

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Publications, Agencies and Staff of the U.S.
Department of Commerce

U.S. Department of Commerce

2005

Genetic estimates of contemporary effective population size: to what time periods do the estimates apply?

Robin Waples

NOAA, robin.waples@noaa.gov

Follow this and additional works at: <https://digitalcommons.unl.edu/usdeptcommercepub>

Waples, Robin, "Genetic estimates of contemporary effective population size: to what time periods do the estimates apply?" (2005). *Publications, Agencies and Staff of the U.S. Department of Commerce*. 462.
<https://digitalcommons.unl.edu/usdeptcommercepub/462>

This Article is brought to you for free and open access by the U.S. Department of Commerce at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Publications, Agencies and Staff of the U.S. Department of Commerce by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Genetic estimates of contemporary effective population size: to what time periods do the estimates apply?

ROBIN S. WAPLES

Northwest Fisheries Science Center, 2725 Montlake Blvd. East, Seattle, WA 98112 USA and Laboratoire d'Ecologie Alpine (LECA), Génomique des Populations et Biodiversité, Université Joseph Fourier, Grenoble, France

Abstract

Although most genetic estimates of contemporary effective population size (N_e) are based on models that assume N_e is constant, in real populations N_e changes (often dramatically) over time, and estimates (\hat{N}_e) will be influenced by N_e in specific generations. In such cases, it is important to properly match \hat{N}_e to the appropriate time periods (for example, in computing N_e/N ratios). Here I consider this problem for semelparous species with two life histories (discrete generations and variable age at maturity — the 'salmon' model), for two different sampling plans, and for estimators based on single samples (linkage disequilibrium, heterozygote excess) and two samples (temporal method). Results include the following. *Discrete generations*: (i) Temporal samples from generations 0 and t estimate the harmonic mean N_e in generations 0 through $t - 1$ but do not provide information about N_e in generation t ; (ii) Single samples provide an estimate of N_e in the parental generation, not the generation sampled; (iii) single-sample and temporal estimates never provide information about N_e in exactly the same generations; (iv) Recent bottlenecks can downwardly bias estimates based on linkage disequilibrium for several generations. *Salmon model*: (i) A pair of single-cohort (typically juvenile) samples from years 0 and t provide a temporal estimate of the harmonic mean of the effective numbers of breeders in the two parental years ($N_{b(0)}$ and $N_{b(t)}$), but adult samples are more difficult to interpret because they are influenced by N_b in a number of previous years; (ii) For single-cohort samples, both one-sample and temporal methods provide estimates of N_b in the same years (contrast with results for discrete generation model); (iii) Residual linkage disequilibrium associated with past population size will not affect single-sample estimates of N_b as much as in the discrete generation model because the disequilibrium diffuses among different years of breeders. These results lead to some general conclusions about genetic estimates of N_e in iteroparous species with overlapping generations and identify areas in need of further research.

Keywords: discrete generations, heterozygote excess, linkage disequilibrium, N_e/N ratio, Pacific salmon, temporal method

Received 22 January 2005; revision accepted 15 June 2005

Introduction

Effective population size (N_e) is a critical parameter in evolutionary biology, but it is difficult to collect enough demographic data from most natural populations to calculate N_e directly. As a consequence, genetic methods for estimating N_e have been used widely by both evolutionary biologists and conservation biologists. For some

N_e estimators, performance measures such as bias and precision have been evaluated (Nei & Tajima 1981; Waples 1989; Pudovkin *et al.* 1996; Wang 2001). However, one fundamental problem has received little attention: most estimators of contemporary N_e are based on models in which N_e is assumed (either implicitly or explicitly) to be constant, but this is highly unlikely in natural populations. These estimates apply to a specific generation(s) or time period(s), and if N_e varies it is important (for two reasons) to determine to which time period(s) the estimate applies. First, several different genetic methods for estimating N_e

Correspondence: R. S. Waples, Fax: 206 860-3335; E-mail: robin.waples@noaa.gov

are available (see below), and in comparing results the time periods relevant to each estimate should be congruent.

Second, if one wants to estimate the ratio N_e/N , failure to properly match the N_e estimate with the appropriate estimate of census population size (N) can result in biologically misleading conclusions. A consensus has not emerged in the scientific literature regarding the 'typical' range of N_e/N values in natural populations (Nunney 1993; Hedgcock 1994; Frankham 1995; Scribner *et al.* 1997; Hauser *et al.* 2002). This remains an active area of research, and genetically based estimates of N_e/N are regularly used to draw conclusions on topics such as mating systems and reproductive success (Kaeuffer *et al.* 2004; Matocq 2004); N_e/N ratios in different taxa (Turner *et al.* 2002; Hoffman *et al.* 2004); patterns of change in N_e/N over time (Ardren & Kapuscinski 2003; Shrimpton & Heath 2003); and appropriate conservation and management actions (Hoffman & Blouin 2004; Johnson *et al.* 2004).

To illustrate some of the implications of properly matching time periods for \hat{N}_e and N , consider the hypothetical situation depicted in Fig. 1: genetic samples are taken in two consecutive generations, and the temporal method is used to estimate N_e , yielding (in this example) $\hat{N}_e = 100$. Does this estimate apply to the effective size in the first genera-

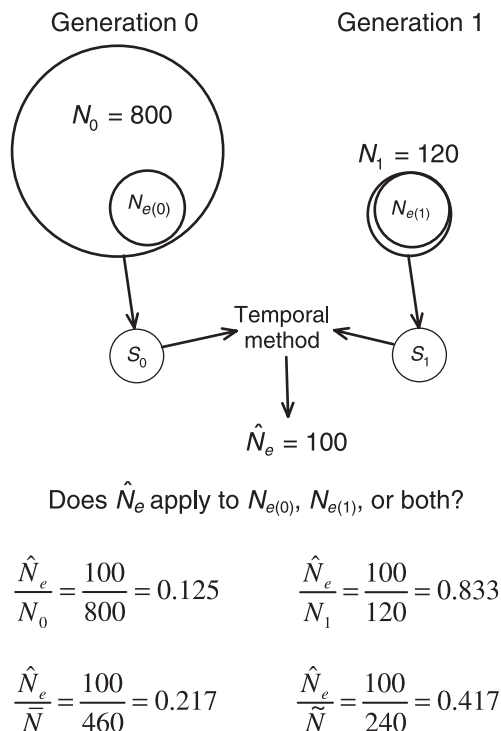


Fig. 1 Schematic diagram of estimation of effective population size based on samples (S_0 , S_1) taken in consecutive generations. In this hypothetical example, census size (N) in the two generations is 800 and 120, and the temporal method yields an estimate $\hat{N}_e = 100$. The estimate of N_e/N varies greatly depending on which generation (or combination of generations) \hat{N}_e is applied to.

tion, the second generation, or some combination (harmonic or arithmetic mean) of the two? In this example, the estimate of N_e/N varies almost sevenfold depending on the generation or combination of generations to which \hat{N}_e is applied. Note also that both samples could be used individually to estimate N_e based on either of the one-sample methods discussed below. To what time period(s) would those estimates of N_e apply, and how would they compare with the time period(s) covered by the temporal estimate?

The issue is complex because results depend on the life history of the species in question, the method used to estimate \hat{N}_e , and the sampling regime. Here I consider this topic for estimates of contemporary N_e based on single samples (linkage disequilibrium, heterozygote excess) and two samples (temporal changes in allele frequency); for typical sampling regimes (before vs. after reproduction; single- vs. multiple-cohort); and for semelparous species with two different life histories (discrete generations and variable age at maturity). Pacific salmon (*Oncorhynchus* spp.) are perhaps the best-known species that exhibit semelparity with variable age at maturity, but this type of life history is also found in a variety of other taxa. Although discrete generations occur in relatively few species, discrete generation models have been widely used to estimate N_e in species with overlapping generations. Finally, I will comment on the relevance of the results of the semelparous models to iteroparous, age-structured species.

Methods

Effective population size

Random genetic processes occur at a rate inversely related to population size. However, it is not the census size that determines the rate of these random processes, but rather the population's effective size. N_e is the size of an 'ideal' population (Wright 1931) that would have the same rate of genetic change as the population under consideration. In an ideal population (one in which mating is random, sex ratio is equal, generations are discrete, and variation in reproductive success is random), N and N_e are the same, but that is rarely the case in nature. In addition to directly affecting rates of genetic change, effective size also determines the relative importance of migration and selection; these forces are deterministic in large populations but can be overwhelmed by random processes in small ones. The models discussed here all assume that populations are closed to migration and the markers considered are selectively neutral. Selection or migration could bias estimates of N_e and, under some circumstances, could also affect the time periods to which estimates of effective size apply.

Several effective sizes have been identified in the literature (Ewens 1979; Crow & Denniston 1988; Caballero 1994);

Table 1 Notation used

A_j	Fraction of breeders that mature at age j ; $\sum A_j = 1$
g	Generation length in years; $g = \sum jA_j$
t	Elapsed time between two samples, scaled in years in the salmon model and in generations in the discrete generation model
N_i	Number of individuals in the population in time period i
$N_{e(i)}$	Effective population size in generation i
$N_{b(i)}$	Effective number of breeders in time period i
\hat{N}_e, \hat{N}_b	An estimate of N_e or N_b
$\bar{N}_e (\bar{N}_b)$	Harmonic mean effective population size (effective number of breeders) over a period of time
P_i	Frequency of an allele in $2N_{e(i)}$ genes representing effective population size in generation i or in $2N_{b(i)}$ genes representing effective breeders in time period i
S_i	Number of individuals sampled at time period i
$X (Y)$	Frequency of an allele in $2S_0$ ($2S_i$) genes in a sample taken in time period 0 (t)
$V(X - Y)$	Variance of the difference in allele frequency between two samples
F	Standardized variance of allele frequency change; $F = V(X - Y)/[P(1 - P)]$
r^2	Squared correlation of allele frequencies at different gene loci
\bar{r}^2	Asymptotic r^2 reached in a population of constant size

here we will be concerned with the two most commonly used: variance effective size (related to the rate of allele frequency change) and inbreeding effective size (related to the rate of increase in inbreeding). Effective size can be thought of as applying to either a single generation or to multiple generations. A single-generation N_e is associated with transmission of genetic material from one generation to the next. Kimura & Crow (1963) showed that inbreeding N_e pertains to the number in the parental generation, whereas variance N_e pertains to the number in the progeny generation. However, as discussed below, this conclusion applies to population parameters; genetic estimates of N_e depend on samples, and the sampling process can affect the time period(s) to which the estimates apply. This paper will focus on estimates of contemporary N_e — that is, one or more generations that span the time frame represented by the samples. I do not consider estimates of long-term effective size (which generally depend on equilibrium relationships between N_e and measures of genetic diversity) or methods specifically designed to detect bottlenecks; however, I do evaluate effects of recent changes in N_e that can affect contemporary estimates of effective size.

