

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

---

2009

## Modelling evolutionary processes in small populations: not as ideal as you think

Robin Waples

NOAA, robin.waples@noaa.gov

James R. Faulkner

Northwest Fisheries Science Center

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

---

Waples, Robin and Faulkner, James R., "Modelling evolutionary processes in small populations: not as ideal as you think" (2009). *Publications, Agencies and Staff of the U.S. Department of Commerce*. 468. <https://digitalcommons.unl.edu/usdeptcommercepub/468>

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.

# Modelling evolutionary processes in small populations: not as ideal as you think

ROBIN S. WAPLES and JAMES R. FAULKNER

Northwest Fisheries Science Center, 2725 Montlake Blvd. East, Seattle, WA 98112, USA

## Abstract

Evolutionary processes are routinely modelled using ‘ideal’ Wright–Fisher populations of constant size  $N$  in which each individual has an equal expectation of reproductive success. In a hypothetical ideal population, variance in reproductive success ( $V_k$ ) is binomial and effective population size ( $N_e$ ) =  $N$ . However, in any actual implementation of the Wright–Fisher model (e.g., in a computer),  $V_k$  is a random variable and its realized value in any given replicate generation ( $V_k^*$ ) only rarely equals the binomial variance. Realized effective size ( $N_e^*$ ) thus also varies randomly in modelled ideal populations, and the consequences of this have not been adequately explored in the literature. Analytical and numerical results show that random variation in  $V_k^*$  and  $N_e^*$  can seriously distort analyses that evaluate precision or otherwise depend on the assumption that  $N_e^*$  is constant. We derive analytical expressions for  $\text{Var}(V_k)$  [ $4(2N - 1)(N - 1)/N^3$ ] and  $\text{Var}(N_e)$  [ $N(N - 1)/(2N - 1) \approx N/2$ ] in modelled ideal populations and show that, for a genetic metric  $G = f(N_e)$ ,  $\text{Var}(\hat{G})$  has two components:  $\text{Var}_{\text{Gene}}$  (due to variance across replicate samples of genes, given a specific  $N_e^*$ ) and  $\text{Var}_{\text{Demo}}$  (due to variance in  $N_e^*$ ).  $\text{Var}(\hat{G})$  is higher than it would be with constant  $N_e = N$ , as implicitly assumed by many standard models. We illustrate this with empirical examples based on  $F$  (standardized variance of allele frequency) and  $r^2$  (a measure of linkage disequilibrium). Results demonstrate that in computer models that track multilocus genotypes, methods of replication and data analysis can strongly affect consequences of variation in  $N_e^*$ . These effects are more important when sampling error is small (large numbers of individuals, loci and alleles) and with relatively small populations (frequently modelled by those interested in conservation).

*Keywords:* bias, binomial sampling, computer simulations, effective population size, genetic drift, precision, pseudoreplication, variance in reproductive success

Received 17 October 2008; revised 14 January 2009; accepted 7 February 2009

## Introduction

Computer modelling has revolutionized many aspects of evolutionary biology, and virtually all evolutionary models include the parameter effective population size ( $N_e$ ). Effective size is defined in terms of demographic parameters, and for a monoecious population with random selfing the effective size is given by (Crow & Denniston 1988; Caballero 1994):

$$N_e = \frac{\bar{k}N - 1}{\bar{k} - 1 + \frac{V_k}{\bar{k}}}, \quad (\text{eqn 1})$$

Correspondence: Robin Waples, Fax: +1 206 860 3335; E-mail: robin.waples@noaa.gov

This article is a US Government work and is in the public domain in the USA

where  $N$  is the number in the parental generation and  $\bar{k}$  and  $V_k$  are the mean and variance among parents in the number of gametes contributed to the next generation.

Coincident with recent, rapid increases in computational power has been an equally rapid escalation of the number of highly polymorphic markers available for genetic analysis, which has greatly enhanced the ability to study contemporary evolution (Pearse & Crandall 2004; Manel *et al.* 2005). As a consequence, computer models of evolutionary processes now commonly track multilocus genotypes of many individuals. The simplest (and by far most common) way to incorporate  $N_e$  into such evolutionary models is to use ‘ideal’ populations, in which  $N$  is constant, generations are discrete and all individuals have an equal opportunity to contribute genes to the next generation. This condition is met, for example, if each individual contributes equally to

a large gamete pool and the next generation is chosen by random union of these gametes. This ‘Wright–Fisher’ model (Fisher 1930; Wright 1931) is straightforward to implement in individual-based computer models: each individual in the progeny generation is formed by choosing two parents, with replacement, from the parental pool, and then randomly choosing one multilocus gamete from each parent. It is easy to control whether selfing is allowed. If sexes are separate, the numbers of males and females must be equal, and each progeny is created by choosing one male and one female parent, with replacement.

In the Wright–Fisher model, the expected variance in reproductive success is the binomial variance

$$E(V_k) = \frac{(N-1)}{N} \bar{k}, \tag{eqn 2}$$

which approaches the Poisson variance ( $\bar{k}$ ) as population size increases. Since an ideal population is also constant in size (hence each individual contributes on average  $\bar{k} = 2$  gametes to the next generation), the Wright–Fisher model implies that  $E(V_k) = 2(N-1)/N$ . Inserting these values into eqn 1 yields  $N_e = N$ . Different effective sizes can therefore be modelled by choosing populations with different numbers of ‘ideal’ individuals.

However, eqn 2 is only an expectation. In any given population that is otherwise ‘ideal’, the random process of gamete formation will lead to a range of realized  $V_k$  values ( $V_k^*$ ) across generations (or across different replicate ideal populations within the same generation). Because effective population size varies directly as a function of  $V_k$  (eqn 1), it follows that realized effective size ( $N_e^*$ ) will also vary randomly in otherwise ‘ideal’ populations that behave according to the Wright–Fisher model. That is, when biologists model ‘ideal’ populations in computer simulations, they do not actually know the effective size of each generation in each replicate, because  $V_k^*$  (and hence  $N_e^*$ ) varies stochastically. This fact has not been adequately dealt with in the literature.

In this paper, we consider the consequences of these random variations in  $V_k^*$  and  $N_e^*$  in modelled populations that are ‘ideal’. In particular, we evaluate the consequences of (falsely) assuming that  $N_e^*$  in a modelled ideal population is constant and equal to the census population size.

**The Model**

We assume a monoecious diploid population with random selfing, which corresponds to the original Wright–Fisher definition of an ideal population. The population is of constant size  $N$ , so  $\bar{k} = 2$  and the inbreeding and variance effective sizes are the same, and eqn 1 simplifies to

$$N_e = \frac{2N-1}{1 + \frac{V_k}{2}}. \tag{eqn 3}$$

The number of gametes contributed by individual  $i$  to the next generation ( $k_i$ ) is a random variable. Reproduction is according to the Wright–Fisher model described above, in which a total of  $2N$  multilocus gametes are chosen for the next generation, with each individual having an equal opportunity to contribute a gamete on each draw. The  $k_i$  thus follow a multinomial distribution with  $2N$  trials, and the probability that a particular individual contributes a gamete to the next generation at each trial is equal to  $1/N$ . In any given generation  $j$ , the variance of  $k_i$  among individuals ( $V_k$ ) varies randomly with expectation equal to the binomial variance  $2(1-1/N)$ . We will use the term  $V_k^*$  to refer generically to realized values of  $V_k$  for a particular generation. Similarly, we will use  $N_e^*$  to refer to realized values of  $N_e$  that apply to a particular generation [ $N_e^* = (2N-1)/(1 + V_k^*/2)$ ; eqn 3]. We used both analytical and numerical methods to evaluate  $\text{Var}(V_k)$  and the associated effects on  $N_e^*$  and genetic indices that depend on  $N_e$ .

**Results**

*Distribution of  $V_k^*$*

The variance of  $V_k$  can be shown analytically to be a simple function of population size (see Appendix S1, Supporting information):

$$\text{Var}(V_k) = \frac{4(2N-1)(N-1)}{N^3} \approx 8/N. \tag{eqn 4}$$

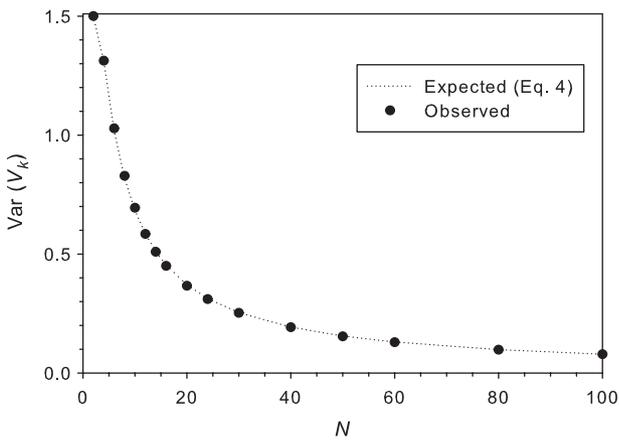
Empirical results from computer simulations using  $N = 2-100$  ideal individuals produced the following results: (i) mean of realized  $V_k^*$  across replicates agreed with the binomial expectation  $E(V_k^*) = 2(1-1/N)$  (Table 1); and (ii) observed variance of  $V_k^*$  confirmed the accuracy of eqn 4 (Table 1, Fig. 1). Because the largest term in eqn 4 is of order  $1/N$ ,  $\text{Var}(V_k)$  is small unless  $N$  is also small. For large populations ( $N \geq 100$ ), the distribution of  $V_k^*$  is approximately normal, but for smaller  $N$  it is skewed due to the boundary constraint that  $V_k^* \geq 0$  (Fig. 2a).

