

# Fixation probability of beneficial mutations in a fluctuating population

STEINAR ENGEN<sup>1\*</sup>, RUSSELL LANDE<sup>2,3</sup> AND BERNT-ERIK SÆTHER<sup>3</sup>

<sup>1</sup>Centre for Conservation Biology, Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway

<sup>2</sup>Division of Biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK

<sup>3</sup>Centre for Conservation Biology, Department of Biology, Norwegian University of Science and Technology, N-7491 Trondheim, Norway

(Received 29 September 2008 and in revised form 5 December 2008)

## Summary

We compute an accurate approximation to the probability of fixation for a beneficial mutation in a population fluctuating with a stationary distribution of population size. The population dynamics are described by the theta-logistic model with environmental variance, assuming that the population size is large enough to ignore demographic variance. We show that stochastic fluctuations of population size reduce the probability of fixation. However, it is not the magnitude of the population fluctuations *per se* that creates this reduction. Only the environmental variance has a substantial effect on the probability of fixation. The strength of density dependence (or expected return time to equilibrium) and the functional form of density-regulation, given by the parameter  $\theta$  in the theta-logistic model, have little effect on the fixation probability. Effective population size based on harmonic mean population size will therefore underestimate the expected fixation rate of beneficial mutations in fluctuating populations.

## 1. Introduction

The classical result of Haldane (1927) that a beneficial mutation with selective advantage  $s$  will reach fixation with probability approximately equal to  $2s$ , has been an important contribution to evolutionary theory. The derivation of this simple result was based on assuming a closed population with given constant size as well as Poisson distributed family size. However, determining probabilities of fixation is a lot more complicated under more realistic assumptions, such as in models including spatial variation (Whitlock, 2002; Cherry & Wakeley, 2003; Wakeley, 2003; Taylor, 2008) and fluctuations in population size (Ewens, 1967; Otto & Whitlock, 1997; Phillips, 1997; Pollak, 2000; Campbell, 2003; Lambert, 2006; Parsons & Quince, 2007). Here we analyse the effect of fluctuating population size described by a stationary stochastic process. Most real populations often show large fluctuations due to environmental effects on all individuals, and the fate of a new beneficial mutation

is mainly determined by the dynamics of the sub-population with the mutation during the initial several generations after it arises. The process describing the number of mutants is therefore complex, with stochasticity in the gene transmission as well as temporally correlated noise due to stochastic fluctuations in population size.

Otto & Whitlock (1997) derived the fixation probability for a beneficial mutation in a fluctuating population. Their derivation and results are, however, based on the assumption that the sequence of population sizes through time is known. Hence, although their theory is stochastic in the sense that it is based on the classical result of Haldane (1927) with stochastic variation in gene frequency (random genetic drift), the population dynamical part of their theory is deterministic. Here we extend this theory by incorporating stochastic population dynamics. We show that the expected fixation rate of beneficial mutations depends primarily on the environmental variance causing stochastic population fluctuations, rather than the harmonic mean population size through time. The classical formula of Wright (1931, 1969) for effective

\* Corresponding author. Tel: +47 73 591747. Fax: +47 73 591038. e-mail: steinaen@math.ntnu.no

population size based on harmonic mean population size through time is expected to underestimate the mean probability of fixation when temporal fluctuations are viewed as stochastic rather than deterministic.

Haldane (1927) assumed that family size is Poisson-distributed with mean 1 for all individuals except those bearing the mutation, for which the Poisson mean is  $1 + s$ . We employ a similar model in which the family size produced by an adult individual is defined as the number of offspring surviving to maturity, assuming that the number of surviving offspring per family is Poisson-distributed for each genotype. This model also would apply if all adults of a given genotype produce large, nearly equal numbers of offspring, which suffer a high non-selective mortality, leaving an approximately Poisson distribution of number of surviving offspring per family. Haldane further assumed a large population size,  $Ns \gg 1$  so that selection is more important than random drift at intermediate frequencies of the mutant allele. For a fluctuating population we accordingly assume that the mean population size is much larger than  $1/s$ . Otto & Whitlock (1997) made the same assumption of Poisson family size and derived the probability  $u_i$  that a mutation in generation  $i$  will be fixed for a given sequence of population sizes  $N_1, N_2, \dots$ . A single mutation in generation  $i$  will have a number of copies next generation that is Poisson-distributed with mean  $\lambda = (1 + s)N_{i+1}/N_i$ . Hence the total probability that the mutation will not be fixed is

$$1 - u_i = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} (1 - u_{i+1})^k = e^{-(s+1)(N_{i+1}/N_i)u_{i+1}} \quad (1)$$

(Ewens, 1967; Otto & Whitlock, 1997). Haldane's result appears by assuming all population sizes equal,  $N_i = N$ , and  $u_i = u$ , in which case the solution is approximately  $u = 2s$  for a small  $s$ . With changing population size eqn (1) can be solved iteratively for a given sequence  $N_i$ , but only as a function of the unknown initial probability  $u_1$ . For cyclic populations, however, the formula defines probabilities of fixation by the fact that  $u_i$  takes the initial value after one cycle. Hence, computing  $u_{i+c}$  as a function of  $u_1$ , where  $c$  is the cycle length, and solving numerically  $u_1 = u_{1+c}$  gives the complete solution. Also for populations approaching a limiting size, the probabilities can be found utilizing the probability at equilibrium, which is the same as for constant population size (Otto & Whitlock, 1997).

If a single mutation produces  $k$  copies in the next generation, the factor  $(1 - u_{i+1})^k$  in eqn (1) is based on the assumption that future contributions from these alleles are independent. This is a realistic assumption if the future population sizes are known. However, with stochastic population fluctuations this assumption of independence is no longer valid.

Environmental stochasticity driving the fluctuations in population size exerts the same impact on all individuals with the mutant gene, violating the assumption of independence among the branching processes for copies from the original mutation.

This paper focuses on deriving an accurate approximation to the probability of fixation for fluctuating populations described by a stationary Markov process. For such populations the probability of fixation of a beneficial mutation at population size  $N$  must depend on  $N$  only, because the Markov assumption ensures that all properties of the future population depend only on the initial population size. To derive these probabilities  $P(A|N) = u(x)$ , where  $A$  denotes the event of fixation and  $x = \ln(N/K)$  is the log of the initial population size scaled by the carrying capacity  $K$ , we first consider the probability of fixation conditioned on all future population sizes. However, our primary goal is not the computation of the probability  $u(x)$ , which is conditioned on the population size at mutation, but rather the expected probability and rate of fixation for mutations averaged over the stationary distribution of population sizes. The expected probability of fixation, denoted  $\bar{u}$ , will only depend on the population dynamical parameters and the selective advantage  $s$ . Using the theta-logistic model (Gilpin & Ayala, 1973), we explore how the environmental variance and the strength and form of density-regulation influence the expected probability and rate of fixation of beneficial mutations.

