

POLYMORPHISMS IN CYCLICALLY-VARYING ENVIRONMENTS*

THOMAS NAGYLAKI†

Department of Medical Genetics, University of Wisconsin, Madison, Wisconsin 53706

Received 5.xi.74

SUMMARY

We analysed both continuous and discrete two-allele models of cyclically-varying environments with an arbitrary degree of dominance. In continuous models, the gene frequency fluctuates with the period of the environmental oscillation. For the discrete case, the calculations were carried out to second order in selection. In contrast to the continuous models, and depending on the amount of dominance and the initial gene frequency, fixation is possible as well as polymorphism.

1. INTRODUCTION

THE maintenance of genetic variability in natural populations is a subject of considerable interest in population genetics and evolutionary theory. At any given time, some proportion of the polymorphisms in a population are transient. Many mechanisms have been proposed which preserve genetic diversity in the equilibrium state. Recurrent mutation will keep deleterious genes in a population at low frequencies, preventing their elimination by natural selection. Kimura and Ohta (1971, Chs. 8 and 9) discuss the amount of genetic variability maintained by mutation and random drift in panmictic and geographically structured populations. These authors also review some of the pertinent experimental data.

While mutation is obviously the ultimate source of all genetic heterogeneity, polymorphic equilibria may be stable without recurrent mutation. The well-known overdominant equilibrium is probably the most important example of such a situation. If a population occupies multiple ecological niches, under suitable conditions a stable polymorphism may occur (Levene, 1967). Heterogeneity will be maintained if rare genotypes are favoured by selection. Numerous cases of this type of frequency-dependent selection are analysed by Wright (1969, Ch. 5), who also treats many other selective mechanisms which can produce stable equilibrium, *e.g.* unequal selection in the two sexes, sex linkage, and meiotic drive (1969, Ch. 3).

Balanced polymorphisms may be due to time-dependent selection. Haldane and Jayakar (1963) proved that if one of two alleles is completely dominant to the other, both alleles will remain in the population if the arithmetic mean of the fitnesses of recessives (relative to those of dominants) in different generations exceeds unity and the geometric mean is less than one. Recently, Hartl and Cook (1973) have shown in a two-allele model that if the fitnesses are linear functions of a random variable and have the same mean, then natural selection favours the genotype with the smallest

* Paper number 1640 from the Genetics Laboratory, University of Wisconsin. Supported by the National Institutes of Health (Grant GM-15422).

† Present address: Dept. of Biophysics, Univ. of Chicago, 920 East 58th St., Chicago, Illinois 60637.

variance in fitness. In particular, there will be a stable polymorphism whenever twice the standard deviation of the heterozygote fitness is less than the sum of the corresponding homozygote standard deviations.

In this paper, we shall study the consequences of periodic variations in selection coefficients on a single autosomal diallelic locus in a population sufficiently large to permit neglect of random drift. An example of this type of phenomenon is the seasonal variation in the frequencies of different gene arrangements of *Drosophila pseudoobscura* discussed by Dobzhansky (1971, pp. 109-133). Larval crowding appears to favour the homokaryotype *ST/ST* relative to *CH/CH*, while its absence does the reverse. Thus, seasonal fluctuations in population size will induce regular oscillations in the selection coefficients. Allard and Workman (1963) have observed cyclic changes in selection coefficients in lima bean populations. Powell (1971) has reported that the average heterozygosity and the average number of alleles maintained at a locus in 13 experimental populations of *Drosophila willistoni* were higher in heterogeneous than in homogeneous environments. It is easy to see from his table 1 that this effect still exists if only the cage temperature is varied at weekly intervals between 19° and 25°C, while the other two factors, which provided spatial heterogeneity, are kept constant.

This problem was investigated theoretically some time ago by Dempster (1955) and Kimura (1955). Our aim is to test the generality of their results by analysing broader classes of models with continuous random births and deaths and with discrete, non-overlapping generations.

2. THE CONTINUOUS MODEL

We shall denote the frequencies of the alleles *A* and *a* by *p* and *q*. Assuming Hardy-Weinberg proportions, the time-derivative of *p* is given by (see, e.g. Crow and Kimura, 1970, p. 192).

$$\dot{p} = p(1-p)[(m_{12}-m_{22})+(m_{11}-2m_{12}+m_{22})p], \quad (1)$$

where m_{11} , m_{12} ($= m_{21}$), and m_{22} represent the Malthusian parameters of the genotypes *AA*, *Aa*, *aa*. Usually, even with random mating, Hardy-Weinberg proportions apply only approximately (Nagylaki and Crow, 1974).