*Distribution of  $N_e^*$  and  $N_e^*/N$*

Each realized  $V_k^*$  value leads to a different  $N_e^*$  (cf. eqn 1), so we can ask how random variation in  $V_k^*$  affects realized  $N_e^*$ . Because of the inverse relationship between  $V_k$  and  $N_e$ , the distributions of these two parameters differ in an important way. Whereas the empirical  $\bar{V}_k^*$  from simulations agrees asymptotically with the expected binomial variance given in eqn 2 (Table 1), the arithmetic mean  $N_e^*$  ( $\bar{N}_e^*$ ) does not equal  $N$ . That is, in an ideal, Wright–Fisher population of finite size  $N$ , the expected value of  $N_e^*$  is not  $N$  but something slightly higher. This result arises because  $N_e^*$  is a convex function of  $V_k^*$ , and by Jensen’s inequality  $f[E(x)] \leq E[f(x)]$  for a convex function  $f$  (Jensen 1906). A

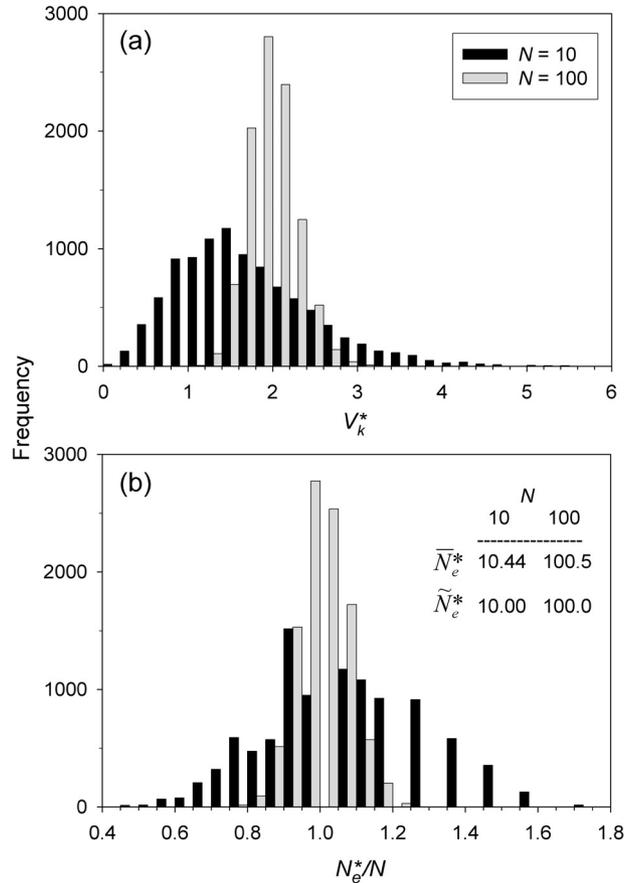
**Table 1** Comparison of expected values for  $V_k^*$  and  $\text{Var}(V_k)$  (eqns 2 and 4, respectively) with observed values for simulated data. Simulations modelled ideal populations of size  $N$ ; observed values are means over 30 000 replicates.  $\bar{N}_e^*$  and  $\tilde{N}_e^*$  are the arithmetic and harmonic means of  $N_e^*$  across replicates, respectively, calculated from the empirical  $V_k^*$  using eqn 1

N	Mean $V_k^*$		Var( $V_k$ )		$\bar{N}_e^*$	$\tilde{N}_e^*$
	Expected	Observed	Expected	Observed		
2	1.0000	1.0067	1.5000	1.5005	2.25	2.00
4	1.5000	1.5080	1.3125	1.3126	4.37	3.99
6	1.6667	1.6674	1.0185	1.0283	6.40	5.98
8	1.7500	1.7517	0.8203	0.8287	8.43	7.99
10	1.8000	1.7988	0.6840	0.6942	10.44	9.98
12	1.8333	1.8303	0.5856	0.5845	12.44	12.02
14	1.8571	1.8549	0.5117	0.5093	14.46	14.00
16	1.8750	1.8681	0.4541	0.4503	16.47	16.00
20	1.9000	1.8966	0.3705	0.3665	20.47	19.99
24	1.9167	1.9159	0.3128	0.3105	24.50	24.02
30	1.9333	1.9330	0.2535	0.2524	30.45	29.99
40	1.9500	1.9524	0.1926	0.1924	40.53	40.00
50	1.9600	1.9585	0.1552	0.1539	50.46	50.01
60	1.9667	1.9674	0.1300	0.1294	60.46	59.99
80	1.9750	1.9739	0.0981	0.0977	80.49	79.94
100	1.9800	1.9800	0.0788	0.0788	100.47	99.98



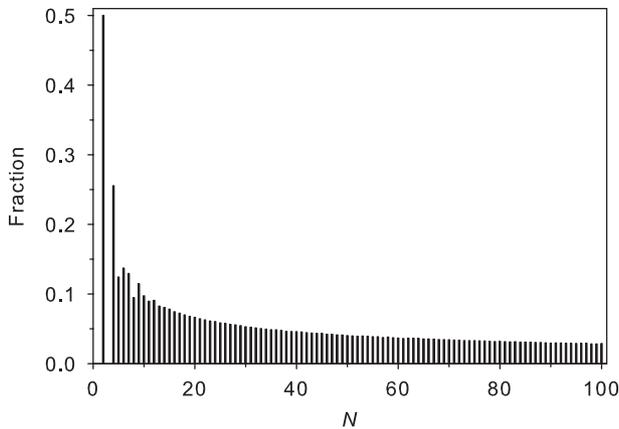
**Fig. 1** Comparison of expected values for  $\text{Var}(V_k)$  (eqn 4) with observed values for simulated data. Simulations model ideal populations of size  $N$ ; observed values are means over 30 000 replicates.

second order Taylor's series approximation to the expected value of  $N_e^*$  gives  $E(N_e^*) \approx N + (N - 1)/(2N - 1)$  (see Appendix S2, Supporting information). The second term is small and asymptotes to  $1/2$  as  $N$  gets large, but the proportional contribution is noticeable for small  $N$ . For example, with  $N = 10$ ,  $\bar{N}_e^*$  was 10.44, over 4% higher than expected for an ideal population, but with  $N = 100$  the mean  $N_e^*$  was less than one-half of one percent too high ( $\tilde{N}_e^* = 100.47$ ; Table 1, Fig. 2).



**Fig. 2** Distribution of realized  $V_k^*$  (panel a) and  $N_e^*/N$  (panel b) values in simulated ideal populations of size  $N = 10$  or  $100$ . In each of 30 000 replicates,  $N_e^*$  was calculated from the realized  $V_k^*$  using eqn 1.

The variance of  $N_e^*$  is a function of  $\text{Var}(V_k)$  and is therefore a function of  $N$ . A first order Taylor's series approximation to the variance of  $N_e^*$  is  $\text{Var}(N_e^*) \approx N(N - 1)/(2N - 1)$  (see Appendix S2, Supporting information), which, to a good approximation, is simply  $N/2$ . The consequences of random variation in  $V_k^*$  on the distribution of effective size can best be evaluated by examining the ratio  $N_e^*/N$ . Figure 2(b) illustrates how much wider the distribution of  $N_e^*/N$  is for  $N = 10$  than for  $N = 100$ . With  $N \leq 10$ , random variation in  $V_k^*$  can lead to random variation in  $N_e^*/N$  of  $\pm 50\%$  of the value expected for an ideal population. Even with  $N = 100$ , however, the effect is not negligible, as realized  $N_e^*$  fluctuates between about 80 and 120, with occasionally greater deviations (Fig. 2b). For  $N > 10$ , the fraction of replicates in which realized  $V_k^*$  exactly equals the binomial variance is less than 10% — which means that  $N_e^* \neq N$  more than 90% of the time (Fig. 3). A sequence of random realizations of  $N_e^*$  with  $N$  constant at 100 is shown in Fig. 4. In this sequence,  $N_e^*$  varied from 88 (generation 18) to 121 (generation 2), and in



**Fig. 3** Empirical fraction of modelled Wright–Fisher populations in which realized  $V_k^*$  exactly equalled the binomial variance,  $2(N-1)/N$ . For each value of  $N$ , results were averaged over 500 000 realizations of the Wright–Fisher process with  $\bar{k}$  fixed at 2.

only one generation (5) was  $V_k^*$  exactly equal to the binomial variance, leading to  $N_e^* = N = 100$ .

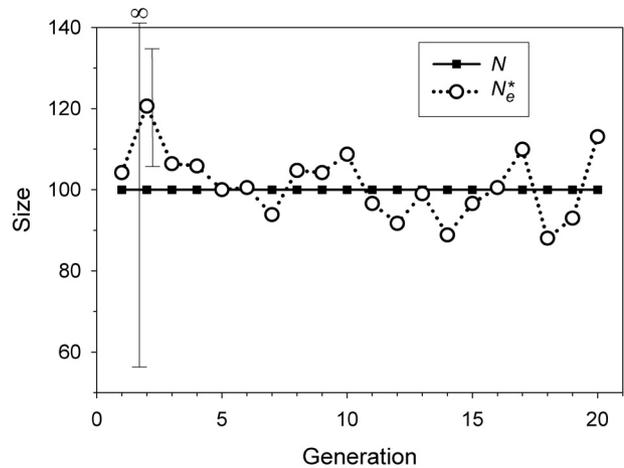
#### Consequences of Variation in $N_e^*$ and $N_e^*/N$

What consequences does random variation in  $N_e^*$  and  $N_e^*/N$  have for evolutionary inference? Below we consider this question with respect to both bias and precision.