## 2. Model

### (i) Formulation and general solution

Writing  $\mathbf{N}_i = (N_i, N_{i+1}, \dots)$  the recurrence formula of Otto & Whitlock (1997) takes the form

$$1 - P(A|\mathbf{N}_i) = e^{-(s+1)(N_{i+1}/N_i)P(A|\mathbf{N}_{i+1})}. \quad (2)$$

This conditional model assumes known future population sizes. Taking the expectation on both sides with respect to the distribution of  $\mathbf{N}_{i+2}$  conditioned on  $N_{i+1}$  yields

$$1 - P(A|N_i, N_{i+1}) = E_{\mathbf{N}_{i+2}|N_{i+1}} e^{-(s+1)(N_{i+1}/N_i)P(A|N_{i+1}, \mathbf{N}_{i+2})}. \quad (3)$$

Conditioned on  $N_{i+1}$  the fixation probability  $P(A|N_{i+1}, \mathbf{N}_{i+2})$  is a stochastic variable because it is a function of the unknown sequence of future population sizes  $\mathbf{N}_{i+2}$ , and the right-hand side of eqn (2) is its moment generating function with argument  $-(s+1)(N_{i+1}/N_i)$ . Numerical calculations show that its variance is quite small and could be set to zero to give a rough first-order approximation. However, a better approximation can also easily be obtained by including the variance of  $P(A|N_{i+1}, \mathbf{N}_{i+2})$  conditioned

on  $N_{i+1}$ . This second-order approximation for eqn (2) would be the exact solution if  $P(A|N_{i+1}, \mathbf{N}_{i+2})$  was normally distributed and is generally an accurate approximation if the variance of this variable is small. Although  $P(A|N_{i+1}, \mathbf{N}_{i+2})$  depends on the unknown future expressed by  $\mathbf{N}_{i+2}$  it is likely to have a small variance (later exemplified by numerical solution in Fig. 2). Writing  $t = -(s+1)(N_{i+1}/N_i)$  and introducing the scaled log population size  $x_i = \ln(N_i/K)$ , we then have

$$1 - P(A|N_i, N_{i+1}) = e^{tu(x_{i+1}) + t^2 \sigma^2(x_{i+1})/2}, \quad (4)$$

where  $u(x_{i+1}) = EP(A|N_{i+1}, \mathbf{N}_{i+2}) = P(A|N_{i+1})$  and  $\sigma^2(x_{i+1}) = \text{var}[P(A|N_{i+1}, \mathbf{N}_{i+2})]$ . The expectation as well as the variance here refer to the distribution of  $\mathbf{N}_{i+2}$  conditioned on  $N_{i+1}$ .

One way of formulating the stationary time series  $N_i$  is to specify the distribution of the scaled population size in the next generation  $y = \ln(N_{i+1}/K)$  given the value  $x = \ln(N_i/K)$  the current generation, say  $f(y|x)$ . Then, multiplying eqn (4) by  $f(y|x)$  and integrating over  $y$  yields

$$1 - u(x) = \int e^{-(s+1)e^{y-x}u(y) + (s+1)^2 e^{2(y-x)}\sigma^2(y)/2} f(y|x) dy. \quad (5)$$

This equation for the unknown function  $u(x)$  can be used to find the first-order approximation, setting  $\sigma^2(x)$  to zero and adopting the fixed point method. We then substitute the  $k$ th approximation for  $u(x)$ , say  $u_k(x)$ , on the right side giving  $1 - u_{k+1}(x)$  on the left, starting with Haldane's (1927) result for a constant population  $u_0(x) = 2s$  and computing subsequently  $\mu_1(x), \mu_2(x), \dots$  approaching the solution of eqn (5). To find the more accurate second-order approximation we introduce the second moment of  $1 - P(A|N_{i+1}, \mathbf{N}_{i+2})$ ,  $v(x) = \sigma^2(x) + [1 - u(x)]^2$ . The same argument as the one used in the derivation of eqn (5), starting by squaring both sides of eqn (1), then yields

$$v(x) = \int e^{-2(s+1)e^{y-x}u(y) + 2(s+1)^2 e^{2(y-x)}\sigma^2(y)} f(y|x) dy. \quad (6)$$

Equations (5) and (6) can now be solved jointly for  $u(x)$  as well as  $\sigma^2(x)$  by calculating both integrals based on the  $k$ th approximation for  $u(x)$  and  $\sigma^2(x)$ , giving on the left-hand side the  $(k+1)$ th approximation for  $u(x)$  and  $v(x)$ , and hence also for  $\sigma^2(x)$ . The iterations can be initiated by choosing  $u_0(x) = 2s$  and  $\sigma_0^2(x) = 0$ .

(ii) *The theta-logistic model*

A population dynamical model must now be specified to derive  $f(y|x)$ . A flexible model describing a wide range of types of density regulation, magnitude of stochasticity and expected return times to equilibrium

is the theta-logistic model (Gilpin & Ayala, 1973; Diserud & Engen, 2000; Lande *et al.*, 2003). Writing  $\Delta N$  for the annual change in population size, the deterministic form of the model is

$$\Delta N = r_0 N \left[ 1 - \left( \frac{N}{K} \right)^\theta \right] = r_1 N \left[ 1 - \frac{N^\theta - 1}{K^\theta - 1} \right],$$

where  $r_0$  is the growth rate at  $N=0$ . However, the model is more flexible if we consider the lowest possible population size to be 1 writing the formula as the expression on the right-hand side, where  $r_1 = r_0(1 - K^{-\theta})$  (Diserud & Engen, 2000) is the growth rate at  $N=1$ . Choosing  $r_1$  as a free parameter and varying  $\theta$  then defines a number of well-known population models. For  $\theta=1$  we recover the classical logistic model, while the limit obtained as  $\theta$  approaches zero is a linear expression in  $\ln N$ , often called the Gompertz model. As  $\theta$  approaches infinity we arrive at a model with a ceiling at  $K$  and no density-regulation below this value. Finally, for  $\theta=-1$  the model is linear in the population size  $N$ . Alternatively, the model may be formulated on the log scale as

