

## THE DYNAMICS OF HYBRID ZONES

N. H. BARTON

*School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, U.K.*

Received 14.iii.79

### SUMMARY

This paper investigates the dynamic behaviour of hybrid zones which are maintained by a balance between dispersal and selection against hybrids. In the first section it is shown that a hybrid zone involving a single locus can move in response to a selective imbalance between the two homozygotes, and also to variation in population density and dispersal rate. It can be trapped by natural barriers, and so an allele which is selected against when rare cannot advance, even if it is advantageous when common. The continuous model used in deriving these results is shown to be a good approximation to the stepping-stone model, provided that the cline contains several demes.

The effect of stochastic forces on multi-locus hybrid zones is then considered. An expression giving the shift in position after an arbitrary perturbation in gamete frequency is derived. Using this formula, it is found that sampling drift is negligible unless the zone includes few organisms and involves few loci. Random variations in population structure are the dominant force, and could allow considerable movement in an even environment. However, natural barriers can still trap hybrid zones, and so it is likely that they will remain roughly where they first formed.

### 1. INTRODUCTION

In recent years, many examples of parapatrically distributed taxa have come to light. Two distinct types, often differing at many loci, abut and interbreed in a narrow "hybrid zone".<sup>1</sup> These zones are of particular interest, since they reveal the interaction between divergent genotypes, and may have some bearing on mechanisms of speciation. Indeed, it has been proposed that they play a central role as barriers to gene flow between speciating groups (White, 1968, 1978).

There are a number of possible explanations for such narrow clines. They might be due simply to a recent mixing of two selectively identical types. There might be a sharp ecotone along the cline, so that different parental types are favoured on opposite sides, or so that hybrids are favoured in a narrow band (Moore, 1977). However, one of the simplest and most obvious explanations is that hybrids are at a disadvantage—in the simplest case, heterozygotes are less fit. They might be inferior in viability, in fertility, in fecundity, or in mating success. Organisms which cross the zone, into an area where their own type is rare, will produce a greater proportion of hybrid offspring than does the more common type, and so will be selected against. In other words, when two populations, each at different, stable, equilibria meet, a narrow zone of intermediates is expected to form (cf. Karlin and McGregor, 1972). Since any complex genetic system should have many stable equilibria, such sharp zones might be expected to occur quite often, provided that the system changes so slowly that it is usually near equilibrium,

<sup>1</sup> The term "hybrid zone" will be used to denote clines maintained by hybrid unfitness, regardless of their origin.

and that there is some mechanism for shifting a substantial region from one equilibrium to another.

In this paper, the movement of simple, single locus zones under deterministic forces will be considered first; the effects of various stochastic forces on more general, multi-locus zones are then examined.

## 2. THE BASIC MODEL

To quantify the balance between dispersal and selection against hybrids, either the stepping-stone or the continuous model of population structure may be used. The former consists of a set of discrete demes, each exchanging members only with nearest neighbours, and having discrete, non-overlapping generations. The latter assumes a continuous, dense distribution of organisms, each of which has a certain probability of leaving offspring any given distance away, per unit time.

The second model is much easier to analyse and is more amenable to generalisation than the first. Therefore, although organisms may often be distributed in discrete demes, and reproduce in discrete generations, the continuous approximation is used; it will be valid where gene frequencies change little between demes and between generations, as is likely if selection is weak.

Second-order selection terms have been neglected, so the theory presented here is less useful where selection is strong (for example, in pest control using translocations). With these restrictions, it is possible to find the effect of both deterministic and stochastic factors on the position of the zone.

### (i) Notation

$x, y$	position
$t$	time
$p, q$	allele frequencies ( $p + q = 1$ )
$s$	fitness deficit of hybrids
$m$	variance in progeny position per generation
$\rho$	population density
$f' \equiv \frac{df}{dx}, f'' \equiv \frac{d^2f}{dx^2}$ , etc.	
$f \equiv \frac{df}{dt}$	
$w \equiv \sqrt{\frac{2m}{s}}$	typical zone width

### (ii) Selection

Suppose heterozygotes have fitness  $1-s$ , compared with homozygotes both with unit fitness. If  $s \ll 1$ , the population will be in approximate Hardy-Weinberg equilibrium, and changes will be almost continuous in time. So,

$$\Delta p = \left( \frac{p^2 + pq(1-s)}{1-2pqs} \right) - p \quad \therefore \dot{p} \approx spq(p-q).$$

(iii) *Dispersal*

An equation expressing the effect of dispersal may be derived either as the limit of the stepping-stone model, with very close demes, or, as in Nagylaki (1975), from the distribution of progeny at  $(y, t + dt)$ , from parents at  $(x, t)$ . Either approach gives, in one dimension,

$$\dot{p} = \frac{m}{2} p'' + m' p' + m \left( \frac{\rho'}{\rho} \right) p.$$

In the stepping stone model,  $m$  is just  $M\varepsilon^2$ , where a fraction  $M$  is exchanged between demes  $\varepsilon$  apart each generation, and  $\rho$  is  $N/\varepsilon$ , where  $N$  is the number of breeding individuals in each deme. Using Nagylaki's formulae, and assuming that dispersal is equally likely in all directions,  $m$  is the variance in progeny position which is produced per unit time. (No generality is lost in assuming that dispersal is isotropic, since the effect of anisotropy would be the same as that of a density gradient.)

Provided that the organism does not disperse too far (the number of progeny left at large distances must fall off at least exponentially), only  $m$ , the variance of the distribution, need be known to describe gene flow.

(iv) *An alternative representation*

Combining the effects of selection and dispersal, a single-locus hybrid zone, in one dimension, can be described by:

$$\dot{p} = \frac{m}{2} p'' + m' p' + m \left( \frac{\rho'}{\rho} \right) p + spq(p - q).$$

It is sometimes more useful not to solve this equation for  $p$  directly. Instead, given that  $p$  obeys the above equation, we can look for a function of  $p$  which never increases, so that it is minimised at equilibrium.

Consider, therefore,

$$H \equiv \int_{-\infty}^{\infty} \rho^2 [m^2 p'^2 + 2msp^2 q^2] dx.$$

Then

$$\dot{H} = - \int_{-\infty}^{\infty} 4m\rho^2 \dot{p}^2 dx$$

is always negative unless  $\dot{p} = 0$  everywhere.

$H$  therefore tends to a minimum.

(v) *Equilibrium solution in one dimension* (see also Bazykin, 1969)

Using the above, if  $m, s, \rho$  are constant, the solution at equilibrium (other than fixation of one or other allele) is:

$$p = \frac{1}{2} \left( 1 + \tanh \left( \sqrt{\frac{s}{2m}} (x - x_0) \right) \right),$$

where  $x_0$  is the (arbitrary) centre of the cline.

