

# Selection in a Subdivided Population With Dominance or Local Frequency Dependence

Joshua L. Cherry<sup>1</sup>

*Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138*

Manuscript received October 21, 2002

Accepted for publication January 2, 2003

## ABSTRACT

The interplay between population structure and natural selection is an area of great interest. It is known that certain types of population subdivision do not alter fixation probabilities of selected alleles under genic, frequency-independent selection. In the presence of dominance for fitness or frequency-dependent selection these same types of subdivision can have large effects on fixation probabilities. For example, the barrier to fixation of a fitter allele due to underdominance is reduced by subdivision. Analytic results presented here relate a subdivided population that conforms to a finite island model to an approximately equivalent panmictic population. The size of this equivalent population is different from (larger than) the actual size of the subdivided population. Selection parameters are also different in the hypothetical equivalent population. As expected, the degree of dominance is lower in the equivalent population. The results are not limited to dominance but cover any form of polynomial frequency dependence.

NATURAL populations are likely to be characterized by some kind of population structure. The population-genetic and evolutionary consequences of such structure have been investigated since the beginnings of population genetics (WRIGHT 1931, 1939, 1943). Much work has centered on the amount of polymorphism maintained in a subdivided population (SLATKIN 1977; MARUYAMA and KIMURA 1980; NAGYLAKI 1998) and on the distribution of allele frequencies (MARUYAMA, 1972a,b,c). A closely related topic is the effective size of a subdivided population (WRIGHT 1939; MARUYAMA 1970a; SLATKIN 1981, 1991; TAKAHATA 1991; NEI and TAKAHATA 1993; SANTIAGO and CABALLERO 1995; WHITLOCK and BARTON 1997; WANG and CABALLERO 1999).

One area of interest is the interaction of population structure with selection. MARUYAMA (1970b, 1974) has shown that fixation probabilities are unaffected by population subdivision under simple genic selection and fairly general conditions of migration patterns. Under a finite island model of subdivision with a large number of demes, the trajectory of allele frequency over time is approximately the same as that in a panmictic population with a different size and different selection coefficient (CHERRY and WAKELEY 2003).

If the assumption of genic selection is relaxed, the problem becomes more difficult. Subdivision affects fixation probabilities when there is dominance for fitness or when the relative fitnesses of genotypes depend on their frequencies. An area of particular interest has

been the case of underdominance (heterozygote disadvantage), which can serve as a barrier to the fixation of a fitter genotype. Subdivision reduces this barrier, and this effect has been interpreted as a simple case of Wright's "shifting balance" theory (LANDE 1985), a theory that has been a topic of debate for many decades. Formally equivalent to dominance is a form of locally frequency-dependent selection. For example, selection that favors an allele when it is locally common, but disfavors it when it is rare, is similar to underdominance, and underdominance may be considered a case of positive frequency dependence. Frequency dependence may have any form; it need not be restricted to the linear case that is formally equivalent to dominance.

Analytic results, including expressions for fixation probabilities, have been obtained for selection with dominance in the low-migration limit (SLATKIN 1981; LANDE 1985). Simulations have provided results for intermediate cases, where the migration rate is neither very low nor sufficiently high to make subdivision irrelevant (SLATKIN 1981; SPIRITO *et al.* 1993). Here I present analytic results that are not restricted to the weak-migration limit for a finite island model of subdivision. These results relate the subdivided population to a hypothetical equivalent panmictic population that differs from the actual population in both its size and its selection parameters. The selection parameter  $h$ , which is a measure of dominance for fitness (or degree of frequency dependence), is in effect moved toward  $\frac{1}{2}$  (additive fitness or no frequency dependence) by subdivision. This result can be generalized to more complicated forms of frequency dependence. The existence of an equivalent panmictic population allows application of established diffusion results to the subdivided popula-

<sup>1</sup>Address for correspondence: 2307 Massachusetts Ave., Cambridge, MA 02140. E-mail: cherry@oeb.harvard.edu

tion. Such quantities of fixation probabilities and expected times to fixation can be calculated.

MODELS AND RESULTS

Consider a finite island model of population structure, in which a finite number of demes (“islands”) exchange migrants with one another. The population consists of  $D$  demes, each containing  $N$  haploid or  $N/2$  diploid individuals. Each generation involves migration, selection, and genetic drift. The order in which these occur makes little difference when selection coefficients and migration rates are small compared to unity, so this order need not be specified. The migration rate (the expected fraction of genes that come from outside the deme in any generation) is given by  $m$ , and selection operates in a manner to be specified below.

Three processes, selection, drift, and migration, alter the frequency of an allele in a deme each generation. Migration is symmetric in the island model, so it does not affect the overall allele frequency  $\bar{x}$  (it does not affect the mean, and its effect on the variance is negligible). The effects of selection and drift are approximately additive. The mean change in  $\bar{x}$  results from selection, whereas the variance is the result of genetic drift.

**Dominance:** CHERRY and WAKELEY (2003) analyzed the case of genic selection in an island model of subdivision. This analysis can be extended to cases where there is dominance for fitness, including over- and underdominance. These cases are equivalent, in terms of their usual diffusion approximations, to a certain form of frequency-dependent selection. Diffusion results for dominance in a panmictic population are well established. I show that, under certain conditions, the diffusion for a subdivided population is equivalent to that for some panmictic population. As in the case of genic selection, this equivalent panmictic population differs from the subdivided population not only in size, but also in fitness parameters.