$$\Delta \ln N = r_1 \left[ 1 - \frac{N^\theta - 1}{K^\theta - 1} \right].$$

For small or moderate fluctuations in the population size, these models are approximately equivalent because  $\Delta \ln N \approx \Delta N/N$ . The stochastic formulation on the log scale takes the form

$$\Delta \ln N = r_1(t) - \bar{r}_1 \frac{N^\theta - 1}{K^\theta - 1},$$

where  $r_1(t)$  is the growth rate at time  $t$  fluctuating between years with mean  $\bar{r}_1$  (May, 1973, 1974; Turelli, 1977; Engen *et al.*, 1998; Lande *et al.*, 2003). For large populations the demographic stochastic component of  $r_1(t)$ , due to within year variation in vital rates between individuals, can be ignored by the law of large numbers and the only stochastic components are those generated by fluctuations in the environment producing approximately a constant environmental variance  $\sigma_e^2$  in  $r_1(t)$  between years. For this model  $\bar{r}_1$  is the mean growth rate on the log scale at small population sizes ( $N=1$ ) called the stochastic growth rate (Lande *et al.*, 2003). Again writing  $x$  and  $y$  for the value of  $\ln(N/K)$  in two sequential generations, we have  $\text{var}(y|x) \approx \sigma_e^2$ , and

$$E(y|x) = x + \gamma(1 - e^{\theta x})/\theta \quad (7)$$

for  $\theta \neq 0$  and  $(1-\gamma)x$  for  $\theta=0$ . The parameter  $\gamma$  is the strength of density regulation defined as  $\bar{r}_1\theta/(1 - K^{-\theta}) = \bar{r}_0\theta$  for  $\theta \neq 0$  and  $\bar{r}_1/\ln K$  for  $\theta=0$ . The expected return time to equilibrium defined by May (1974) for this model is  $T = 1/\gamma$ .

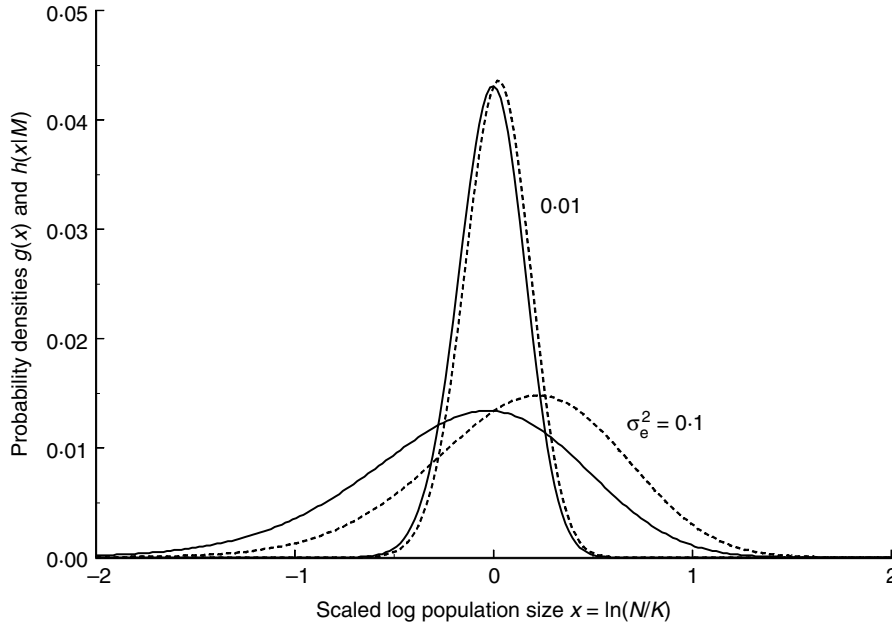


Fig. 1. Stationary distributions of the scaled log population size  $g(x)$  (solid line) and the scaled log population size at mutation  $h(x|M)$  (dashed line) for two values of the environmental variance  $\sigma_e^2$  in the theta-logistic model. The other parameters are  $\theta=1$  and  $\gamma=0.2$ .

Finally, the change in log population size between generations can often be approximated by a normal distribution. These assumptions then specify  $f(y|x)$  as the normal distribution with the mean given by eqn (7) and constant variance  $\sigma_e^2$ , which we apply in the recurrence relations for the fixed point iterations defined by eqns (5) and (6).

### (iii) Mean fixation rate

The rate of production of new mutations is proportional to the population size. Writing  $g(x)$  for the stationary distribution of the scaled log population size, the distribution of population size given that a mutation occurs ( $M$ ) must be proportional to  $e^x$ , that is,

$$h(x|M) = e^x g(x) / \int e^z g(z) dz. \quad (8)$$

Hence, for a random beneficial mutation with selective advantage  $s$  the probability of fixation is  $\bar{u} = \int u(x) h(x|M) dx$ .

Although the stationary distribution is known for the theta-logistic model based on the diffusion approximation (Diserud & Engen, 2000), more accurate values for the fixation probability for our model with discrete generations can be found numerically by using the fixed point method also for  $g(x)$ , calculating the  $(k+1)$ th approximation as  $g_{k+1}(x) = \int f(x|y) \times g_k(y) dy$ , which converges quickly. Figure 1 shows examples of  $g(x)$  as well as  $h(x|M)$ .