Here we will be concerned with genetic data for two samples taken in different time periods (samples S_0 and S_i ; see Table 1 for notation). In the discrete generation model, the time periods are generations; in the variable age at maturity model (henceforth known as the 'salmon model'), the time periods are referred to as years. Two samples provide a basis for using the temporal method to estimate N_e ; depending on the species' life history, the estimate applies either to the effective population size per generation (N_e) or the effective number of breeders per year (N_b). In addition, each of the two samples can be analysed separately for information about effective size. The analyses below assume the species is diploid, but simple modifications could be made for haploid or polyploid species.

Population genetic models

Discrete generations. The model used here (Fig. 2) is similar to that of Waples (1989), the only important difference being that here we want to distinguish effective sizes that may differ among generations. In each generation i , there are N_i adult breeders having effective size $N_{e(i)}$, which (to be completely general) can be thought of as representing $2N_{e(i)}$ genes (having frequency P_i) rather than a particular number of 'ideal' individuals. The $N_{e(i)}$ effective breeders produce a large pool of gametes from which individuals in the next generation are drawn. The gamete pool is assumed to be large enough that its allele frequency is also P_i . Drift occurs in choosing the next generation's total population (N_{i+1}) and effective population ($N_{e(i+1)}$) from the gamete pool.

Salmon model. Species with this life history reproduce in only one time period but can mature at a variety of ages j with specified probability A_j ($\sum A_j = 1$); as a consequence, the population allele frequency in time period i depends not only on the frequency in time period $i - 1$, but also on frequencies in one or more previous time periods. For salmon, the relevant time period is 1 year, so t represents the elapsed number of years between samples. A model of genetic change that incorporates these life history features is shown in Fig. 3. Each large circle represents the N_i adults making up the pool of potential breeders in 1 year. Smaller circles each year are the effective number of breeders, which can be thought of as representing $2N_{b(i)}$ genes that are the equivalent of $N_{b(i)}$ ideal individuals. In this example, adults can mature at ages $j = 3, 4$, or 5 . Thus (for example), adults in year 0 are derived from breeders in years $-3, -4$, and -5 , and adults in year 2 are derived from breeders in years $-1, -2$, and -3 . Breeders in years 0 and 2 share one common source — breeders in year -3 — but the degree of overlap could

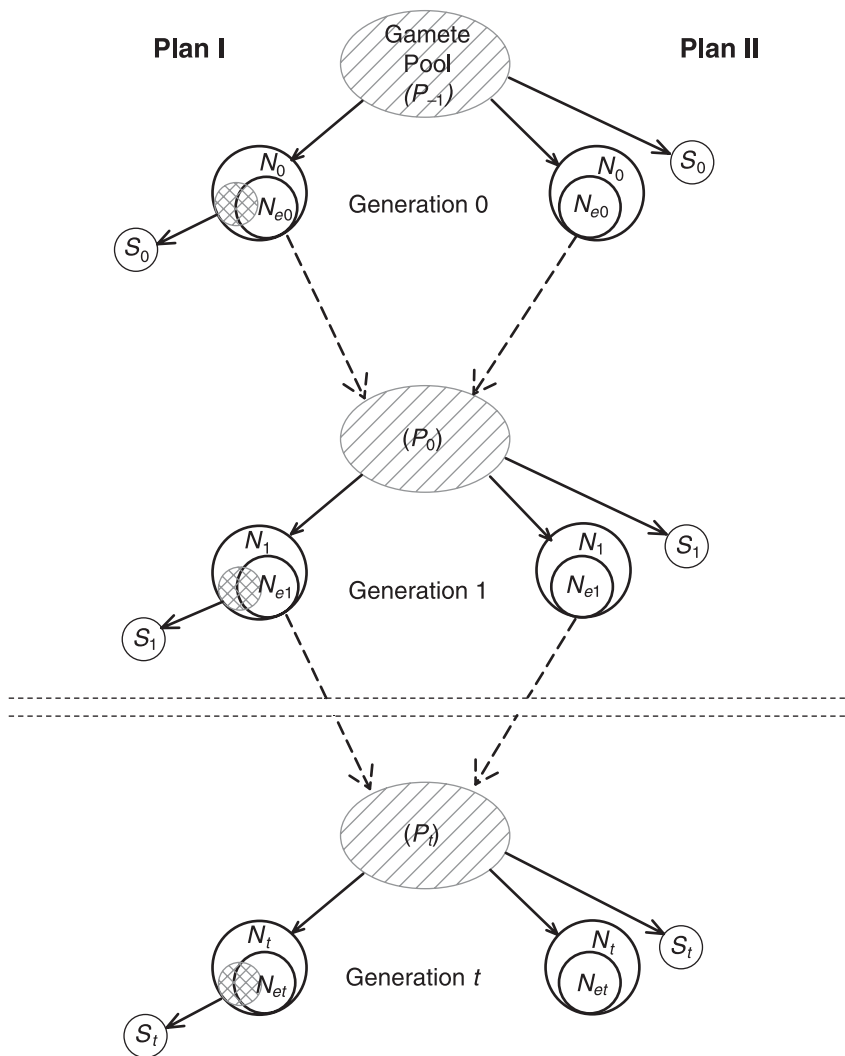


Fig. 2 Model of genetic change over time and sampling from a population with discrete generations. Point of reference for all genetic changes is the allele frequency (P_{-1}) in the gamete pool preceding generation 0. N_i represents the census population size in generation i , and the $2N_{e(i)}$ genes representing the effective population size in generation i produce the gamete pool from which the next generation is formed. Samples (of S_i individuals) for genetic analysis are taken at two or more points in time according to two sampling plans. In Plan I, individuals (typically adults) are sampled after reproduction or nonlethally before reproduction, so that some individuals in the sample may also contribute to future generations (potential overlap indicated by shaded circle). In Plan II, individuals (typically juveniles) are sampled without replacement before reproduction, so there is no overlap between individuals comprising the sample and the census or effective population size in that generation. By convention (Nei & Tajima 1981; Waples 1989), both Plan I and Plan II samples are associated with the generation in which the sampled individuals would mature.

be more or less depending on the timing of the samples and the age structure.

Waples (1990a, b) showed that in this model, effective size per generation (N_e) is approximately equal to gN_b , where g is the generation length (average age of parents = average age at maturity weighted by age-specific fecundity). When N_b varies, $N_e \approx g\bar{N}_b$, where \bar{N}_b is the harmonic mean N_b across the years making up the generation (Waples 1990b, 2002). The model used here is similar, except that care will be taken to identify which specific years most strongly influence the estimate of N_b .

N_e estimators

Single-sample estimates. Single-sample estimators of N_e depend on departures from one- or two-locus genetic equilibrium that arise when progeny are derived from a finite number of parents. These samples provide an estimate of the effective number of parents that produced the

progeny from which the sample is drawn; hence, they can be associated with the inbreeding effective size (Laurie-Ahlberg & Weir 1979). In the discrete generation model, the sample can be thought of as being produced by the effective population size of the parental generation. In the salmon model, the single-sample estimates apply to a particular number of breeders that generally do not represent a generation.

The heterozygote excess method (Pudovkin *et al.* 1996) measures increases in the observed proportion of heterozygotes (in comparison with expected Hardy-Weinberg proportions) due to random differences in allele frequency between males and females. Because a single generation of random mating is sufficient to restore HW proportions, the heterozygote excess method provides an estimate of the effective number of parents that produced the sample — that is, N_e (or N_b) in the parental generation (or year).

The linkage disequilibrium method, which measures departures of digenic gametic frequencies from those

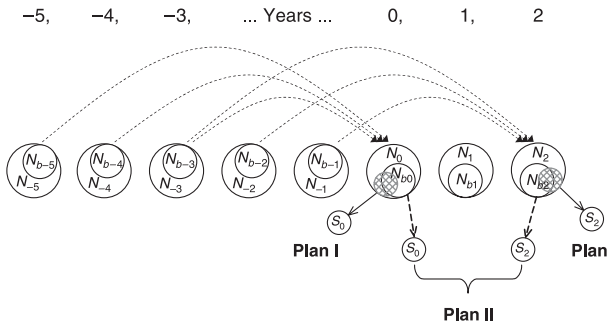


Fig. 3 Model of genetic change over time and sampling from a semelparous population with variable age at maturity (the salmon model). Large circles (labelled N_t) represent the breeding population in successive time periods (in this example, years); smaller nested circles (labelled $N_{b(i)}$) represent the effective number of breeders in year i . No individual is a breeder in more than 1 year. Because age at maturity varies, the breeders in year i represent progeny of parents from more than one previous year. In this example, the breeders in years 0 and 2 are both derived from parents that reproduced 3, 4, and 5 years before (as indicated by curved, dashed lines with arrows). As in Figure 2, samples are taken according to two different plans. Plan I samples (typically adults) include individuals from more than one cohort. Plan II samples (typically juveniles) for year i are from a single cohort that represents progeny of adults reproducing in year i .

expected based on random mating and binomial sampling, depends on chance associations of alleles at different gene loci. Unlike single-locus HW departures, linkage disequilibria do not disappear after a single generation of random mating; instead, they decay at an exponential rate determined by the recombination fraction, c ($0 \leq c \leq 0.5$). In a population of constant size, an asymptotic level of disequilibria due to drift is eventually reached, in which decay of existing disequilibria is balanced by random creation of new disequilibria each generation. For neutral gene loci that are unlinked ($c = 0.5$), the asymptotic linkage disequilibrium is approximated by

$$E(\hat{r}^2) \approx \frac{1}{3N_e} + \frac{1}{S} \quad (\text{Hill 1981}), \quad (\text{eqn 1})$$

where \hat{r}^2 is an estimate (from a sample of S individuals) of the mean squared correlation of allele frequencies at different gene loci. The first term in equation 1 is disequilibrium due to drift; the second is due to sampling a finite number of progeny for genetic analysis. This equation can be rearranged to provide an estimator of N_e (Weir & Hill 1980; Waples 1991):

$$\hat{N}_e = \frac{1}{3(\hat{r}^2 - 1/S)}. \quad (\text{eqn 2})$$

Equation 1 assumes that second order terms in S and N_e can be ignored (Weir & Hill 1980), and as a result equation

2 is biased unless $S \approx 2N_e$ (R. Waples & P. England, unpublished). Although a correction can be applied to remove most of the bias (R. Waples, unpublished), the bias is not substantial enough to affect conclusions of this paper, so equations 1 and 2 are used in the examples below. The above formula all assume that linkage disequilibrium is estimated from genotypic data and mating is random without permanent pair bonds; see Weir & Hill (1980) for adjustments for other mating systems and for haplotype data.

Temporal estimates. Temporal estimates depend on the premise that the rate of genetic change over time is a function of N_e ; this method thus provides an estimate of the variance effective population size. If X and Y are allele frequencies in samples S_0 and S_t , the quantity $V(X - Y)$ can be expressed as $V(X) + V(Y) - 2 \text{cov}(X, Y)$, where V denotes a variance and COV a covariance (Nei & Tajima 1981). $V(X - Y)$ is of interest because the standardized variance, $F = V(X - Y)/P(1 - P)$, can be expressed as a function of N_e and the time between samples:

$$E(\hat{F}) \approx \frac{t}{2N_e} + \frac{1}{\bar{S}},$$

where \bar{S} is the harmonic mean of S_0 and S_t (Waples 1989) and t is the number of generations between samples. This leads to an estimator of N_e as:

$$\hat{N}_e = \frac{t}{2(\hat{F} - 1/\bar{S})}. \quad (\text{eqn 3})$$

Equation 3 assumes Plan II sampling (see next section); a slight modification is needed if sampling is according to Plan I.