Suppose that the fitnesses of genotypes  $aa$ ,  $Aa$ , and  $AA$  are  $1$ ,  $1 + 2hs$ , and  $1 + 2s$ . The fitness difference between the two alleles is  $2hs$  when paired with an  $a$  allele and  $2s - 2hs$  when paired with an  $A$  allele. Thus the mean selective difference between the two alleles  $\hat{s}$ , which might be called the marginal selection coefficient (by analogy to the marginal fitness), depends on the allele frequency and is given by

$$\hat{s}(x) = 2hs(1 - x) + (2s - 2hs)x \tag{1}$$

in an unstructured population with nonassortative mating. This can be rewritten as  $\hat{s}(x) = 2hs + (2s - 4hs)x$ , which makes it clearer that  $\hat{s}(x)$  has the form  $k_0 + k_1x$ , with  $k_0 = 2hs$  and  $k_1 = 2s - 4hs$ .

Now consider a subdivided population. We are interested in the mean and variance of the change in overall allele frequency from one generation to the next,  $M_{\Delta\bar{x}}$  and  $V_{\Delta\bar{x}}$ . Expressions for these as functions of  $\bar{x}$  will allow use of diffusion approximations. To obtain these we

need to consider the probability distribution of the allele frequency in the  $i$ th deme,  $x_i$ . If drift within a deme is strong compared to selection, the population as a whole in effect serves in the short term as a source population for any subpopulation, with constant allele frequency  $\bar{x}$ . Under these conditions the distribution of within-deme allele frequency is a beta distribution whose probability density function is

$$\frac{1}{B(a, b)} x_i^{a-1} (1 - x_i)^{b-1},$$

where  $a = 2Nm\bar{x}$ ,  $b = 2Nm(1 - \bar{x})$ , and the beta function  $B$  is defined by  $B(a, b) = \int_0^1 x^{a-1} (1 - x)^{b-1} dx$  (WRIGHT 1931; DOBZHANSKY and WRIGHT 1941). The moments of this distribution follow from the recursion properties of the beta function, namely that  $B(a + 1, b) = (a / (a + b)) B(a, b)$ . The first, second, and third moments of this distribution are  $a / (a + b)$ ,  $(a + 1)a / (a + b + 1)(a + b)$ , and  $(a + 2)(a + 1)a / (a + b + 2)(a + b + 1)(a + b)$ , respectively. These are all that we need to know to treat the case of dominance or linear frequency dependence.

The beta-distribution approximation is valid when selection is weak compared to drift in a subpopulation, in the sense that  $|\hat{s}| \ll 1/N$ . This condition must hold for all allele frequencies. In the present case,  $\hat{s}(x_i) = 2hs(1 - x_i) + (2s - 2hs)x_i$  takes on its most extreme values at allele frequencies of zero or one. Thus the constraint on the strength of selection becomes

$$|2Nhs| \ll 1 \tag{2}$$

and

$$|2N(1 - h)s| \ll 1. \tag{3}$$

This condition allows selection to be strong compared to drift in the population as a whole and hence allows selection to have a large effect on the fate of an allele.

The stochastic change in allele frequency in a deme comes from the binomial sampling of alleles. Thus the variance of the change in allele frequency in the  $i$ th deme is  $\sim (1/N)x_i(1 - x_i)$ . Using expressions for the first two moments of the beta distribution we can show that the variance of the change in population-wide frequency  $\bar{x}$  is given by

$$\begin{aligned} V_{\Delta\bar{x}} &= \frac{1}{D^2} \sum_i (1/N)x_i(1 - x_i) \approx \frac{1}{ND} E x_i(1 - x_i) \\ &= \frac{1}{ND} \frac{2Nm}{2Nm + 1}. \end{aligned} \tag{4}$$

The variance definition of effective population size  $N_e$  is given by  $V_{\Delta\bar{x}} = (1/N_e)\bar{x}(1 - \bar{x})$ . Thus we have

$$N_e = DN / \left( \frac{2Nm}{2Nm + 1} \right) = \left( 1 + \frac{1}{2Nm} \right) DN.$$

This is a well-known expression for the effective size of a subdivided population conforming to an island model

(WRIGHT 1943; NEI and TAKAHATA 1993). It is identical to the effective size obtained in the analogous context for the case of genic selection (CHERRY and WAKELEY 2003).

The mean change, on the other hand, is affected by the more complex selection scheme. For the  $i$ th deme, the change in allele frequency  $x_i$  due to selection is approximately  $\hat{s}(x_i)x_i(1 - x_i) = (k_0 + k_1x_i)x_i(1 - x_i)$ . We are interested in the mean of this quantity. Using the approximation that the  $x_i$  are beta distributed with  $a = 2Nm\bar{x}$  and  $b = 2Nm(1 - \bar{x})$ , along with expressions for the first three moments of a beta distribution, we can show that the mean change in  $\bar{x}$  is given by

$$M_{\Delta\bar{x}} = \left[ \left( k_0 + \frac{k_1/2}{Nm + 1} \right) + \left( \frac{Nm}{Nm + 1} \right) k_1\bar{x} \right] \left( \frac{2Nm}{2Nm + 1} \right) \bar{x}(1 - \bar{x}). \tag{5}$$

This expression has the form  $(k_{0e} + k_{1e}\bar{x})\bar{x}(1 - \bar{x})$ . Together with the form of  $V_{\Delta\bar{x}}$  (Equation 4), this shows that the diffusion for the subdivided population is equivalent to that for a panmictic population with altered parameters. The constants  $k_{0e}$  and  $k_{1e}$  are to be interpreted as the effective values of  $k_0$  and  $k_1$ , *i.e.*, the values that, in a hypothetical Wright-Fisher population with size  $N_e$ , yield roughly the same trajectory of allele frequency as does the actual subdivided population. From Equation 5 it follows that

$$\begin{aligned} k_{0e} &= \left( \frac{2Nm}{2Nm + 1} \right) \left( k_0 + \frac{k_1/2}{Nm + 1} \right) \\ k_{1e} &= \left( \frac{2Nm}{2Nm + 1} \right) \left( \frac{Nm}{Nm + 1} \right) k_1. \end{aligned} \tag{6}$$