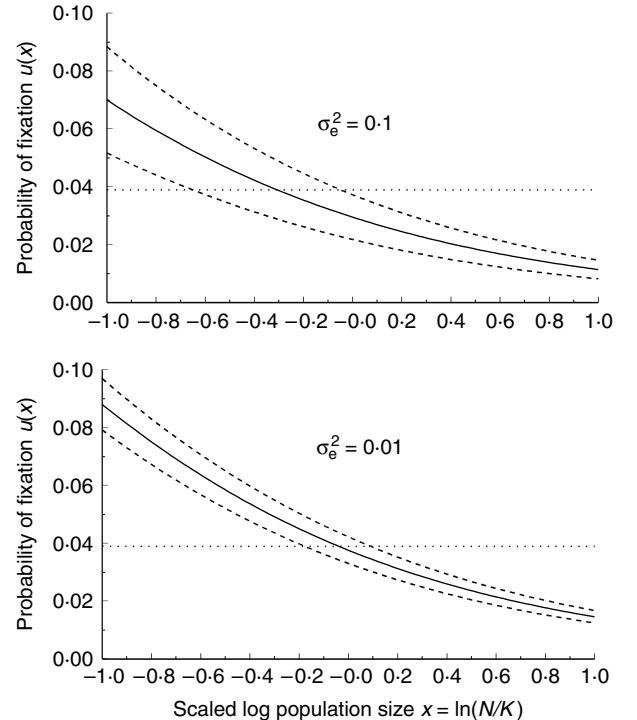


Fig. 2. The approximation to the unconditional probability of fixation (solid lines)  $u(x)$  for a mutation at population size  $N$ , where  $x = \ln(N/K)$ . Dashed lines show  $u(x)$  plus or minus one standard deviation  $\sigma(x)$ . The environmental variance is  $\sigma_e^2 = 0.1$  in the upper panel and  $0.01$  in the lower. Other parameters are  $s = 0.02$ ,  $\theta = 1$  and  $\gamma = 0.2$ . Dotted horizontal lines show the exact probability of fixation ( $0.03896 \approx 0.04 = 2s$ ) for a constant population size.

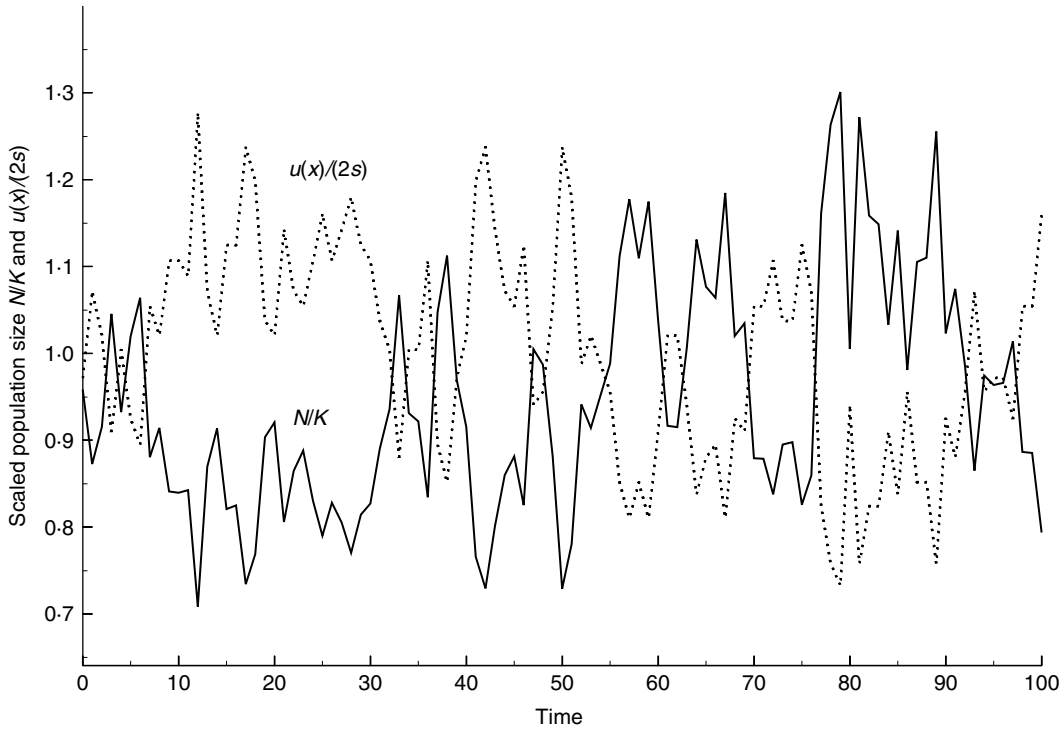


Fig. 3. Simulation of the theta-logistic process scaled by the carrying capacity (solid line) with parameters  $\sigma_c^2=0.01$ ,  $\gamma=0.2$ , and  $\theta=1$  (logistic model). Unconditional probabilities of fixation scaled by  $2s$ , that is,  $u(x)/(2s)$ , evaluated for the population size each generation (dashed line). The selection coefficient for the new beneficial mutation is  $s=0.02$ .

For the Gompertz model with  $\theta=0$ , the population process is a simple first-order autoregressive time series and the stationary distribution  $f(x)$  is normal with zero mean and variance  $v=\sigma_c^2/(2\gamma-\gamma^2)$ . The population size then has a lognormal stationary distribution with coefficient of variation  $CV=\sqrt{e^v-1}$ . When  $\theta\neq 0$  the mean population size in the next generation may be linearized at  $x=0$  to give the approximation  $E(y|x)\approx(1-\gamma)x$ . Accordingly, the above normal distribution can also be used as an approximation for the stationary distribution in non-linear models defined by  $\theta\neq 0$ .

#### (iv) Fixation rate at a given population size

Since  $u(x)$  is the probability of fixation for a single mutation at scaled population size  $x$ , the rate of fixation for mutations at population size  $N$  is  $R(x)=2\mu Nu(x)$ . Here  $2\mu$  is the genomic mutation rate for beneficial mutations, and we assume that the product of the per locus mutation rate and the population size is small enough to render unlikely any interference among beneficial mutations that could occur if multiple mutations simultaneously segregate at a single locus or among tightly linked loci. For a constant population size  $K$ , the fixation rate is approximately  $R_0=4sK\mu$ . Accordingly, for population size  $N=Ke^x$ , we define the scaled fixation rate  $R(x)/R_0=u(x)N/(2sK)$ , which measures the reduction in fixation rate due to

stochastic fluctuations in population size, relative to the fixation rate for a constant population of size  $N=K$ .

### 3. Results

The proposed fixed point method for the joint solution of  $u(x)$  and  $\sigma^2(x)$ , based on eqns (5) and (6) with initial conditions  $u_0(x)=2s$  and  $\sigma_0^2(x)=0$ , converges for relevant values of the parameters  $\gamma$ ,  $\theta$  and  $\sigma_c^2$ . Examples of the function  $u(x)$  plus or minus one standard deviation  $\sigma(x)$  are shown in Fig. 2. The results agree with Otto & Whitlock's (1997) findings in the sense that the probability of fixation is larger for populations that are expected to grow ( $x<0$ ) than for those expected to decrease ( $x>0$ ). For mutations at equilibrium the probability of fixation is usually a little smaller than the probability  $2s$  of fixation at constant population size.