**Bias.**  $\bar{N}_e^*$  and  $\bar{N}_e^*/N$  are slightly biased upwards compared to values expected for ideal populations. By itself, however, this bias is unlikely to present a problem for most applications. Because most evolutionary processes are a function of  $1/N_e$ , response of a population over time is determined by the harmonic mean  $N_e^*$  ( $\bar{N}_e^*$ ), not the arithmetic mean ( $\bar{N}_e^*$ ). Even in small populations with considerable random variation in  $V_k^*$ ,  $\bar{N}_e^*$  agrees asymptotically with the nominal  $N_e$  for an ideal population (Table 1). This means that the cumulative amount of genetic drift experienced by an ideal population of size  $N$ , even with randomly fluctuating  $V_k^*$ , has an expectation equal to that of a population with constant effective size  $N_e = N$ .

Conclusions about potential biases, however, must be qualified with two important caveats about replication. First,  $\bar{N}_e^*$  only approaches  $N$  asymptotically as the number of replicates increases. If replication is limited, random variation in  $N_e^*$  will generally cause the mean drift signal  $1/\bar{N}_e^*$  to be higher or lower than expected for an ideal population of size  $N$ . If this source of random variation is not accounted for, misleading conclusions can result.

Second, in computer modelling of ideal populations, the type of replication can strongly affect results. To see this, assume that a biologist is interested in modelling a single generation of genetic drift. The stochastic process has a high intrinsic variance, so a large number of replications are



**Fig. 4** Random sequence of realized  $N_e^*$  values in a modelled ideal population with constant size  $N = 100$ . For the data point  $N_e^* = 121$  in generation 2, the vertical lines indicate parametric 95% confidence intervals (CIs) for a single sample, linkage disequilibrium estimate (Waples & Do 2008) with  $\bar{N}_e = 121$ . CIs are shown for two different sampling regimes: limited data (left; 30 individuals, eight loci with eight alleles each), and plentiful data (right; 60 individuals, 16 loci with 16 alleles each).

necessary to characterize ‘average’ behaviour. Replication can be achieved in two general ways. One might focus on a single gene locus and replicate the demographic process of reproduction across many ideal populations of the same size  $N$ . Alternatively, one might follow a single ideal population (size  $N$ ) and replicate the process of sampling genes across many independent gene loci. Neutral theory tells us that the expected variance in allele frequency among replicates is the same under the two scenarios. However, this is not the case when one considers the consequences of random, demographic variation in  $V_k^*$  and hence  $N_e^*$ . In the first scenario, each population will have an independent realization of  $V_k^*$  and  $N_e^*$ ; if enough replicate populations are considered, the harmonic mean  $N_e^*$  and hence the variance in allele frequency across populations will be close to that expected for ideal populations of size  $N$ . In the second scenario, there is no replication of the stochastic process that determines the realized variance in reproductive success; instead, there is a single fixed  $V_k^*$  (and  $N_e^*$ ) that applies equally to all gene loci. As a consequence, replication across a large number of gene loci will not, in general, produce a drift signal equivalent to that in a population of effective size  $N_e = N$ ; instead, replication will cause the drift signal to converge on the fixed  $1/N_e^*$ , which is determined by the value of  $V_k^*$  that (by chance) occurred in the single episode of reproduction. The consequences of this are discussed in more detail in the Discussion.

**Precision.** Sometimes, one is primarily interested in the variance (rather than the mean) of a quantity related to

genetic drift. Consider a quantity  $G$  that is influenced by genetic drift and is parametrically related to effective size by  $G = f(1/N_e)$  (e.g., see the example below using  $F$ ). If  $N_e$  is fixed (and hence  $G$  is fixed), and an unbiased estimator of  $G(\hat{G})$  is used to compute independent estimates from empirical genetic data, then observed variation in repeated estimates is simply due to the sampling variability of the estimator  $\hat{G}$ . This variance approaches zero as the sampling effort and efficiency informing each estimate gets very large. For a given realized parametric value of  $G$ , this conditional theoretic variance of the estimator can be written as  $\text{Var}(\hat{G} | G)$ , or alternatively  $\text{Var}(\hat{G} | N_e)$ . Here we condition on the parameter  $G$  to allow for cases when the variance of the estimator depends on the value of  $G$ .

However, realized  $N_e^*$  is not constant in an ideal population of constant size  $N$ , but instead fluctuates randomly. Typically, one is concerned with a number of replicate generations of drift, each characterized by a different realized parametric value of  $G$ . To get the overall variance of  $\hat{G}$  across replicate generations *and* across replicate samples within generations, it is necessary to account for variation in parametric values of  $G$  due to varying  $N_e^*$  across generations as well as sampling variation of  $\hat{G}$  within generations. An expression for this partitioning of separate variance components can be obtained by using the formula for the total unconditional variance of a random variable that is jointly distributed with another random variable (e.g., see Casella & Berger 1990, p. 158). For any two random variables  $X$  and  $Y$ , that formula for the variance of  $Y$  is  $\text{Var}(Y) = \text{Var}(E(Y | X)) + E(\text{Var}(Y | X))$ , provided the expectations exist. Note that  $E(Y | X)$  and  $\text{Var}(Y | X)$  are both random variables that are functions of the random variable  $X$ . This formula can be utilized here because the estimator  $\hat{G}$  is a random variable and is jointly distributed with the random variable  $N_e^*$ . Applied to our case, given a constant population size  $N$ , the variance of the estimator  $\hat{G}$  across replicate generations and replicate samples within generations is

$$\text{Var}(\hat{G}) = \text{Var}(E(\hat{G} | N_e^*)) + E(\text{Var}(\hat{G} | N_e^*)). \quad (\text{eqn 5})$$

In the special case (considered here) where  $\hat{G}$  is an unbiased estimator,  $E(\hat{G} | N_e^*) = G = f(1/N_e^*)$ , and  $\text{Var}(E(\hat{G} | N_e^*)) = \text{Var}(f(1/N_e^*))$ , which is the variance of realized parametric values of  $G$  across generations.

Eqn 6 (below) applies this case to eqn 5 and shows that overall variance of  $\hat{G}$  has two components: one related to variation of  $G$  itself, and another related to sampling variation of individual estimates around a given parameter value of  $G$ . This first term we will refer to as  $\text{Var}(\hat{G})_{\text{Demo}}$  since it quantifies the component of overall variance due to random variation in the demographic process of reproduction. The second term in eqn 6, the expected value of the variance among replicate samples (of individuals and/or

gene loci) within a population or generation, we will refer to as  $\text{Var}(\hat{G})_{\text{Gene}}$ .

$$\begin{aligned} \text{Var}(\hat{G}) &= \text{Var}(E(\hat{G} | N_e^*)) + E(\text{Var}(\hat{G} | N_e^*)) \\ &= \text{Var}(G) + E(\text{Var}(\hat{G} | G)) \\ &= \text{Var}(\hat{G})_{\text{Demo}} + \text{Var}(\hat{G})_{\text{Gene}}. \end{aligned} \quad (\text{eqn 6})$$

Analyses that consider only  $\text{Var}(\hat{G})_{\text{Gene}}$  under the implicit assumption that  $N_e$  is fixed thus can underestimate variability associated with modelled ideal populations. As a check, note that if  $N_e$  is fixed at a constant value, then  $\text{Var}(f(1/N_e)) = 0$  and  $E(\text{Var}(\hat{G} | N_e))$  equals the constant  $\text{Var}(\hat{G} | N_e)$ , leaving  $\text{Var}(\hat{G}) = \text{Var}(\hat{G} | N_e) = \text{Var}(\hat{G})_{\text{Gene}}$ . An additional point to note is that the total variance in eqns 5 and 6 is conditional on a constant value of population size  $N$ , even though it is not explicit in the notation (see Discussion for more on this point).

In practice, parameter values will not be known and estimators of these variance components will have to be used. Examples of variance component estimators can be found in many standard statistical texts that address analysis of variance (e.g., Kuehl 2000). In Appendix S3 (Supporting information) we provide the estimators we used in calculating variance components for the simulated data in the following examples, which illustrate points made above.

## Two Empirical Examples

### *Random changes in allele frequency*

Perhaps the most commonly modelled evolutionary process is random fluctuation of allele frequency at neutral gene loci (genetic drift). If we denote by  $P_0$  the population frequency of a neutral allele at a given gene locus at time 0, then the variance of allele frequency  $t$  generations later is given by (e.g., Crow & Kimura 1970)

$$\text{Var}(P_t | P_0) = E(P_t - P_0)^2 = P_0(1 - P_0) \left[ 1 - \left( 1 - \frac{1}{2N_e} \right)^t \right].$$

Variance effective size is defined as the value of  $N_e$  that, when substituted in the above equation, produces the realized variance in allele frequency. Since this variance is a function of initial allele frequency, it is often convenient to define a standardized variance,  $F$ , as

$$F = \frac{(P_t - P_0)^2}{P_0(1 - P_0)}; E(F | N_e) = \frac{E(P_t - P_0)^2}{P_0(1 - P_0)} = 1 - \left( 1 - \frac{1}{2N_e} \right)^t.$$

For simplicity, we deal with population frequencies (so  $F$  is a parameter) and we consider the special case of  $t = 1$ , so that  $E(F) = 1/(2N_e)$ .

