Heredity (1979), **43** (3), 333-339

GENE FLOW PAST A CLINE

N. H. BARTON

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

Received 14.iii.79

SUMMARY

The effect of a cline as a barrier to gene flow at closely linked, weakly selected loci is investigated using a continuous diffusion model. It is shown that a linked cline induces a step in the frequency of a neutral allele, which is proportional to the gradient in neutral allele frequency and to the inverse of the recombination rate. A barrier to dispersal or a local region of low abundance has a similar effect (Nagylaki, 1976). The cline will block the flow of neutral alleles over a region of chromosome roughly $2\sqrt{s/t}$ map units long. However, a slightly advantageous allele will be little affected, and must be very tightly linked to be delayed for long.

1. INTRODUCTION

IT has often been suggested that clines can act as important barriers to gene flow (*e.g.*, Key, 1974). A geographic pattern at one locus will tend to be imposed on other linked loci, and so the spread of linked genes past a cline will be impeded. If geographic variation is common, this effect may be important in determining both the rate of spread of new alleles, and the pattern of variation at equilibrium. It has great bearing on some theories of speciation; if, for example, clines set up between two reuniting forms can reduce gene flow enough at other loci, divergence to give fully isolated species could occur.

Suppose there is a cline at some locus, maintained at equilibrium by natural selection in the face of dispersal; alleles at any locus closely linked to this cline, and themselves more or less neutral, will tend to show a similar pattern of variation. Any gradient in these alleles' frequency, due to past history or to weak selection, will generate linkage disequilibrium in the cline, because two dissimilar populations are being mixed (Li and Nei, 1974). It will be shown that this disequilibrium, interacting with selection at the primary cline, will in turn produce a step in allele frequency, and gene flow will be reduced. Furthermore, the size of this step, at equilibrium, will be proportional to the gradient in allele frequency outside the region of interest. This is just the same effect as that produced by a barrier to dispersal (Nagylaki, 1976).

Such a barrier is characterised by its strength and by its degree of asymmetry. The strength is determined by the ratio between the step size and the external gradient. Since gene flow is proportional to the gradient in allele frequency, the amount of gene flow will be proportional to the step size and inversely proportional to the strength. One can either view the external gradient as producing the step, or, conversely, view the step as allowing a certain amount of gene flow. The asymmetry of gene flow will be determined by the difference in gradient on either side, relative to the average gradient. A barrier may allow more gene flow, and hence introgression, in one direction than in the other.

2. The basic model

To find the strength of the barrier, the continuum model of dispersal and selection will be used (Nagylaki, 1975). This is reasonable if gene frequencies change little between demes and between generations. Only the one-dimensional case will be considered; the two-dimensional case will give similar results if the cline runs in a straight line.

Let x be the geographic distance, and let $p(x)$, $q(x)$ be the frequencies of the two alleles at the primary cline, which is maintained by selection $s(x)$, such that $p(x)$ goes from 0 to 1 as x increases through zero. If the dispersal rate (defined as m, the variance in offspring position produced per generation) is constant, then:

$$\dot{p} = \frac{m}{2} p'' + s(x)pq \quad \left(\dot{f} \equiv \frac{\partial f}{\partial t} \quad f' \equiv \frac{\partial f}{\partial x} \text{ etc.} \right).$$

The cline will have width $W = \sqrt{m/\bar{s}}$, where $\bar{s}$ is some measure of the average selection acting. Let $u(x,t)$, $v(x,t)$ be the frequencies of alleles at the neutral locus, and let $D(x,t)$ be the disequilibrium between the two loci. If the recombination rate is r, it can be shown that:

$$\dot{u} = \frac{m}{2} u'' + s(x)D \tag{A}$$

$$\dot{D} = -rD + \frac{m}{2} D'' - s(x)D(p-q) + mp'u' \tag{B}$$

These two equations will be solved for the two extreme cases of tight and loose linkage; only the equilibrium solution will be found, but since the neutral allele frequency will usually only change slowly, this results in little error.

