicates the true value of  $N_e$  in the simulation. Accuracy is the deviation of  $\hat{N}_e$  (squares) from  $N_e$  (dashed line); precision is the width of the lower bound 95% CI (y-error bar), such that more precise estimates have narrower CIs. The upper bound 95% CI in all cases was infinity.

values ostensibly important for conservation and/or management of a species (Turner et al. 1999). To illustrate, we evaluated the statistical power to distinguish an estimated effective population size of 500 from a benchmark value of  $N_e$  of 100 using red drum mtDNA and microsatellite data. This benchmark value was selected because at  $N_e < 100$  genetic drift is expected to remove significant amounts of genetic diversity each generation (Frankel and Soulé 1981) and spontaneous deleterious mutations are expected to increase risk of population extinction via 'mutational meltdown' (Schultz and Lynch 1997). By calculating  $F$  with Equations 3 or 4 (depending on marker class)

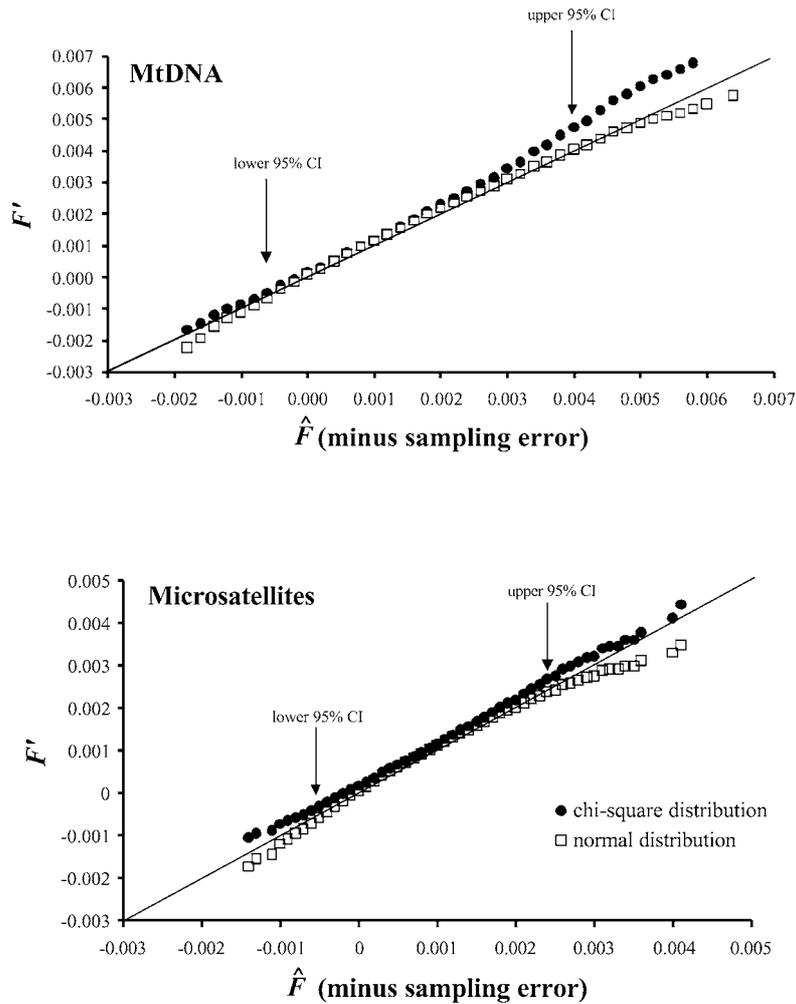


Figure 4. Bivariate plots of  $F'$  expected under chi-square and normal distributions (y-axis), and the distribution of  $\hat{F}$  (minus sampling error) observed in 10,000 simulation replicates (x-axis). Initial conditions of simulation were  $N_e = 500$ ,  $S = 200$ , and the computational method of Nei and Tajima (1981) was used. Similar results were obtained for initial population sizes of  $N_e = 50$  and  $N_e = 100$ , for  $S = 50$ , and using the method of Pollak (1983). The diagonal line represents equality of expected and observed results. Note departures of chi-square and normal approximations at the tails of the distribution. Chi-square values tend to produce slightly wider 95% CIs around estimates of  $\hat{F}$  and  $\hat{N}_e$  in all cases examined in this study.

for  $N_e = 500$ , and then evaluating Equations 5 or 6 at  $\alpha = 0.025$ , it is possible to evaluate the lower bound 95% CI of  $\hat{F}$ . Holding either  $n$  or  $S$  constant allows a determination of the minimum number of independent alleles or sample size needed for the lower bound 95% CI to exceed the benchmark  $N_e$  value of 100.