For species with salmon life history, the analogue to equation 3 is (Waples 1990a)

$$\hat{N}_b = \frac{b}{2(\hat{F} - 1/\bar{S})}, \quad (\text{eqn 4})$$

where b is a constant that varies according to age structure and the number of years between samples. Waples (1990a) obtained values of b via simulation, but Tajima (1992) showed that they can be derived analytically.

For both the discrete generation and salmon models, it will be necessary to revisit previous derivations of \hat{F} to determine which generations or years of genetic drift contribute to \hat{F} and hence to the estimate of N_e (or N_b).

Sampling

Two sampling plans have been identified in the literature for each life history model (Figs 2 and 3; Table 2). Although the names of the plans (Plans I and II) are the same in both models, some important differences will be noted briefly here; more details appear below.

Table 2 Features of two sampling plans discussed in the text

Model	Plan I	Plan II
Discrete generations		
Features	Sampling before reproduction, with replacement; or sampling after reproduction	Sampling before reproduction, without replacement
Examples	Adults sampled after mating Pre-mating adults sampled nonlethally	Lethal sample of juveniles
Salmon model		
Features	Sample contains > 1 cohort	Sample consists of a single cohort
Examples	Adult spawners Mixed-age juveniles	Same-age juveniles Adults grouped by cohort

Discrete generations (Nei & Tajima 1981; Waples 1989; Fig. 2). In sampling Plan I, individuals are sampled after reproduction or nonlethally (with replacement) before reproduction. This type of sample thus could include individuals that also contribute offspring to the next generation. In Plan II, samples are taken without replacement before they can reproduce and do not contribute to any future generations. The distinction is important because in Plan I the allele frequencies in the two temporal samples are positively correlated to the extent that some genes appear both in S_0 and $N_{e(0)}$ (Nei & Tajima 1981). Plan II samples are generally juveniles and Plan I sampling typically involves adults; however, with the increasing popularity of nonlethal biopsy sampling for DNA analyses, Plan I would apply to many juvenile samples as well.

Salmon model (Fig. 3). Waples (1990a) used this same 'before'/'after' reproduction framework to define Plan I and Plan II samples in Pacific salmon. However, as discussed in more detail below, with variable age at maturity the distinction between sampling before or after reproduction is less important than it is with discrete generations. In the salmon model, a more important consideration is whether the sample represents a single cohort (progeny of a single year's breeding population) or multiple cohorts. In the following therefore Plan I will refer to samples that include multiple cohorts, and Plan II will refer to samples from a single cohort. In Pacific salmon, juvenile samples are typically Plan II because different cohorts are spatially and temporally isolated, whereas adults samples are typically Plan I because of variable age at maturity. However, exceptions occur in both directions. Mixed-cohort samples of juveniles are common in species such as *Oncorhynchus mykiss*, and these should be analysed according to Plan I. Conversely, it may be possible to partition an adult sample into individual cohorts, in which case each of these reconstructed cohorts can be analysed as a Plan II sample.

Reference generation. For analytical purposes it is essential to clarify which generation or time period a particular sample is associated with, and in this I have followed conventions established by previous authors. In the discrete generation model, Nei & Tajima (1981) identified both Plan I and Plan II samples with the progeny generation — that is, the generation at which the sampled individuals would become mature (Fig. 2). This is logical when generations are discrete, and it is also logical for samples of adult salmon, which can readily be associated with a specific year of spawning. But this framework is problematical when applied to a sample of juvenile salmon, which in general cannot reliably be associated with any single year of breeders in the future. However, a juvenile sample generally can be identified with a single parental population. For this reason, Waples (1990a) considered Plan II salmon samples to be associated with the parental generation — that is, they are progeny of the same year's breeders that an adult sample would be associated with (Fig. 3).

It should be noted that although the reference generation is important to clarify for bookkeeping purposes, choice of the reference generation does not affect the underlying processes of genetic drift that influence F and estimates of effective size. The more important question biologically is, 'What time period(s) affect an estimate of N_e (or N_b) developed using a particular sampling regime?' This question is the focus of the rest of this paper.

Simulated data

Linkage disequilibrium. Because equations 1 and 2 assume that \hat{r}^2 has reached its asymptotic value, \hat{N}_e can be biased if the population has recently changed in size, and the biases will differ depending on whether the population has recently increased or decreased. Some limited theoretical evaluation has been carried out of the effects of population history on r^2 , but none has considered the consequences for estimates of effective population size. Therefore, computer

simulations were used to model the rate of increase of disequilibrium due to drift in a declining population. I used a Wright–Fisher model (two sexes, random mating, random variation in reproductive success) to track multilocus genotypes for eight independent, diallelic gene loci over a number of discrete generations. Initial genotypes were drawn binomially at each locus with parametric allele frequency 0.5 (equivalent to assuming a population of infinite size in generation 0); each subsequent generation was produced by mating a fixed number of parents (N_e = either 40 or 200). Samples of $S = N_e$ progeny were taken each generation to compute the mean \hat{r}^2 across all pairs of loci (using Weir's 1979 unbiased estimator of the Burrows method). Because the interest here is the general behaviour of \hat{r}^2 as a function of changing population size (rather than on variation in \hat{r}^2 among replicate samples), mean \hat{r}^2 was averaged across a large number of replicates to generate an empirical approximation to $E(\hat{r}^2)$, which was then used to calculate \hat{N}_e using equation 2.

Temporal estimates in the salmon model. To pinpoint specific years to which an estimate of \hat{N}_b applies in the temporal salmon model, I conducted simulations similar to those described by Waples (1990a, b), with an initial population consisting of 5 years with $P = 0.5$ at 20 diallelic loci. Reproduction was Wright–Fisher, so $N_{b(i)} = N_i$ every year, and the age of parents of each progeny produced was determined by random variation around a fixed probability of age at maturity ($A_j, j = 1, 5$). In each replicate, a 'warm-up' period of 40 years was run at a constant (Baseline) N_b (in most cases, Baseline N_b was 100). In year 41, and for 1 year only, the population size was changed to $N_b = N_1$, and at year 41 + j , $j = 1, 5$, the size was changed to $N_b = N_2$. In all other years, effective size was the Baseline N_b . Plan II samples (generally of $S = 50$) were taken annually for 15 years beginning in year 35, and \hat{N}_b was computed for all pairs of samples 1–5 years apart, using equation 4. This format allowed evaluation of the following combinations of true N_b values in the two years sampled: $N_1 \times N_2$, $N_1 \times$ Baseline, $N_2 \times$ Baseline, and Baseline \times Baseline. The latter group was split into comparisons in which both sampled years were prior to the first year of population change ('before'), and those for which at least one of the samples was taken after year 41 ('after'). Within each group of comparisons, a distribution of \hat{N}_b values was generated. For each parameter set, a large number of replicates (10 000) was run to ensure that results reflected real age-structure effects and not random noise.

Results

Discrete generations

Single-sample estimates. As discussed above, in general single-sample estimates provide information about effective

size in the parental generation — that is, the population of breeders that produced the sample. From Fig. 2, it is apparent that both Plan I and Plan II samples in generation i are produced by the effective breeders in generation $i - 1$. Thus, juvenile and adult samples from generation i both provide information about N_e in generation $i - 1$.

When effective size is not constant, the heterozygote excess method will still provide an estimate of $N_{e(i-1)}$, but in the linkage disequilibrium method \hat{r}^2 and \hat{N}_e will also be affected by N_e in generations prior to $N_{e(i-1)}$. Sved (1971; see also Sved & Feldman 1973) provided a formula for evaluating the rate at which drift disequilibria accumulate in a population that initially begins with $r = 0$ (i.e. as would occur in a population of infinite size) and then stabilizes at a constant, finite N_e . This scenario thus represents an extreme example of population decline. If $c = 0.5$, then in generation t the expected disequilibria in a sample can be approximated by:

$$E(\hat{r}_t^2) \approx [1 - (1/4)^t](\bar{r}^2)_{\text{drift}} + \left(\frac{1}{S}\right)_{\text{sample}} \quad (\text{eqn 5})$$

where \bar{r}^2 is the asymptotic r^2 due to drift for the new effective size (first term in equation 1) and $1/S$ is the contribution to $E(\hat{r}^2)$ from sampling (second term in equation 1). Since the term $[1 - (1/4)^t] \rightarrow 1$ very rapidly as t increases (being 0.94 for $t = 2$ and 0.98 for $t = 3$), the drift term rapidly approaches its asymptotic value, meaning that if population size has recently decreased, the disequilibrium method will primarily reflect the new (smaller) N_e rather than N_e in previous generations.

Results for simulated populations under the population decline scenario modelled by Sved are shown in Fig. 4A. With new $N_e = 40$, equation 1 slightly underestimates $E(\hat{r}^2)$, primarily because the term $1/S$ does not completely account for disequilibria due to sampling (R. Waples, unpublished data). However, for both new effective sizes the empirical rate of increase in \hat{r}^2 closely matches the expectation from equation 5. More importantly, the estimate \hat{N}_e is strongly upwardly biased only in the first generation after the population change and quickly stabilizes at approximately the new effective size. Even under this extreme scenario therefore the effects of the historically larger population size on \hat{N}_e do not persist beyond 2–3 generations.