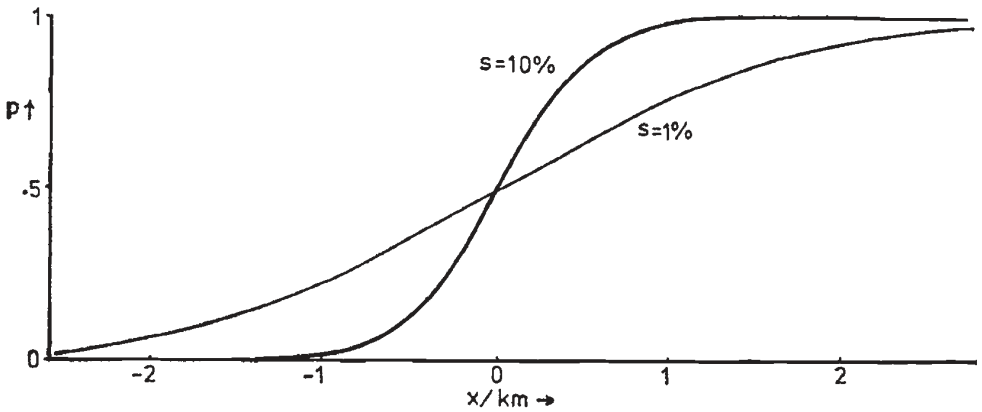


FIG. 1.—Hybrid zones produced by selection against heterozygotes; dispersal  $\sqrt{m} = 112 \text{ m.gen.}^{-\frac{1}{2}}$

Thus, there is a transition from one allele to the other over a distance of about  $w = \sqrt{\frac{2m}{s}}$ ; as selection increases and dispersal decreases, the hybrid zone narrows, as would be expected intuitively. This solution will still hold, approximately, if dispersal, selection against hybrids, and population density change little within the zone. However, since  $H$ , with the above solution, is  $\frac{m\rho^2\sqrt{2ms}}{3}$ , there will be a tendency for the zone to move towards local regions

of low dispersal, low hybrid unfitness, and low population density; obviously if homozygotes are not equally fit, it will tend to move in favour of the fitter allele.

It should be noted that it has been assumed throughout that dispersal, selection and population density are independent of allele frequency; this is not necessarily so. One type might be more mobile; selection might be frequency dependent; or, most likely, infertility or inviability of hybrids might reduce population density at the zone. In the extreme “hard” selection case, where there is no competition and so selective deaths cause a proportionate reduction in population size,  $\rho = \rho_0(1 - 2spq/r)$ , where  $r$  is the intrinsic rate of increase. This reduction in density would tend to sharpen the zone somewhat. The effect of genotype frequencies on population size is complicated, and so it is reasonable to use the simple assumption of independence as a first approximation.

### 3. MOVEMENT OF THE ZONE

We have seen that a cline maintained by heterozygote inferiority will move in response to a number of factors, whilst maintaining its integrity. How fast is this movement, and which factors are likely to be most important?

#### (i) Selective advantage of one allele

Let the fitnesses of  $PP, PQ, QQ$  be  $1 + 2S, 1 - s + S, 1$ ; let  $S = \alpha s$ ; ( $S, s \ll 1$ ). Then:

$\dot{p} = \frac{m}{2} p'' + spq(p - q) + \alpha spq$ , which has a solution:

$$p = \frac{1}{2} \left( 1 + \tanh \left( \left( x + \frac{\alpha t}{2} \right) \sqrt{\frac{s}{2m}} \right) \right).$$

Thus, the zone has the same width as before  $\left( \sqrt{\frac{2m}{s}} \right)$ , but moves at a speed  $\frac{S}{2} \sqrt{\frac{2m}{s}}$  towards the less fit type. For example, if the typical dispersal distance,  $\sqrt{m}$ , is 100 m gen<sup>-1/2</sup>, and selection against heterozygotes,  $s$ , is 10 per cent, the zone will be about 500 m across. If one homozygote has a 1 per cent superiority, it will move at 25 m per generation. As  $s \rightarrow 0$ , this tends to Fisher's equation for the wave of advance of an advantageous gene; the above solution ceases to be stable for weak underdominance, and is replaced by an asymmetric wavefront moving at  $c = \sqrt{2m(S-s)}$  when  $S > 2s$  (Fisher 1937, Stokes 1976).

(ii) *Density variations*

If population density is not constant, the zone will move into regions of low density. Such effects may be quite strong, since local variations in population size may be large. (Fine scale changes in density *within* the zone, will, however, not be important, even if they are quite large. See later.)

Since  $H \approx \frac{m\rho^2\sqrt{2ms}}{3}$  for smooth changes in  $\rho$ , and

$$\dot{H} = - \int_{-\infty}^{\infty} 4m\rho^2 \dot{p}^2 dx,$$

the speed may be calculated using  $\frac{\partial H}{\partial t} = \frac{dy}{dt} \frac{\partial H}{\partial y}$ , where  $y$  is the zone position.

Letting

$$p = \frac{1}{2} \left( 1 + \tanh \left( (x - y) \sqrt{\frac{s}{2m}} \right) \right), \quad \dot{H} = -\frac{2}{3} y^2 \rho^2 \sqrt{2ms},$$

$$\frac{\partial H}{\partial y} = \frac{2}{3} m \rho \rho' \sqrt{2ms}.$$

Hence,  $\dot{y} = -m \left( \frac{\rho'}{\rho} \right)$ , with *density gradient*  $\rho'$ .

The effect of changes in dispersal rate, or selection against hybrids, can be found by the same method.

$$\dot{y} = -\frac{3}{4} m', \quad \text{with } \textit{dispersal gradient } m',$$

$$\dot{y} = -\frac{m}{4} \left( \frac{s'}{s} \right), \quad \text{with } \textit{selection against hybrids varying at } s'.$$

What size of density gradient can prevent the spread of an allele which is advantageous in the homozygote? Including the effect of differences in homozygote fitness

$$H(y) = \int_{-\infty}^{\infty} m\rho^2 \left( m^2 p'^2 + 2msp^2 q^2 - \frac{2mSp^2}{3} (1+2q) \right) dx$$

$$\therefore \frac{\partial H}{\partial y} = \frac{m\rho^2 s}{3} \left( 2\sqrt{\frac{2m}{s}} \frac{\rho'}{\rho} - \frac{S}{s} \right), \text{ if } p \approx \frac{1}{2} \left( 1 + \tanh \left( \sqrt{\frac{s}{2m}} (x-y) \right) \right).$$

Thus, if the density gradient is greater than  $\frac{1}{2w} \left( \frac{S}{s} \right)$ , the fitter allele will prevail. An allele advantageous when homozygous, but not in the heterozygote, may not spread if it meets strong local changes in population size. (Note that if  $S > s$ , the allele is at an advantage even in the heterozygote, and so local density gradients will not prevent a few long distance migrants from spreading the allele. See Hadelér, 1976.)

For example, if there is a density doubling in 2 km, a zone, maintained by 10 per cent selection and with 100 m dispersal as before, will move down the gradient at 35 m per generation. A form with an advantage of less than 1.7 per cent will not be able to advance up the gradient, but will be pushed back.