Of course, the formulae above are only expectations that will be approached in the limit with a large number of

replications. We will consider two types of replication: multiple samples of gene loci within populations/generations and multiple populations. Each replicate population goes through an independent process of reproduction and hence has a different realized  $V_k^*$  and a different realized  $N_e^*$ . Let  $\hat{F}$  be the estimator of the value of  $F^* = 1/(2N_e^*)$  realized in a given generation. An analogue of eqn 6 can now be used to calculate the total standardized variance in allele frequency across replicate samples of genes in replicate ideal populations:

$$\begin{aligned} \text{Var}(\hat{F}) &= \text{Var}(\hat{F})_{\text{Demo}} + \text{Var}(\hat{F})_{\text{Gene}} \\ &= \text{Var}(E(\hat{F} | N_e^*)) + E(\text{Var}(\hat{F} | N_e^*)). \end{aligned} \tag{eqn 7}$$

Using  $E(\hat{F} | N_e^*) = 1/(2N_e^*)$ , we find  $\text{Var}(\hat{F})_{\text{Demo}} = \text{Var}[1/(2N_e)] = \text{Var}(1/N_e)/4$ . Using  $\text{Var}(V_k)$  from eqn 4 we find

$$\text{Var}\left(\frac{1}{N_e}\right) = \text{Var}\left[\frac{2 + V_k}{4N - 2}\right] = \frac{\text{Var}(V_k)}{(4N - 2)^2} = \frac{N - 1}{N^3(2N - 1)}. \tag{eqn 8}$$

Therefore, the first term in eqn 7 is given by  $\text{Var}(\hat{F})_{\text{Demo}} = (N - 1)/[4N^3(2N - 1)]$ .

The second term in eqn 7,  $\text{Var}(\hat{F})_{\text{Gene}} = E(\text{Var}(\hat{F} | N_e^*))$ , has attracted considerable interest in the population genetics literature. Lewontin & Krakauer (1973) proposed that this variance (conditional on a fixed  $N_e$ ) could be used in a test of neutrality, based on the premise that the quantity  $n_e\phi_F$  should be approximately 2 for neutral loci.  $n_F$  is the number of independent alleles used in the computation of  $\hat{F}$  and  $\phi_F = \text{Var}(\hat{F} | N_e)/E^2(\hat{F} | N_e)$  is the squared coefficient of variation of  $\hat{F}$  for a given  $N_e^*$ . If the underlying assumptions are satisfied, then  $n_F\hat{F}/E(\hat{F} | N_e)$  is expected to be distributed as Chi square with  $n_F$  degrees of freedom, in which case  $E(n_F\phi_F) = 2$ . Application of the Lewontin–Krakauer test to the study of population subdivision (using  $F_{ST}$ ) was criticized (Robertson 1975; Ewens & Feldman 1976) because other factors can lead to inter-locus variance in  $F_{ST}$ , and as a result this test has largely fallen into disuse (but see Beaumont 2005). However, the basic premise of the Lewontin–Krakauer test is sound for the study of temporal variation within a single population (Gaines & Whittam 1980).

Each  $F^*$  value is the realization from one episode of reproduction, so the Lewontin–Krakauer result is applicable to the variance of  $\hat{F}$  around a given fixed  $F^*$ . Using the assumption that  $n_F\phi_F \approx 2$ , we get  $\text{Var}(\hat{F} | N_e^*) \approx (2/n_F)E^2(\hat{F} | N_e^*)$ . Finding  $\text{Var}(\hat{F})_{\text{Gene}}$  requires taking the expectation of  $\text{Var}(\hat{F} | N_e^*)$  over the distribution of possible  $N_e^*$  values for a given constant  $N$  (see Appendix S4, Supporting information, for derivation). The result is the following expression for the second term in eqn 7:

$$\text{Var}(\hat{F})_{\text{Gene}} = E[\text{Var}(\hat{F} | N_e^*)] = \frac{2N^2 - 1}{2n_F N^3(2N - 1)}.$$

Combining this with the expression above for  $\text{Var}(\hat{F})_{\text{Demo}}$  we now have an expression for the overall variance in  $\hat{F}$  that accounts for average sampling variation of  $\hat{F}$  and variation of the parametric values of  $F$  due to variation in  $N_e^*$ :

$$\text{Var}(\hat{F}) = \frac{N - 1}{4N^3(2N - 1)} + \frac{2N^2 - 1}{2n_F N^3(2N - 1)}. \tag{eqn 9}$$

If the ‘-1’ terms in the above are ignored, eqn 9 simplifies to

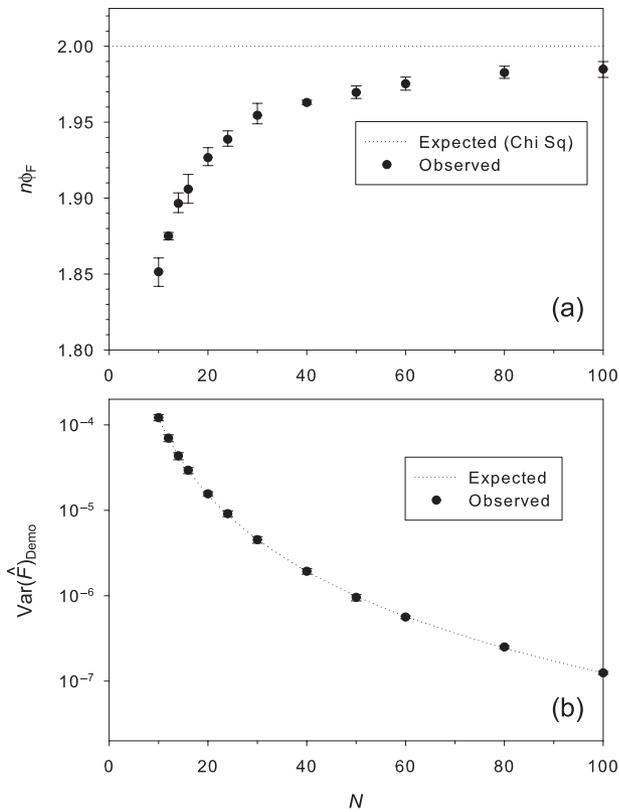
$$\text{Var}(\hat{F}) \approx \frac{1}{8N^3} + \frac{1}{2n_F N^2}. \tag{eqn 9a}$$

The first term in eqn 9a represents the increase in  $\text{Var}(\hat{F})$  compared to a model that assumes constant  $N_e = N$ . Based on this approximation, the two terms are equal when  $n_F = 4N$ . If  $n_F$  is smaller than this,  $\text{Var}(\hat{F})_{\text{Gene}}$  dominates the total variance; if  $n_F > 4N$ ,  $\text{Var}(\hat{F})_{\text{Demo}}$  is relatively more important. We have treated  $F$  as a population parameter and ignore variance associated with sampling individuals; if this factor were also included, we would see that  $\text{Var}(\hat{F})_{\text{Gene}}$  also gets smaller as the number of sampled individuals increases, as shown in the next example.

To evaluate eqn 9, we simulated one generation of genetic drift in  $R = 1$ –1000 replicate ideal populations (with random selfing), ranging in size from  $N = 10$ –100 individuals. Within each replicate population of size  $N$ , we generated many (1000) replicate estimates of  $\hat{F}$ , each based on allele frequencies at  $n_F = 1$ –100 independent, diallelic gene loci.  $E(F)$  is a function of the population allele frequency in the initial generation,  $P_0$  (see above). Since the population allele frequency is typically unknown, most estimators of  $F$  based on genetic samples estimate  $P_0$  as the mean of the sample frequencies in the initial and subsequent samples (e.g., Nei & Tajima 1981; Pollak 1983). In our simulations, however, the population allele frequency in the parental generation was fixed at  $P_0 = 0.5$  and we sampled the entire population, so we calculated  $\hat{F}$  as  $(P_1 - P_0)^2/[P_0(1 - P_0)] = (P_1 - 0.5)^2/[0.25]$ , where  $P_1$  is the population frequency in the progeny generation. We used the estimators described in Appendix S3 (Supporting information) to estimate the variance components  $\text{Var}(\hat{F})_{\text{Demo}}$  and  $\text{Var}(\hat{F})_{\text{Gene}}$  as well as the total  $\text{Var}(\hat{F})$ .

Results of these analyses showed the following. Overall mean  $\hat{F}$  asymptotically converged on  $1/(2N)$  for a large number of replicate populations. With a limited number of replicates, however, by chance, the grand mean  $\hat{F}$  deviated from the expectation, and the magnitude of the deviation was inversely related to the number of replicates (data not shown). This result demonstrates that lack of replication at the demographic (population/generation) level cannot be compensated for by more replication at the gene level.

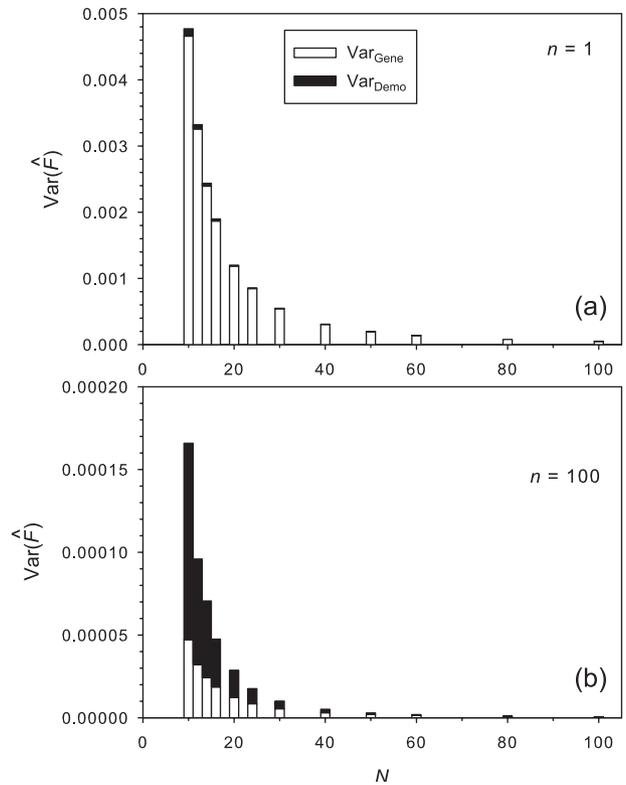
The variance of  $\hat{F}$  in the simulations agreed with expectations from eqn 9. For the  $\text{Var}(\hat{F})_{\text{Gene}}$  component,  $n_F\phi_F$  was slightly less than 2 for small  $N$  but asymptotically approached 2 as  $N$  increased (Fig. 5a). This result is