In Fig. 3, we show a simulation of  $N/K$  for a stochastic logistic model together with the probabilities of fixation scaled by  $2s$ ,  $u(x)/(2s)$ . This graph is qualitatively similar to Otto & Whitlock's (1997) Fig. 4, showing the fluctuations of the cyclic snowshoe hare population over a period of 20 generations (Krebs *et al.*, 1995) together with probabilities of fixation fluctuating in an opposite way, assuming deterministic cyclic fluctuations. We see from Fig. 3 that the unconditional probability, that is, the probability

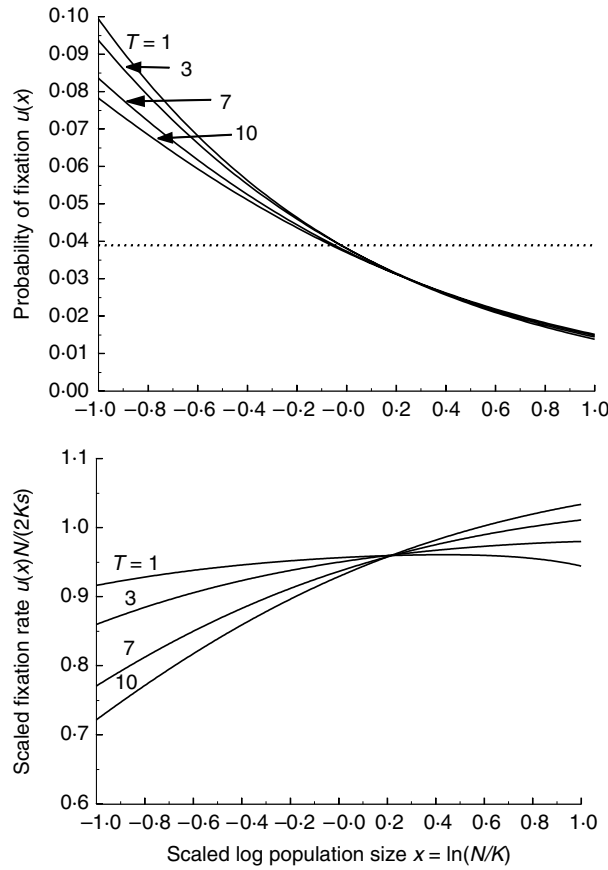


Fig. 4. The probability of fixation  $u(x)$  (upper panel) and the scaled probability of fixation  $u(x)N/(2Ks)$  (lower panel) as functions of the scaled log population size  $x = \ln(N/K)$  for different values of mean return time to equilibrium  $T = 1/\gamma$ . Other parameters are  $\sigma_e^2 = 0.01$ ,  $\theta = 1$  and  $s = 0.02$ .

of fixation only conditioned on the population size at mutation and averaged over the future stochastic population fluctuations, also fluctuates opposite to the scaled population size.

Figures 2, 4 (upper panel) and 5 (upper panel) show how the environmental variance (Fig. 2), the return time to equilibrium (Fig. 4) and the form of density-regulation defined by  $\theta$  (Fig. 5) affect  $u(x)$ . The overall message from these examples is that changing  $T = 1/\gamma$  and  $\theta$  has surprisingly small effects except at very small population sizes. Linearizing the theta-logistic model at equilibrium gives the first-order autoregressive model with stationary variance of  $\ln(N/K)$  equal to  $\sigma_e^2/(2\gamma - \gamma^2)$ . Hence, in Fig. 4 the stationary distribution of  $x = \ln(N/K)$  has standard deviations of approximately 0.1 for  $T = 1$  and 0.22 for  $T = 10$ , which implies that the probability of fixation is almost unaffected by  $T$  for this magnitude of variation in  $x$ . The same argument holds for the effect of changing  $\theta$ .

Figures 4–6 show examples of reduction in fixation rate measured by  $R(x)/R_0 = u(x)N/(2sK)$  as functions of the scaled log population size for different values of the environmental variance  $\sigma_e^2$  (Fig. 6), the return time to equilibrium  $T = 1/\gamma$  (Fig. 4, lower panel) and the

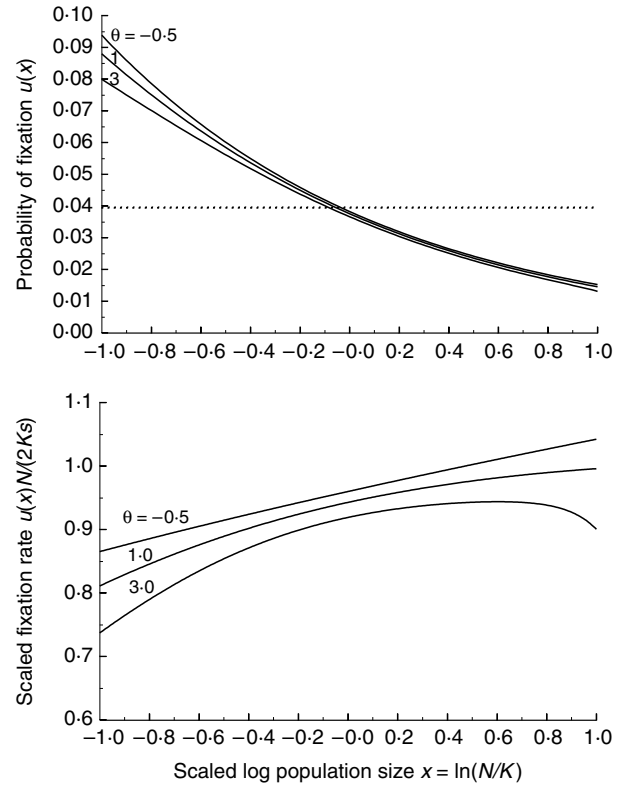


Fig. 5. The probability of fixation  $u(x)$  (upper panel) and the scaled probability of fixation  $u(x)N/(2Ks)$  as functions of the scaled log population size  $x = \ln(N/K)$  for different values of  $\theta$ . Other parameters are  $\sigma_e^2 = 0.01$ ,  $\gamma = 0.2$  and  $s = 0.02$ .