(i) *Loose linkage*

If $r \gg s$, there will be little disequilibrium, and the cline will have little effect on the linked gene. Since D is small,

$$D \simeq \frac{mp'u'}{r} \; (\ll 1)$$

(from (B), since all terms in $\dot{D}$ or D'' are of order $\bar{s}D$).
Hence, at equilibrium,

$$0 = u'' - u' \left(\frac{mp'p''}{rpq} \right).$$

This can be integrated to give:

$$u(x) = u'_- \int_{-\infty}^{x} \left(\int_{-\infty}^{y} \left(\frac{mp'p''}{rpq} \right) dz + 1 \right) dy.$$

So,

$$(u'_+ - u'_-) = \frac{m}{r} u'_- \int_{-\infty}^{\infty} \left(\frac{p'p''}{pq} \right) dx$$

and

$$(u_+ - u_-) = \frac{m}{r} u'_- \int_{-\infty}^{\infty} \left(\int_{-\infty}^{y} \left(\frac{p'p''}{pq} \right) dz - \tfrac{1}{2} \int_{-\infty}^{\infty} \left(\frac{p'p''}{pq} \right) dz \right) dy \sim \frac{mu'_-}{rW}$$

(where u'_+, u'_- are the gradients immediately outside the perturbed region).

The smooth external gradient has produced a small step in allele frequency of order $\frac{mu'_-}{rW}$. If the cline is asymmetric, so that $\int_{-\infty}^{\infty} \frac{p'p''}{pq} dx \neq 0$, there will be a difference in the gradients, and hence in the gene flow, in either direction.

(ii) *Tight linkage*

If $r \ll s$, the cline will have a strong effect on the linked gene. Disequilibrium will be close to its maximum value, and gene flow will be low. At equilibrium, from (A) and (B):

$$rD = \frac{m}{2} D'' + mp'u' - sD(p-q) \tag{1}$$

$$0 = \frac{m}{2} u'' + sD \tag{2}$$

$$\therefore rD = \frac{m}{2} D'' + \frac{m}{2} (u'(p-q))' \qquad ((1)+(2)(p-q))$$

Integrating over the region of interest (from $-A$ to $+A$, say),

$$r \int_{-A}^{A} D dx = \frac{m}{2} (u'_+ + u'_-).$$

Since linkage is tight, D will be close to, and less than, $pq(u_+ - u_-)$, its maximum value. So,

$$(u_+ - u_-) \leqq \frac{m}{2r} (u'_+ + u'_-) \left(\int_{-\infty}^{\infty} pq dx \right)^{-1} \sim \frac{m}{2rW} (u'_+ + u'_-).$$

Thus, the strength of the barrier (the ratio of step size to external gradient) is given by the same formula as for weak linkage, except that the width of the cline is defined by a different integral. The *difference* in the gradient, and hence the asymmetry of introgression, may be quite large.

3. The effect of a barrier

We will now apply the formulae just derived to find out how long it takes for an allele, which was initially confined to one side of the cline, to spread past it and become established in the other genome; the following solutions can also be used to find the effect of a barrier to dispersal.

(i) *A neutral allele*

Away from the cline, the allele diffuses freely;

$$\dot{u} = \frac{m}{2} u'' \tag{3}$$

At the cline ($x = 0$, say), there is a sudden jump in frequency proportional to the gradient $u'(0, t)$. Consider $x < 0$, and assume a symmetrical barrier, so that $u(x) = 1 - u(-x)$. Then, we have the boundary condition:

$$u(0, t) = \tfrac{1}{2}\left(1 - \frac{mu'(0, t)}{rW}\right)$$

$$\left(W = \int_{-\infty}^{\infty} pqdx \text{ for tight linkage,} \qquad W = \left(\int_{-\infty}^{\infty}\int_{-\infty}^{z} \frac{p'p''}{pq}\, dydz\right)^{-1} \text{for loose linkage}\right).$$