For populations that increase in size, the opposite effect must be considered — decay of existing disequilibria that are higher than the asymptotic value for the new N_e . Assuming again that the loci are unlinked, existing disequilibria decay at a rate of 75% per generation, leading to a rapid convergence rate similar to (but from the opposite direction as) that described by equation 5. At the same time, new drift disequilibria are generated each generation that will converge on the asymptotic value for the new, larger N_e . If we ignore sampling for the moment, the overall effects on drift disequilibria can be approximated using

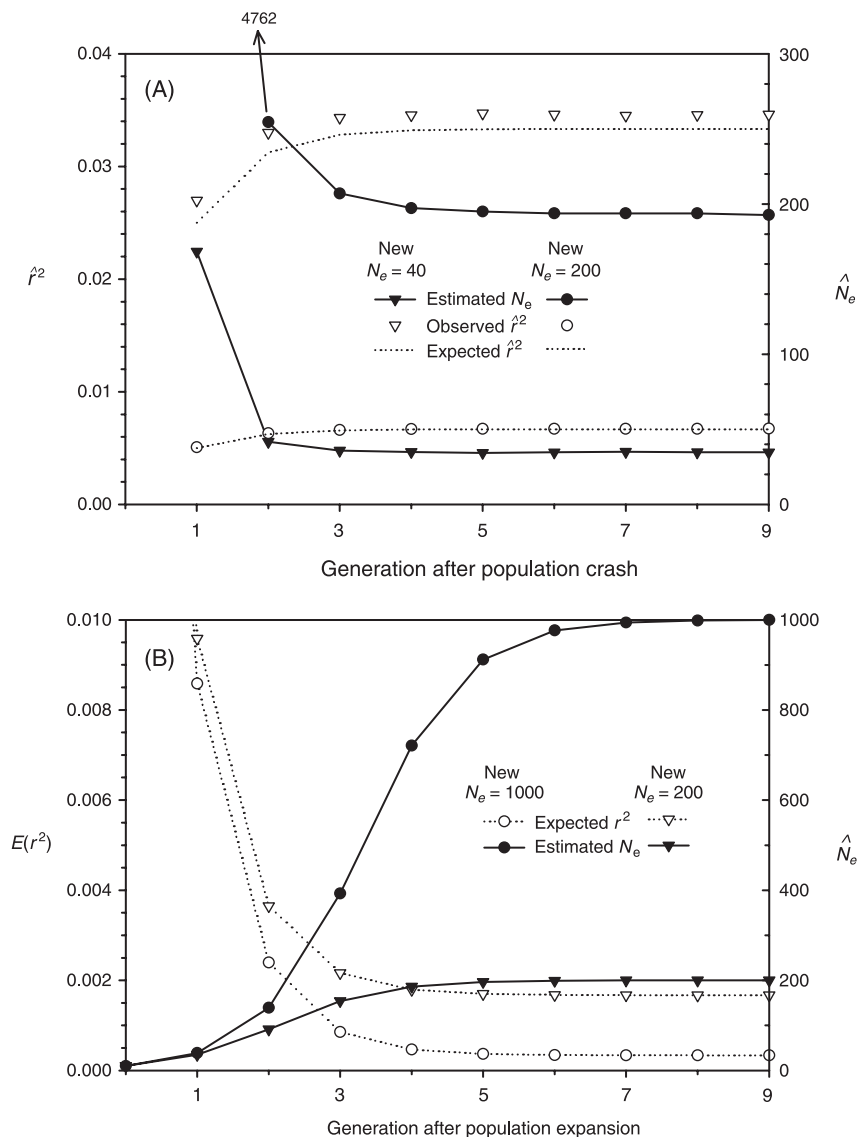


Fig. 4 Rate of approach of linkage disequilibrium to its new asymptotic value after a change in population size, and resulting effects on \hat{N}_e . Discrete generations are assumed. (A) (Population decline; comparison of simulated data with theoretical expectation). The population begins with $r^2 = 0$ in generation 0 and subsequently has constant $N_e = 40$ (triangles) or 200 (circles). Open symbols are empirical \hat{r}^2 from simulations; dotted lines are expected values based on equations 2 and 5. Filled symbols show approach of \hat{N}_e to the new effective size as disequilibria accumulate in the population. (B) (Population expansion; theoretical dependence of \hat{N}_e on rate of decrease in r^2). Results shown are parametric expectations for the population (equivalent to assuming a sample of an infinite number of progeny). The population experiences a bottleneck of $N_e = 10$ ending in generation 0 and increases to constant $N_e = 200$ (triangles) or $N_e = 1000$ (circles) in generations 1–9. Open symbols show expected rate of decrease of r^2 as residual linkage disequilibria break down; filled symbols show convergence of \hat{N}_e (calculated using equation 2 assuming $1/S = 0$) to the new, larger N_e .

a recursive equation to generate $E(r_t^2)$ for a given number of generations (t) following a population increase:

$$E(r_t^2) \approx (1/4)^t (r_0^2) + [1 - (1/4)^t] (\bar{r}^2), \quad (\text{eqn 6})$$

where r_0^2 is disequilibrium at the time of the bottleneck and \bar{r}^2 again is the asymptotic r^2 for the new (larger) effective size. If $r_0^2 = 0$, equation 6 reduces to the drift term in equation 5.

As an example, assume that a population experienced a bottleneck of $N_e = 10$ and subsequently grew in one generation to a new, constant N_e (200 or 1000; Fig. 4B). Assume also that the bottleneck was long enough (at least 2–3 generations; see above) that the drift disequilibrium approached its asymptotic value [$\bar{r}^2 = 1/(3N_e) = 0.033$ for $N_e = 10$; equation 1]. One generation after the bottleneck,

the residual disequilibrium from the small population size is still much higher than the asymptotic values for the new effective sizes (open symbols in Fig. 4B), leading to downward bias in \hat{N}_e (closed symbols). If the new N_e is 200, the estimate \hat{N}_e will be within 20% of the true value by post-bottleneck generation 3. If the new N_e is 1000, however, at generation 3, \hat{N}_e still will be downwardly biased by over 50%, and not until generation 5 will the bias due to the bottleneck drop below 10%. Thus, even with rapid decay the residual disequilibrium from a small bottleneck can affect r^2 and \hat{N}_e for a number of generations. Bias from less severe bottlenecks would not be as substantial for as many generations (Table 3).

We can conclude that population declines are not likely to seriously affect single-sample estimates of contemporary N_e if the effective size in the generation producing the

Table 3 Rate of approach of \hat{N}_e to the new (constant) N_e following bottlenecks of $N_e = 10, 40$, and 100 , with post bottleneck effective sizes (new N_e) = $1000, 200, 100$, and 40 . Values in body of table are the ratio $\hat{N}_e/\text{new } N_e$. \hat{N}_e was calculated from equation 2 based on $E(r^2)$ calculated from equation 6. See Fig. 4B for a plot of some of these data

Gen	New N_e	Bottleneck size								
		$N_e = 10$				$N_e = 40$			$N_e = 100$	
		1000	200	100	40	1000	200	100	1000	200
1		0.04	0.17	0.31	0.57	0.14	0.50	0.73	0.31	0.80
2		0.14	0.46	0.64	0.84	0.40	0.80	0.91	0.64	0.94
3		0.39	0.77	0.88	0.96	0.73	0.94	0.98	0.88	0.98
4		0.72	0.93	0.97	0.99	0.91	0.98	0.99	0.97	1.00
5		0.91	0.98	0.99	1.00	0.98	1.00	1.00	0.99	1.00
6		0.98	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00
7		0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

sample is small, the methods will (on average) detect that small size, regardless whether the population was larger in previous generations. For populations that have recently increased, \hat{N}_e can be biased downwards for several generations, with duration and magnitude of bias proportional to severity of the bottleneck and the relative increase in N_e .

These conclusions apply to unlinked loci. Linked loci have greater precision for estimating N_e provided the recombination fraction c is known (Hill 1981), but interpretation of the estimate is more complicated because historical population size has a stronger effect. If the loci used to calculate \hat{r}^2 are tightly linked, \hat{N}_e can be more strongly affected by N_e in the distant past than by contemporary effective size.

Temporal estimates. A general formula for $V(X - Y)$ for both Plans I and II is (Nei & Tajima 1981; Waples 1989):

$$V(X - Y) = P_{-1}(1 - P_{-1}) \left[\frac{1}{2S_0} + 1 - \left(1 - \frac{1}{2N_e}\right)^t \left(1 - \frac{1}{2S_t}\right) \right] - 2\text{cov}(X, Y) \\ \approx P_{-1}(1 - P_{-1}) \left[\frac{1}{2S_0} + \frac{1}{2S_t} + 1 - \left(1 - \frac{1}{2N_e}\right)^t \right] - 2\text{cov}(X, Y), \quad (\text{eqn 7})$$

where P_{-1} is the allele frequency in the pool of gametes preceding generation 0. The covariance term differs in the two sampling plans. In Plan I, $\text{cov}(X, Y) = P_{-1}(1 - P_{-1})/(2N_0)$ (because both samples are derived from the same N_0 individuals in generation 0), whereas in Plan II the samples at time 0 and time t are independently derived from the initial gamete pool, so $\text{cov}(X, Y) = 0$ (Waples 1989).

If the samples at times 0 and t are taken at the same life stage, they are separated by exactly t generations of genetic drift. On the other hand, the time period from generation 0 through generation t , inclusive, encompasses $t + 1$ genera-

tions of effective sizes, and it is not so obvious exactly which of these time periods affect the estimate of N_e . To pinpoint the generation(s) to which the estimates of N_e apply when effective size varies over time, it is necessary to decompose the term $[1 - 1/(2N_e)]^t$ into a series of t terms of magnitude $1 - 1/(2N_{e(i)})$. It is straightforward but tedious (see details in Appendix) to obtain the following results:

- 1 If one considers the population as a whole, the magnitude of allele frequency change depends on effective population sizes in generations 1 through t but does not depend on $N_{e(0)}$. This illustrates a point made by Kimura & Crow (1963) — the variance effective size is determined by the number in the progeny generation(s).
- 2 When sampling is considered, however, the generations that affect $V(X - Y)$ are not 1 through t but rather 0 through $t - 1$. This result can be understood from examination of Fig. 2. With Plan II sampling, the difference in allele frequencies between the samples at times 0 and t cannot depend on $N_{e(t)}$, since the sample in generation t is taken independently of $N_{e(t)}$ and before any breeding occurs. However, the magnitude of $V(X - Y)$ will depend on $N_{e(0)}$ and all subsequent generations through $t - 1$. Plan I sampling appears more complicated but leads to the same conclusion: the sample in generation t does not provide information about $N_{e(t)}$, even though some individuals in the sample may also be represented in $N_{e(t)}$. For both sampling plans therefore,

$$V(X - Y) \approx P_{-1}(1 - P_{-1}) \left[\frac{1}{2S_0} + \frac{1}{2S_t} + \sum_{i=0}^{t-1} \frac{1}{2N_{e(i)}} \right] - 2\text{cov}(X, Y). \quad (\text{eqn 8})$$

Substituting the $\text{cov}(X, Y)$ terms from above, and noting that $E(\hat{F}) \approx V(X - Y)/P_{-1}(1 - P_{-1})$ and \bar{S} = harmonic mean of S_0 and S_t , leads to the following:

$$\text{Plan I: } E(\hat{F}) \approx \frac{1}{\bar{S}} + \sum_{i=0}^{t-1} \frac{1}{2N_{ei}} - \frac{1}{N_0}; \quad \hat{N}_e = \frac{t}{2(\hat{F} - 1/\bar{S} + 1/N_0)}.$$

$$\text{Plan II: } E(\hat{F}) \approx \frac{1}{\bar{S}} + \sum_{i=0}^{t-1} \frac{1}{2N_{ei}}; \quad \hat{N}_e = \frac{t}{2(\hat{F} - 1/\bar{S})}.$$

Note that $\sum_{i=0}^{t-1} [1/2N_{ei}] = (t)/(2\bar{N}_e)$, demonstrating that $E(\hat{F})$ and \hat{N}_e are a function of the harmonic mean N_e from generations 0 through $t-1$, inclusive. In Plan I sampling, the estimate is also affected by the population size in generation 0 (N_0). Population size in subsequent generations does not affect the estimates under either sampling plan.