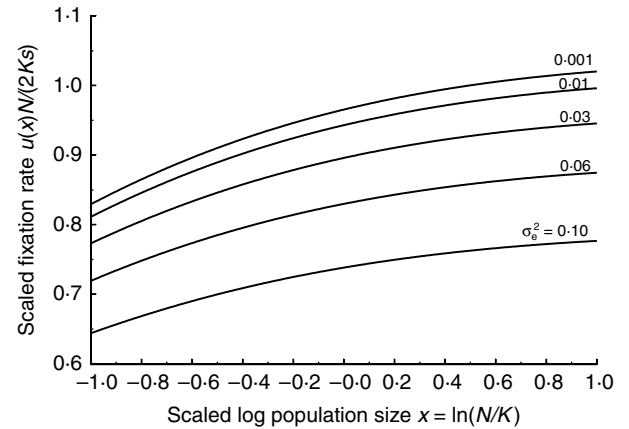


Fig. 6. The scaled fixation rate  $u(x)N/(2sK)$  as a function of the scaled log population size  $x$  for different values of the environmental variance  $\sigma_e^2$ . Other parameters are  $T = 5$  ( $\gamma = 0.2$ ),  $\theta = 1$  and  $s = 0.02$ .

shape of the density-regulation  $\theta$  (Fig. 5, lower panel), respectively. Increasing  $\sigma_e^2$  causes a substantial reduction in the scaled fixation rate for all population sizes, whereas the effects of  $T$  and  $\theta$  are quite small, especially for scaled population sizes  $x = \ln(N/K)$  in the range  $(-0.5, 0.5)$ , corresponding to a population size  $N$  in the range  $(0.61K, 1.65K)$ .

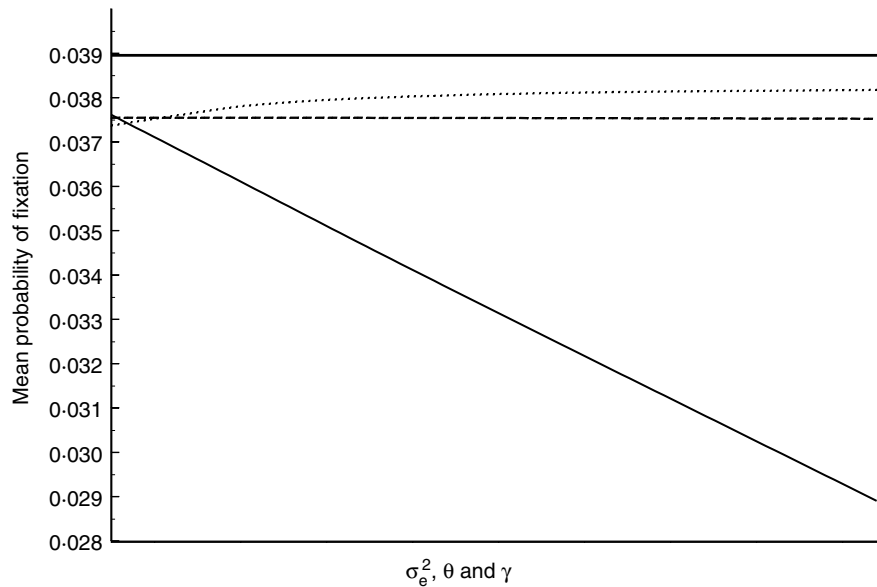


Fig. 7. The mean fixation probability  $\bar{u}$  for a random mutation as a function of one of the parameters  $\sigma_e^2$  (solid line),  $\theta$  (dashed line), and  $\gamma$  (dotted line), keeping the others constant. The baseline parameters are  $\sigma_e^2=0.01$ ,  $\theta=1$  and  $\gamma=0.2$ . Ranges of parameters shown in the graph are  $[0.001, 0.1]$  for  $\sigma_e^2$ ,  $[-0.5, 1.0]$  for  $\theta$  and  $[0.1, 1.0]$  for  $\gamma$ . The CV in population size ranges from 0 to approximately 0.8 in the graph, the thick solid horizontal line being the probability of fixation  $0.03896 \approx 2s = 0.04$  for a constant population size.

In Fig. 7, we show how the mean probability of fixation for a random beneficial mutation  $\bar{u}$  (averaged over the stationary first-moment distribution  $h(x|M)$ ) depends on  $\sigma_e^2$ ,  $\gamma$  and  $\theta$  for realistic ranges of these parameters. This confirms that the only substantial influence arises from the environmental variance, an increase of which reduces the probability of fixation. Interestingly, we see indirectly that the stationary variance of the population fluctuations, which is approximately  $v = \sigma_e^2 / (2\gamma - \gamma^2)$  for the scaled log population size, is not by itself important. Figure 7 shows practically no influence of  $\gamma$  even if a change in  $\gamma$  leads to a large change in  $v$ . On the other hand, an increase in the environmental variance substantially decreases the mean fixation probability. The practical conclusion from these findings is that short-term fluctuations in population size between generations produced by  $\sigma_e^2$  are much more important than the long-term stationary fluctuations in population size itself (Figs 6 and 7).

As a further illustration we consider the Gompertz model ( $\theta=0$ ) with an exactly lognormal stationary distribution. In Fig. 8 this distribution is kept constant while varying the strength of density regulation  $\gamma$  and the environmental variance  $\sigma_e^2$  so that  $v = \sigma_e^2 / (2\gamma - \gamma^2)$  of log population size and the CV  $= \sqrt{e^v - 1}$  of the population size  $N$  are kept constant. Figure 8 shows  $\bar{u}$  plotted against return time to equilibrium  $T = 1/\gamma$  (lower panel) as well as against  $\sigma_e^2$  (upper panel) for realistic parameter ranges. There is a decrease in  $\bar{u}$  up to about 25% relative to the maximum value for the chosen parameter intervals although the

stationary distribution remains unchanged. Figure 8 (upper panel) also shows the approximation for  $\bar{u}$  based on using the harmonic mean  $N_e$ , that is, the exact value of the probability of fixation for a constant population size multiplied by  $2N_e/\bar{N}$ , where  $\bar{N}$  is the mean population time through time. It appears that this approximation underestimates  $\bar{u}$  for large values of CV.