Converting back to notation involving  $h$  and  $s$  and their analogs in the equivalent panmictic population  $h_e$  and  $s_e$ , and recalling our expression for  $N_e$ , we obtain

$$\begin{aligned} N_e &= \left( 1 + \frac{1}{2Nm} \right) DN \\ s_e &= \left( \frac{2Nm}{2Nm + 1} \right) s \\ h_e &= h - \frac{h - 1/2}{Nm + 1} = \frac{1}{2} + \left( \frac{Nm}{Nm + 1} \right) \left( h - \frac{1}{2} \right). \end{aligned} \tag{7}$$

As in the case of genic selection, subdivision increases  $N_e$  by a certain factor and decreases  $s_e$  by this same factor, so that  $N_e s_e$  is unaffected by subdivision. The effect of subdivision on  $h_e$  is to move it toward  $1/2$  and therefore to decrease the effective strength of dominance or frequency dependence, as expected from previous work (SLATKIN 1981; LANDE 1985; SPIRITO *et al.* 1993). This is made clear by the relationship

$$h_e - \frac{1}{2} = \left( \frac{Nm}{Nm + 1} \right) \left( h - \frac{1}{2} \right).$$

**Alternative parameterizations:** There are several natural parameterizations of the selection model used above.

The parameterization involving  $s$  and  $h$  cannot be applied to symmetric over- or underdominance:  $s$  would have to be zero for symmetry, while  $hs$  would have to be nonzero for over- or underdominance. One parameterization that can represent the symmetric cases is that involving  $k_0$  and  $k_1$ . A more commonly used notation for dominance gives the fitnesses of  $aa$ ,  $Aa$ , and  $AA$  as  $1$ ,  $1 + s_1$ , and  $1 + s_2$ . Let  $s_{1e}$  and  $s_{2e}$  be the effective values of  $s_1$  and  $s_2$ . Using the fact that  $s_1 = 2hs$  and  $s_2 = 2s$ , we obtain

$$\begin{aligned} s_{1e} &= \left( \frac{2Nm}{2Nm + 1} \right) \left[ \left( \frac{Nm}{Nm + 1} \right) s_1 + \frac{1}{2} \left( \frac{1}{Nm + 1} \right) s_2 \right] \\ s_{2e} &= \left( \frac{2Nm}{2Nm + 1} \right) s_2. \end{aligned}$$

Note that  $s_{1e}$  depends on  $s_2$  as well as  $s_1$ .

In the symmetric case  $s_2 = 0$ . If we let  $s' = s_1$  then both homozygotes have fitness 1 and the heterozygote has fitness  $1 + s'$ . If  $s'_e$  is the effective value of  $s'$  under subdivision, we have

$$s'_e = \left( \frac{2Nm}{2Nm + 1} \right) \left( \frac{Nm}{Nm + 1} \right) s'.$$

**More complicated forms of selection:** The analysis given above applies when the frequency dependence of the selective difference between the alleles has the form  $\hat{s}(x) = k_0 + k_1x$ . This case includes an arbitrary degree of dominance when the relative fitnesses of the diploid genotypes do not depend on their frequencies. This is not the only possible form of frequency dependence;  $\hat{s}(x)$  can in principle have any form whatsoever. Can we obtain diffusion approximations for more complicated forms of  $\hat{s}(x)$  when the population is subdivided, again assuming that  $|N\hat{s}(x)|$  is always small compared to 1?

The results obtained above are consequences of the forms of the first three moments of the beta distribution. The useful properties of these moments extend to the higher-order moments. These properties permit the analysis of cases where the frequency dependence is described by any polynomial, *i.e.*, where  $\hat{s}(x) = P(x) = k_0 + k_1x + k_2x^2 + \dots + k_nx^n$ . We see that the diffusion for a subdivided population with this form of selection is equivalent to that for a panmictic population with a different size and with  $\hat{s}_e(x) = P_e(x) = k_{0e} + k_{1e}x + k_{2e}x^2 + \dots + k_{ne}x^n$ . Simple dominance is a special case of this, with the degree of the polynomials  $P$  and  $P_e$  equal to 1.

Recall that we are interested in the expected value of  $\hat{s}(x)x(1 - x)$  when  $x$  has a beta distribution with parameters  $a = 2Nm\bar{x}$  and  $b = 2Nm(1 - \bar{x})$ . For polynomial  $\hat{s}(x)$ ,  $\hat{s}(x)x(1 - x)$  is the sum of terms of the form  $k_i x^{i+1}(1 - x)$ . For any beta distribution

$$E[x^{i+1}(1 - x)] = \frac{(a + i)(a + i - 1) \dots (a + 1)ab}{(a + b + i + 1)(a + b + i) \dots (a + b + 1)(a + b)}.$$