Salmon model

Single-sample estimates. Plan II samples associated with year i are progeny of parents reproducing in year i and thus provide information directly about $N_{b(i)}$. As in the discrete generation model, estimates based on linkage disequilibrium can also be affected by N_b in prior years. However, since breeding populations in the salmon model are not connected by a simple Markov chain, the effect is less one of residual disequilibrium persisting through time than of mixture disequilibrium created each year by breeding among individuals of different ages derived from parents with slightly different allele frequencies. Waples & Smouse (1990) evaluated this interyear effect and found that it was generally small compared to disequilibrium due to a mixture of different salmon populations. Unless recent fluctuations in $N_{b(i)}$ have been very extensive, therefore, a Plan II sample in year i should primarily be affected by $N_{b(i)}$.

Plan I samples are more difficult to interpret because they are derived from parents in two or more previous years. Two different factors, opposite in sign and of uncertain magnitude, will tend to bias Plan I estimates of N_b . On the one hand, the total number of parents producing a typical Plan I sample will be larger – perhaps two to three or more times larger – than the number that produces a Plan II (single-cohort) sample. This will tend to upwardly bias a Plan I estimate compared to the effective number of breeders in any single year, toward an estimate that is intermediate between N_b and N_e . On the other hand, a Plan I sample includes individuals from two or more cohorts produced by parents with allele and genotypic frequencies that vary randomly among years. This will produce a kind of Wahlund effect (heterozygote deficiency at single loci and mixture disequilibrium at pairs of loci; Sinnock 1975) that will tend to downwardly bias estimates of N_b . The relative importance of these two effects is likely to vary considerably depending on a number of factors, and until this issue is evaluated quantitatively it is difficult to assess their net effect on \hat{N}_b based on Plan I sampling.

Temporal estimates. Previous results for the salmon model (Waples 1990a, b) were obtained by computer simulations because analytical solutions for $V(X - Y)$ and $E(\hat{F})$ are difficult owing to the complexity of the system. Here, I have extended previous results to obtain approximate analytical solutions, which were then tested with simulated data. To better understand population-level processes, for the moment we will ignore sampling.

In the salmon model, the variance in parametric population allele frequency between years 0 and t is given by

$$V(P_0 - P_t) \approx P(1-P) \left[1 - \left(1 - \frac{1}{2N_e} \right)^k \right] + C \approx P(1-P) \frac{k}{2N_e} + C \quad (\text{eqn 9})$$

(Waples 1990b), where k is the elapsed number of generations ($k = t/g$) and C is a constant. Thus, the rate of allele frequency change in the population is a function of the effective size per generation, just as in the discrete generation model; however, over any given time period the observed variance is elevated by an additional, constant magnitude C (Fig. 5A). This latter effect occurs because a single year of spawners represents only a portion of a generation, and comparing two years of spawners is equivalent to taking two samples from the overall population. Because the value of C depends on ‘sampling’ $2N_b$ genes from the population as a whole to form the breeders in years 0 and t , its value is approximately given by $C \approx P(1-P)/2N_{b(0,t)} + P(1-P)/2N_{b(t)} = P(1-P)/\bar{N}_{b(0,t)}$, where $\bar{N}_{b(0,t)}$ is the harmonic mean N_b in years 0 and t . Thus,

$$V(P_0 - P_t) \approx P(1-P) \left[\frac{k}{2N_e} + \frac{1}{\bar{N}_{b(0,t)}} \right] \quad \text{and} \quad E(F) \approx \frac{k}{2N_e} + \frac{1}{\bar{N}_{b(0,t)}}. \quad (\text{eqn 10})$$

If we note that $N_e = g\bar{N}_b$, where \bar{N}_b is the harmonic mean N_b over the time period under consideration, equation 10 can be expressed as:

$$E(F) \approx \frac{k}{2N_e} + \frac{1}{\bar{N}_{b(0,t)}} = \frac{k}{2g\bar{N}_b} + \frac{1}{\bar{N}_{b(0,t)}}. \quad (\text{eqn 11})$$

Therefore, $E(F)$ in the salmon model is determined by two drift components, one related to a general ‘background’ effective size and one related to N_b in the two years under consideration.

It is useful to examine the relative importance of these two terms. If $k = 2g$, the first term becomes $1/\bar{N}_b$ and two terms will be approximately equal as long as the population is fluctuating around a long-term mean value, so that $\bar{N}_{b(0,t)} \approx \bar{N}_b$. If $g = 4$ years, a common generation length for salmon, the contribution to F from overall population change will be as large as the contribution from N_b in the two sampled years only after approximately $k = 8$ generations (over

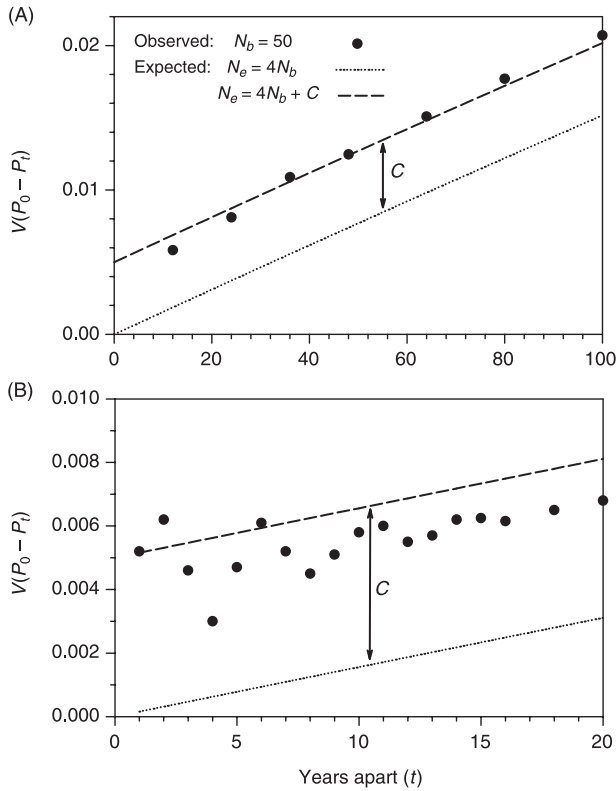


Fig. 5 Temporal variance of allele frequency $V(P_0 - P_t)$ in a simulated Pacific salmon population with $N_b = 50$ each year, $g = 4$ years ($A_j = 0, 0, 0.25, 0.5, 0.25$), and initial allele frequency (P_0) = 0.5. (A) Long-term observed rate of increase of $V(P_0 - P_t)$ (filled circles) is as predicted assuming $N_e = 4N_b$ (dotted line), but absolute magnitude at any point in time is inflated by the approximate amount $C = P_0(1 - P_0)/N_b = 0.25/50 = 0.005$. (B) On shorter time frames (a few generations), yearly fluctuations in $V(P_0 - P_t)$ due to age structure are relatively more important. Modified from Waples (1990a, 2004).

30 years) elapse between the samples. Most temporal comparisons of salmon populations span only 1–2 generations or less, and with $k = 1$ the overall F is strongly dominated by N_b in the two sampled years ($1/\tilde{N}_{b(0,t)}$ vs. $1/(8\tilde{N}_b)$ for the background drift effect). This suggests that temporal estimates spanning up to about two generations in the salmon model primarily reflect the harmonic mean of the effective numbers of breeders in the two years sampled:

$$\hat{N}_b(t < 2g) \approx \tilde{N}_{b(0,t)}. \quad (\text{eqn 12})$$

We expect a priori that this approximation should be best when effective size in the sampled years is small relative to the background $N_{b'}$ in which case the $1/\tilde{N}_{b(0,t)}$ term will be relatively large.

To evaluate the accuracy of equation 12 for typical (short-term) salmon comparisons, I conducted simulations (see Methods) that compared \hat{N}_b with $\tilde{N}_{b(0,t)}$. Results (Fig. 6,

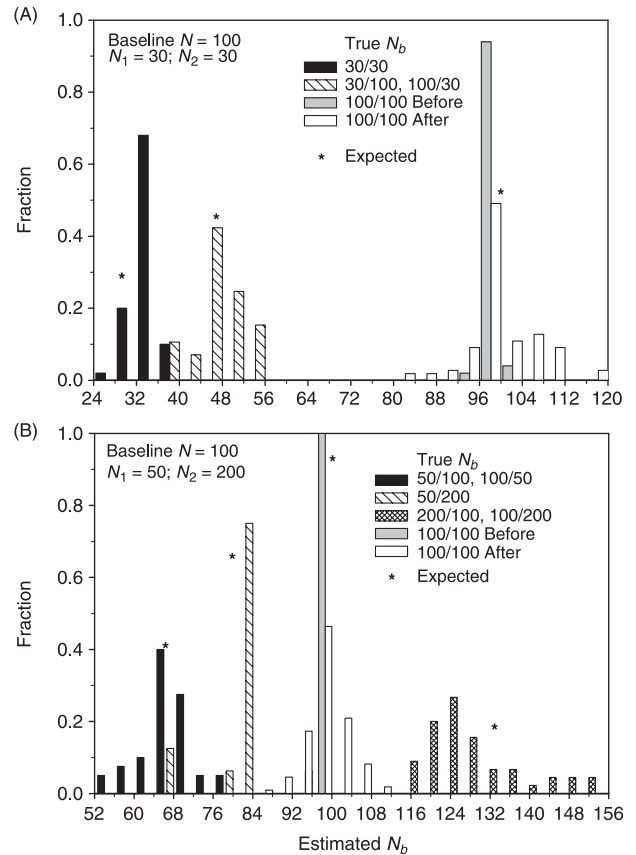


Fig. 6 Temporal estimates of N_b in the salmon model compared with the harmonic mean N_b in the two years sampled (indicated with an *). Simulated populations had a constant population size ($N_b = \text{Baseline } N = 100$) except for 2 years with higher or lower N_b [$N_1, N_2 = 30/30$ (Panel A) or $50/200$ (Panel B)]. Plan II samples were taken each year for 15 years spanning the period of population change. Estimates for which both samples were from years with $N_b = 100$ are split into those for which both samples were 'before' the first year of population change and those for which at least one sample was taken 'after' the population change. Vertical bars show the distribution of N_b estimates based on samples taken 1–5 years apart, with the true N_b in the years of sampling as shown. Sample size was 50 and the fraction maturing at ages $j = 1-5$ was $A_j = 0, 0, 0.25, 0.5, 0.25$.