We shall impose the biologically reasonable restriction that the two selective differences in (1) have the same time dependence:

$$m_{12}-m_{22} = af(t), \quad (2)$$

$$m_{11}-2m_{12}+m_{22} = bf(t), \quad (3)$$

where *a* and *b* are constants and *f(t)* is some function of time. If the selective differences are affected by the environmental variation as specified by (2) and (3), the degree of dominance,

$$\frac{(m_{12}-m_{22})-(m_{11}-m_{12})}{m_{12}-m_{22}} = -\frac{b}{a},$$

is constant. The variables in (1) separate and we find by straightforward integration

$$F[p(t)] - F[p(0)] = g(t), \quad (4)$$

where

$$F(p) = \frac{\ln p}{a} - \frac{\ln(1-p)}{a+b} - \frac{b \ln |a+bp|}{a(a+b)}, \quad (5)$$

$$g(t) = \int_0^t f(t') dt'. \quad (6)$$

Let us focus our attention now on regularly oscillating environments by supposing that $f(t)$ has period T ,

$$f(t+T) = f(t), \quad \text{for all } t, \quad (7)$$

and mean zero,

$$\int_0^T f(t') dt' = 0. \quad (8)$$

Manifestly, (7) and (8) imply that $g(t)$ is periodic with period T . Formally, we have, from (6)

$$g(t+T) - g(t) = \int_t^{t+T} f(t') dt' \quad (9)$$

$$= - \int_0^t f(t') dt' + \int_0^T f(t') dt' + \int_T^{t+T} f(t') dt' \quad (10)$$

$$= 0. \quad (11)$$

The final line follows because the second integral in (10) vanishes due to (8), and the third one cancels the first, as is readily proved by changing variables to $\tau = t' - T$ and employing (7).

From (4) and (11), we conclude that the gene frequency oscillates with period T . The simplest special case of $f(t)$ is

$$f(t) = \cos \omega t, \quad \omega = \frac{2\pi}{T}, \quad (12)$$

whence

$$g(t) = \frac{\sin \omega t}{\omega}. \quad (13)$$

If there is no dominance, $b = 0$, and

$$F(p) = \frac{1}{a} \ln \left(\frac{p}{1-p} \right). \quad (14)$$

The specific solution (13), (14) was derived (with a different phase convention) and discussed by Kimura (1955). Since (11) shows that $g(t)$ is always finite, (4) and (5) imply that the usual time-independent equilibrium point $p = -a/b$ is never reached. The gene frequency is stationary only where $f(t) = 0$. Finally, the continuous haploid model is merely a particular case of our general solution with $af(t) = m_1 - m_2$ and $b = 0$, m_1 and m_2 being the Malthusian parameters of the alleles A and a , respectively.

3. THE DISCRETE CASE

(i) *Haploid population*

We shall let w_1 and w_2 denote the fitnesses of the alleles A and a , and permit them to depend on a discrete time variable, $t = 0, 1, 2, \dots$. Then

$$p' = w_1 p / \bar{w}, \quad (15)$$

where

$$\bar{w} = w_1 p + w_2 q \quad (16)$$

is the mean fitness of the population and the prime refers to generation $t+1$. Dempster (1955) solved the recursion relation (15). From his result, with a slight rearrangement, we obtain

$$p(t) = \frac{p(0)}{p(0) + k(t)q(0)}, \quad (17)$$

where

$$k(t) = \prod_{n=0}^{t-1} \frac{w_2(n)}{w_1(n)}. \quad (18)$$

Owing to (17) and (18), for any haploid model in which the fitnesses of the two alleles are interchanged every generation, the gene frequency will return to its original value every second generation.

(ii) *Diploid organisms*

Random mating, in contrast to the continuous case, will lead to exact Hardy-Weinberg proportions provided the fertility of any union is equal to a product of factors associated with each of the two genotypes (Bodmer, 1965).