(iii) *Movement in two dimensions*

So far, only a one-dimensional model has been used. This will apply to short, straight zones, but takes no account of the effect of curvature. One would expect "bulges" in the zone to be pushed back by weight of numbers, thus keeping the zone more or less straight, and eventually eliminating small isolates.

If  $m, s, \rho$  vary slowly,  $H \approx \int_c \frac{m\rho^2 \sqrt{2ms}}{3} dl$ , where the integral is taken round the hybrid zone. For example, consider a circular zone, of radius  $r$  much greater than the zone width. Then:

$$H = \frac{2\pi r}{3} m\rho^2 \sqrt{2ms}, \quad \dot{H} = -\frac{8\pi r}{3} \dot{r}^2 \rho^2 \sqrt{2ms} = \frac{\dot{r}H}{r}.$$

Hence,  $\dot{r} = -\frac{m}{4r}$ , if  $m, s, \rho$  are constant. A zone with radius  $r$  moves with speed  $m/4r$  in such a way as to reduce its radius. A circular zone will contract, slowly at first, until it disappears in about  $2r_0^2/m$  generations. For example, if there is a dispersal rate of 100 m per gen<sup>+1</sup>, and there is an isolate distributed as a disc 10 km across, it will be eliminated in about 5000 generations. Of course, very slight inhomogeneities in population structure, or slight selective advantages, can prevent this. If the form contained in the isolate had a 0.1 per cent advantage, and the zone were maintained by 10 per cent selection against heterozygotes, then it would advance unless it were initially in an area less than 5 km across, in which case it would disappear. This advance would probably soon be halted by variations in population density, however.

(iv) *Resistance to movement due to discrete population structure*

In reality, organisms are often not continuously distributed, but live in discrete demes. If the zone does not include many demes, the assumptions

previously used break down; even if there are several demes in the zone, one would expect a slight effect on its behaviour. The equilibrium will no longer be precisely neutral, and certain zone positions will be favoured. Thus, a definite amount of force will be needed to shift the zone, and it will not move unless the density gradient or selective advantage driving it forwards exceeds a certain critical value.

Let demes be spaced  $\epsilon$  apart, at  $x = n\epsilon$ ,  $-\infty < n < \infty$ . Then, the potential function  $H$  can still be used to describe the zone, but must be calculated as a sum, taken at demes, rather than a continuous integral.

$$H = \sum_i 2\epsilon m \rho^2 \left( \frac{m}{2\epsilon^2} (p_i - p_{i-1})^2 + s p_i^2 q_i^2 \right) \text{ will be minimised,}$$

if  $m$  and  $\rho$  are constant. If the effect of discrete population structure is small, then the allele frequency is still given, to sufficient accuracy, by

$$p = \frac{1}{2} \left( 1 + \tanh \left( (x-y) \sqrt{\frac{s}{2m}} \right) \right).$$

$H$  can now be calculated as a function of the zone position,  $y$ ; it will vary cyclically, and the zone will tend to rest at a local minimum. The critical force needed to move it from this equilibrium is that which will just counteract the maximum slope in  $H$ ,  $\left. \frac{\partial H}{\partial y} \right|_{\max}$ . Taking new coordinates

$$x' = x \sqrt{\frac{s}{2m}}, \quad y' = y \sqrt{\frac{s}{2m}}, \quad \epsilon' = \epsilon \sqrt{\frac{s}{2m}}$$

and substituting for  $p$

$$H(y') = \sum_{-\infty}^{\infty} \frac{\epsilon'}{4} m \rho^2 \sqrt{2ms} (1 - \tanh^2 (n\epsilon' + y'))^2$$

Taking the Fourier transform *w.r.t.*  $y$ ,

$$\begin{aligned} \tilde{H}(w) &= \sum_{n=-\infty}^{\infty} \frac{m \rho^2 \sqrt{2ms}}{4} \epsilon' \left( \frac{\pi w(4+w^2)}{3 \sinh(\pi w/2) \sqrt{8\pi}} \right) e^{-inw\epsilon'} \\ &= \sum_{j=-\infty}^{\infty} \frac{\pi}{2} m \rho^2 \sqrt{2ms} \left( \frac{w(4+w^2)}{3 \sinh(\pi w/2) \sqrt{8\pi}} \right) \delta \left( w - \frac{2j\pi}{\epsilon'} \right) \\ &\quad (j = \dots -1, 0, 1 \dots) \end{aligned}$$

( $\delta(x)$  is the Dirac delta function).

There are thus a series of components of frequency  $w = \frac{2j\pi}{\epsilon'}$ ; if  $\epsilon' \ll \pi^2$  (when the zone includes many demes), the higher frequency components become negligible, and so one need only deal with terms up to  $j = \pm 1$ . Taking the inverse Fourier transform,

$$\begin{aligned} H(y') &= m \rho^2 \sqrt{2ms} \left( \frac{1}{3} + \frac{4}{3} \frac{\pi^4}{\epsilon'^3} e^{-\pi^2/\epsilon'} \cos \left( \frac{2\pi y'}{\epsilon'} \right) \dots \right) \\ \therefore \left. \frac{\partial H}{\partial y'} \right|_{\max} &= m \rho^2 \sqrt{2ms} \left( \frac{8\pi^5}{3\epsilon'^4} e^{-\pi^2/\epsilon'} \right). \end{aligned}$$

Compare, for example, with  $\frac{\partial H}{\partial y} = \frac{Smp^2\sqrt{2ms}}{3s}$ , for a selective imbalance  $S$ .

Letting  $S^*$  be the critical selective difference between homozygotes needed to move the zone,

$$S^* = \frac{8\pi^5 s}{\epsilon'^4} e^{-\pi^2/\epsilon'}; \text{ this declines rapidly as } \epsilon' \text{ becomes } \ll 1.$$

$N$	1	2	3	10
$S^*$	1.2%	$10^{-6}$	$2 \times 10^{-9}$	$2 \times 10^{-37}$

( $N = 1/\epsilon'$ , the number of demes in the zone;  $s = 10$  per cent)

Thus, the effect of discrete population structure may be neglected when there are more than one or two demes in the zone; the effect drops off very rapidly if demes are more closely spaced than this. If there are fewer than one or two demes in the zone, it would be feasible to use the stepping-stone model directly; in considering zone movement, at least, the two approximations overlap.

#### 4. STOCHASTIC EFFECTS

A hybrid zone could move quite rapidly if, as is likely in nature, the two parental types were not equally fit. However, local barriers to dispersal, or reductions in population density, will prevent such persistent, directed movements, even if the forces driving the zone forwards are quite large. In practice, population structure is not constant from year to year, as has been assumed so far. Can random changes in population sizes, dispersal rates, or gene frequencies enable a zone to escape from local traps, and move large distances?

To proceed, one first determines the distance the zone will move in response to an arbitrary, small perturbation. If the covariance function\* of the perturbations is known (without necessarily knowing their complete distribution), then the variance in zone position produced per unit time can be calculated. Large perturbations cannot be dealt with, since their effects are not additive. This is not a serious restriction, at least for wide zones which include many demes. The relative importance of different perturbing factors can thus be assessed, and the conditions which will allow the zone to be mobile can be indicated.