Table 4) show that under most circumstances \hat{N}_b is close to $\tilde{N}_{b(0,t)}$. For example, with Baseline $N_b = 100$ and $N_1 = N_2 = 30$, median \hat{N}_b for the $N_1 \times N_2$ (30×30) comparisons was 32.8 (Table 4), close to the expected value of 30. Similarly, median \hat{N}_b for $N_1 \times \text{Baseline} = N_2 \times \text{Baseline}$ (30×100) comparisons was 46.8, almost identical to the harmonic mean of 30 and 100 (46.2). For the Baseline \times Baseline comparisons, the different distributions of the 'before' and 'after' estimates illustrate the effects of age structure. The 'before' estimates are very tightly clustered, with 80% falling between 97.4 and 99.5, reflecting the ability of equation 4 to accurately account for age structure in estimating the

Table 4 Temporal estimates of N_b in the salmon model. Simulated populations had a constant population size (N_b = Baseline) except for two years with higher or lower N_b (N_1, N_2). Plan II samples were taken each year for 15 years spanning the period of population change. Each line shows the median and percentiles of N_b estimates based on samples taken 1–5 years apart, with true N_b in the years of sampling as shown (see Fig. 6 for a graphical presentation of data for two of the parameter sets). A (B) indicates that sampling occurred after (before) change in N_b . Expected N_b is the harmonic mean of N_b in the two years of sampling. In the 'standard' parameter set, Baseline N was 100, N_1 and N_2 were 30, sample size was 50, and the fraction maturing at ages $j = 1-5$ was $A_j = 0, 0, 0.25, 0.5, 0.25$. In other simulations, changes from this standard set are shown in bold

Baseline (N_1, N_2)	N_b in years sampled	Estimated N_b			Expected N_b
		Median	Percentiles		
			10%	90%	
100 (30, 30)	30/30	32.8	30.0	36.0	30.0
	30/100	46.8	40.7	53.2	46.2
	100/100 B	98.3	97.4	99.5	100.0
	100/100 A	98.8	97.7	108.1	100.0
$A_j = 0, 0, 0.33, 0.34, 0.33$	30/30	31.4	26.9	32.9	30.0
	30/100	43.5	41.8	49.3	46.2
	100/100 B	98.4	97.3	99.4	100.0
	100/100 A	98.7	92.8	105.7	100.0
$A_j = 0, 0.1, 0.2, 0.3, 0.4$	30/30	30.6	27.3	31.5	30.0
	30/100	45.2	40.4	47.5	46.2
	100/100 B	98.1	97.2	99.2	100.0
	100/100 A	98.2	94.9	107.6	100.0
$A_j = 0, 0, 0.2, 0.6, 0.2$	30/30	35.3	32.9	39.7	30.0
	30/100	50.5	36.0	56.5	46.2
	100/100 B	98.8	97.4	99.8	100.0
	100/100 A	98.6	93.7	108.6	100.0
$A_j = 0, 0, 0.1, 0.8, 0.1$	30/30	47.6	45.5	54.1	30.0
	30/100	64.1	32.5	70.6	46.2
	100/100 B	100.6	99.1	101.8	100.0
	100/100 A	99.2	95.5	105.3	100.0
Sample size = 100	30/30	30.0	32.7	36.1	30.0
	30/100	46.7	40.5	53.0	46.2
	100/100 B	99.3	98.6	99.9	100.0
	100/100 A	99.4	93.9	108.4	100.0
100 (30, 70)	30/70	45.3	41.5	48.8	42.0
	30/100	46.4	37.1	57.9	46.2
	70/100	81.3	78.0	87.2	82.4
	100/100 B	98.5	97.2	99.5	100.0
	100/100 A	98.5	93.9	106.0	100.0
100 (200, 200)	200/200	171.0	157.3	199.4	200.0
	100/200	127.4	120.7	137.0	133.3
	100/100 B	98.5	97.5	99.4	100.0
	100/100 A	97.7	91.6	103.1	100.0
100 (500, 500)	500/500	318.9	257.0	545.0	500.0
	100/500	153.5	136.5	183.4	166.7
	100/100 B	98.7	97.4	99.6	100.0
	100/100 A	98.1	82.1	108.5	100.0
100 (50, 200)	50/200	80.9	74.0	83.3	80.0
	50/100	66.3	57.7	72.0	66.7
	200/100	126.4	120.8	144.6	133.3
	100/100 B	98.4	97.5	99.6	100.0
	100/100 A	98.3	93.2	103.9	100.0
200 (100, 100)	100/100	100.5	87.8	112.8	100.0
	100/200	134.0	122.6	143.4	133.3
	200/200 B	195.8	191.4	199.0	200.0
	200/200 A	196.1	187.0	206.6	200.0

true N_b (100) in populations of constant size. The 'after' estimates are more dispersed, although not biased on average. This modest dispersion (80% of estimates within 8% of the true N_b) reflects the uneven effects of changing size in age-structured populations. All 'after' estimates were potentially affected by low N_b (30) in at least one prior year, but these effects can take many generations to even out in the population; on shorter time frames, such as considered here, the effects can be either negligible or not, depending on the age structure and elapsed years between samples.

In general, when N_1 and N_2 were both $<$ Baseline N_b , \hat{N}_b agreed well with $\hat{N}_{b(0,t)}$ (Table 4). In contrast, when N_1 and/or $N_2 >$ Baseline N_b , estimates involving samples in years with higher N_b tended to be lower than $\hat{N}_{b(0,t)}$. For example, median \hat{N}_b for the $N_b = 100 \times 200$ comparisons shown in Fig. 6B was 126.4, about 5% less than the harmonic mean of 100 and 200 (133.3). If both sampled years had much higher N_b than the Baseline, the difference between \hat{N}_b and the true N_b was larger (median \hat{N}_b about 15% too low for $N_1 = N_2 = 200$, Baseline = 100 and about 35% too low for $N_1 = N_2 = 500$, Baseline $N_b = 100$). This result is due to the relatively weak drift signal when effective size is large in the sampled years, compared to background fluctuations in allele frequency associated with the smaller Baseline N_b .

The above results were obtained using an age structure with 50% of adults maturing at age 4 and 25% each at ages 3 and 5 (so $A_j = 0, 0, 0.25, 0.5, 0.25$), which has been used in previous salmon models (Waples 1990a, b, 2002). I also examined the scenario shown in Fig. 6A (Baseline $N = 100$; $N_1 = N_2 = 30$) with a variety of different age structures (Table 4). Performance was comparable to that shown in Fig. 6A as long as no more than 50% of the adults matured at a single age, even if the age distribution was highly asymmetrical (e.g. $A_j = 0, 0.1, 0.2, 0.3, 0.4$). When 60% of the adults matured at one age, small biases were apparent in median \hat{N}_b and dispersion increased; with a strongly unimodal age structure ($A_j = 0, 0, 0.1, 0.8, 0.1$), behaviour of \hat{N}_b became more erratic. This pattern — estimates of \hat{N}_b strongly affected by demographic fluctuations in prior years when a single age at maturity strongly dominates — has been noted previously (Waples 2002).

Results discussed above apply to Plan II; Plan I sampling is more complex and results are quite different. In this case, each (generally adult) sample can be represented as a composite of individuals produced by two or more previous breeding populations (Fig. 3). Allele frequencies in adult samples thus will be directly affected by N_b in a series of previous years. Two general approaches are possible for dealing with this complication. If each individual in the sample can be assigned to a parental breeding population (e.g. by marking, paternity analysis, or ageing), a series of new samples can be constructed that can be treated under Plan II. Alternatively, it could be recognized that the Plan I sample is affected by N_b in multiple prior years, in which

case \hat{N}_b can best be interpreted as estimating the harmonic mean N_b over the previous few generations.

In the salmon model, correlations in allele frequencies between temporal samples can arise in two ways: either directly (if the breeders in year 0 make a direct contribution of progeny to the breeders in year t) or indirectly (through cascading effects of prior years' breeders on years 0 and t). Covariances associated with the latter effect are captured in the coefficient b that depends on age structure and years between samples. Direct effects differ according to the sampling plan. Under Plan II, both the sample in year 0 and the breeders in year t are derived independently by binomial sampling from the breeders in year 0; therefore, $\text{cov}(X, Y) = 0$, just as it is in the discrete generation model.

Plan I samples could include some individuals that reproduce and contribute to subsequent generations; however, because the salmon model is not a simple Markov chain, this factor is less important than it is for species with discrete generations. For example, in Fig. 3 breeders in year 0 make no genetic contribution to the population in year 2, so whether adults sampled in year 0 include any individuals that also reproduce has no effect on the difference in allele frequencies in samples S_0 and S_2 . If the second sample were taken in year 3 instead of year 2, there would be some direct contribution from adults in year 0 to the second sample, but that would apply to only a fraction of the breeding population. Waples (1990a) did not quantitatively evaluate adult sampling in the salmon model but suggested that a correction of magnitude $1/N_0$ (comparable to the adjustment for Plan I sampling in the discrete generation model) might be appropriate. The above arguments indicate that, at a maximum, the effect will be smaller than this and in many cases will be zero. So, $\text{cov}(X, Y)$ can probably be safely ignored even for Plan I samples in the salmon model.

Discussion

Results for both models are summarized in a generalized form in Table 5. In the discrete generation model, genetic drift is a simple Markov process, and it is possible to specify precisely the generations that contribute to a temporal estimate of N_e . The heterozygote excess method also provides information about a specific time period (the parental generation). In contrast, Plan I samples in the salmon model are derived from a mixture of cohorts and are affected by N_b in a variety of previous brood years. Intermediate to these extremes are estimates based on Plan II samples in the salmon model and estimates based on linkage disequilibrium in the discrete generation model. In these cases, the estimates can be affected by effective size in previous time periods, but these effects are funnelled through (and modulated by) specific years or generations. As a result, in these cases it is possible to identify the time

Table 5 Summary of time periods to which an estimate of N_e or N_b applies. In the discrete generation model, single-sample estimates based on linkage disequilibrium can also be influenced by N_e in prior generations. In the salmon model, j takes a value equal to each age at which adults reproduce

Time of sample	Discrete generations (\hat{N}_e)		Salmon model (\hat{N}_b)	
	Gen (0)	Gen (t)	Year (0)	Year (t)
One sample estimates (Inbreeding effective size)				
Plan I	Gen (-1)	Gen ($t - 1$)	Years ($0 - j$) ⁶	Years ($t - j$)
Plan II	Gen (-1)	Gen ($t - 1$)	Year (0)	Year (t)
Temporal estimates (Variance effective size)				
Plan I	Gen (0) to Gen ($t - 1$)		Years ($0 - j$) and Years ($t - j$)	
Plan II	Gen (0) to Gen ($t - 1$)		Year (0) and Year (t)	

periods that under most conditions can be expected to have the greatest influence on estimates of effective size. Specific conclusions include the following:

Discrete generations

- 1 A temporal estimate based on either Plan I or Plan II samples taken in generations 0 and t provides information about the harmonic mean of the effective sizes in generations 0 through $t - 1$.
- 2 In sampling Plan I, temporal estimates also depend on census size in generation 0 (N_0). N_0 should be enumerated at the time of sampling, so if nonlethal sampling occurs at the juvenile stage, N_0 can be quite large.
- 3 Single samples, either Plan I or Plan II, never provide information about N_e in the current generation; instead, they estimate N_e in the parental generation. In the disequilibrium method, \hat{N}_e can also be influenced by effective size several generations earlier. These historical effects are asymmetrical, with recent bottlenecks having a more profound and longer-lasting effect on \hat{N}_e than recent episodes of larger population size.
- 4 Given a pair of genetic samples in generations 0 and t , the temporal method and single-sample methods never estimate N_e in exactly the same generations. Both approaches jointly estimate $N_{e(t-1)}$ but otherwise are influenced by effective size in different generations. And for generation $t - 1$, the single-sample methods estimate the inbreeding $N_{e'}$ whereas the temporal method is a function of the variance N_e . With respect to the example in Fig. 1 therefore the temporal \hat{N}_e estimates $N_{e(0)}$, regardless of the sampling plan, and the appropriate estimate of N_e/N is $100/800 = 0.125$. A disequilibrium or heterozygote excess estimate based on S_1 also estimates $N_{e(0)}$ but the sample S_0 provides an estimate of $N_{e(-1)}$.
- 5 Results presented here for the F -based temporal model should also be applicable to recently developed likelihood methods (Williamson & Slatkin 1999; Wang 2001;

Berthier *et al.* 2002), which also depend on allele frequency change and should be affected in a similar way by effective size in specific time periods. However, the present results would not be applicable to equilibrium or asymptotic N_e estimated, for example, by the coalescent approach of Beerli & Felsenstein (2001). Beaumont (2003) described a coalescent-based temporal method which can provide separate estimates of the initial and final N_e in a population experiencing exponential growth; therefore, the time periods for \hat{N}_e discussed here would not necessarily be concordant with those applicable to Beaumont's method.