Taking Fourier transforms, $u(x) \to \tilde{u}(w)$, (3) becomes:

$$\dot{\tilde{u}} = -\frac{m}{2}\tilde{u}w^2 \quad \therefore \tilde{u} = f(w)\exp\left(-\frac{mtw^2}{2}\right); f(w) \text{ is arbitrary.}$$

The boundary condition becomes:

$$1 = \int_{-\infty}^{\infty} \frac{f(w)}{\sqrt{2\pi}} \exp\left(-\frac{mtw^2}{2}\right)\left(2 + \frac{miw}{rW}\right) dw \quad (i = \sqrt{-1}).$$

The only $f(w)$ for which the above is true $\forall t$ is:

$$f(w) = \frac{-i}{\sqrt{2\pi}}\left(\frac{1}{w} - \frac{1}{(w - 2irW/m)}\right).$$

Taking the inverse Fourier transform, and changing to dimensionless coordinates, (X,θ), we find:

$$u(X, \theta) = \left(P\left(\frac{X}{\sqrt{\theta}}\right) - e^{2(\theta - X)}P\left(\frac{X - 2\theta}{\sqrt{\theta}}\right)\right) \quad (X < 0)$$

$$X = x\left(\frac{rW}{m}\right),\ \theta = \left(\frac{r^2W^2}{m}\right)t,\ P(x) = \int_{-\infty}^{x} \frac{e^{-\varepsilon^2/2}}{\sqrt{2\pi}}\, d\varepsilon.$$

It can be seen from fig. 1 that most of the differentiation is lost by $\theta = 1$, that is, by $t = \dfrac{m}{r^2W^2} = \dfrac{\bar{s}}{r^2}$. There will be a small region of differentiated chromosome around the cline locus, whose size $\left(\simeq 2\sqrt{\dfrac{\bar{s}}{t}}\right)$ will decrease with time.

(Note that since the original formula holds only for slowly changing u, this solution is not valid in the initial stages of the flow, when $\theta < (r/\bar{s})^2$. This restriction is only important for weak barriers, which have little effect anyway.)

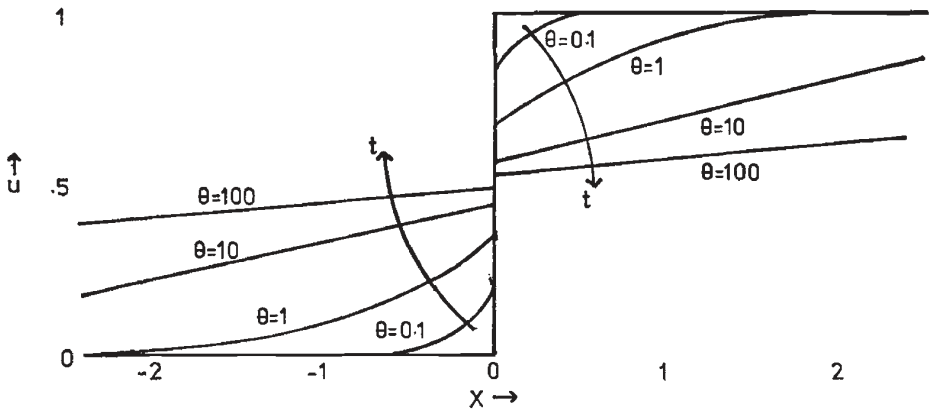

FIG. 1.—This graph shows the shape of a neutral cline, whose decay is being impeded by a strong cline at a linked locus. Time and distance are measured by the dimensionless coordinates, X, θ. See text.

(ii) *An advantageous allele*

Suppose that an advantageous mutant occurs on one side of a cline; if it is tightly linked to the cline locus, and if the selection maintaining the cline is much stronger than the mutant's advantage, there will be a considerable delay in its progress. It will gradually leak through, until enough accumulate on the other side for their intrinsic rate of increase due to selection to become greater than the leakage rate. The new allele will then become fixed in a time which depends on the frequency at which leakage becomes negligible (fig. 2). The total delay will be approximately equal to the time