This exercise revealed expected differences in precision and statistical power between mtDNA and microsatellite markers. For the red drum data set, microsatellite markers required twice the sample size of mtDNA markers to have the same power to distinguish the estimated  $N_e$  from benchmark effective

size when the number of independent alleles is held constant. Likewise, when sample sizes were held constant ( $S = 200$ ), mtDNA required 19 haplotypes to distinguish estimated from benchmark effective size, whereas microsatellites required approximately four independent loci with ten alleles each. In this case, mtDNA markers offered more power for hypothesis testing than microsatellites because the effective size of a haploid, uniparentally inherited marker is one-fourth that of a diploid, biparentally inherited marker. Provided assumptions (Waples 1989) of temporal-method estimation are met, and sample sizes and

the number of alleles are equal, mitochondrial DNA appears to provide more precise, but less accurate estimates of  $N_e$  than microsatellites. However, there are drawbacks to exclusive use of mtDNA data. An important assumption of temporal-method estimation is that allele frequencies have not been influenced by natural selection. Because mtDNA is inherited as a single locus, it is difficult to evaluate whether  $\hat{F}$  based on mtDNA data has been influenced by natural selection. Estimates of  $F$  based on several independent loci may permit evaluation of the action of natural selection by comparing  $\hat{F}$  values among loci using the test proposed by Lewontin and Krakauer (1973). Although this test has been criticized for geographically structured populations (Robertson 1975a, b), it is applicable to variance in  $F$  across loci within a population. Finally, given the same number of total alleles, mtDNA markers are expected to yield less accurate estimates of  $N_e$  because they have a higher proportion of rare alleles than microsatellites. Adding more independent loci in a microsatellite study increases precision without necessarily decreasing accuracy.

Overall, the temporal method of estimating  $N_e$  appeared to perform reasonably well when variance in shifts of allele frequencies were calculated from highly polymorphic loci with large numbers of rare alleles. Examination of the approximation method employed by Waples (1989) to derive the temporal method estimate demonstrates that bias in the approximation is small, unless allele frequencies are very close to 0 or to 1. Equation 7 provides a correction factor to reduce bias when numerous low-frequency alleles are present. Both chi-square and normal approximations appeared suitable for confidence interval estimation and hypothesis testing, although chi-square produced more conservative estimates of 95% CIs. The magnitude of bias is slight, or can be corrected in the cases we examined; thus, the benefit to precision may outweigh the cost to accuracy when employing highly polymorphic loci for temporal-method estimation of  $N_e$ .

### Acknowledgements

We thank E. Bedrick, C. Baer, J. Bielawski, E. Heist, and L. Richardson, and especially J. Wares for helpful discussion and/or comments on the manuscript. J. Wares assisted with simulation analyses. The manuscript was greatly improved by suggestions from

R. Waples and two anonymous reviewers. This study is number XXV in the series 'Genetics Studies in Marine Fishes' and is contribution number 79 from the Center for Biosystematics and Biodiversity at Texas A&M University. Simulations were conducted using the RESAMPLING STATS computer package, 612 N. Jackson St., Arlington, VA 22201 USA. The project was supported by the Texas Agricultural Experiment Station Project (Project H-6703) and by the Advanced Research Program (ARP) of the State of Texas (Project 999902-103).