Salmon model

- 1 Plan I and Plan II samples never provide information about N_b in the same suite of years. This is true for both one-sample and temporal estimates of N_b . This means that in most cases, juvenile and adult samples will not estimate N_b in the same time periods.
- 2 Plan I samples are particularly hard to interpret because they can be affected by N_b in a wide range of prior years. Ageing individuals so that separate cohorts can be reconstructed can help reduce this complication.
- 3 A pair of Plan II samples taken in years 0 and t provides an estimate of the harmonic mean of $N_{b(0)}$ and $N_{b(t)}$, whether the estimates are computed using the single-sample or temporal method. This congruence between the one-sample and temporal methods contrasts with the situation for discrete generations and should facilitate combining the estimates into a single overall estimate of effective size.
- 4 Effects of residual linkage disequilibrium due to changes in population size in the recent past should have less effect on \hat{N}_b than is the case in the discrete generation model.
- 5 In Plan I sampling for the temporal method, initial population size (N_0) has less effect on \hat{F} and \hat{N}_b than it does in

Table 6 Hypothetical example illustrating how general results shown in Table 5 can be applied to real data for the two life history models. Plan II samples are taken in time periods 0 and 4. Arrows point to the effective size in the time period(s) to which each estimate applies. \hat{N}_e , \hat{N}_b , and \tilde{N} are harmonic means over the time periods indicated. In computing the ratios \hat{N}_e/N and \hat{N}_b/N , it is assumed that \hat{N}_e and \hat{N}_b accurately estimate the true effective size

Discrete generations					
Gen	Sample	N	N_e	Single sample	Temporal
-1	S_0	200	140	$N_{e(-1)}$ $\hat{N}_e/N_{(-1)} = 0.7$	$\hat{N}_e \rightarrow \tilde{N}_{e(0-3)} = 70$ $\hat{N}_e/\tilde{N}_{(0-3)} = 70/509 = 0.14$
0		1150	970		
1		420	24		
2		710	310		
3	S_4	312	89	$N_{e(3)}$ $\hat{N}_e/N_{(3)} = 0.29$	
4		80	55		
5		890	611		
Salmon model					
Year	Sample	N	N_b	Single sample	Temporal
-1	S_0	200	140	$N_{b(0)}$ $\hat{N}_b/N_{(0)} = 0.84$	$\hat{N}_b \rightarrow \tilde{N}_{b(0,4)} = 104$ $\hat{N}_b/\tilde{N}_{(0,4)} = 104/501 = 0.21$
0		1150	970		
1		420	24		
2		710	310		
3	S_4	312	89	$N_{b(4)}$ $\hat{N}_b/N_{(4)} = 0.69$	
4		80	55		
5		890	611		

the discrete generation model, and less than has been suggested in the literature for the salmon model. Variable age at maturity dilutes any correlations that occur between the genes in S_0 and the genes sampled at a later time (t).

- Although the salmon model assumes semelparity, the conclusions should be generally applicable to populations in species such as Atlantic salmon (*Salmo salar*) and steelhead (anadromous *Oncorhynchus mykiss*) in which the incidence of iteroparity is low.
- Results for the salmon model should also be generally applicable to other semelparous, age-structured species (including monocarpic plants; Nunney 2002; Waples 2002; Vitalis *et al.* 2004), although this would be important to evaluate directly. In evaluating genetic data for these species, it will be important (as it is with salmon) to determine whether the samples represent single or multiple cohorts.

Practical implications

Results discussed here can be of direct practical relevance in two ways. First, genetic estimates of N_e have intrinsically high stochastic variance, so it is important to maximize the amount of usable data to generate estimates with acceptable precision. One-sample and two-sample methods provide independent information about effective size (Waples 1991),

but these estimates cannot be combined properly without knowledge of the specific time periods to which each applies. Combining estimates in the discrete generation model should be carried out with considerable caution, as the one-sample and two-sample methods generally do not estimate N_e in the same time periods. Table 6 illustrates this point using a hypothetical data set. In this example, the same time series of census and effective sizes (N_e or N_b) is used for both models, and Plan II samples are taken in the same time periods (0 and 4). With discrete generations, the two individual samples provide estimates of $N_{e(-1)}$ and $N_{e(3)}$, respectively, whereas the temporal method estimates the harmonic mean N_e in generations 0 thru 3, inclusive. In contrast, a pair of Plan II samples in the salmon model estimates the effective numbers of breeders in years 0 and 4 regardless which method is used.

Second, properly matching estimates of N_e with census size in the appropriate time periods is essential if one is to develop a meaningful estimate of N_e/N . In Table 6 for the discrete generation model, note that if the single-sample estimate taken in generation 4 were matched with the census size in generation 4 rather than generation 3, the resulting estimate of N_e/N ($89/80 = 1.1$) would be substantially biased upwards and biologically implausible. When temporal samples span a number of generations (and hence \hat{N}_e applies to the harmonic mean N_e), it is important to consider how the denominator of N_e/N is calculated. Using

the harmonic mean N (rather than the arithmetic mean, as has been suggested by some authors) avoids the statistical artefact of mixing harmonic and arithmetic means in a ratio and leads to an estimate that can be interpreted as a function of the single generation N_e/N values (Kalinowski & Waples 2002). In the example in Table 6, the discrete-generation temporal estimate of N_e was compared with the harmonic mean N in generations 0 through 3 ($\tilde{N}_{0-3} = 509$), leading to an estimated N_e/N ratio of 0.14. If \hat{N}_e had instead (and improperly) been matched with the harmonic mean N in generations 1 through 4 ($\tilde{N}_{1-4} = 235$), the estimate of N_e/N would have more than doubled.

Overlapping generations

Those adopting genetic methods to estimate N_e commonly use discrete generation models but apply them to species with overlapping generations. Although a number of issues related to age structure and sampling complicate the analysis of species with overlapping generations, some general observations can be made.

Single-sample estimates. In many cases, parents of a single sample will be drawn from only one or a limited number of cohorts rather than the generation as a whole; if so the estimate reflects not N_e but rather N_b or a quantity intermediate to N_b and N_e . The relationship between N_b and N_e in iteroparous species cannot be determined without information about the degree to which individuals that contribute to N_b in one time period (within a generation) also contribute to N_b in a subsequent time period — that is, the distribution of lifetime reproductive success. Still, the estimate of N_b can be useful in some cases (Waples 1991).

Temporal estimates. In the discrete generation model, the temporal method assumes that samples S_0 and S_t (whether Plan I or Plan II) are drawn randomly from the gamete pool produced by the parental generation as a whole. Again, it is difficult to envision a sampling regime that would accomplish the same thing for a species with overlapping generations. Instead, samples often will represent progeny of breeders in one or more years that make up a fraction of a generation. In this case, $V(X - Y)$ is expected to be larger than given in equation 7, and the standard temporal method would tend to upwardly bias F (because the adjustment $1/S$ is no longer adequate to account for sampling variance), leading to a downward bias in \hat{N}_e . Those who have applied the temporal method to species with overlapping generations (e.g. Miller & Kapuscinski 1997; Hauser *et al.* 2002) have typically assumed that, if enough generations have elapsed between samples, cumulative drift will be large enough that this bias can be ignored. Determining under what circumstances this is a reasonable assumption must await a quantitative treatment of this subject.

Acknowledgements

I would like to thank Eric Anderson, Craig Busack, Mike Ford, Michael Hansen, Lorenz Hauser, Alexander Pudovkin, David Tallmon, and two anonymous reviewers for useful comments on earlier drafts and Jinliang Wang for interesting discussions. The research was initiated while the author was a visiting scientist at LECA, and I am grateful to Gordon Luikart for making that possible.

References

- Ardren WR, Kapuscinski AR (2003) Demographic and genetic estimates of effective population size (N_e) reveals genetic compensation in steelhead trout. *Molecular Ecology*, **12**, 35–49.
- Beaumont M (2003) Estimation of population growth or decline in genetically monitored populations. *Genetics*, **164**, 1139–1160.
- Berli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences, USA*, **98**, 4563–4568.
- Berthier P, Beaumont MA, Cornuet J-M, Luikart G (2002) Likelihood-based estimation of the effective population size using temporal changes in allele frequencies: a genealogical approach. *Genetics*, **160**, 741–751.
- Caballero A (1994) Developments in the prediction of effective population size. *Heredity*, **3**, 657–679.
- Crow JF, Denniston C (1988) Inbreeding and variance effective population numbers. *Evolution*, **42**, 482–495.
- Ewens WJ (1979) *Mathematical Population Genetics*. Springer-Verlag, Berlin.
- Frankham R (1995) Effective population size/adult population size ratios in wildlife: a review. *Genetical Research (Cambridge)*, **66**, 95–107.
- Hauser L, Adcock GJ, Smith PJ, Ramírez JHB, Carvalho GR (2002) Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). *Proceedings of the National Academy of Sciences, USA*, **99**, 11742–11747.
- Hedgecock D (1994) Does variance in reproductive success limit effective population sizes of marine organisms? In: *Genetics and Evolution of Aquatic Organisms* (ed. Beaumont AR), pp. 122–134. Chapman & Hall, London.
- Hill WG (1981) Estimation of effective population size from data on linkage disequilibrium. *Genetical Research (Cambridge)*, **38**, 209–216.
- Hoffman EA, Blouin MS (2004) Historical data refute recent range contraction as cause of low genetic diversity in isolated frog populations. *Molecular Ecology*, **13**, 271–276.
- Hoffman EA, Schueler FW, Blouin MS (2004) Effective population sizes and temporal stability of genetic structure in *Rana pipiens*, the northern leopard frog. *Evolution*, **58**, 2536–2545.
- Huhn M, Piepho H-P (2004) Inbreeding coefficients for stochastically varying small population sizes — bias of calculation based on effective numbers. *Journal of Theoretical Biology*, **226**, 467–475.
- Jagers P, Sagitov S (2004) Convergence to the coalescent in populations of substantially varying size. *Journal of Applied Probability*, **41**, 358–368.
- Johnson JA, Bellinger MR, Toepfer JE, Dunn P (2004) Temporal changes in allele frequencies and low effective population size in greater prairie-chickens. *Molecular Ecology*, **13**, 2617–2630.

