Theoretical and Applied Genetics 40, 245—250 (1970)

Effects of Selection and Drift on the Dynamics of Finite Populations

IV. Gene Frequency dependent Selection

R. F. NASSAR

Department of Statistics and Computer Science, Kansas State University, Manhattan (USA)

Summary. A genetic model of a two allelic locus involving gene frequency dependent selection with overdominance or underdominance was investigated with regard to the probability of and the expected time to fixation of an allele in the face of stochastic variations arising from finite population size. Results showed the following:

1. If the probability of fixation is larger than for no selection, the case of fequency dependent selection with overdominance at equilibrium gave rise to the highest probability of fixation. If, however, the probability of fixation is less than for no selection, the case of frequency dependent selection and underdominance at equilibrium gives rise to the highest probability of fixation.
2. The expected time to fixation is largest for the case of overdominance at equilibrium and smallest for the case of underdominance at equilibrium if $0.2 < \hat{p} < 0.8$.
3. For equilibrium gene frequencies ($\hat{p}$) above 0.8 or below 0.2 there was acceleration in the time to fixation. This acceleration occurred over a wide range of initial gene frequencies that includes the $\hat{p}$ value.

1. Introduction

It is well known that natural populations contain a large amount of genetic variability. Many genetic models have been invoked to explain the maintenance of this variability. Of the oldest and perhaps most popular is the overdominance model. Recently, there has been increasing evidence that the relative selective values of certain genotypes depend on their relative frequencies (Tobari and Kojima, 1967; Kojima and Yarbrough, 1967; Kaul and Parsons, 1965; Erhman, 1966 and others). Evidence also exists (unpublished results) to show that the mode of selection with regard to a third chromosome inversion in *Drosophila melanogaster* involves frequency-dependency and overdominance. It is also possible that at a stable equilibrium, when selection is frequency-dependent, the relative selective values could show underdominance. Thus, a general model that provides for frequency-dependent selection with overdominance and underdominance is of interest to consider in explaining the maintenance of genetic variability in natural populations.

In studies of the nature of forces maintaining variation in natural populations, population geneticists have, to varying degrees, used information from small laboratory populations. It is deemed useful, therefore, to analyse the genetic behavior of small populations under genetic models that serve to explain the maintenance of variability in infinite populations. Knowledge of the dynamics of small populations is also useful in plant and animal breeding as well as in experimental quantitative genetics. It is also known that gene frequency changes under the combined effects of selection and drift due to finite population size are approximately determined by $N s$ (N = effective population number, s = selection coefficient) which makes it possible to apply results from small populations to large populations.

In finite populations, barring mutation and migration, the genetic variability is transient, no stable equilibrium is possible and the end result is fixation or loss of an allele. In this regard there are two problems of main interest:

1. the ultimate probability with which an allele is fixed and 2. the time in generations it takes to fixation. In this paper points 1 and 2 are investigated in the case of a two allelic locus and frequency-dependent selection with overdominance or underdominance.

2. Materials and Methods

Consider that at a locus with two alleles the relative selective values of the three genotypes AA, Aa and aa are $1 + s - b_1 p$, $1 + h s$ and $1 - s + b_2 p$, respectively (s is the selection coefficient and h is the dominance factor). The model assumes that the selective values of the homozygotes decrease linearly as the corresponding chromomosome frequencies increase. The heterozygote takes on different values depending on h. At a stable gene frequency equilibrium in an infinite population ($\hat{p}$) there are three cases to distinguish with regard to the relative selective values of the three genotypes. The first is selective neutrality, the second overdominance and the third underdominance. For $b_1 = b_2$, selective neutrality occurs when $h = h_n = 0$, for $b_1 < b_2$ ($b_1 \geq 0$) it occurs for $0 < h_n \leq 1$ (at $b_1 = 0$, $h_n = 1$); and for $b_1 > b_2$ ($b_2 \geq 0$) selective neutrality occurs in the range $-1 \leq h_n < 0$ (at $b_2 = 0$, $h_n = -1$). In all cases, if $h > h_n$, there is overdominance at equilibrium and if $h < h_n$, there is underdominance at equilibrium.