**Fig. 5** (a):  $n_F \phi_F$  for simulated populations with different numbers of ideal individuals ( $N$ ). Data shown are means (filled circles) and ranges (bars) of  $n_F \phi_F$  for simulations using 1000 replicate samples of  $n_F = 1, 2, 5, 10, 25, 50$  and 100 independent alleles.  $\phi_F$  is the estimated squared coefficient of variation of  $\hat{F}$ , based only on the estimated component  $\hat{V}ar(\hat{F})_{Gene}$ . Dotted line shows the Chi-squared expectation ( $n_F \phi_F = 2$ ). (b) Mean and range of  $\hat{V}ar(\hat{F})_{Demo}$  for the same values of  $n_F$  as in the top panel.  $\hat{V}ar(\hat{F})_{Demo}$  is the estimated variance of  $E(\hat{F} | N_e^*)$  across  $R = 1000$  replicate populations. Dotted line is the expectation based on the first term in eqn 9.

consistent with previous simulations by Nei & Tajima (1981), Pollak (1983) and Waples (1989), who found generally good agreement with the Lewontin–Kraekauer prediction that  $n_F \phi_F \approx 2$  but that in many cases the actual variance of  $\hat{F}$  was slightly lower than expected. We also found very good agreement between the empirical  $\hat{V}ar(\hat{F})_{Demo}$  and the theoretical expectation  $(N - 1) / [4N^2(2N - 1)]$  (Fig. 5b).

Inspection of eqn 9 shows that  $\hat{V}ar(\hat{F})_{Gene}$  becomes smaller as  $n_F$  increases (sampling variance of  $\hat{F}$  is smaller when more alleles are used to compute the mean), but the  $\hat{V}ar(\hat{F})_{Demo}$  term depends only on  $N$ . This means that as  $n_F$  increases,  $\hat{V}ar(\hat{F})_{Demo}$  becomes relatively more important to overall  $\hat{V}ar(\hat{F})$ . This effect is illustrated in Fig. 6. When each  $\hat{F}$  was computed using a single allele, overall variance of  $\hat{F}$  was large and  $\hat{V}ar(\hat{F})_{Gene}$  dominated (Fig. 6a). In contrast, with  $n_F = 100$  (easily achieved with highly variable markers like microsatellites),  $\hat{V}ar(\hat{F})_{Gene}$  is smaller than  $\hat{V}ar(\hat{F})_{Demo}$



**Fig. 6** Changes in the relative magnitude of the two estimated components of  $\hat{V}ar(\hat{F})$  as a function of  $n_F$  and  $N$ . Data are averaged over  $R = 1000$  replicate populations. (a) With only  $n_F = 1$  independent allele,  $\hat{V}ar(\hat{F})_{Gene}$  is large and dominates the total  $\hat{V}ar(\hat{F})$ . (b) When  $n_F$  is increased to 100,  $\hat{V}ar(\hat{F})_{Gene}$  is dramatically reduced and  $\hat{V}ar(\hat{F})_{Demo}$  (which does not vary with  $n_F$ ) becomes relatively more important. Note the change in scale of the vertical axis from panel a to b.

particularly for small  $N$  (Fig. 6b). We can also see that with  $n_F = 100$ ,  $\hat{V}ar(\hat{F})_{Gene}$  and  $\hat{V}ar(\hat{F})_{Demo}$  are approximately the same magnitude for  $N = 25$ , in agreement with the expectation (eqn 9a) that the two variance components are equal when  $n_F = 4N$ .  $n_F \phi_F$  is considerably inflated compared to the Lewontin–Kraekauer theoretical expectation for  $n_F = 100$  (Fig. 7; filled symbols and dashed line).

### Linkage disequilibrium

As a second example, consider a scenario in which genetic data are used to estimate effective population size using an index of linkage disequilibrium,  $\hat{r}^2$ . This method is based on the theoretical expectation (Weir & Hill 1980; Hill 1981) that, for random mating, unlinked loci and a sample of  $S$  diploid individuals for which multilocus genotypes can be determined but gametic phase is unknown,

$$E(\hat{r}^2 | N_e) \approx \frac{1}{3N_e} + \frac{1}{S}. \tag{eqn 10}$$

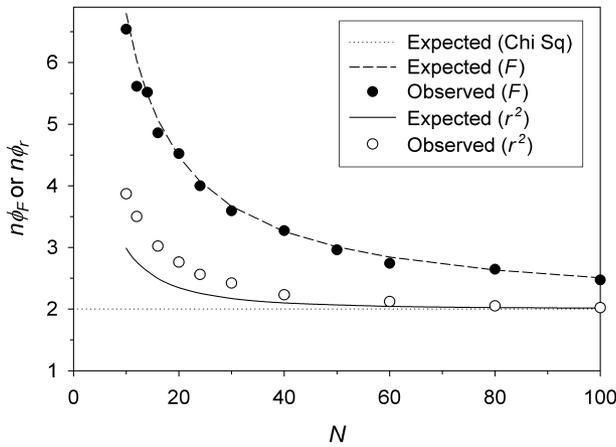


Fig. 7 The ratio  $n_{\hat{\phi}_F}$  or  $n_{\hat{\phi}_r}$  as a function of  $N$ .  $\hat{\phi}$  includes both estimated variance components,  $\text{Var}_{\text{Demo}}$  and  $\text{Var}_{\text{Gene}}$ . Filled circles: empirical data for  $\hat{F}$  from simulated populations of  $N = 10\text{--}100$  ideal individuals. Mean  $\hat{F}$  was averaged over 1000 replicates of  $n_r = 100$  independent alleles and 1000 replicate populations. Open circles: Empirical data for  $\hat{r}^2$  from Waples (2006). In each of 180 000 replicate generations,  $\hat{r}^2$  was computed as the mean across all  $n_r = 28$  pairwise comparisons of eight diallelic gene loci using a sample of 100 individuals. Dotted line: the Chi-squared expectation ( $n\phi = 2$ ). Dashed curve and solid curve: expectations calculated from eqns 9 and 12, respectively.

If information is available for  $L$  diallelic loci, then an average  $\hat{r}^2$  can be calculated as the mean across all  $n_r = L(L - 1)/2$  pairs of loci for a particular sample of individuals. The variance of  $\hat{r}^2$  is large, so it is important to evaluate precision that can be expected from a sample scored for a plausible number of gene loci. Note that from here on, we use the notation  $\hat{r}^2$  to represent the mean  $\hat{r}^2$  across  $n_r$  pairs of loci. In theory, the quantity  $n_r \hat{r}^2 / E(\hat{r}^2 | N_e)$  follows a Chi-squared distribution with  $n_r$  degrees of freedom (Hill 1981), in which case we expect that  $n_r \phi_r \approx 2$ , where  $\phi_r = \text{Var}(\hat{r}^2 | N_e) / E^2(\hat{r}^2 | N_e)$ . It follows that  $\text{Var}(\hat{r}^2 | N_e) \approx (2/n_r)E^2(\hat{r}^2 | N_e)$ . Figure 7 (open symbols) shows the empirical distribution of  $n_r \hat{\phi}_r$  for simulated ideal populations of different size  $N$ . In this example, drawn from data in Waples (2006),  $\hat{r}^2$  in each replicate was calculated as the mean over  $n_r = 28$  pairwise comparisons of eight diallelic gene loci. The estimated total variance and overall mean of  $\hat{r}^2$  across replicated samples and generations was used to compute  $\hat{\phi}_r$  for a given  $N$ . Although agreement with theoretical expectations was good for  $N \geq 100$ , the variance of  $\hat{r}^2$  was higher than expected for smaller  $N$ , and as a consequence  $n_r \hat{\phi}_r$  was higher than the Chi-square expectation of 2 and rose sharply when  $N$  dropped below about 50.

Waples (2006) speculated that the inflated variance for small  $N$  was a consequence of random fluctuation in realized  $N_e^*$  across replicates. To evaluate quantitatively whether this additional source of random variation can

explain the results shown in Fig. 7, we can use an analogue of eqn 7:

$$\begin{aligned} \text{Var}(\hat{r}^2) &= \text{Var}(\hat{r}^2)_{\text{Demo}} + \text{Var}(\hat{r}^2)_{\text{Gene}} \\ &= \text{Var}[E(\hat{r}^2 | N_e^*)] + E(\text{Var}(\hat{r}^2 | N_e^*)). \end{aligned} \quad (\text{eqn 11})$$

Using the formula for  $E(\hat{r}^2 | N_e)$  from eqn 10 and the expression for  $\text{Var}(1/N_e)$  from eqn 8,

$$\text{Var}(\hat{r}^2)_{\text{Demo}} = \text{Var}\left(\frac{1}{3N_e^*} + \frac{1}{S}\right) = \frac{1}{9} \text{Var}\left(\frac{1}{N_e^*}\right) = \frac{N-1}{9N^3(2N-1)}.$$

Finding  $\text{Var}(\hat{r}^2)_{\text{Gene}}$  requires finding the expectation of  $\text{Var}(\hat{r}^2 | N_e)$  over the distribution of possible  $N_e^*$  values for a given  $N$ . The derivation is shown in Appendix S4 (Supporting information), the result being

$$\text{Var}(\hat{r}^2)_{\text{Gene}} = \frac{2}{n_r} \left[ \frac{2N^2 - 1}{9N^3(2N-1)} + \frac{3N + 2S}{3NS^2} \right].$$

Putting  $\text{Var}(\hat{r}^2)_{\text{Demo}}$  and  $\text{Var}(\hat{r}^2)_{\text{Gene}}$  together gives

$$\text{Var}(\hat{r}^2) = \frac{N-1}{9N^3(2N-1)} + \frac{2}{n_r} \left[ \frac{2N^2 - 1}{9N^3(2N-1)} + \frac{3N + 2S}{3NS^2} \right]. \quad (\text{eqn 12})$$

If the ‘-1’ terms are ignored, eqn 12 simplifies to

$$\begin{aligned} \text{Var}(\hat{r}^2) &\approx \frac{1}{18N^3} + \frac{2}{n_r} \left[ \frac{1}{9N^2} + \frac{3N + 2S}{3NS^2} \right] \\ &= \frac{1}{18N^3} + \frac{2}{n_r} [E^2(\hat{r}^2)]. \end{aligned}$$

The term  $1/(18N^3)$  thus quantifies the component of  $\text{Var}(\hat{r}^2)$  due to random variation in  $N_e^*$ . Note that  $\text{Var}(\hat{r}^2)_{\text{Gene}}$  is inversely related to both  $n$  and  $S$ , so  $\text{Var}(\hat{r}^2)_{\text{Demo}}$  becomes relatively more important as samples of loci, alleles, and/or individuals increase.