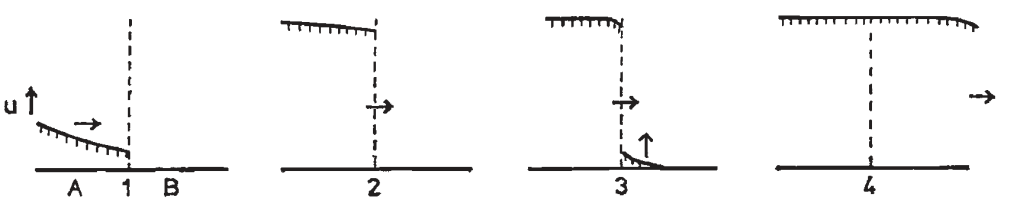

FIG. 2.—The flow of an advantageous allele past a linked cline; see text for explanation.

to go from (2) to (4); for a strong barrier, the time from (2) to (3) will be much less than that from (3) to (4).

Let the allele have selective advantage $S \ll (m/W^2) = \bar{s}$. On side B, when leakage predominates, $u(0) = \frac{1}{2} - e^{2\theta}P(-2\sqrt{\theta})$. For small θ, this is approximately $\sqrt{\frac{2\theta}{\pi}}$. The rate of change of allele frequency due to leakage (at $x = 0$) is therefore of order $u/2t$, whilst that due to selection is Su. Hence, leakage becomes negligible when $t > 1/2S$, however strong the barrier. At this time, $u(0) \simeq \frac{rW\sqrt{2}}{\sqrt{\pi m S}}$. The time to fixation after selection takes over is about $-\frac{1}{S}\ln(u_0)$ (Fisher, 1930), so that the delay is

$$\frac{1}{2S}\ln\left(\frac{\pi S}{2r^2}\frac{m}{W^2}\right) = \frac{1}{2S}\ln\left(\frac{\pi S\bar{s}}{2r^2}\right).$$

This is much greater than the time for leakage to occur ($1/2S$), unless linkage is relatively tight ($r^2 \ll \pi\bar{s}S/2$); the delay is then much less than the time required for the initial leakage. If the selective advantage S is weak, $\left(S \ll \frac{2r^2}{\pi\bar{s}}\right)$, however, the allele behaves as though it were neutral, and the delay is primarily due to the initial leakage. For example, if $r = 0{\cdot}1$ per cent

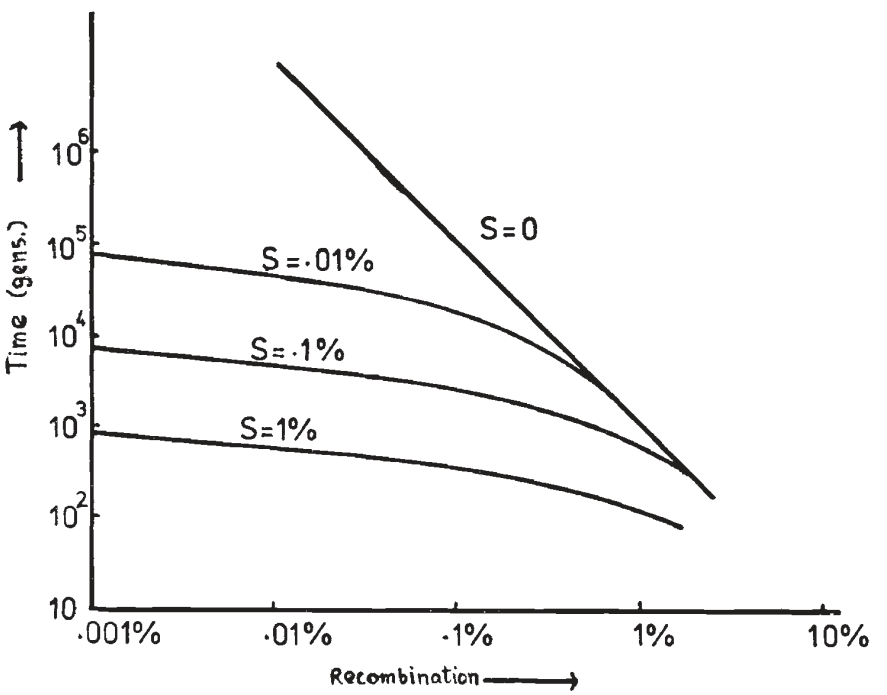

FIG. 3.—This graph shows the time taken for an allele with advantage S to get past a linked cline, maintained by selection $_2$ = 10 per ecnt.