At equilibrium

$$\Delta p = \frac{\begin{bmatrix} b_1 p^3 (1-p) + b_2 p^2 (1-p)^2 \\ - s\,(p\,(1-p) + hp\,(1-2\,p)\,(1-p)) \end{bmatrix}}{\overline{w}} = 0$$

(where $\overline{w}$ = the mean fitness)
which reduces to

$$\hat{p}\,(b_1\hat{p} + b_2\,(1-\hat{p})) = s\,(h + 1 - 2\,\hat{p}h) \tag{2.1}$$

The equilibrium is stable if

$$\left.\frac{\partial\,\Delta\,p}{\partial\,p}\right|_{\hat{p}} < 0 \tag{2.2}$$

(2.2) reduces to

$$A\,\hat{p}^3 + C\,\hat{p} - 2\,D < 0 \tag{2.3}$$

where

$$A = b_1 - b_2 = \Delta\,b$$
$$C = s\,(1 + 3\,h) + b_2$$
$$D = s\,(1 + h)$$

when $\Delta\,b$ is small such that Ap^3 is negligible, $\hat{p}$ satisfies the following inequality

$$\hat{p} < \frac{2}{3 + \dfrac{\frac{b_2}{s} - 2}{1 + h}}$$

This shows that if $b_2 \geq 2\,s$, the value of the stable equilibrium gene frequency is $\hat{p} < 0.67$. The condition $b_2 \geq 2\,s$ is probably not stringent if we consider that in natural populations s is likely to be small and b_2 relatively large.

In what follows the word equilibrium refers to a stable equilibrium in an infinite population.

Consider a random mating diploid population of fixed size N with non-overlapping generations. Assume that the $2\,N$ gametes in adults of a given generation are a random sample of the infinite number produced (or potentially produced) by adults of the previous generation. In a previous paper (Carr and Nassar, 1970) it was shown that the stochastic process $(X\,(0), X(1), \ldots\ldots X(t)), X(t)$ being the number of A alleles in generation t, is a finite Markov chain with state space $0, 1, \ldots\ldots, 2\,N$ and stochastic matrix P, having as its $(i\,j)th$ element

$$P_{ij} = \binom{2N}{j} P_i^j\,(1 - P_i)^{2N-j} \quad \begin{matrix} i = 0, 1, \ldots\ldots, 2\,N \\ j = 0, 1, \ldots\ldots, 2\,N \end{matrix} \tag{2.4}$$

For this model

$$p_i = \frac{x_i + x_i^2\,(s - b_1\,x_i) + hsx_i\,(1 - x_i)}{\begin{bmatrix} 1 + x_i^3\,(b_2 - b_1) + 2\,x_i\,(1 - x_i)\,hs + \\ + (1 - 2\,x_i)\,(b_2 x_i - s) \end{bmatrix}}\,;\; x_i = \frac{i}{2\,N} \tag{2.5}$$

is the probability of allele A in the infinite population after selection and before the random sample of N individuals or $2\,N$ gametes.

The chain is stationary with two absorbing states $(0,2\,N)$ and with remaining $2\,N - 1$ transient states.

Let

$Q = [P_{ij}]$, a $(2\,N - 1)$ by $(2\,N - 1)$ matrix of transition probabilities
$e\,(T) = e\,(t_i)$, a vector of order $2\,N - 1$ whose elements are the expected time to fixation or loss from an initial state $X(0) = i$ $(i = 1, 2, \ldots, 2\,N - 1)$
$r = [p_i, 2\,N]$, a $(2\,N - 1)$ vector of transition probabilities from transient state i $(i = 1, 2, \ldots, 2\,N - 1)$ to the absorbing state $2\,N$.
$f_i =$ the probability of ultimately fixing the A allele given that the initial state was i.
$f = [f_i]$, a $2\,N - 1$ vector of the probabilities of ultimate fixation

It can be shown (see Carr and Nassar, 1970) that

$$f = (I - Q)^{-1}\,r \tag{2.6}$$

and

$$e(T) = (\mathrm{I} - Q)^{-1}\,\delta \tag{2.7}$$

where I is the identity matrix and $\delta' = (1, 1, \ldots 1)$ is the unit vector of order $2\,N - 1$.