It is apparent from Fig. 7 (solid curve) that eqn 12 predicts the general form of the inflated variance in  $\hat{r}^2$  but cannot account for all of the effect seen in this example. We quantitatively evaluated two factors related to the model that might have affected results. First, eqn 9 is biased except for certain values of  $S/N_e$  because the derivation ignores second-order terms (England *et al.* 2006; Waples 2006). We derived expectations for  $\text{Var}(\hat{r}^2)_{\text{Demo}}$  and  $\text{Var}(\hat{r}^2)_{\text{Gene}}$  that account for these second-order terms, but this did not improve agreement with observed results (data not shown). Because use of second-order terms required tedious derivations of  $\text{Var}(1/N_e^2)$  and  $\text{Cov}(1/N_e, 1/N_e^2)$ , in the example above we present analytical results for the simpler eqn 9. Second, ‘observed’ data shown in

Fig. 7 were obtained from a model with separate sexes, whereas the above analyses assume random selfing. To evaluate this factor, we simulated new multilocus genetic data using a different computer program and a mating system that allowed random selfing. We calculated ( $\hat{f}^2$ ) as described in Waples (2006) and obtained qualitatively similar results (higher  $\text{Var}(\hat{f}^2)$  in the simulations than expected from eqn 12; data not shown); therefore, the discrepancy evident in Fig. 7 cannot be attributed to the mating system or vagaries of the simulation algorithm. In the example shown in Fig. 7,  $\text{Var}(\hat{f}^2)$  was computed over a very large number ( $R = 180\,000$ ) of replicate generations, so a lack of demographic replication was not responsible for the results.

One factor likely to be important is that, because linkage disequilibrium at unlinked loci decays by only 50% per generation,  $\hat{f}^2$  depends not only on  $N_e^*$  in the parental generation, but also to a lesser extent on  $N_e^*$  in the previous few generations. As a consequence,  $\text{Var}(\hat{f}^2)_{\text{Demo}}$  is not fully accounted for by the above formulas for  $\text{Var}(V_k)$  and  $\text{Var}(N_e)$ . Furthermore, as decay of disequilibrium is a random process at each pair of loci, it can be expected that for any given sequence of  $N_e^*$  values, the amount of residual disequilibrium will vary among pairs of loci, which would tend to inflate  $\text{Var}(\hat{f}^2)_{\text{Gene}}$ . Thus, both components of  $\text{Var}(\hat{f}^2)$  are probably larger than expected based on eqn 12.

Another possible factor is the sensitivity of  $\hat{f}^2$  to extreme allele frequencies. Although  $\hat{f}^2$  is a standardized form of linkage disequilibrium, this does not entirely remove effects of allele frequency (Maruyama 1982; Hedrick 1987). Use of rare alleles biased tests of linkage disequilibrium reported by Waples & Smouse (1990). For the simulated data considered here, Waples (2006) excluded alleles with frequencies  $> 0.95$  or  $< 0.05$  and found little overall effect of allele frequency. However, based on results obtained by Waples & Do (2008), use of alleles with frequency less than about 0.1 can downwardly bias  $\hat{f}^2$ . Hence, some locus pairs in the simulations conducted by Waples (2006) were probably differentially affected by rare alleles, and this would increase the variance among replicate  $\hat{f}^2$  values. Notably, this effect should be largest with small  $N$ , when genetic drift is strongest and alleles can rapidly drift to extreme frequencies.

## Discussion

That genetic drift is a stochastic process, and hence its consequences can be predicted only in a statistical sense, are central tenets of theoretical and applied population genetics. Over a single generation, the frequency of a selectively neutral allele might increase by a small or large amount, decrease by a small or large amount, or remain unchanged. Although what occurs in a single episode of drift cannot be predicted reliably, strong statements about average behaviour (over many replicate episodes of

genetic drift) can be made, based on the effective size of the population in question.

Here, we consider a slightly different type of random variation that occurs in modelled, 'ideal' Wright–Fisher populations. As originally envisioned by Fisher and Wright, the concept of an ideal population is elegantly simple: each individual contributes equally to a large pool of gametes which unite at random to produce the next generation. If the gamete pool is large enough, sampling can be assumed to be binomial, and a binomial distribution of offspring number produces an effective size equal to the number of ideal individuals. Theoretical population genetics relies heavily on ideal populations of  $N$  individuals, and in these hypothetical populations it is easy to imagine that  $N_e = N$ , always and by definition.

It is generally accepted that ideal populations are unlikely to occur in nature, but it is easy to model the Wright–Fisher process in a computer, in which case the resulting populations are no longer hypothetical but acquire specific characteristics. This paper deals with properties of these *in silico* ideal populations – specifically, the fact that even though conditions of a Wright–Fisher population are exactly satisfied, the result is random variation in realized  $V_k^*$  and hence  $N_e^*$  in replicate generations. Curiously, this topic has received little or no attention in the literature – a potentially important omission, as use of the Wright–Fisher model to create populations of different 'known' effective sizes is pervasive in evolutionary biology. In recent years, several factors (ready availability of large numbers of highly variable genetic markers; rapid increases in computational power; development of sophisticated software programs for population genetic data analysis) have conspired to produce an information explosion and a need for (and interest in) exploring contemporary evolutionary processes using individual-based models that track multilocus genotypes. Notably, the effects described here are relatively more important when other sources of sampling error are small (e.g., with large numbers of individuals, loci and alleles) and with relatively small populations (which are frequently modelled by those interested in conservation).

Does effective size actually vary in these *in silico* ideal populations? Or is parametric effective size fixed at  $N_e = N$ , in which case the random demographic processes occurring in individual generations can be viewed simply as manifestations of the inherently stochastic process of random genetic drift? This is a rather philosophical question that has little practical relevance to the phenomena described here, since they occur regardless how this issue is resolved. We favour the former view, because the quantity we call 'realized  $N_e^*$ ' is the value one gets by inserting the actual  $V_k^*$  for a particular generation into a standard formula for effective size (e.g., eqn 1 or comparable equations in Crow & Kimura 1970; Crow & Denniston 1988; Caballero 1994, etc.).

Furthermore, as data are averaged across more and more genes, the rates of inbreeding and allele-frequency changes for that generation converge on a function of  $1/N_e^*$ , not  $1/N$ .

A multilocus computer model of the Wright–Fisher process involves two kinds of replication (Appendix S5, Supporting information): replicating the process of reproduction within a population of fixed size  $N$ , and replicating the process of sampling genes contingent on a particular demography. The first type of replication leads to variation in realized  $N_e^*$ , which is primarily responsible for the phenomena described here. Variation in  $N_e^*$  is most apparent for small  $N$  but can also be relatively important when large amounts of data make other sources of variation relatively small. If  $N_e^*$  were equal to  $N$  every generation (which is essentially true for very large  $N$ ), the two types of replication would be equivalent, in which case data generated for  $A$  replicate generations and  $B$  replicate loci in each generation would be comparable to data for  $A \times B$  replicate genes in a single generation, or to data for a single gene in  $A \times B$  replicate generations. In a modelled Wright–Fisher population, however, the two types of replication are not interchangeable, since within a given generation the behaviour of all genes is affected by the same fixed pedigree. That is, the different genes are not true replicates of a process generated by  $N$  ideal individuals; they are replicates of a process determined by realized  $N_e^*$  in a single generation. This leads to a type of pseudoreplication (Hurlbert 1984), which adds some complications but also provides additional opportunities for analyzing subsets of the simulated data (see Appendix S5, Supporting information).

Here is another way of thinking about the two-step process of modelling ideal populations. At the start of each replicate generation, the reference population is defined by the parameters  $N$  and  $P$ . From this point of reference, the expected value of  $F$  after a single generation of drift is  $E(F) = 1/(2N)$ . In the first step of the model, choose the parents of the next generation by random sampling, with replacement. This will produce a vector of  $k_i$  values and a realized  $V_k^*$ . The second step is to choose the genes each parent will contribute to the next generation. For this step,  $E(F)$  is no longer  $1/(2N)$  but rather  $1/(2N_e^*)$ , where  $N_e^*$  is determined by the  $V_k^*$  for that replicate generation. As a consequence of this random variation in  $N_e^*$ , overall variance of  $\hat{F}$  across multiple loci and multiple generations in a modelled Wright–Fisher population will be higher than it would be if  $N_e^*$  were equal to  $N$  every generation. The quantity we call  $\text{Var}_{\text{Demo}}$  quantifies this additional variance component. Although  $\text{Var}_{\text{Demo}}$  is implicit in the Wright–Fisher model, the fact that its relative importance increases with  $n$  and  $S$  has not been explored in the literature.