For any number of alleles A_i , with frequency p_i , if the genotype $A_i A_j$ has a fitness expressible in the form

$$w_{ij} = c_i c_j, \quad (19)$$

for some set of numbers, c_i , then the mean fitness of the allele A_i is

$$w_i = \sum_j w_{ij} p_j = \bar{c} c_i, \quad (20)$$

with

$$\bar{c} = \sum_i c_i p_i. \quad (21)$$

The mean fitness, therefore, is given by

$$\bar{w} = \sum_i w_i p_i = \bar{c}^2. \quad (22)$$

We may conclude immediately that the scheme (19) is equivalent to a haploid model since

$$p'_i = w_i p_i / \bar{w} = c_i p_i / \bar{c}, \quad (23)$$

in which the c_i function as allelic fitness.

Let us consider now the two-allele discrete case analogous to (1). In one generation, the gene frequency changes by (Crow and Kimura, 1970, p. 182).

$$\Delta p = p(1-p)[(w_{12} - w_{22}) + (w_{11} - 2w_{12} + w_{22})p] / \bar{w}, \quad (24)$$

where the mean fitness is

$$\bar{w} = w_{11}p^2 + 2w_{12}pq + w_{22}q^2. \quad (25)$$

Kimura (1955) examined the problem assuming the absence of dominance and setting

$$w_{12} - w_{22} = (-1)^t s, \quad s > 0. \quad (26)$$

To derive his Eq. (36),

$$\Delta p = (-1)^t s p(1-p), \quad (27)$$

it is also necessary to require

$$\bar{w} = 1. \quad (28)$$

Kimura calculated the change in p in two generations. We reproduce his result in slightly more general form, correcting two misprints:

$$\Delta^2 p = s^2 p(1-p)[(2p-1) + s(-1)^t p(1-p)]. \quad (29)$$

For even t , the right-hand side of (29) is positive or negative according as $p > p_0$ or $p < p_0$, where

$$p_0 = [(2+s) - (4+s^2)^{1/2}]/(2s) \quad (30)$$

$$\sim 1/2 - s/8 + O(s^2), \quad \text{as } s \rightarrow 0. \quad (31)$$

Therefore, we conclude with Kimura that if the initial frequency of A is less than p_0 , A is eventually lost, while if it is greater, A approaches fixation.

The condition (28) signifies that Kimura's example only applies to frequency-dependent selection. It cannot be interpreted as an approximate analysis of selection depending on time only because (29) is of the second order in selection (due to the sign reversals), and hence (24) must be expanded at least to $O(s^2)$. The difference is of biological importance since an exact calculation, employing (24), (25), and (26), with $w_{12} = 1$ and still postulating no dominance, yields, instead of (29),

$$\Delta^2 p = s^2 p(1-p)\{(1-2p) + s(-1)^t[3p(1-p) - 1]\}/\{[\bar{w}(t)]^2 \bar{w}(t+1)\}. \quad (32)$$

Now $\Delta^2 p(2t) > 0$ if $p < p_1(s)$ and $\Delta^2 p(2t) < 0$ if $p > p_1(s)$, where

$$p_1(s) = [(3s-2) + (4-3s^2)^{1/2}]/(6s) \quad (33)$$

$$\sim 1/2 - s/8 + O(s^2), \quad \text{as } s \rightarrow 0. \quad (34)$$

Furthermore, $\Delta^2 p(2t+1) > 0$ if $p < p_1(-s)$ and $\Delta^2 p(2t+1) < 0$ if $p > p_1(-s)$. We can verify easily that $p_1(-s) > 1/2 > p_1(s)$. Consequently, p oscillates around $1/2$. Dempster (1955) reached this conclusion by the following elegant semi-intuitive argument. For either homozygote, the product of the fitnesses for two consecutive generations is $(1-s^2)$, while for the heterozygote the corresponding quantity is unity. Therefore, disregarding single-generation fluctuations, we expect the usual symmetric overdominant stable equilibrium at $p = 1/2$. Since the gene frequency changes every generation, this reasoning is not completely rigorous, and the examination of the more general problem of arbitrary dominance will disclose the extent of its validity. From our detailed treatment, it follows that at equilibrium alternately $p = p_1(s)$ and $p = p_1(-s)$.