This follows from the recursion properties of the beta

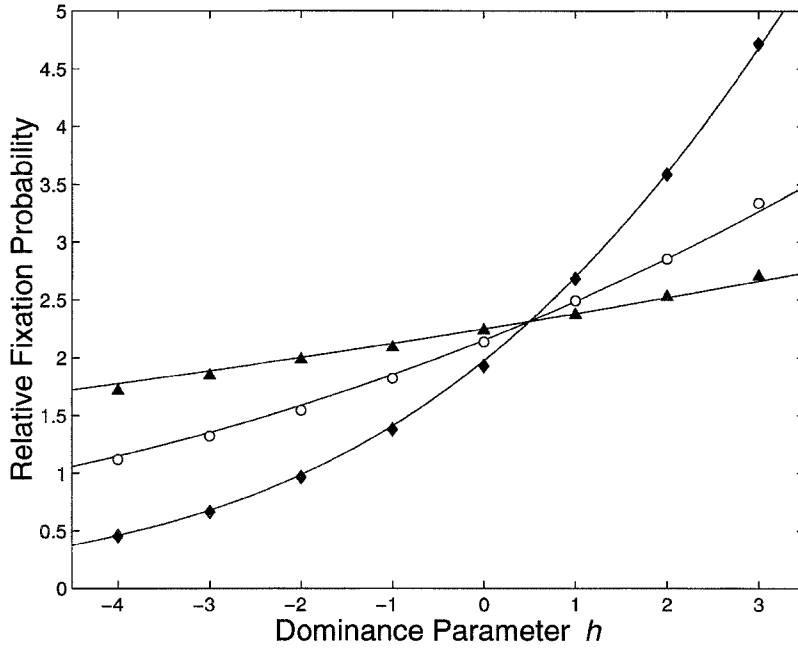


FIGURE 1.—Predicted and observed fixation probabilities as functions of  $h$  and  $m$ . Predicted values of fixation probabilities (curves), relative to that for a neutral allele, are compared to values estimated by simulations (points) for a range of values of the dominance parameter  $h$ . In all cases the number of demes ( $D$ ) is 100, the deme size ( $N$ ) is 100, and  $s = 10^{-4}$ . The allele was initially present in a single copy. Results are presented for three migration rates:  $m = 0.001$  (solid triangles and curve),  $m = 0.003$  (open circles and curve), and  $m = 0.01$  (solid diamonds and curve).

function mentioned earlier. In the present case,  $a + b = 2Nm$ , so the denominator is independent of  $\bar{x}$ . The numerator expands to an expression of the form  $Q(\bar{x})\bar{x}(1 - \bar{x})$ , where  $Q(\bar{x})$  is a polynomial of degree  $i$  in  $\bar{x}$ . It follows that  $E\{P(x)x(1 - x)\}$  has the form  $P_c(\bar{x})\bar{x}(1 - \bar{x})$  for some polynomial  $P_c$  of the same degree as  $P$ . The coefficients of this polynomial ( $k_{0c}, k_{1c}, k_{2c}, \dots, k_{nc}$ ), which depend on the coefficients of  $P$  and on  $N$  and  $m$ , may be obtained by some tedious but straightforward algebra. Consider, for example, the case of quadratic frequency dependence, where  $\hat{s}(x) = P(x) = k_0 + k_1x + k_2x^2$ . We have

$$\begin{aligned} E\{x^3(1 - x)\} &= \frac{(a + 2)(a + 1)ab}{(a + b + 3)(a + b + 2)(a + b + 1)(a + b)} \\ &= \frac{(2Nm\bar{x} + 2)(2Nm\bar{x} + 1)2Nm\bar{x}2Nm(1 - \bar{x})}{(2Nm + 3)(2Nm + 2)(2Nm + 1)2Nm} \\ &= \left(\frac{2Nm}{2Nm + 1}\right) \frac{(2Nm)^2\bar{x}^2 + 6Nm\bar{x} + 2}{(2Nm + 3)(2Nm + 2)(2Nm + 1)} \bar{x}(1 - \bar{x}). \end{aligned}$$

Thus the quadratic term of  $P$  contributes quadratic, linear, and constant terms to  $P_c$ , so the value of  $k_2$  affects not only the value of  $k_{2c}$ , but also the values of  $k_{1c}$  and  $k_{0c}$ . Combining this result with those given for first-degree frequency dependence we obtain

$$\begin{aligned} k_{0c} &= \left(\frac{2Nm}{2Nm + 1}\right) \left(k_0 + \frac{k_1/2}{Nm + 1} + \frac{k_2}{(2Nm + 3)(Nm + 1)}\right) \\ k_{1c} &= \left(\frac{2Nm}{2Nm + 1}\right) \left(\frac{Nm}{Nm + 1}\right) \left(k_1 + \frac{3Nm}{2Nm + 3}k_2\right) \\ k_{2c} &= \left(\frac{2Nm}{2Nm + 3}\right) \left(\frac{Nm}{Nm + 1}\right) \left(\frac{2Nm}{2Nm + 1}\right) k_2 \end{aligned} \tag{8}$$

(compare with Equation 6). The same principle could

be applied to a polynomial  $\hat{s}(x)$  of any degree or to a Taylor series of an analytic function.

### COMPUTER SIMULATIONS

To test the approximations used above, I have run computer simulations and compared the results to theoretical predictions. In these simulations frequency-dependent selection acts on a haploid population. Most of the simulations are for cases of linear frequency dependence and can also be interpreted in terms of diploidy with dominance. I utilize the parameterization involving  $s$  and  $h$  for these cases.