### Appendix

We give here the details concerning the computation of bias in Equation 7 under Sampling Plan II in Waples (1989). Following the notation in Waples (1989), let  $x$  and  $y$  be the allele frequencies in samples of sizes  $S_0$  and  $S_t$  drawn at generations 0 and  $t$ , respectively. Nei and Tajima (1981) proposed estimation of the temporal variance  $F$  using

$$\hat{F}_C = \frac{1}{K} \sum_{i=1}^K \frac{(x_i - y_i)^2}{(x_i + y_i)/2 - x_i y_i}, \quad (\text{A1})$$

where  $K$  is the number of alleles in the samples. By approximating the expectation of  $\hat{F}_C$ , Waples (1989) showed that

$$E(\hat{F}_C) \approx \frac{1}{2S_0} + \frac{1}{2S_t} + \frac{t}{2N_e}, \quad (\text{A2})$$

when  $t/(2N_e)$  is small, which then allows the effective population size to be estimated by replacing the expectation of  $\hat{F}_C$  with the observed value, as shown in Equation 3 in the text. Waples (1989) obtained the result in Equation A2 using a first order Taylor Series approximation for the expectation of the summand in Equation A1,

$$E\left(\frac{(x - y)^2}{(x + y)/2 - xy}\right) \approx \frac{E((x - y)^2)}{E((x + y)/2 - xy)}. \quad (\text{A3})$$

It is worth noting that the above expression is not a function of the allele frequencies, which allows for a simple estimator of effective population size. We consider approximating the expectation of the ratio on the left hand side of Equation A3 using the second order Taylor Series expansion to determine the effect of allele frequency on the adequacy of the first order approximation. To do this, we note that a second order

approximation of the expectation of the ratio of two random variables  $R$  and  $S$  is given by

$$E\left(\frac{R}{S}\right) \approx \frac{\mu_R}{\mu_S} + \sigma_S^2 \frac{\mu_R}{\mu_S^3} - \frac{\sigma_{RS}}{\mu_S^2}, \quad (\text{A4})$$

where  $\mu_R$  is the mean of  $R$ ,  $\mu_S$  is the mean of  $S$ ,  $\sigma_S^2$  is the variance of  $S$ , and  $\sigma_{RS}$  is the covariance of  $R$  and  $S$  (Rice 1995). Letting  $R = (x - y)^2$  and  $S = \frac{x+y}{2} - xy$  and substituting the appropriate expressions into Equation A4, we have the following general expression for the second order approximation of the expectation,

$$\begin{aligned} E\left(\frac{(x-y)^2}{(x+y)/2 - xy}\right) &\approx \frac{E((x-y)^2)}{E((x+y)/2 - xy)} \\ &+ \text{Var}((x+y)/2 - xy) \frac{E((x-y)^2)}{[E((x+y)/2 - xy)]^3} \\ &- \frac{\text{cov}((x-y)^2, (x+y)/2 - xy)}{[E((x+y)/2 - xy)]^2}. \end{aligned} \quad (\text{A5})$$

We next evaluate the expectation and variance terms in Equation A5 to develop an expression for the expectation as a function of  $S_0$ ,  $S_t$ ,  $N_e$ ,  $P$ , and  $t$ . Note that under Sampling Plan II of Waples (1989) allele frequencies  $x$  and  $y$  are independent (since sampling is before reproduction). Thus, samples at times 0 and  $t$  can be considered to be independent Binomial samples from the pool of gametes preceding generation 0. Under these conditions, several of the necessary expressions have already been obtained by Waples (1989). In particular, Waples showed that

$$\begin{aligned} \text{a) } E((x-y)^2) &= \text{Var}(x-y) = \\ &P(1-P) \left[ \frac{1}{2S_0} + 1 - \left(1 - \frac{1}{2N_e}\right)^t \left(1 - \frac{1}{2S_t}\right) \right] \\ \text{b) } E\left(\frac{x+y}{2} - xy\right) &= P(1-P) \\ \text{c) } \text{Var}(x) &= \frac{P(1-P)}{2S_0} \\ \text{d) } \text{Var}(y) &= P(1-P) \left[ 1 - \left(1 - \frac{1}{2N_e}\right)^t \left(1 - \frac{1}{2S_t}\right) \right] \end{aligned}$$

Expressions (a) and (b) allow for calculation of most of the terms in Equation A5. In addition, it is straightforward to show that

$$\begin{aligned} \text{Var}\left(\frac{x+y}{2} - xy\right) &= \\ &\left(P - \frac{1}{2}\right)^2 (\text{Var}(x) + \text{Var}(y)) \\ &+ \text{Var}(x)\text{Var}(y). \end{aligned} \quad (\text{A6})$$