Equations (2.6) and (2.7) were solved for $N = 10$ and for various values of the parameters s_1, b_1, b_2 and h. The solution was obtained numerically on the computer by inverting the matrix $(I - Q)$ using the iterative procedure discussed in the paper by Carr and Nassar (1970).

To investigate the three possible relations between b_1 and b_2 $(b_1 = b_2,\ b_1 > b_2,\ b_1 < b_2)$ the following representative cases were considered: $b_1 = b_2$, $b_1 = 1/2\,b_2$, $b_1 = 0$, $b_2 = 1/2\,b_1$ and $b_2 = 0$. For each set of b_1, b_2 values three h values were chosen such that at equilibrium the relative selective values of the three genotypes would show overdominance, under-dominance or selective neutrality. The values of h that give rise to selctive neutrality at equilibrium were:

$h_n = 0$ $(b_1 = b_2)$, $h_n = .3333$ $(b_1 = 1/2\,b_2)$, $h_n = 1$ $(b_1 = 0)$, $h_n = -.3333$ $(b_2 = 1/2\,b_1)$, $h_n = -1$ $(b_2 = 0)$.

3. Results

Expected time to fixation

Figures 1 and 2a, b. . e have as ordinates the ratio $R_{\hat{p},s(i)} = E_{\hat{p},s}(t_i)/E_0(t_i)$, where $E_0(t_i)$ is the expected time to fixation in the case of no selection given that the initial gene frequency is $\frac{i}{2\,N}$ $(i = 1, 2\,.\,., 2N\text{-}1)$; and $E_{\hat{p},s}(t_i)$ is the expected time to fixation with selection (s) and equilibrium gene frequency $(\hat{p})$. The abscissas are the original gene frequencies $\left(\frac{i}{2\,N}\right)$. The horizontal line is $E_0(t_i)/E_0(t_i) = 1$. If $R_{\hat{p},s}(i) > 1$, then this combination of s and $\hat{p}$ has the effect of retarding the expected time to fixation in a population with initial gene frequency $\frac{i}{2\,N}$. If $R_{\hat{p},s}(i) < 1$, acceleration in the expected time to fixation is in effect.

In the case of relatively weak selection $(Ns = 1)$ figure 1a, . . . e, it is seen that for $b_1 = b_2$, $b_2 = 1/2 b_1$, $b_2 = 0$ and $\hat{p}$ of 0.2 and 0.9 there was an acceleration in the expected time to fixation over a range of initial gene frequencies. In the case of $b_1 = 0$, the acceleration was only for $\hat{p} = .9$. In all cases the acceleration occurred for overdominance at equilibrium. Acceleration in the case of underdominance occurred only at $b_2 = 1/2\,b_1$ and $b_2 = 0$, and in the case of selective neutrality only for $b_2 = 0$. For all the b_1 and b_2 relations that were considered the retardation in the expected time to fixation was largest for the case of overdominance and smallest for the case of underdominance at equilibrium. In the former case the retardation was larger for high than for low initial gene frequencies when $\hat{p}$ was less than intermediate in value. This was reversed for $\hat{p}$ larger than intermedi-