##### (i) Response to a perturbation

We will first consider the stability of a hybrid zone to small fluctuations in allele frequency. This section will deal with the more general case of a hybrid zone maintained by interactions between several loci; the previous deterministic results were derived from a study of a simple single-locus zone, but they can easily be extended to the multi-locus case. The effects of population structure are precisely the same, whereas the effects of asym-

\* The covariance function of a stochastic process is the covariance between the values of the process at two given points in space and time;

$$r(x, x' : t, t') = \langle f(x, t) f(x', t') \rangle$$

(see Soong, 1973).



metric selection coefficients are qualitatively similar. The stability of a zone will only be calculated explicitly in certain simple cases—a single locus, weakly interacting loci, or loci having clines of similar shape. For these examples, we shall find the rate of movement of a zone due to sampling drift, or to variations in population structure.

Let allele frequencies (or, if linkage is important, gamete frequencies) be denoted by the vector  $\theta$ , or by its elements  $\theta_i$ . Let selection be weak, and be defined by  $\dot{\theta} = s(\theta)$ .

If the simple system  $\dot{\theta} = s(\theta)$  has two or more stable equilibria, then a hybrid zone may be possible, for some value of density gradient just sufficient to counteract any selective imbalance (it may not be possible to form a stable hybrid zone, however; for example, a recombinant formed in the zone may prove superior to all other types).

From the basic equation,

$$\dot{\theta}_\varepsilon = \frac{m}{2} \theta''_\varepsilon + s(\theta_\varepsilon) = 0$$

at equilibrium, taking population density and dispersal rate constant for simplicity. Now consider small perturbations,  $\gamma_i$ , from this neutral equilibrium,  $\theta_\varepsilon$ ; let  $\theta_i = \theta_{ei} + \gamma_i \theta'_{ei}$ . A constant  $\gamma_i(x) = D \forall i$  corresponds to a shift in zone position of  $-D$ .

Transforming to  $\gamma$ , the linearised equation around the equilibrium is (in matrix notation):

$$\dot{\gamma} = \frac{m}{2} \gamma'' + mT\gamma' + A\gamma,$$

$$\text{where } T_{ij} = \frac{\theta''_i}{\theta'_i} (i = j), = 0 (i \neq j)$$

$$A_{ij} = \frac{\partial s_i}{\partial \theta_j} \frac{\theta'_j}{\theta'_i} (i \neq j), = - \sum_{k, i \neq k} A_{ik} (i = j)$$

( $T$  is a diagonal matrix depending on cline shape, whilst  $A$  represents the interactions between loci).

Any initial perturbation of a stable zone will result in a constant final shift of  $\gamma_i(x) = -D$ . Hence, if some integral of the form  $I = \int_{-\infty}^{\infty} g \cdot \gamma dx$  can be found which is invariant during zone movement, the final shift in zone position will be given by:

$$D = - \int_{-\infty}^{\infty} g \cdot \gamma dx / \int_{-\infty}^{\infty} \sum_i g_i dx$$

since, eventually,  $I = - \int_{-\infty}^{\infty} \sum_i g_i(x) D dx$ .

Now, integrating by parts,

$$I = \int_{-\infty}^{\infty} g \cdot \dot{\gamma} dx = \int_{-\infty}^{\infty} \left( \frac{m}{2} g'' - m(gT)' + gA \right) dx.$$

So, for  $I$  to be invariant for all possible  $\gamma$ :

$$\frac{m}{2} \underline{g}'' - m(\underline{g}T)' + \underline{g}A = 0.$$

(ii) *Solution for  $g$  in particular cases*

(a) *A single locus*: If only one locus is involved, then  $g_i = \theta_i'^2$  is the solution, since  $A = 0$ . When several loci interact to produce a hybrid zone, it is hard to determine the stability except in especially simple cases.

(b) *Clines of similar shape*: If

$$\theta_i = \theta_j \forall i, j, \quad \text{and} \quad \sum_j \left( \frac{\partial s_i}{\partial \theta_j} - \frac{\partial s_j}{\partial \theta_i} \right) = 0 \forall i,$$

(which will hold for most selection schemes) then the  $gA$  term cancels and the solution is again  $g_i = \theta_i'^2$ . For example, suppose there are clines at many loci, maintained to varying degrees by selection against heterozygotes,  $S_i$ , and by epistasis between loci,  $E_{ij}$ . As long as  $S_i + E_{ij}$  is constant for all loci, linkage is loose, and  $E_{ij} = E_{ji}$ , the above conditions are satisfied. In this case, the way in which each cline is maintained does not affect the contribution it makes to zone movement.

(c) *Epistasis between two linked loci*: Suppose that a hybrid zone is maintained by epistatic selection between two loci (as in Bazykin, 1973). The two clines will have the same shape (by symmetry), and so if linkage is loose, the above conditions are satisfied, and  $g_1 = g_2 = \theta_1'^2$ . However, if recombination is comparable to selection, disequilibrium becomes important, and a third variable is needed to describe the system. Let  $\theta_1, \theta_2$  be the frequencies of alleles  $P_1, P_2$  respectively, and let  $\theta_3$  be the sum of the frequencies of gametes  $P_1P_2$  and  $Q_1Q_2$ . Then, with epistasis  $\varepsilon$ , and recombination  $r$ :

$$\left. \begin{aligned} \dot{\theta}_1 &= \frac{m}{2} \theta_1'' + \frac{\varepsilon}{2} (2\theta_1 - 1)(1 - \theta_3) \\ \dot{\theta}_2 &= \frac{m}{2} \theta_2'' + \frac{\varepsilon}{2} (2\theta_2 - 1)(1 - \theta_3) \\ \dot{\theta}_3 &= \frac{m}{2} \theta_3'' + \varepsilon(1 - \theta_3)\theta_3 + r[\theta_1\theta_2 + (1 - \theta_1)(1 - \theta_2) - \theta_3] \end{aligned} \right\}.$$

By symmetry,  $g_1 = g_2$ . Hence, the equation for  $g$  is:

$$\left. \begin{aligned} \frac{m}{2} g_1'' - m \left( g_1 \frac{\theta_1''}{\theta_1'} \right)' + (2\theta_1 - 1) \left( r g_3 \frac{\theta_1'}{\theta_3'} + \frac{\varepsilon \theta_1'}{2\theta_1'} 3g_1 \right) &= 0 \\ \frac{m}{2} g_3'' - m \left( g_3 \frac{\theta_3''}{\theta_3'} \right)' - 2(2\theta_1 - 1) \left( r g_3 \frac{\theta_1'}{\theta_3'} + \frac{\varepsilon \theta_1'}{2\theta_1'} 3g \right) &= 0 \end{aligned} \right\}.$$

So, whatever the value of  $r$ ,

$$g_3 = -\frac{\varepsilon}{2r} \theta_1'^2, \quad g_1 = g_2 = \theta_1'^2$$

is a solution. The equation for  $g$  is only as simple as this when the zone is maintained purely by epistasis; the introduction of selection against heterozygotes renders it insoluble.