The remaining term to be computed is the covariance term in Equation A5. After some simplification, this term can be shown to be

$$\begin{aligned} \text{Cov}\left((x-y)^2, \frac{x+y}{2} - xy\right) &= \\ &\left(\frac{1}{2} - P\right) (E(x^3) + E(y^3)) \\ &+ \left(3P^2 - \frac{3}{2}P\right) (\text{Var}(x) + \text{Var}(y)) \\ &+ 2\text{Var}(x)\text{Var}(y) - P^3 + 2P^4. \end{aligned} \quad (\text{A7})$$

The variance terms in Equations A6 and A7 can be evaluated using (c) and (d) above. To complete the computation of the covariance, we provide expressions for the expectation of the cubes of the allele frequencies. First, note that under Sampling Plan II, the  $2S_0$  genes sampled at time 0 follow a Binomial distribution, and hence

$$\begin{aligned} E(x^3) &= \frac{1}{8S_0^3} [2S_0(2S_0 - 1)(2S_0 - 2)P^3 \\ &+ 6S_0(2S_0 - 1)P^2 + 2S_0P]. \end{aligned} \quad (\text{A8})$$

Finding the expectation of  $y^3$  is more difficult because we must take into account genetic drift between generations. If we let  $P_t$  be the allele frequency in the gamete pool from which the sample at generation  $t$  is drawn, the expectation of  $y^3$  given  $P_t$  is

$$\begin{aligned} E(y^3|P_t) &= \frac{1}{8S_t^3} [2S_t(2S_t - 1)(2S_t - 2)P_t^3 \\ &+ 6S_t(2S_t - 1)P_t^2 + 2S_tP_t]. \end{aligned} \quad (\text{A9})$$

Taking the expectation of Equation A9 with respect to  $P_t$  will then give the unconditional expectation of  $y^3$ , i.e.

$$\begin{aligned} E(y^3) &= E(E(y^3|P_t)) = \\ &\frac{1}{8S_t^3} [2S_t(2S_t - 1)(2S_t - 2)E(P_t^3) \\ &+ 6S_t(2S_t - 1)E(P_t^2) + 2S_tE(P_t)]. \end{aligned} \quad (\text{A10})$$

To evaluate Equation A10, we note that conditional expectation can be further used to show that  $E(P_t) = P$ , and

$$E(P_t^2) = \left(1 - \frac{1}{2N_e}\right)^t P^2 + \left[1 - \left(1 - \frac{1}{2N_e}\right)^t\right] P, \quad (\text{A11})$$

and finally that

$$E(P_t^3) = \frac{(2N_e - 1)^t (2N_e - 2)^t}{(2N_e)^{2t}} P^3 + \left[ \sum_{j=1}^t \left( \frac{(2N_e - 1)^{j-1} (2N_e - 2)^{j-1}}{(2N_e)^{2(j-1)}} \right) \left( \frac{6N_e(2N_e - 1)}{8N_e^3} \right) \right] P^2 + \left[ \frac{1}{4N_e^2} \left( \sum_{j=1}^t \frac{(2N_e - 1)^{j-1} (2N_e - 2)^{j-1}}{(2N_e)^{2(j-1)}} \right) + I(t > 1) \sum_{j=2}^t \left( \frac{(2N_e - 1)^{j-2} (2N_e - 2)^{j-2}}{(2N_e)^{2(j-2)}} \right) \right] P, \quad (A12)$$

where  $I(t > 1)$  is an indicator function that is 1 when  $t > 1$  and 0 otherwise. Using Equations A6 through A12, it is possible to write a general expression for the expectation as a function of  $S_0$ ,  $S_t$ ,  $N_e$ ,  $P$ , and  $t$ . Since the expression for  $E(y^3)$  is cumbersome, we have left this term in the expression for the expectation below,