In these simulations the state of the population is represented by an array of  $D$  integers, each corresponding to a deme. Each integer indicates the number of copies of allele  $A$  in the deme and hence ranges from 0 to  $N$ . Each generation the new value for each deme is drawn from a binomial distribution. The index parameter  $n$  of this binomial (number of “trials”) is equal to  $N$ . The probability parameter  $p$  (probability of “success”) is determined by the current allele frequency in the deme  $x_i$ , the population-wide mean allele frequency  $\bar{x}$ , the migration rate  $m$ , and the selection parameters  $s$  and  $h$ . Let  $\hat{p} = (1 - m)x_i + m\bar{x}$ . This would be the expected allele frequency in the  $i$ th deme in the next generation if there were no selection. The (marginal) selection coefficient is given by  $\hat{s}(\hat{p})$ . Therefore  $p = (1 + \hat{s})\hat{p}/(1 + \hat{s}\hat{p})$ .

Predictions of fixation probabilities follow from combination of the theory presented here with classical diffusion results. KIMURA (1957, Equation 5.4) gave an expression for fixation probability in a panmictic Wright-Fisher population with an arbitrary form of fre-

TABLE 1

Predicted and observed fixation probabilities for  $N = D = 100$  and  $s = 3 \times 10^{-4}$  or  $s = 10^{-3}$

$s$	$m$	$h$	Relative fixation probability	
			Predicted	Observed
$3 \times 10^{-4}$	0.001	-4	3.24	$3.03 \pm 0.03$
		-3	3.76	$3.50 \pm 0.03$
		-2	4.34	$4.13 \pm 0.03$
		-1	4.97	$4.76 \pm 0.05$
		0	5.65	$5.59 \pm 0.07$
		1	6.39	$6.26 \pm 0.08$
	0.01	2	7.17	$7.34 \pm 0.09$
		3	7.99	$8.30 \pm 0.09$
		-4	0.05	$0.05 \pm 0.001$
		-3	0.18	$0.18 \pm 0.003$
		-2	0.60	$0.59 \pm 0.01$
		-1	1.73	$1.69 \pm 0.02$
		0	4.19	$4.10 \pm 0.03$
		1	8.20	$8.16 \pm 0.09$
		2	13.37	$13.51 \pm 0.12$
$10^{-3}$	0.001	3	19.09	$19.59 \pm 0.14$
		-4	7.10	$6.26 \pm 0.08$
		-3	9.53	$8.24 \pm 0.09$
		-2	12.26	$10.79 \pm 0.10$
		-1	15.22	$14.14 \pm 0.12$
		0	18.36	$17.64 \pm 0.13$
	0.01	1	21.63	$21.92 \pm 0.15$
		3	28.39	$33.20 \pm 0.18$
		-2	0.04	$0.05 \pm 0.002$
		-1	1.49	$1.51 \pm 0.04$
		0	11.55	$11.14 \pm 0.33$
		1	29.26	$29.45 \pm 0.54$
		2	48.61	$51.50 \pm 0.23$
		3	68.27	$75.40 \pm 1.22$

Theoretical predictions are compared to the results of computer simulations. Fixation probabilities are expressed relative to that for a neutral allele ( $1/ND$ ).

quency dependence. Replacement of the parameters in this expression with the effective values derived here (Equations 7 and 8), and numerical evaluation of the resulting expression, yields the desired numerical predictions. Analogous use of results of KIMURA and OHTA (1969, Equation 12) yields predictions of mean times to fixation.

Figure 1 compares theoretical predictions of fixation probabilities to the results of simulations for  $N = 100$ ,  $D = 100$ ,  $s = 10^{-4}$ , and various values of  $m$  and  $h$ . The predictions are all very close to the observed values: all predictions are within 3.6% of the simulation results, and a majority are within 1%. The plot illustrates that the theory captures the effects of both the degree of dominance and the migration rate on the probability of fixation. Furthermore, all the observed mean fixation times are within a few percent of the theoretical predictions (maximum deviation 2.5%; data not shown), indicating that the diffusion is a good description of the

TABLE 2

Fixation probabilities for various numbers of demes

$D$	$s$	$m$	$h$	Relative fixation probability	
				Predicted	Observed
30	$3.3 \times 10^{-4}$	0.001	-4	1.78	$1.63 \pm 0.02$
			-3	1.89	$1.80 \pm 0.02$
			-2	2.00	$1.90 \pm 0.02$
			-1	2.12	$2.08 \pm 0.02$
			0	2.25	$2.23 \pm 0.03$
			1	2.38	$2.37 \pm 0.03$
		0.01	2	2.52	$2.61 \pm 0.03$
			3	2.66	$2.76 \pm 0.03$
			-4	0.46	$0.45 \pm 0.01$
			-3	0.68	$0.67 \pm 0.01$
			-2	0.99	$0.92 \pm 0.02$
			-1	1.41	$1.36 \pm 0.02$
			0	1.97	$1.94 \pm 0.02$
			1	2.70	$2.70 \pm 0.03$
			2	3.60	$3.62 \pm 0.03$
10	$10^{-3}$	0.001	3	4.68	$4.84 \pm 0.04$
			-4	1.78	$1.35 \pm 0.01$
			-3	1.89	$1.54 \pm 0.01$
			-2	2.00	$1.74 \pm 0.01$
			-1	2.12	$1.91 \pm 0.01$
			0	2.25	$2.18 \pm 0.01$
		0.01	1	2.38	$2.45 \pm 0.02$
			2	2.51	$2.74 \pm 0.02$
			3	2.66	$3.07 \pm 0.02$
			-4	0.46	$0.42 \pm 0.01$
			-3	0.68	$0.64 \pm 0.01$
			-2	0.99	$0.92 \pm 0.01$
			-1	1.41	$1.35 \pm 0.01$
			0	1.97	$1.95 \pm 0.01$
			1	2.69	$2.76 \pm 0.02$
2	3.59	$3.81 \pm 0.02$			
3	4.67	$5.13 \pm 0.02$			

Values of  $s$  were chosen such that  $N_e s_e \equiv NDs = 1$ . The deme size ( $N$ ) is 100 in all cases.

trajectory of allele frequency as well as the probability of ultimate fixation.