(iii) *Generalisation to two dimensions*

In one dimension, a quantity  $I$  can be found which is invariant during the evolution of small perturbations, and so defines the final displacement of the zone. If the zone is very long, however, the treatment must be extended to two dimensions. A small perturbation,  $\gamma(x, y)$  will rapidly lead to local shifts in zone position, which will vary in magnitude along the zone. Since curvature is minimised, these local deviations will eventually be smoothed out to give a uniform final displacement.

Defining  $I(y, t)$  as before, as  $\int_{-\infty}^{\infty} \sum_i \gamma_i(x, y) g_i dx$ , one finds that (for a fairly straight zone)  $\dot{I} = \frac{m}{2} \frac{\partial^2 I}{\partial y^2}$ . One can now find the final shift due to continual small perturbations  $\gamma(x, y, t)$  by using the Green's function for diffusion and integrating over time.

$$I(y, t) = \int_0^t \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\exp [-(y-z)^2/2mt']}{\sqrt{2\pi mt'}} \gamma(x, z, t') \cdot g(x) dx dz dt'$$

The local shift in position,  $D(y, t)$ , will, disregarding the effect of recent disturbances, be proportional to  $I(y, t)$ .

$$D(y, t) = -I(y, t) / \left( \int_{-\infty}^{\infty} \sum_i g_i(x) dx \right)$$

(iv) *Displacement due to random perturbations*

We now have a formula which gives the displacement,  $D$ , of the zone after some small change in allele frequency from its equilibrium value. The covariance function of  $D$ ,  $\lambda$

$$r_D(\Delta y, t, \Delta t) = \langle D(y, t) D(y + \Delta y, t + \Delta t) \rangle; \quad r_D(0, t, 0) = \langle \dot{D}^2 \rangle,$$

(the rate of drift of the zone), can now be found if the covariance function of the disturbance,  $r_{\gamma\gamma'}(t, x, \Delta x, \Delta y, \Delta t)$ , is known. This latter function describes the variance in allele frequency produced by the perturbations  $\gamma$ , and also their correlation over time and space. The zone may be disturbed by sampling drift, by variations in population structure, or by varying selection pressure. The first two possibilities will be considered below.

(a) *Sampling drift*: If there are a small number of individuals in the zone, there will be some drift in allele frequency from generation to generation. Consider a small area  $\delta A$ . Inside this region there will be an effective population of  $\rho_E \delta A$ , where  $\rho_E$  is the effective population density (Crow and Kimura, 1970). Thus, the variance in allele frequency  $i$  produced per generation is  $\theta_i(1 - \theta_i)/2\rho_E \delta A$ .\* Fluctuations are not correlated between

\* There will also be a contribution to the drift due to variation gene flow between demes; this is of order  $mb^2/\rho$ , however, which is negligible for weak selection (see Nagylaki, 1978).

areas or between generations, so that as  $\delta A$  tends to zero, and many generations elapse,  $r_{\theta_i\theta_j}$  tends to  $(\theta_i(1-\theta_i)/2\rho_E)\delta(\Delta x, \Delta y, \Delta t)$ .  $\delta(\Delta x, \Delta y, \Delta t)$  is the Dirac delta function, which is zero except for a unit singularity at the origin. If loci are loosely coupled, fluctuations will not be correlated between loci, and  $r_{\theta_i\theta_j} = 0$  for  $i \neq j$ . However, if disequilibrium is important, as may often be the case, the covariance function will have the form  $R_{ij}\delta(\Delta x, \Delta y, \Delta t)$ ;  $R$  is some matrix function of  $\theta$ , and hence of  $x$ .

So, we now know the effect of sampling drift on allele frequencies. The effect on  $\gamma$  is given by the relation:

$$r_{\gamma_i\gamma_j} = r_{\theta_i\theta_j}/\theta'_i\theta'_j.$$

Taking the expectation of  $D(y, t)D(y + \Delta y, t + \Delta t)$ , one finds that:

$$r_D(\Delta y, t, \Delta t) = \int_0^{t+\Delta t} \int_0^t \iiint_{-\infty}^{\infty} \frac{\exp [-(\Delta y - \Delta z)^2/2m(t_1+t_2)]}{\sqrt{2\pi m(t_1+t_2)}} \times \left\{ \frac{\sum_{i,j} g_i(x)r_{\gamma_i\gamma_j}(x, \Delta x, \Delta z, \Delta t)g_j(x+\Delta x)dx d\Delta x dz dt_1 dt_2}{\left(\int_{-\infty}^{\infty} \sum_i g_i(x)dx\right)^2} \right\}$$

Substituting  $r_{\gamma_i\gamma_j}(x, \Delta x, \Delta z, \Delta t) = R(x)\delta(\Delta x, \Delta z, \Delta t)$  and integrating,

$$r_D(\Delta y, t, \Delta t) = \frac{\int_0^t \int_{-\infty}^{\infty} \frac{\exp(-\Delta y^2/4mt')g_i^T(x)R(x)g_j(x)dx dt'}{\sqrt{4\pi mt'}}}{\left(\int_{-\infty}^{\infty} \sum g_i(x)dx\right)^2} \quad (\Delta t > 0).$$

Several general questions can now be asked. Firstly, what sort of shape will the zone have? Selection against hybrids will tend to straighten the zone, whilst perturbations at different places will tend to move and bend it. A series of superimposed bulges will develop, of typical length  $\sqrt{2mt}$ , and size

$$\left(\frac{t}{m\pi} \left(\int_{-\infty}^{\infty} \underline{g}^T R \underline{g} dx\right)^2\right)^{\frac{1}{4}} / \left(\int_{-\infty}^{\infty} \sum_i g_i dx\right).$$

Since  $g$  is zero outside the hybrid zone, and  $R$  is proportional to  $w^2/\rho_E$ , where  $w$  is the zone width, the zone movement is proportional to  $\left(\frac{w^2 t}{m\pi\rho_E^2}\right)^{\frac{1}{4}}$  after  $t$  generations. Eventually, however, the bulges become comparable in length to the zone length,  $L$  (when  $2mt \geq L^2$ ). Then, it begins to move as if in one dimension, and the typical displacement is  $(wt/\rho_E L)^{\frac{1}{2}}$ . Thus, movement is fastest for a short, wide zone containing few individuals. The movement in one generation is approximately equal to the zone width divided by the square root of the total number of hybrids; the effect of sampling drift on a hybrid zone is likely to be very small.

Secondly, how does the number of loci involved in the zone affect its sensitivity to sampling drift? If disequilibrium is small, only the diagonal elements of  $R_{ij}$  will be significant, and so  $\sum_i \sum_j g_i R_{ij} g_j$  will be proportional

to the number of loci,  $n$ . However, the denominator,  $\left(\int_{-\infty}^{\infty} \sum_i g_i dx\right)^2$  is proportional to  $n^2$ , and so the rate of drift is proportional to  $1/n$ . As more independent loci become involved, the effects of drift at each locus tend to cancel, so that zone movement decreases. If linkage is tight, however, fluctuations may be correlated between loci, so that the drift rate may not depend so strongly on the number of loci.