#### 4. Discussion

Ewens (1967) considered a model where the population size cycles deterministically and showed that such fluctuations decreased the probability of fixation for a beneficial mutant. This result was later confirmed by Otto & Whitlock (1997), who concluded that fixation is more likely if the mutation occurs during increase than during decrease in population size. This was supported by a logistic diffusion model of Lambert (2006) with only demographic stochasticity. He found that fixation probabilities were larger than Haldane's (1927) result  $2s$  for constant populations if mutations occurred below the carrying capacity and were smaller above. This result was confirmed by Parsons & Quince (2007) using a logistic type of birth and death process for the total population size. On the other hand, in a population with immediate return to equilibrium, Campbell (2003) found that fixation probabilities could be larger than for constant population size. Cyclic fluctuations were also dealt with by Pollak (2000), who found, in accordance with Otto & Whitlock (1997), that the probability of fixation was

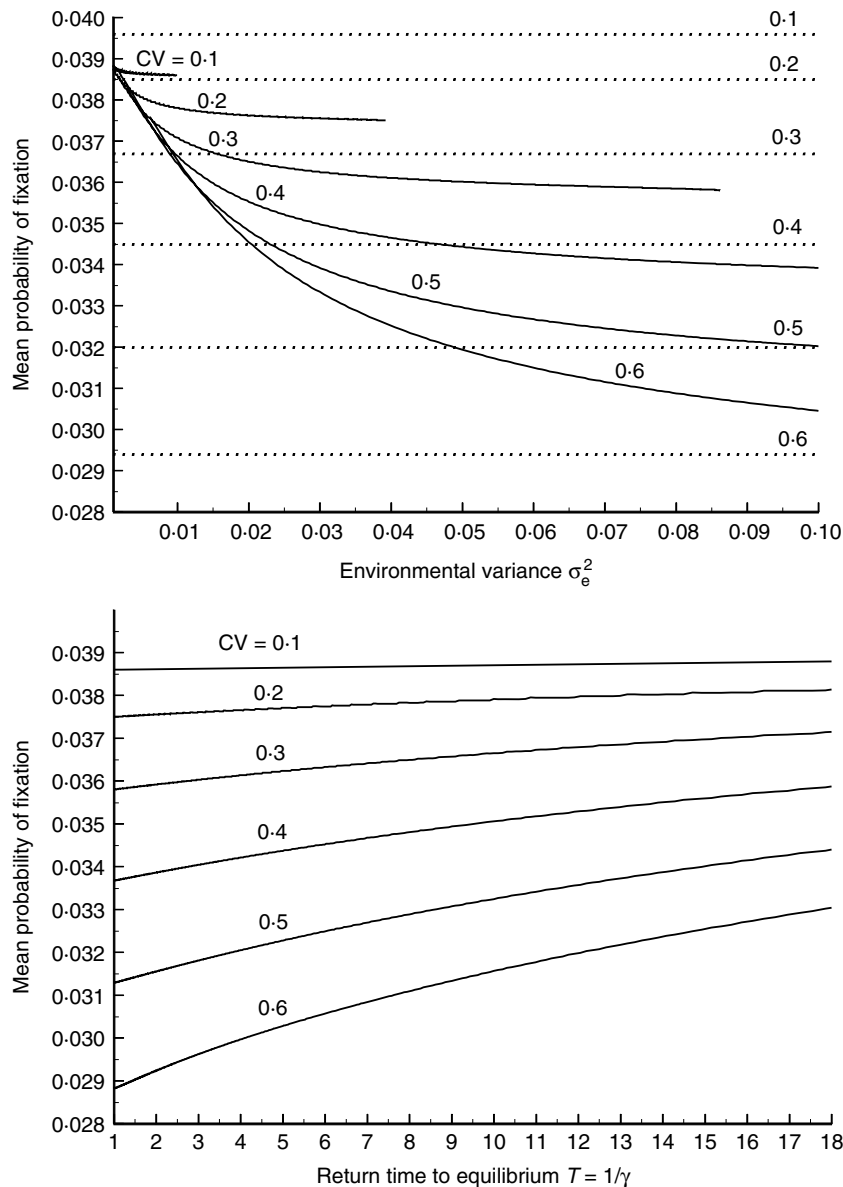


Fig. 8. The upper panel shows the mean fixation probability  $\bar{u}$  plotted against the environmental variance  $\sigma_e^2$  for the Gompertz model with  $\theta=0$  (solid lines). For each CV in population size  $T$  (and  $\gamma=1/T$ ) changes with  $\sigma_e^2$  so that  $v=\sigma_e^2/(2\gamma-\gamma^2)$  is kept constant, that is, the stationary distribution is the same along each line. Only values of  $\sigma_e^2$  corresponding to  $1 < T < 1000$  are used. Dotted lines show the corresponding values obtained by using the harmonic mean, that is, the exact value for a constant population, which is  $0.03896$  multiplied by  $N_e/\bar{N}$ , where  $N_e$  is the mean of  $1/N$  and  $\bar{N}$  is the mean of  $N$  in the stationary distribution. The lower panel shows the same mean fixation probability  $\bar{u}$  plotted against  $T=1/\gamma$ .

approximately proportional to the inverse of population size  $1/N$  at mutation, provided that  $s$  multiplied by the cycle length was small.

Many populations undergo fluctuations over long periods of time which often can be realistically described by a stationary stochastic process. Here, we have used the theta-logistic population model, assuming population sizes are large enough for demographic stochasticity to be ignored. This model is flexible, covering a wide range of possible types of population fluctuations, including large fluctuations of large populations generated by a fluctuating environment. The models of Lambert (2006) and Parsons

and Quince (2007), on the other hand, have only demographic stochasticity defined through the independent birth and death of individuals and will therefore have a CV in total population size that approaches zero as the carrying capacity increases. Fluctuating environments that affect all individuals in the population in the same way generate stochastic dependence between individuals that leads to a more complicated dynamics than under demographic stochasticity alone.

The accurate approximation to the probability of fixation derived here is the ‘unconditional’ probability, conditioned only on the population size at

mutation. This is rather different from the probability used by Ewens (1967) and Otto & Whitlock (1997), which is conditioned on all future population sizes. Nevertheless, the results are rather similar. We find a large reduction in fixation probability at population sizes above average at the time of mutation ( $x > 0$ ), as these are expected to decrease towards the mean size. Considering the stochastic variation in population size for different models, exemplified by a simulation in Fig. 3, we also see that the variation in population size is large enough for the fixation probabilities to change considerably through time. The shape of these curves for unconditional probabilities is approximately proportional to  $1/N$ , implying a nearly exponential function of  $x$ . This is typically the form of the curves shown in Fig. 2 and Figs 4–5 (upper panels) although the shape changes with the population dynamical parameters, especially the environmental variance  $\sigma_e^2$ . Deviation from proportionality to  $1/N$  is seen most clearly from Figs 6 and 8 showing the scaled fixation rates, as these should be constant functions of  $x$  if  $u(x)$  is proportional to  $1/N$  so that  $u(x)N$  is constant.