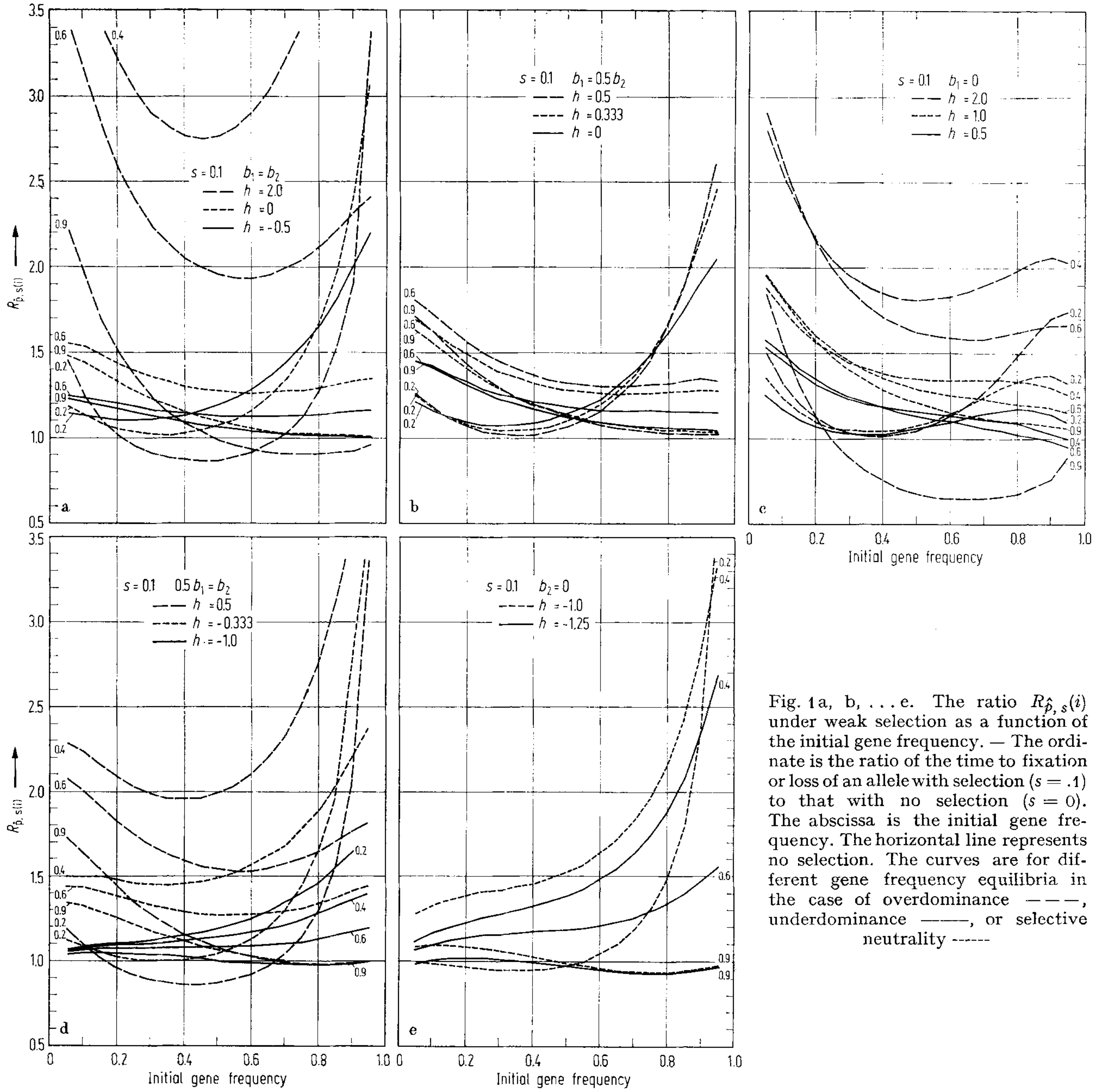

Fig. 1a, b, ... e. The ratio $R_{\hat{p},s}(i)$ under weak selection as a function of the initial gene frequency. — The ordinate is the ratio of the time to fixation or loss of an allele with selection ($s = .1$) to that with no selection ($s = 0$). The abscissa is the initial gene frequency. The horizontal line represents no selection. The curves are for different gene frequency equilibria in the case of overdominance — — —, underdominance ———, or selective neutrality ------

ate value. The slopes of the $R_{\hat{p},s}(i)$ curves for all $\hat{p}$ larger than intermediate value, were reduced by going from overdominance to selective neutrality and further to underdominance. This tended to reduce the differences in $R_{\hat{p},s}(i)$ between low and high initial gene frequencies. At underdominance these differences were least pronounced.