I will now examine two special cases, in order to illustrate and confirm the above points. The equations for  $g$  can be solved for a single locus zone maintained by heterozygote disadvantage, and for a zone maintained by epistasis between two linked loci. The rate of drift can then be found, although only for weak linkage in the second case.

*Heterozygote disadvantage:* It has already been shown that if  $s(\theta) = \bar{s}\theta(1-\theta)(2\theta-1)$ , a hybrid zone can form, with  $\theta = (1+T)/2$ , where  $T = \tanh(x/w)$ , and  $w = (2m/\bar{s})^{1/2}$ . For a single locus zone,  $g = \theta'^2 = (1-T^2)^2/4w^2$  and  $r_y = (\theta(1-\theta)/2\rho_E\theta'^2)\delta(\Delta x, \Delta z, \Delta t) = w^2/(2\rho_E(1-T^2))\delta$ . Hence, using the above formulae, the rate of zone movement is:

$\langle \dot{D}^2 \rangle = 3wt/10\rho_E L$  for a short zone of length  $L$ , after  $t$  generations, and

$$\langle \dot{D}^2 \rangle = \frac{3w}{10\rho_E} \cdot \left(\frac{t}{m\pi}\right)^{\frac{1}{2}} \text{ for a long zone, } 2mt \ll L^2$$

The typical displacement is proportional to  $t^{1/2}$  for a short zone, but only to  $t^{\frac{1}{4}}$  for a long zone, since bulges in opposite directions tend to retard each other. For example, consider a long zone 500 m wide, maintained by 10 per cent selection against heterozygotes, and having a population density of one per 100 m<sup>2</sup>; after 100 generations it will have moved (at any one point) about 30 m, but after 10,000 generations it will have moved only about 100 m. If the zone were in fact only 2.5 km long, it would speed up after about 100 generations, and after 10,000 generations it would have moved about 250 m. It seems that, unless the total number of hybrids is quite small, sampling drift is of negligible importance.

*Epistasis between two linked loci:* The solution for  $g$  has already been obtained for this case; it remains to find  $R$  and  $\theta$ . If linkage is loose, only the allele frequencies need be considered; by symmetry,  $\theta_1 = \theta_2$ , and so  $\theta_1 = \theta_2 = (1+t)/2$ . Since there is no disequilibrium,  $R_{12} = 0$ . The solution is then just the same as in the previous example of heterozygote disadvantage, except that, since two loci are involved, the distance moved by the zone decreases by a factor  $\sqrt{2}$ .

If linkage is important, however, three variables are needed to describe the system.  $g$  can be found in terms of the gamete frequencies, as can  $R$ , and so the drift rate can be expressed in terms of  $\theta$ .

$$\langle \dot{D}^2 \rangle = \int_{-\infty}^{\infty} g^T \underline{R} g dx / \left( \int_{-\infty}^{\infty} \sum_i g_i dx \right)^2$$

Now,

$$g = (\theta_1'^2, \theta_1'^2, -(\epsilon/2r)\theta_3'^2)$$

(see previous section)

and, using the multinomial distribution,

$$\underline{R} = \begin{vmatrix} \theta_1(1-\theta_1)/\theta_1'^2 & \frac{1}{2}(\theta_3-\theta_1^2-(1-\theta_1)^2)/\theta_1'^2 & \frac{1}{2}(2\theta_1-1)(1-\theta_3)/\theta_1'\theta_3' \\ \cdot & \theta_1(1-\theta_1)/\theta_1'^2 & \frac{1}{2}(2\theta_1-1)(1-\theta_3)/\theta_1'\theta_3' \\ \cdot & \cdot & \theta_3(1-\theta_3)/\theta_3'^2 \end{vmatrix} \frac{1}{2\rho_E L}$$

Hence,

$$\langle \dot{D}^2 \rangle = \frac{\int_{-\infty}^{\infty} \left[ \theta_1'^2(\theta_3 - (2\theta_1 - 1)^2) - \frac{\epsilon}{r} \theta_1' \theta_3' (2\theta_1 - 1)(1 - \theta_3) + \frac{\epsilon^2}{4r^2} \theta_3'^2 \theta_3 (1 - \theta_3) \right] \frac{dx}{2\rho_E L}}{\left( \int_{-\infty}^{\infty} \left( 2\theta_1'^2 - \frac{\epsilon}{2r} \theta_3'^2 \right) dx \right)^2}$$

For most values of  $r$ , the equations for  $\theta$  are insoluble. However, the first order effect of slight linkage can be found. If  $\frac{\epsilon}{r} \ll 1$ ,

$$\theta_1 = (1 + T)/2 - \frac{\epsilon}{4r} (1 - T^2) \frac{x}{w}$$

$$\theta_3 = (1 + T^2)/2 + \frac{\epsilon}{2r} (1 - T^2) \left( 1 - T^2 - \frac{x}{w} T \right)$$

and so

$$\langle \dot{D}^2 \rangle = (3w/20\rho_E L) \left( 1 + \frac{137\epsilon}{105r} \right).$$

As linkage tightens, the zone becomes broader, and the rate of drift increases.

(b) *Variations in population structure:* We have seen that genetic drift is probably of little importance in moving the zone. However, other forces may be more important. Variations in dispersal rates and population density from year to year are often large, and will cause random movement of the zone. Treatment of these effects is rather more complicated than is that of genetic drift, since deviations may well be correlated between years, and between places. Therefore, only stochastic variation in population density will be considered.

Let  $L = \ln(\rho)$ , the logarithm of population density. Then, if  $L$  is small there is an extra term in  $\theta_i$  of  $mL'\theta_i'$ . Therefore,  $\gamma_i = mL'\forall i$ ; the perturbation of the zone by changes in population structure is the same for all loci. This leads to a considerable simplification; letting  $r_L(\Delta x, \Delta y, \Delta t)$  be the covariance function of the log population density,  $L$ , and assuming that the expected gradient is zero, one finds:

$$r_D(\Delta y, \Delta t) = \frac{\int_0^{t_1} \int_0^{t_2} \iiint_{-\infty}^{\infty} m^2 r_L(\Delta x, \Delta z, \Delta t) \exp \left( -(\Delta y - \Delta z)^2 / 2m(t_1 + t_2) \right) G'(x)G'(x + \Delta x) dx d\Delta x dz dt}{\left( \int_{-\infty}^{\infty} G(x) dx \right)^2}$$

where  $G(x) = \sum_i g_i(x)$ .

The terms involving  $\Delta y$  and  $t$  are the same as before; the movement in two dimensions will therefore be similar to that with sampling drift. Displacements will be correlated over distances of order  $\sqrt{2mt}$ , whilst short zones will move much faster than long ones. The dependence of drift rate on the type of selection maintaining the zone is, however, quite different.