If all individuals are equally prone to mutate, it is realistic to assume that mutations are more likely to arise in large populations, or more precisely, the probability of a single mutation is proportional to population size. Hence, the population size at mutation does not have the stationary distribution of the population process, but rather the so-called first-moment distribution multiplied by  $N$  and then scaled. For the scaled log population size  $x$ , the distribution, correspondingly, has to be multiplied by  $e^x$ . Integrating the fixation probabilities  $u(x)$  over this distribution leads to the mean fixation probability for a random mutation  $\bar{u}$  with advantage  $s$ . In this model the mean fixation probability  $\bar{u}$  is rather insensitive to substantial changes in the form of the density-regulation and the return time to equilibrium (Fig. 7). The most important parameter is the environmental variance  $\sigma_e^2$  (Fig. 7) rather than the variance of the stationary distribution of population size itself. This occurs because the number of descendants of a beneficial mutation during the first few generations after it originates is crucial for its fate, as pointed out by Crow & Kimura (1970) and Phillips (1997). Short-term variation in population size is therefore of great importance, and is mainly determined by the environmental variance.

These conclusions are further illustrated in Fig. 7. The environmental variance is of the order of 0.01 for most populations, corresponding to a standard deviation of  $\sqrt{0.01} = 0.1$  for fluctuations in the population growth rate between generations (Lande *et al.*, 2003). With this environmental variance we see that large changes in CV caused by differences in  $\gamma$  produce quite small changes in  $\bar{u}$ , while changes in CV

produced by differences in  $\sigma_e^2$  create much larger changes in  $\bar{u}$ .

It is commonly argued that long-term genetic drift and fixation probabilities are determined by the effective population size  $N_e$  approximated by the harmonic mean population size  $N$  through time (Wright 1931, 1969; Waples, 2006). Crow & Kimura (1970) claimed that the mean probability of fixation is approximately  $2s(N_e/\bar{N})$ . As increasing the amplitude of population fluctuations tends to produce a smaller harmonic mean, the mean fixation probability of a beneficial mutation should then decrease with increasing population fluctuations. We see from Fig. 8 that this is correct, but  $\bar{u}$  may still be sensitive to changes in environmental variance  $\sigma_e^2$  when  $N_e/\bar{N}$  is kept constant. This occurs because the fixation of a new mutant depends heavily on its fate during the first several generations after its origination. Thus mean fixation probabilities may depend rather strongly on  $\sigma_e^2$  even if the stationary distribution of population size and the harmonic mean are kept constant.

This study was generously supported by grants from the Research Council of Norway (STORFORSK), the Royal Society of London and cord funding to the Centre for Conservation Biology from the Norwegian University of Science and Technology.

## References

- Campbell, R. B. (2003). A logistic branching process for population genetics. *Journal of Theoretical Biology* **225**, 195–293.
- Cherry, J. L. & Wakelaj, J. (2003). A diffusion approximation for selection and drift in a subdivided population. *Genetics* **163**, 421–428.
- Crow, J. F. & Kimura, M. (1970). *An Introduction to Population Genetics Theory*. Minneapolis, MN: Burgess Publishing Company/Alpha Editions.
- Diserud, O. H. & Engen, S. (2000). A general and dynamic species abundance model, embracing the lognormal and the gamma models. *The American Naturalist* **155**, 497–511.
- Engen, S., Bakke, Ø. & Islam, A. (1998). Demographic and environmental stochasticity – concepts and definitions. *Biometrics* **54**, 39–45.
- Ewens, W. J. (1967). The probability of survival of a new mutant in a fluctuating environment. *Heredity* **22**, 438–443.
- Gilpin, M. E. & Ayala, F. J. (1973). Global models for growth and competition. *Proceedings of the National Academy of Sciences, USA* **70**, 3590–3593.
- Haldane, J. B. S. (1927). A mathematical theory of natural and artificial selection. V. Selection and mutation. *Proceedings of the Cambridge Philosophical Society* **23**, 838–844.
- Krebs, C., Boutin, S., Boonstra, R., Sinclair, A.-R. E. & Smith, J. N. M. (1995). Impact of food and predation on the snowshoe hare cycle. *Science* **269**, 1112–1115.
- Lambert, A. (2006). Probability of fixation under weak selection: a branching process unifying approach. *Theoretical Population Biology* **69**, 419–441.

- Lande, R., Engen, S. & Sæther, B.-E. (2003). *Stochastic Population Dynamics in Ecology and Conservation*. Oxford: Oxford University Press.
- May, R. M. (1973). Stability in randomly fluctuating versus deterministic environments. *The American Naturalist* **107**, 621–650.
- May, R. M. (1974). *Stability and Complexity in Model Ecosystems*, 2nd edn. Princeton, NJ: Princeton University Press.
- Otto, S. P. & Whitlock, M. C. (1997). The probability of fixation in populations of changing size. *Genetics* **146**, 723–733.
- Parsons, T. L. & Quince, C. (2007). Fixation in haploid populations exhibiting density-dependence. I: The non-neutral case. *Theoretical Population Biology* **72**, 121–135.
- Phillips, P. C. (1997). The rise and fall of new mutations. *Trends in Ecology and Evolution* **12**, 466–468.
- Pollak, E. (2000). Fixation probabilities when the population size undergoes cyclic fluctuations. *Theoretical Population Biology* **57**, 51–58.
- Taylor, J. E. (2008). Environmental variation, fluctuating selection and genetic drift in subdivided populations. *Theoretical Population Biology* **74**, 233–250.
- Turelli, M. (1977). Random environments and stochastic calculus. *Theoretical Population Biology* **12**, 140–178.
- Wakeley, J. (2003). Polymorphism and divergence for island-model species. *Genetics* **163**, 411–420.
- Waples, R. S. (2006). Seed banks, salmon, and sleeping genes: effective population size in semelparous, age-structured species with fluctuating abundance. *The American Naturalist* **167**, 118–135.
- Whitlock, M. C. (2002). Selection, load and inbreeding depression in a large metapopulation. *Genetics* **160**, 1191–1202.
- Wright, S. (1931). Evolution in Mendelian populations. *Genetics* **16**, 98–160.
- Wright, S. (1969). *Evolution and the Genetics of Populations. Vol. 2. The Theory of Gene Frequencies*. Chicago, IL: University of Chicago Press.