In the case of overdominance at equilibrium the retardation is largest at intermediate $\hat{p}$ values. Among the different b_1 and b_2 relations it was largest for $b_1 = b_2$ and smallest for $b_1 = 1/2\, b_2$. For $\hat{p}$ values above 0.8 there was little effect of the change in the relative values of b_1 and b_2 on the $R_{\hat{p},s}(i)$ and for $\hat{p} < .02$ the effect was only for high initial gene frequencies. In the latter event the retardation was again largest for $b_1 = b_2$ and smallest for $b_1 = 1/2\, b_2$ and $b_1 = 0$. The effect on $R_{\hat{p},s}(i)$ of varying the relative values of b_1 and b_2 was small in the case of selective neutrality and underdominance. As in the case of retardation, acceleration was also smallest for $b_1 < b_2$.

Stronger selection ($N s = 4$) caused little qualitative changes in the results. The main effect of stronger selection, figure 2a, ... e, was to increase retardation for intermediate $\hat{p}$ values and to simultaneously increase acceleration for $\hat{p}$ above 0.8 and below 0.2. For all the b_1 and b_2 relations acceleration occurred at $\hat{p} = .8$ or above for overdominance and at $\hat{p} = .9$ or above for underdominance and selective neutrality.

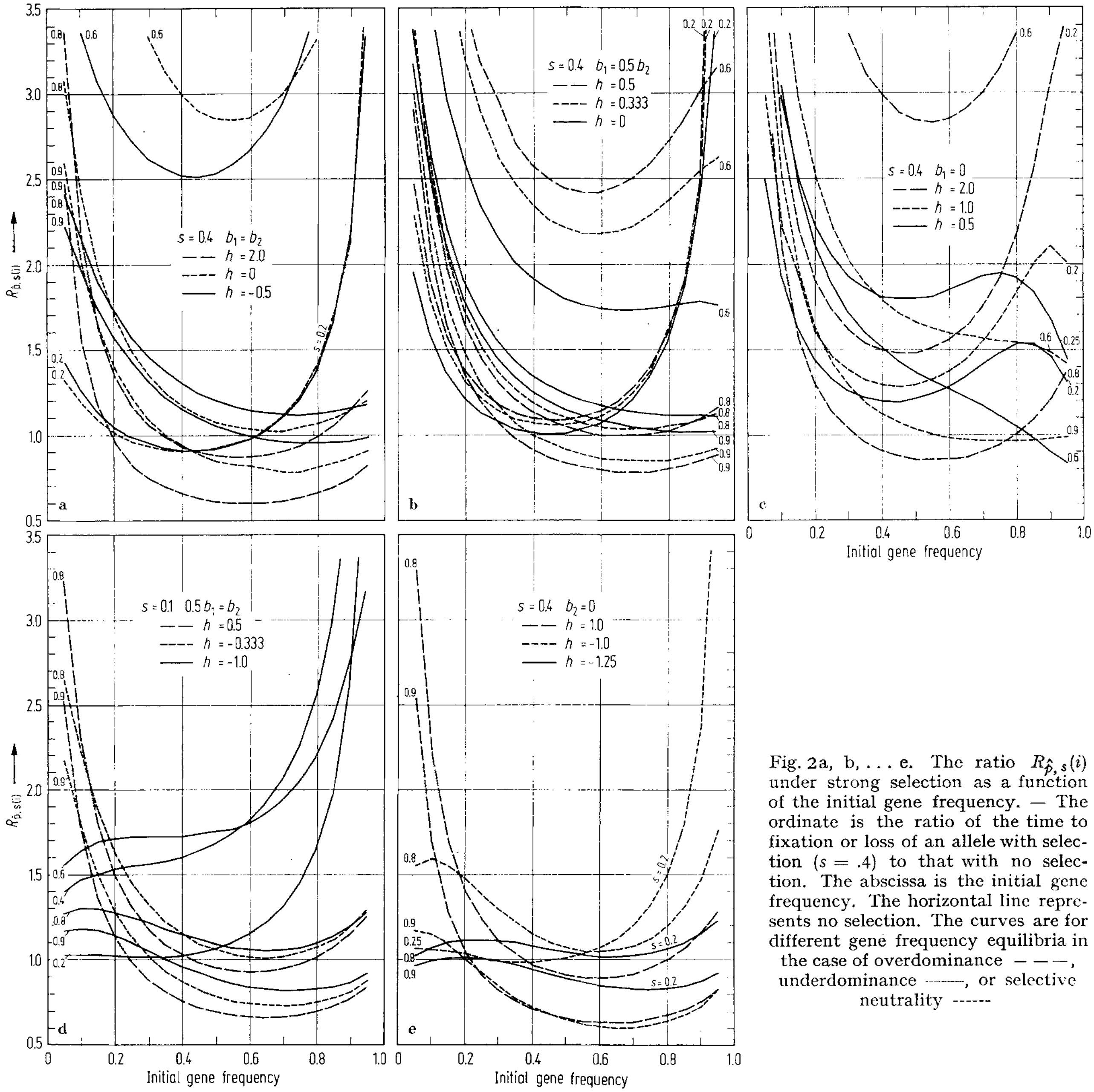

Fig. 2a, b, . . . e. The ratio $R_{\hat{p},s}(i)$ under strong selection as a function of the initial gene frequency. — The ordinate is the ratio of the time to fixation or loss of an allele with selection ($s = .4$) to that with no selection. The abscissa is the initial gene frequency. The horizontal line represents no selection. The curves are for different gene frequency equilibria in the case of overdominance — — —, underdominance ———, or selective neutrality -----