$$E\left(\frac{(x-y)^2}{(x+y)/2-xy}\right) \approx \frac{1}{2S_0} + 1 - \left[ \left(1 - \frac{1}{2N_e}\right)^t \left(1 - \frac{1}{2S_t}\right) \right] + \left(\frac{1}{P(1-P)}\right) \left(P - \frac{1}{2}\right)^2 \left[ \frac{1}{2S_0} + 1 - \left(1 - \frac{1}{2N_e}\right)^t \left(1 - \frac{1}{2S_t}\right) \right]^2 - \left(\frac{1}{P(1-P)}\right) \left(3P^2 - \frac{3}{2}P\right) \left[ \frac{1}{2S_0} + 1 - \left(1 - \frac{1}{2N_e}\right)^t \left(1 - \frac{1}{2S_t}\right) \right] + \left(\frac{1}{2S_0}\right) \left[ 1 - \left(1 - \frac{1}{2N_e}\right)^t \left(1 - \frac{1}{2S_t}\right) \right] \left(\frac{1}{2S_0} - \left[ 1 - \left(1 - \frac{1}{2N_e}\right)^t \left(1 - \frac{1}{2S_t}\right) \right]\right) - \frac{(1/2 - P)}{P^2(1-P)^2} \left(\frac{1}{8S_0^3} \left[ 2S_0(2S_0 - 1)(2S_0 - 2)P^3 + 6S_0(2S_0 - 1)P^2 + 2S_0P \right] + E(y^3)\right) - \frac{2P^4 - P^3}{P^2(1-P)^2}, \quad (A13)$$

where the expectation of  $y^3$  is given in Equation A8. Equation A13 appears to be quite complex, but it is a relatively simple matter to evaluate it for known  $S_0$ ,  $S_t$ ,  $N_e$ ,  $P$ , and  $t$ , which allows for examination of the bias in our simulation studies.

To relate Equation A13 to the temporal method estimate of  $N_e$ , we note that this expression gives the expectation for each term  $i$  in the summand of Equation A1. We then use the approximation Waples (1989) used for the first three terms (first line) in Equation A13, and call the rest of the terms in Equation A13 (the last five lines) the bias for term  $i$ . This

yields the following second order approximation for the expectation of  $\hat{F}_c$ ,

$$E(\hat{F}_c) \approx \frac{1}{2S_0} + \frac{1}{2S_t} + \frac{t}{2N_e} + \frac{1}{K} \sum_{i=1}^K Bias_i(S_0, S_t, N_e, P_i, t). \quad (A14)$$

Finally, letting

$$Bias(S_0, S_t, N_e, P, t) = \frac{1}{K} \sum_{i=1}^K Bias_i(S_0, S_t, N_e, P_i, t), \quad (A15)$$

and solving Equation A13 for  $N_e$  gives the expression in Equation 7 in the text. It is worth noting that the true bias term should include third- through  $n$ th-order terms in the expansion, plus any remainder. However, these terms are likely to be very small. Finally, a more general expression for bias would also include the component that is due to approximating the first three terms in Equation A13 with  $I/(2S_0) + I/(2S_t) + t/(2N_e)$ . While this would not be difficult to do (one simply needs to add to the bias expression the difference between the approximation and the actual term), we have chosen not to include it here for several reasons. The first is that, for the majority of studies, the term  $t/(2N_e)$  will be sufficiently small so that bias due to this approximation will be negligible. Second, our purpose here is to examine bias as a function of allele frequency. Bias introduced by this portion of the approximation is independent of allele frequency, and will not affect the analysis presented here. For example, in the case where  $N_e = 50$ ,  $S = 50$ , and  $t = 1$ , the bias due to this part of the approximation is 0.0001 for all values of  $P$ .

## References

- Avise JC (2000) *Phylogeography*. Sinauer Associates, Sunderland, MA.
- Brooker AL, Cook D, Bentzen P, Wright JM, Doyle RW (1994) Organization of microsatellites differs between mammals and cold-water teleost fishes. *Can. J. Fish. Aquat. Sci.*, **51**, 1959–1966.
- Broughton RE, Gold JR (1997) Microsatellite development and survey of variation in northern bluefin tuna (*Thunnus thynnus*). *Mol. Mar. Biol. Biotech.*, **6**, 308–314.
- Brown WM (1983) Evolution of animal mitochondrial DNA. In: *Evolution of Genes and Proteins* (eds. Nei M, Koehn RK), pp. 62–88. Sinauer, Sunderland, Mass.