- Kaeuffer R, Pontier D, Devillard S, Perrin N (2004) Effective size of two feral domestic cat populations (*Felis catus* L.): effect of the mating system. *Molecular Ecology*, **13**, 483–490.
- Kalinowski ST, Waples RS (2002) Relationship of effective to census size in fluctuating populations. *Conservation Biology*, **16**, 129–136.
- Kimura M, Crow JF (1963) The measurement of effective population number. *Evolution*, **17**, 279–288.
- Laurie-Ahlberg CC, Weir BS (1979) Allozyme variation and linkage disequilibrium in some laboratory populations of *Drosophila melanogaster*. *Genetics*, **92**, 1295–1314.
- Matocq MD (2004) Reproductive success and effective population size in woodrats (*Neotoma macrotis*). *Molecular Ecology*, **13**, 1635–1642.
- 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.
- Motro U, Thompson G (1982) On heterozygosity and the effective size of populations subject to size changes. *Evolution*, **36**, 1059–1066.
- Nei M, Tajima F (1981) Genetic drift and estimation of effective population size. *Genetics*, **98**, 625–640.
- Nunney L (1993) The influence of mating system and overlapping generations on effective population size. *Evolution*, **47**, 1329–1341.
- Nunney L (2002) The effective size of annual plant populations: The interaction of a seed bank with fluctuating population size in maintaining genetic variation. *American Naturalist*, **160**, 195–204.
- Pudovkin AID, Zaykin V, Hedgecock D (1996) On the potential for estimating the effective number of breeders from heterozygote-excess in progeny. *Genetics*, **144**, 383–387.
- Scribner KT, Arntsen JW, Burke T (1997) Effective number of breeding adults in *Bufo bufo* estimated from age-specific variation in minisatellite loci. *Molecular Ecology*, **6**, 707–712.
- Shrimpton JM, Heath DD (2003) Census vs. effective population size in chinook salmon: large- and small-scale environmental perturbation effects. *Molecular Ecology*, **12**, 2571–2583.
- Sinnock P (1975) The Wahlund effect for the two-locus model. *American Naturalist*, **109**, 565–570.
- Sved JA (1971) Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical Population Biology*, **2**, 125–141.
- Sved JA, Feldman MW (1973) Correlation and probability methods for one and two loci. *Theoretical Population Biology*, **4**, 129–132.
- Tajima F (1992) Statistical method for estimating the effective population size in Pacific salmon. *Journal of Heredity*, **83**, 309–311.
- Turner TF, Wares JP, Gold JR (2002) Genetic effective size is three orders of magnitude smaller than adult census size in an abundant, estuarine-dependent marine fish (*Sciaenops ocellatus*). *Genetics*, **162**, 1329–1339.
- Vitalis R, Glémin S, Olivieri I (2004) When genes go to sleep: The population genetic consequences of seed dormancy and monocarpic perenniality. *American Naturalist*, **163**, 295–311.
- Wang J (2001) A pseudo-likelihood method for estimating effective population size from temporally spaced samples. *Genetical Research*, **78**, 243–257.
- Waples RS (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics*, **121**, 379–391.
- Waples RS (1990a) Conservation genetics of Pacific salmon. III. Estimating effective population size. *Journal of Heredity*, **81**, 277–289.
- Waples RS (1990b) Conservation genetics of Pacific salmon. II. Effective population size and the rate of loss of genetic variability. *Journal of Heredity*, **81**, 267–276.
- Waples RS (1991) Genetic methods for estimating the effective size of cetacean populations. *Report of the International Whaling Commission (Special Issue 13)*, 279–300.
- Waples RS (2002) The effective size of fluctuating salmon populations. *Genetics*, **161**, 783–791.
- Waples RS (2004) Salmonid insights into effective population size. In: *Evolution illuminated: Salmon and their relatives* (eds Hendry AP, Stearns SC), pp. 295–314. Oxford University Press, Oxford, UK.
- Waples RS, Smouse PE (1990) Gametic disequilibrium analysis as a means of identifying mixtures of salmon populations. *American Fisheries Society Symposium*, **7**, 439–458.
- Weir BS (1979) Inferences about linkage disequilibrium. *Biometrics*, **35**, 235–254.
- Weir BS, Hill WG (1980) Effect of mating structure on variation in linkage disequilibrium. *Genetics*, **95**, 477–488.
- Williamson EG, Slatkin M (1999) Using maximum likelihood to estimate population size from temporal changes in allele frequencies. *Genetics*, **152**, 755–761.
- Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.

This research was motivated by an interest in applying population genetic principles to real-world problems in applied conservation. The author's research focuses on population genetics and conservation genetics of marine and anadromous fish, and he is particularly interested in the genetic consequences of different life history strategies.

Appendix

Let P_0 and P_t be the population allele frequencies in generations 0 and t , respectively. The elapsed time is t generations, but a total of $t + 1$ generations of reproduction are involved (generations 0 ... t). We want to evaluate $E(P_0 - P_t)^2 = V(P_0 - P_t) = V(P_0) + V(P_t) - 2\text{cov}(P_0, P_t)$ to determine exactly which time periods contribute to $V(P_0 - P_t)$ and hence F . For the moment, we are interested in parametric population frequencies so we ignore sampling. If we take as a point of reference the allele frequency in the gamete pool preceding generation 0 (P_{-1} , which is also the frequency in the $2N_{e(-1)}$ effective breeders in generation -1), we want to find $V(P_0 | P_{-1})$, $V(P_t | P_{-1})$, and $\text{cov}(P_0, P_t | P_{-1})$. The $2N_{e(0)}$ genes with frequency P_0 in generation 0 can be considered a random binomial draw from the gametes produced by generation -1, so $V(P_0 | P_{-1}) = P_{-1}(1 - P_{-1})/(2N_{e(0)})$.

$V(P_t | P_{-1})$ reflects $t + 1$ generations of genetic drift (i.e. in drawing the $2N_{e(i)}$ genes for the effective population sizes in generations 0, 1, 2, ... t). It is well known that if N_e is constant, this variance is given by $V(P_t | P_{-1}) = P_{-1}(1 - P_{-1})\{1 - [1 - 1/(2N_e)]^{t+1}\}$. When N_e varies, this expression can be rewritten as

$$V(P_t | P_{-1}) = P_{-1}(1 - P_{-1}) \left\{ 1 - \prod_{i=0}^t [1 - 1/(2N_{e(i)})] \right\}, \quad (\text{eqn A1})$$

which is often approximated using \tilde{N}_e , the harmonic mean of the single-generation values: $V(P_t | P_{-1}) \approx P_{-1}(1 - P_{-1})\{1 - [1 - 1/(2\tilde{N}_e)]^{t+1}\}$. The harmonic mean approximation has been shown to be robust in most practical circumstances (Motro & Thompson 1982; Huhn & Piepho 2004; Jagers & Sagitov 2004). In the present case, we are interested in N_e in specific years, so a more useful way of expressing the approximation for equation A1 is

$$V(P_t | P_{-1}) \approx P_{-1}(1 - P_{-1}) \left[\sum_{i=0}^t 1/(2N_{e(i)}) \right].$$

P_0 and P_t are correlated with respect to P_{-1} , because they share a common pathway through the $2N_{e(0)}$ genes in generation 0; therefore, $\text{cov}(P_0, P_t | P_{-1}) = P_{-1}(1 - P_{-1})/(2N_{e(0)})$. We can therefore approximate the overall drift variance $V(P_0 - P_t)$ as

$$\begin{aligned} V(P_0 - P_t) &= V(P_0) + V(P_t) - 2\text{cov}(P_0, P_t) \\ &\approx P_{-1}(1 - P_{-1}) \left[\frac{1}{2N_{e(0)}} + \sum_{i=0}^t \frac{1}{2N_{e(i)}} - \frac{1}{N_{e(0)}} \right] \\ &= P_{-1}(1 - P_{-1}) \left[\sum_{i=1}^t \frac{1}{2N_{e(i)}} \right]. \end{aligned} \quad (\text{eqn A2})$$

Thus, the difference in parametric population allele frequencies between generations 0 and t is due to t gener-

ations of genetic drift, beginning with generation 1 and ending with selection of the $2N_{e(t)}$ genes representing the effective population size in generation t .

In real populations, of course, parametric allele frequencies are not known. Even if complete enumeration and sampling of all individuals in the population were possible, in general P_t could not be calculated directly because of the unknown relationship between $N_{e(i)}$ and N_i . Consider the same population scenario as above, but now we are interested in the quantity $V(X - Y) = V(X) + V(Y) - 2\text{cov}(X, Y)$, where X and Y are allele frequencies in samples S_0 and S_t . Let us first examine the sample associated with generation 0. The $2S_0$ genes representing allele frequency X can be considered to be a binomial draw from the gamete pool preceding generation 0 (frequency P_{-1}). This is obvious for Plan II sampling but also true for Plan I. The sample of adults in generation 0 is randomly drawn from a finite population (size N_0) that is itself (assumed to be) randomly drawn from the gamete pool preceding generation 0. Nei & Tajima (1981) pointed out that this two-step sampling process is equivalent to taking a binomial draw of $2S_0$ genes from the gamete pool preceding generation 0. Therefore, for both sampling plans, $V(X)$ is the binomial variance $P_{-1}(1 - P_{-1})/2S_0$. $V(Y)$ is also the same in the two sampling plans (Waples 1989) and in the constant size model is given by $V(Y) = P_{-1}(1 - P_{-1})\{1 - [1 - 1/(2N_e)]^t[1 - 1/(2S_t)]\}$. Once again, the sample S_t can be considered a binomial draw from the gamete pool in the parental generation ($t - 1$), so it is not influenced by P_t or $N_{e(t)}$. Therefore, the t generations of drift contributing to $V(Y)$ are generations 0 through $t - 1$, and $V(Y)$ can be expressed explicitly as

$$V(Y) = P_{-1}(1 - P_{-1}) \left\{ 1 - \prod_{i=0}^{t-1} [1 - 1/(2N_{e(i)})][1 - 1/(2S_t)] \right\},$$

which, ignoring second order and higher terms, leads to

$$V(Y) \approx P_{-1}(1 - P_{-1}) \left[\sum_{i=0}^{t-1} 1/(2N_{e(i)}) + 1/(2S_t) \right] \quad \text{and}$$

$$V(X - Y) \approx P_{-1}(1 - P_{-1}) \left[\frac{1}{2S_0} + \sum_{i=0}^{t-1} \frac{1}{2N_{e(i)}} + \frac{1}{2S_t} \right] - 2\text{cov}(X, Y).$$

Thus, for neither sampling plan does the sample in generation t provide information about $N_{e(t)}$. In the context of the sampling plans described by Nei & Tajima (1981) and Waples (1989), $N_{e(0)}$ relates to the progeny generation for the sample at time 0, whereas $N_{e(t-1)}$ relates to the parental generation for the sample taken at time t . Thus, the temporal method actually estimates N_e in one parental generation and one progeny generation, as well as N_e in the intervening generations (if any). If $t = 1$, the parental generation for the second sample is the progeny generation for the first sample, and together they provide an estimate of $N_{e(0)}$.