We shall now study (24) in the presence of dominance by writing

$$w_{11} = 1 + (-1)^t s, \quad (35)$$

$$w_{12} = 1 + (-1)^t h s, \quad (36)$$

$$w_{22} = 1 - (-1)^t s, \quad (37)$$

with $0 < s \leq 1$ and $0 \leq h \leq 1/s$. As for the continuous scheme, (2), (3), the degree of dominance,

$$\frac{2w_{12} - w_{11} - w_{22}}{w_{12} - w_{22}} = \frac{2h}{1+h},$$

is constant. Observe that this selection pattern is *not* symmetric under interchange of the two alleles and replacement of t by $t+1$ unless $h=0$. The fitness of AA always deviates from unity in the same direction as that of Aa . Thus, one must not expect the analysis to produce symmetrical results. In order to simplify the calculation, we shall confine ourselves to fairly weak selection: $s, hs \ll 1$. By direct computation, we obtain

$$\Delta^2 p \sim s^2 p(1-p)[1+h(1-2p)][(1-2p)-h(1-2p+2p^2)] + O(s^3). \quad (38)$$

All the results below follow from an elementary analysis of the sign of the right-hand side of (38). The zeros of this function in the interval $0 \leq p \leq 1$ are

$$p_2 = [(h-1) + (1-h^2)^{1/2}]/(2h), \quad 0 \leq h \leq 1 \quad (39)$$

$$p_3 = (1+h)/(2h), \quad h \geq 1. \quad (40)$$

Interestingly, p_3 is just the overdominant equilibrium one would deduce from our selection pattern by naive application of the standard formula (see, e.g. Crow and Kimura, 1970, p. 270). We observe that p_2 decreases monotonically from $1/2$ to 0 as h increases from 0 to 1 , and p_3 decreases monotonically from 1 to $1/2$ as h increases from 1 to ∞ . The behaviour of the system depends on the value of h :

(a) $0 \leq h < 1$ (no dominance or incomplete dominance)

The gene frequency p oscillates around p_2 . If $h=0$, there is no dominance and $p_2 = 1/2$, in agreement with the leading term in (34).

(b) $h = 1$ (complete dominance)

The dominant allele, A , is lost, as shown previously by Dempster's (1955) exact calculation.

(c) $h > 1$ (overdominance)

If the initial frequency of A is less than p , A is lost; if it is greater, A is fixed. From (35), (36), and (37) it is evident that the limit $s \rightarrow 0$, $h \rightarrow \infty$, $hs \rightarrow k$ (a constant) yields a fluctuating heterozygote selection coefficient $(-1)^t k$ and equally selected homozygotes. In this symmetric case, $p_3 = 1/2$.

These conclusions are exhibited in fig. 1.

It is very instructive to apply Dempster's biological argument here. The fitness products for the homozygotes are $(1-s^2)$, as above, but for the heterozygote we have now $(1-h^2 s^2)$. Manifestly, with complete dominance ($h=1$), the situation is too delicate to be analysed in this manner. With

incomplete dominance ($h < 1$) or overdominance ($h > 1$), we arrive at the correct qualitative result regarding fixation, loss, or polymorphism. The critical gene frequency, however, is not $1/2$, which corresponds to the symmetrical selection scheme $(1-s^2)$, $(1-h^2s^2)$, $(1-s^2)$ for two generations, but rather p_2 or p_3 . The discrepancy is due to the change in gene frequency from one generation to the next.

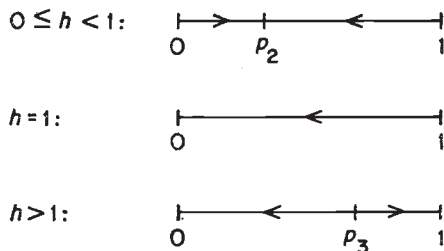


FIG. 1.—The direction of dominant gene frequency change in fluctuating environments. The symbols are defined in the text.

4. DISCUSSION

We have examined the effects of cyclic variations in selection coefficients on gene frequency. For a continuous model, if the selection coefficients have mean zero, the gene frequency follows the environmental oscillation with the same period. Fitnesses in natural populations, however, will never be exactly periodic with mean zero. If the deviations from this assumption predominantly favour one allele, the other will be reduced to a low frequency in a time of the order of the inverse of the selection coefficients. If the deviations generally cancel, our results imply that polymorphism will be maintained for times orders of magnitude longer than the above.

For a discrete model with no dominance, we have shown that Kimura's (1955) model, which leads to an homozygous population, ought to be interpreted as frequency-dependent selection. An exact analysis of the corresponding model with selection coefficients which depend on time only yields a polymorphic population in which the equilibrium gene frequency alternates between the values $p_1(s)$ and $p_1(-s)$, $p_1(-s) > 1/2 > p_1(s)$, given by (33).