- Chakraborty R, Fuerst PA, Nei M (1980) Statistical studies on protein polymorphism in natural populations. III. Distribution of allele frequencies and the number of alleles per locus. *Genetics*, **94**, 1039–1063.
- Chambers SM (1995) Spatial structure, genetic variation, and the neighborhood adjustment to effective population size. *Conservation Biology*, **9**, 1312–1315.
- Crow JF (1986) *Basic concepts in population, quantitative, and evolutionary genetics*. W.H. Freeman and Co., New York.
- Donnelly P, Tavaré S (1995) Coalescents and genealogical structure under neutrality. *Ann. Rev. Genetics*, **29**, 401–421.
- DeWoody JA, Avise JC (2000) Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *J. Fish Biol.*, **56**, 461–473.
- Ewens WJ (1972) The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.*, **3**, 87–112.
- Frankel OH, Soulé ME (1981) *Conservation and Evolution*. Cambridge University Press, Cambridge.
- Frankham R (1995) Conservation genetics. *Ann. Rev. Genet.*, **29**, 305–327.
- Gold JR, Richardson LR, Turner TF (1999) Temporal stability and spatial divergence of mitochondrial DNA haplotype frequencies in red drum (*Sciaenops ocellatus*) from coastal regions of the western Atlantic Ocean and Gulf of Mexico. *Mar. Biol.*, **133**, 593–602.
- Hedgecock D, Chow V, Waples RS (1992) Effective population numbers of shellfish broodstocks estimated from temporal variance in allelic frequencies. *Aquaculture*, **108**, 215–232.
- Husband BC, Barrett SCH (1992) Effective population size and genetic drift in tristylous *Eichornia paniculata* (Pontederiaceae). *Evolution*, **46**, 1875–1890.
- Jarne P, Lagoda PJJ (1996) Microsatellites, from molecules to populations and back. *Trends Ecol. Evol.*, **11**, 424–429.
- Jorde PE, Ryman N (1996) Demographic genetics of brown trout (*Salmo trutta*) and estimation of effective population size from temporal change in allele frequencies. *Genetics*, **143**, 1369–1381.
- Laikre L, Jorde PE, Ryman N (1998) Temporal change of mtDNA haplotype frequencies and female effective size in a brown trout population. *Evolution*, **52**, 910–915.
- Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.
- Luikart G, Cornuet JM, Allendorf FW (1999) Temporal changes in allele frequencies provide estimates of population bottleneck size. *Conserv. Biol.*, **13**, 523–530.
- Miller LM, Kapuscinski AR (1997) Historical analysis of genetic variation reveals low effective population size in a northern pike (*Esox lucius*) population. *Genetics*, **147**, 1249–1258.
- Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York.
- Nei M, Tajima F (1981) Genetic drift and the estimation of effective population size. *Genetics*, **98**, 625–640.
- Pollak E (1983) A new method for estimating the effective population size from allele frequency changes. *Genetics*, **104**, 531–548.
- Rice JA (1995) *Mathematical Statistics and Data Analysis, 2nd Ed.* Duxbury Press, Belmont, California.
- Richards C, Leberg PL (1996) Temporal changes in allele frequencies and a population's history of severe bottle necks. *Cons. Biol.*, **10**, 832–839.
- Robertson A (1975a) Remarks on the Lewontin-Krakauer test. *Genetics*, **80**, 396.
- Robertson A (1975b) Gene frequency distributions as a test of selective neutrality. *Genetics*, **81**, 775–785.
- Schultz ST, Lynch M (1997) Mutation and extinction: the role of variable mutation effects synergistic epistasis, beneficial mutations, and degree of outcrossing. *Evolution*, **51**, 1363–1371.
- Turner TF, Richardson LR, Gold JR (1998) Polymorphic microsatellite DNA markers in red drum (*Sciaenops ocellatus*). *Mol. Ecol.*, **7**, 1771–1773.
- Turner TF, Richardson LR, Gold JR (1999) Temporal genetic variation of mtDNA and effective female population size of red drum in the northern Gulf of Mexico. *Mol. Ecol.*, **8**, 1223–1230.
- Waples RS (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics*, **121**, 379–391.
- Williamson EG, and Slatkin M (1999) Using maximum-likelihood to estimate population size from temporal changes in allele frequencies. *Genetics*, **152**, 755–761.