Acceleration also occurred at $\hat{p} = .2$ or below, but only for $b_1 \geq b_2$. The magnitude and range of acceleration was largest for overdominance and smallest for underdominance.

Probability of ultimate fixation

Figure 3a, b . . . f, has the probability of fixation, f_i, as ordinate and the initial gene frequency, $i/2N$, as abscissa. The diagonal line in each graph represents the case of no selection ($s = 0$, $b_1 = b_2 = 0$).

From results in figure 3 a, b. . . e it is seen that for weak selection ($Ns = 1$) and for $\hat{p}$ in the neighbourhood of 0.5 ($\hat{p}$ between .4 and .6) there was a point x in initial gene frequency (point of intersection of a curve with the diagonal line) above which f_i was less than it would be under no selection. For $\hat{p}$ below 0.5, x lies below 0.5 and vice versa (the only exception was for the case of underdominance at $b_1 = 0$). At $\hat{p} = .5$, x as expected, was equal to 0.5 regardless of the mode of selection at equilibrium. For $\hat{p} > .6$ and $\hat{p} < .4$ the probability of fixation was above and below the diagonal line respectively. The probability of fixation above the diagonal line was largest for overdominance and smallest for underdominance. This was reversed when f_i was below the diagonal line. The ultimate probability of fixation was an increasing function of $\hat{p}$ as well as of the initial gene frequency.