Table 1 presents results for larger values of  $s$ . For  $s = 3 \times 10^{-4}$ , the predictions are still close to the observed values, differing by at most 7.4%. For  $s = 10^{-3}$ , some of the predictions differ significantly from the simulation results for the smaller migration rate, especially at the extremes of over- and underdominance. This is to be expected because the weak selection assumption (Equations 2 and 3) is not met: for  $h = -4$ ,  $2N(1 - h)s = -1$ , and for  $h = 3$ ,  $2Nhs = 0.6$ , neither of which is very small in magnitude compared to 1. Even at these extremes, however, the predictions come within 16% of the observed values. For the less extreme values of  $h$  the theoretical predictions are quite good.

Table 2 shows some results for different numbers of demes ( $D$ ). With 30 demes, the predictions are again quite good: all of them are within 10% of the simulation

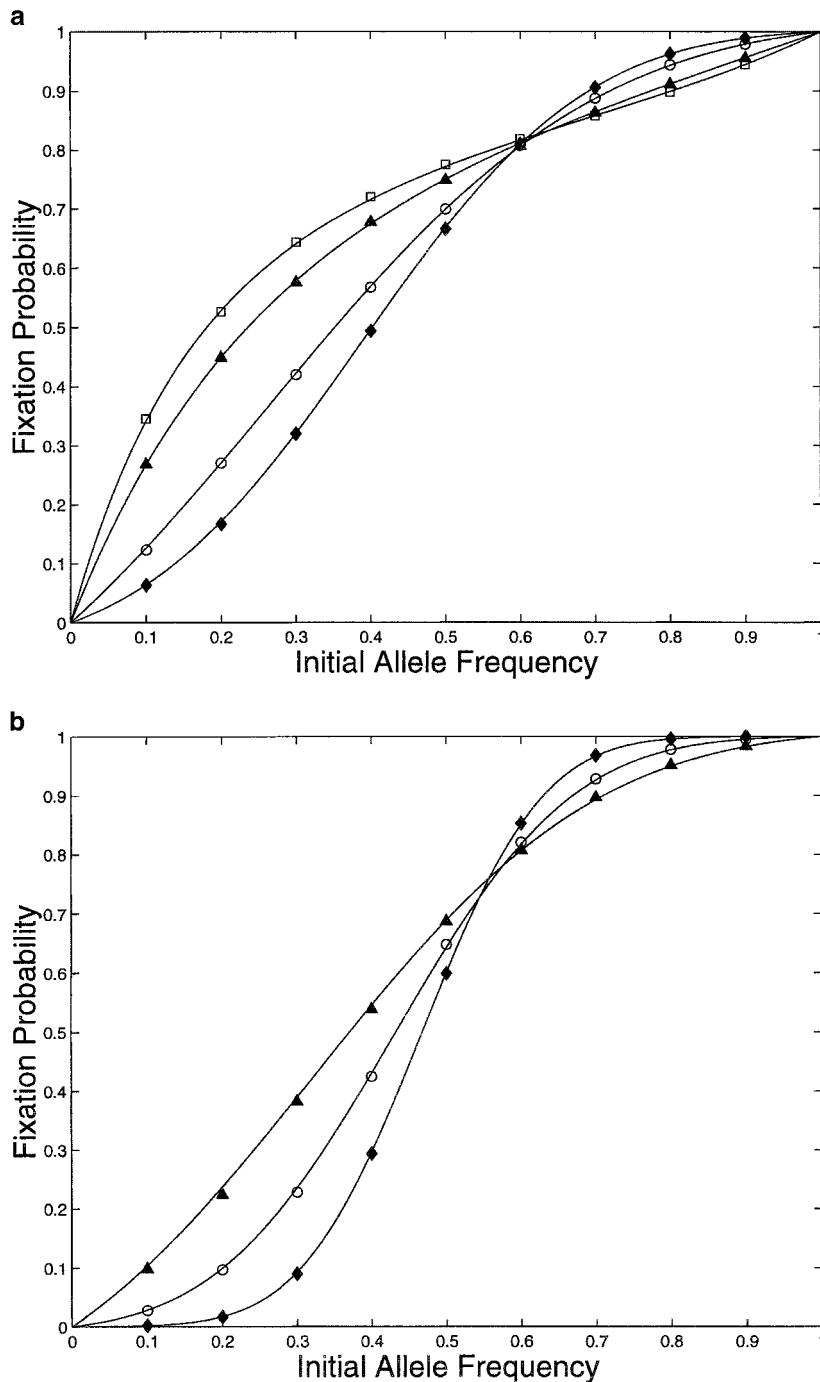


FIGURE 2.—Fixation probabilities as functions of initial allele frequency. Theoretical curves for various values of  $h$  and  $m$  are compared to simulation results (points), with  $N = 100$ ,  $D = 100$ , and  $s = 10^{-4}$ . (a)  $h = 3$  and  $m = 0.01$  (open squares and curve), or  $h = 3$  and  $m = 0.003$  (solid triangles and curve), or  $h = -4$  and  $m = 0.003$  (open circles and curve), or  $h = -4$  and  $m = 0.01$  (solid diamonds and curve). (b)  $h = -15$  and  $m = 0.001$  (solid triangles and curve),  $m = 0.003$  (open circles and curve), or  $m = 0.01$  (solid diamonds and curve).

results. With as few as 10 demes the assumption of a large number of demes is seriously violated. Although many of the predictions are close to the observations, some differ from them by as much as 32%.