The behaviour of the zone in response to varying population structure depends only on  $G(x)$ , rather than on the individual  $g_i(x)$ . In particular, if variations in density gradient are correlated over distances much greater than the zone width, so that  $r_L(\Delta x, \Delta y, \Delta t)$  is approximately constant through the zone, the rate of drift does not depend at all on the form of  $G$ . At the other extreme, when fluctuations are not correlated from place to place, the drift rate is proportional to

$$\left( \int_{-\infty}^{\infty} G'^2(x) dx \right) / \left( \int_{-\infty}^{\infty} G(x) dx \right)^2.$$

This is of the same order as  $1/w^3$ , and so decreases as the zone gets wider. It does not depend directly on the number of loci involved, so that multilocus systems will be affected by population structure in much the same way as single locus systems.

Combining the results of this section with those from the last, one finds that sampling drift will be most important for wide zones, maintained by weak selection, and involving few loci. These will be poor barriers to gene flow, and so population structure is likely to be the dominant factor for zones of biological interest.

To give a concrete example, consider a single locus hybrid zone, moving in one dimension, and maintained by heterozygote disadvantage. Suppose fluctuations in population density are correlated over distances of about  $X$ , and times of about  $T$ :

$$r_L(\Delta x, \Delta t) = v \exp(-|\Delta x|/X) \exp(-|\Delta t|/T) \quad (v = \text{variance in } L).$$

Then, the rate of drift of the zone is:

$$\begin{aligned} \langle \dot{D}^2 \rangle &= \frac{128XTm^2v}{35w^3} \quad (X \ll w) \\ &= \frac{2Tm^2v}{X^2} \quad (X \gg w). \end{aligned}$$

Thus, the most effective types of perturbation are those which persist for some time, and are correlated over distances similar to the zone width. For example, suppose the zone is 500 m wide, and is maintained by 10 per cent selection. If  $X$  is 1 km,  $T$  is 10 generations, and the variance in population

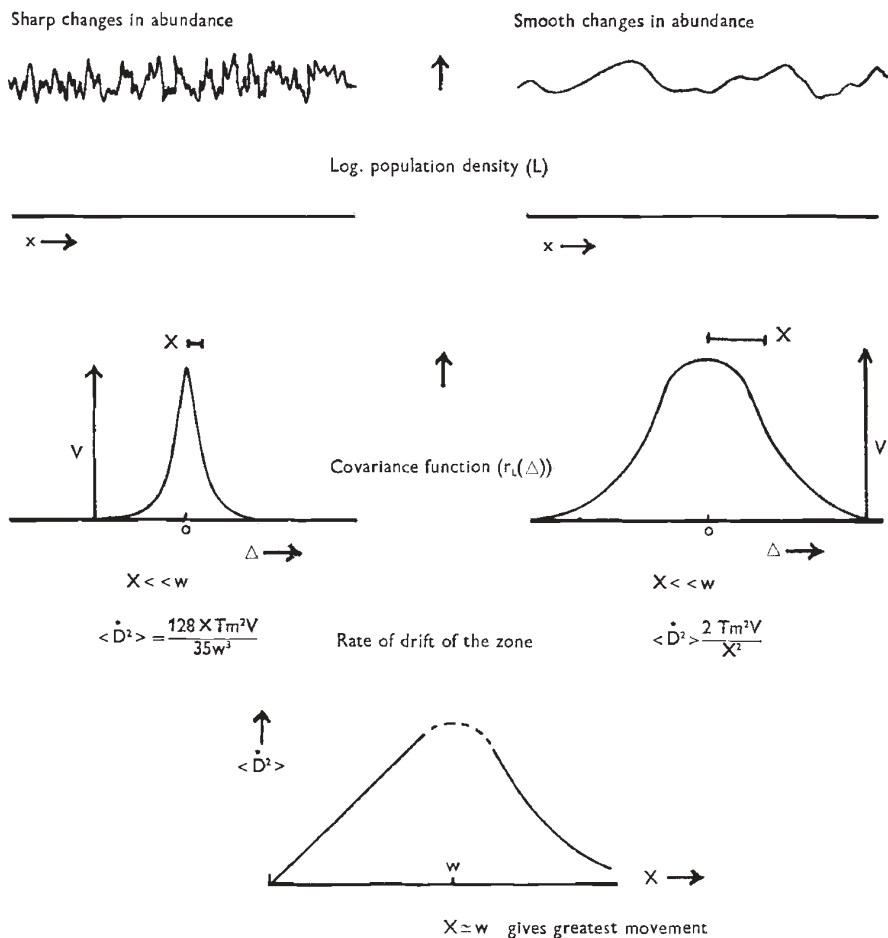


FIG. 2.—Zone movement due to fluctuations in population density.

density is 0.5, then the zone will move about 20 m per generation. This is much more than would be possible with sampling drift alone; variations in population structure are the main influence on strong hybrid zones.

(v) *Density troughs*

In the previous discussion, it was assumed that the zone was free to move in response to small perturbations. This is highly unlikely; the two forms will not have the same fitness, and so the zone will move until it is trapped in a region of, for example, low population density. We have seen that quite large selection differentials may be needed to allow it to escape from such a trough. Can random forces free it, and allow consistent movement to continue?

Here, only variations in population density will be considered, as before. Changes in dispersal rate or selection pressure will have similar effects. Suppose the zone is trapped in a shallow density trough, so that:



$$\langle L \rangle = L_0 + \frac{L''D^2}{2}, \text{ where } L'' \text{ is a small constant, } \ll \frac{1}{w^2}.$$

Then, the zone will tend to move towards  $D = 0$ :

$$\dot{D} = -mL''D + X(t), \text{ where } X(t) \text{ is the stochastic perturbation in } D.$$

Integrating this equation,

$$D(t) = \exp(-mL''t) \int_0^t \exp(mL''t')X(t')dt'.$$

Let changes in  $L$  be independent of  $D$ , and have a covariance function proportional to  $\exp(-|\Delta t|/T)$ , as in the previous example. Then, letting  $r_X(\Delta t) = k \cdot \exp(-|\Delta t|/T)$ , where  $k$  depends on  $r_L$  and  $G(x)$ , by taking the second moments of the last equation,

$$r_D(\Delta t) = \frac{k(\exp(-|\Delta t|/T) - \exp(-mL''\Delta t))}{(m^2L''^2 - 1/T^2)}$$

$$\text{var}(D) = k/(mL''(mL'' + 1/T))$$

$$\langle L \rangle = L_0 + k/2m(mL'' + 1/T).$$

The hybrid zone will not be at the point of minimum population density, but will be perturbed away from it by random forces. The expected height above the equilibrium position is, if  $mL''T \ll 1$ , just  $kT/2m$ . Since the rate of movement in the absence of constraints is  $2kT$ , this excess density ( $\langle L \rangle - L_0$ ) is  $\langle \dot{D}^2 \rangle_{FREE}/4m$ . For example, if  $\sqrt{m} = 100$  metres  $\text{gen}^{-1}$ , a zone which would normally drift 10 m in a generation will, when trapped in a shallow density trough, usually be at a point with 0.25 per cent greater density than that expected in the absence of stochastic forces.