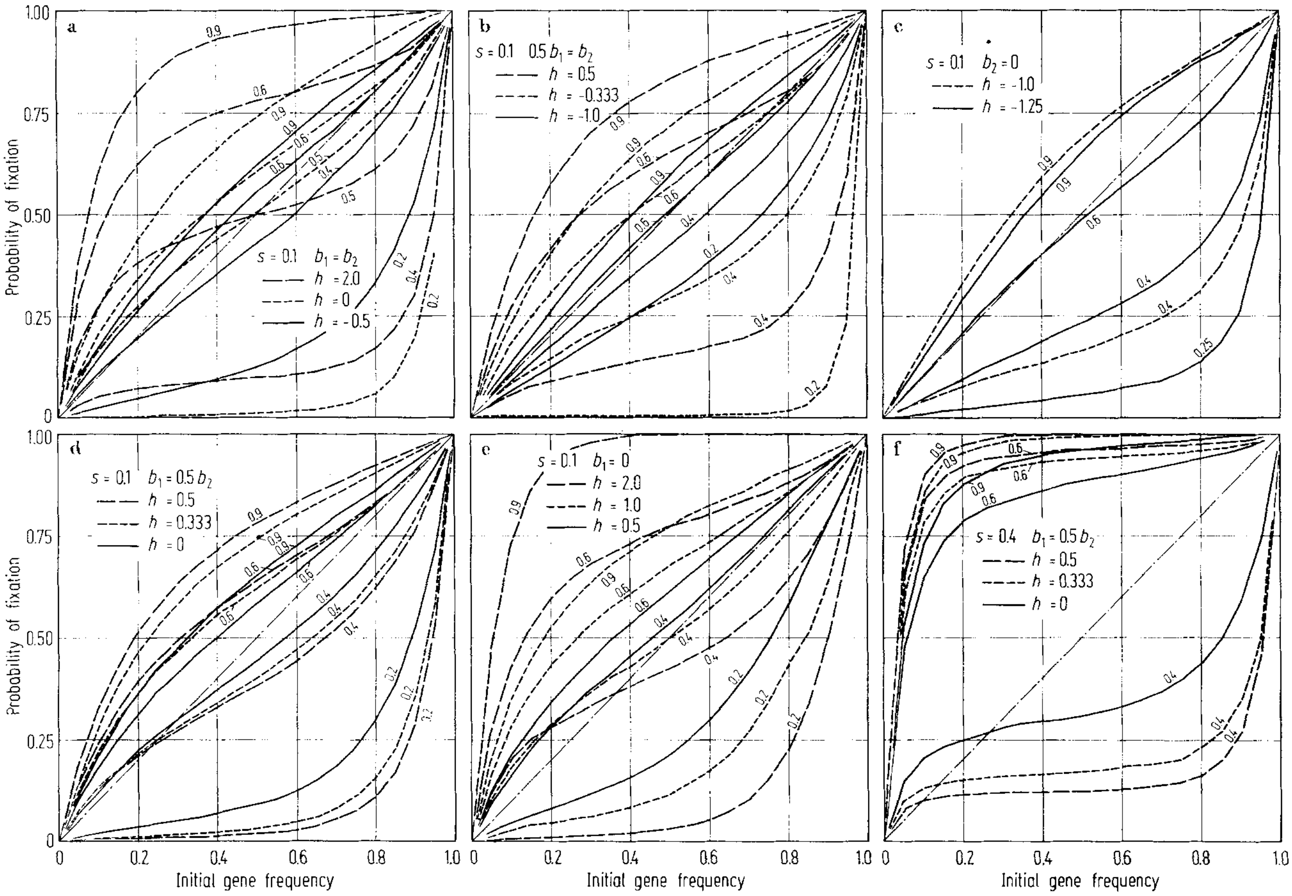

Fig. 3a, b, ...f. The probability of fixation of an allele as a function of the initial gene frequency. — The ordinate is the probability of fixation of an allele with selection. The abscissa is the initial gene frequency. The diagonal line represents the case of no selection. The curves are for different gene frequency equilibria in the case of overdominance — — —, underdominance - - —, or selective neutrality ------

The effect of increasing the selection intensity ($N s = 4$) was, for all b_1 and b_2 relations, to increase the ultimate probability of fixation when above the diagonal line and to decrease it when below the line (this is typically shown in figure 3f). As $N s$ gets larger the end result of course is that f_i goes to 0 or 1. With an increase in the selection intensity the tendency was for x to decrease if it were below 0.5 and to increase if above 0.5. The change in the b_1 and b_2 relations from $b_1 = b_2$ to $b_1 < b_2$ increased in general f_i. The effect of reducing b_2 in relation to b_1 did in general decrease the ultimate probability of fixation.

4. Discussion

The most salient features of these results with regard to the ultimate probability of fixation is the fact that this probability is not only determined by the value of the stable gene frequency equilibrium in the infinite population, but also by the mechanism that maintains the stable equilibrium. This can be helpful in using results from small populations to infer about the genetic prpperties of large populations. If the probability of ultimate fixation is larger than in the case of no selection, the case of overdominance at equilibrium gives rise to the highest probability of fixation. If, however, f_i is less than in the case of no selection, the case of underdominance at equilibrium gives rise to the highest probability of fixation. This relationship holds for all relative values of b_1 and b_2. The differences in f_i between overdominance, selective neutrality and underdominance are most pronounced at intermediate initial gene frequencies.

The retardation in the expected time to fixation, as in the case of f_i, is a function both of $\hat{p}$ and the mechanism maintaining equilibrium in the infinite population. Retardation is largest for the case of overdominance and smallest for the case of underdominance. Also the difference in the magnitude of the retardation between that for small and large initial gene frequencies is largest for overdominance at equilibrium and smallest for underdominance at equilibrium. This is particularly so when the equilibrium gene frequency is below 0.2 or above 0.8.