In the simulation results presented so far the allele was initially present in a single copy. Figure 2 shows results for a range of initial allele frequencies. Figure 2a shows results for overdominance ( $h = 3$ ) and underdominance ( $h = -4$ ) with different migration rates. All of the points (simulation results) fall along the theoretical curves. All of the predicted fixation probabilities are within 3% of the simulation estimates, and the probabili-

ties of loss also agree within 3%. Figure 2b shows results for a case of strong underdominance ( $h = -15$ ). Results for such extreme underdominance were not given for an allele starting at a single copy because some fixation probabilities would be so low that they would be difficult to estimate by simulation. For higher initial allele frequencies the fixation probability is much larger and can be measured more easily. Figure 2b illustrates that the theory correctly predicts the reduced probability of fixation of a rare allele at high migration rates and the increased fixation probabilities at lower  $m$ . The predicted fixation probabilities are all within 6% of the

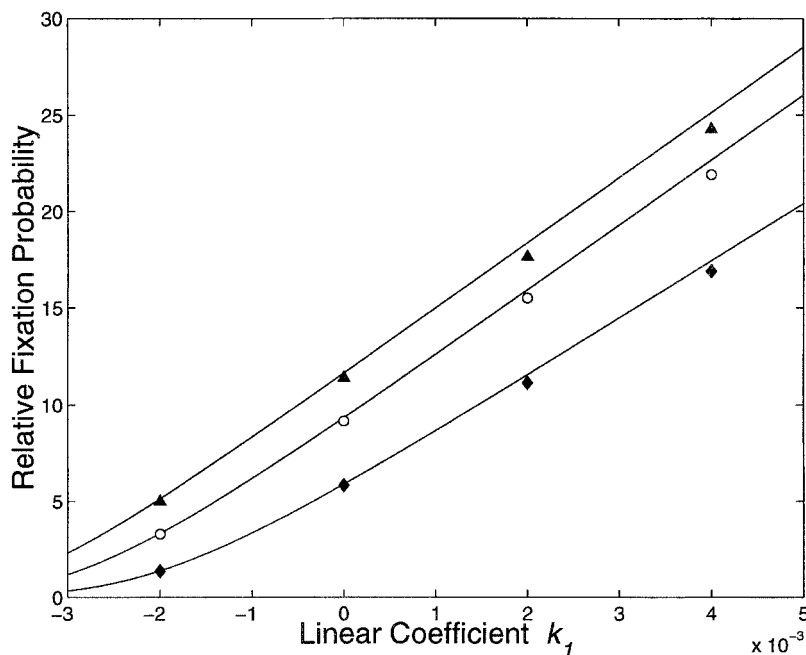


FIGURE 3.—Fixation probabilities for quadratic frequency dependence. Theoretical predictions (curves) are compared to simulation results (points) over a range of values of selection parameters. In all cases  $k_0 = 0$ ,  $k_2 = 0.002 - k_1$ ,  $N = 100$ ,  $D = 100$ , and the allele was initially present in a single copy. Results are presented for three migration rates:  $m = 0.001$  (solid triangles and curve),  $m = 0.003$  (open circles and curve), and  $m = 0.01$  (solid diamonds and curve).

estimates from simulation. Furthermore, all of the predicted probabilities of loss are within 5% of the simulation results.

The theory presented here covers any frequency dependence described by a polynomial. In the simulations discussed above this was a first-degree polynomial. Figure 3 compares predictions and results for cases of quadratic frequency dependence. The agreement of the predictions with the results is excellent: all of the predictions are within a few percent of the simulation results (the largest difference is 3.5%).

#### DISCUSSION

The theory presented here relates a subdivided population with frequency-dependent selection (including the case of dominance) to an equivalent panmictic population characterized by different parameters. Subdivision alters both the size of the equivalent panmictic population (the effective population size  $N_e$ ) and the effective values of all of the parameters describing selection. In the case of dominance, the dominance parameter  $h$  is in effect moved toward  $\frac{1}{2}$  by subdivision; *i.e.*, fitness is made effectively closer to additive. For frequency dependence described by any polynomial in allele frequency, the effective values of all of the polynomial coefficients are altered by subdivision.

The theoretical treatment assumed that selection was weak in one sense, but allowed that it was strong in another sense. The requirement is that selection is weak compared to drift in a subpopulation, *i.e.*, that the product of deme size ( $N$ ) and (marginal) selection coefficient is always small in magnitude compared to unity. This allows selection to be quite strong with respect to

drift in the population as a whole, so that selection may have a large effect on the fate of an allele without the assumptions being violated.

Computer simulations confirm that the theory closely predicts fixation probabilities so long as the parameters meet the stated conditions, namely that  $D$  is large and  $|N\hat{s}(x)| \ll 1$  for all  $x$  between zero and one. As expected, when the parameter values violate these conditions the predictions are less reliable. Nonetheless, when the conditions are moderately violated, for example, when selection and within-deme drift are of comparable strength, the predictions are still quite good.

A special case covered by the theoretical results is that of dominance for fitness in the absence of other sources of frequency dependence. It was shown that subdivision in effect decreases the degree of dominance, as measured by the deviation of the dominance parameter  $h$  from  $\frac{1}{2}$ , by a factor that depends only on the deme size  $N$  and the migration rate  $m$ . Specifically, subdivision changes the effective deviation of  $h$  from  $\frac{1}{2}$ ,  $h_e - \frac{1}{2}$ , by a factor of  $Nm/(Nm + 1)$ . The direction of this effect is in accord with the well-established fact that subdivision decreases the effect of dominance on fixation probabilities (WRIGHT 1940, 1941; SLATKIN 1981). It should be noted that this effect is not simply a matter of a reduction in the fraction of heterozygotes. If this were the sole source of the effect, one would expect a factor of  $1 - F_{st} = 2Nm/(2Nm + 1)$  reduction in the effect of dominance. If the same degree of inbreeding were achieved by, for example, brother-sister mating, while selection operated globally, then  $1 - F$  would indeed be the factor by which the effect of dominance was modified. The model analyzed here differs from that case in that competition is local to each subpopulation.