The discrete model with an arbitrary degree of dominance specified by (35), (36), and (37) has the interesting feature that partial or no dominance in a single generation produces overdominance for two consecutive generations. The frequency of the dominant at the expected stable equilibrium is p_2 of (39), and not $1/2$, which one would obtain from the fitness products for consecutive generations by ignoring the change in gene frequency between these generations. If a polymorphism is found for which the cyclic model is plausible, one could estimate the degree of dominance from the value of p_2 . With complete dominance, the recessive allele is eventually established in the population. As (38) shows, weak constant selection against the recessive, of the second order in the strength of the oscillating selection, would result in a stable equilibrium. For numerical examples, the reader should refer to the paper of Dempster (1955), where this problem was first considered. If the single-generation selection scheme is overdominant, the consecutive-generation fitness products lead us to expect an unstable equilibrium at $p = 1/2$.

This equilibrium does exist, but at $p = p_3$, given by (40). The reason for the discrepancy is again the gene frequency shift between generations. Some other selective agency would have to intervene to produce a balanced polymorphism.

The main qualitative conclusion of this paper is that in a continuous model with random births and deaths and regularly oscillating selection coefficients, polymorphism is maintained with the gene frequency following the environmental cycles through values depending on the initial gene frequency, while if generations are discrete and non-overlapping, cyclically-varying selection coefficients lead to stable or unstable equilibrium points independent of the initial gene frequency. The reason for the very different behaviour of the two models is that in the first case the gene frequency can "track" the environment continuously while in the second each environmental cycle operates on the population after a discrete change in gene frequency. The continuous time model applies accurately to some lower organisms such as bacteria, whereas the discrete, non-overlapping generations scheme represents well populations like annual plants. If a population with overlapping generations reproduces continuously, as do human beings, for example, we would expect the first model to be a better approximation. If generations overlap but there is a definite breeding season, as in many mammals, gene frequencies will change discretely, but in closer phase with the environment than in the second model. For detailed predictions, one must know the nature of the environmental fluctuation as well as the demographic structure of the population.

Acknowledgment.—It is a pleasure to thank Professor J. F. Crow for several points of interpretation and interesting discussions.

5. REFERENCES

- ALLARD, R. W., AND WORKMAN, P. L. 1963. Population studies in predominantly self-pollinated species. IV. Seasonal fluctuations in estimated values of genetic parameters in lima bean populations. *Evolution*, 17, 470-480.
- BODMER, W. F. 1965. Differential fertility in population genetics models. *Genetics*, 51, 411-424.
- CROW, J. F., AND KIMURA, M. 1970. *An Introduction to Population Genetics Theory*. Harper and Row, New York.
- DEMPSTER, E. R. 1955. Maintenance of genetic heterogeneity. *Cold Spring Harbor Symp. Quant. Biol.*, 20, 25-32.
- DOBZHANSKY, TH. 1971. Evolutionary oscillations in *Drosophila pseudoobscura*. In: *Ecological Genetics and Evolution* (ed. R. Creed). Blackwell, Oxford.
- HALDANE, J. B. S., AND JAYAKAR, S. D. 1963. Polymorphism due to selection of varying direction. *Jour. Genet.*, 58, 237-242.
- HARTL, D. L., AND COOK, R. D. 1973. Balanced polymorphisms of quasi-neutral alleles. *Theor. Pop. Biol.*, 4, 163-172.
- KIMURA, M. 1955. Stochastic processes and distribution of gene frequencies under natural selection. *Cold Spring Harbor Symp. Quant. Biol.*, 20, 33-53.
- KIMURA, M., AND OHTA, T. 1971. *Theoretical Aspects of Population Genetics*. Princeton University Press, Princeton.
- LEVENE, H. 1967. Genetic diversity and diversity of environment: Mathematical aspects. *Proc. Fifth Berkeley Symp. Math. Stat. Prob.*, 4, 305-316.
- NAGYLAKI, T., AND CROW, J. F. 1974. Continuous selective models. *Theor. Pop. Biol.*, 5, 257-283.
- POWELL, J. R. 1971. Genetic polymorphisms in varied environments. *Science*, 174, 1035-1036.
- WRIGHT, S. 1969. *Evolution and the Genetics of Populations*, vol. 2. The University of Chicago Press, Chicago.