When $\hat{p}$ is between 0.2 and 0.8, it can be inferred from these results that gene frequency dependent selection with overdominance at equilibrium is most effective in maintaining variability in natural populations.

At $\hat{p}$ values above .8 and below 0.2 and when the initial gene frequency is in the neighbourhood of the equilibrium frequency none of the above selection systems is very effective in maintaining variability in the face of stochastic variation. In this regard, however, underdominance is better than overdominance or selective neutrality by virtue of the fact that it does not accelerate the time to fixation by as much as the others do.

Overdominance without gene frequency dependent selection is more effective in maintaining variability than overdominance with gene frequency dependent selection. However, the segregational load in the former case can be much larger than in the latter. In the latter case, intermediate $\hat{p}$ values can be maintained with slight overdominance effect (slight segregational load at equilibrium). This is in contrast with a large overdominance effect (large segregational load) needed to maintain intermediate $\hat{p}$ values when overdominance is not coupled with gene frequency dependent selection. Frequency dependent selection with selective neutrality at equilibrium gives rise to no or little segregational load and is next in line to overdominance as an effective way of maintaining variability in the face of stochastic variations.

Zusammenfassung

Es wird das genetische Modell eines biallelen Locus mit genfrequenzabhängiger Selektion bei Vorliegen von Über- oder Unterdominanz im Hinblick auf die Wahrscheinlichkeit der Fixierung und die erwartete Zeit bis zur Fixierung eines Allels im Zusammenhang mit stochastischer Variation, die aus einem begrenzten Populationsumfang resultiert, untersucht. Folgende Ergebnisse wurden erhalten:

1. Wenn die Fixierungswahrscheinlichkeit größer ist als ohne Selektion, dann führt der Fall häufigkeitsabhängiger Selektion in Verbindung mit Superdominanz im Gleichgewicht zur größten Fixierungswahrscheinlichkeit. Ist jedoch die Fixierungswahrscheinlichkeit niedriger als ohne Selektion, dann führt der Fall häufigkeitsabhängiger Selektion in Verbindung mit Unterdominanz im Gleichgewicht zur größten Fixierungswahrscheinlichkeit.
2. Die erwartete Zeit bis zur Fixierung ist am größten im Fall der Superdominanz im Gleichgewicht und am kleinsten für den Fall der Unterdominanz im Gleichgewicht, wenn $0{,}2 < \hat{p} < 0{,}8$ ist.
3. Für Gleichgewichtsfrequenzen ($\hat{p}$) über 0,8 oder unter 0,2 gibt es eine Akzeleration in der Fixierungszeit. Diese Akzeleration tritt über einen breiten Bereich ursprünglicher Genfrequenzen auf, die den $\hat{p}$ Wert einschließen.

Diese Ergebnisse werden im Hinblick auf die Erhaltung der Variabilität in natürlichen Populationen, die einer stochastischen Variation unterliegen, diskutiert.

Acknowledgement

This work was supported by a Senior Research Fellowship from the Alexander von Humboldt-Stiftung.

References

1. Carr, N. R., Nassar, R. F.: Effects of selection and drift on the dynamics of finite populations. I. Ultimate probability of fixation of an allele. Biometrics (1970 in press). — 2. Erhman, L.: Mating success and genotype frequency in *Drosophila*. Animal Behaviour **14**, 332—339 (1966). — 3. Kaul, D., Parsons, P. A.: Genotype control of mating speed. Heredity **20**, 381—392 (1965). — 4. Kojima, K., Yarbrough, K. M.: Frequency dependent selection at the Esterase 6 locus in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U.S. **57**, 645—649 (1967). — 5. Tobari, Y. N., Kojima, K.: Selective modes associated with inversion karyotypes in *Drosophila ananassae* I. Frequency dependent selection. Genetics **57**, 179—188 (1967).

Received April 15, 1970
Communicated by W. Seyffert

R. F. Nassar
Institut für Forstgenetik
Sieker Landstr. 2
207 Schmalenbeck (Germany/BDR)