The diffusion approximation derived here for a finite island model completely describes the trajectory of allele frequency over time in the presence of dominance or frequency-dependent selection. The results followed from the moments of the distribution of within-deme allele frequencies. This is a beta distribution for the island model. Under other models of subdivision this distribution may have a different form, but the moments of the distribution nonetheless characterize the population. These moments could be derived theoretically or measured empirically. So long as  $F_{st}$  is independent of allele frequency and ordinary genetic drift is the only stochastic force operating, these moments can be used to relate the subdivided population to an equivalent panmictic population even in the presence of frequency-dependent selection.

I thank Christina Muirhead and Jon Wilkins for comments on the manuscript. John Wakeley encouraged me to work on this topic. This work was supported by National Science Foundation grant DEB-9815367 to John Wakeley.

#### LITERATURE CITED

- CHERRY, J. L., and J. WAKELEY, 2003 A diffusion approximation for selection and drift in a subdivided population. *Genetics* **163**: 421–428.
- DOBZHANSKY, T., and S. WRIGHT, 1941 Genetics of natural populations. V. Relations between mutation rate and accumulation of lethals in populations of *Drosophila pseudoobscura*. *Genetics* **26**: 23–51.
- KIMURA, M., 1957 Some problems of stochastic processes in genetics. *Ann. Math. Stat.* **28**: 882–901.
- KIMURA, M., and T. OHTA, 1969 The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**: 763–771.
- LANDE, R., 1985 The fixation of chromosomal rearrangements in a subdivided population with local extinction and colonization. *Heredity* **54** (3): 323–332.
- MARUYAMA, T., 1970a Effective number of alleles in a subdivided population. *Theor. Popul. Biol.* **1**: 273–306.
- MARUYAMA, T., 1970b On the fixation probability of mutant genes in a subdivided population. *Genet. Res.* **15**: 221–225.
- MARUYAMA, T., 1972a Distribution of gene frequencies in a geographically structured finite population. I. Distribution of neutral genes and of genes with small effect. *Ann. Hum. Genet.* **35**: 411–423.
- MARUYAMA, T., 1972b Distribution of gene frequencies in a geographically structured population. III. Distribution of deleterious genes and genetic correlation between different localities. *Ann. Hum. Genet.* **36**: 99–108.
- MARUYAMA, T., 1972c Distribution of gene frequencies in a geographically structured population. II. Distribution of deleterious genes and of lethal genes. *Ann. Hum. Genet.* **35**: 425–432.
- MARUYAMA, T., 1974 A simple proof that certain quantities are independent of the geographical structure of population. *Theor. Popul. Biol.* **5**: 148–154.
- MARUYAMA, T., and M. KIMURA, 1980 Genetic variability and effective population size when local extinction and recolonization of subpopulations are frequent. *Proc. Natl. Acad. Sci. USA* **77**: 6710–6714.
- NAGYLAKI, T., 1998 The expected number of heterozygous sites in a subdivided population. *Genetics* **149**: 1599–1604.
- NEI, M., and N. TAKAHATA, 1993 Effective population size, genetic diversity, and coalescence time in subdivided populations. *J. Mol. Evol.* **37**: 240–244.
- SANTIAGO, E., and A. CABALLERO, 1995 Effective size of populations under selection. *Genetics* **139**: 1013–1030.
- SLATKIN, M., 1977 Gene flow and genetic drift in a species subject to frequent local extinctions. *Theor. Popul. Biol.* **12**: 253–262.
- SLATKIN, M., 1981 Fixation probabilities and fixation times in a subdivided population. *Evolution* **35**: 477–488.
- SLATKIN, M., 1991 Inbreeding coefficients and coalescence times. *Genet. Res.* **58**: 167–175.
- SPIRITO, F., M. RIZZONI and C. ROSSI, 1993 The establishment of underdominant chromosomal rearrangements in multi-deme systems with local extinction and colonization. *Theor. Popul. Biol.* **44**: 80–94.
- TAKAHATA, N., 1991 Genealogy of neutral genes and spreading of selected mutations in a geographically structured population. *Genetics* **129**: 585–595.
- WANG, J., and A. CABALLERO, 1999 Developments in predicting the effective size of subdivided populations. *Heredity* **82**: 212–226.
- WHITLOCK, M. C., and N. H. BARTON, 1997 The effective size of a subdivided population. *Genetics* **146**: 427–441.
- WRIGHT, S., 1931 Evolution in Mendelian populations. *Genetics* **16**: 97–159.
- WRIGHT, S., 1939 *Actualités Scientifiques et Industrielles 802: Exposes de Biometrie et de la Statistique Biologique XIII*. Hermann et Cie, Paris.
- WRIGHT, S., 1940 Breeding structure of populations in relation to speciation. *Am. Nat.* **74**: 232–248.
- WRIGHT, S., 1941 On the probability of fixation of reciprocal translocations. *Am. Nat.* **75**: 513–522.
- WRIGHT, S., 1943 Isolation by distance. *Genetics* **28**: 114–138.

Communicating editor: N. TAKAHATA