and $\bar{s} = 10$ per cent, any alleles with $S > 0{\cdot}001$ per cent are significantly non-neutral and get past the barrier much faster than do neutral alleles. Fig. 3 shows the time taken for alleles to flow past a cline maintained by 10 per cent selection. It can be seen that for the delay to be, say, 1000 generations, linkage must be tighter than 0·5 per cent, and the advantage must be less than 0·5 per cent. Note that these results do not apply at low population densities, when only a few genes leak across every generation ($\rho Wr \sim 1$); drift is then important (see Slatkin, 1976).

4. DISCUSSION

We have seen that a cline can be a significant barrier to gene flow at closely linked loci; the amount of the linkage map affected is proportional to the square root of the selection maintaining the cline. However, though

neutral alleles can be considerably restricted by such a barrier, even mildly advantageous alleles can easily penetrate it.

Only the effect of a single cline has been examined here. However, a strong barrier to flow at many loci can be formed only if there are many coincident clines. The cumulative effect of a number of clines is therefore of interest, but it is a much harder problem. One must consider not only the effect of first-order linkage disequilibria, but also that of the higher order disequilibria. The strength of the single locus barrier is inversely proportional to the recombination, and so one might expect the cumulative effect of distant loci to be important $\left(\int_x^\infty \frac{1}{r}\,dr = \infty\right)$. However, interactions between loci will alter this long-range effect; the multi-locus case needs to be considered in more detail.

Although only stationary clines have been considered here, advancing alleles will have a similar effect. For a wavefront of given width, the strength of the barrier will be about the same, but there will be a marked asymmetry of gene flow, into the advancing type.

How will evolution be affected? Firstly, the spread of new alleles will be slowed down if there is geographic variation at many loci. In effect, the dispersal rate will be reduced. This will be most important for nearly neutral alleles, since moderately advantageous genes are affected little by barriers to gene flow. Local heterozygosity will be reduced, and divergence between regions increased. Secondly, clines will tend to be attracted to each other, if they overlap. The effect will be slight for clines whose position is determined by the external environment, but clines maintained by internal interactions (hybrid zones) will be more mobile, and may therefore clump together (see Slatkin, 1975). Thirdly, and perhaps most important, a very strong barrier to gene flow may be formed if there is a large number of coincident clines, as a result, for example, of the reunion of two diverged isolates. If the total selection involved is as large or larger than the total recombination rate, gene flow at most loci will be considerably reduced, and so divergence can continue. However, advantageous genes will still be able to get through such a barrier, unless selection is so strong that almost all hybrids between the two types die or are sterile. Thus, the two putative species must already be quite different if isolation is to be effective when they meet again.

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

5. References

FISHER, R. A. 1930. *The Genetical Theory of Natural Selection.* Oxford University Press.

KEY, K. H. L. 1974. Speciation in the Australian morabine grasshoppers—taxonomy and ecology. In *Genetic Mechanisms of Speciation in Insects*, ed. M. J. D. White. Austr. and New Zealand Book Co.

LI, W. H., AND NEI, M. 1974. Stable linkage disequilibrium without epistasis in subdivided populations. *Theor. Pop. Biol., 6,* 173-183.

NAGYLAKI, T. 1975. Conditions for the existence of clines. *Genetics, 80,* 595-615.

NAGYLAKI, T. 1976. Clines with variable migration. *Genetics, 83,* 867-886.

SLATKIN, M. 1975. Gene flow and selection in a two locus system. *Genetics, 75,* 733-756.

SLATKIN, M. 1976. The rate of spread of an advantageous allele in a subdivided population. In *Population Genetics and Ecology*, ed. S. Karlin and E. Nevo. Academic Press, New York.