Not all computer models involving effective population size will produce these phenomena. Coalescent models (Kingman 1982; Wakeley 2008) differ from those considered here in that, even when multiple genes are considered,

each is typically modelled as a separate realization of the coalescent process, and  $N_e$  is used as a fixed scaling parameter. Models that employ the standard coalescent do not simulate multilocus genotypes transmitted across generations according to a specific pedigree and hence should not be affected by the phenomena described above. In other cases, the effects will occur but be insignificantly small. Consider, for example, the simplest way to model genetic drift in a population with a specified  $N_e$ : for each of many replicate generations, draw  $2N$  genes binomially (and independently) using a fixed allele frequency from the parental generation as the binomial parameter (see Scenario A in Fig. S5 (Supporting information) for a depiction of this model). This is the type of model used by Lewontin & Krakauer (1973), Nei & Tajima (1981), Pollak (1983), and Waples (1989) to evaluate variance of  $\hat{F}$ . In this model, the intrinsic variance of single-locus estimates  $\hat{F}$  is so large that the  $\text{Var}_{\text{Demo}}$  component is small and easily overlooked (Fig. 6a), which explains why the authors mentioned above found generally good agreement with the theoretical expectation (assuming a fixed  $N_e$ ) that  $n_e \phi_F \approx 2$ .

In what follows, we briefly summarize the principal results of our study and discuss some practical consequences. Although equations derived above and in the Supporting information follow the original definition of an ideal population and assume random selfing in monoecious populations, the same issues apply to species with separate sexes, for which only minor modifications are required to the quantitative results presented here.

### Bias

In a Wright–Fisher model, mean  $V_k^*$  is an unbiased estimator of the binomial variance. However, because of the inverse relationship between  $V_k$  and  $N_e$ , the expected value of  $N_e^*$  is slightly higher than  $N$ . This effect is small and of little practical consequence, because most evolutionary processes scale with  $1/N_e$  and  $E(1/N_e^*) = 1/N$ . Modellers are typically interested in ‘average’ behaviour, which is determined by the harmonic mean  $N_e^*$  and converges to  $N$  with adequate replication. Note, however, that lack of replication of the process of reproduction cannot be compensated for by extensive replication of genetic sampling on a fixed pedigree. Furthermore, as discussed below, even with adequate replication of reproduction, some evaluations are strongly affected by realized  $N_e^*$  in individual generations.

### Precision

Random variation in  $V_k^*$  and  $N_e^*$  can strongly affect precision, but the consequences for evolutionary inference depend heavily on the context. Below, we identify situations where this phenomenon is (and is not) particularly important to consider (see also Appendix S5, Supporting information).

*Simulated data.* When Wright–Fisher populations are modelled by computer, random variation in realized  $N_e^*$  generates a component to the total variance of a genetic index ( $\text{Var}_{\text{Demo}}$ ) that is not generally accounted for in standard treatments (cf. Supporting information, eqns d.3 and d.6). As a consequence, variances of genetic indices related to  $N_e$  will be higher in modelled ideal populations than predicted based on the implicit assumption that  $N_e^* = N$  in every generation. Because random variations in  $N_e^*$  arise directly from the modelling process, they are embedded in the simulated data and will affect performance evaluations of any method used to analyze such data [e.g., moment-based, likelihood-based, or Approximate Bayesian Computation (ABC) methods].

Two factors exacerbate these effects by increasing the ratio  $\text{Var}_{\text{Demo}}/\text{Var}_{\text{Gene}}$  and hence the relative contribution of  $\text{Var}_{\text{Demo}}$  to the overall variance: (i) small  $N$ ; and (ii) high replication of genetic sampling (large samples of loci/alleles and/or individuals). The latter is a key point because it means this effect will become increasingly important to consider as technical advances continually increase the number of genes and individuals that can be assayed in the field and modelled by computer. These effects can be strongly influenced by methods of simulating data and replicating the process of genetic drift; see Appendix S5 (Supporting information) for discussion and examples.

A common performance measure is mean squared error (MSE), which reflects both precision and bias. For example, Faubet *et al.* (2007) used simulated data to evaluate performance of a Bayesian method for studying contemporary migration (Wilson & Rannala 2003). One of the performance criteria Faubet *et al.* (2007) used was MSE of the estimated migration rate. This metric probably was affected at least indirectly by random variation in  $N_e^*$ , which would tend to produce more variable estimates than would occur if the implicit assumption that  $N_e^* = N$  in every generation were actually true.

The problems that can result from assuming that  $N_e^* = N$  in every replicate also apply to evaluation of another common performance metric – the fraction of confidence intervals or credible intervals (CIs) for a point estimate that contain the true value. In general, one expects, for example, that 95% of such 95% CIs will contain the true value. With simulated Wright–Fisher data, however, the CIs are computed with respect to the point estimate associated with realized  $N_e^*$  for that generation, but they are being compared with the nominal  $N_e = N$ . As we have seen (Figs 3 and 4), only rarely, and only by chance, will  $N_e^*$  exactly equal  $N$  in any given generation. This means that, in modelled ideal populations, random variation in  $N_e^*$  will lead to overly pessimistic conclusions about performance of methods for generating CIs. Furthermore, as more and more data are collected (samples of individuals, loci and alleles), CIs will converge more narrowly on  $N_e^*$  and have less chance of including the parametric  $N_e = N$ . This will produce the

counter-intuitive result that as the estimator becomes more precise, CIs will include the ‘true’ value a smaller fraction of the time. Note that this problem will occur even if the estimator is unbiased; it arises from falsely assuming that realized  $N_e^*$  is fixed and equal to  $N$  in every generation.

To illustrate the problem this poses, assume one is interested in using simulated data to evaluate performance of methods for estimating  $N_e$  from a single sample (Tallmon *et al.* 2008; Waples & Do 2008; Zhdanova & Pudovkin 2008). Figure 4 shows a typical series of replicated ideal populations with  $N = 100$ . Harmonic mean  $N_e^*$  (100.7) was close to the nominal expectation that  $N_e = N$ ; however,  $N_e^*$  in individual generations varied considerably (e.g., in generation 2,  $V_k^* = 1.3$  and  $N_e^* = 121$ ). For that generation, with limited amounts of data available, the CI for the point estimate is wide and incidentally includes the nominal ‘true’ value  $N_e = 100$ . However, with sample sizes of individuals, loci and alleles that are routinely achievable today, the parametric CI is narrow and does not include 100 (Fig. 4). We have found that this issue has made it challenging to evaluate performance of confidence intervals for linkage-disequilibrium based estimates of  $N_e$  (Waples & Do 2008 and unpublished data).

At a minimum, researchers generating or analyzing simulated data should be aware of the consequences of random variation in  $N_e^*$ . In some cases, depending on how the data are simulated and analyzed and what the objectives are (see Appendix S5, Supporting information for discussion), it might be desirable to modify the algorithm to generate populations with constant realized  $N_e^* = N$ . Here is one way this might be done: (i) identify a target  $N_e$  to model; (ii) create a population of  $N = N_e$  individuals; (iii) choose  $N$  random numbers from a Poisson distribution with  $\mu = 2$ , each number being the  $k_i$  value for one individual; (iv) use a random procedure to adjust the vector of  $k_i$  values to produce  $\bar{k}_i = 2$  and  $\text{Var}(k) = 2(N - 1)/N$ ; this will produce  $N_e^* = N$  exactly; (v) develop a system of random mating (two parents for each offspring) that will produce this vector of  $k_i$  values; (vi) apply this algorithm to produce the offspring generation; and (vii) repeat step vi for each successive generation, randomly choosing individuals to be matched with the same vector of  $k_i$  values. Note that for each  $N$ , steps (i)–(v) only need to be performed once. Steps (iv) and (v) require a little thought but should be feasible. For example, for any given number of ideal individuals except  $N = 3$ , it is possible to find one or more vectors of  $k_i$  values that satisfy both  $\bar{k} = 2$  and  $\text{Var}(k) = 2(N - 1)/N$  (Fig. 3). A ‘brute-force’ variation of this approach would be to simulate ideal populations with variable  $N_e^*$  but only accept generations in which  $N_e^* = N$ .

*Data from natural populations.* The ability to massively replicate demographic and evolutionary processes is a core strength of computer simulations, and replicating

the Wright–Fisher process of reproduction across multiple generations leads to the  $\text{Var}_{\text{Demo}}$  component that is responsible for most of the phenomena described above. In contrast, any given natural population has only one realization of the process of reproduction (and one realized  $N_e^*$ ) in each generation. Therefore, with a couple of exceptions, these phenomena should not affect analyses of samples from natural populations. The first exception would occur if data for a number of natural populations (or a number of generations within one population) are being treated and analyzed as replicates. In this case, realized  $N_e^*$  should vary even more than it does in ideal populations because  $N$  would almost certainly vary among replicate natural populations as well.