(vi) *Probability of escape from a trough*

How long will it take for a zone to escape from a barrier? In order to answer this question, some form must be assumed for the probability distribution of density variations. The rate of escape will be calculated for normally distributed fluctuations, since this is the only tractable case; in practice, the distribution may be quite different.

For a gaussian stochastic process, the rate of crossing  $x = \pm x_0$  is:

$$\frac{2}{\pi} \sqrt{\frac{\langle \dot{x}^2 \rangle}{\langle x^2 \rangle}} \exp(-x_0^2/2\langle x^2 \rangle) \text{ (Soong, 1973).}$$

The rate of *escape* from the region  $|x| \leq x_0$  is one half this value. Applying this to a hybrid zone trapped in a shallow density trough of width  $X$ , and depth  $L$ , the rate of escape is approximately (using previous section):

$$\frac{2}{\pi X} \sqrt{\frac{2m\delta L}{T}} \exp(-\delta L/2\phi),$$

where  $\phi$  is the expected "height" above equilibrium. ( $\phi = \langle L \rangle - L_0 = \langle \dot{D}^2 \rangle_{FREE}/(4m)$ ).

So, as long as the zone is narrow, the rate of escape only depends on the nature of the zone through the ratio  $\delta L/\phi$ . If, as is likely, the zone only

moves at a small fraction of the dispersal rate, quite small density troughs can form very strong barriers. For example, suppose  $m^{\frac{1}{2}} = 100 \text{ m.gen}^{-\frac{1}{2}}$ ,  $T = 10$  generations, and the free rate of drift is 10 m per generation. If such a zone is trapped in a region 1 km across where density is reduced by 20 per cent, it will take about 4000 generations to escape in either direction. It seems unlikely that a narrow zone could move far unless the species were distributed very evenly.

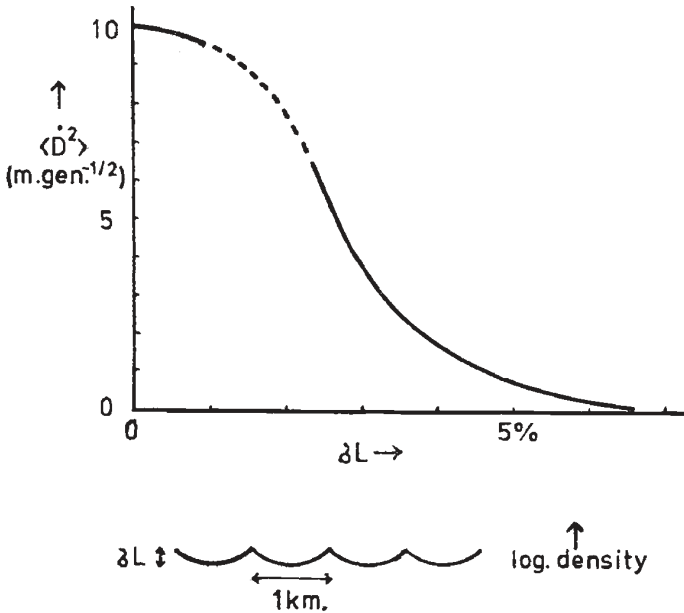


FIG. 3.—Rate of drift of a zone across a series of density troughs; suppose a zone moves 10 m in a generation because of gaussian random forces, when it is in a uniform environment. The random forces have a persistence,  $T$ , of 10 generations. If the population density varies gradually in a series of troughs, 1 km wide and  $\delta L$  deep (see diagram), then the rate of drift will be reduced from the free value of  $10 \text{ m.gen.}^{-\frac{1}{2}}$

### 5. CONCLUSION

We have seen that hybrid zones can move at substantial speeds if there is a selective imbalance between the two forms involved. However, zones are also very sensitive to population structure, and can easily be trapped by natural barriers. Random variations in population structure, and, to a far lesser extent, sampling drift, will tend to free them, but unless the organism leads a very uncertain existence, it seems unlikely that a zone, once formed, will move far.

If hybrid zones are indeed so static, it is hard to see how they could ever become established. On the other hand, once a zone encloses a large area, it should persist for a considerable time, and one might expect such sharp clines to be found quite often. These questions, and their implications for ideas on divergence and speciation, will be examined in a later paper (Barton and Hewitt, 1980).

*Acknowledgments.*—I would like to thank Dr G. M. Hewitt for helpful discussion, and for his comments on the manuscript. This work was supported by the Natural Environment Research Council.

## 6. REFERENCES

- BARTON, N. H., AND HEWITT, G. M. 1980. Hybrid zones and speciation. In *Essays on Evolution and Speciation in Honour of M. J. D. White*, eds. W. R. Atchley, D. S. Woodruff. Cambridge Univ. Press (in press).
- BAZYKIN, A. D. 1969. A hypothetical mechanism of speciation. *Evolution*, 23, 685-687.
- BAZYKIN, A. D. 1972. Disadvantage of heterozygotes in a population within a continuous area. *Genetika*, 8, 162-167.
- BAZYKIN, A. D. 1973. Population genetic analyses of disruptive and stabilizing selection. II. Systems of adjacent populations and populations within a continuous area. *Genetika*, 9, 156-166.
- CROW, J. F., AND KIMURA, M. 1970. *An Introduction to Population Genetics Theory*. Harper and Row.
- FISHER, R. A. 1937. The wave of advance of an advantageous gene. *Ann. Eugen.*, 7, 355-369.
- HADELER, K. P. 1976. Travelling population fronts. In *Population Genetics and Ecology*, eds. S. Karlin, E. Nevo. Academic Press.
- KARLIN, S., AND MCGREGOR, J. 1972. Application of the method of small parameters to multi-niche population genetic models. *Theor. Pop. Biol.*, 3, 186-209.
- KEY, K. 1974. Speciation in the Australian Morabine grasshoppers. In *Genetic Mechanisms of Speciation in Insects*, ed. M. J. D. White. Aust. and New Zealand Book Co.
- MOORE, W. S. 1977. An evaluation of narrow hybrid zones in vertebrates. *Quart. Rev. Biol.*, 52, 263-277.
- NAGYLAKI, T. 1975. Conditions for the existence of clines. *Genetics*, 80, 595-615.
- NAGYLAKI, T. 1978. Random genetic drift in a cline. *Proc. Natl. Acad. Sci., U.S.A.*, 75, 423-426.
- SOONG, T. T. 1973. *Random Differential Equations*. Academic Press.
- STOKES, A. N. 1976. On two types of moving front in quasilinear diffusion. *Math. Biosci.*, 31, 307-315.
- WHITE, M. J. D. 1968. Models of speciation. *Science*, 159, 1065-1070.
- WHITE, M. J. D. 1978. *Modes of Speciation*. W. H. Freeman Ltd.