The second exception involves ABC methods (e.g., Beaumont *et al.* 2002), where one simulates data for a wide range of values of the parameter(s) of interest and finds the value that produces results that agree best with the observed data. For example, Tallmon *et al.* (2004, 2008) used ABC to estimate  $N_e$  using variations of the temporal and linkage disequilibrium methods, respectively. These ABC methods are designed to estimate effective size based on samples from real populations, so they are estimating the realized  $N_e^*$  that applies to the generation(s) that produced the data. Ideally, then, an ABC program would simulate data using a range of fixed and known  $N_e$  values to generate a posterior distribution of  $\hat{N}_e$ . However, simulating multilocus genotypes with a Wright–Fisher model does not produce a fixed  $N_e^*$ , which introduces an additional source of variation. To take a concrete example, let's say one has a single-generation sample from a natural population, and that the true (realized but unknown) effective size for that generation is  $N_e^* = 76$ . Therefore, one would expect that when ABC methods are used to simulate data, the runs with  $N = 76$  ideal individuals would produce the best match to the observed data. That should be true if effective size was exactly 76 in every one of the simulated generations. However, simulating the Wright–Fisher process with  $N = 76$  ideal individuals produces a distribution of realized  $N_e^*$  values (Figs 2 and 4). Although this random variation might not bias the point estimate  $\hat{N}_e$  (since every nominal value of  $N_e$  has the same uncertainty), it should affect the posterior distribution of  $\hat{N}_e$ , making it broader than it would be if  $N_e^*$  actually were fixed at 76 in every generation.

Whereas it is easy to constrain  $N$  to a fixed value in a computer, size varies over time in all natural populations, which means that realized  $N_e^*$  across time will be a complex function of several random (and perhaps some directional) processes. A simple framework that incorporates two types of stochastic processes – random variation in  $V_k^*$  and random variation in  $N$  – could be developed and might serve as a useful reference point for more detailed evaluations of this topic. Assume that in generation  $t$  in a given natural population of size  $N_t$  each individual is equally

likely to contribute genes to the next generation and that  $E(k_i) = 2$ . The distribution of progeny number could thus be modelled (as suggested above) by choosing  $N$  values of  $k_i$  from a Poisson distribution with  $\mu = 2$ . This would produce a realized  $V_k^*$  which is the variance of the  $k_i$ , a realized  $\bar{k}^* = \sum k_i/N$ , and a realized  $N_e^*$  which is defined by eqn 1, given  $V_k^*$  and  $\bar{k}^*$ . Furthermore, this process would produce a realized population size in the next generation, defined by  $N_{t+1} = \sum k_i/2$ . In this model, replication across time (multiple sequential generations) produces a Markov chain of realized  $N_t$  and  $N_e^*$  values, driven by random variation in  $V_k^*$  and  $\bar{k}^*$ .

## Conclusions

We can summarize the major points that emerge from our evaluations:

- In a Wright–Fisher computer model, realized  $N_e^*$  varies randomly among generations with variance  $\sim N/2$ .
- Harmonic mean  $N_e^*$  converges on  $N$  with increasing replication of the process of reproduction, so the long-term evolutionary behaviour in modelled populations will be comparable to that for populations with constant  $N_e^* = N$ .
- Replication of the process of sampling genes on a fixed pedigree cannot compensate for lack of replication of the process of reproduction.
- Analyses of simulated data that consider variances of quantities related to  $N_e$  can be strongly affected by random variation in  $N_e^*$ .
- Analyses or performance measures that depend on the value of realized  $N_e^*$  in individual generations can be strongly affected.
- These latter effects are more pronounced when  $N$  is small (as it often is for populations of conservation interest) and/or the amount of data (loci, alleles, individuals) is large.

Related phenomena occur in modelling other evolutionary forces characterized by fixed parameters that describe an underlying stochastic process. For example, it is common to model migration among subpopulations using fixed values of the migration rate,  $m$ . However, in many implementations,  $m$  only describes a probability of migration, and, if so, the realized fraction of migrants ( $m^*$ ) will vary randomly across populations and generations. Variable migration can have consequences that differ from constant migration having the same mean (Whitlock & McCauley 1999), and in some circumstances these effects could be important to consider. Similarly, if  $N$  otherwise ideal individuals are modelled and the sex ratio is allowed to vary randomly (with each individual having an equal probability of being male or female), effective size will be reduced by an average of one individual (see Appendix in Waples & Do 1994).

We recommend that scientists who use or generate simulated Wright–Fisher data should evaluate whether these factors are likely to affect results. Furthermore, it would be prudent for anyone who uses computer programs whose performance has been evaluated with simulated data for ideal populations to consider to what extent those evaluations might have been affected by these factors.

## Acknowledgements

We thank Nick Barton, Joe Felsenstein, Bill Hill and, in particular, Eric Anderson for stimulating discussions and Oscar Gaggiotti, Nils Ryman, Steve Smith and three anonymous reviewers for useful comments on an earlier draft of the manuscript. This paper benefitted from discussions with the Genetic Monitoring (GeM) Workgroup, co-sponsored by the National Center for Ecological Analysis and Synthesis (NCEAS) and the National Evolutionary Synthesis Center (NESCENT).

## References

- Beaumont MA (2005) Adaptation and speciation: what can  $F_{ST}$  tell us? *Trends in Ecology and Evolution*, **20**, 435–440.
- Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.
- Caballero A (1994) Developments in the prediction of effective population size. *Heredity*, **73**, 657–679.
- Casella G, Berger RL (1990) *Statistical Inference*. Duxbury Press, Belmont, CA.
- Crow JF, Denniston C (1988) Inbreeding and variance effective population numbers. *Evolution*, **42**, 482–495.
- Crow JF, Kimura M (1970) *An Introduction to Population Genetics Theory*. Harper & Row, New York.
- England PR, Cornuet J-M, Berthier P, Tallmon DA, Luikart G (2006) Estimating effective population size from linkage disequilibrium: severe bias using small samples. *Conservation Genetics*, **7**, 303–308.
- Ewens MJ, Feldman MW (1976) The theoretical assessment of selective neutrality. In: *Population Genetics and Ecology* (eds Karlin S, Nevo E), pp. 303–338. Academic Press, New York.
- Faubet F, Waples RS, Gaggiotti OE (2007) Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. *Molecular Ecology*, **16**, 1149–1166.
- Fisher RA (1930) *The Genetical Theory of Natural Selection*. Oxford University Press, New York.
- Gaines MS, Whittam T (1980) Genetic changes in fluctuating vole populations: selective vs. nonselective forces. *Genetics*, **96**, 767–778.
- Hedrick PH (1987) Gametic disequilibrium: proceed with caution. *Genetics*, **117**, 331–341.
- Hill WG (1981) Estimation of effective population size from data on linkage disequilibrium. *Genetical Research*, **38**, 209–216.
- Hurlbert SH (1984) Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, **54**(2), 187–211.
- Jensen JLWV (1906) Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, **30**, 175–193.
- Kingman JFC (1982) On the genealogy of large populations. *Journal of Applied Probability*, **19A**, 27–43.
- Kuehl RO (2000) *Design of Experiments: Statistical Principles of Research Design and Analysis*, 2nd edn. Duxbury Press, Pacific Grove, CA.
- Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.
- Manel S, Gaggiotti O, Waples RS (2005) Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology and Evolution*, **20**, 136–142.
- Maruyama T (1982) Stochastic integrals and their application to population genetics. In: *Molecular Evolution, Protein Polymorphism, and the Neutral Theory* (Ed. Kimura M), pp. 151–166. Springer-Verlag, Berlin.
- Nei M, Tajima F (1981) Genetic drift and estimation of effective population size. *Genetics*, **98**, 625–640.
- Pearse DE, Crandall KA (2004) Beyond  $F_{ST}$ : population genetic analysis for conservation. *Conservation Genetics*, **5**, 585–602.
- Pollak E (1983) A new method for estimating the effective population size from allele frequency changes. *Genetics*, **104**, 531–548.
- Robertson A (1975) Gene frequency distribution as a test of selective neutrality. *Genetics*, **81**, 775–785.
- Tallmon DA, Luikart G, Beaumont MA (2004) Comparative evaluation of a new effective population size estimator based on approximate Bayesian computation. *Genetics*, **167**, 977–988.
- Tallmon DA, Koyuk A, Luikart G, Beaumont MA (2008) ONESAMP: a program to estimate effective population size using approximate Bayesian computation. *Molecular Ecology Resources*, **8**, 299–301.
- Wakeley J (2008) *Coalescent Theory: An Introduction*. Roberts and Company, Denver, CO.
- Waples RS (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics*, **121**, 379–391.
- Waples RS (2006) A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conservation Genetics*, **7**, 167–184.
- Waples RS, Do C (1994) Genetic risk associated with supplementation of Pacific salmonids: captive broodstock programs. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**(Suppl. 1), 310–329.
- Waples RS, Do C (2008)  $LdNe$ : a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources*, **8**, 753–756.
- 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, Hill WG (1980) Effect of mating structure on variation in linkage disequilibrium. *Genetics*, **95**, 447–488.
- Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration:  $F_{ST}$  doesn't equal  $1/(4Nm + 1)$ . *Heredity*, **82**, 117–125.
- Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177–1191.
- Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.
- Zhdanova OL, Pudovkin AI (2008)  $N_B\_HETEx$ : a program to estimate the effective number of breeders. *Journal of Heredity*, **99**, 694–695.

---

A major thrust of Robin Waples' work is to apply population genetic principles to real-world problems in conservation and management. He never ceases to be amazed by how an elegant concept like effective population size becomes hideously complex as soon as practical realities intrude on idealized models. Jim Faulkner is interested in developing statistical methods with applications in ecology and conservation biology. Much of his research focuses on modeling and estimation related to population processes of anadromous fishes.

---

### Supporting Information

Additional supporting information may be found in the online version of this article.

**Appendix S1** Derivation of the Variance of  $V_k$

**Appendix S2** Approximations to the Expected Value and Variance of  $N_e^*$

**Appendix S3** Variance Component Estimators used in Simulations

**Appendix S4** Derivation of  $\text{Var}(\hat{F})_{Gene}$  and  $\text{Var}(\hat{r}^2)_{Gene}$

**Appendix S5** Scenarios that can arise in modeling ideal populations

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.